THÈSE

présentée pour l'obtention du titre de

DOCTEUR DE L'ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES

Spécialité : Mathématiques appliquées

par

Antoine GLORIA

Sujet : *Modélisation et méthodes numériques multi-échelles en élasticité non linéaire*

(Multiscale modeling and numerical methods in nonlinear elasticity)

Soutenance le 20 juin 2007 devant le jury composé de :

Président :	Hervé Le Dret
Rapporteurs :	Yalchin Efendiev Patrick Le Tallec
Examinateurs :	Grégoire Allaire Gilles Francfort Karam Sab
Directeurs de thèse :	Claude Le Bris Jean-Frédéric Gerbeau

A mes grands-parents

Remerciements

Quand on commence une thèse, on a du temps, mais on ne sait pas forcément bien s'y prendre. Quand arrive la fin, on a des idées, mais on manque de temps. Mon passage de jeune doctorant à jeune docteur s'est fait doucement mais sûrement, encouragé et assisté d'autres doctorants et docteurs de tous horizons. Je voudrais remercier tout d'abord Claude Le Bris. C'est au travers de ses cours que j'ai découvert l'attrait des mathématiques appliquées et c'est naturellement que j'ai choisi de faire une thèse sous sa direction au CERMICS. Son travail et sa façon d'appréhender les problèmes ont été une grande source de motivation et d'inspiration. J'ai également passé une grande partie de ces trois années à l'INRIA, encadré par Jean-Frédéric Gerbeau, qui m'a laissé une grande liberté dans le travail et dont les conseils, de toute nature, sont précieux. Je l'en remercie.

De nombreuses personnes m'ont aidé durant ces trois années. J'ai une pensée particulière pour Marina Vidrascu, dont l'énergie, la puissance de travail, la bonne humeur et la disponibilité, tant du point de vue professionnel que personnel, sont extraordinaires. Je lui dois beaucoup. Elle a toute ma gratitude et mon affection. Merci également à Miguel Fernandez sans qui une partie de ma thèse n'existerait pas. Cette thèse a aussi été pour moi l'occasion de voyager. J'ai eu la chance de passer trois à UCLA. Grâce au réseau européen MULTIMAT et au financement de l'Union Européenne, j'ai pu travailler trois mois avec Andrea Braides à l'université de Rome Tor Vergata. È molto piaciavole di lavorare con Andrea. È una persona molta calorosa che dà fiducia a se stessi. Lo ringrazio per questi tre mesi e per il suo sostegno. Durant ce séjour, j'ai aussi eu l'opportunité de rencontrer deux anciens élèves d'Andrea, Roberto Alicandro et Marco Cicalese, avec lesquels une longue collaboration a commencé. È semplicemente un piacere di andare a Roma o Napoli per lavorare, discutere di tutto e giocare con Luca o il piccolo Andrea ! Je voudrais aussi remercier le centre Ennio de Giorgi qui m'a accueilli un mois à Pise lors d'un trimestre thématique.

Merci à Hervé Le Dret d'avoir bien voulu présider mon jury de thèse. Patrick Le Tallec et Yalchin Efendiev m'ont fait l'amitié de rapporter sur mes travaux de thèse. Je les remercie pour le temps qu'ils y ont consacré, pour leurs conseils et leurs encouragements. Leurs travaux sont la base de cette thèse. Je suis aussi très heureux que Grégoire Allaire, Gilles Francfort et Karam Sab aient accepté de faire partie du jury. C'est dans les cours de Grégoire Allaire que je me suis initié à l'homogénéisation et dans ceux de Gilles Francfort que j'ai découvert le calcul des variations 'appliqué'. Grazie mille a Pippo Geymonat per la sua presenza alla mia tesi, è un onore.

Le CERMICS et le bâtiment 16 de l'INRIA sont des endroits propices à la recherche. Je remercie Bernard Lapeyre et Serge Piperno pour leur accueil et leurs conseils. Merci à l'équipe de dynamique moléculaire, Eric Cancès, Gabriel Turinici, Tony Lelièvre, Frédéric Legoll, Mathieu Lewin, Amélie Deleurence, Mathias Rousset, et tous les autres, d'avoir fait une place à l'étranger scientifique ! Les discussions sur l'homogénéisation avec Sébastien Boyaval étaient plus que les bienvenues à cet égard ! Alexandre Ern m'a confié une partie du cours de calcul scientifique de l'ENPC, je le remercie de sa confiance renouvelée. Merci également à Sylvie Berte, Khadija Elouali et Martine Ouhanna pour le soutien administratif et logistique. Le CERMICS regorge de personnalités scientifiques et de doctorants sympathiques, merci à tous. Gabriel a presque réussi à me faire tenir mes bonnes résolutions de courrir, manger équilibré... maintenant que j'ai un pilier à Bonn, et un roc à Gand, je n'ai plus d'excuse pour le trépied ; et merci Hermann Hesse pour sa révélation à Muscle Beach ! Le bâtiment 16 a un palmarès qui ferait pâlir bien des locaux d'autres laboratoires ! J'ai eu le plaisir de partager l'ancien bureau de Jacques-Louis Lions avec Nuno, Najib et Matteo... ambiance méditerranéenne à Rocquencourt ! Sans parler des pauses café, précédées ou suivies de discussions

sur les coques, le cinéma, l'assimilation de données, Desperate Housewives, la triangulation de Delaunay, l'iPod, Modulef, l'iMac, FreeFEM, l'iPhone, LifeV... Merci à Dominique, Philippe, Sonia, Grégoire, Jacques, Mathieu, Eric, Paul-Louis, Jean-François, Benoît, Marie-Odile, Marie, Marc, Laurent, Elsie, Irène, Vincent, et tous les autres. Last but not least, parce qu'il faut ce qu'il faut et que quand on a dépassé les bornes, il n'y a plus de limites... merci à Céline, Iria et Marina pour les jeudis midi !

Enfin, et ce seront mes derniers mots, merci à mes amis, ma famille et Matthias d'avoir supporté bien plus que mes intempestives envolées mathématiques au ciné, à table, à la plage, dans la voiture ou au téléphone ! Pourvu que ça dure !

Titre : Modélisation et méthodes numériques multiéchelles en élasticité non linéaire.

Résumé : Ce travail porte principalement sur l'étude mathématique de méthodes numériques pour l'homogénéisation de fonctionnelles intégrales utilisées en élasticité non linéaire. Ces méthodes couplent, au niveau mésoscopique, un matériau hyperélastique hétérogène ou un réseau de liens en interaction, avec, au niveau macroscopique, un modèle d'élasticité non linéaire. La loi de constitution macroscopique est obtenue par la résolution de problèmes mésoscopiques, continus ou discrets. Aux chapitres 1, 2 et 3 on introduit les modèles mécaniques et les outils mathématiques et numériques utilisés par la suite. Aux chapitres 5, 6 et 7, on présente une méthode directe de résolution numérique du comportement homogénéisé d'un matériau composite périodique en grandes déformations et un cadre général pour l'analyse des méthodes d'homogénéisation numérique. On démontre notamment la convergence de méthodes numériques classiques sous des hypothèses générales ainsi qu'un résultat de correcteur numérique. On étend enfin les résultats au couplage avec des méthodes de sur-échantillonnage. Aux chapitres 8, 9 et 10, nous considérons une modélisation mésoscopique par un système discret. Nous étudions d'abord un problème de G-fermeture pour un réseau de résistances. Au chapitre suivant nous démontrons un résultat de représentation intégrale pour l'énergie d'un système de spins en interaction. Enfin, nous dérivons un modèle hyperélastique continu à partir d'un réseau stochastique de points en interaction, et l'appliquons pour démontrer la convergence de modèles discrets développés en mécanique. Dans une dernière partie, chapitre 11, nous présentons une nouvelle méthode numérique pour résoudre des problèmes d'interaction fluide structure, où la structure est décrite par une coque tridimensionnelle.

Mots clés : homogénéisation numérique, élasticité non linéaire, méthodes variationnelles, équations aux dérivées partielles, Γ-convergence, passage discret continu, interaction fluide structure.

Title : Multiscale modeling and numerical methods for nonlinear elasticity.

Abstract : The most important part of this work deals with the mathematical analysis of numerical methods for the homogenization of multiple integrals widely used in nonlinear elasticity. These methods couple, at the mesoscopic scale, a heterogeneous hyperelastic material or a network of interacting bonds with, at the macroscopic scale, a nonlinear elasticity model. The macroscopic constitutive law is obtained by solving mesoscopic problems, either continuous or discrete. In chapters 1, 2, and 3, we introduce the mechanical models and mathematical tools we use in the sequel. In chapters 5, 6, and 7, we present a direct method for the numerical solution of the homogenized behavior of a periodic composite material at finite strains, and a general framework to study numerical homogenization methods. We prove the convergence of such methods within general hypotheses and provide a numerical corrector convergence result. We also extend the analysis to cover the cases of oversampling and windowing. In chapters 8, 9, and 10, we consider a mesoscopic model based on discrete systems of bonds. We first study a G-closure problem for a network of conductances. In the next chapter, we prove an integral representation result for a system of interacting spins. We then address the rigorous derivation of a continuous hyperelastic model starting from a stochastic network of interacting points. We apply this result to prove the convergence of discrete models for rubber developed in mechanics. In the last chapter, we introduce a new solution method for fluid structure interaction problems with three dimensional shell elements to describe the structure.

Key words : numerical homogenization, nonlinear elasticity, variational methods, partial differential equations, Γ-convergence, discrete to continuum, fluid structure interaction.

Mathematical subject classification (2000) : 35J2, 35J60, 49J45, 49J55, 65N12, 65N30, 70C20, 70G75, 74A40, 74B20, 74E30, 74F10, 74G15, 74G65, 74Q05, 74Q15, 74Q20, 82B20, 82B21, 82D60.

Table des matières

Partie I Introduction

1 Présentation des problématiques et des modèles 7
1.1 Motivation (*in English below*) 7
1.2 Quelques problématiques de la modélisation multiéchelle (*in English below*) 8
1.2.1 Dérivation mathématique de modèles effectifs 8
1.2.2 Couplage discret-continu 8
1.2.3 Couplage continu-continu 9
1.2.4 Objectifs de la thèse 9
1.3 Modélisation en mécanique des milieux continus 13
1.3.1 Energie, déformation et contraintes 13
1.3.2 Equations de l'élastostatique et de l'élastodynamique tridimensionnelle 16

2 Analyse mathématique et numérique des équations de l'élastostatique 19
2.1 Analyse mathématique 19
2.1.1 Elasticité linéaire 19
2.1.2 Application du théorème des fonctions implicites 20
2.1.3 Méthode directe du calcul des variations 21
2.1.4 Quasiconvexité et semi-continuité inférieure des fonctionnelles intégrales . . . 22
2.1.5 Polyconvexité et théorie d'existence en élasticité non linéaire 23
2.1.6 Γ-convergence, application aux fonctionnelles intégrales et à la dérivation de modèles 24
2.2 Méthodes numériques en élasticité 26
2.2.1 Résultat d'approximation et estimation d'erreur 26
2.2.2 Méthode des éléments finis en élasticité 27
2.2.3 Méthodes de décomposition de domaine 29
2.2.4 Discrétisation temporelle pour l'élastodynamique 31

3 Homogénéisation - approches théoriques et numériques 33
3.1 Théorie de l'homogénéisation 33
3.1.1 Développement formel à échelles multiples 33
3.1.2 Homogénéisation périodique et fonctions oscillantes de Tartar 35
3.1.3 Homogénéisation périodique par convergence à deux échelles 36
3.1.4 Homogénéisation des opérateurs elliptiques linéaires par H-convergence . . . 38
3.1.5 Homogénéisation des intégrales multiples par Γ-convergence 39
3.1.6 Principe de localisation et problèmes de G-fermeture 42
3.2 Méthodes numériques pour l'homogénéisation et homogénéisation numérique 42
3.2.1 Eléments finis mutiéchelles dans le cas linéaire 43
3.2.2 Approximation de l'opérateur homogénéisé 44

4 Contributions de la thèse (*in English below*) . . . 47
4.1 Méthodes numériques en homogénéisation . . . 47
4.1.1 Homogénéisation périodique en élasticité non linéaire . . . 47
4.1.2 Cadre général pour l'analyse des méthodes d'homogénéisation numérique . . . 48
4.2 Modélisation multiéchelle . . . 50
4.2.1 Sur un problème de G-fermeture pour un passage discret-continu . . . 50
4.2.2 Passage discret-continu pour des énergies d'interaction de spin . . . 51
4.2.3 Dérivation variationnelle d'une énergie caoutchoutique à partir d'un modèle discret . . . 52
4.2.4 Application aux problématiques de la mécanique . . . 52
4.3 Méthodes partitionnées en interaction fluide-structure . . . 53

Part II Numerical homogenization of elliptic equations

5 A direct approach to numerical homogenization in finite elasticity . . . 63
5.1 Physical motivation . . . 63
5.2 A quick review of periodic homogenization theory . . . 64
5.2.1 Convexity and minimization problems . . . 64
5.2.2 Basic homogenization result . . . 65
5.2.3 Homogenization for connected media . . . 66
5.2.4 Homogenization of elliptic operators in divergence form . . . 67
5.2.5 Some open issues . . . 68
5.3 Approximation result for the standard homogenization problem . . . 69
5.3.1 Approximation theory for standard energy densities . . . 70
5.3.2 Approximation result for a homogenized energy density . . . 73
5.3.3 Error estimates in the convex case . . . 76
5.4 Numerical method . . . 82
5.4.1 Presentation . . . 82
5.4.2 Computation of the stress tensor and the stiffness matrix . . . 83
5.4.3 Implementation of the algorithm in a nonlinear elasticity software . . . 86
5.5 Alternative method: multiscale finite elements (MsFEM) . . . 87
5.5.1 Description of the method . . . 87
5.5.2 Comparison of the two methods in the periodic setting . . . 87
5.6 Numerical tests . . . 88
5.6.1 The convex case . . . 89
5.6.2 Buckling of the cell-problem in the standard case . . . 89
5.6.3 Shear band instabilities . . . 91
5.6.4 Tests on a wider class of energies . . . 92

6 Analytical framework for numerical homogenization . . . 95
6.1 Setting of the problem and statement of the main results . . . 95
6.1.1 Homogenization of convex energy densities . . . 97
6.1.2 Main results . . . 98
6.1.3 Extension to the quasiconvex case . . . 102
6.2 Proof of the main results . . . 105
6.2.1 The one-dimensional linear case . . . 105
6.2.2 Proof of Theorem 40 . . . 106
6.2.3 Proof of Theorem 42 . . . 109
6.2.4 Proof of Theorem 41 . . . 113
6.2.5 Proof of Proposition 5 . . . 117
6.2.6 Proof of Proposition 6 . . . 121
6.3 Relation to some existing numerical approaches . . . 123
6.3.1 HMM . . . 124

6.3.2 MsFEM . . . 126
6.3.3 Summary of some numerical analysis results . . . 130
6.4 Conclusion . . . 130
6.5 Appendix . . . 131
6.5.1 $\boldsymbol{\Gamma}$-limit and boundary conditions . . . 131
6.5.2 Proof of Lemma 6.7 . . . 131
6.5.3 Proof of Lemma 6.12 . . . 132
6.5.4 Proof of Lemma 6.13 . . . 134
6.5.5 Sketch of proof of Proposition 6 . . . 135

7 Numerical homogenization with oversampling . . . 137
7.1 Introduction . . . 137
7.2 Numerical homogenization methods . . . 137
7.2.1 Main notations . . . 138
7.2.2 Minimization problem . . . 138
7.2.3 Averaged energy densities . . . 139
7.2.4 Numerical corrector and fine scale features . . . 140
7.2.5 HMM . . . 141
7.2.6 MsFEM . . . 141
7.3 Windowing in the periodic case . . . 142
7.3.1 Setting of the problem . . . 142
7.3.2 Mathematical formulation . . . 143
7.3.3 A remark on boundary conditions . . . 144
7.3.4 A remark on the volume element $C(x,\eta)$ . . . 145
7.3.5 Interpretation of the MsFEM in Petrov-Galerkin formulation . . . 145
7.4 Windowing for general heterogeneities . . . 148
7.4.1 Scaling of the windowing . . . 148
7.4.2 Convergence results . . . 149
7.4.3 Fine scale reconstruction . . . 153
7.5 Conclusion . . . 154

Part III Discrete to continuum limits

8 Exact bounds for the effective behaviour of a ‘discrete’ polycrystal . . . 157
8.1 Introduction . . . 157
8.2 Main result and derivation of bounds in two dimensions . . . 161
8.2.1 Optimal bounds . . . 161
8.2.2 Simple bounds . . . 162
8.2.3 Bounds from arithmetic means . . . 162
8.2.4 Bounds from harmonic means . . . 162
8.3 Optimality of the bounds . . . 165
8.3.1 Arithmetic bound . . . 165
8.3.2 Harmonic bound . . . 166
8.3.3 Optimal bounds . . . 167
8.4 Extension to higher dimension . . . 167
8.4.1 Bounds from arithmetic means . . . 168
8.4.2 Bounds from harmonic means . . . 168
8.4.3 Optimality of the bounds . . . 171
8.5 Interpretation in terms of G-closure . . . 171
8.5.1 Derivation of the trivial bounds . . . 171
8.5.2 Interpretation in terms of quasiconvexification . . . 172
8.5.3 Extensions to any number of conductivities and dimensions . . . 173

9 Variational description of spin systems . . . 175
9.1 Introduction . . . 175
9.2 Notation and preliminary results . . . 176
9.3 Compactness and integral representation results for spin systems . . . 177
9.3.1 Pairwise-interaction energies . . . 177
9.3.2 Case $1 < p < \infty$ . . . 178
9.3.3 Case $p = \infty$ . . . 179
9.3.4 Proof in L^p, $1 < p < \infty$ . . . 181
9.3.5 Proof in L^∞ . . . 185
9.4 Minimum problems . . . 186
9.5 Homogenization . . . 188
9.5.1 Homogenization in L^p, $1 < p < \infty$ . . . 188
9.5.2 Homogenization in L^∞ . . . 192
9.6 Ferromagnetic-antiferromagnetic systems: existence of the bulk limit . . . 193
9.7 Non-pairwise-interaction energies . . . 196

10 Mathematical derivation of a rubber-like stored energy functional . . . 201
10.1 Mesoscopic model . . . 201
10.1.1 Free energy of a polymeric chain . . . 201
10.1.2 Mesoscopic modeling of a rubber-like polymer . . . 202
10.1.3 Stochastic networks and mesoscopic energies . . . 204
10.2 Derivation of a macroscopic model . . . 205
10.2.1 Main result . . . 205
10.2.2 Sketch of the proof . . . 206
10.3 Mechanical properties of the macroscopic model . . . 207
10.3.1 Objectivity . . . 207
10.3.2 Isotropy . . . 207
10.3.3 Behaviour at small strains . . . 208
10.3.4 Compressibility issues . . . 208
10.3.5 Physical setting . . . 209
10.4 Numerical tests and further issues . . . 210
10.4.1 Numerical validation of the model . . . 210
10.4.2 Comments and further developments . . . 211

Part IV Fluid-structure interaction problems

11 Domain decomposition based Newton methods for fluid-structure interaction problems . . . 215
11.1 Introduction . . . 215
11.1.1 Equations de Navier-Stokes . . . 215
11.1.2 Théorie des coques . . . 216
11.1.3 Couplage fluide-structure . . . 217
11.1.4 Aim of the study . . . 218
11.2 Classical solution methods . . . 219
11.2.1 Monolithic formulation . . . 220
11.2.2 Dirichlet to Neumann formulations . . . 220
11.2.3 Symmetric Steklov-Poincaré formulation . . . 220
11.3 Mechanical setting . . . 221
11.3.1 The coupled problem . . . 222
11.3.2 Weak formulation . . . 223
11.4 Semi-discretized weak formulation . . . 224
11.4.1 Implicit coupling scheme . . . 225
11.4.2 Abstract formulations . . . 225

11.4.3 Steklov-Poincaré operators 226
11.5 A partitioned Newton method 227
11.5.1 Weak state operators derivatives 227
11.5.2 Domain decomposition method 229
11.5.3 Complexity analysis 230
11.6 Description of a 3D shell for fluid-structure interaction 230
11.7 Preliminary numerical results 232
11.8 Conclusion 232

Bibliographie 235

Publications

Liste des articles parus ou acceptés dans des revues à comité de lecture

[P1] A. Gloria. A direct approach to numerical homogenization in finite elasticity. *Netw. Heterog. Media.*, Vol. 1 (2006), 109–141, and 513–514.

[P2] A. Gloria. An analytical framework for the numerical homogenization of monotone elliptic operators and quasiconvex energies. *Multiscale Model. Simul.*, Vol. 5 (2006), No. 3, 996–1043.

[P3] A. Braides and A. Gloria. Exact bounds for the effective behaviour of a 'discrete' polycrystal. *Accepted for publication, Multiscale Model. Simul.*

Liste des articles soumis

[S1] A. Gloria. An analytical framework for numerical homogenization - Part II : the case of oversampling. *Submitted.*

[S2] R. Alicandro, M. Cicalese and A. Gloria. Mathematical derivation of a rubber-like stored energy functional. *Submitted.*

Autres travaux

[A1] M. Fernández, J.-F. Gerbeau, A. Gloria and M. Vidrascu. Domain decomposition based Newton methods for fluid-structure interaction problems. *Accepted for publication, ESAIM Proc., CANUM 2006.*

[A2] M. Fernández, J.-F. Gerbeau, A. Gloria and M. Vidrascu. A partitioned Newton method for the interaction of a fluid and a 3D shell structure. *In preparation.*

[A3] R. Alicandro, M. Cicalese and A. Gloria. Variational description and homogenization of bulk energies for bounded and unbounded spin systems. *In preparation.*

[A4] R. Alicandro, M. Cicalese and A. Gloria. Integral representation results for energies defined on stochastic lattices and application to nonlinear elasticity. *In preparation.*

Partie I

Introduction

La science des matériaux est par essence multiéchelle : des équations de Boltzmann à la mécanique des fluides ou de la mécanique quantique à l'élasticité non linéaire, tout est question d'échelles d'espace et de temps. La physique statistique et la physique mathématique ont longtemps fait de ces problématiques leur spécialité. Dans des disciplines aussi variées que la chimie quantique, la biologie, la mécanique, les sciences de l'environnement ou la médecine, le besoin de coupler différentes échelles ou différents modèles s'est développé de manière cloisonnée. Il n'y a pas de science des aspects multiéchelles, en revanche il y a des techniques. Les mathématiques constituent un langage et offrent des outils prévilégiés pour étudier et développer ces techniques, comme en témoigne par exemple la récente création de deux journaux de mathématiques appliquées : *Multiscale Modeling and Simulation* (SIAM, 2003) et *Networks and Heterogeneous Media* (AIMS, 2006). Ces journaux ont pour ambition d'être transversaux et de faciliter le transfert de compétences d'une science à l'autre pour les aspects multiéchelles, ainsi que de promouvoir les outils d'analyse mathématique et de simulation numériques associés.

Plus qu'un effet de mode, l'engouement pour ce genre de techniques vise à répondre aux questions des sciences fondamentales, appliquées et industrielles. Les phénomènes en jeu dans le *design* de matériaux aux propriétés variées, la description et le contrôle de phénomènes biologiques ou l'étude de la conformation spatiale de molécules pharmacologiques par exemple, sont de plus en plus complexes et mettent de plus en plus fréquemment en défaut l'intuition et le savoir-faire des meilleurs ingénieurs et chercheurs. Il y a un besoin de modélisation, d'analyse des modèles et de développement de méthodes numériques efficaces pour aller plus loin. De multiples facettes des mathématiques ont leur rôle à jouer.

Dans cette thèse, nous nous focalisons sur quelques problèmes d'analyse de modèles, de développement et d'analyse de méthodes numériques, pour l'essentiel liés à l'élasticité et pour lesquels différentes échelles spatiales ou modèles interviennent. L'échelle d'intérêt est l'échelle macroscopique, certaines de ses propriétés étant obtenues à partir de principes physiques ou mécaniques à une échelle inférieure.

Autant que possible, nous nous sommes efforcés de traiter les cas les plus généraux dans la classe de problèmes considérés. Le cas échéant, dans la volonté de répondre - ne serait-ce que partiellement - aux questions posées, nous avons été amenés à faire des simplifications, parfois plus guidées par la technologie mathématique utilisée que par le souci de la cohérence mécanique. Ainsi résoudre des problèmes sur la conduction thermique est considéré dans cette thèse comme une résolution partielle des mêmes problèmes sur l'élasticité linéaire. Les résultats obtenus sur des problèmes simplifiés (passage d'une inconnue vectorielle à une inconnue scalaire par exemple) sont envisagés comme des résultats préliminaires. C'est ainsi que doit être interprétée leur présence dans cette thèse.

Material sciences are multiscale by nature : from Boltzmann equations to fluid mechanics or from quantum mechanics to nonlinear elasticity, everything is a matter of timescales and lengthscales. These issues have been the specialities of statistical mechanics and mathematical physics for a long time. In quantum chemistry, medicine, biology, mechanics or even environmental sciences, the need to couple different scales or different models has appeared indepently in each discipline. There is no science of multiscale aspects, but there are techniques. Mathematics constitutes a common language and offers promising tools to study and develop these techniques. Two journals have recently been created in order to facilitate the communication between the different sciences, on multiscale aspects and methods : Multiscale Modeling and Simulation (SIAM, 2003) and Networks and Heterogeneous Media (AIMS, 2006). Both journals are dedicated to the mathematical and numerical aspects of multiscale issues, and their applications.

The development of such techniques aims at answering questions raised by the fundamental, applied and industrial research communities. The phenomena at stake in the design of materials with specific properties, the description and the control of biological phenomena or the study of the spatial configuration of a pharmacological molecule, e.g., are getting more and more complex. They may even question the intuition and the ability of engineers and researchers. There is a need of modeling, analysis of models and development of efficient numerical methods to go further.

In the present thesis, we focus on some issues related to the analysis of models, and the development and analysis of numerical methods in the framework of nonlinear elasticity, regarding multiscale and multiphysics aspects. The scale of interest is the macroscale whereas some properties are obtained from physical or mechanical principles at a finer scale.

1

Présentation des problématiques et des modèles

1.1 Motivation (*in English below*)

Les échelles de temps et les échelles d'espace peuvent toutes deux être source d'un aspect multiéchelle. D'un point de vue mathématique, les outils et méthodes sont très différents, comme en témoignent les ouvrages [119] et [159]. Cette thèse n'aborde que des aspects multiéchelles en espace, le plus souvent dans un cadre statique. Le principe physique central est ainsi la minimisation de l'énergie du système. Le cadre mathématique adéquat est donc le calcul des variations et les équations aux dérivées partielles : l'étude des propriétés mathématiques des énergies et l'étude des équations d'équilibre associées.

D'un point de vue numérique, les problèmes multiéchelles sont très souvent difficiles à aborder car ils nécessitent soit de traiter un nombre bien trop important de degrès de liberté, soit de faire un nombre bien trop important de pas de temps, alors qu'*in fine*, seuls certaines sorties moyennées ou comportements effectifs ont un intérêt. Avant de passer à la description des outils et méthodes utilisées dans la thèse, nous présentons un exemple concret auquel de nombreux travaux, qui ont servi de point de départ pour la Partie II, ont été consacrés (voir notamment [104–106]). Il s'agit de l'application de la méthode des éléments finis multiéchelles qui est présentée au chapitre 3.

Le domaine d'application originel de cette méthode concerne l'ingénierie pétrolière. L'industrie pétrolière est amenée aujourd'hui à extraire du pétrole très difficile d'accès, ce qu'elle ne faisait pas auparavant. Un exemple typique est l'extraction du pétrole de roches poreuses et accidentées (fissurées). Une des méthodes utilisée consiste à injecter de l'eau à certains endroits pour faire jaillir le pétrole. Ceci conduit naturellement à un problème de commande optimale : comment faire jaillir le plus de pétrole en fonction de l'injection d'eau ?

Classique du point de vue de la commande optimale, la difficulté du problème réside dans sa dimension. Le sous-sol étant très accidenté, la simulation numérique des effets d'advection et de diffusion requiert l'utilisation d'un maillage très fin, ce qui rend les calculs très coûteux voire impossibles - même si l'on dispose d'une grande puissance de calcul. Cet effet est très accentué dans l'optique de la commande optimale (où le problème complet est résolu à chaque évaluation de la fonction objectif). La construction des éléments finis multiéchelles est un pré-calcul dans le cas linéaire. Cette étape est certes coûteuse, mais *offline*. Une fois la base précalculée, la résolution numérique du problème d'optimisation est alors beaucoup plus abordable puisque les hétérogénéités sont déjà prises en compte dans la base de Galerkin.

Les éléments finis multiéchelles ne sont cependant qu'une des briques de la résolution de ce type de problème, qui fait l'objet de recherches intensives.

D'autres motivations peuvent conduire au même type de formulation, par exemple la dépollution d'un sous-sol ou encore le stockage souterrain de déchets nucléaires.

Des modèles plus complexes de transport et diffusion dans les milieux poreux peuvent imposer l'emploi d'éléments finis multiéchelles dans un cadre non linéaire, comme celui considéré aux

chapitres 6 et 7. Par ailleurs, le cas de l'élasticité permet de traiter l'exemple d'une mousse de caoutchouc au delà du cas idéalisé périodique. Cependant, les coûts de calcul sont encore assez prohibitifs. Une des questions essentielles pour l'applicabilité de cette approche est la quantification de la taille du volume élémentaire représentatif à considérer dans les applications concrètes, ce qui requiert la construction d'un outil de type estimation a posteriori (dans le cadre de l'homogénéisation stochastique par exemple).

1.2 Quelques problématiques de la modélisation multiéchelle (*in English below*)

1.2.1 Dérivation mathématique de modèles effectifs

Un des objectifs de la modélisation multiéchelle est de dériver rigoureusement des modèles effectifs à l'échelle macroscopique à partir de modèles décrivant des échelles inférieures.

Le dérivation de modèles effectifs est un exemple particulier de modèles multiéchelles : le modèle dit microscopique est alors souvent synonyme de modèle à nombre de degrés de liberté élevé et le modèle dit macroscopique, de modèle à nombre de degrés de liberté réduit. Contrairement à son appellation, le problème microscopique est posé sur un domaine macroscopique. Le fait que la description soit microscopique implique un nombre très grand de dégrés de liberté (idéalement infini dans le passage à la limite). Lorsque le lien entre les descriptions macroscopique et microscopique est direct, que la description du modèle soit fine (microscopique) ou plus grossière (macroscopique), on s'attend à ce que le principe de sélection des solutions ne dépende pas du type de description, microscopique ou macroscopique. Ceci est déjà moins évident lorsque les deux descriptions sont très différentes : le lien entre une fonction d'onde $\psi \in L^2(\mathbb{R}^{6N}, \mathbb{C})$ avec $N \simeq 10^{23}$ et une déformation $u \in H^1(\Omega, \mathbb{R}^3)$ avec Ω un ouvert borné de $\mathbb{R}^3$ n'est pas aussi clair que le lien entre la déformation d'un matériau hétérogène et celle du matériau homogénéisé associé.

Etant donné le cadre de travail variationnel, le principe physique central de minimisation et le lien direct entre les modèles micro et macro que nous nous donnerons, une bonne notion de "dérivation" de modèles est une notion qui assure la convergence des infima d'énergies et de ses éventuels minimiseurs (dans des topologies à définir). On est ainsi assuré que le principe physique de minimisation de l'énergie du modèle fin se traduit bien en principe de minimisation de l'énergie du modèle effectif. Par conséquent, minimiser l'énergie du modèle effectif donne des bonnes informations sur le comportement effectif du modèle fin, qui est *le modèle de départ*.

D'un point de vue mathématique, une notion naturelle est alors la Γ-convergence pour le calcul des variations et la H-convergence pour les équations aux dérivées partielles en général (ou G-convergence dans le cadre variationnel). De nombreux résultats de compacité permettent de considérer de manière abstraite des modèles effectifs. La seconde partie du travail consiste à caractériser l'objet dont on a démontré l'existence. De ce point de vue, le principe physique de minimisation de l'énergie peut être un très bon guide. Sans anticiper sur la suite, on verra qu'une interprétation possible des résultats d'homogénéisation numérique consiste à remplacer la minimisation du problème fin sur un grand espace par deux minimisations emboîtées sur des espaces plus petits. Pour d'autres applications, la bonne notion n'est pas le minimum global de l'énergie, mais certains points critiques. D'autres critères de sélection doivent alors être développés, comme dans [57] par exemple.

Démontrer des résultats de H- ou Γ-convergence peut être très difficile. Parfois, seuls des résultats plus faibles comme une convergence ponctuelle de l'énergie, ou tout autre type de description partielle du comportement effectif à caractériser, sont accessibles à l'heure actuelle.

1.2.2 Couplage discret-continu

Un premier type de couplage est le couplage de modèles discrets et continus. Le mathématicien appliqué est souvent confronté à la configuration inverse : on discrétise un problème continu pour

donner une approximation de la solution. Ici, partant d'un problème discret à une petite échelle, on cherche à le remplacer par un modèle continu "équivalent" à une échelle plus grande. L'étape suivante est souvent la discrétisation du modèle effectif obtenu, où, cette fois, on maîtrise le paramètre de discrétisation.

Plus précisément, cette approche est typique de la dérivation des modèles de mécanique des milieux continus à partir de modèles discrets. Le cadre le plus simple est une chaîne de ressorts élastiques unidimensionnelle, pour laquelle toutes les limites sont caractérisées rigoureusement en termes de Γ-convergence pour la plupart des énergies physiquement raisonnables [40–42]. Au contraire, un exemple bien plus difficile est celui du passage des modèles de la chimie quantique tridimensionnelle à des modèles continus, pour lequel l'état de l'art est loin d'être complet [24–26]. Entre ces deux exemples, il existe tout un ensemble de problèmes intéressants plus ou moins difficiles.

Nous n'abordons pas dans cette thèse les aspects liés aux échelles quantiques. Les problématiques de passage du discret au continu ont un champ d'application mécanique large, le couplage pouvant être de type méso-macro, au sens où le problème fin peut être un problème discret idéalisé à une échelle petite devant celle du continuum mais grande devant l'échelle atomique. On peut par exemple penser à un réseau tridimensionnel de chaînes de polymères considérées comme élastiques. La taille du réseau tendant vers zéro, on retrouve un modèle continu au sens de la Γ-convergence pour certains types d'interaction, comme on le verra au chapitre 10. Le lien entre les modèles micro et macro est alors clair : aux déplacements $\{u_i\}_{i\in I}$ d'un nombre fini (et très grand) de points $\{x_i\}_{i\in I}$ d'un domaine Ω, on associe le déplacement du domaine $u \in H^1(\Omega)$.

Nous avons considéré dans cette optique deux types de modèle : un modèle de spins et un modèle d'élasticité. Au chapitre 9, nous dérivons un modèle continu en partant d'un réseau périodique de spins en interaction. Au chapitre 10, nous présentons l'étude de la dérivation d'un modèle continu d'hyperélasticité en partant d'un réseau stochastique de points en interaction. Ces deux types d'énergie correspondent à des inconnues scalaires dans L^p, et vectorielles dans $W^{1,p}$, respectivement.

1.2.3 Couplage continu-continu

Un autre exemple qui sera abordé en détail dans la suite est l'homogénéisation en élasticité (chapitres 5, 6 et 7). Considérons un matériau composite, obtenu par l'assemblage régulier - idéalement périodique - de deux matériaux aux propriétés mécaniques différentes. A l'échelle microscopique, le matériau est inhomogène, ses propriétés mécaniques varient d'un point à l'autre au grès de l'assemblage des deux phases. A l'échelle macroscopique cependant, ce matériau semble homogène. La caractérisation mathématique et numérique de ce comportement homogène fait l'objet d'une grande partie de cette thèse sous des hypothèses de modélisation variées (énergies convexes, quasiconvexes, cadres périodique, stochastique *etc.*).

Le problème posé consiste ainsi à remplacer un modèle continu à grande variabilité spatiale par un modèle continu plus homogène spatialement. On dérive alors une densité d'énergie continue effective à partir d'une densité d'énergie continue très hétérogène à des petites échelles. L'objectif est ensuite de caractériser *numériquement* la réponse d'un tel matériau à des sollicitations extérieures par la réponse du matériau effectif à ces mêmes sollicitations. La théorie de l'homogénéisation est l'une des théories multiéchelles les plus abouties.

1.2.4 Objectifs de la thèse

La majeure partie du travail effectué durant la thèse a consisté à introduire ou étudier des méthodes analytiques et numériques permettant de dériver des modèles macroscopiques en hyperélasticité non linéaire à partir de descriptions microscopiques, notamment par des couplages continu-continu et discret-continu dans les cadres périodique, stochastique ou plus généralement "compact" (pour lequel on sait a priori que le modèle effectif existe).

Les questions de couplage de modèles continus-continus se reformulent en termes d'homogénéisation. La théorie de l'homogénéisation des opérateurs elliptiques et des énergies hyperélastiques est bien avancée, nous avons plutôt travaillé sur des objectifs numériques aux chapitres 5, 6 et 7. Au contraire, les questions de couplage discret-continu sont plus récentes, et un large espace est ouvert pour améliorer les résultats théoriques existants. Dans le cadre de réseaux périodiques, beaucoup a déjà été fait, notamment par Lions et collaborateurs, et Braides et collaborateurs, sous des hypothèses de modélisation diverses (chimie quantique, énergies à croissance super-linéaire à l'infini *etc.*). C'est dans le cadre des réseaux stochastiques introduits par Blanc, Le Bris et Lions dans [25,26] que nous avons commencé à aborder le passage discret-continu pour des matériaux de type caoutchouc au chapitre 10. La plupart des résultats concernant les méthodes numériques pour l'homogénéisation s'applique *mutatis mutandis* aux méthodes numériques associées au passage discret-continu.

Du point de vue du couplage de modèles, nous proposons, au chapitre 11, une méthode de résolution numérique de problèmes d'interaction fluide-structure par décomposition de domaines pour coupler l'élastodynamique des coques 3D et les équations de Navier-Stokes, en vue d'applications bio-médicales.

La suite de cette introduction vise à rappeler les concepts physiques, mécaniques et mathématiques nécessaires à la mise en perspective du reste de la thèse, tant des points de vues de la modélisation que des résultats mathématiques et numériques. On décrit ainsi brièvement les concepts de base de la mécanique des milieux continus, les résultats et outils d'analyse mathématique des équations de l'élastostatique, les méthodes numériques associées et la théorie de l'homogénéisation. L'objectif est de présenter, au chapitre 4, les résultats de la thèse dans un cadre assez général et de mettre en lumière l'unité des approches utilisées et/ou développées.

Motivation

Time and space scales may both be the source of multiscale aspects. From a mathematical point of view, the tools and methods are very different, as it can be seen in [119] and [159]. In this PhD thesis, we only deal with multiscale aspects in space, most of the time in a static frame. The physical principle at stake is mainly energy minimization. The right mathematical tools are the calculus of variations and partial differential equations : the study of mathematical properties of energies and the study of the associated equilibrium equations.

From a numerical viewpoint, multiscale problems are often very difficult to tackle because one needs to deal either with an important number of degrees of freedom or with an important number of time steps, whereas only averaged quantities or effective behaviors are interesting. Let us give a prototypical example before going further in the description of the tools and methods used in the thesis. This very example is the starting point of Part II. Much work has been devoted to this example (see [104–106]), the multiscale finite element method, which will be detailed in Chapter 3.

This method has been originally designed to answer a problematic of the oil industry. Oil companies face the issue of extracting oil which is hardly accessible. A typical example is the extraction of oil from porous geological areas. One possible method consists in injecting water at given places to make the oil escape. This leads quite naturally to optimal control problems : how to inject water in order to extract the maximum of oil ?

Classical from the optimal control point of view, a major part of the difficulty is a consequence of the dimension. The underground being very heterogeneous, the numerical simulation of advection effects and of diffusion requires the use of very refined mesh, which is usually out of reach due to computational cost issues. This effect is even more stressed within the optimal control frame (for which the complete problem has to be solved at each evaluation of the objective function). The construction of the multiscale finite element basis is a precomputation in the linear case. This step is expensive, however it may be done *offline*. Once the basis is precomputed, the numerical solution

of the optimal control problem is far more affordable since the heterogeneities have already been taken into account in the Galerkin basis.

Multiscale finite elements are only one part of the solution of this kind of problem, which is currently being investigated a lot.

Other motivations can lead to similar problems and formulation. Depollution issues or underground storage of nuclear waste, for instance.

More complex transport and diffusion models in porous media may require the use of multiscale finite elements in a nonlinear framework, as the one considered in Chapters 6 and 7. In addition, the case of elasticity allows us to deal with rubber foams beyond the idealized periodic case. Yet, the computational costs are still rather prohibitive. One important issue related to these methods is the quantification of the size of the representative volume element to consider in practice. This requires in particular the design of an a posteriori estimate tool (in the stochastic homogenization framework for intance).

Some issues of multiscale modeling

Mathematical derivation of effective models

To derive rigorously effective models at the macroscopic scale from a microscopic description is one of the aims of multiscale modeling.

In this case, the microscopic model often involves an important number of degrees of freedom, as opposed to the macroscopic model. When the link between the macroscopic and microscopic descriptions is straighforward, one may expect that the selection principle of the equilibrium configuration does not depend on the description (micro or macro). This is less clear when the two descriptions are very different : the link beween a wave function $\psi \in L^2(\mathbb{R}^{6N}, \mathbb{C})$ with $N \simeq 10^{23}$ and a deformation $u \in H^1(\Omega, \mathbb{R}^3)$ with Ω an open bounded domain of $\mathbb{R}^3$ is not as obvious as the link between the deformations of a heterogeneous material and of the associated homogenized material.

Given a variational framework, the minimization principle and the direct link we will take between the micro and macro models, a good notion for the "derivation" of models is a notion that ensures the convergence of infima and minimizers. Such a derivation implies that minimizing the effective model gives reliable information on the effective behavior of the fine model.

From a mathematical point of view, Γ-convergence is a natural notion for the calculus of variations, and H-convergence for partial differential equations in general (or G-convergence in a variational context). Compactness results allow us to consider, in an abstract manner, effective models. The second part of the work consists in characterizing the very object we have proved to exist. To this extent, the minimization principle may be very useful, as will be seen in the chapters on numerical homogenization. For other applications, good notions are not necessarily global minimizers, but some critical points. Other selection criteria should be developed, as in [57] for instance.

To prove H- or Γ-convergence results may be very difficult. Sometimes, only weaker results such as the pointwise convergence of the energy, or any other partial description of the effective behavior, is available today.

Discrete/continuum coupling

Coupling discrete to continuum descriptions of matter is a first type of issues. The applied mathematician often faces the opposed problem : one discretizes a continuous equation to obtain a numerical approximation of its solution. In the present case, the starting problem is described at

a microscopic scale by a discrete system. We aim at replacing this discrete system by a continuous one, which would be "equivalent" at the macroscopic scale. The following step is usually the discretization of the obtained model, for which we may choose the discretization parameter, as opposed to the original model.

More precisely, this approach is typical of the derivation of continuum mechanics models from discrete models. The simplest framework is an unidimensional chain of elastic springs, for which all the limits are rigorously characterized in terms of Γ-convergence for most of the physically reasonable energies [40–42]. On the contrary, the passage from quantum chemistry to three dimensional continuous models is very tough, and the landscape is far from being complete [24–26]. Between these two sides of the spectrum, one can find a whole set of interesting problems more or less difficult to address.

We do not address aspects related to atomic scales and effects. The problematic to go from discrete to continuous models cover many applications. The coupling may be more meso-macro than really micro-macro : the mesoscopic description can be an idealized discrete problem at a scale which is small with respect to the continuum but large with respect to the atomic lengthscale. One may think of a three dimensional elastic polymeric chain forming a newtork. Letting the typical size of the network go to zero, we recover a continuous model in the sense of Γ-convergence for some interaction energies, as will be seen in Chapter 10. In what follows we will make no difference between a mesoscopic and a microscopic scale. The link between the micro and macro models is clear : to the displacements $\{u_i\}_{i\in I}$ of a finite (but large) number of points we associate the displacement of a domain $u \in H^1(\Omega)$.

In this framework we have considered two types of models : a spin model and an elasticity model. In Chapter 9, we derive a continuous model starting from a periodic network of spins in interaction. In Chapter 10, we present the derivation of a hyperelastic continuous model from a stochastic network of points in interaction. These two types of energies correspond respectively to scalar unknowns in L^p and vectorial unknowns in $W^{1,p}$.

Continuum/continuum coupling

Another example will be dealt with in detail in Chapters 5, 6 and 7, namely numerical homogenization in finite elasticity. Let us consider a composite material obtained by the regular - ideally periodic - assembling of two materials with different mechanical properties. At the microscopic scale, the material is heterogeneous and its properties depend on space according to the assembling of the two phases. At the macroscopic scale however, the material seems to be homogeneous. The mathematical and numerical characterization of the latter homogeneous behavior is the aim of the first part of the thesis, under different modeling assumptions (convex, quasiconvex energies, periodic or stochastic framework *etc.*).

The homogenization method consists in replacing a continuous model whose spatial dependence is very high by another continuous model more homogeneous spatially. To this aim we derive an effective energy density starting from an energy density highly heterogeneous at small scales. A second issue concerns the *numerical* characterization of such a homogenized material to external loads and boundary conditions, which is an approximation of the behavior of the original material under the same loads and boundary conditions.

Coupling of equations

We have addressed the problem of fluid-structure interaction within the framework of the coupling of the Navier-Stokes equations and the nonlinear elastodynamics of shells. In particular, in Chapter 11 we focus on the design of numerical methods. There are basically two points of view : either we aim at using fluid and structure solvers as black boxes or we aim at building a numerical code *ex nihilo* to address at the same time both models. Our method is somewhere in between : we keep the basic numerical solution methods suitable for each type of problem but we also make them communicate a lot using a domain decomposition framework.

Objectives of the thesis

Most of the work of the thesis focuses on introducing or analyzing analytical and numerical methods designed to derive macroscopic models in nonlinear elasticity starting from microscopic descriptions. We have addressed examples in both discrete and continuum to continuum derivation, in a periodic, stochastic and more generally compact framework (for which it is known *a priori* that the effective model exists).

Issues related to continuum to continuum derivation may be recast in terms of homogenization. The homogenization theory for elliptic operators and hyperelastic energies is rather complete. We have addressed more numerically oriented issues in Chapers 5, 6 and 7. On the contrary, deriving rigorously continuum models from discrete descriptions is more recent and there is place for new theoretical results. In the framework of periodic networks, much has already been done, by Lions and collaborators on the one hand, and by Braides and collaborators on the other hand, under various assumptions (quantum chemistry, energy with superlinear growth at infinity*etc.*). Using the stochastic lattices framework introduced by Blanc, Le Bris et Lions in [25, 26], we have begun to investigate in Chapter 10 the discrete to continuum derivation of rubber-like materials. Most of the results on the numerical methods for homogenization apply *mutatis mutandis* on numerical methods associated to the discrete to continuum coupling.

From the point of view of coupling different models, we have proposed in Chapter 11 a solution method for fluid-structure interaction problems based on domain decomposition algorithm. We aim at using this method to solve numerically the coupling of the elastodynamics of 3D shells with the Navier-Stokes equations, for biomedical applications.

In the remainder of the introduction, we quickly recall the physical, mechanical and mathematical concepts that may allow the reader to put in perspective the core of the thesis, from the modeling, numerical and mathematical points of view. We first describe basic notations and concepts of continuum mechanics, the main results and mathematical tools for the analysis of the elastostactics problem, and numerical methods for the homogenization of linear elliptic operators. Our objective is to present, in Chapter 4, the main results of the thesis in a rather general framework that may allow to put in evidence the similarities of the approaches used and/or developed.

1.3 Modélisation en mécanique des milieux continus

Ce paragraphe s'inspire des traités de Ciarlet [52] et Le Tallec [125] sur l'élasticité.

1.3.1 Energie, déformation et contraintes

Un problème central en élasticité non linéaire tridimensionnelle consiste à déterminer la position d'équilibre d'un corps élastique qui occupe une configuration de référence Ω (ouvert borné de $\mathbb{R}^3$) en l'absence de forces appliquées (un tel état est appelé état d'équilibre naturel). Soumis à des forces extérieures et/ou déplacements imposés de sa frontière, le corps occupe une configuration déformée $\phi(\Omega)$ (représentée Figure 1.1), caractérisée par une application $\phi : \Omega \to \mathbb{R}^3$ qui doit en particulier préserver l'orientation et être injective pour être physiquement acceptable. Ces applications sont appelées des déformations et sont des champs de vecteurs dans $\mathbb{R}^3$. Parmi toutes les déformations possibles, les déformations d'intérêt sont celles qui minimisent l'énergie du système ou satisfont les équations d'équilibre (première variation formelle du problème de minimisation). Le champ de déplacement associé u mesure l'écart entre la configuration déformée et la configuration de référence. Il s'écrit $u : \Omega \to \mathbb{R}^3, X \mapsto \phi(X) - X$.

La description du système dans la configuration déformée est appelée description eulérienne. Les différents champs (déformation, vitesse, accélération,...) sont alors exprimés au point géométrique $x \in \phi(\Omega)$. En utilisant l'application inverse ϕ^{-1}, on peut se ramener à la configuration de

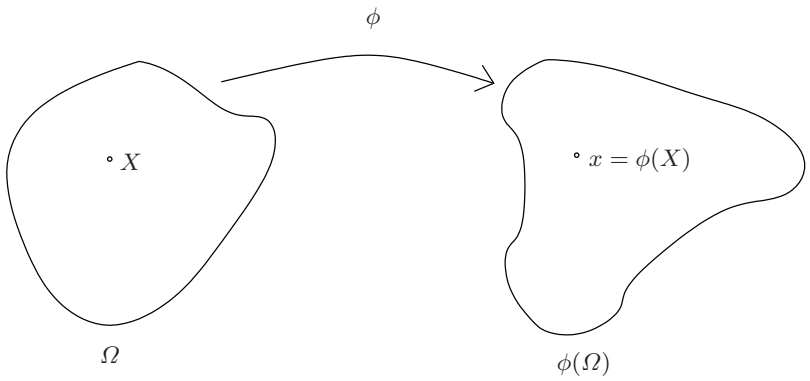

Fig. 1.1. Représentation d'une déformation d'un milieu continu

référence. Le transport des champs eulériens dans la configuration de référence permet de définir les champs lagrangiens (associés à un point matériel et non plus géométrique). De manière générale, la formulation eulérienne est bien adaptée à la description des fluides, tandis que la formulation lagrangienne est mieux adaptée à la description des solides (la densité d'énergie est généralement donnée dans la configuration de référence, formulation qui permet de démontrer l'existence de solutions au problème d'équilibre d'un corps élastique).

L'énergie I du système est un scalaire qui s'exprime comme l'intégrale sur la configuration d'une densité d'énergie. Dans le cas d'un matériau hyperélastique - seul cas considéré par la suite, la densité d'énergie au point matériel $X \in \Omega$ dans la configuration de référence ne dépend, par définition, que de X et du gradient de déformation $\nabla\phi(X)$. On la note $W(X, \nabla\phi(X)) \in \bar{\mathbb{R}}$. Ainsi,

$$I(\phi) = \int_\Omega W(X, \nabla\phi(X)).$$

La densité vérifie en particulier des propriétés de compatibilité géométrique liées à des hypothèses de modélisation physique comme l'objectivité (aussi appelée indifférence matérielle) ou l'isotropie.

La dérivée de la densité d'énergie par rapport au gradient de déformation (variable muette ξ) définit le premier tenseur de Piola-Kirchhoff :

$$\pi(X, \nabla\phi(X)) = \frac{\partial W}{\partial \xi}(X, \nabla\phi(X)).$$

Le premier tenseur de Piola-Kirchhoff est un tenseur du deuxième ordre (non nécessairement symétrique). Il correspond au transport par ϕ^{-1} du tenseur (symétrique) des contraintes de Cauchy $\sigma(x) = \sigma(X, \nabla\phi(X))$ défini sur la configuration déformée au point $x = \phi(X)$:

$$\pi(X, \nabla\phi(X)) = (\det \nabla\phi(X))\sigma(X, \nabla\phi(X))\nabla\phi(X)^{-T}.$$

L'indifférence matérielle traduit le fait que, toute quantité observable (donc dans la configuration déformée) ayant un caractère intrinsèque (telle que la densité de masse, la quantité d'accélération, *etc.*) doit être indépendante de l'observateur. Ceci s'applique en particulier au tenseur des contraintes de Cauchy et impose

$$\sigma(X, Q\xi) = Q\sigma(X, \xi)Q^T$$

pour tout $\xi \in \mathbb{R}^{3\times 3}$ tel que $\det(\xi) > 0$ et toute transformation orthogonale $Q \in O^3(\mathbb{R})$. Transportée sur la configuration de référence, cette égalité s'écrit

$$\pi(X, Q\xi) = Q\pi(X, \xi).$$

Elle se traduit par des restrictions concrètes. L'objectivité est caractérisée par la propriété suivante sur la densité d'énergie : $W(X, Q\xi) = W(X, \xi)$. De manière équivalente, l'énergie doit pouvoir s'écrire sous la forme $W(X, \xi) = \tilde{W}(X, \xi^T\xi)$, ce qui, en toute généralité, exclut la convexité (voir [52, Th 4.8-1]). Cette symétrie a des conséquences importantes sur la richesse des phénomènes mécaniques (par exemple : instabilités Figure 5.1 et bifurcations Tableau 5.2), la complexité de l'analyse mathématique et des méthodes numériques.

On peut particulariser encore plus l'énergie en considérant d'autres propriétés. L'isotropie, par exemple, correspond à l'idée intuitive qu'un matériau répond aux sollicitations de la même manière dans toutes les directions. Mathématiquement, cette propriété se traduit sur le tenseur de Cauchy par

$$\sigma(X, \xi Q) = \sigma(X, \xi)$$

pour tout $\xi \in \mathbb{R}^{3\times 3}$ tel que $\det(\xi) > 0$ et toute matrice de rotation Q : le tenseur de Cauchy est inchangé quand la configuration de référence est sujette à une rotation arbitraire autour du point $X \in \Omega$. Le transport de cette égalité sur la configuration de référence impose alors

$$\pi(X, \xi Q) = \pi(X, \xi)Q.$$

En termes de dépendance, le théorème de représentation de Rivlin-Eriksen montre que la densité d'énergie d'un matériau hyperélastique objectif et isotrope s'écrit sous la forme :

$$W(X, \xi) = \hat{W}(X, I_1(\xi^T\xi), I_2(\xi^T\xi), I_3(\xi^T\xi))$$

où I_1, I_2 et I_3 sont les invariants principaux de la matrice symétrique $\xi^T\xi$ donnés par les expressions suivantes

$$\begin{cases} I_1(A) = \mathrm{tr}(A) \\ I_2(A) = \mathrm{tr}(\mathrm{Cof}A) \\ I_3(A) = \det(A) \end{cases}$$

et Cof$A$ est la matrice des cofacteurs de A, pour toute matrice symétrique $A \in \mathcal{S}_3(\mathbb{R})$.

Les exemples les plus courants de densité d'énergie hyperélastique objective, homogène et isotrope sont les lois de Saint Venant-Kirchhoff, de Ciarlet-Geymonat (cas particulier des lois d'Ogden), et de Mooney-Rivlin rappelées ci-dessous.

$$\begin{aligned} W_{StV-K}(\xi) &= \frac{\lambda}{2}(\mathrm{tr}\, E)^2 + \mu \mathrm{tr}\, E^2 \\ W_{C-G}(\xi) &= C_1(I_1 - 3) + C_2(I_2 - 3) + a(I_3 - 1) - (C_1 + 2C_2 + a)\log(I_3) \\ W_{M-R}(\xi) &= \begin{cases} C_1(I_1 - 3) + C_2(I_2 - 3) & \text{si } I_3 = 1 \\ +\infty & \text{si } I_3 \neq 1 \end{cases}. \end{aligned}$$

Le tenseur E est le tenseur de Green-Saint Venant défini par $E = \frac{1}{2}(C - \mathrm{Id}) = \frac{1}{2}(\nabla u^T + \nabla u + \nabla u^T \nabla u)$, avec $\nabla u = \xi - \mathrm{Id}$. Les constantes du modèle de Saint Venant-Kirchhoff satisfont les relations $\lambda, \mu > 0$ et ont des valeurs typiques de l'ordre de $10^5 \mathrm{kg/cm}^2$ pour les métaux. Ce type d'énergie a un domaine d'application restreint aux petites déformations. Le modèle de Saint Venant-Kirchhoff souffre à la fois d'inconvénients mécaniques et mathématiques. Il trouve cependant tout son intérêt dans la remarque suivante.

Remarque **1** *[52, Th. 4.5-1.] Considérons un matériau hyperélastique, objectif, isotrope et homogène (la densité d'énergie ne dépend pas du point $X \in \Omega$), pour lequel la configuration de référence est un état d'équilibre naturel. Sa densité d'énergie est de la forme :*

$$W(\xi) = \tilde{W}(\xi^T\xi) = \bar{W}(E).$$

Si la densité d'énergie exprimée en fonction des invariants est deux fois dérivable en $\xi = Id$, alors

$$\bar{W}(E) = \frac{\lambda}{2}(tr\, E)^2 + \mu tr\, E^2 + o(\|E\|^2), \quad \xi^T\xi = Id + 2E.$$

Cette remarque implique qu'en régime de petites déformations, tout matériau hyperélastique isotrope homogène est "proche" d'un matériau de Saint Venant-Kirchhoff. L'utilisation de la loi de Saint Venant-Kirchhoff constitue une linéarisation mécanique (en termes de tenseur de Green-Lagrange). La linéarisation du tenseur de Cauchy associé à la loi de Saint Venant-Kirchhoff est de nature géométrique et consiste à remplacer E par son développement à l'ordre un : $e = \frac{1}{2}(\nabla u^T + \nabla u)$. Ainsi, $\sigma = \lambda \mathrm{tr}\, e \mathrm{Id} + 2\mu e$. Si on identifie le domaine déformé au domaine de référence, la densité d'énergie associée s'écrit

$$W_l(\xi) = \frac{\lambda}{2}(\mathrm{tr}\, e)^2 + \mu \mathrm{tr}\, e^2.$$

La justification rigoureuse de la linéarisation géométrique et du lien entre élasticité linéaire et élasticité non linéaire en petits déplacements et petites déformations est très récente [60]. Le résultat principal est énoncé au chapitre 2.

Revenons maintenant aux deux autres exemples de densité d'énergie. Les matériaux de Ciarlet-Geymonat satisfont deux propriétés "majeures" de modélisation mécanique. La première concerne le comportement quand $\det \xi \to 0^+$:

$$\lim_{I_3(\xi^T\xi)\to 0^+} W_{C-G}(\xi) = +\infty. \tag{1.1}$$

En particulier, cette propriété modélise le fait qu'on ne peut comprimer infiniment la matière sans lui donner infiniment d'énergie (I_3 mesure les changements de volume). Cette propriété, qui n'est pas vérifiée par les matériaux de Saint Venant-Kirchhoff, est à l'origine de nombreuses difficultés mathématiques. La seconde propriété concerne le comportement quand $|\xi| \to \infty$ et sert à donner de la coercivité à l'énergie dans des espaces fonctionnels adéquats. Elle s'exprime par

$$\lim_{|I_1|+|I_2|+|I_3|\to+\infty} W(\xi) = +\infty.$$

Comme le souligne P.G. Ciarlet [52, Sect. 4.6], tout comportement en grandes déformations (au sens de la modélisation mathématique $|\xi| \to \infty$) relève essentiellement de l'hypothèse mathématique puisque les expériences concrètes de mécanique ne peuvent être réalisées que pour des valeurs de $|\xi|$ dans un compact. Par ailleurs, d'autres phénomènes comme la rupture interviennent bien avant ce genre de comportement en "très grandes" déformations. Ainsi les hypothèses de coercivité de la densité d'énergie - dans le régime de la mécanique des milieux continus - peuvent être considérées comme raisonnables tant que les résultats restent dans des plages de déformation réalistes et sans rupture. Dans d'autres régimes cependant, comme la rupture, ces hypothèses doivent être relaxées. Le troisième exemple, les matériaux de Mooney-Rivlin, modélise un comportement incompressible. Les modèles de Ciarlet-Geymonat et Mooney-Rivlin s'appliquent particulièrement bien aux caoutchoucs pour des valeurs typiques de $C_1 = 0.5$MPa et $C_2 = 0.0056$MPa.

Au chapitre 10, nous vérifierons que la densité d'énergie obtenue par passage discret-continu est hyperélastique, objective, homogène et isotrope.

1.3.2 Equations de l'élastostatique et de l'élastodynamique tridimensionnelle

Le problème de la détermination de la (ou d'une) position d'équilibre d'un corps élastique sollicité, qui occupait une configuration de référence Ω (ouvert borné de $\mathbb{R}^3$) en l'absence de forces appliquées, est appelé le problème de l'élastostatique. Il peut s'écrire sous deux formes : un problème aux limites composé d'un système de trois équations aux dérivées partielles du second ordre ou un problème de minimisation de l'énergie associée. Le système d'équations aux dérivées partielles correspond formellement aux équations d'Euler-Lagrange du problème de minimisation. La question de l'équivalence mathématique de ces formulations est un problème difficile et en grande partie ouvert (voir [181]).

Dans les développements qui suivent, nous ne considérons que des forces exercées indépendantes de la solution elle-même. Il s'agit de fortes dites mortes, par opposition aux forces vives qui dépendent de la configuration déformée (penser par exemple à un champ de gravité non homogène).

Problème de minimisation

L'interprétation physique de cette formulation est l'application du "principe de minimisation" : parmi toutes les configurations possibles, le système occupe la configuration d'énergie minimale. Soit $\mathbf{\Phi}$ un espace de déformations admissibles et L une forme linéaire sur $\mathbf{\Phi}$ modélisant les conditions aux limites de Neumann et les forces volumiques. Typiquement, $L(\psi) = \int_\Omega f\psi + \int_{\partial\Omega_N} T \cdot n\psi$ où f est une force de volume et T une contrainte imposée sur la partie $\partial\Omega_N$ du bord de Ω. Dans ce cas, le problème de minimisation s'écrit :

$$\inf_{\psi\in\mathbf{\Phi}} \left\{ \int_\Omega W(X, \nabla\psi(X)) + L(\psi) \right\}. \tag{1.2}$$

La définition précise de $\mathbf{\Phi}$ sera donnée au paragraphe 2.1.

Par la suite, nous nous focaliserons sur des problèmes de minimisation de ce type, dans les contextes de l'homogénéisation (chapitres 5-7) et du passage discret-continu (chapitres 8-10).

Equations d'équilibre

L'interprétation physique de cette formulation fait intervenir le concept de contraintes. La formule de Stokes montre que la nullité de la divergence du tenseur des contraintes sur un petit cube matériel est équivalent pour le cube à compenser, par sa déformation, les forces exercées sur sa surface par les cubes adjacents. Le cas échéant, la déformation peut aussi compenser une force volumique. Cette formulation traduit qu'un corps est à l'équilibre si les efforts intérieurs compensent exactement les forces extérieures. Dans le cas où les équations d'équilibre sont les équations d'Euler-Lagrange du problème de minimisation et où l'espace tangent est bien défini, une déformation qui satisfait ces équations est un point critique de l'énergie, un équilibre local. Les équations d'équilibre *formellement* impliquées par (1.2) s'écrivent : trouver $\phi \in \mathbf{\Phi}$ telle que

$$\begin{cases} -\text{div}\ \pi(X, \nabla\phi(X)) = f(X) & \text{dans } \Omega \\ \phi(X) = \phi_0(X) & \text{sur } \partial\Omega \setminus \partial\Omega_N \\ \pi(X, \nabla\phi(X)) = T(X) & \text{sur } \partial\Omega_N \end{cases}. \tag{1.3}$$

Nous avons volontairement choisi la même notation pour l'espace de solutions $\mathbf{\Phi}$. Cependant, les espaces dans lesquels les minimiseurs existent ne sont pas assez réguliers pour donner un sens aux équations aux dérivées partielles et aux conditions aux limites. Par ailleurs, la condition aux limites de Dirichlet est redondante et déjà présente dans $\mathbf{\Phi}$.

Elastodynamique

Les équations de l'élastodynamique sont reliées aux équations d'équilibre et non pas au problème de minimisation. Ces équations sont de type système hyperbolique de lois de conservation. Soit $\tau > 0$, l'inconnue est la déformation $\phi : [0, \tau[\times\Omega \to \mathbb{R}^3, (t, X) \mapsto \phi(t, X)$. Si on néglige les forces de volume, les équations s'écrivent

$$\begin{cases} \rho\dfrac{\partial^2\phi}{\partial t^2} - \text{div}\ \pi(\nabla\phi) = 0 & \text{dans } \Omega \\ \phi(0, X) = \phi_0(X) & \text{dans } \Omega \\ \phi(t, X)_{|\partial\Omega_D} = \phi_D(t, X) & \text{sur } \partial\Omega_D \\ \pi(\nabla\phi)_{|\partial\Omega_N}(t, X) = \pi_N(t, X) & \text{sur } \partial\Omega_N \end{cases}, \tag{1.4}$$

où u_0 est la condition initiale, et u_D et π_N des conditions aux limites.

2

Analyse mathématique et numérique des équations de l'élastostatique

Ce chapitre est composé de deux parties. La première partie rappelle les résultats mathématiques principaux sur les équations de l'élastostatique. Elle s'inspire largement des traités de Ciarlet [52], et Braides et Defranceschi [38]. La méthode directe du calcul des variations et les propriétés de semi-continuité inférieure des fonctionnelles intégrales sont omniprésentes dans les développements sur l'homogénéisation numérique des chapitres 5, 6 and 7, de même que la théorie de la Γ-convergence.

Dans la deuxième partie de ce chapitre, on décrit des méthodes numériques pour les équations de l'élastostatique. Cette partie s'inspire des ouvrages de Le Tallec [124, 125]. Ces notions sont à la base des chapitres 5 et 11.

Dans la suite, Ω désigne un ouvert borné régulier et connexe de $\mathbb{R}^3$.

2.1 Analyse mathématique

Dans cette section, nous nous restreignons au problème de l'élastostatique avec forces mortes. Certains des résultats ci-dessous se généralisent au cas de forces vives. On renvoie à l'ouvrage [52] pour ces questions.

2.1.1 Elasticité linéaire

Le système de l'élasticité linéaire est donné par

$$\begin{cases} -\mathrm{div}\,\{\lambda(\mathrm{tr}e(u))\mathrm{Id} + 2\mu e(u)\} = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega_D \\ \{\lambda(\mathrm{tr}e)\mathrm{Id} + 2\mu e\} \cdot n = g & \text{sur } \partial\Omega_N \end{cases}, \tag{2.1}$$

où $e(u) = \frac{1}{2}(\nabla u + \nabla u^T)$.

L'étude du problème de l'élasticité linéaire est similaire à l'étude de l'équation de Laplace. Il s'agit de minimiser l'énergie quadratique

$$E(u) = \int_\Omega \frac{1}{2}\lambda(\mathrm{tr}e(u))^2 + \mu e(u) : e(u) - gu,$$

sur l'espace de Hilbert $\{u \in H^1(\Omega, \mathbb{R}^3), u_{|\partial\Omega_D} = 0\}$. L'intégrande étant convexe, la fonctionnelle d'énergie est semi-continue inférieurement pour la topologie faible de $H^1(\Omega, \mathbb{R}^3)$. Ainsi pour obtenir l'existence d'une solution, il suffit de montrer que l'énergie est coercive, à savoir que l'énergie domine la norme $H^1(\Omega, \mathbb{R}^3)$. Cette propriété de coercivité est surprenante car on n'a d'information que sur le symétrisé $e(u)$ de ∇u et pas sur toutes les composantes de ∇u. C'est la deuxième inégalité de Korn qui donne ce contrôle, et par conséquent l'existence et l'unicité de la solution du problème linéaire.

Lemme 1 *[151] Il existe une constante $c > 0$ telle que*

$$\|v\|_{H^1(\Omega,\mathbb{R}^3)} \leq c(\|v\|^2_{L^2(\Omega,\mathbb{R}^3)} + \|e(v)\|^2_{L^2(\Omega,\mathbb{R}^3)})^{1/2}$$

pour tout $v \in H^1(\Omega,\mathbb{R}^3)$. L'application $v \mapsto (\|v\|^2_{L^2(\Omega,\mathbb{R}^3)} + \|e(v)\|^2_{L^2(\Omega,\mathbb{R}^3)})^{1/2}$ est donc une norme équivalente à $\|\cdot\|_{H^1(\Omega,\mathbb{R}^3)}$.

Par la suite, nous désignerons indifféremment par $L^p(\Omega)$ l'espace de Lebesgue $L^p(\Omega,\mathbb{R}^3)$ et par $W^{1,p}(\Omega)$ l'espace de Sobolev $W^{1,p}(\Omega,\mathbb{R}^3)$. Sauf si précisé explicitement, tous les résultats qui suivent sont valables à la fois pour les cas scalaires $W^{1,p}(\Omega,\mathbb{R})$ et les cas vectoriels $W^{1,p}(\Omega,\mathbb{R}^3)$, quitte à remplacer la double contraction ":" au sens des tenseurs d'ordre deux par le produit scalaire "$\cdot$" au sens des vecteurs.

Une approche directe du système d'équations aux dérivées partielles consiste à appliquer le théorème de représentation de Riesz, grâce auquel on obtient

Théorème 1 *[52, Th. 6.3-5] Soient $\lambda > 0$, $\mu > 0$, $f \in L^{6/5}(\Omega)$, $g \in L^{4/3}(\partial\Omega_N)$ et $\mathcal{L}^2(\partial\Omega_D) > 0$. Il existe une et une seule solution (faible) de (2.1) dans $V = \{v \in H^1(\Omega), v = 0 \text{ sur } \partial\Omega_D\}$.*

Par ailleurs, le système de l'élasticité est régularisant et on a

Théorème 2 *[52, Th. 6.3-6] Si $\partial\Omega$ est de classe $\mathcal{C}^2$, $\partial\Omega_D = \partial\Omega$ et $f \in L^p(\Omega)$ avec $p \geq \frac{6}{5}$ alors l'unique solution faible $u \in H^1_0(\Omega)$ du système de l'élasticité linéaire en déplacement pur (2.1) est dans l'espace $W^{2,p}(\Omega)$ et satisfait*

$$-\mathrm{div}\,\{\lambda(\mathrm{tr}e)\mathrm{Id} + 2\mu e\} = f \qquad \text{dans } L^p(\Omega).$$

De plus, soit $m \geq 1$ un entier. Si le bord $\partial\Omega$ est de classe $\mathcal{C}^{m+2}$ et si $f \in W^{m,p}(\Omega)$, alors la solution faible $u \in H^1_0(\Omega)$ est de classe $W^{m+2,p}(\Omega)$.

Ce résultat de régularité est particulièrement intéressant pour le paragraphe suivant car il permet d'obtenir l'inversibilité d'une application linéaire, requise pour appliquer le théorème des fonctions implicites.

2.1.2 Application du théorème des fonctions implicites

Le théorème des fonctions implicites est l'un des deux outils principaux d'analyse des équations de l'élastostatique. Partant d'une configuration d'équilibre dépendant d'un paramètre de contrôle (la force volumique par exemple) par rapport auquel le système d'équations est différentiable et la différentielle inversible, on peut démontrer l'existence et l'unicité locale de solutions régulières pour des configurations proches. L'inversibilité de la différentielle se ramène à l'étude d'un problème de type (2.1). Le théorème suivant fait notamment le lien entre l'élasticité linéaire et l'élasticité non linéaire en petites déformations et petits déplacements.

Le résultat peut s'écrire comme suit.

Théorème 3 *[52, Th. 6.7-1] Supposons $\partial\Omega$ de classe $\mathcal{C}^2$ et que l'application $\Sigma \in \mathcal{C}^2(\mathbb{V}(0),\mathbb{S}^3)$ satisfait*

$$\Sigma(E) = \lambda(\mathrm{tr}E)\mathrm{Id} + 2\mu E + O(\|E\|^2), \quad \text{avec } \lambda,\mu > 0,$$

où $\mathbb{S}^3$ est l'ensemble des matrices d'ordre 3 symétriques et $\mathbb{V}(0)$ un voisinage de l'origine dans $\mathbb{S}^3$. Alors, pour tout $p > 3$, il existe un voisinage $\mathbf{F}^p$ de l'origine dans l'espace $L^p(\Omega)$ et un voisinage $\mathbf{U}^p$ de l'origine dans l'espace

$$\mathbf{V}^p(\Omega) = \{v \in W^{2,p}(\Omega), v = 0 \text{ sur } \partial\Omega\}$$

tels que pour tout $f \in \mathbf{F}^p$, le problème aux limites

$$-\mathrm{div}\,[(\mathrm{Id} + \nabla u)\Sigma(E(u))] = f \tag{2.2}$$

a exactement une solution u dans $\mathbf{U}^p$.

Le Théorème 3 montre que le problème d'élastostatique en déplacement pur (2.2) admet une et une seule solution si le chargement f est petit. Le problème (2.2) est écrit en utilisant le tenseur Σ, appelé deuxième tenseur de Piola-Kirchhoff. C'est un tenseur symétrique relié au premier tenseur de Piola-Kirchhoff par la relation

$$\pi(u) = (\mathrm{Id} + \nabla u)\Sigma(E(u)).$$

D'après la Remarque 1, ce théorème couvre de manière générale les lois hyperélastiques isotropes homogènes. Il peut se généraliser au cas non homogène moyennant une dépendance spatiale régulière. Cependant, l'application de ce théorème est limitée en termes de conditions aux limites par l'hypothèse fondamentale de régularité $W^{2,p}(\Omega)$: on ne peut par exemple par traiter le cas de conditions aux limites mixtes Neumann et Dirichlet ou le cas de conditions aux limites de Dirichlet sur un ouvert peu régulier.

L'analyse de la méthode de résolution numérique d'un problème homogénéisé proposée au chapitre 2 repose sur le théorème des fonctions implicites, qui permet de démontrer - dans les cas simples - que les formules (5.76) et (5.77) donnant les dérivées première et seconde de l'énergie homogénéisée sont justes.

Le calcul des variations permet de démontrer l'existence de solutions dans le cas général, en passant par la formulation variationnelle plutôt que par le système d'équations aux dérivées partielles.

2.1.3 Méthode directe du calcul des variations

Le deuxième outil principal d'analyse du problème de l'élastostatique est la méthode directe du calcul des variations. Elle peut s'énoncer ainsi.

Soit $J : V \to \overline{\mathbb{R}}^+$ une fonctionnelle sur un espace métrique $(V, \|\cdot\|_V)$. Soit $R > \inf_V J + 1$. On suppose que V et J sont tels que $V_R = \{v \in V \,|\, J(v) \leq R\}$ est précompact pour une topologie $\mathcal{T}$ sur V. On considère alors une suite minimisante v_n de J sur V. Comme $v_n \in V_R$ pour n assez grand, v_n converge à extraction près vers $u \in V$ pour la topologie $\mathcal{T}$. Si J est (séquentiellement) semi-continue inférieurement pour la topologie $\mathcal{T}$, alors

$$J(u) \leq \liminf_{n\to\infty} J(v_n) = \inf_V J,$$

donc u est un minimiseur de J sur V.

Nous avons déjà appliqué la méthode directe du calcul des variations pour résoudre le problème de l'élasticité linéaire (2.1).

Afin d'être en mesure d'utiliser cette méthode, il convient de trouver des conditions nécessaires et/ou suffisantes de semi-continuité inférieure des fonctionnelles qui nous intéressent, à savoir les fonctionnelles intégrales.

Par la suite, on utilisera la coercivité.

Définition 1 *[38, Def 1.8] $J : V \to \overline{\mathbb{R}}^+$ est coercive si pour tout $t \in \mathbb{R}^+$, l'ensemble $V_t = \{v \in V, J(v) \leq t\}$ est précompact s'il est non vide (c'est-à-dire qu'à extraction près, toute suite de V_t converge vers une fonction $v \in V$ pour la topologie faible de V).*

La méthode directe du calcul des variations se résume donc à

$$\text{Semicontinuité inférieure} \;+\; \text{Coercivité} \;\Longrightarrow\; \text{Existence de minimiseur(s).}$$

Cette méthode est à la base des résultats des chapitres 5, 6, 7, 9, et 10.

2.1.4 Quasiconvexité et semi-continuité inférieure des fonctionnelles intégrales

Une condition nécessaire à la semi-continuité inférieure (sci) des fonctionnelles intégrales pour la topologie faible des espaces de Sobolev $W^{1,p}(\Omega, \mathbb{R}^m)$, $p \geq 1$, $m \geq 1$ est la notion de quasiconvexité introduite par Morrey [141]. Cette notion est très intéressante mais elle a un inconvénient pratique majeur : il est aussi difficile - dans le cas général - de vérifier qu'une densité d'énergie est quasiconvexe que de démontrer que la fonctionnelle intégrale est sci.

Ceci est une différence fondamentale avec la notion de convexité (qui est plus forte et coïncide avec la quasiconvexité dans le cas scalaire $m = 1$) pour laquelle une caractérisation en terme de dérivée seconde existe si la fonction est C^2.

Définition 2 *[38, Def 4.2] Une fonction $f : \mathcal{M}_{m\times n}(\mathbb{R}) \to \overline{\mathbb{R}}^+$ est dite $W^{1,p}$-quasiconvexe en $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ si il existe un ouvert borné $E \subset \mathbb{R}^n$ lipschitzien tel que*

$$f(A) = \inf \left\{ \frac{1}{|E|} \int_E f(A + \nabla\phi(x))dx : \phi \in W_0^{1,p}(E, \mathbb{R}^m) \right\}. \tag{2.3}$$

La fonction f est $W^{1,p}$-quasiconvexe si elle vérifie (2.3) pour tout $A \in \mathcal{M}_{m\times n}(\mathbb{R})$.

La proposition suivante fait le lien entre la $W^{1,p}$-quasiconvexité et la semi-continuité inférieure des fonctionnelles intégrales.

Proposition 1 *[38, Prop 4.3] Si la fonctionnelle intégrale $I(u) = \int_\Omega f(\nabla u(x))dx$ est (séquentiellement) semi-continue inférieurement pour la topologie faible de $W^{1,p}(\Omega, \mathbb{R}^m)$ (ou faible-* si $p = \infty$), alors f est $W^{1,p}$-quasiconvexe.*

Avant de donner une réciproque partielle à la Proposition 1, nous introduisons la notion de quasiconvexité :

Définition 3 *[38, Def 5.14] Une fonction $f : \mathcal{M}_{m\times n}(\mathbb{R}) \to \overline{\mathbb{R}}^+$ est dite quasiconvexe si f est continue et si pour tout $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, et tout ouvert borné $E \subset \mathbb{R}^n$ lipschitzien,*

$$f(A) = \inf \left\{ \frac{1}{|E|} \int_E f(A + \nabla\phi(x))dx : \phi \in C_0^\infty(E, \mathbb{R}^m) \right\}. \tag{2.4}$$

La réciproque s'énonce alors

Proposition 2 *[38, Th. 5.16] Si $1 \leq p < \infty$, et si $f : \mathcal{M}_{m\times n}(\mathbb{R}) \to \overline{\mathbb{R}}^+$ est une fonction quasiconvexe satisfaisant la propriété de domination*

$$0 \leq f(A) \leq C(1 + |A|^p) \qquad \text{pour tout } A \in \mathcal{M}_{m\times n}(\mathbb{R}), \tag{2.5}$$

la fonctionnelle intégrale $I(u) = \int_\Omega f(\nabla u(x))dx$ est (séquentiellement) semi-continue inférieurement pour la topologie faible de $W^{1,p}(\Omega, \mathbb{R}^m)$.

La Proposition 2 s'étend aux fonctions $f : \Omega \times \mathcal{M}_{m\times n}(\mathbb{R}) \to \overline{\mathbb{R}}^+$ dès que f est Carathéodory, à savoir mesurable en la première variable et continue en la seconde ([2]).

Du point de vue de la modélisation, la Proposition 2 est intéressante car elle permet de concilier l'objectivité des densités d'énergie mécaniques (il n'y a pas d'incompatibilité entre la quasiconvexité et l'objectivité, contrairement à la convexité) avec l'existence de solutions au problème de l'élastostatique (les fonctionnelles d'énergie sont semi-continues inférieurement, comme pour les densités d'énergie convexes). C'est pourquoi nous avons considéré des densités d'énergie quasiconvexes dans la suite. En revanche, la condition de croissance (2.5) est incompatible avec la condition (1.1) quand $\det \xi \to 0^+$.

Remarque **2** *Si $f : \mathcal{M}_{m\times n}(\mathbb{R}) \to \overline{\mathbb{R}}^+$ est quasiconvexe et continue alors elle est rang-1 convexe : pour tout $A, B \in \mathcal{M}_{m\times n}$, la fonction $t \mapsto f(A + tB)$ est convexe.*

Un résultat décisif a été obtenu par Ball dans [14] pour s'affranchir de la condition (2.5) en introduisant une classe particulière de fonctions quasiconvexes.

2.1.5 Polyconvexité et théorie d'existence en élasticité non linéaire

La notion introduite par Ball dans [14] est la polyconvexité :

Définition 4 *[38, Def 5.6] Une fonction $f : \mathcal{M}_{m\times n}(\mathbb{R}) \to (-\infty, +\infty]$ est dite polyconvexe s'il existe une fonction convexe $g : \mathbb{R}^{\tau(n,m)} \to \overline{\mathbb{R}}$ telle que*

$$f(A) = g(M(A)) \qquad \text{pour tout } A \in \mathcal{M}_{m\times n}(\mathbb{R}), \tag{2.6}$$

où $M(A)$ représente le vecteur ordonné de tous les mineurs d'ordre $1, 2, \ldots, n \wedge m = \min(n, m)$, et

$$\tau(n, m) = \sum_{k=1}^{n\wedge m} \binom{m}{k}\binom{n}{k}.$$

Le résultat central est alors le suivant.

Théorème 4 *[38, Th. 5.10] Soit $f : \mathcal{M}_{m\times n}(\mathbb{R}) \to (-\infty, +\infty]$ une fonction polyconvexe telle qu'il existe une fonction g convexe positive, semi-continue inférieurement et satisfaisant (2.6). Alors la fonctionnelle $I(u) = \int_\Omega f(\nabla u(x))dx$ est (séquentiellement) semi-continue inférieurement pour la topologie faible de $W^{1,n\wedge m}(\Omega, \mathbb{R}^m)$.*

Un des intérêts de cette notion dans le cadre de l'élasticité non linéaire est qu'elle recouvre en particulier les fonctions convexes du déterminant du gradient de déformation. On comprend alors mieux le choix des énergies de Ciarlet-Geymonat qui modélisent le comportement (1.1) quand $\det \xi \to 0^+$ par une fonction convexe du troisième invariant.

Contrairement à la quasiconvexité, la polyconvexité est une propriété "explicite", au sens où on peut exhiber une fonction convexe g satisfaisant (2.6). Il n'y a cependant pas unicité d'une telle décomposition. La polyconvexité est une propriété plus forte que la quasiconvexité et il existe des fonctions quasiconvexes non polyconvexes, comme l'illustre l'exemple de Šveràk [179].

Remarque **3** *[38, Prop 6.2] Soit $f : \mathcal{M}_{m\times n}(\mathbb{R}) \to (-\infty, +\infty]$ une fonction polyconvexe continue, alors f est quasiconvexe.*

Cette notion a permis à Ball [14] de démontrer le théorème suivant relatif au problème de l'élastostatique en grandes déformations.

Théorème 5 *[52, Th. 7.7-1] Soit $W : \Omega \times \mathcal{M}_3(\mathbb{R})_+ \to \mathbb{R}$ une densité d'énergies satisfaisant les propriétés suivantes :*

(a) Polyconvexité : pour presque tout $x \in \Omega$, il existe une fonction convexe $\mathbb{W}(x, \cdot) : \mathcal{M}_3(\mathbb{R}) \times \mathcal{M}_3(\mathbb{R}) \times]0, +\infty[\to \mathbb{R}$ telle que

$$\mathbb{W}(x, F, \mathrm{Cof}F, \det F) = W(x, F) \quad \text{pour tout } F \in \mathcal{M}_3(\mathbb{R})_+;$$

la fonction $\mathbb{W}(\cdot, F, H, \delta) : \Omega \to \mathbb{R}$ est mesurable pour tout $(F, H, \delta) \in \mathcal{M}_3(\mathbb{R}) \times \mathcal{M}_3(\mathbb{R}) \times]0, +\infty[$.

(b) Comportement quand $\det F \to 0^+$: pour presque tout $x \in \Omega$,

$$\lim_{\det F \to 0^+} W(x, F) = +\infty.$$

(c) Coercivité : il existe des constantes α, β, p, q, r telles que

$$\begin{aligned} &\alpha > 0, p \geq 2, q \geq \frac{p}{p-1}, r > 1, \\ &W(x, F) \geq \alpha(|F|^p + |\mathrm{Cof}F|^q + (\det F)^r) + \beta \\ &\qquad \text{pour preque tout } x \in \Omega \text{ et pour tout } F \in \mathcal{M}_3(\mathbb{R})_+. \end{aligned}$$

Soit $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ *une partition de la frontière de* Ω *telle que* $\mathcal{L}^2(\partial\Omega_D) > 0$ *et soit* $\phi_0 : \partial\Omega_D \to \mathbb{R}^3$ *une fonction mesurable telle que l'ensemble*

$$\begin{aligned}\mathbf{\Phi} = \big\{&\psi \in W^{1,p}(\Omega), \mathrm{Cof}\nabla\psi \in L^q(\Omega), \det\nabla\psi \in L^r(\Omega),\\ &\psi = \phi_0 \ d\mathcal{L}^2 - \textit{presque partout sur } \partial\Omega_D, \det\nabla\psi > 0 \textit{ presque partout sur } \Omega\big\}\end{aligned} \tag{2.7}$$

n'est pas vide. Soient $f \in L^p(\Omega)$ *et* $g \in L^\sigma(\partial\Omega_N)$ *tels que la forme linéaire*

$$L : \psi \in W^{1,p}(\Omega) \mapsto L(\psi) = \int_\Omega f\psi + \int_{\partial\Omega_N} g\psi$$

est continue. Soit

$$I(\psi) = \int_\Omega W(x, \nabla\psi(x))dx - L(\psi),$$

et supposons que $\inf_{\psi\in\mathbf{\Psi}} I(\psi) < +\infty$. *Alors il existe au moins une fonction* ϕ *telle que*

$$\phi \in \mathbf{\Phi} \textit{ et } I(\phi) = \inf_{\psi\in\mathbf{\Phi}} I(\psi).$$

Ce résultat d'existence de minimiseurs permet donc de résoudre le problème de l'élastostatique pour des densités d'énergie hyperélastiques de type Ciarlet-Geymonat. Il peut être étendu au cas incompressible.

Les problèmes ouverts relatifs à ce résultat concernent notamment la régularité des minimiseurs obtenus.

On peut par ailleurs noter que la densité d'énergie de Saint Venant-Kirchhoff n'est pas polyconvexe [165]. Elle n'est pas non plus quasiconvexe. Cependant, son enveloppe quasiconvexe est explicite comme l'ont démontré Le Dret et Raoult [121].

2.1.6 Γ-convergence, application aux fonctionnelles intégrales et à la dérivation de modèles

Définition et application aux fonctionnelles intégrales

Nous renvoyons à [36, 37, 56] pour une introduction à la Γ-convergence. Par la suite, nous ne considérons que des espaces métriques (ou métrisables et séparables) pour lesquels la Γ-convergence est métrisable et séquentielle.

Définition 5 *[37, Th. 2.1] Soient* (X, d) *un espace métrique et* $F_\epsilon, F : X \to [-\infty, +\infty]$. *La suite* F_ϵ $\Gamma(d)$*-converge vers* F *au point* $x \in X$, *si et seulement si l'une des deux conditions suivantes est satisfaite*

(a) $F(x) = \inf\{\liminf_{\epsilon\to 0} F_\epsilon(x_\epsilon) : x_\epsilon \to x\} = \inf\{\limsup_{\epsilon\to 0} F_\epsilon(x_\epsilon) : x_\epsilon \to x\}$,

(b) (i) (inégalité de la liminf) pour toute suite $x_\epsilon \to x$,

$$F(x) \le \liminf_{\epsilon\to 0} F_\epsilon(x_\epsilon),$$

(ii) (inégalité de la limsup) il existe une suite $x_\epsilon \to x$ *telle que*

$$F(x) \ge \limsup_{\epsilon\to 0} F_\epsilon(x_\epsilon).$$

La suite F_ϵ Γ*-converge vers* F *si* $F_\epsilon(x)$ Γ*-converge vers* $F(x)$ *pour tout* $x \in X$.

Une propriété immédiate de la Γ-convergence est la stabilité par rapport aux perturbations continues pour la métrique d.

On définit également la $\Gamma - \liminf$ et la $\Gamma - \limsup$ par

$$\Gamma{-}\liminf F_\epsilon(x) = \inf\{\liminf_{\epsilon\to 0} F_\epsilon(x_\epsilon) : x_\epsilon \to x\}, \quad \Gamma{-}\limsup F_\epsilon(x) = \inf\{\limsup_{\epsilon\to 0} F_\epsilon(x_\epsilon) : x_\epsilon \to x\}.$$

Ces deux quantités existent *a priori*, et la suite $F_\epsilon(x)$ Γ-converge vers $F(x)$ si et seulement si $\Gamma - \liminf F_\epsilon(x) = \Gamma - \limsup F_\epsilon(x) = F(x)$.

La classe des fonctions semicontinues inférieurement est stable par Γ-convergence :

Proposition 3 *[37, Prop. 2.4] La Γ-liminf et la Γ-limsup d'une suite F_ϵ sont d-semicontinues inférieurement.*

Le théorème fondamental de la Γ-convergence, qui se déduit aisément de la Définition 5, est le suivant.

Théorème 6 *[37, Th. 2.10] Soient (X, d) un espace métrique et (F_ϵ) une suite équi-coercive de fonctions sur X. Soit $F = \Gamma - \lim_{\epsilon\to 0} F_\epsilon$, alors*

$$\exists \min_X F = \lim_{\epsilon\to 0} \inf_X F_\epsilon.$$

De plus, si (x_ϵ) est une suite précompacte telle que $\lim_{\epsilon\to 0} F_\epsilon(x_\epsilon) = \lim_{\epsilon\to 0} \inf_X F_\epsilon$, alors toute valeur d'adhérence de (x_ϵ) est un point de minimum de F.

Au chapitre 3, Théorème 18, nous verrons que la classe des fonctionnelles intégrales satisfaisant une condition de croissance standard (2.5) est compacte pour la $\Gamma(L^p)$-convergence. Des propriétés utiles et plus spécifiques seront rappelées au chapitre 6.

Le Théorème 6 permet de formaliser le concept de convergence variationnelle introduit au paragraphe 1.2.1. La Γ-convergence est l'outil principal utilisé d'une part aux chapitre 6 et 7 pour démontrer la convergence des méthodes d'homogénéisation numérique, et d'autre part aux chapitres 9 et 10 pour dériver des modéles continus à partir de modèles discrets. Au paragraphe suivant, nous formalisons le lien entre élasticité linéaire et élasticité non linéaire en termes de Γ-convergence.

Dérivation de l'élasticité linéaire

Considérons un corps élastique à l'équilibre naturel sur Ω. Imposons alors une force volumique $\tilde{f}_\epsilon$ au système. L'énergie s'écrit pour une déformation u :

$$E_\epsilon(u) = \int_\Omega W(\mathrm{I} + \nabla u) - \int_\Omega \tilde{f}_\epsilon u.$$

Supposons que $\tilde{f}_\epsilon = \epsilon f$ avec ϵ petit, on s'attend alors à ce que l'énergie $\inf E_\epsilon$ soit proche de l'énergie de l'élasticité linéaire, et que les minimiseurs u_ϵ puissent s'écrire $u_\epsilon \simeq \epsilon u$. Avec cette mise à l'échelle, il convient d'étudier la convergence de la fonctionnelle d'énergie

$$F_\epsilon : W^{1,p}(\Omega) \ni u \mapsto \frac{1}{\epsilon^2}\left(\int_\Omega W(\mathrm{I} + \epsilon\nabla u) - \epsilon^2 \int_\Omega f u\right).$$

A u fixé, un développement limité (cf. Théorème 3) montre que

$$\lim_{\epsilon\to 0} F_\epsilon(u) = \frac{1}{2}\int_\Omega Ae(u) : e(u) - \int_\Omega f u := F_0(u),$$

où $A = \frac{\partial^2 W}{\partial \xi^2}(\mathrm{I})$. Cependant, ceci ne donne aucune information sur le comportement des minimiseurs de F_ϵ, ni sur la convergence des infima d'énergie. Le théorème suivant énonce la Γ-convergence de F_ϵ vers F_0 sous des hypothèses compatibles avec les modèles présentés précédemment, et permet ainsi de rendre rigoureux le lien entre élasticité linéaire en petites déformations et élasticité non linéaire en grandes déformations. Ce résultat récent [60] est basé sur une version quantitative du théorème de Liouville due à Friesecke, James et Müller [89].

Théorème 7 *[60, Th. 2.1] Soit $W : \mathcal{M}_3(\mathbb{R}) \mapsto \mathbb{R}^+$ une densité d'énergie régulière objective telle que I est un état d'équilibre naturel ($W(I) = 0$ et $\partial_\xi W(I) = 0$), coercive et à croissance superlinéaire à l'infini, et satisfaisant une propriété de croissance standard d'ordre 2 sur un voisinage de l'identité. Soit alors $\partial\Omega_1$ une partie de la frontière de Ω de mesure non nulle où sont imposées des conditions de Dirichlet, $f \in L^2(\Omega)$ une force volumique, $g \in W^{1,\infty}(\Omega)$ un relèvement de la condition aux limites, et $H^1_{\partial\Omega_1} = \{u \in H^1(\Omega), u_{|\partial\Omega_1} = g_{|\partial\Omega_1}\}$. Si $(u_\epsilon) \in H^1_{\partial\Omega_1}$ est telle que*

$$F_0(u_\epsilon) = \inf_{H^1_{\partial\Omega_1}} F_\epsilon(v) + o(1),$$

alors $u_\epsilon \rightharpoonup u_0$ *dans* $H^1(\Omega)$*, où* u_0 *est l'unique minimiseur de* F_0 *sur* $H^1_{\partial\Omega_1}$*, et* $F_\epsilon(u_\epsilon) \to F_0(u_0)$*.*

On obtient bien ainsi la convergence de l'énergie et des minimiseurs quand la force volumique (ϵf) et la déplacement aux bords (ϵg) sont petits.

2.2 Méthodes numériques en élasticité

Cette section s'inspire des traités de Ciarlet [52], et Le Tallec [124, 125], ainsi que de la thèse de Mouro [144]. Elle constitue le pendant numérique de l'étude théorique du problème de l'élastostatique : étant donnée une solution du problème d'élastostatique, comment en obtenir une approximation ?

2.2.1 Résultat d'approximation et estimation d'erreur

L'analyse numérique des problèmes d'élasticité non linéaire est loin d'être complète. Le Tallec a démontré en 1981 [123] un résultat d'approximation, autrement dit la convergence de la méthode de Galerkin, en suivant le schéma de preuve de Ball pour l'existence de solutions.

Théorème 8 *[125, Th. 16.1] Avec les notations et sous les hypothèses du Théorème 3, on considère* $\phi \in \mathbf{\Phi}$ *une solution isolée du problème de minimisation* $\inf_{\psi\in\mathbf{\Phi}} I(\psi)$*, pour laquelle il existe* $r_0 > 0$ *tel que* $I(\psi) > I(\phi)$ *pour tout* $\psi \in \mathbf{\Phi} \cap B_{p,q}(\phi, r_0)$ *où la distance* $d_{p,q}$ *est donnée par*

$$d_{p,q}(\psi_1, \psi_2) = \|\psi_1 - \psi_2\|_{1,p} + \|\mathrm{cof}(\mathrm{Id} + \nabla\psi_1) - \mathrm{cof}(\mathrm{Id} + \nabla\psi_2)\|_{0,q}, \quad \textit{pour tout } \psi_1, \psi_2 \in \mathbf{\Phi},$$

et $B_{p,q}(\phi, r_0)$ *est la boule de centre* ϕ *et de rayon* r_0 *pour la distance* $d_{p,q}$*. Soit alors* $\mathbf{\Phi}_h$ *une famille de sous-espaces vectoriels de* $\mathbf{\Phi}$ *de dimension finie pour laquelle il existe* $\tilde{\psi}_h \in \mathbf{\Phi}_h \cap B_{p,q}(\phi, r_0)$ *telle que*

$$\lim_{h\to 0} I(\tilde{\psi}_h) = I(\phi).$$

Alors on peut construire une suite ϕ_h *de solutions du problème discret* $\inf_{\psi_h\in\mathbf{\Phi}_h} I(\psi_h)$ *convergeant fortement vers* ϕ *au sens*

$$\lim_{h\to 0} d_{p,q}(\phi, \phi_h) = 0.$$

Un résultat similaire au Théorème 8 existe dans le cas d'énergies incompressibles, voir [125, Th. 15.1].

Il faut attendre 2004 [47] avant d'avoir une première estimation d'erreur a priori pour un problème d'élasticité non linéaire en petites déformations pour une énergie polyconvexe. En substance, le résultat est le suivant.

Théorème 9 *[47, Th. 4.3] Sous des hypothèses de régularité de* Ω*, de la densité d'énergie polyconvexe homogène* W *et de petits déplacements imposés sur la frontière* $\partial\Omega$ *(conditions de Dirichlet non homogènes mais petites), on a l'estimation d'erreur suivante entre la solution* $\phi \in W^{2,r}(\Omega)$ *du problème de l'élastostatique et la solution approchée* $\phi_h \in W^{1,r}(\Omega_h)$ *sur une triangulation régulière de finesse* h *de* Ω_h *(avec* $\mathcal{L}^3(\Omega \setminus \Omega_h) \le Ch^2$ *et* $\mathrm{dist}(z, \partial\Omega) \le Ch^2$ *pour tout noeud* z *du bord de* Ω_h*) du problème approché*

$$\|\nabla\phi - \nabla\phi_h\|^2_{L^2(\Omega)} + \|\nabla\phi - \nabla\phi_h\|^r_{L^r(\Omega)} \le c(h^{4(1-1/r)} + h^2)$$

où la constante c *ne dépend pas de* h*, mais peut dépendre du déplacement et du chargement.*

Nous renvoyons le lecteur à [47] pour le détail des hypothèses. Carstensen et Dolzmann ont basé leur démonstration sur deux faits : l'équivalence avec le système d'équations aux dérivées partielles pour utiliser la régularité de ϕ et la structure algébrique des fonctions polyconvexes. Leur preuve ne s'étend par exemple pas aux fonctions quasiconvexes générales à croissance superlinéaire.

En particulier, pour $r = 2$, le Théorème 9 donne le même ordre de convergence que pour l'élasticité linéaire avec des éléments finis $P1$.

Par ailleurs, les auteurs mentionnent les difficultés suivantes pour l'extension de ce résultat aux grandes déformations : le manque d'informations sur la régularité des solutions, la non unicité des minimiseurs et les bifurcations (et donc la perte de stabilité utilisée dans la démonstration), l'impossibilité numérique de cavitation, le phénomène de Lavrentiev (les minima peuvent être différents sur $W^{1,\sigma_1}(\Omega)$ et $W^{1,\sigma_2}(\Omega)$ pour $\sigma_1 \neq \sigma_2$ alors que l'approximation numérique des deux problèmes de minimisation peut coïncider). L'analyse numérique est donc limitée par l'analyse des propriétés des solutions.

De nouveaux travaux ont vu le jour plus récemment encore, avec l'application des méthodes de Galerkin discontinues à l'élasticité non linéaire, notamment par le groupe de Süli à Oxford. Des estimations d'erreur sont alors obtenues en exploitant la stricte rang-1 convexité de la densité d'énergie (ou stricte stabilité au sens de Legendre-Hadamard) moyennant des hypothèses de régularité plus fortes que celles de [47].

Au chapitre 5, nous démontrons un résultat d'approximation similaire au Théorème 8 dans le cas de l'homogénéisation de fonctionnelles d'énergies (Théorème 34) et faisons une analyse d'erreur a priori dans le cas particulier d'une densité d'énergie strictement convexe (Théorèmes 35 et 36).

2.2.2 Méthode des éléments finis en élasticité

La méthode des éléments finis est très bien adaptée au problème de l'élastostatique car elle permet une gestion "aisée" des géométries complexes (la problématique de génération de maillages est cependant elle-même très complexe, notamment sous certaines restrictions comme l'utilisation d'éléments hexaédriques). Le Théorème 8 nous assure par ailleurs la convergence de solutions discrètes (à condition de trouver un minimiseur global discret).

Méthode de Newton, de continuation, et résolution des systèmes linéaires

Le problème qu'on résout numériquement n'est pas le problème de minimisation, mais le système d'équations aux dérivées partielles (1.3) en formulation faible sur un espace de dimension finie $\mathbf{\Phi}_h$. Le problème étant non linéaire, des méthodes particulières de résolution du problème discret doivent être utilisées. Par ailleurs, le problème est très sensible à la qualité du calcul de la hessienne. Seule la méthode de Newton-Raphson exacte est assez robuste (comparée aux algorithmes itératifs comme BFGS ou aux méthodes de quasi-Newton). Cette méthode est effectivement utilisée en pratique. Pour un problème de la forme $\mathcal{F}(X) = 0$, elle consiste à résoudre itérativement les systèmes linéaires

$$\mathcal{F}(X^n) + \frac{\mathrm{D}\mathcal{F}(X^n)}{\mathrm{D}(X)}(X^{n+1} - X^n) = 0, \tag{2.8}$$

jusqu'à convergence. Il est donc nécessaire de calculer les dérivées première et seconde de la densité d'énergie par rapport au gradient de déformation. Dans l'implémentation concrète (voir [177] par exemple), on utilise souvent la formulation de l'énergie dépendant des invariants (dont la dérivation est aisée) et on code à part la dérivation des invariants par rapport au gradient de déformation. Le détail de la formulation de la méthode des éléments finis peut se lire dans [125, Sec. 18 et 19] et de l'implémentation dans [177].

A chaque itération de l'algorithme de Newton, il convient de résoudre le problème linéaire (2.8). En élasticité, la matrice de raideur est souvent très mal conditionnée (de l'ordre de 1 à 10^6), ce

qui rend difficile l'utilisation de méthodes itératives (comme le gradient conjugué, ou GMRES) en l'absence de bons préconditionneurs. Les méthodes de résolution utilisées sont alors les méthodes directes, de factorisation LU, ou de Choleski. Ces méthodes sont lourdes à la fois en place de mémoire (les matrices sont construites) et en temps de calcul (la factorisation d'une matrice bande (L, N) requiert NL^2 opérations), mais inévitables. La résolution numérique d'un problème d'élastostatique est donc très coûteuse.

La convergence de la méthode de Newton est quadratique quand la déformation de départ est proche de la solution recherchée. Pour résoudre un problème d'élastostatique avec un chargement important, il est naturel de subdiviser le chargement et de résoudre plusieurs problèmes d'élastostatique avec des chargements plus petits (et donc des déformations plus petites et un état initial plus proche de la solution recherchée). La subdivision a priori du chargement n'est pas forcément efficace, tant des points de vue de la qualité de convergence que de sa vitesse. Une méthode plus automatique et robuste, la continuation (*arc-length continuation*), consiste à considérer l'application $s \mapsto \{\phi(s), \lambda(s)\}$ telle que $\mathcal{F}(\phi(s), \lambda(s)f, \lambda(s)g) = 0$, où $\mathcal{F}$ représente le système d'équations aux dérivées partielles avec le second membre en deuxième variable et la condition aux limites de Neumann en troisième variable. Le paramétrage par λ est alors défini en imposant la relation différentielle $\|\frac{\mathrm{D}\phi}{\mathrm{D}s}\|^2 + |\frac{\mathrm{D}\lambda}{\mathrm{D}s}|^2 = 1$. Ce paramétrage (inconnue intermédiaire) définit le facteur d'échelle de chargement et permet d'avoir une subdivision uniforme en "abscisse curviligne" de la famille de solutions. En particulier, la relation différentielle permet d'approximer $s \mapsto \lambda(s)$. Etant donné $\lambda(s)$, $\phi(s)$ est prédite par une formule d'Euler explicite qui sert d'initial guess à l'algorithme de Newton qui corrige alors la prédiction. L'algorithme est décrit en détail dans [125, Sec. 21] et testé dans [177].

Au chapitre 5, pour traiter la non-linéarité de la loi de constitution homogénéisée et des grands déplacements et grandes déformations, nous introduisons un algorithme de Newton qui couple à la fois les niveaux micro et macro. Il est détaillé Section 5.4.

Formulation mixte pour les matériaux incompressibles

Quand la densité d'énergie est incompressible ou quasi-incompressible (par exemple si le coefficient a de la loi de Ciarlet-Geymonat est grand par rapport à C_1 et C_2), il convient d'introduire un problème mixte en déplacement/pression et d'utiliser des espaces d'éléments finis compatibles, pour éviter le vérouillage numérique. La compatibilité entre les espaces d'éléments finis assure la validité de la condition inf-sup de Ladyzenskaya-Babuška-Brezzi, et ainsi la solvabilité du problème linéarisé mixte discret (voir [125, Sec. 14] ou [44]).

Il peut paraître étrange à première vue d'utiliser une formulation mixte pour les matériaux quasi-incompressibles. On ne remplace en fait pas la quasi-incompressibilité par l'incompressibilité et on ne modifie pas la loi de comportement à la limite quand h tend vers zéro, mais on traite de manière spéciale le terme dépendant du troisième invariant. On projette en effet le gradient de déformation sur un espace d'approximation plus grossier (l'espace d'approximation des pressions) avant de calculer le troisième invariant associé. Le problème linéaire a alors les mêmes propriétés que dans le cas incompressible, et est en particulier inversible. Une autre méthode consiste à "sous-intégrer" la partie dépendant du troisième invariant (utiliser moins de points d'intégration) plutôt que de projeter le gradient de déformation. Sous certaines hypothèses les deux méthodes sont équivalentes.

Concrètement, il y a peu d'éléments finis qui satisfont la condition inf-sup. L'un des plus stables (et couramment utilisé) est l'élément fini hexaédrique isoparamétrique $Q2-P1_{disc}$ à 27 noeuds, qui consiste à approximer le déplacement dans l'espace $Q2$ et la pression dans un espace $P1$ discontinu (voir élément HEXA3QCC dans [177]). Les inconvénients sont le degré du polynôme (alors que les solutions peuvent ne pas être régulières) et surtout l'élément géométrique : tous les domaines ne peuvent pas être maillés en hexaèdres, et il est en général moins difficile de mailler en tétraèdres par la méthode de Delaunay.

2.2.3 Méthodes de décomposition de domaine

Le principe de la décomposition de domaine est le suivant. Considérons l'équation de Poisson sur un domaine borné Lipschitzien Ω, complétée de conditions aux limites de Dirichlet homogènes sur $\partial\Omega$. Soit $\{\Omega_1, \Omega_2\}$ une partition de Ω en domaines Lipschitziens. On note $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$ et on suppose $|\partial\Omega \cap \partial\Omega_i| > 0$ pour $i = 1, 2$. Soit $f \in L^2(\Omega)$, alors

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \tag{2.9}$$

est équivalent au problème couplé

$$\begin{cases} -\Delta u_1 = f & \text{dans } \Omega_1 \\ u_1 = 0 & \text{sur } \partial\Omega_1 \setminus \Gamma \\ u_1 = u_2 & \text{sur } \Gamma \\ \frac{\partial u_1}{\partial n_1} = -\frac{\partial u_2}{\partial n_2} & \text{sur } \Gamma \\ -\Delta u_2 = f & \text{dans } \Omega_2 \\ u_2 = 0 & \text{sur } \partial\Omega_2 \setminus \Gamma \end{cases} \tag{2.10}$$

Seules les méthodes de décomposition de domaine sans recouvrement seront considérées dans la thèse, en particulier l'algorithme de Neumann-Neumann. Elles consistent à résoudre la formulation (2.10) du problème (2.9) de manière itérative. Dans ce paragraphe, nous décrivons une telle méthode pour deux domaines, d'abord au niveau continu, puis au niveau discret. Cette partie reprend assez littéralement l'introduction de [175]. Nous renvoyons également à [163] et [124] pour l'analyse précise des méthodes de décomposition de domaines.

Formulation continue.

La méthode de Neumann-Neumann revient à résoudre le problème couplé (2.10) de la manière itérative suivante. A l'étape $n \geq 0$, résoudre

$$(D_i) \left\{ \begin{array}{ll} -\Delta u_i^{n+1/2} = f & \text{dans } \Omega_i, \\ u_i^{n+1/2} = 0 & \text{sur } \partial\Omega_i \setminus \Gamma, \\ u_i^{n+1/2} = u_\Gamma^n & \text{sur } \Gamma \end{array} \right\}, i = 1, 2$$

$$(N_i) \left\{ \begin{array}{lll} -\Delta \psi_i^{n+1} = 0 & & \text{dans } \Omega_i, \\ \psi_i^{n+1} = 0 & & \text{sur } \partial\Omega_i \setminus \Gamma, \\ \dfrac{\partial \psi_i}{\partial n_i}^{n+1} = \dfrac{\partial u_1}{\partial n_1}^{n+1/2} & + \dfrac{\partial u_2}{\partial n_2}^{n+1/2} & \text{sur } \Gamma, \end{array} \right\}, i = 1, 2 \tag{2.11}$$

$$u_\Gamma^{n+1} = u_\Gamma^n - \theta(\psi_1^{n+1} + \psi_2^{n+1}) \qquad \text{sur } \Gamma,$$

avec θ un paramètre de relaxation dont dépend la convergence.

Formulation discrète.

La formulation discrète de l'algorithme de Neumann-Neumann permet d'interpréter *facilement* la méthode comme la résolution du problème d'interface préconditionné, et de dimensionner efficacement θ. L'interprétation continue correspondante fait intervenir les opérateurs de Steklov-Poincaré, qui seront introduits au chapitre 11. Après discrétisation des équations par éléments finis et avec des notations évidentes, l'équation (2.9) s'écrit

$$Au = f, \quad A = \begin{pmatrix} A_{II}^{(1)} & 0 & A_{I\Gamma}^{(1)} \\ 0 & A_{II}^{(2)} & A_{I\Gamma}^{(2)} \\ A_{\Gamma I}^{(1)} & A_{\Gamma I}^{(2)} & A_{\Gamma\Gamma} \end{pmatrix}, \quad u = \begin{pmatrix} u_I^{(1)} \\ u_I^{(2)} \\ u_\Gamma \end{pmatrix}, \quad f = \begin{pmatrix} f_I^{(1)} \\ f_I^{(2)} \\ f_\Gamma \end{pmatrix}, \tag{2.12}$$

où les degrés de liberté ont été partitionnés en degrés intérieurs de Ω_1 et Ω_2 d'une part et sur Γ d'autre part.

En décomposant $A_{\Gamma\Gamma}$ et f_Γ selon les contributions des domaines Ω_1 et Ω_2,

$$A_{\Gamma\Gamma} = A^{(1)}_{\Gamma\Gamma} + A^{(2)}_{\Gamma\Gamma}, \quad f_\Gamma = f^{(1)}_\Gamma + f^{(2)}_\Gamma,$$

et en utilisant la formule de Green

$$\int_\Gamma \frac{\partial u_i}{\partial n_i}\phi_j ds = \int_{\Omega_i} (\Delta u_i \phi_j + \nabla u_i \nabla \phi_j) dx = \int_{\Omega_i} (-f\phi_j + \nabla u_i \nabla \phi_j) dx,$$

on peut écrire le système (2.10) sous la forme

$$\begin{cases} A^{(1)}_{II} u^{(1)}_I + A^{(1)}_{I\Gamma} u^{(1)}_\Gamma = f^{(1)}_I, \\ u^{(1)}_\Gamma = u^{(2)}_\Gamma = u_\Gamma, \\ (A^{(1)}_{\Gamma I} + A^{(1)}_{\Gamma\Gamma} u^{(1)}_\Gamma - f^{(1)}_\Gamma) = (A^{(2)}_{\Gamma I} + A^{(2)}_{\Gamma\Gamma} u^{(2)}_\Gamma - f^{(2)}_\Gamma) = \lambda_\Gamma, \\ A^{(2)}_{II} u^{(2)}_I + A^{(2)}_{I\Gamma} u^{(1)}_\Gamma = f^{(2)}_I. \end{cases} \tag{2.13}$$

Après factorisation par blocs

$$A = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ A^{(1)}_{\Gamma I} {A^{(1)}_{II}}^{-1} & A^{(2)}_{\Gamma I} {A^{(2)}_{II}}^{-1} & I \end{pmatrix} \begin{pmatrix} A^{(1)}_{II} & 0 & A^{(1)}_{I\Gamma} \\ 0 & A^{(2)}_{II} & A^{(2)}_{I\Gamma} \\ 0 & 0 & S \end{pmatrix},$$

et quelques manipulations algébriques, le système linéaire (2.12) s'écrit

$$\begin{pmatrix} A^{(1)}_{II} & 0 & A^{(1)}_{I\Gamma} \\ 0 & A^{(2)}_{II} & A^{(2)}_{I\Gamma} \\ 0 & 0 & S \end{pmatrix} u = \begin{pmatrix} f^{(1)}_I \\ f^{(2)}_I \\ g_\Gamma \end{pmatrix}, \tag{2.14}$$

où

$$S = S^{(1)} + S^{(2)} = A^{(1)}_{\Gamma\Gamma} - A^{(1)}_{\Gamma I} {A^{(1)}_{II}}^{-1} A^{(1)}_{I\Gamma} + A^{(2)}_{\Gamma\Gamma} - A^{(2)}_{\Gamma I} {A^{(2)}_{II}}^{-1} A^{(2)}_{I\Gamma},$$
$$g_\Gamma = g^{(1)}_\Gamma + g^{(2)}_\Gamma = (f^{(1)}_\Gamma - A^{(1)}_{\Gamma I} {A^{(1)}_{II}}^{-1} f^{(1)}_I) + (f^{(2)}_\Gamma - A^{(2)}_{\Gamma I} {A^{(2)}_{II}}^{-1} f^{(2)}_I).$$

L'équation de l'interface (ou système du complément de Schur) s'écrit alors

$$S u_\Gamma = g_\Gamma. \tag{2.15}$$

Dès que u_Γ est connu, on obtient les composantes intérieures de u par la formule

$$u^{(i)}_I = {A^{(i)}_{II}}^{-1} (f^{(i)}_I - A^{(i)}_{I\Gamma} u_\Gamma), \quad i = 1, 2.$$

En notant les vecteurs de degrés de liberté intérieurs $v_i = u^{(i)}_I$ et $w_i = \phi^{(i)}_I$, on peut écrire l'algorithme de Neumann-Neumann sur le système discret : à l'étape $n \geq 0$, résoudre

$$\begin{cases} (D_i) \quad A^{(i)}_{II} v^{n+1/2}_i + A^{(i)}_{I\Gamma} u^n_\Gamma = f^i_I, \quad i = 1, 2 \\ (N_i) \quad \begin{pmatrix} A^{(i)}_{II} & A^{(i)}_{I\Gamma} \\ A^{(i)}_{\Gamma I} & A^{(i)}_{\Gamma\Gamma} \end{pmatrix} \begin{pmatrix} w^{n+1}_i \\ \eta^{n+1}_i \end{pmatrix} = \begin{pmatrix} 0 \\ r_\Gamma \end{pmatrix}, \quad i = 1, 2 \\ u^{n+1}_\Gamma = u^n_\Gamma - \theta(\eta^{n+1}_1 + \eta^{n+1}_2), \end{cases} \tag{2.16}$$

où le résidu r_Γ est défini comme

$$r_\Gamma = (A^{(1)}_{\Gamma I} v_1^{n+1/2} + A^{(1)}_{\Gamma\Gamma} u^n_\Gamma - f^{(1)}_\Gamma) + (A^{(2)}_{\Gamma I} v_1^{n+1/2} + A^{(2)}_{\Gamma\Gamma} u^n_\Gamma - f^{(2)}_\Gamma).$$

En éliminant les variables $v_i^{n+1/2}$ et w_i^{n+1}, et après quelques manipulations algébriques, on obtient :

$$u_\Gamma^{n+1} - u_\Gamma^n = \theta({S^{(1)}}^{-1} + {S^{(2)}}^{-1})(g_\Gamma - S u_\Gamma^n), \tag{2.17}$$

ce qui montre que l'algorithme de Neumann-Neumann s'interprète comme une méthode de Richardson sur le complément de Schur, préconditionnée par $P = \left[{S^{(1)}}^{-1} + {S^{(2)}}^{-1}\right]^{-1}$. L'inversion de la matrice $S^{(i)}$ correspond à la résolution d'un problème de Neumann sur le domaine Ω_i, d'où le nom de la méthode. Spectralement, le préconditionneur est optimal dans le cas de deux sous-domaines [163, Rem 4.4.1], et $\mathrm{Cond}\left(P^{-1}S\right) = \mathrm{Cond}\left((S^{(1)^{-1}} + S^{(2)^{-1}})(S^{(1)} + S^{(2)})\right) = O(1)$ indépendamment de la finesse du maillage. Dans le cas de plus de deux sous-domaines, la méthode est quasi-optimale [125, Th. 4.8](le conditionnement dépend de $\log \frac{|\Omega_i|}{h}$ où h est la finesse du maillage).

Enfin, plutôt qu'une méthode de Richardson, il est plus efficace d'utiliser une méthode de gradient conjugué (ou GMRES si la matrice de Schur n'est pas symétrique) sur le complément de Schur, ce qui permet de choisir θ de manière automatique à chaque itération et ce qui change légèrement la troisième équation de (2.16). On renvoie à [124, Sec. 3.2.5] pour une version détaillée de l'algorithme.

2.2.4 Discrétisation temporelle pour l'élastodynamique

Dans ce paragraphe, nous présentons une méthode de discrétisation temporelle pour les équations de l'élastodynamique (1.4). Les schémas de type Euler ne sont pas souvent utilisés en dynamique des structures car ils sont trop dissipatifs. Il convient plutôt d'utiliser des schémas linéairement conservatifs comme les schémas de point milieu ou de Newmark, qui conservent l'énergie mécanique en élasticité linéaire. Le schéma se fait en deux étapes. D'abord une étape utilisant un schéma d'Euler implicite sur un demi-pas de temps, puis une extrapolation sur le demi-pas de temps suivant pour corriger l'excès de dissipation du schéma d'Euler. Etant donnés ϕ^n, $\phi^{n+1/2}$, $\dot\phi^n$ et $\dot\phi^{n+1/2}$ aux temps n et $n+1/2$, on définit ϕ^{n+1} comme la solution de

$$\rho\left(\frac{\phi^{n+1} - \phi^{n+1/2}}{\delta t/2} - \dot\phi^{n+1/2}\right)\frac{2}{\delta t} - \mathrm{div}\ \pi(\phi^{n+1}) = f^{n+1},$$

et $\dot\phi^{n+1} = \frac{\phi^{n+1} - \phi^{n+1/2}}{\delta t/2}$. Ensuite, on extrapole comme suit

$$\phi^{n+3/2} = 2\phi^{n+1} - \phi^n \qquad \text{et} \qquad \dot\phi^{n+3/2} = 2\dot\phi^{n+1} - \dot\phi^n.$$

Le schéma ainsi obtenu est stable et précis à l'ordre deux pour des problèmes linéaires.

Nous renvoyons à [126,128,144] pour plus de détails et à [93,100] pour l'étude d'autres schémas.

3

Homogénéisation - approches théoriques et numériques

3.1 Théorie de l'homogénéisation

Cette section rassemble quelques résultats d'homogénéisation des opérateurs elliptiques. On introduit différentes techniques de preuve, la H-convergence, la Γ-convergence et la convergence à deux échelles. Chaque technique a ses spécifités et ses intérêts propres, comme la proximité à l'intuition, la généralité, la concision. Même si seule l'homogénéisation des intégrales multiples est utilisée aux chapitres 6 et 7, les techniques de G-convergence et de convergence à deux échelles ont été originellement utilisées pour démontrer les résultats d'analyse numérique de la Section 3.2. L'homogénéisation a une longue histoire mathématique. Nous prenons le parti de présenter les résultats pour eux-mêmes, souvent hors de leur contexte historique. La littérature est trop vaste et les contributions trop nombreuses pour donner en quelques pages sa juste place à chacun.

Cette synthèse s'inspire des articles et ouvrages de Murat et Tartar [149], Bensoussan, Lions et Papanicolaou [21], Braides [33], Müller [145], Nguetseng [150], Allaire [6, 7], Braides et Defranceschi [38], Cioranescu et Donato [54], et Jikov, Kozlov et Oleinik [113].

L'homogénéisation par Γ-convergence et par H-convergence sont cousines. On peut dire que la première se focalise sur l'énergie tandis que la seconde se focalise sur l'opérateur différentiel. Ainsi, si on considère des énergies dont les équations d'Euler-Lagrange sont de type monotone, alors les deux notions sont équivalentes (de même que G- et H-convergences dans ce cas). En revanche, si on considère des énergies quasiconvexes ou des équations aux dérivées partielles qui ne sont pas issues d'un principe variationnel alors seulement l'une des deux méthodes s'applique (Γ-convergence dans le premier cas, H-convergence dans le second). La convergence à deux échelles, quant à elle, est une technique qui peut être utilisée dans les deux contextes de Γ- et H-convergence.

Par souci de clarté et en référence aux problèmes d'élasticité, nous présentons les techniques de preuve en dimension 3, cas scalaire d'abord, puis vectoriel pour la Γ-convergence. Le lettre Ω désigne toujours un domaine (ouvert borné connexe) de $\mathbb{R}^3$.

3.1.1 Développement formel à échelles multiples

La méthode du développement formel permet de comprendre assez intuitivement les enjeux de l'homogénéisation. Elle est présentée dans le cadre périodique.

Soit $Y = (0,1)^3$ et $M(\alpha, \beta, Y)$ l'ensemble des matrices $A : \mathbb{R}^3 \to \mathcal{M}_3(\mathbb{R})$ Y-périodiques telles que pour presque tout $x \in Y$ et pour tout $\xi \in \mathbb{R}^3$,

$$\begin{cases} \langle A(x)\xi, \xi\rangle \geq \alpha|\xi|^2 \\ |A(x)\xi| \leq \beta|\xi| \end{cases}.$$

Etant donnée une matrice Y-périodique $A \in M(\alpha, \beta, Y)$, l'opérateur différentiel associé $\mathcal{A}_\epsilon = -\mathrm{div}\, A(\frac{\cdot}{\epsilon})\nabla$ et un second membre $f \in L^2(\Omega)$, le problème de base qu'on veut résoudre est la détermination du comportement asymptotique du problème linéaire suivant

$$\begin{cases} \mathcal{A}_\epsilon u_\epsilon = f & \text{dans } \Omega \\ u_\epsilon = 0 & \text{sur } \partial\Omega \end{cases} \tag{3.1}$$

quand ϵ tend vers zéro.

En physique et en mécanique, on introduit alors naturellement une solution de (3.1) sous la forme d'un développement asymptotique à deux échelles

$$u_\epsilon(x) = u_0\left(x, \frac{x}{\epsilon}\right) + \epsilon u_1\left(x, \frac{x}{\epsilon}\right) + \epsilon^2 u_2\left(x, \frac{x}{\epsilon}\right) + \dots, \tag{3.2}$$

où tous les termes sont supposés réguliers et Y-périodiques en la deuxième variable.

La première étape formelle consiste à insérer ce développement dans (3.1) en utilisant la définition suivante : soit $\phi = \phi(x, y)$ une fonction de deux variables de $\mathbb{R}^3$, on note ϕ_ϵ la fonction de une variable définie par

$$\phi_\epsilon(x) = \phi\left(x, \frac{x}{\epsilon}\right).$$

La règle de dérivation implique

$$\frac{\partial \phi_\epsilon}{\partial x} = \frac{\partial \phi}{\partial x_i}\left(x, \frac{x}{\epsilon}\right) + \frac{1}{\epsilon}\frac{\partial \phi}{\partial y_i}\left(x, \frac{x}{\epsilon}\right).$$

A cette fin, on introduit les opérateurs différentiels suivants

$$\begin{cases} \mathcal{A}_0 = -\displaystyle\sum_{i,j=1}^{3} \frac{\partial}{\partial y_i}\left(a_{ij}(y)\frac{\partial}{\partial y_j}\right) \\ \mathcal{A}_1 = -\displaystyle\sum_{i,j=1}^{3} \frac{\partial}{\partial x_i}\left(a_{ij}(y)\frac{\partial}{\partial y_j}\right) - \sum_{i,j=1}^{3} \frac{\partial}{\partial y_i}\left(a_{ij}(y)\frac{\partial}{\partial x_j}\right) \\ \mathcal{A}_2 = -\displaystyle\sum_{i,j=1}^{3} \frac{\partial}{\partial x_i}\left(a_{ij}(y)\frac{\partial}{\partial x_j}\right) \end{cases}$$

En annulant chaque terme en puissance de ϵ fixée, l'équation (3.1) se récrit alors comme une cascade infinie d'équations, dont les trois premières sont

$$\begin{cases} \mathcal{A}_0 u_0 = 0 & \text{dans } Y \\ u_0 \ Y\text{-périodique en } y \end{cases} \tag{3.3}$$

$$\begin{cases} \mathcal{A}_0 u_1 = -\mathcal{A}_1 u_0 & \text{dans } Y \\ u_1 \ Y\text{-périodique en } y \end{cases} \tag{3.4}$$

$$\begin{cases} \mathcal{A}_0 u_2 = f - \mathcal{A}_1 u_1 - \mathcal{A}_2 u_0 & \text{dans } Y \\ u_2 \ Y\text{-périodique en } y \end{cases} \tag{3.5}$$

L'équation (3.3) implique que u_0 ne dépend que de x et pas de y, c'est la solution du problème homogénéisé et le premier terme du développement. La deuxième équation donne le premier correcteur en fonction du premier terme u_0. En effet, en introduisant $\hat{\chi}_j$ solution de

$$\begin{cases} \mathcal{A}_0 \hat{\chi}_j = \displaystyle\sum_{i=1}^{3} \frac{\partial a_{ij}}{\partial y_i} & \text{dans } Y \\ \hat{\chi}_j \ Y\text{-périodique en } y \end{cases}, \tag{3.6}$$

on obtient $u_1(x, y) = -\sum_{j=1}^{3} \hat{\chi}_j(y)\frac{\partial u_0}{\partial x_j}$. Enfin, l'alernative de Fredholm impose que $\int_Y f - \mathcal{A}_1 u_1 - \mathcal{A}_2 u_0 = 0$ pour que (3.5) ait une solution. Ceci est équivalent à avoir ponctuellement dans Ω

$$\begin{cases} \mathcal{A}_{hom} u_0 = f & \text{dans } \Omega \\ u_0 = 0 & \text{sur } \partial\Omega \end{cases} \tag{3.7}$$

où A_{hom} est donnée par $a_{ij}^{hom} = \int_Y a_{ij} - \sum_{k=1}^3 a_{ik} \frac{\partial \hat{\chi}_j}{\partial y_k}$. En définissant $\hat{\theta}^{kl}$ comme l'unique solution de

$$\begin{cases} \mathcal{A}_0 \hat{\theta}^{kl} = -a_{kl}^0 - \displaystyle\sum_{i,j=1}^3 \frac{\partial a_{ij} \delta_{kj} \hat{\chi}_l}{\partial y_i} - \sum_{j=1}^3 a_{kj} \frac{\partial (\hat{\chi}_l - y_l)}{\partial y_i} & \text{dans } Y \\ \hat{\theta}^{kl} \ Y\text{-périodique en } y \end{cases}, \tag{3.8}$$

on obtient $u_2(x,y) = \sum_{k,l=1}^3 \hat{\theta}^{kl}(y) \frac{\partial^2 u_0}{\partial x_k \partial x_l}$.

Reste encore à justifier le développement. Sous des hypothèses de régularité, on a le résultat suivant

Théorème 10 *[54, Th. 6.3] En supposant Ω régulier, A de classe $\mathcal{C}^\infty$ et f régulière, toutes les manipulations formelles sont licites et on a l'estimation d'erreur suivante*

$$\left\| u_\epsilon - \left(u_0 + \epsilon u_1 \left(x, \frac{x}{\epsilon} \right) + \epsilon^2 u_2 \left(x, \frac{x}{\epsilon} \right) \right) \right\|_{H^1(\Omega)} \leq C\sqrt{\epsilon}. \tag{3.9}$$

Dans les cas moins réguliers, le développement n'est en général pas valable à l'ordre deux et d'autres techniques de preuve sont nécessaires.

3.1.2 Homogénéisation périodique et fonctions oscillantes de Tartar

La méthode des fonctions oscillantes de Tartar est un moyen de justifier les deux premiers termes du développement asymptotique formel (3.2). Cette méthode consiste à introduire le problème de cellule adjoint. En choisissant alors des fonctions tests spéciales, l'addition des deux formulations donne lieu à la compensation de deux termes composés du produit de deux suites faiblement convergentes, qu'on ne maîtrise pas *a priori*. C'est un exemple de compacité par compensation. Le problème de cellule adjoint est le suivant : on définit χ_j comme l'unique solution périodique dans $H^1_\#(Y)/\mathbb{R}$ de

$$\begin{cases} -\text{div } A^T \chi_j = \displaystyle\sum_{i=1}^3 \frac{\partial a_{ij}}{\partial y_i} & \text{dans } Y \\ \chi_j \ Y\text{-périodique en } y \end{cases}. \tag{3.10}$$

Pour tout $\xi \in \mathbb{R}^3$ on introduit la fonction $w_\xi \in H^1(Y)$ définie par

$$w_\xi(y) = \xi \cdot y + \sum_{j=1}^3 \chi_j \xi_j.$$

Le résultat de convergence s'énonce alors ainsi :

Théorème 11 *[7, Th. 1.3.18] [54, Th. 6.1] Soit $f \in H^{-1}(\Omega)$ et u_ϵ la solution de* (3.1) *pour $A \in M(\alpha, \beta, Y)$. Alors*

$$\begin{cases} i) \ \ u_\epsilon \rightharpoonup u_0 & \textit{faiblement dans } H_0^1(\Omega) \\ ii) \ A_\epsilon \nabla u_\epsilon \rightharpoonup A_{hom} \nabla u_0 & \textit{faiblement dans } (L^2(\Omega))^3 \end{cases} \tag{3.11}$$

où u_0 est l'unique solution dans $H_0^1(\Omega)$ du problème homogénéisé (3.7) *pour lequel les coefficients sont donnés de manière équivalente par*

$$(A_{hom})^T \xi = \int_Y A^T \nabla w_\xi.$$

De plus, $A_{hom} \in M(\alpha, \frac{\beta^2}{\alpha}, Y)$.

Ce théorème se démontre comme suit. Les estimations *a priori* montrent qu'il existe $u^0 \in H_0^1(\Omega)$ et $\lambda_0 \in L^2(\Omega)^3$ telles que, à extraction près,

$$\begin{cases} i) \quad u_\epsilon \rightharpoonup u^0 & \text{dans } H^1(\Omega) \\ ii) \quad u_\epsilon \to u^0 & \text{dans } L^2(\Omega) \\ iii) \ \lambda_\epsilon \rightharpoonup \lambda_0 & \text{dans } (L^2(\Omega))^3 \end{cases}$$

où $\lambda_\epsilon = A_\epsilon \nabla u_\epsilon$ et satisfait

$$\int_\Omega \lambda_\epsilon \cdot \nabla v = \langle f, v \rangle, \qquad \forall v \in H_0^1(\Omega). \tag{3.12}$$

Il suffit de démontrer $u^0 = u_0$ et $\lambda_0 = A_{hom} \nabla u_0$.

On pose $w_\xi^\epsilon = \epsilon w_\xi(\frac{\cdot}{\epsilon})$. Cette fonction oscille à l'échelle ϵ et vérifie

$$\begin{cases} i) \quad w_\xi^\epsilon \rightharpoonup \xi \cdot x & \text{dans } H^1(\Omega) \\ ii) \ w_\xi^\epsilon \to \xi \cdot x & \text{dans } L^2(\Omega) \end{cases}$$

par propriété des fonctions périodiques (convergence faible vers la moyenne).

Pour tout $\xi \in \mathbb{R}^3$, on introduit alors la fonction $\eta_\xi^\epsilon = A_\epsilon^T \nabla w_\xi^\epsilon$. Par définition de w_ξ (en utilisant en particulier la périodicité),

$$\int_\Omega \eta_\xi^\epsilon \cdot \nabla v = 0, \qquad \forall v \in H_0^1(\Omega). \tag{3.13}$$

Par ailleurs, cette fonction est ϵY-périodique, elle satisfait donc $\eta_\xi^\epsilon \rightharpoonup \int_Y A^T \nabla w_\xi = (A_{hom})^T \xi$.

Soit $\phi \in \mathcal{D}(\Omega)$, on prend alors comme fonction test (oscillante) ϕw_ξ^ϵ dans (3.12) et ϕu_ϵ dans (3.13). Le choix du problème adjoint fait que les premiers termes des deux problèmes variationnels sont les mêmes. En les soustrayant, on obtient

$$\int_\Omega \lambda_\epsilon \cdot (\nabla \phi)\, w_\xi^\epsilon - \eta_\xi^\epsilon \cdot (\nabla \phi)\, u_\epsilon = \langle f, \phi w_\xi^\epsilon \rangle, \qquad \forall \phi \in \mathcal{D}(\Omega).$$

On peut alors passer à la limite dans chaque terme par produit de convergences forte/faible. On obtient d'abord $\lambda_0 = (A_{hom})^T \nabla u^0$, puis $\int_\Omega (A_{hom})^T \nabla u^0 \nabla v = \int_\Omega \lambda_0 \nabla v = \langle f, v \rangle$ pour tout $v \in H_0^1(\Omega)$, ce qui démontre les résultats du Théorème 11.

La justification du deuxième terme du développement est le résultat dit du correcteur. Il s'énonce ainsi

Théorème 12 *[54, Th. 8.6 et Rem 8.8] Soit $C_{ij}^\epsilon : \Omega \ni x \mapsto C_{ij}^\epsilon(x) = \delta_{ij} + \frac{\partial \hat{\chi}_j}{\partial y_i}(\frac{x}{\epsilon})$, où χ_j est définie par* (3.6)*. Alors*

– *$C^\epsilon \rightharpoonup Id$ faiblement dans $L^2(\Omega)^{3\times 3}$,*

– *$A^\epsilon C^\epsilon \rightharpoonup A_{hom}$ faiblement dans $L^2(\Omega)^3$.*

De plus,

$$\nabla u_\epsilon - C^\epsilon \nabla u_0 \to 0 \text{ fortement dans } L^2(\Omega).$$

3.1.3 Homogénéisation périodique par convergence à deux échelles

La convergence à deux échelles, introduite plus de 10 ans après la méthode de Tartar (Nguetseng [150] et Allaire [6]), fait une synthèse entre les développements asymptotiques et la méthode des fonctions oscillantes. Par définition, la convergence à deux échelles introduit deux échelles x et $\frac{x}{\epsilon}$ comme dans le développement asymptotique. Cette méthode utilise également des fonctions tests oscillantes, comme la méthode de Tartar. Cependant, celles-ci ne sont pas spécifiques à l'opérateur $\mathcal{A}_\epsilon$.

Définition 6 *[54, Th. 9.3] Soit v_ϵ une suite de fonctions dans $L^2(\Omega)$. On dit que v_ϵ converge à deux échelles vers $v_0 = v_0(x,y)$ avec $v_0 \in L^2(\Omega \times Y)$ si pour toute fonction $\psi = \psi(x,y) \in \mathcal{D}(\Omega, \mathcal{C}^\infty_\#(Y))$*

$$\lim_{\epsilon \to 0} \int_\Omega v_\epsilon(x)\psi\left(x, \frac{x}{\epsilon}\right) dx = \frac{1}{|Y|} \int_\Omega \int_Y v_0(x,y)\psi(x,y)dydx \tag{3.14}$$

Le lien entre le développement asymptotique formel et la convergence à deux échelles est le suivant : si u_ϵ admet un développement de la forme (3.2), alors u_ϵ converge à deux échelles vers u_0. Ceci permet de justifier *a posteriori* le développement.

Les trois théorèmes suivants donnent les propriétés majeures de la convergence à deux échelles : la compacité dans $L^2(\Omega)$, le comportement du produit deux suites qui convergent à deux échelles (ce qui explicite en fait le produit de deux suites faiblement convergentes) et une caractérisation de la compacité dans $H^1(\Omega)$.

Théorème 13 *[54, Th. 9.7] Soit v_ϵ une suite bornée dans $L^2(\Omega)$. Alors il existe $v_0 \in L^2(\Omega \times Y)$ telle que, à extraction près, v_ϵ converge à deux échelles vers v_0.*

Théorème 14 *[54, Th. 9.8] Soit v_ϵ une suite de $L^2(\Omega)$ qui converge à deux échelles vers $v_0 \in L^2(\Omega \times Y)$. Si de plus*

$$\lim_{\epsilon \to 0} \int_\Omega [v_\epsilon(x)]^2 dx = \frac{1}{|Y|} \int_\Omega \int_Y [v_0(x,y)]^2 dydx \tag{3.15}$$

alors, quelle que soit w_ϵ convergeant à deux échelles vers une limite $w_0 \in L^2(\Omega \times Y)$, on a

$$v_\epsilon w_\epsilon \to \frac{1}{|Y|} \int_Y v_0(\cdot, y) w_0(\cdot, y) dy \qquad \text{dans } \mathcal{D}'(\Omega). \tag{3.16}$$

Théorème 15 *[54, Th. 9.9] Soit v_ϵ une suite de fonctions de $H^1(\Omega)$ telle que*

$$v_\epsilon \rightharpoonup v_0 \qquad \text{dans } H^1(\Omega).$$

Alors v_ϵ converge à deux échelles vers v_0 et il existe une fonction $v_1(x,y) \in L^2(\Omega, H^1_\#(Y))$ telle qu'à extraction près,

$$\nabla v_\epsilon \qquad \text{converge à deux échelles vers} \qquad \nabla_x v_0 + \nabla_y v_1.$$

Les espaces fonctionnels introduits et étudiés, la justification des deux premiers termes du développement asymptotique est directe et naturelle. Les grandes étapes de la démonstration du Théorème 11 sont les suivantes.

1) Les estimations a priori permettent d'obtenir la convergence faible dans $H^1(\Omega)$ de u_ϵ vers une fonction $u^0 \in H^1(\Omega)$ à extraction près ;
2) Quitte à extraire une autre sous-suite, le Théorème 15 montre qu'il existe une fonction $u_1 \in L^2(\Omega, H^1_\#(Y))$ telle que ∇u_ϵ converge à deux échelles vers $\nabla_x u_0 + \nabla_y u_1$;
3) Identification de la limite : $u^0 = u_0$. Pour cela on utilise comme fonction test dans la formulation faible de 3.1 $v_0 \in \mathcal{D}(\Omega)$ et $v_1 \in \mathcal{D}(\Omega, \mathcal{C}^\infty_\#(Y))$:

$$\int_\Omega A_\epsilon \nabla u_\epsilon \left[\nabla v_0(x) + \epsilon \nabla_x v_1\left(x, \frac{x}{\epsilon}\right) + \nabla_y v_1\left(x, \frac{x}{\epsilon}\right)\right] dx = \left\langle f, v_0(\cdot) + \epsilon v_1\left(\cdot, \frac{\cdot}{\epsilon}\right)\right\rangle_{H^{-1}, H^1_0};$$

4) En remarquant que $A_\epsilon^T[\nabla v_0 + \nabla_y v_1(x, \frac{x}{\epsilon})]$ est une fonction test pour la convergence à deux échelles, on peut passer à la limite dans chacun des termes ;
5) Il convient enfin de reconnaître dans l'équation limite le problème homogénéisé.

Le théorème de correcteur peut aussi se démontrer en utilisant la méthode de la convergence à deux échelles : ceci revient à démontrer que la convergence à deux échelles du gradient ∇u_ϵ, qui implique sa convergence faible, implique aussi sa convergence forte (voir [54, Th. 9.12]).

La convergence à deux échelles peut aussi s'utiliser dans d'autres contextes que celui des espaces de Lebesgue. La méthode a été notamment étendue au cas de la convergence des mesures dans [32].

3.1.4 Homogénéisation des opérateurs elliptiques linéaires par H-convergence

Dans cette section, nous présentons la généralisation des fonctions oscillantes de Tartar au cas non périodique. La compacité par compensation reste ici encore au centre de la démonstration mais les fonctions oscillantes ne sont plus si explicites.

Définition 7 *[149, Def 1] Une suite A_ϵ d'éléments de $M(\alpha, \beta, \Omega)$ H-converge vers un élément A_{hom} de $M(\alpha', \beta', \Omega)$ si et seulement si, pour tout $\omega \subset\subset \Omega$ et pour tout $f \in H^{-1}(\omega)$, la solution u_ϵ de*

$$\begin{cases} -\text{div}\,(A_\epsilon \nabla u_\epsilon) = f & \text{dans } \omega, \\ u_\epsilon \in H_0^1(\omega), \end{cases} \tag{3.17}$$

est telle que

$$\begin{cases} u_\epsilon \rightharpoonup u_{hom} & \text{faiblement dans } H_0^1(\omega), \\ A_\epsilon \nabla u_\epsilon \rightharpoonup A_{hom} \nabla u_{hom} & \text{faiblement dans } \in L^2(\omega)^3, \end{cases} \tag{3.18}$$

où u_{hom} est solution de

$$\begin{cases} -\text{div}\,(A_{hom} \nabla u_{hom}) = f & \text{dans } \omega, \\ u_{hom} \in H_0^1(\omega). \end{cases}$$

Remarque **4** *Si la suite A_ϵ satisfait $A_\epsilon = A_\epsilon^T$ alors A_{hom} est symétrique et la deuxième convergence de* (3.18) *est une conséquence du reste de la Définition 7. Dans ce cas, la H-convergence coïncide avec la G-convergence. La H-convergence généralise la G-convergence au cas des matrices non symétriques, et plus généralement aux équations aux dérivées partielles ne correspondant pas aux équations d'Euler-Lagrange de la minimisation d'une fonctionnelle d'énergie.*

L'introduction du problème adjoint et l'utilisation de la compacité par compensation permet de démontrer le résultat fondamental suivant de compacité séquentielle.

Théorème 16 *[149, Th. 2] Soit A_ϵ une suite de $M(\alpha, \beta, \Omega)$. Il existe une matrice $A_{hom} \in M(\alpha, \frac{\beta^2}{\alpha}, \Omega)$ telle que, à extraction près, A_ϵ H-converge vers A_{hom}.*

Le Théorème 16 est un exemple de compacité de la suite des opérateurs et des solutions d'équations aux dérivées partielles. Si de plus, la limite (l'opérateur et la solution) ne dépend pas de l'extraction - comme c'est le cas en homogénéisation périodique, alors toute la suite converge. Un autre résultat intéressant concerne les conditions aux limites : la H-convergence est indépendante des conditions aux limites de l'équation aux dérivées partielles, comme le montre la proposition suivante.

Proposition 4 *[149, Th. 1] Supposons que la suite $A_\epsilon \in M(\alpha, \beta, \Omega)$ H-converge vers $A_{hom} \in M(\alpha', \beta', \Omega)$. Si*

$$\begin{cases} u_\epsilon \in H^1(\Omega), \\ f_\epsilon \in H^{-1}(\Omega), \\ -\text{div}\,(A_\epsilon \nabla u_\epsilon) = f_\epsilon & \text{dans } \Omega, \\ u_\epsilon \rightharpoonup u_{hom} & \text{faiblement dans } H^1(\Omega), \\ f_\epsilon \to f_0 & \text{fortement dans } H^{-1}(\Omega). \end{cases}$$

Alors

$$A_\epsilon \nabla u_\epsilon \rightharpoonup A_{hom} \nabla u_{hom} \qquad \text{faiblement dans } L^2(\Omega)^3.$$

Tout comme dans le cas périodique, on peut définir un correcteur.

Définition 8 *[149, Def 2] Soit $A_\epsilon \in M(\alpha, \beta, \Omega)$ une suite qui H-converge vers $A_{hom} \in M(\alpha, \frac{\beta^2}{\alpha}, \Omega)$. On définit la matrice de correcteur $P_\epsilon \in L^2(\omega)^{3\times 3}$ pour tout $\omega \subset\subset \Omega$ par*

$$P_\epsilon \lambda = \nabla w_\epsilon^\lambda, \qquad \lambda \in \mathbb{R}^3, \tag{3.19}$$

où w_ϵ^λ satisfait

$$\begin{cases} w_\epsilon^\lambda \in H^1(\omega), \\ w_\epsilon^\lambda \rightharpoonup \lambda \cdot x & \text{dans } H^1(\omega), \\ -\text{div}\,(A_\epsilon \nabla w_\epsilon^\lambda) \to -\text{div}\,(A_{hom} \lambda) & \text{dans } H^{-1}(\omega). \end{cases}$$

Le résultat de convergence s'énonce alors

Théorème 17 *[149, Th. 3] Supposons que la suite $A_\epsilon \in M(\alpha, \beta, \Omega)$ H-converge vers $A_{hom} \in M(\alpha', \beta', \Omega)$. Si*

$$\begin{cases} u_\epsilon \in H^1(\omega), \\ f_\epsilon \in H^{-1}(\omega), \\ -\mathrm{div}\,(A_\epsilon \nabla u_\epsilon) = f_\epsilon \qquad \textit{dans } \omega, \\ u_\epsilon \rightharpoonup u_{hom} \qquad \textit{faiblement dans } H^1(\omega), \\ f_\epsilon \to f_0 \qquad \textit{fortement dans } H^{-1}(\omega), \end{cases}$$

où $\omega \subset\subset \Omega$. Alors,

$$\begin{cases} \nabla u_\epsilon = P_\epsilon \nabla u_{hom} + z_\epsilon, \\ z_\epsilon \to 0 \qquad \textit{fortement dans } L^1_{loc}(\omega)^3. \end{cases}$$

Moyennant des hypothèses d'intégrabilité (qu'on peut par exemple obtenir en utilisant les estimations de Meyers [134]), on peut démontrer que la convergence de z_ϵ est forte dans $L^p_{loc}(\omega)$ pour $p > 1$.

Aux chapitres 6 et 7, nous démontrons la convergence d'un correcteur *numérique*, qui est une approximation du correcteur introduit par Tartar.

3.1.5 Homogénéisation des intégrales multiples par Γ-convergence

Avant de présenter la méthode à proprement parler, il convient de faire quelques remarques d'ordre général sur la Γ-convergence. La définition de la Γ-convergence est un cadre dans lequel de nombreux résultats des paragraphes précédents peuvent se reformuler. De ce point de vue, la Γ-convergence est un formalisme. Ce point de vue est cependant, à plusieurs titres, très partiel. Cette notion trouve un premier sens quand elle s'applique à une "classe précise" de problèmes. L'exemple fondamental est l'étude de familles de problèmes de minimisation de fonctionnelles intégrales dans les espaces de Sobolev. Il devient alors possible d'étudier plus précisément les topologies associées aux différentes Γ-convergences, d'en démontrer des propriétés abstraites intéressantes (compacité par exemple). Une telle boîte à outils (voir par exemple le traité de Dal Maso [56], ou encore le chapitre 17 de [113] pour les résultats utiles à l'homogénéisation) est alors factorisée pour un ensemble de problèmes. Dans ce cas, la Γ-convergence est une technique. Le deuxième intérêt réside dans la robustesse de son formalisme qui permet d'aborder une grande variété de problèmes en conférant une certaine structure au résultat mais aussi à la démonstration. Le livre introductif de Braides illustre très bien ce deuxième point [36, 37].

La Γ-convergence recouvre deux significations différentes : c'est d'abord une notion abstraite, celle de la convergence au sens variationnel (convergence des infima et des minimiseurs et stabilité par rapport aux perturbations), mais aussi une technique de démonstration (point de vue variationnel par opposition aux équations aux dérivées partielles, introduction de la Γ-liminf, Γ-limsup, études de propriétés de classes d'objet mathématiques par rapport à la Γ-convergence).

Dans ce paragraphe, on présente d'abord le résultat d'homogénéisation *au sens de* la Γ-convergence et on décrit rapidement les étapes de démonstration de ce résultat *par* Γ-convergence. Dans la démonstration, on utilise alors la compacité d'une classe de problèmes variationnels pour une certaine topologie de Γ-convergence. Parallèllenent, il est possible - et parfois plus intéressant - de donner des démonstrations directes de ces mêmes résultats. Nous donnons quelques références de ce type.

La classe d'énergies à laquelle nous nous intéressons est la classe des densités d'énergie standard :

Définition 9 *Une densité d'énergie standard d'ordre $p > 1$ est une fonction $W : \mathbb{R}^3 \times \mathcal{M}_3(\mathbb{R}) \to \mathbb{R}$ telle que*

i. $W(\cdot, \xi)$ est mesurable pour tout $\xi \in \mathcal{M}_3(\mathbb{R})$,

ii. $W(x,\cdot)$ *est quasiconvexe pour presque tout* $x \in \mathbb{R}^3$,
iii. W *satisfait la condition de continuité et de coercivité : il existe* $C \geq c > 0$ *telles que*

$$c(|\xi|^p - 1) \leq W(x,\xi) \leq C(|\xi|^p + 1), \tag{3.20}$$

pour presque tout $x \in \mathbb{R}^3$ *et pour tout* $\xi \in \mathcal{M}_3(\mathbb{R})$.

Donnons quelques exemples d'énergies standard. L'énergie de Saint Venant-Kirchhoff n'est pas standard car elle n'est pas quasiconvexe. En revanche, son enveloppe quasiconvexe [121] est une énergie stantard pour $p = 4$. Les énergies de Ciarlet-Geymonat ne sont pas standard car elles satisfont la propriété (1.1), ce qui est incompatible avec la majoration de (3.20). Si on supprime la dépendance en le troisième invariant, l'énergie obtenue est standard pour $p = 2$. Plus généralement, les énergies d'Ogden sont standard si le terme volumique de l'énergie n'explose pas quand $I_3 \to 0$ et satisfait une propriété de croissance à l'infini compatible avec les termes en I_1 et I_2. Un exemple non trivial est donné par l'énergie introduite par Müller et rappelée Section 5.6.2. La seule limitation importante des énergies standard est l'incompatibilité avec la condition (1.1) modélisant le comportement des matériaux lorsque $I_3 \to 0$.

Le résultat fondamental en homogénéisation des intégrales multiples par Γ-convergence est le résultat de compacité suivant (qui peut-être vu comme le pendant du Théorème 16 de compacité pour la H-convergence) pour les densités d'énergie standard.

Théorème 18 *[38, Prop 12.3] Soit* W_ϵ *une famille de densités d'énergie standard d'ordre* $p > 1$ *satisfaisant (3.20) sur* Ω *domaine de* $\mathbb{R}^3$. *Alors il existe une densité d'énergie standard d'ordre* p *notée* $W_{hom} : \Omega \times \mathcal{M}_3(\mathbb{R}) \to \mathbb{R}$ *satisfaisant (3.20) et telle que, à extraction près en* ϵ, $I_\epsilon : W^{1,p}(\Omega) \to \mathbb{R}, u \mapsto \int_\Omega W_\epsilon(x,\nabla u)$ $\Gamma(L^p)$*-converge vers* $I_{hom} : W^{1,p}(\Omega) \to \mathbb{R}, u \mapsto \int_\Omega W_{hom}(x,\nabla u)$.

Les Théorèmes 18 et 6 impliquent en particulier la convergence des infima et des minimiseurs indépendamment des conditions aux limites dans $W^{1,p}(\Omega)$ (Dirichlet, Neumann, mixtes). Dans le cas de l'homogénéisation périodique on a le résultat suivant.

Théorème 19 *[38, Th. 14.5] Si de plus* $W_\epsilon(\cdot,\xi) = W(\frac{\cdot}{\epsilon},\xi)$ *pour tout* $\xi \in \mathcal{M}_3(\mathbb{R})$ *avec* W *densité d'énergie* 1*-périodique en espace, alors la densité d'énergie homogénéisée* W_{hom} *est homogène et satisfait la formule asymptotique suivante*

$$W_{hom}(\xi) = \lim_{N\to\infty} \frac{1}{N^3} \inf\left\{ \int_{(0,N)^3} W(x,\xi+\nabla v), v \in W^{1,p}_{\#}((0,N)^3) \right\}. \tag{3.21}$$

De plus, toute la suite I_ϵ $\Gamma(L^p)$*-converge vers* I_{hom}.

Remarque **5** *[38, Th. 14.7] Si en outre* $W(x,\cdot)$ *est strictement convexe, alors la formule asymptotique* (3.21) *s'écrit également*

$$W_{hom}(\xi) = \inf\left\{ \int_{(0,1)^3} W(x,\xi+\nabla v), v \in W^{1,p}_{\#}((0,1)^3) \right\}.$$

Nous donnons ici les grandes étapes de démonstration du résultat de compacité, en suivant la démarche de [33] et [38].

1) Démontrer la compacité de la classe de fonctionnelles d'ensembles abstraites $\tilde{I}_\epsilon : W^{1,p}(\Omega) \times \mathcal{O}(\Omega) \to \mathbb{R}, (u,U) \mapsto \int_U W_\epsilon(x,\nabla u(x))dx$ pour la $\Gamma(L^p)$-convergence, où $\mathcal{O}(\Omega)$ est l'ensemble des ouverts de Ω,
2) Démontrer qu'après extraction la limite obtenue vérifie des propriétés assurant une représentation intégrale (critètes de De Giorgi-Letta [38, Th. 10.2], [46]),
3) Dans le cas périodique (ou stochastique ergodique), démontrer que la densité d'énergie ne dépend pas de $x \in \Omega$,
4) Déduire, de la convergence des infima, la formule asymptotique.

Les 4 points seront détaillés au chapitre 9 dans les démonstrations des résultats sur les passages d'énergies discrètes aux énergies continues, pour lesquels l'articulation des preuves est identique. Le premier point s'obtient généralement par la stabilité des propriétés de croissance par un certain passage à la limite et un argument d'extraction diagonale. Le deuxième point concerne les formules de représentation intégrale de fonctionnelles dans les espaces de Lebesgue, Sobolev ou encore des fonctions à variation bornée [46]. Le troisième point est véritablement à la base des théories de l'homogénéisation et peut-être désigné par le terme générique d'ergodicité (voir [59, Prop. 1] pour le cas stochastique). Enfin le dernier point est la conclusion logique de l'argumentaire précédent.

Plusieurs autres démonstrations existent dans le cas périodique. La méthode développée dans [33] et [38] est générale car elle démontre d'abord un résultat de compacité avant de particulariser au cas périodique. La méthode de Müller [145] est plus directe que la précédente. La méthode de l'éclatement périodique [53] permet également de démontrer ce résultat et simplifie la preuve au sens où aucun argument relatif à la Γ-convergence n'est utilisé. Cependant ce théorème ne peut pas se démontrer sans faire appel à des arguments assez fins du calcul des variations, qu'ils soient d'un type ou d'un autre : la preuve de [53], bien que directe, utilise des propriétés fines de mesurabilité de fonctions multivaluées. Enfin, une dernière formulation et démonstration introduite par Babadjian, Baìa et Santos dans [12] (avec une formule originale pour le problème de cellule) utilise les mesures d'Young à deux échelles. Le lien entre l'approche par éclatement périodique et les mesures d'Young à deux échelles se déduit directement de [12, Th. 1.2 i)], la formule de cellule de [12] impliquant celle de [53].

Revenons maintenant à l'élasticité. La théorie de l'homogénéisation des intégrales multiples s'applique aux densités d'énergies quasiconvexes, et donc aux énergies objectives. Cependant elle ne s'applique pas aux énergies de Ciarlet-Geymonat à cause de la propriété de domination (3.20). On ne peut donc pas traiter avec le Théorème 18 les densités d'énergie modélisant correctement le comportement (1.1) quand $\det \xi \to 0^+$. Un intérêt de la polyconvexité utilisée dans [14] est d'assurer la semi-continuité inférieure des fonctionelles intégrales tout en autorisant le comportement mécanique (1.1) quand $\det \xi \to 0^+$. L'étude de l'homogénéisation des énergies polyconvexes permettrait de traiter de manière plus satisfaisante l'homogénéisation en élasticité non linéaire. Cependant un tel résultat est un problème ouvert pour lequel les premiers signes ne sont pas encourageants. En effet, Braides a démontré dans [34] que la polyconvexité est une propriété qui n'est pas conservée par homogénéisation des densités d'énergie à croissance non standard, seule la quasiconvexité est préservée. Barchiesi a récemment étendu ce résultat aux densités d'énergie à croissance standard (3.20) dans [16], en se concentrant sur le level set à densité d'énergie nulle. Ces résultats indiquent donc que si les densités d'énergie modélisant correctement le comportement (1.1) quand $\det \xi \to 0^+$ sont "homogénéisables", la fonctionelle d'énergie associée serait semi-continue inférieurement, aurait le bon comportement quand $\det \xi \to 0^+$ mais ne serait vraisemblablement pas polyconvexe. Nous sommes confrontés à un problème similaire au chapitre 10, dans lequel la dérivation d'énergies hyperlélastiques incompressibles à partir de modèles discrets est incomplète (Théorème 58).

Bien que le Théorème 18 ne soit pas complètement satisfaisant au regard de l'élasticité non linéaire, de nombreuses propriétés intéressantes sont déjà présentes, dont les phénomènes de bifurcation. Ainsi, dans le cas général, le formule asymptotique (3.21) ne peut pas se simplifier comme dans le cas convexe. L'exemple de Müller [145] est basé sur le concept mécanique de flambement d'une barre. Il considère un matériau composite bidimensionnel lamellé composé d'une partie rigide (la barre) entourée d'un matériau mou. Quand la raideur du matériau mou tend vers zéro, il démontre que l'énergie de la formule asymptotique (pour une compression verticale fixe) tend également vers zéro (le matériau rigide flambe et son énergie tend vers zéro) alors que l'énergie d'une cellule est bornée inférieurement indépendant de la raideur du matériau mou (par comparaison à un mouvement rigidifiant). Ainsi le minimum n'est pas atteint pour un nombre fini de cellules de périodicité. Cet exemple est étudié numériquement au chapitre 5.

Un dernier phénomène intéressant est lié à la condition de stabilité de Legendre-Hadamard, qui peut être perdue au cours du passage à la limite comme le montre l'étude de Geymonat,

Müller et Triantafyllidis [94]. Ceci se traduit mathématiquement par la perte de stricte rang un convexité de la densité d'énergie homogénéisée : les minimiseurs de l'énergie homogénéisée ne sont alors plus isolés. Un exemple est aussi illustré numériquement au chapitre 5. La difficulté liée à la polyconvexité et à son lien avec l'homogénéisation est la raison principale pour laquelle seules les énergies standard de la Définition 9 sont considérées aux chapitres 5, 6, 7 et 10.

3.1.6 Principe de localisation et problèmes de G-fermeture

Dans le cadre des équations elliptiques linéaires, le principe de localisation établit que toute matrice limite donnée par le résultat abstrait de compacité est aussi la limite d'une suite de matrices homogénisées au sens de l'homogénéisation périodique. La formulation rigoureuse de ce résultat est donnée dans l'exemple suivant. Les problèmes de G-fermeture ont été introduits par Tartar, et le principe de localisation par Tartar, Dal Maso et Kohn.

Considérons un mélange bidimensionnel de deux matériaux de conductivité A et B dans $]0,1[^2$, et les fonctions caractéristiques associées 1_A et 1_B $:]0,1[^2 \to \{0,1\}$ qui décrivent la composition spatiale du mélange, et telles que $1_A + 1_B = 1$. Soit $(1_A^n, 1_B^n)$ une suite de tels matériaux composites telle que $1_A^n \rightharpoonup^* 1_A^\infty$ et $1_B^n \rightharpoonup^* 1_B^\infty \in L^\infty(]0,1[^2,(0,1))$, et $C^* :]0,1[^2 \to \mathcal{M}_2(\mathbb{R})$ une matrice limite associée (dont l'existence est assurée par le résultat de compacité). On définit la proportion limite de matériau A par

$$\theta(x) = \lim_{\rho \to 0} \frac{1}{|]0,1[^2 \cap B(x,\rho)|} \int_{]0,1[^2 \cap B(x,\rho)} 1_A^\infty(y)dy,$$

qui existe en tout point de Lebesgue $x \in]0,1[^2$ de 1_A^∞ et donc presque partout.

Pour tout $\alpha \in (0,1)$, on appelle $\mathcal{H}(\alpha)$ l'ensemble des matrices obtenues par l'homogénéisation d'un composite périodique des matériaux A et B en proportions α et $\beta = 1-\alpha$ respectivement, et $\overline{\mathcal{H}(\alpha)}$ la fermeture de $\mathcal{H}(\alpha)$ dans $\mathcal{M}_2(\mathbb{R})$.

Le principe de localisation [37, Prop 5.9] s'écrit alors : pour presque tout $x \in]0,1[^2$, $C^*(x) \in \overline{\mathcal{H}(\theta(x))}$.

On appelle problème de G-fermeture la caractérisation de $\overline{\mathcal{H}(\alpha)}$ ou de tout ensemble de matrices obtenues par l'homogénéisation périodique de matériaux sous diverses contraintes de composition.

Le principe de localisation s'étend facilement aux énergies convexes comme nous le démontrons au chapitre 6. Dans le cas quasiconvexe, la réponse n'est pas claire. Au chapitre 8, nous faisons le lien entre un problème de G-fermeture classique et un problème de G-fermeture pour une énergie issue d'un passage discret-continu.

3.2 Méthodes numériques pour l'homogénéisation et homogénéisation numérique

Cette section est une introduction aux méthodes utilisées pour approcher numériquement la solution d'un problème aux limites elliptique très hétérogène, qui seront analysées dans la Partie II. Développées pour traiter des problèmes non périodiques (homogénéisation numérique en général), ces méthodes sont souvent testées et analysées dans le cadre périodique (méthodes numériques pour l'homogénéisation). Dans la littérature, on trouve principalement deux types de méthodes qu'on pourrait appeler duales dans un cadre variationnel. La première tente d'homogénéiser *numériquement* l'opérateur et de calculer classiquement une solution du problème homogénéisé, tandis que la seconde essaie de particulariser l'espace des fonctions tests en y ajoutant des fonctions liées à l'opérateur (par exemple des fonctions tests à deux échelles, au sens des fonctions oscillantes de Tartar). La formulation variationnelle est très intéressante pour comparer les deux approches et les analyser, aussi bien dans les cas d'énergies quadratiques, convexes que quasiconvexes. Dans la suite, nous nous focalisons sur deux méthodes particulières. D'autres travaux existent et proposent des approches numériques différentes qui ne seront pas détaillées ici, notamment [133], [45] et [143].

3.2.1 Eléments finis mutiéchelles dans le cas linéaire

Ce paragraphe présente les résultats de Hou [104–106] sur l'homogénéisation numérique d'équations aux dérivées partielles elliptiques linéaires basée sur une méthode d'éléments finis multiéchelles.

On considère le problème (3.1), en supposant de plus que Ω est un polygône convexe de $\mathbb{R}^3$, et que $A \in \mathcal{C}^\infty(Y, \mathcal{M}_3(\mathbb{R}))$. On définit un espace d'éléments finis associé à l'opérateur $\mathcal{A}_\epsilon$ comme suit.

Construction d'une base d'éléments finis mutiéchelles

Soit $\mathcal{T}_H$ une triangulation régulière de Ω. Soit $\{x_j\}_{j=1,J}$ les noeuds intérieurs au maillage et $\{\Psi_j\}_{j=1,J}$ la base de fonctions $P1$ associées engendrant l'espace d'éléments finis $W_H \subset H_0^1(\Omega)$. On note $S_i = \text{supp}(\Psi_i)$ et on définit Φ^i à support dans S_i comme suit :

$$\begin{cases} \mathcal{A}_\epsilon \Phi_\epsilon^i = 0 \quad \text{dans } K \\ \Phi_\epsilon^i = \Psi_i \text{ sur } \partial K \quad \forall K \in \mathcal{T}_h, K \subset S_i \end{cases} \tag{3.22}$$

On obtient donc des fonctions $\Phi_\epsilon^i \in H_0^1(S_i)$. On définit $V_H^\epsilon \subset H_0^1(\Omega)$ comme l'espace engendré par le prolongement naturel des Φ_ϵ^i à Ω tout entier.

On résout alors le problème (3.1) sur l'espace V_H^ϵ : trouver $u_H^\epsilon \in V_H^\epsilon$ tel que

$$\left\langle A(\frac{x}{\epsilon})\nabla u_H^\epsilon, \nabla v_H^\epsilon \right\rangle = \langle f, v_H^\epsilon \rangle \quad \forall v_H^\epsilon \in V_H^\epsilon \tag{3.23}$$

où $\langle \cdot, \cdot \rangle$ désigne le produit scalaire de $L^2(\Omega)$.

En chaque noeud du maillage, on doit résoudre (au moins une fois) le problème réel à l'échelle ϵ sur le support de la fonction de base associée au noeud.

Les résultats d'homogénéisation permettent d'obtenir des estimations a priori sur ce type de discrétisation.

Estimations a priori

Comme au paragraphe 3.1.1, u_0 désigne la solution du problème homogénéisé et u_1 la solution de (3.4). Il est clair que $u_0 \in H^2(\Omega)$ car Ω est un polygône convexe. Soit θ_ϵ la solution du problème

$$\begin{cases} \mathcal{A}_\epsilon \theta_\epsilon = 0 & \text{dans } \Omega \\ \theta_\epsilon(x) = u_1(x, \dfrac{x}{\epsilon}) & \text{sur } \Gamma \end{cases} \tag{3.24}$$

L'analyse a priori est basée sur le résultat suivant dû à Moskow et Vogelius

Lemme 2 *[142] Soit $u_0 \in H^2(\Omega)$ la solution de (3.7), u_1 la solution de (3.4), et $\theta_\epsilon \in H^1(\Omega)$ la solution de (3.24). Alors il existe une constante C, ne dépendant pas de u_0, ϵ et Ω, telle que*

$$\|u - u_0 - \epsilon(u_1 - \theta_\epsilon)\|_{1,\Omega} \leq C\epsilon(|u_0|_{2,\Omega} + \|f\|_{0,\Omega}).$$

A partir de ces résultats, on obtient les estimations a priori **indépendantes de** ϵ suivantes :

Cas $H < \epsilon$

Théorème 20 *Soit $u_\epsilon \in H^2(\Omega)$ solution de (3.1) et $u_H^\epsilon \in V_H^\epsilon$ solution de (3.23). Il existe alors une constante C indépendante de u et H telle que,*

$$\|u_\epsilon - u_H^\epsilon\|_{1,\Omega} \leq CH(|u_\epsilon|_{2,\Omega} + \|f\|_{0,\Omega}) \tag{3.25}$$

Il est à noter que l'estimation (3.25) explose comme H/ϵ quand $\epsilon \to 0$ car $|u_\epsilon|_{2,\Omega} = O(1/\epsilon)$. Cette propriété montre que la méthode de construction des éléments finis est une méthode d'approximation consistante de la solution. Cependant, cette estimation n'est pas intéressante puisque $H > \epsilon$ en pratique.

Cas $H > \epsilon$

Théorème 21 *Soit $u_\epsilon \in H^2(\Omega)$ solution de (3.1) et $u_H^\epsilon \in V_H^\epsilon$ solution de (3.23). Il existe alors une constante C indépendante de u_ϵ, ϵ et H telle que,*

$$\|u_\epsilon - u_H^\epsilon\|_{1,\Omega} \leq C(H+\epsilon)\|f\|_{0,\Omega} + C\left(\frac{\epsilon}{H}\right)^{1/2}\|u_0\|_{1,\infty,\Omega}, \tag{3.26}$$

où $u_0 \in H^2(\Omega) \cap W^{1,\infty}(\Omega)$ est la solution de l'équation homogénéisée (3.7).

Cette estimation est intéressante dans les cas pratiques où on ne veut pas simuler à l'échelle ϵ. Cependant, le terme $(\epsilon/H)^{1/2}$ n'est pas forcément très petit pour autant. Pour une microstructure de l'ordre du micron et une maille de l'ordre du millimètre, l'estimation (3.26) est dominée par le terme $(\epsilon/H)^{1/2}$ qui est de l'ordre de 3.10^{-2}. Quand H se rapproche de ϵ, il y a des phénomènes de résonance. Il est possible de pallier cet inconvénient par des techniques de sur-échantillonnage, qui seront présentées et analysées au chapitre 7. Par ailleurs, au chapitre 6, nous généralisons certaines estimations d'erreur au cas des opérateurs monotones et nous démontrons la convergence de la méthode sans hypothèse de périodicité dans le cas des densités d'énergie quasiconvexes.

Intérêt pratique

Le calcul des fonctions de base mutiéchelles peut être considéré comme un précalcul : il ne dépend pas des conditions aux limites, ni du second membre, il est fait une fois pour toutes. Si on est intéressé par faire plusieurs calculs avec différents conditions aux limites et seconds membres, cette méthode devient donc très performante.

Comparé à une approche directe, typiquement par décomposition de domaines, la méthode des éléments finis mutiéchelles est encore intéressante car elle minimise beaucoup le nombre de données échangées entre les blocs (qui sont ici les supports S_i). En effet, en décomposition de domaines, on impose que le résidu variationnel à l'interface est nul, avec les éléments finis mutiéchelles on impose seulement une unique relation à chaque interface.

Enfin, dans les cas non périodiques, s'il y a une séparation des échelles suffisantes, il est alors inutile de calculer entièrement Ψ_i^ϵ et il suffit de calculer une restriction sur une partie de S_i et de prolonger la restriction par périodicité à tout S_i.

3.2.2 Approximation de l'opérateur homogénéisé

Le lien entre l'opérateur, les éléments finis mutiéchelles, l'opérateur homogénéisé et les éléments finis classiques sera clarifié au chapitre 6. Dans la méthode des éléments finis mutiéchelles, on conserve l'opérateur et on calcule une base particulière "hétérogène" (variations à l'échelle ϵ) alors que dans d'autres méthodes d'homogénéisation numérique, on conserve la base homogène (variations à l'échelle 1) et on essaie d'approcher l'opérateur homogénéisé. Le point clé de la comparaison est de remarquer que le problème (3.22) permet certes de définir Φ_i^ϵ mais permet aussi d'approcher l'opérateur homogénéisé A_{hom} par A_{hom}^ϵ selon :

$$\langle \mathcal{A}_{hom}^\epsilon \nabla\Psi_i, \nabla\Psi_j\rangle := \langle \mathcal{A}_\epsilon \nabla\Phi_i^\epsilon, \nabla\Phi_j^\epsilon\rangle \simeq \langle \mathcal{A}_{hom}\nabla\Psi_i, \nabla\Psi_j\rangle,$$

pour tous i et j.

Dans la suite de cette section, nous présentons les résultats de [1], correspondant aux résultats de la Section 3.2.1 dans cet autre cadre.

Calcul approché de l'opérateur homogénéisé

L'opérateur homogénéisé est approché par la formule suivante

$$A^{\epsilon,h,\eta}(x)_{ij} = \frac{1}{|C(x,\eta)|}\int_{C(x,\eta)} A_\epsilon(y)(e_i + \nabla u_i^{\epsilon,h,\eta})(e_j + \nabla u_j^{\epsilon,h,\eta}),$$

où $C(x,\eta)$ est un voisinage de x de diamètre η et $u_i^{\epsilon,h,\eta}$ est l'unique solution dans $V_{h,\eta}$, sous-espace de dimension finie de $H_0^1(C(x,\eta))$, de

$$\int_{C(x,\eta)} A_\epsilon(y)(e_i + \nabla u_i^{\epsilon,h,\eta})\nabla v = 0 \qquad \text{pour tout } v \in V_{h,\eta}.$$

Dans le cas périodique, on a l'estimation d'erreur suivante

Théorème 22 *Soit $V_{h,\eta}$ un sous-espace d'éléments finis $P1$ de $H_0^1(C(x,\eta))$ sur un maillage régulier de finesse ηh, alors il existe C_1 et C_2 indépendants de ϵ, h, η et x tels que*

$$|A_{hom} - A^{\epsilon,h,\eta}| \leq C_1\frac{\epsilon}{\eta} + C_2 h.$$

Résolution du problème homogénéisé approché

Ayant calculé $A^{\epsilon,h,\eta}$, on peut résoudre le problème homogénéisé approché. On appelle $u_{0,H}^{\epsilon,h,\eta} \in W_H$ l'unique solution de

$$\langle \mathcal{A}^{\epsilon,h,\eta}\nabla u_{0,H}^{\epsilon,h,\eta}, \nabla\Psi\rangle = \langle f, \Psi\rangle \qquad \text{pour tout } \Psi \in W_H.$$

On a alors l'estimation *a priori* suivante

Théorème 23 *[1, Th. 3.6] Sous les hypothèses du Théorème 22, il existe C_1, C_2 et C_3 tels que*

$$\|u_{0,H}^{\epsilon,h,\eta} - u_0\|_{H^1(\Omega)} \leq C_1\frac{\epsilon}{\eta} + C_2 h^2 + C_3 H.$$

Reconstruction des échelles fines

La reconstruction $u_H^{\epsilon,h,\eta}$ de la solution u_ϵ à l'échelle ϵ sur chaque cube $C(x,\eta)$ est définie par

$$\left(u_H^{\epsilon,h,\eta}\right)_{|C(x,\eta)} = \sum_i \left(\frac{1}{|C(x,\eta)|}\int_{C(x,\eta)} \nabla u_{0,H}^{\epsilon,h,\eta}\right)_i (u_i^{\epsilon,h,\eta} + \Psi_i).$$

On a alors l'estimation d'erreur suivante :

Théorème 24 *[1, Th. 3.11] Sous les hypothèses du Théorème 22, il existe C_1, C_2, C_3 et C_4 tels que*

$$\|u_H^{\epsilon,h,\eta} - u_\epsilon\|_{H^1(\Omega)} \leq C_1\frac{\epsilon}{\eta} + C_2 h + C_3 H + C_4\sqrt{\epsilon}.$$

Au chapitre 6, nous développons un cadre qui permet d'unifier l'analyse des approches des paragraphes 3.2.1 et 3.2.2, et généraliser les estimations d'erreur à des opérateurs non linéaires ainsi que la convergence à l'élasticité non linéaire.

4

Contributions de la thèse (*in English below*)

Nous reproduisons dans la suite les travaux issus de la thèse. Certaines parties sont redondantes, notamment en homogénéisation, et chaque chapitre est relativement indépendant. Le coeur de la thèse est constitué des Parties II et III, sur l'homogénéisation numérique et la dérivation de modèles continus à partir de modèles discrets. Dans la Partie IV, nous présentons un travail sur des problèmes numériques en interaction fluide-structure.

4.1 Méthodes numériques en homogénéisation

4.1.1 Homogénéisation périodique en élasticité non linéaire [P1]

Au chapitre 5, nous présentons une approche directe de la simulation numérique d'un composite hyperélastique. Plus précisément, nous considérons l'assemblage périodique de deux matériaux hyperélastiques - typiquement du caoutchouc et une inclusion ou renfort métallique, ou encore une mousse de caoutchouc. La plupart du temps, le problème de l'élastostatique pour la densité d'énergie hétérogène n'est pas abordable du point de vue numérique : la séparation des échelles macroscopique et mésoscopique (l'échelle de l'assemblage périodique) requiert souvent un maillage trop fin pour permettre une simulation complète du matériau. Par ailleurs, les détails de la déformation à l'échelle des hétérogénéités peuvent ne pas être d'un intérêt primordial (tout au moins en première approximation) et seul le comportement à l'échelle macroscopique peut être recherché. On peut alors plutôt simuler numériquement le comportement du matériau homogénéisé, et utiliser la formule asymptotique (3.21) du paragraphe 3.1.5. Cette formule n'étant pas analytique, nous nous proposons de la discrétiser. L'énergie homogénéisée n'est pas hérérogène et on peut approximer (classiquement) le problème homogénéisé approché avec un maillage de finesse maîtrisée.

Le premier niveau de discrétisation consiste à remplacer la formule (3.21) par une quantité $W^{N,h}$ calculable. Il convient alors de faire deux approximations : d'abord considérer un nombre fini N de cellules de périodicité par dimension, et ensuite de résoudre le problème de minimisation sur un espace de dimension finie $V_{N,h}$, typiquement par une méthode d'éléments finis,

$$W^{N,h}(A) = \frac{1}{N^3} \inf \left\{ \int_{(0,N)^3} W(x, A + \nabla v(x)) dx : v \in V_{N,h} \right\}.$$

Le second niveau de discrétisation, plus standard, consiste à rechercher la solution du problème de l'élastostatique pour le matériau homogénéisé approché $W^{N,h}$, dans un espace de dimension finie V_H. Nous démontrons un résultat d'approximation dans l'esprit du Théorème 8, avec cette fois trois niveaux de discrétisation emboîtés (N, $V_{N,h}$ et V_H) :

Théorème 25 *Soit $u_H^{N,h}$ une suite de points de minimum de $v \mapsto \int_\Omega W^{N,h}(\nabla u_0 + \nabla v)$ sur V_H, alors il existe un point de minumum u de $v \mapsto \int_\Omega W_{hom}(\nabla u_0 + \nabla v)$ sur $W_0^{1,p}(\Omega)$ tel que, à extractions près,*

$$\lim_{H\to 0} \lim_{N\to\infty} \lim_{h\to 0} u_H^{N,h} = u \qquad \text{dans } W^{1,p}(\Omega),$$

si u est un minimiseur isolé.

Concrètement, à coût total fixé, nous sommes alors confrontés au choix de la répartition de la puissance de calcul à chacune des approximations. Dans le cas simple, mais non trivial, d'une énergie convexe (non quadratique), nous donnons une estimation d'erreur du type

$$\|u_H^h - u\|_{1,p} \leq C_1 h^\alpha + C_2 H^\beta$$

qui met en évidence le lien entre les différents paramètres de discrétisation h et H, et leur impact sur la solution approchée.

Pour appliquer les méthodes décrites au paragraphe 2.2.2, la connaissance des deux quantités suivantes est nécessaire : le tenseur des contraintes de Piola-Kirchhoff $\frac{\partial W^{N,h}}{\partial \xi}$ et la matrice de rigidité $\frac{\partial^2 W^{N,h}}{\partial \xi^2}$, soit les dérivées prémières et secondes de l'approximation de la formule asymptotique (3.21). Nous introduisons alors une formule pour chacune des ces quantités, (5.76) et (5.77), que nous justifions dans le cas convexe. Dans le cas quasiconvexe de l'élasticité en grandes déformations, ces formules sont formelles.

Pour juger l'intérêt de l'approche numérique proposée, nous présentons également plusieurs cas-tests. Les deux premiers exemples illustrent des résultats mécaniques réalistes, mis en évidence expérimentalement et reproduits mathématiquement par le modèle. Le premier exemple présente une étude de l'influence du nombre de périodes considérées dans l'approximation de la formule asymptotique (3.21) en termes d'énergie et de minimiseur de l'équation homogénéisée. Il reprend et quantifie l'exemple de Müller [145] en trois dimensions, basé sur le flambement d'une barre rigide (voir Tableau 5.2). Le deuxième exemple est une illustration de la perte de stabilité qui peut résulter du processus d'homogénéisation (mis en évidence dans [94], et qui est par ailleurs bien connu mécaniquement pour les structures en nid d'abeille [118]). Ceci se traduit numériquement par une instabilité de la solution et une dépendance au maillage, comme illustré Figure 5.1. Ces deux exemples montrent les limites de l'approche développée, limites essentiellement liées à des phénomènes mécaniques intéressants et leur interprétation mathématique. Le dernier exemple montre un cas où ces difficultés n'apparaissent pas et où la méthode converge et donne des résultats réalistes en compression et extension. Elle peut ainsi permettre de calculer rigoureusement une densité d'énergie approchée pour une mousse de caoutchouc.

4.1.2 Cadre général pour l'analyse des méthodes d'homogénéisation numérique [P2,S1]

Si on s'affranchit de l'hypothèse de périodicité, la théorie de l'homogénéisation reste valable "à extraction près", comme le montrent les résultats de compacité des Théorèmes 18 et 16. Dans de nombreux cas, la limite peut ne pas dépendre de l'extraction (cas stochastique stationnaire [59], cas localement périodique [13] *etc.*). Ce sont les cas qui nous intéressent ici, les cas "homogénéisables" pour lesquels la limite existe mais n'est pas donnée par un problème de cellule exploitable numériquement. Notre hypothèse de départ est donc la suivante : mécaniquement, à l'échelle macroscopique le matériau semble "localement" homogène.

La stratégie adoptée dans les chapitres 6 et 7 consiste à introduire une densité d'énergie moyennée (par minimisation locale)

$$W_{\eta,\epsilon}(x,\xi) = \inf \Big\{ \langle W_\epsilon(\cdot, \xi + \nabla v(\cdot)) \rangle_{C(x,\eta)} \mid v \in W_\#^{1,p}(C(x,\eta), \mathbb{R}^d) \Big\},$$

calculable numériquement, et qui converge aussi vers la densité homogénéisée W_{hom} du problème de départ avec W_ϵ. Cette énergie moyennée a l'avantage d'être beaucoup plus homogène en espace

que W_ϵ. On peut interpréter cette énergie en termes mécaniques de volume élémentaire représentatif $C(x,\eta)$. Une fois cette densité d'énergie introduite, plutôt que résoudre le problème de l'élastostatique avec la densité d'énergie très hétérogène W_ϵ, on résout le problème de l'élastostatique avec la densité d'énergie moyennée $W_{\eta,\epsilon}$, de la même façon qu'on remplaçait le problème avec densité d'énergie périodique par le problème avec densité d'énergie homogénéisée (3.21) au chapitre 5.

Au chapitre 6, nous démontrons la Γ-convergence de l'énergie associée à $W_{\eta,\epsilon}$. Par ailleurs, l'approche numérique développée au chapitre 5 s'applique *mutatis mutandis* à la résolution numérique du problème de l'élastostatique avec la densité d'énergie $W_{\eta,\epsilon}$. L'objectif du chapitre n'est donc pas d'introduire une méthode numérique mais d'abord de démontrer que $W_{\eta,\epsilon}$ converge vers W_{hom} quand ϵ et η tendent vers zéro, au sens du Théorème 18, la convergence étant variationnelle. Nous démontrons le résultat pour des énergies quasiconvexes et le particularisons au cas convexe (notamment aux équations aux dérivées partielles de type monotone).

Le deuxième point qui a été abordé concerne la reconstruction des échelles fines de la solution u_ϵ à partir de la solution $u_{\eta,\epsilon}$ du problème homogénéisé approché. Il s'agit d'un correcteur numérique, qui est une approximation du correcteur de la Définition 8. Soit $\{Q_{H,i}\}_{i\in[\![1,I_H]\!]}$ une partition de Ω en sous-domaines disjoints de diamètre d'ordre H. On définit les correcteurs numériques $v^{H,i}_{\eta,\epsilon}$, pour une densité d'énergie strictement convexe, comme les uniques minimiseurs (à une constante près) de

$$\inf\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla v)\,|\,v\in W^{1,p}(Q_{H,i}),\langle\nabla v\rangle_{Q_{H,i}}=\langle\nabla u_{\eta,\epsilon}\rangle_{Q_{H,i}}\right\}.$$

Nous démontrons que cette généralisation naturelle du correcteur périodique donne une approximation consistante en norme L^p du gradient du problème originel :

$$\lim_{H,\eta\to 0}\lim_{\epsilon\to 0}\left\|\sum_i \nabla v^{H,i}_{\eta,\epsilon}1_{Q_{H,i}}-\nabla u_\epsilon\right\|_{0,p}=0,$$

dans le cas monotone sans hypothèse supplémentaire sur la nature des hétérogénéités. Ceci généralise ainsi un résultat récent démontré dans le cas stochastique stationnaire [75].

Cette méthode, conçue originellement pour les cas non périodiques, s'applique également au cas périodique, ce qui nous permet de faire une première analyse d'erreur quantitative entre $W_{\eta,\epsilon}$ et W_{hom} ainsi que sur les solutions des problèmes de l'élastostatique associés.

Enfin, selon la discrétisation du problème homogénéisé approché utilisée, on retrouve deux types de méthodes d'homogénéisation numérique connues : la méthode des éléments finis multiéchelle (MsFEM) et la méthode multiéchelle hétérogène (HMM). La démonstration de convergence faite au niveau continu s'étend naturellement au niveau discret, ce qui prouve la convergence des deux méthodes dans un cadre très général, notamment celui de l'élasticité non linéaire.

Au chapitre 7, nous poursuivons l'analyse des méthodes numériques issues de la discrétisation du problème homogénéisé approché. Les méthodes d'homogénéisation numérique sont souvent couplées à des techniques de sur-échantillonnage : la densité d'énergie $W_{\eta,\epsilon}$ est en fait obtenue comme la moyenne locale sur une boule de rayon η d'une fonction calculée sur un domaine un peu plus plus grand :

$$W^{over}_{\eta,\epsilon,\zeta}(x,\xi)=\langle W_\epsilon(y,\xi+\nabla v^{over}_{\eta,\epsilon,\zeta}(y))\rangle_{C(x,\eta)},$$

où $v^{over}_{\eta,\epsilon,\zeta}$ est la restriction à $C(x,\eta)$ d'une solution $\tilde{v}^{over}_{\eta,\epsilon,\zeta}$ du problème de minimisation suivant posé sur $C(x,\eta+\zeta)$

$$\inf\left\{\langle W_\epsilon(\cdot,\xi+\nabla v(\cdot))\rangle_{C(x,\eta+\zeta)}\,|\,v\in W^{1,p}_0(C(x,\eta+\zeta))\right\}.$$

Quelques exemples numériques simples (voir Tableaux 7.1 et 7.2) dans le cas périodique nous permettent d'identifier précisément l'intérêt potentiel du sur-échantillonnage. Nous généralisons ensuite les resultats de convergence du chapitre 6 au cas du sur-échantillonnage, donnons une

interprétation simple de l'application de cette méthode aux éléments finis multiéchelles en termes variationnels au paragraphe 7.3.5, ce qui nous permet de démontrer la convergence de la formulation de Petrov-Galerkin discontinue des éléments finis multiéchelles, couramment utilisés en pratique.

4.2 Modélisation multiéchelle

Dans la partie III, nous abandonnons momentanément l'aspect numérique pour se concentrer sur des aspects de modélisation, notamment l'étude de propriétés de certains types de modèles en termes de passage du discret au continu.

4.2.1 Sur un problème de G-fermeture pour un passage discret-continu [P3]

Considérons un réseau carré composé de résistances de deux types, comme représenté Figure 4.1.

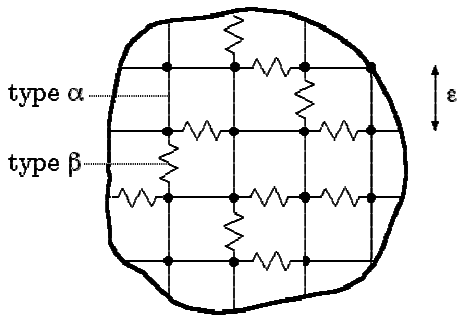

Fig. 4.1. Réseau discret de résistances

Quand on laisse la taille ϵ du réseau tendre vers zéro, l'énergie associée converge vers une énergie continue quadratique avec une matrice de conductivité hétérogène A^*. Cette propriété macroscopique A^* dépend de la structure du réseau sous-jacent. Une question naturelle, tant des points de vue physique et mécanique (travaux de Hill [102], Hashin et Shtrikman [99] *etc.*) que mathématique (travaux de Tartar [171] et [172]), consiste à déterminer l'ensemble des propriétés effectives qu'on peut atteindre à partir d'une quantité fixée des deux types de résistances (ou matériaux) utilisées au niveau discret.

Cette question a été en grande partie résolue par Braides et Francfort dans [39]. De manière surprenante l'ensemble obtenu contient strictement les bornes classiques de Tartar obtenues au niveau continu, comme illustré Figure 4.2.

L'objectif du chapitre 8 consiste à éclaircir le lien entre le matériau composite discret et les matériaux composites continus. Nous obtenons une caractérisation complète de la conductivité effective d'une famille de polycristaux conducteurs anisotropes et exhibons les configurations optimales.

Nous interprétons également les résultats obtenus en termes et méthodes classiquement utilisés en G-fermeture.

Il existe un intérêt pratique aux problèmes de G-fermeture. Le plus évident est la construction de composites aux propriétés optimales. De nombreuses géométries nouvelles pour les composites ont ainsi été proposées suite à des études analytiques et numériques de problèmes de G-fermeture (voir par exemple [170] et [97]). La connaissance d'un certain nombre de géométries optimales permet également de construire des cas-tests efficaces pour les algorithmes d'optimisation de la structure des composites. On renvoie à la monographie de Milton [135] pour ces aspects.

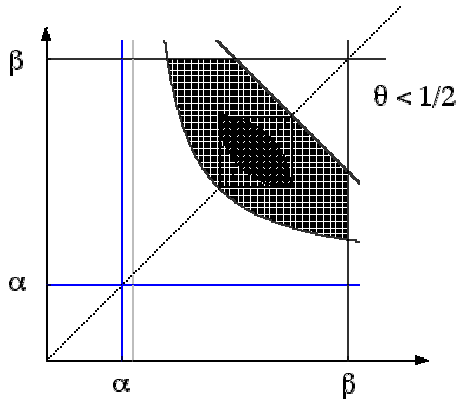

Fig. 4.2. Bornes de Tartar et bornes obtenues dans [39]

4.2.2 Passage discret-continu pour des énergies d'interaction de spin [A3]

Au chapitre 8, nous sommes partis d'un résultat de compacité (voir [4]), qui nous a permis de donner un sens variationnel au passage de l'énergie du réseau discret à l'énergie du milieu continu. Au chapitre 9, nous démontrons un tel résultat de compacité pour un type d'énergie différent, relatif aux systèmes de spin. Ce type de problème est intéressant dans une approche micro-macro des propriétés magnétiques ou micromagnétiques des matériaux. Un des objectifs est de comprendre l'origine des microstructures caractéristiques des états fondamentaux de tels systèmes. Du point de vue continu, la présence de microstructures est comprise comme la conséquence de la non-existence des minimiseurs dans les espaces fonctionnels naturels (et donc physiques !) et la taille des microstructures comme résultat de la compétition entre les énergies de volume et de surface (voir par exemple les articles de synthèse [55, 62]). D'un point de vue plus micro-macro, la communauté de la physique statistique (comme le montre l'article de Giuliani, Lebowitz et Lieb [98]) vise à comprendre l'origine des mésostructures des états fondamentaux (grandes devant la taille caractéristique du réseau de spin, mais petites devant les échelles de la limite thermodynamique) en fonction de la compétition entre les interactions ferromagnétiques à courte distance et anti-ferromagnétiques à longue distance.

Partant de l'énergie considérée dans [98], nous énonçons et démontrons un résultat de compacité pour le passage des énergies discrètes de type

$$E_\epsilon(u) = \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha,\alpha+\epsilon\xi\in(0,1)^N} \epsilon^N f_\epsilon^\xi(\alpha, u(\alpha), u(\alpha+\epsilon\xi))$$

aux énergies continues qui s'écrivent

$$E(u) = \int_{(0,1)^N} f(x, u(x))dx,$$

ainsi qu'un résultat d'homogénéisation. Ces résultats sont des étapes préliminaires à l'étude plus fine des exemples physiques de [98], dans l'esprit de [3] où ces questions sont abordées pour les systèmes de spin avec interactions à courte portée. L'approche du chapitre 8 est très complémentaire de [98] et pourrait permettre de démontrer la stabilité des résultats vis à vis de champs magnétiques extérieurs par exemple.

Les techniques de preuve s'inspirent de [4], sans la contrainte de gradient. Chaque étape de démonstration est détaillée, contrairement au chapitre 10, pour lequel seules les grandes lignes de la démonstration sont données.

4.2.3 Dérivation variationnelle d'une énergie caoutchoutique à partir d'un modèle discret [S2,A4]

Le dernier chapitre de la partie III s'inspire du chapitre 9 tout en se rapprochant enfin des problématiques mécaniques. Il s'agit de la dérivation variationnelle d'un modèle de caoutchouc

$$E(u,D) = \int_D W_{hom}(\nabla u)$$

à partir d'un modèle discret constitué d'un réseau de chaînes de polymères en interaction

$$E_\epsilon(u,D)(\omega) = \sum_{x_i \in \epsilon\mathcal{L}(\omega)\cap D} \epsilon^d \sum_{\substack{x_j \neq x_i \in \epsilon\mathcal{L}(\omega)\cap D \\ [x_i,x_j]\subset D}} J\left(\frac{x_j - x_i}{\epsilon}\right) f\left(\frac{u(x_j)-u(x_i)}{|x_j - x_i|}\right),$$

où $\epsilon\mathcal{L}(\omega)$ représente une réalisation d'un réseau stochastique de points x_i. Ce chapitre constitue la clé de voûte entre les parties II et III. La mise à bout des chapitres 10, 6 et 5 permet de concevoir une dérivation micro-macro complète d'un modèle hyperélastique : modélisation microscopique, formulation rigoureuse de la dérivation micro-macro selon le principe de minimisation, étude des propriétés mécaniques du modèle continu obtenu et simulation numérique du modèle continu. En effet, que la formule donnant $W_{\eta,\epsilon}$ soit obtenue par minimisation d'un problème micro continu ou discret ne change rien, ni à l'approche considérée, ni à la méthodologie numérique (formules donnant la contrainte et la matrice de raideur par exemple). D'où la pertinence des chapitres 5 et 6 dans ce cadre.

Mathématiquement, le chapitre 10 énonce la généralisation des résultats de [4] au cas d'un réseau stochastique $\mathcal{L}$ au sens de Blanc, Le Bris et Lions [26]. La démonstration détaillée et les extensions du résultat principal seront présentées dans [A2]. Au chapitre 10, on insiste plutôt sur la motivation, les hypothèses de modélisation et les propriétés mécaniques et mathématiques de la densité d'énergie continue obtenue.

Partant de la loi de comportement d'une chaîne de polymère (obtenue par des arguments de physique statistique) donnant J et f, et d'une description (probabiliste simple) du réseau de chaînes $\omega \mapsto \mathcal{L}(\omega)$, nous faisons une dérivation variationnelle d'une densité d'énergie continue (déterministe) à partir de l'énergie du réseau discret. La dérivation est uniquement basée sur un principe de minimisation et sur aucune autre hypothèse de type statistique (moyenner la réponse d'une famille de chaînes de différentes longueurs et orientations) ou géométrique (déformation affine ou orientation spontanée d'une cellule représentative selon les directions principales de déformation). Nous démontrons notamment que l'énergie limite est hyperélastique, objective, quasiconvexe et isotrope. Nous proposons également un résultat partiel pour les matériaux de type incompressible.

Ces résultats sont préliminaires. Ils permettent néanmoins de démontrer que le modèle mécanique développé et testé numériquement par Böl et Reese dans [27] converge vers un modèle continu hyperélastique, objectif, homogène et isotrope quand le paramètre de maille tend vers zéro. Les questions plus ambitieuses comme l'interprétation de l'effet de Muellins, de l'hystérésis, de la visco-élasticité ou de la fatigue en termes de passage micro-macro semblent bien plus difficiles et requièrent sans doute l'utilisation ou le développement d'autres techniques mathématiques ou d'autres approches. Le chapitre 10 constitue cependant une bonne base de départ, il s'agit de la dérivation d'un caoutchouc idéalisé sain... première étape vers la dérivation d'un matériau plus réaliste.

4.2.4 Application aux problématiques de la mécanique

Les développements de la partie III sont très unifiés du point de vue des techniques mathématiques et très éclectiques du point de vue des applications. Des chapitres 8 à 10, nous avons affaibli

les hypothèses sur le type d'énergie (quadratique, convexe puis quasiconvexe) et sur la description microscopique (du périodique à l'aléatoire). Chaque chapitre pose (ou reformule) des questions de modélisation et de mathématiques. Le lien entre des modèles de spin pour le magnétisme ou des modèles de polymères pour l'élasticité non linéaire semble très ténu du point de vue physique, il est beaucoup plus clair du point de vue des techniques d'analyse utilisées. Comme indiqué en introduction, traiter un problème dégradé (quasiconvexe → convexe → quadratique, aléatoire → périodique) est un premier pas vers la résolution du problème d'origine (quasiconvexe, aléatoire).

De nombreux défis de la modélisation mécanique partagent les problématiques abordées dans cette partie. Une première classe de problèmes regroupe la dérivation de loi de comportement pour les matériaux du vivant (membrane cellulaire, tissus) à partir de considérations mécaniques à l'échelle microscopique mais aussi bien souvent chimiques ou biochimiques. La deuxième classe de problèmes couvre les matériaux du génie civil en général : les bétons, les boues, les talus, tas de sable et autres matériaux "multiéchelles" à comportement macroscopique complexe.

4.3 Méthodes partitionnées en interaction fluide-structure [A1,A2]

Dans cette dernière partie, nous abordons un autre type de problème, il s'agit de méthodes numériques pour l'interaction fluide-structure.

Le problème de l'interaction fluide-structure dans les vaisseaux sanguins de grande taille est un sujet qui connaît de nombreuses contributions des communautés de la mécanique et des mathématiques appliquées, mais aussi des physiologistes et des médecins.

Dans la partie IV, nous faisons une présentation succincte des différentes méthodes numériques développées pour ce type d'interaction fluide-structure. Nous mettons l'accent sur le partitionnement des méthodes, à savoir leur niveau d'intrusion dans les codes déjà existants (utilisation en boîte noire, sorties non standard, imbrication des niveaux des différentes boucles *etc.*). Nous introduisons également une nouvelle méthode qui s'inspire des méthodes numériques utilisées en élasticité non linéaire. En effet, dans ce cas, il est plus efficace de linéariser puis utiliser des techniques de décomposition de domaine pour résoudre le système linéaire que d'utiliser une formulation de décomposition de domaine non linéaire puis linéariser. Or, en interaction fluide-structure c'est souvent cette dernière alternative qui est choisie. Nous avons proposé au contraire une méthode pour linéariser le problème fluide-structure avant d'utiliser des techniques de décomposition de domaine. Plus les lois de comportement de la structure et du fluide sont complexes (penser à un fluide non newtonien, une structure multiéchelle ou multicouche pour la paroi du vaisseau), plus la méthode est potentiellement intéressante.

La complexification des modèles va de paire avec la finesse de la description physiologique. Quand des modèles de coque mince sont utilisés pour la paroi des vaisseaux sanguins, la variété des lois de comportement utilisables est limitée, et les lois développées par les biomécaniciens sont généralement incompatibles. Ces lois sont en effet pour la plupart tridimensionnelles. Il est ainsi légitime d'essayer de compléfixier la structure dans cette direction. C'est pourquoi nous avons remplacé, au chapitre 11, l'élément de coque (mince) par une élément de coque 3D, ce qui permet d'utiliser toutes les lois de comportement tridimensionnelles.

Contributions of the thesis

We reproduce in Parts II, III and IV the results obtained during the thesis. Some parts are redundant, especially in homogenization, and every chapter is relatively independent.

Numerical methods in homogenization

Periodic homogenization in finite elasticity [P1]

In Chapter 5, we present a direct approach for the numerical simulation of a hyperelastic composite material. More specifically we consider the periodic assembling of two hyperelastic materials - typically rubber with inclusions or reinforcing metallic sheets, or even a rubber foam. Most of the time, the numerical solution of a nonlinear elasticity problem for such a heterogeneous material is not affordable : the separation of the microscopic and macroscopic scales (the scale of the periodic assembling) often requires the use of a mesh which is to refined to allow a complete simulation of the material. On the other hand, the details of the deformation at the scale of the heterogeneities are not really necessary and useful in practice (at least as a first approximation) and only the macroscopic behavior may be looked for. One can then numerically simulate the behavior of the homogenized material instead of the heterogeneous material, and use the asymptotic formula (3.21) of Paragraph 3.1.5. This formula is not analytical, and we need to discretize it. The homogenized energy is not heterogeneous and we may approximate classically the homogenized problem using a coarser mesh.

The first level of discretization consists in replacing formula (3.21) by a quantity $W^{N,h}$ that we may compute. There are two approximations to be done : first we need to consider a finite number N of periodic cells per dimension, and then solve the minimization problem on a finite dimensional space $V_{N,h}$, typically by a finite element method,

$$W^{N,h}(A) = \frac{1}{N^3} \inf \left\{ \int_{(0,N)^3} W(x, A + \nabla v(x)) dx : v \in V_{N,h} \right\}.$$

The second level of discretization is more standard and consists in looking for the solution of the nonlinear elasticity problem with the approximated homogenized energy $W^{N,h}$ in a finite dimensional space V_H. We prove an approximation result in the spirit of Theorem 8, with three levels of discretizations (N, $V_{N,h}$ and V_H) :

Theorem 26 *Let $u_H^{N,h}$ be a sequence of minimum points for $v \mapsto \int_\Omega W^{N,h}(\nabla u_0 + \nabla v)$ on V_H, then there exists a minimum point u for $v \mapsto \int_\Omega W_{hom}(\nabla u_0 + \nabla v)$ on $W_0^{1,p}(\Omega)$ such that, up to extraction,*

$$\lim_{H \to 0} \lim_{N \to \infty} \lim_{h \to 0} u_H^{N,h} = u \qquad \text{in } W^{1,p}(\Omega),$$

if u is an isolated minimizer.

In practice, with fixed computional resources, we face the choice of the repartition of the computational power between the different approximations. In the simple but nontrivial cases of convex energies, we provide with an error estimate of the type

$$\|u_H^h - u\|_{1,p} \leq C_1 h^\alpha + C_2 H^\beta$$

which shows the link between the discretization parameters h and H, and their influence on the approximated solution.

To use the solution methods recalled in Paragraph 2.2.2, we need to know two quantities : the Piola-Kirchhoff stress tensor $\frac{\partial W^{N,h}}{\partial \xi}$ and the stiffness matrix $\frac{\partial^2 W^{N,h}}{\partial \xi^2}$, respectively the first and second derivatives of the approximation of the asymptotic formula (3.21). For each quantity, we introduce a formula, (5.76) and (5.77), that we justify in the convex case. In the quasiconvex case of nonlinear elasticity, these formulae are formal.

To show the interest of the method, we also present some benchmark tests. The first two examples illustrate realistic mechanical results, put in evidence experimentally and reproduced mathematically by the model. In the first example, we study the influence of the number of periods to be considered for the approximation of the asymptotic formula (3.21) in terms of energy and minimizers of the homogenized problem. We quantify in particular the example of Müller [145] in three dimensions, based on the buckling of a rigid bar (see Table 5.2). The second example illustrates the loss of stability that may occur during the homogenization process (put in evidence in [94], and also well-known for honeycomb structures [118]). Numerically the associated solution strongly depends on the mesh, as it is can be seen on Figure 5.1. This is a consequence of the mechanical instability. These two examples show the limits of the present approach, which are mainly due to interesting mechanical phenomena and their mathematical interpretations. The last numerical test is an example for which no such difficulty occurs and for which the method converges and gives realistic results in compression and extension. This allows us to compute rigorously the energy density of a rubber foam.

General framework for the analysis of numerical homogenization methods [P2,S1]

If we relax the periodicity assumption, the homogenization theory holds "up to extraction", as it is showed by the compactness results of Theorems 18 and 16. In many cases, the limit may not depend on the extraction (stochastic case [59], locally periodic case [13] *etc.*). We are interested in such materials, that we may call "homogenizable" materials, for which the limit exists but is not given by any numerically tractable cell problem. Our starting assumption is the following : mechanically, at the macroscopic scale, the material seems "locally" homogeneous.

Our strategy in Chapters 6 and 7 consists in introducing an averaged energy density (by local minimizations)

$$W_{\eta,\epsilon}(x,\xi) = \inf\Big\{\langle W_\epsilon(\cdot,\xi+\nabla v(\cdot))\rangle_{C(x,\eta)} \mid v \in W^{1,p}_{\#}(C(x,\eta),\mathbb{R}^d)\Big\},$$

that we may compute numerically, and that converges to the same homogenized energy density W_{hom} as the original problem with W_ϵ. This averaged energy has the advantage to be far more homogeneous in space than W_ϵ. One may interpret this energy with the mechanical concept of representative volume element $C(x,\eta)$. Once this energy density is introduced, instead of solving the nonlinear elasticity problem with the heterogeneous energy density W_ϵ, we solve the nonlinear elasticity problem with the averaged energy density $W_{\eta,\epsilon}$, exactly as we have replaced the periodic energy density by the homogenized energy density (3.21) in Chapter 5.

In Chapter 6, we prove the Γ-convergence of the energy associated to $W_{\eta,\epsilon}$. It is worth noticing that the numerical approach introduced in Chapter 5 applies *mutatis mutandis* to the numerical solution of the nonlinear elasticity problem with the energy density $W_{\eta,\epsilon}$. We do not aim in Chapter 6 at developing numerical methods but rather at proving that $W_{\eta,\epsilon}$ converges to W_{hom}

when ϵ and η go to zero, as in Theorem 18. We prove the result for quasiconvex energies and make it more specific for convex energies (and monotone partial differential equations).

The second issue we have addressed is the fine scales reconstruction of the solution u_ϵ starting from the solution $u_{\eta,\epsilon}$ of the approximated homogenized problem. Such a reconstruction is called a numerical corrector and is an approximation of the corrector of Definition 8. Let $\{Q_{H,i}\}_{i\in[\![1,I_H]\!]}$ be a partition of Ω in non overlapping subdomains of diameters of order H. The numerical correctors $v_{\eta,\epsilon}^{H,i}$ are defined, for a strictly convex energy density, as the unique minimizers (up to a constant) of

$$\inf\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla v)\,|\,v\in W^{1,p}(Q_{H,i}),\langle\nabla v\rangle_{Q_{H,i}}=\langle\nabla u_{\eta,\epsilon}\rangle_{Q_{H,i}}\right\}.$$

We show that this natural generalization of the periodic corrector provides with a consistant approximation in L^p of the gradient of the solution of the original problem :

$$\lim_{H,\eta\to 0}\lim_{\epsilon\to 0}\left\|\sum_i \nabla v_{\eta,\epsilon}^{H,i}1_{Q_{H,i}}-\nabla u_\epsilon\right\|_{1,p}=0,$$

in the monotone case, without any further assumption on the nature of the heterogeneities. This generalizes a recent result in the stochastic case [75].

This method, initially designed for nonperiodic cases, also applies to the periodic case, which allows us to perform a first quantitative error analysis between $W_{\eta,\epsilon}$ and W_{hom}, and between the solutions of the associated nonlinear problems.

Finally, according to the type of discretization used, we recover two types of numerical homogenization methods : the multiscale finite element method (MsFEM) and the heterogeneous multiscale method (HMM). The convergence proof at the continuous level is easily extended to the discrete level, which proves the convergence of both methods in a rather general context, including nonlinear elasticity.

In Chapter 7, we go further in the analysis of numerical methods obtained by the discretization of the approximate homogenized problem. In practice, numerical homogenization methods are usually coupled with oversampling techniques : the energy density $W_{\eta,\epsilon}$ is actually obtained as the local mean on a ball of radius η of a function computed on a wider domain :

$$W_{\eta,\epsilon,\zeta}^{over}(x,\xi)=\langle W_\epsilon(y,\xi+\nabla v_{\eta,\epsilon,\zeta}^{over}(y))\rangle_{C(x,\eta)},$$

where $v_{\eta,\epsilon,\zeta}^{over}$ is the restriction on $C(x,\eta)$ of a solution $\tilde{v}_{\eta,\epsilon,\zeta}^{over}$ of the following minimization problem posed on $C(x,\eta+\zeta)$

$$\inf\left\{\langle W_\epsilon(\cdot,\xi+\nabla v(\cdot))\rangle_{C(x,\eta+\zeta)}\,|\,v\in W_0^{1,p}(C(x,\eta+\zeta))\right\}.$$

Some simple examples (see Tables 7.1 and 7.2) in the periodic case allow us to identify precisely the potential interest of oversampling. We then generalize the convergence results of Chapter 6 to the oversampling case. We also give a simple interpretation of the multiscale finite element method in variational terms in Paragraph 7.3.5, which allows us to prove the convergence of the discontinuous Petrov-Galerkin formulation of the MsFEM.

Multiscale modeling

In Part III, we temporarily leave numerical aspects to focus on modeling issues, and more specifically on the study of some properties of models obtained by a discrete to continuum derivation.

On a G-closure problem for a discrete to continuum model [S3]

Let us consider a square network made of two types of conductors, as sketched Figure 4.1.

If we let the size ϵ of the network go to zero, the associated energy converges to a continuous quadratic energy with a heterogeneous conductivity matrix A^*. This macroscopic property A^* depends on the structure of the underlying network. A natural question, both from the physical and mechanical (works of Hill [102], Hashin et Shtrikman [99] *etc.*) and from the mathetical viewpoints (works of Tartar [171] and [172]), consists in determining the set of effective properties that can be reached starting from a fixed quantity of the two types of conductors (or materials) used at the discrete level.

This question has been mainly solved by Braides and Francfort in [39]. Surprisingly, the obtained set strictly contains the famous bounds of Tartar obtained at the continuous level, as illustrated Figure 4.2.

In Chapter 8, we aim at clarifying the link between the discrete composite material and continuous composite materials. We obtain a complete characterization of the effective properties of a family of anisotropic conductive polycrystals and provide with optimal configurations.

We also interpret the results in terms and methods ususally used in G-closure.

There is a practical interest in G-closure problems. The most obvious one is the design of composite materials with optimal properties. Numerous new geometries have been proposed using analytical and numerical studies of G-closure problems (see for instance [170] and [97]). The knowledge of some optimal geometries also allows to build benchmark tests for the algorithms dedicated to the optimization of composite structures. We refer the reader to the monography by Milton [135] for these aspects.

Discrete to continuum derivation for spin systems in interaction [A3]

In Chapter 8, we have used a compactness result (see [4]), which has allowed us to give a variational sense to the passage from an energy on a discrete network to an energy of a continuous medium. In Chapter 9, we prove such a result for a different type of energies, related to spin systems. This approach aims at understanding better the origin of magnetic and micromagnetic properties of materials, and especially the origins of the microstructures that characterize the ground states of these systems. From the continuous point of view, the presence of microstructures is understood as a consequence of the non attainment of infima in the natural functional (and physical!) spaces, and the size of the microstructures as a result of the competition between the bulk energies and the surface energies (see for instance the surveys [55, 62]). From a micro-macro point of view, the statistical mechanics community (see the article of Giuliani, Lebowitz and Lieb [98] for instance) aims at understanding the origin of mesostructures of ground states (large with respect to the characteristic lengthscale of the network and small with respect to the scale of the thermodynamic limit) according to the competition between short range ferromagnetic interactions and long range anti-ferromagnetic interactions.

Starting from the energy considered in [98], we state and prove a compactness result for the passage from discrete energies of type

$$E_\epsilon(u) = \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha,\alpha+\epsilon\xi\in(0,1)^N} \epsilon^N f_\epsilon^\xi(\alpha, u(\alpha), u(\alpha+\epsilon\xi))$$

to continuous energie

$$E(u) = \int_{(0,1)^N} f(x, u(x))dx,$$

as well as a homogenization result. These results are preliminary steps to study more accurately the physical examples of [98], in the spirit of [3] where these questions are addressed for spin

systems with short range interactions. The approach of Chapter 8 is very complementary to [98] and could allow us to prove the stability of their results with respect to small exterior magnetic fields.

The techniques of proof are similar to that of [4], without the gradient constraint. Every step of the proof is detailed, as apposed to Chapter 10, in which only a sketch of the proof is given.

Variational derivation of a rubber-like energy from a discrete model [S2,A4]

The last chapter of Part III addresses issues similar to Chapter 9 but related to mechanical modeling. We derive a continuous model for rubber

$$E(u,D) = \int_D W_{hom}(\nabla u)$$

starting from the energy of a discrete network of elastic polymeric chains in interaction

$$E_\epsilon(u,D)(\omega) = \sum_{x_i \in \epsilon\mathcal{L}(\omega)\cap D} \sum_{\substack{x_j \neq x_i \in \epsilon\mathcal{L}(\omega)\cap D \\ [x_i,x_j]\subset D}} J\left(\frac{x_j - x_i}{\epsilon}\right) f\left(\frac{u(x_j)-u(x_i)}{|x_j - x_i|}\right),$$

where $\epsilon\mathcal{L}(\omega)$ is a realization of a stochastic network of points x_i. This chapter makes the link between Parts II and III. The concatenation of Chapters 10, 6 and 5 allow us to conceive a complete micro-macro derivation of a hyperelastic model : microscopic modeling, rigorous formulation of the micro-macro derivation relying on the minimization principle, analysis of the mechanical properties of the obtained model, and numerical simulation of the continuous model. Actually, the way the formula giving $W_{\eta,\epsilon}$ is obtained has no influence on the numerical methodology, may this formula be obtained by the minimization of a discrete or continuous microscopic problem. The formulae for the Piola-Kirchhoff stress tensor and stiffness matrix derived in Chapter 5 still hold.

Mathematically, Chapter 10 generalizes the results of [4] to the case of a stochastic network $\mathcal{L}$ introduced by Blanc, Le Bris and Lions [26]. The detailed proof and some extensions of the main result will be presented in [A2]. In Chapter 10, we focus on the motivation, the modeling assumptions and the mechanical properties of the continuous energy density obtained.

Starting from the constitutive law of a polymeric chain (obtained by statistical mechanics arguments) giving J and f, and from a geometric (and probabilistic) description of the network $\omega \mapsto \mathcal{L}(\omega)$ of polymeric chains, we derive a continuous (and deterministic) energy density. This derivation is only based on minimization arguments, excluding any *ad hoc* assumption of geometric type (affine deformation, spontaneous orientation of a representative cell). We prove that the limit energy density is hyperelastic, frame-invariant, quasiconvex and isotropic. In addition, we give a partial result concerning quasi-incompressible materials.

These results are preliminary. Nevertheless, they allow us to prove that the mechanical model developed and numerically tested by Böl and Reese in [27] converges to a hyperelastic, frame-invariant, homogeneous and isotropic continuous model when the mesh parameter goes to zero. More ambitious questions such as the modeling of the Muellins effect, of hysteresis, viscoelasticity or fatigue in terms of micro-macro derivation seem more complex and require the use or development of new techniques or approaches. Chapter 10 is a first step, where we derive an idealized rubber.

Application to mechanics

The developments of Part III are very similar from the point of view of the mathematical techniques used and very eclectic from the point of view of applications. From Chapter 8 to 10, we have weakened the assumptions on the type of energy (quadratic, convex and quasiconvex) and

on the microscopic description (from periodic to random). In each chapter we ask or reformulate modeling and mathematical questions. The link between spin models for micromagnetism or polymeric models for nonlinear elasticity seem very tight from the physical point of view. It is far clearer from the point of view of the analytical techniques.

Numerous challenges of mechanical modeling share the same kind of problematics addressed in this part. A first class of problems consists in deriving constitutive laws for living materials (cellular membranes, tissues) starting from mechanical considerations at the microscopic scale but also from chemical or biochemical aspects. The second class of problems covers civil engineering materials : concretes, muds, pastes, talus, sand piles and other multiscale materials whose macroscopic behavior is complex.

Partitioned methods in fluid-structure interaction [A1,A2]

Fluid-structure interaction problems in large blood vessels are rather classical. There are numerous contributions from both the mechanics and applied mathematics communities, but also from physiologists and physicians.

In Chapter 11, we make a quick classification of the different numerical methods developed for this type of fluid-structure interaction problems. We stress the partitionning of the methods in the classification, namely their level of intrusion in the codes that already exist (black box, non standard outputs, mixing of the different loops *etc.*). We also introduce a new method, which adapts to fluid-structure interaction what is usually done in nonlinear elastostatics. Actually, in the latter, it is more efficient to linearize the problem first, before solving the linear systems by domain decomposition techniques, rather than introducing a nonlinear domain decomposition formulation and solving the resulting nonlinear problem by a Newton method. Yet, in fluid-structure interaction problems, the nonlinear domain decomposition is often introduced. On the contrary, we have chosen to linearize first and solve the resulting linear problems by efficient domain decomposition algorithms. The more complex the constitutive laws for the structure and the fluid are (think of a non-Newtonian fluid, a multiscale or multi-shell model for the vessel), the more interesting may be this method.

The complexification of the models allow to reach more accurate physiological descriptions. When shell models are used for the vessels, the set of constitutive laws we are allowed to use is limited, and the constitutive relations developed in biomechanics cannot be used in general. Most of the latter are tridimensional, which is incompatible with shells. Therefore it may seem interesting to complexify the numerical model in this direction. This is why we have replaced, in Chapter 11, the classical shell element by a 3D shell element, which allows us to use any 3D constitutive law. The results of the simulations with the method introduced in Chapter 11 and the comparison with other approaches will be presented in [A1].

The applications of the methods recalled or proposed here cover the simulation of living materials. One of the major objectives of the simulation of living materials is the help for diagnosis or for surgery (the cardiovascular system in the present setting).

Part II

Numerical homogenization of elliptic equations

5

A direct approach to numerical homogenization in finite elasticity

Summary. We describe, analyze, and test a direct numerical approach to a homogenized problem in nonlinear elasticity at finite strain. The main advantage of this approach is that it does not modify the overall structure of standard softwares in use for computational elasticity. Our analysis includes a convergence result for a general class of energy densities and an error estimate in the convex case. We relate this approach to the multiscale finite element method and show our analysis also applies to this method. Microscopic buckling and macroscopic instabilities are numerically investigated. The application of our approach to some numerical tests on an idealized rubber foam is also presented. For consistency a short review of the homogenization theory in nonlinear elasticity is provided.

5.1 Physical motivation

Whereas the development of computational tools has helped engineers to design pieces with specific mechanical properties, chemists and physicists have developed new types of materials enjoying new types of properties and characterized by a high heterogeneity. Because of this heterogeneity the numerical methods commonly used by engineers cannot directly deal with these new materials. The reason is that classical analytical constitutive laws do not model correctly all the regimes encountered by these materials at the macroscopic scale.

A computational approach to circumvent the difficulty related to macroscopic constitutive laws could be to use a finite element method (FEM) at a scale for which classical constitutive laws are relevant. Unfortunately this is often out of reach of computers to date since the meshsize would have to be of the order of the micrometer e.g., which is prohibitive.

The landscape is then the following: direct computations at the microscopic scale are too expensive whereas computations at the macroscopic scale are delicate because of the lack of relevant analytical constitutive laws. An alternative track is provided by the homogenization approach.

The article is organized as follows. To start with, some results of the mathematical theory of periodic homogenization for nonlinear energy densities are recalled. The reader familiar with the state of the art of the homogenization theory for minimization problems and elliptic operators in divergence form can easily skip Section 5.2. Our specific contribution is detailed throughout Sections 5.3 to 5.6. Section 5.3 is devoted to an approximation result for a nonlinear elasticity problem with a homogenized constitutive law and to the derivation of an error estimate in the convex case. The numerical method introduced in Section 5.4 consists in replacing an unknown analytical constitutive law at the macroscopic scale by a numerical constitutive law computed at each macroscopic point by the resolution of a so-called *cell-problem* at the microscopic level. This approach is well developed and has been applied to linear materials (FE^2 method [85]) and nonlinear materials at small strain ([143]). It is adapted here to the finite strain case, for which convergence properties of Newton algorithms are very sensitive to the approximation of the second derivative of the constitutive law ([177], [125]). The computation of such a stiffness matrix has not been addressed in the literature to the knowledge of the author. This is one purpose of this

article. In Section 5.5 this direct approach is related to the multiscale finite element method introduced by Hou and co-authers and the error estimate of Section 5.3 is proved to apply to the MsFEM, at least in the periodic case. Finally the question first introduced by Geymonat, Müller and Triantafyllidis in [94] concerning buckling in the cell-problem and instabilities of the homogenized energy density is numerically addressed. The convergence properties of the method for a class of energy densities which is not covered by the mathematical theory is also investigated.

5.2 A quick review of periodic homogenization theory

For consistency, some well-known results of the periodic homogenization theory applied to nonlinear energy densities with specific growth properties are recalled here. Such theoretical results guide the numerical strategy and tell in what sense mechanical quantities are approximated. To fully illustrate the situation, and for the sake of comparison, a synthesis of what is theoretically known for energy densities of several types and for general elliptic operators in divergence form is given. The following results of homogenization of nonconvex energies can be found in the original work of Braides [33] and Müller [145]. For convenience, references are borrowed from the book of Braides and Defranceschi [38]. The theoretical point of view preferably uses the energy minimization problem whereas the PDE approach is more relevant for the numerical practice.

This theory makes use of the growth condition (5.1) introduced in Definition 12 below. In practice, several problems do not satisfy this condition. This is the case for porous materials and Ogden materials. To model porous media, we consider a perforated domain where the energy density satisfies (5.1) and we let the size of the holes, where the energy vanishes, go to zero. In a way this homogenization is also geometric since the domain is not fixed. Unlike porous materials, Ogden materials can violate (5.1) almost everywhere. Therefore the approach introduced in Section 5.4 is still to be justified mathematically. These limitations are summarized in the last paragraph of this section.

5.2.1 Convexity and minimization problems

Throughout the article Ω denotes an open bounded connected subset of $\mathbb{R}^3$. The following definitions and results (see e.g. [161]) will be extensively used in the sequel.

Definition 10 *Given an integer $p \geq 1$, a function $W : \mathcal{M}_3(\mathbb{R}) \to [0, +\infty]$ is $W^{1,p}$-*quasiconvex *if for all $A \in \mathcal{M}_3(\mathbb{R})$ (set of real square matrices of size 3), there exists an open bounded subset E of $\mathbb{R}^3$ with $|\partial E| = 0$ such that:*

$$W(A) = \min \left\{ \frac{1}{|E|} \int_E W(A + \nabla\phi(x))dx \,|\, \phi \in W_0^{1,p}(E; \mathbb{R}^3) \right\}$$

The function W is polyconvex *when it can be expressed as a convex function of the minors of orders 1,2,3 of A.*

Property 1 *If W is polyconvex then W is quasiconvex.*

Definition 11 *Let $(x, A) \mapsto W(x, A)$ be a quasiconvex energy density defined on $\Omega \times \mathcal{M}_3(\mathbb{R})$, for which there exist an integer $p \geq 1$, positive constants c and C, such that for almost all $x \in \Omega$ and for all $A \in \mathcal{M}_3(\mathbb{R})$,*

$$c|A|^p \leq W(x, A) \leq C(1 + |A|^p) \tag{5.1}$$

The function W is then said to satisfy a standard growth condition *(of order p).*

Definition 12 *The function $W : \Omega \times \mathcal{M}_3(\mathbb{R}) \to \mathbb{R}, (x, A) \mapsto W(x, A)$ is a* standard energy density *if W is a quasiconvex Carathéodory function, that is:*

- *$W(\cdot, \cdot)$ is measurable in its first variable and continuous in its second variable*
- *$W(x, \cdot)$ is quasiconvex for almost every $x \in \Omega$*

and if W satisfies (5.1).

Definition 13 *A* standard minimization problem *refers in the literature to a minimization problem associated to a standard energy density that reads: given $\bar{u} \in W^{1,p}(\Omega, \mathbb{R}^3)$, solve*

$$\inf \left\{ \int_\Omega W(x, \nabla(u + \bar{u}))dx \,|\, u \in W_0^{1,p}(\Omega, \mathbb{R}^3) \right\} \tag{5.2}$$

The direct method of the calculus of variations shows

Theorem 27 *For $p > 1$, the minimization problem (5.2) admits at least a minimizer in $W_0^{1,p}(\Omega, \mathbb{R}^3)$.*

Theorem 27 is a consequence of the following lemma.

Lemma 5.1. *If $1 \leq p < \infty$, and $W : \mathcal{M}_3(\mathbb{R}) \to \mathbb{R}$ is a quasiconvex function satisfying*

$$0 \leq W(A) \leq C(1 + |A|^p) \qquad \text{for all } A \in \mathcal{M}_3(\mathbb{R}),$$

then the functional $J(u) = \displaystyle\int_\Omega W(\nabla u)$ is weakly lower semi-continuous on $W^{1,p}(\Omega)$.

5.2.2 Basic homogenization result

Periodic functions are defined as follows.

Definition 14 *A function $\psi : \mathbb{R}^3 \to \mathbb{R}$ is said N-periodic, $N \in \mathbb{N}$, if for almost every (ae) $x \in \mathbb{R}^3$, and for all $(i, j, k) \in \mathbb{N}^3$,*

$$\psi(x + iNe_1 + jNe_2 + kNe_3) = \psi(x)$$

with $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$.

For convenience, in the sequel of the article, only 1-periodic energy densities are considered, instead of general periodic functions. Theoretical results still hold *mutatis mutandis* for periodic functions whose periodic cells have shapes with piecewise regular boundaries.

Periodic homogenization aims at studying problems for which the energy density W_ϵ is of the form

$$W_\epsilon(x, A) = W(\frac{x}{\epsilon}, A),$$

where W is periodic in space. This heterogeneous energy density is commonly used to model composite materials.

The limit $\epsilon \to 0$ of the minimization problem (5.2) with W_ϵ is described by

Theorem 28 *([38], Section 14.2) Let $W : \mathbb{R}^3 \times \mathcal{M}_3(\mathbb{R}) \to [0, +\infty)$ be a standard energy density satisfying the periodicity assumption*

$$W(\cdot, A) \text{ is 1-periodic for all } A \in \mathcal{M}_3(\mathbb{R})$$

and the growth condition (5.1) of order $p \geq 1$.

For Ω a bounded open set of $\mathbb{R}^3$, $u \in W^{1,p}(\Omega, \mathbb{R}^3)$ and $\epsilon > 0$, we set

$$J_\epsilon(u) = \int_\Omega W(\frac{x}{\epsilon}, \nabla u(x))dx$$

Then, for all $\bar{u} \in W^{1,p}(\Omega, \mathbb{R}^3)$,

$$\lim_{\epsilon \to 0} \inf\{J_\epsilon(u + \bar{u}) \,|\, u \in W_0^{1,p}(\Omega)\} = \inf\{J_{hom}(u + \bar{u}) \,|\, u \in W_0^{1,p}(\Omega)\}, \tag{5.3}$$

where $J_{hom}(u) = \int_\Omega W_{hom}(\nabla u(x))dx$ *and* $W_{hom} : \mathcal{M}_3(\mathbb{R}) \to [0, +\infty)$ *is a standard energy density functional defined by the asymptotic homogenization formula*

$$W_{hom}(A) = \lim_{N\to\infty} \frac{1}{N^3} \inf \left\{ \int_{(0,N)^3} W(x, A + \nabla v(x))dx \,|\, v \in W_0^{1,p}((0,N)^3, \mathbb{R}^3) \right\} \tag{5.4}$$

for all $A \in \mathcal{M}_3(\mathbb{R})$. *The function* W_{hom} *satisfies in particular condition (5.1) of order* p.

In addition, if u_ϵ *is a minimizing sequence of* $J_\epsilon(\cdot + \bar{u})$ *on* $W_0^{1,p}(\Omega, \mathbb{R}^3)$ *weakly converging to some* u *in* $W_0^{1,p}(\Omega, \mathbb{R}^3)$, *then* u *is a minimizer of* $J_{hom}(\cdot + \bar{u})$ *on* $W_0^{1,p}(\Omega, \mathbb{R}^3)$.

Remark **6** *([38], Remark 14.6) The asymptotic homogenization formula (5.4) can be replaced by*

$$W_{hom}(A) = \lim_{N\to\infty} \frac{1}{N^3} \inf \left\{ \int_{(0,N)^3} W(x, A + \nabla v(x))dx \,|\, v \in W_\#^{1,p}((0,N)^3, \mathbb{R}^3) \right\} \tag{5.5}$$

where $W_\#^{1,p}((0,N)^3, \mathbb{R}^3)$ *is the set of the restrictions of* N*-periodic functions* v *of* $W_{loc}^{1,p}(\mathbb{R}^3, \mathbb{R}^3)$ *such that* $\int_{(0,N)^3} v = 0$. *Note that the limit* $N \to \infty$ *can be replaced by an infimum on* $N \in \mathbb{N}$ *in (5.4) and (5.5).*

5.2.3 Homogenization for connected media

Connected media are defined as follows.

Definition 15 *Let* E *be an infinite 1-periodic, connected, open subset of* $\mathbb{R}^3$ *(that is in particular a periodic replication of a subset of* $(0,1)^3$*) with a Lipschitz boundary, and* Ω *be a bounded open subset of* $\mathbb{R}^3$. *Given* $\epsilon \in \mathbb{R}_+$, $\Omega \cap \epsilon E$ *is called a connected medium.*

Theorem 28 also holds in a weaker form for connected media:

Theorem 29 *([38], Section 19.1) Let* E *and* Ω *be as in Definition 15. Given a standard energy density* $W : E \times \mathcal{M}_3(\mathbb{R}) \to \mathbb{R}$ *with* $W(\cdot, A)$ *1-periodic for every* $A \in \mathcal{M}_3(\mathbb{R})$, *satisfying condition (5.1) with* $p > 1$, *let* $J_\epsilon : L^p(\Omega; \mathbb{R}^3) \to [0, +\infty]$ *be the functional defined for every* $\epsilon > 0$ *by*

$$J_\epsilon(u) = \begin{cases} \int_{\Omega\cap\epsilon E} W(\frac{x}{\epsilon}, \nabla u(x))dx & \text{if } u_{|\Omega\cap\epsilon E} \in W^{1,p}(\Omega \cap \epsilon E; \mathbb{R}^3), \\ +\infty & \text{otherwise.} \end{cases}$$

Then there exist a constant $k_1 > 0$ *depending on* E *and* p, *and a standard energy density functional* $W_{hom} : \mathcal{M}_3(\mathbb{R}) \to \mathbb{R}$ *satisfying*

$$\frac{c}{k_1}|A|^p \le W_{hom}(A) \le C|(0,1)^3 \cap E|(1 + |A|^p),$$

for all $A \in \mathcal{M}_3(\mathbb{R})$, *and such that, defining the functional* $J_{hom} : L^p(\Omega; \mathbb{R}^3) \to [0, +\infty]$ *by*

$$J_{hom}(u) = \begin{cases} \int_\Omega W_{hom}(\nabla u(x))dx & \text{if } u \in W^{1,p}(\Omega; \mathbb{R}^3), \\ +\infty & \text{otherwise,} \end{cases}$$

we have, for all $\bar{u} \in W^{1,p}(\Omega)$,

$$\lim_{\epsilon\to 0} \inf\{J_\epsilon(u+\bar{u}) \,|\, u \in W_0^{1,p}(\Omega)\} = \inf\{J_{hom}(u+\bar{u}) \,|\, u \in W_0^{1,p}(\Omega)\}.$$

The functional W_{hom} is given by the asymptotic homogenization formula

$$W_{hom}(A) = \lim_{N\to\infty} \inf\left\{ \frac{1}{N^3} \int_{(0,N)^3\cap E} W(x, \nabla v(x) + A)dx \,|\, v \in W_0^{1,p}((0,N)^3;\mathbb{R}^3) \right\}$$

for all $A \in \mathcal{M}_3(\mathbb{R})$.

In addition, if u_ϵ is a minimizing sequence of $J_\epsilon(\cdot + \bar{u})$ on $W_0^{1,p}(\Omega, \mathbb{R}^3)$ weakly converging to some u in $W_0^{1,p}(\Omega, \mathbb{R}^3)$, then u is a minimizer of $J_{hom}(\cdot + \bar{u})$ on $W_0^{1,p}(\Omega, \mathbb{R}^3)$.

5.2.4 Homogenization of elliptic operators in divergence form

If W is differentiable and the minimization problems (5.2) for W_ϵ and W_{hom}, and (5.5) or (5.4) are attained, the minimizers satisfy the Euler-Lagrange equations. In that case we denote by

$$a(x,\xi) = \frac{\partial W}{\partial \xi}(x,\xi).$$

Keeping the notation of Theorem 28, the Euler-Lagrange equation for the minimization of J_ϵ reads

$$\begin{cases} -\text{div}\left(a(\frac{x}{\epsilon}, \nabla u_\epsilon)\right) = f \text{ in } \Omega \\ u_\epsilon = \bar{u} \text{ on } \partial\Omega. \end{cases} \tag{5.6}$$

On the other hand, if $W_\epsilon(x,\cdot)$ is strictly convex for almost every $x \in \Omega$, the one for J_{hom} is

$$\begin{cases} -\text{div}\left(a_{hom}(\nabla u)\right) = f \text{ in } \Omega \\ u = \bar{u} \text{ on } \partial\Omega, \end{cases} \tag{5.7}$$

where a_{hom} is defined by $\mathcal{M}_3(\mathbb{R}) \ni \xi \mapsto a_{hom}(\xi) = \displaystyle\int_{(0,1)^3} a(y, \nabla v_\xi(y) + \xi)dy$ and v_ξ is the periodic solution in $W_\#^{1,p}((0,1)^3, \mathbb{R}^3)$ of

$$-\text{div}\left(a(y, \nabla v_\xi(y) + \xi)\right) = 0, \tag{5.8}$$

the latter equation being called the *cell-problem* (see [38]).

It may be noticed that (5.8) is the Euler-Lagrange equation of (5.5) for $N = 1$. In fact the infimum in (5.5) is attained for $N = 1$, due to convexity. We abusively say that $N = 1$ in the cell-problem (5.8).

Considering non-symmetric operators a, monotonicity assumptions (see [130] and [155]) extend the results of Theorem 28 and provide more precise results on the homogenized operator, as stated in the following theorem.

Theorem 30 *([155], Sections 3.2.4 and 3.3.2) Assume $p \geq 2$. Let p' satisfy $\frac{1}{p} + \frac{1}{p'} = 1$. Let $a : \mathbb{R}^3 \times \mathcal{M}_3(\mathbb{R}) \to \mathcal{M}_3(\mathbb{R}), (x,\xi) \mapsto a(x,\xi)$ be Carathéodory and 1-periodic in x. Assume also that $a(\cdot, 0)$ is bounded and that the following continuity and monotonicity properties hold*

$$\begin{aligned} &\exists\, 0 \leq \alpha \leq p-1, C > 0 \quad | \quad \textit{for ae } x \in \mathbb{R}^3, \forall \xi_1, \xi_2 \in \mathcal{M}_3(\mathbb{R}) \\ &\qquad |a(x,\xi_1) - a(x,\xi_2)| \leq C(1 + |\xi_1| + |\xi_2|)^{p-1-\alpha}|\xi_1 - \xi_2|^\alpha, \end{aligned} \tag{5.9}$$

$$\begin{aligned} &\exists\, 2 \leq \beta < +\infty, c > 0 \quad | \quad \textit{for ae } x \in \mathbb{R}^3, \forall \xi_1, \xi_2 \in \mathcal{M}_3(\mathbb{R}) \\ &\qquad (a(x,\xi_1) - a(x,\xi_2), \xi_1 - \xi_2) \geq c(1 + |\xi_1| + |\xi_2|)^{p-\beta}|\xi_1 - \xi_2|^\beta. \end{aligned} \tag{5.10}$$

Then, given $f \in L^{p'}(\Omega, \mathbb{R}^3)$*, the solution* $u_\epsilon \in W_0^{1,p}(\Omega, \mathbb{R}^3)$ *of*

$$-\text{div}\ (a(\frac{x}{\epsilon}, \nabla u_\epsilon)) = f$$

weakly converges in $W_0^{1,p}(\Omega, \mathbb{R}^3)$ *to the solution* $u \in W_0^{1,p}(\Omega, \mathbb{R}^3)$ *of*

$$-\text{div}\ (a_{hom}(\nabla u)) = f,$$

where $a_{hom} : \mathcal{M}_3(\mathbb{R}) \to \mathcal{M}_3(\mathbb{R})$ *is defined by*

$$a_{hom}(A) = \int_{(0,1)^3} a(y, A + \nabla u_A(y))dy$$

and $u_A \in W_\#^{1,p}((0,1)^3, \mathbb{R}^3)$ *is the solution of*

$$-\text{div}\ (a(y, A + \nabla u_A(y))) = 0.$$

In addition a_{hom} *satisfies (5.10) with the same coefficients as* a *and (5.9) with* $\gamma = \alpha/(\beta - \alpha)$ *instead of* α*.*

The assumptions of Theorem 30 can be modified to deal with $1 < p \leq 2$, see [58].

The existence of a corrector, which allows to obtain a strong convergence instead of the weak convergence of Theorem 30, is given

Theorem 31 *([38, Chapter 23] [58]) Under the hypotheses and notation of Theorem 30, let* $(M_\epsilon)_\epsilon$ *be the set of mean operators defined by*

$$M_\epsilon : L^p(\Omega) \to L^p(\Omega),\ \phi(x) \mapsto M_\epsilon \phi(x) = 1/\epsilon^3 \int_{(\epsilon[Y,(Y+1)])^3} \phi(y),$$

with $Y \in \mathbb{Z}^3$ *such that* $x \in (\epsilon[Y, (Y+1)])^3$*. The set* $\{M_\epsilon \phi\}_\epsilon$ *is a set of piecewise constant functions strongly converging to* ϕ *in* $L^p(\Omega)$*.*

The corrector C_ϵ *associated with* u *is then given by* $C_{\epsilon|\epsilon[Y,(Y+1)]^3}(x) = \nabla v_\epsilon(\frac{x}{\epsilon}) + M_\epsilon \nabla u(x)$ *where*

$$v_\epsilon : \Omega \to \mathbb{R}^3, x \in (\epsilon[Y, (Y+1)])^3 \mapsto v_\epsilon(x) = v_{\epsilon,Y}(\frac{x}{\epsilon}),$$

where $v_{\epsilon,Y} \in W_\#^{1,p}((0,1)^3, \mathbb{R}^3)$ *is the periodic solution of*

$$-\text{div}\ a(y, M_\epsilon \nabla u(x) + \nabla v_{\epsilon,Y}(y)) = 0. \tag{5.11}$$

And the following strong convergence in $L^p(\Omega, \mathbb{R}^3)$ *holds,*

$$\|C_\epsilon - \nabla u_\epsilon\|_{0,p,\Omega} \to 0,$$

$\|\cdot\|_{0,p,\Omega}$ *standing for the norm of* $L^p(\Omega)$*.*

In the remainder of the article all the monotone operators considered will be symmetric and associated to a strictly convex energy density.

5.2.5 Some open issues

The conclusions of Theorem 28 are specific to the growth condition (5.1) and the minimization space $W_0^{1,p}(\Omega, \mathbb{R}^3)$. They can be extended to Neumann boundary conditions and mixed Dirichlet and Neumann boundary conditions in weak form. However, two open questions prevent from applying rigorously the above results to general nonlinear elasticity whose energy densities do not satisfy (5.1), e.g. general polyconvex energies.

First, it is not known whether Theorem 28 holds if $W_0^{1,p}(\Omega)$ is replaced by the set $\{u \in W_0^{1,p}(\Omega) \mid \det(\nabla u + \nabla \bar{u}) = 1 \text{ ae}\}$. This variational set models incompressible materials.

The second open question deals with the more general problem of Γ-convergence of sequentially lower semicontinuous functionals. Γ-convergence is an approach which can be applied to prove Theorem 28 (see e.g. [38]). In this case, it requires the growth condition (5.1), which is also used to prove the lower semicontinuity of the integral functional J_ϵ of Theorem 28. Given that other mathematical properties than the growth condition (5.1) can ensure the lower semicontinuity of the functional (typically the polyconvexity), a natural issue would be to try to generalize the application of the Γ-convergence theory to general sequentially lower semicontinuous functionals, which is still an open issue today. It is to be noticed that polyconvexity can be lost by homogenization, as shown in [34].

The answers to several questions related to what has been recalled in this section for different types of energies and operators in divergence form are collected in Table 5.1. The number N of cells to consider in (5.4) and (5.5) to attain convergence depends on the problem at stake and on the functional space. The question relative to the existence of correctors is of importance since it allows to recover strong convergence of minimizers. The existence of correctors is extensively used to derive error estimates for numerical methods, such as e.g. the multiscale finite element method proposed by Hou and co-authors ([105], [72]). For minimization problems with quasiconvex energies, the minimizers are not necessarily unique, thus equation (5.11) has not a unique solution and does not properly define a corrector.

Type		Homogenized	Existence of correctors	Number of cells N to consider in	
Operator		**Operator**		problem (5.8)	
Linear	+cc[a]	Linear+cc	True	1	
Monotone	+(5.10)+(5.9)	Monotone +(5.10)+(5.9)	True	1	
Energy density		**Energy density**		formula (5.4)	formula (5.5)
Convex	+sgc[b]	Convex+sgc	True	∞	1
Quasiconvex	+sgc	Quasiconvex+sgc	? [c]	∞	∞
	+polyconvex +sgc	Quasiconvex+sgc	?	∞	∞
	+polyconvex	?	?	?	?

Table 5.1. Summary of some homogenization results available to date

[a] coercive and continuous on $W^{1,2}$, in order to apply Lax-Milgram lemma
[b] standard growth condition of order p
[c] unknown today

A last comment concerns the issue of non-periodic homogenization. General compactness results exist for general homogenization problems without periodicity assumptions in the framework of Γ-convergence ([38]). However the Γ-limit may depend on the extraction considered and is not given by any homogenization formula as (5.4). Therefore it cannot be computed by a direct method as the one developed throughout the present work. The question of numerical homogenization of non-periodic elliptic problems will be addressed in Chapter 6.

5.3 Approximation result for the standard homogenization problem

Two results of approximation are recalled in details for nonlinear elasticity boundary problems in the standard case. They show that isolated minimizers u of (5.2) can be approximated by minimizers u_h of

$$\inf\left\{\int_\Omega W(x,\nabla(u_h+\bar{u}))dx \,|\, u_h\in V_h\right\},$$

where (V_h) are finite dimensional subspaces of $W_0^{1,p}(\Omega,\mathbb{R}^3)$. Under an assumption on the form of the energy density W (Theorem 32) or up to adding a vanishing perturbing term to the energy density (Theorem 33), u_h converges to u in $W_0^{1,p}(\Omega,\mathbb{R}^3)$. In Section 5.3.2 a similar result is proven for the standard homogenization problem (Theorem 34), in the context of Theorem 28. In Section 5.3.3 an error estimate is derived for a nonlinear elasticity problem with a strictly convex energy density in the context of Section 5.2.4.

5.3.1 Approximation theory for standard energy densities

The first result of this section is classical. The proof, given for completeness, is simpler than that of Le Tallec in [125] due to the restricted class of energies considered. The second result exploits the idea of Pedregal in [160] to get rid of the assumption on the form of the energy density, by adding a vanishing perturbing term.

Definition 16 *Let $W:\Omega\times\mathcal{M}_3(\mathbb{R})\to\mathbb{R}$ be a standard energy density. The integral functional J is defined by*

$$\begin{aligned} J: W^{1,p}(\Omega) &\to \mathbb{R}\\ v \quad &\mapsto \int_\Omega W(x,\nabla v(x))dx. \end{aligned}$$

Given $\bar{u}\in W^{1,p}(\Omega)$, we consider an isolated minimizer (strict local minimizer) u of

$$\inf\{J(v+\bar{u}), v\in W_0^{1,p}(\Omega)\}$$

on $B(u,\bar{r})$, $\bar{r}>0$, that is a minimizer such that

$$J(u+\bar{u})<J(v+\bar{u}) \quad \forall v\in W_0^{1,p}(\Omega,\mathbb{R}^3) \quad | \quad \|u-v\|_{1,p}<\bar{r} \text{ and } v\neq u.$$

In the remainder of the paper, for $v\in W^{1,p}(\Omega)$ and $\rho>0$, the open ball centered at v and of radius ρ in $W^{1,p}(\Omega)$ is denoted by $B(v,\rho)$.

The following lemma on Carathéodory functions will be used in the proofs of the approximation results.

Lemma 5.2. *[117] Let Φ be a function of the type*

$$\Phi(y)(x)=\zeta(x,y(x)),$$

where ζ is Carathéodory and such that Φ sends $L^p(\Omega)$ into $L^q(\Omega)$. Then Φ is continuous from $L^p(\Omega)$ into $L^q(\Omega)$.

Theorem 32 *Let W, $\bar{u}$, J, u and $B(u,\bar{r})$ be as in Definition 16. Let W satisfy:*

$$W(x,A)=c(x)|A|^p+W_1(x,A) \tag{5.12}$$

with $c(x)\geq c/2$ a.e. in Ω and W_1 a standard energy density. Assume that there exist discrete spaces $V_h\subset W_0^{1,p}(\Omega,\mathbb{R}^3)$ and a sequence $\{w_h\}_h$, $w_h\in V_h\cap B(u,\bar{r})$, satisfying $u=\lim_{h\to 0} w_h$ in $W^{1,p}(\Omega,\mathbb{R}^3)$.
Then the minimum values

$$\inf\{J(v_h+\bar{u})\,|\,v_h\in V_h\cap B(u,\bar{r})\} \tag{5.13}$$

are attained, and any sequence u_h of minimizers of (5.13) converges to u in $W^{1,p}(\Omega)$.

Proof

Since V_h is finite dimensional, the subset $V_h \cap B(u,\bar{r}) \ni w_h$ is non empty, bounded and closed in $W^{1,p}(\Omega)$, it is therefore a compact subset. As J is continuous on $W^{1,p}(\Omega)$ (Lemma 5.2) and $V_h \cap B(u,\bar{r}) \ni w_h$ is compact, $J(\cdot + \bar{u})$ attains its minimum on $V_h \cap B(u,\bar{r})$. Let u_h denote one of the minimizers.

The sequence u_h is bounded by $\|u\|_{1,p} + \bar{r}$ in $W^{1,p}(\Omega)$. Thus there exists an extracted sequence, still denoted by u_h, that converges weakly in $W^{1,p}(\Omega)$ to some $u_\infty \in B(u,\bar{r})$.

By definition of u_h,

$$J(u_h + \bar{u}) \leq J(w_h + \bar{u}) \text{ for all } h.$$

As J is lower semi-continuous for the weak topology (Lemma 5.1) and continuous for the strong topology of $W^{1,p}(\Omega)$, the inequalities

$$\begin{aligned} J(u_\infty + \bar{u}) &\leq \liminf J(u_h + \bar{u}) \\ &\leq \lim J(w_h + \bar{u}) \\ &= J(u + \bar{u}) \end{aligned}$$

hold.

Since u is the unique minimizer of $J(\cdot + \bar{u})$ on $B(u,\bar{r})$, the latter inequality implies

$$u_\infty = u \text{ and } \lim J(u_h + \bar{u}) = J(u + \bar{u})$$

The limit $u_\infty = u$ being independent from the extraction, the whole original sequence u_h converges weakly to u in $W^{1,p}(\Omega)$.

Next, the strong convergence comes from the particular form of the energy functional: since $(x,A) \mapsto c(x)|A|^p$ and $(x,A) \mapsto W_1(x,A)$ are quasiconvex for almost every $x \in \Omega$ and satisfy (5.1), the integral functionals associated to these two energies are lower semi-continuous (Lemma 5.1), which implies

$$\int_\Omega c(x)|\nabla(u(x) + \bar{u}(x))|^p dx \leq \liminf \int_\Omega c(x)|\nabla(u_h(x) + \bar{u}(x))|^p dx$$

and

$$\int_\Omega W_1(x, u(x) + \bar{u}(x))dx \leq \liminf \int_\Omega W_1(x, u_h(x) + \bar{u}(x))dx.$$

As in addition $\lim J(u_h + \bar{u}) = J(u + \bar{u})$, necessarily

$$\lim \int_\Omega c(x)|\nabla(u_h(x) + \bar{u}(x))|^p dx = \int_\Omega c(x)|\nabla(u(x) + \bar{u}(x))|^p dx \tag{5.14}$$

Since Ω is bounded, combining the weak convergence of $\nabla(\bar{u} + u_h)$ to $\nabla(\bar{u} + u)$ in $L^p(\Omega)$ with (5.14), the strong convergence of $\nabla(\bar{u} + u_h)$ holds, and consequently:

$$\|u - u_h\|_{1,p} \to 0$$

□

Energy densities satisfying (5.12) are indeed the ones that are most encountered in practice. If an energy density does not satisfy (5.12), the addition of a vanishing term allows to recover strong convergence, as stated by

Theorem 33 *Let W, $\bar{u}$, J, u and $B(u,\bar{r})$ be as in Definition 16. Assume that there exist discrete spaces $V_h \subset W_0^{1,p}(\Omega,\mathbb{R}^3)$ and a sequence $\{w_h\}_h$, $w_h \in V_h \cap B(u,\bar{r})$, satisfying $u = \lim_{h\to 0} w_h$ in $W^{1,p}$. Let $\{J_\eta\}$ be the set of perturbed energy functionals defined by*

$$J_\eta(v) = J(v) + \eta \int_\Omega |\nabla(v)|^p, \tag{5.15}$$

for $\eta > 0$ and $v \in W^{1,p}(\Omega)$.
The minimum values

$$\inf\{J_\eta(v_h + \bar{u}) \mid v_h \in V_h \cap B(u, \bar{r})\} \tag{5.16}$$

are attained and minimizers of $J_\eta(\bar{u} + \cdot)$ on $V_h \cap B(u, \bar{r})$ are denoted by $u_{\eta,h}$.
For any extracting function ϕ_η such that the sequence $u_{\eta,\phi_\eta(h)}$ weakly converges in $W^{1,p}(\Omega, \mathbb{R}^3)$ as h goes to zero, the convergence is actually strong and

$$\lim_{\eta\to 0} \lim_{h\to 0} u_{\eta,\phi_\eta(h)} = u \quad \text{in } W^{1,p}(\Omega, \mathbb{R}^3).$$

Proof

Following [160] the perturbed energy density is obtained by adding the term $\eta \int_\Omega |\nabla v|^p$ to J, for $\eta \in \mathbb{R}^+$,

$$J_\eta(v) = J(v) + \eta \int_\Omega |\nabla v|^p.$$

The following three assertions hold:

(i) $J_\eta(\cdot + \bar{u})$ has at least a minimizer u_η in $B(u, \bar{r})$
(ii) u_η is a minimizing sequence of $J(\cdot + \bar{u})$ on $B(u, \bar{r})$
(iii) $u_\eta \to u$ strongly in $W^{1,p}(\Omega)$ as $\eta \to 0$

Assertion (i) is a consequence of the lower semi-continuity of J_η, as the sum of two lower semi-continuous functionals.

Assertion (ii) follows from the inequalities

$$J(u + \bar{u}) \le J(u_\eta + \bar{u}) \le J_\eta(u_\eta + \bar{u}) \le J_\eta(u + \bar{u}), \tag{5.17}$$

where have been successively used that u minimizes $J(\cdot + \bar{u})$ and u_η minimizes $J_\eta(\cdot + \bar{u})$ on $B(u, \bar{r})$. Next, for all $v \in W^{1,p}(\Omega)$, $\lim_{\eta\to 0} J_\eta(v) = J(v)$, thus (5.17) implies that $J(u_\eta + \bar{u}) \to J(u + \bar{u})$ as $\eta \to 0$, that is (ii).

To prove (iii), it may first be noticed that (ii) implies that

$$u_\eta \rightharpoonup u \quad \text{in } W^{1,p}(\Omega), \tag{5.18}$$

arguing as in the proof of Theorem 32 since u is the unique minimizer of $J(\cdot + \bar{u})$ on $B(u, \bar{r})$.

The following observation

$$\begin{aligned}
\eta \int_\Omega |\nabla u_\eta + \nabla \bar{u}|^p &\le \eta \int_\Omega |\nabla u_\eta + \nabla \bar{u}|^p + \frac{1}{\eta}(J(u_\eta + \bar{u}) - J(u + \bar{u})) \\
&= \frac{1}{\eta} J_\eta(u_\eta + \bar{u}) - \frac{1}{\eta} J(u + \bar{u}) \\
&\le \frac{1}{\eta} J_\eta(u + \bar{u}) - \frac{1}{\eta} J(u + \bar{u}) \\
&= \eta \int_\Omega |\nabla u + \nabla \bar{u}|^p,
\end{aligned} \tag{5.19}$$

combined with (5.18), implies (iii).

As in the proof of Theorem 32, the set of minimizers $\{u_{\eta,h}\}$ of $J_\eta(\bar{u} + \cdot)$ on $V_h \cap B(u, \bar{r})$ is weakly compact. Thus there exists a subsequence $\{u_{\eta,\phi_\eta(h)}\}_h$ which weakly converges to some $u_\eta \in B(u, \bar{r})$. Due to the perturbing term $\eta \int_\Omega |\nabla v|^p$, J_η satisfies (5.12) and (5.14) holds with $c(x) = \eta$. This implies the strong convergence of the subsequence in $W^{1,p}(\Omega)$. Combined with assertion (iii), it proves

$$\lim_{\eta\to 0} \lim_{h\to 0} u_{\eta,\phi_\eta(h)} = u \quad \text{in } W^{1,p}(\Omega).$$

□

5.3.2 Approximation result for a homogenized energy density

This section is devoted to the proof of a result of approximation for a problem of type (5.2) with the energy density (5.4), when $W_0^{1,p}(\Omega,\mathbb{R}^3)$ in (5.2) and $W_0^{1,p}((0,N)^3,\mathbb{R}^3)$ in (5.4) are replaced by finite dimensional subspaces.

Definition 17 *For $N \in \mathbb{N}$, $\{V_{N,h}\}$ is a family of finite dimensional subspaces of $W_0^{1,p}((0,N)^3,\mathbb{R}^3)$ satisfying*

$$h_2 \le h_1 \quad \Longrightarrow \quad V_{N,h_1} \subset V_{N,h_2},$$

and such that $\overline{\cup_h V_{N,h}} = W_0^{1,p}((0,N)^3,\mathbb{R}^3)$.
Similarly, $\{V_{\Omega,H}\}$ is a family of finite dimensional subspaces of $W_0^{1,p}(\Omega,\mathbb{R}^3)$ such that $\overline{\cup_H V_{\Omega,H}} = W_0^{1,p}(\Omega,\mathbb{R}^3)$.

Given a standard energy density W and the homogenized energy density W_{hom} associated by formula (5.4), for any (N,h) the approximate homogenized energy density $W^{N,h} : \mathcal{M}_3(\mathbb{R}) \to \mathbb{R}$ is defined by

$$A \mapsto W^{N,h}(A) = \frac{1}{N^3} \inf \left\{ \int_{(0,N)^3} W(x, A + \nabla v(x))dx : v \in V_{N,h} \right\}. \tag{5.20}$$

Its associated approximate energy functional is

$$J^{N,h}(v) = \int_\Omega W^{N,h}(\nabla v) \quad \text{on } W^{1,p}(\Omega). \tag{5.21}$$

The approximation result is given by

Theorem 34 *Let W, W_{hom}, $V_{N,h}$, $V_{\Omega,H}$ and $J^{N,h}$ be as in Definition 17. Assume that W_{hom} also satisfies (5.12) (see however Remark 7 below).*
Given $\bar{u} \in W^{1,p}(\Omega)$, u is defined as an isolated minimizer of

$$\inf \left\{ \int_\Omega W_{hom}(\nabla(v + \bar{u}))dx \,|\, v \in W_0^{1,p}(\Omega) \right\}. \tag{5.22}$$

The minimum values

$$\inf \left\{ J^{N,h}(v + \bar{u})dx \,|\, v \in V_{\Omega,H} \cap B(u,\bar{r}) \right\} \tag{5.23}$$

are attained and let $\{u_H^{N,h}\}_{H,N,h}$ denote sequences of minimizers of (5.23).

Then, for all extracted sequences in N and h, still denoted by $u_H^{N,h}$, such that $\lim_{N\to\infty} \lim_{h\to 0} u_H^{N,h}$ exists,

$$\lim_{H\to 0} \lim_{N\to\infty} \lim_{h\to 0} u_H^{N,h} = u \quad \text{in } W^{1,p}(\Omega). \tag{5.24}$$

Proof

Step 1: convergence in H

In view of Theorem 28, W_{hom} is a quasiconvex function satisfying (5.1). In addition, W_{hom} satisfies (5.12) by assumption. The application of Theorem 32 to the minimization problem (5.22) implies that any sequence u_H of minimizers of

$$\inf \left\{ \int_\Omega W_{hom}(\nabla(v + \bar{u}))dx \,|\, v \in V_{\Omega,H} \cap B(u,\bar{r}) \right\}, \tag{5.25}$$

converges to u:

$$\lim_{H\to 0} u_H = u \quad \text{in } W_0^{1,p}(\Omega). \tag{5.26}$$

Step 2: convergence in N

In the limit taken in (5.4), let us consider the approximate energy density

$$W^N(A) = \inf\left\{\frac{1}{N^3}\int_{(0,N)^3} W(x, \nabla v(x) + A)dx : v \in W_0^{1,p}((0,N)^3, \mathbb{R}^3)\right\}, \tag{5.27}$$

and define an approximation of problem (5.25) by replacing W_{hom} by W^N:

$$\inf\left\{\int_\Omega W^N(\nabla(v + \bar{u})) \,|\, v \in V_{\Omega,H} \cap B(u, \bar{r})\right\}. \tag{5.28}$$

In the sequel, $J^N(v) = \displaystyle\int_\Omega W^N(\nabla v)dx$.

The second step consists in proving that for all $N \in \mathbb{N}^*$ the minimum value (5.28) is attained and that any converging subsequence of minimizers u_H^N of (5.28) converges in $W^{1,p}(\Omega)$ to a minimizer u_H of (5.25) as N goes to infinity.

Let us prove that W^N is a continuous function on $\mathcal{M}_3(\mathbb{R})$. Let $(A_i)_{i\in\mathbb{N}} \in (\mathcal{M}_3(\mathbb{R}))^{\mathbb{N}}$ satisfy $A_i \to A \in \mathcal{M}_3(\mathbb{R})$ and let $(u_{A_i})_i$ and and u_A be minimizers of

$$\inf\left\{\frac{1}{N^3}\int_{(0,N)^3} W(x, \nabla v(x) + B)dx : v \in W_0^{1,p}((0,N)^3, \mathbb{R}^3)\right\},$$

with $B = A_i$ and $B = A$ respectively.
From Theorem 27 such minimizers exist. Thanks to (5.1), $(u_{A_i})_i$ is bounded in $W_0^{1,p}((0,N)^3)$. There exists a subsequence, still denoted by u_{A_i}, which weakly converges to some $u_{A_\infty} \in W_0^{1,p}((0,N)^3)$. Lemma 5.2 ensures that $v \mapsto \displaystyle\int_\Omega W(x, \nabla v)$ is continuous on $W^{1,p}(\Omega)$, which implies

$$\int_{(0,N)^3} W(x, A + \nabla u_A) = \lim_{i\to\infty}\int_{(0,N)^3} W(x, A_i + \nabla u_A). \tag{5.29}$$

By definition of u_{A_i} and after taking the inferior limit, we have

$$\liminf_{i\to\infty}\int_{(0,N)^3} W(x, A_i + \nabla u_A) \geq \liminf_{i\to\infty}\int_{(0,N)^3} W(x, A_i + \nabla u_{A_i}). \tag{5.30}$$

The lower semi-continuity of $v \mapsto \displaystyle\int_{(0,N)^3} W(x, \nabla v)$ for the weak topology of $W^{1,p}((0,N)^3)$ implies

$$\liminf_{i\to\infty}\int_{(0,N)^3} W(x, A_i + \nabla u_{A_i}) \geq \int_{(0,N)^3} W(x, A + \nabla u_{A_\infty}). \tag{5.31}$$

Combining (5.29), (5.30) and (5.31) gives

$$\int_{(0,N)^3} W(x, A + \nabla u_A) \geq \int_{(0,N)^3} W(x, A + \nabla u_{A_\infty}). \tag{5.32}$$

The definition of u_A and (5.32) then imply

$$W^N(A) = \int_{(0,N)^3} W(x, A + \nabla u_A) = \int_{(0,N)^3} W(x, A + \nabla u_{A_\infty}).$$

Therefore $\lim\limits_{i\to\infty} W^N(A_i) = W^N(A)$ does not depend either on the sequence A_i nor on the subsequence u_{A_i} considered, which proves the continuity of W^N. Consequently, W^N is Carathéodory and satisfies (5.1). The same result and proof hold for $W^{N,h}$.

Lemma 5.2 implies that J^N and $J^{N,h}$ are continuous on $W^{1,p}(\Omega)$. The same property holds for J_{hom} since W_{hom} is also a Carathéodory function satisfying (5.1) (Theorem 28). As $V_{\Omega,H} \cap B(u, \bar{r})$ is a compact set of $W_0^{1,p}(\Omega)$ and J^N is continuous, the minimum value (5.28) is attained.

Let $\{u_H^N\}_N$ be a sequence of minimizers of (5.28). As $V_{\Omega,H}$ is compact, there exists a subsequence $u_H^{\phi_H(N)}$ which converges to some u_H^∞ in $W^{1,p}(\Omega)$. To prove that u_H^∞ is a minimizer of (5.25), it suffices to show that J^N converges to J_{hom} uniformly on $V_{\Omega,H} \cap B(u,\bar{r})$.

For all $\chi \in W_0^{1,p}((0,2^N)^3)$, let $\chi^* \in W_0^{1,p}((0,2^{N+1})^3)$ denote the function obtained by the periodization of χ. Consequently $W^{2^{N+1}}(A) \leq W^{2^N}(A)$ for all $A \in \mathcal{M}_3(\mathbb{R})$, which implies $J^{2^{N+1}}(v) \leq J^{2^N}(v)$ for all $v \in W^{1,p}(\Omega)$. As J^{2^N} is a decreasing sequence of continuous functions which converges to a continuous function J_{hom} on the compact set $V_{\Omega,H} \cap B(u,\bar{r})$, Dini's theorem implies that J^{2^N} converges uniformly to J_{hom} on $V_{\Omega,H} \cap B(u,\bar{r})$. Actually this shows that the whole sequence J^N converges uniformly on $V_{\Omega,H} \cap B(u,\bar{r})$, as proved below.

For all $\epsilon > 0$, there exists $I \in \mathbb{N}$ such that for all $v \in V_{\Omega,H} \cap B(u,\bar{r})$,

$$|J^{2^I}(v) - J_{hom}(v)| \leq \epsilon. \tag{5.33}$$

For all $M \geq 2^I$ and $v \in V_{\Omega,H} \cap B(u,\bar{r})$, *either* $J^M(v) \leq J^{2^I}(v)$ and

$$|J^M(v) - J_{hom}(v)| \leq \epsilon, \tag{5.34}$$

since $J_{hom} \leq J^N$ for all $N \in \mathbb{R}$ (Remark 6), *or*

$$0 \leq J^M(v) - J^{2^I}(v). \tag{5.35}$$

For all $\chi \in W_0^{1,p}((0,2^I)^3)$, let $\chi^{**} \in W_0^{1,p}((0,M)^3)$ be defined by $\left[\dfrac{M}{2^I}\right]^3$ replications of χ on $\left(0, \left[\dfrac{M}{2^I}\right] 2^I\right)^3$ and be extended by zero elsewhere in $(0,M)^3$, where $[\cdot]$ stands for the integer part. Thus

$$\int_{(0,M)^3} W(x, A + \chi^{**}(x)) = \left[\frac{M}{2^I}\right]^3 \int_{(0,2^I)^3} W(x, A + \chi(x)) + \int_{(\left[\frac{M}{2^I}\right]2^I, M)^3} W(x, A),$$

which implies, using (5.1),

$$J^M(v) - J^{2^I}(v) \leq \frac{M^3 - \left(\left[\frac{M}{2^I}\right] 2^I\right)^3}{M^3} C(1 + \|v\|_{1,p}^p). \tag{5.36}$$

As $\|v\|_{1,p} \leq \|u\|_{1,p} + \bar{r}$, the right hand side of (5.36) converges to zero uniformly on $V_{\Omega,H} \cap B(u,\bar{r})$ when M goes to infinity. Therefore, there exists $N^* \geq 2^I$ such that for all $v \in V_{\Omega,H} \cap B(u,\bar{r})$ and $M \geq N^*$ either (5.34) holds or

$$|J^M(v) - J_{hom}(v)| \leq |J^{2^I}(v) - J_{hom}(v)| + |J^{2^I}(v) - J^M(v)| \leq 2\epsilon,$$

by combining (5.33), (5.35) and (5.36). This implies the uniform convergence of J^N to J_{hom} on $V_{\Omega,H} \cap B(u,\bar{r})$.

We are now in position to prove that u_H^∞ is a minimizer of J_{hom} on $V_{\Omega,H} \cap B(u,\bar{r})$. The triangle inequality implies

$$|J^{\phi_H(N)}(u_H^{\phi_H(N)}) - J_{hom}(u_H^\infty)| \leq |J^{\phi_H(N)}(u_H^{\phi_H(N)}) - J_{hom}(u_H^{\phi_H(N)})| + |J_{hom}(u_H^{\phi_H(N)}) - J_{hom}(u_H^\infty)|.$$

The first term goes to zero independently of $u_H^{\phi_H(N)}$ thanks to the uniform convergence of $J^{\phi_H(N)}$ whereas the second term goes to zero thanks to the continuity of J_{hom}.

Thus, $\lim_{N\to\infty} J^{\phi_H(N)}(u_H^{\phi_H(N)}) = J_{hom}(u_H^\infty)$. In addition, for all $v \in V_{\Omega,H} \cap B(u,\bar{r})$ and $N \in \mathbb{N}$, $J^{\phi_H(N)}(v) \geq J^{\phi_H(N)}(u_H^{\phi_H(N)})$. Passing to the limit, we obtain $J_{hom}(v) \geq J_{hom}(u_H^\infty)$, which implies that u_H^∞ is a minimizer of J_{hom} on $V_{\Omega,H} \cap B(u,\bar{r})$.

For any extraction function ϕ_H such that the sequence $u_H^{\phi_H(N)}$ of minimizers of (5.28) converges as N goes to infinity, (5.26) then shows

$$\lim_{H\to 0} \lim_{N\to\infty} u_H^{\phi_H(N)} = u \quad \text{in } W^{1,p}(\Omega). \tag{5.37}$$

Step 3: convergence in h

Step 3 consists in determining an adequate approximation of u_H^N by restricting (5.25) to a finite dimensional subspace as prescribed by Definition 17.

For $h_1 \geq h_2$, $V_{N,h_1} \subset V_{N,h_2}$, thus, for all $A \in \mathcal{M}_3(\mathbb{R})$, $W^{N,h_2}(A) \leq W^{N,h_1}(A)$, showing that $\{J^{N,h}\}_h$ is a decreasing sequence of functions. As $J^{N,h}$ and J^N are continuous on $W^{1,p}(\Omega)$, hypotheses of Dini's theorem hold and $J^{N,h}$ converges uniformly to J^N on $V_{\Omega,H} \cap B(u,\bar{r})$. Arguing as in *Step 2*, there exists a subsequence $u_H^{N,\psi_{N,H}(h)}$ of minimizers of (5.23) which converges to a minimizer $u_H^{N,0}$ of (5.28) in $W^{1,p}(\Omega)$ as h goes to zero.

For any extraction function $\psi_{N,H}$ such that the sequence $u_H^{N,\psi_{N,H}(h)}$ of minimizers of (5.22) converges as h goes to zero, (5.37) finally shows

$$\lim_{H\to 0} \lim_{N\to\infty} \lim_{h\to 0} u_H^{\phi_H(N),\psi_{\phi_H(N),H}(h)} = u \quad \text{in } W^{1,p}(\Omega). \tag{5.38}$$

□

Let err be an error range, Theorem 34 implies that there exist N, H and h depending on err such that

$$\|u - u_H^{N,h}\|_{1,p} \leq err. \tag{5.39}$$

Whereas minimizers u_H of (5.25) cannot be computed directly (W_{hom} is not available analytically), $u_H^{N,h}$ can **actually** be computed by a finite element method.

Remark **7** *The energy density W_{hom} does not satisfy (5.12) in general (see [94]), even if $W(x,\cdot)$ does satisfy it almost everywhere in x. Theorem 34 has been stated this way for the sake of simplicity. In general the homogenized energy density has to be modified as in (5.15), which leads to a straighforward adaptation of Theorem 34; Theorem 33 then allows to pass to the limit as η goes to zero, showing, with obvious notation,*

$$\lim_{\eta\to 0} \lim_{H\to 0} \lim_{N\to\infty} \lim_{h\to 0} u_{\eta,H}^{N,h} = u. \tag{5.40}$$

The question of the existence of an isolated minimizer u is partly discussed in Section 5.6 in view of [94].

Remark **8** *Theorem 34 has been stated and proved in the framework of formula (5.4). The proofs and result also hold with formula (5.5) and straightforward adaptations, replacing $W_0^{1,p}((0,N)^3)$ by $W_\#^{1,p}((0,N)^3)$. In particular, when dealing with a convex energy density and $W_\#^{1,p}((0,N)^3)$, the limit in N in (5.38) can be skipped.*

5.3.3 Error estimates in the convex case

Theorem 34 does not provide the explicit dependence of H, N and h upon err in (5.39). For general quasiconvex energy densities, no error estimate can be derived to completement the approximation result since there is no general error estimate related to Theorem 32. Theorem 34 remains thus abstract. However, in some particular cases, it turns out to be possible to give an error estimate.

The analysis is more difficult for N than for H and h, as N is indeed closely linked to buckling phenomena (see [94] and [145]) in the cell-problem and strongly depends on the load A in (5.4) and not only on err. This issue is investigated numerically in Section 5.6.

In the example of a convex energy density treated here, the general analysis for N need not be handled, since formula (5.5) applies with $N = 1$ (see Table 5.1). It is thus enough to deal with $W^{1,h}$ and $J^{1,h}$ defined by (5.20) and (5.21), where $V_{1,h}$ is a subspace of $W^{1,p}_{\#}((0,1)^3, \mathbb{R}^3)$ (see Remark 8). With the notation of Theorem 34 and Remark 7, (5.40) is replaced by

$$\lim_{\eta\to 0}\lim_{H\to 0}\lim_{h\to 0} u^{1,h}_{\eta,H} = u \quad \text{in } W^{1,p}(\Omega, \mathbb{R}^3). \tag{5.41}$$

It will be seen next that modifying the homogenized energy as in Remark 7 is indeed unnecessary in this case.

Two different error estimates for convex energy densities associated to symmetric monotone operators are presented: one relying on the continuity property (5.9) with $\alpha > 0$ and another one also valid for $\alpha = 0$ (see Theorems 35 and 36 below). The proofs of these theorems are based on regularity properties of the solutions to monotone elliptic systems.

Optimal regularity results of Savaré ([169]) and of Ebmeyer-05 et al. ([71]) are recalled for symmetric monotone systems on Lipschitz and convex domains. An error estimate is then obtained for the cell-problem and finally a global error estimate is derived for problem (5.23).

Hypotheses 1 *The energy density $W : \mathbb{R}^3 \times \mathcal{M}_3(\mathbb{R}) \ni (y,\xi) \mapsto W(y,\xi) \in \mathbb{R}$ is a continuous function, 1-periodic in y and convex in ξ for almost every $y \in (0,1)^3$ that satisfies (5.1) with $p \geq 2$. The operator $a := \dfrac{\partial W}{\partial \xi}$ is monotone and continuous in the sense of (5.10) and (5.9), and $a(y,0) = 0$ for all $y \in (0,1)^3$ (without loss of generality). In addition W enjoys the following uniform Lipschitz property with respect to y,*

$$\exists L > 0 : |W(y_1,\xi) - W(y_2,\xi)| \leq L|y_1 - y_2|(1 + |\xi|^p), \quad \forall y_1, y_2 \in \mathbb{R}^3, \quad \forall \xi \in \mathcal{M}_3(\mathbb{R}).$$

Lemma 5.3. *Under Hypotheses 1 and for $p > 2$, let $\mathcal{O}$ be an open bounded domain of $\mathbb{R}^3$, $\bar{u} \in W^{1+2/p,p}(\mathcal{O}, \mathbb{R}^3)$ and $u \in W^{1,p}_0(\mathcal{O}, \mathbb{R}^3)$ be the solution of*

$$-\text{div } a(x, \nabla(u + \bar{u})) = 0.$$

Then,

- *if $\mathcal{O}$ is Lipschitz, $u \in W^{1+\lambda/p,p}_0(\mathcal{O}, \mathbb{R}^3)$ for all $\lambda \in [0,1[$ ([169], Theorem 2),*
- *if $\mathcal{O}$ is convex, $u \in W^{1+\lambda/p,p}_0(\mathcal{O}, \mathbb{R}^3)$ for all $\lambda \in [0,2[$ ([71], Theorem 2.1 and Remark 2.2).*

Remark **9** *Lemma 5.3 also holds when $W^{1,p}_0$ and $W^{1+\lambda/p,p}_0$ are respectively replaced by $W^{1,p}_{\#}$ and $W^{1+\lambda/p,p}_{\#}$.*

For simplicity, in the remainder of the section, Ω is supposed to be convex.

Definition 18 *With the notation of Hypotheses 1 and forall $A \in \mathcal{M}_3(\mathbb{R})$, u_A is defined as the unique solution in $W^{1,p}_{\#}((0,1)^3, \mathbb{R}^3)$ of*

$$-\text{div } a(x, \nabla u_A + A) = 0. \tag{5.42}$$

Given a family $\{V_h\}_h$ of finite dimensional subspaces of $W^{1,p}_{\#}((0,1)^3, \mathbb{R}^3)$ such that $\overline{\cup_h V_h} = W^{1,p}_{\#}((0,1)^3, \mathbb{R}^3)$, u^h_A denotes an approximation of u_A in V_h defined as the unique solution in V_h of the variational problem

$$\int_{(0,1)^3} a(x, \nabla u^h_A + A) \cdot \nabla v_h = 0 \qquad \forall\, v_h \in V_h. \tag{5.43}$$

Lemma 5.4. *Assume Hypotheses 1 and in addition $\alpha \geq 0$ in (5.9) and $p \geq \beta \geq 2$ in (5.10), and let $A \in \mathcal{M}_3(\mathbb{R})$, $u_A \in W^{1,p}_{\#}((0,1)^3, \mathbb{R}^3)$ be the solution of (5.42) and u^h_A be the solution of (5.43). Then there exists a constant $C > 0$ independent of h such that*

$$\|u_A - u^h_A\|_{1,p} \leq C \inf \left\{ \|u_A - v_h\|^s_{1,p}, v_h \in V_h \right\}, \tag{5.44}$$

with $s = \dfrac{1}{\beta \wedge p - \alpha}$.

Proof

Since u_A and u^h_A are solutions to (5.42) and (5.43), for any $v_h \in V_h$,

$$\begin{aligned} & \int_{(0,1)^3} \big(a(x, A + \nabla u_A) - a(x, A + \nabla u^h_A)\big) \cdot (u_A - u^h_A) \\ = & \int_{(0,1)^3} \big(a(x, A + \nabla u_A) - a(x, A + \nabla u^h_A)\big) \cdot (u_A - v_h) \\ \leq & \, C(1 + 2|A| + \|u_A\|_{1,p} + \|u^h_A\|_{1,p})^{p-1-\alpha} \|u_A - u^h_A\|^{\alpha}_{1,p} \|u_A - v_h\|_{1,p}, \end{aligned} \tag{5.45}$$

using (5.9).

The monotonicity property (5.10) also implies

$$\int_{(0,1)^3} \big(a(x, A + \nabla u_A) - a(x, A + \nabla u^h_A)\big) \cdot (u_A - u^h_A) \geq c\|u_A - u^h_A\|^{\beta}_{1,p} \tag{5.46}$$

if $\beta \leq p$.

Combining (5.45) with (5.46) and taking the infimum on $v_h \in V_h$, (5.44) follows. For $\beta > p$, the use of [38, Lemma 23.9] replaces (5.46) and gives the result

□

Lemmata 5.3 and 5.4 allow to prove the

Theorem 35 *Assume Hypotheses 1 and in addition $\alpha > 0$ in (5.9) and $p \geq \beta \geq 2$ in (5.10). Let $\mathcal{T}_h$ be a regular triangulation of $(0,1)^3$, $\mathcal{T}_H$ be a regular triangulation of Ω and V_h and V_H be linear finite element subspaces of $W^{1,p}_{\#}((0,1)^3, \mathbb{R}^3)$ and of $W^{1,p}_0(\Omega, \mathbb{R}^3)$ respectively associated to $\mathcal{T}_h$ and $\mathcal{T}_H$. Let $\bar{u} \in W^{1+2/p,p}(\Omega, \mathbb{R}^3)$ and u be the minimizer of (5.22). Theorem 34 and formula (5.41) provide a minimizer $u^h_H = u^{1,h}_H$ of (5.23).*
Then there exist positive constants C_1 and C_2 independent of h and H, such that

$$\|u - u^h_H\|_{1,p} \leq C_1 h^{\frac{2}{p} \frac{\alpha}{(p-1)(p-\alpha)}} + C_2 H^{\frac{2}{p} \frac{\beta-\alpha}{p(\beta-\alpha)-\alpha}}. \tag{5.47}$$

Proof

Theorems 28, 30 and 32, Remark 8 and property (5.41) ensure the existence and uniqueness of u and u^h_H, minimizers of (5.22) and (5.23).

We denote by W^h the energy density $W^{1,h}$ since there is no ambiguity. For $A \in \mathcal{M}_3(\mathbb{R})$, let us introduce the notation:

$$a_{hom}(A) = \frac{\partial W_{hom}}{\partial \xi}(A), \tag{5.48}$$

and

$$a_h(A) = \frac{\partial W^h}{\partial \xi}(A). \tag{5.49}$$

Functions a_{hom} and a_h may be shown to be well-defined respectively using the homogenization theory of elliptic operators and the implicit function theorem (see Section 5.4.2 for details).

The proof is divided in two steps, which aim at estimating $a_{hom}(A) - a_h(A)$ for $A \in \mathcal{M}_3(\mathbb{R})$ and $\int_\Omega \Big(a_{hom}(\nabla v) - a_h(\nabla v)\Big) \cdot \nabla w$ for $v \in W^{1,p}(\Omega, \mathbb{R}^3)$ and $w \in W_0^{1,p}(\Omega, \mathbb{R}^3)$ respectively. The concluding argument uses monotonicity.

Step 1

By definition,

$$W_{hom}(A) = \inf \left\{ \int_{(0,1)^3} W(x, \nabla v(x) + A) dx : v \in W_\#^{1,p}((0,1)^3, \mathbb{R}^3) \right\}, \tag{5.50}$$

$$W^h(A) = \inf \left\{ \int_{(0,1)^3} W(x, \nabla v(x) + A) dx : v \in V_h \right\}. \tag{5.51}$$

As W is strictly convex, both minima are uniquely attained at v_A and v_A^h. Inequality (5.9) then implies

$$|a_{hom}(A) - a_h(A)| \leq \left| \int_{(0,1)^3} a(x, A + \nabla v_A) - a(x, A + \nabla v_A^h) \right| \tag{5.52}$$

$$\begin{aligned} &\leq C \|\nabla v_A - \nabla v_A^h\|_{0,p}^\alpha (1 + \|A + \nabla v_A\|_{0,p} + \\ &\qquad \|A + \nabla v_A^h\|_{0,p})^{p-1-\alpha}. \end{aligned} \tag{5.53}$$

On one hand, as v_A is the minimizer of (5.50) and $\{x \mapsto 0\} \in W_\#^{1,p}((0,1)^3, \mathbb{R}^3)$, (5.1) implies

$$C(1 + |A|^p) \geq \int_{(0,1)^3} W(x, A) \geq \int_{(0,1)^3} W(x, A + \nabla v_A) \geq c\|A + \nabla v_A\|_{0,p}^p.$$

Thus

$$\frac{C}{c}(1 + |A|^p) \geq \|A + \nabla v_A\|_{0,p}^p. \tag{5.54}$$

The same inequality holds for v_A^h.

On the other hand, Lemma 5.4 applied to v_A and v_A^h with the regularity given by Lemma 5.3 provides the error estimate

$$\|v_A - v_A^h\|_{1,p} \leq C h^{s\lambda/p}, \tag{5.55}$$

for all $\lambda \in [0, 2[$ and for $s = \dfrac{\alpha + 1}{\beta}$, using the interpolation theory for P1-finite elements.

Combining inequalities (5.52), (5.54) and (5.55) gives

$$|a_{hom}(A) - a_h(A)| \leq C h^{s\alpha\lambda/p}(1 + |A|^{p-1-\alpha}). \tag{5.56}$$

Step 2

For all $v \in W^{1,p}(\Omega, \mathbb{R}^3)$ and $w \in W_0^{1,p}(\Omega, \mathbb{R}^3)$, (5.56) and the Hölder inequality imply

$$\begin{aligned} \left| \int_\Omega \Big(a_{hom}(\nabla v) - a_h(\nabla v)\Big) \cdot \nabla w \right| &\leq C h^{s\alpha\lambda/p} \int_\Omega (1 + |\nabla v(x)^{p-1-\alpha}|)|\nabla w| dx \\ &\leq C h^{s\alpha\lambda/p}(1 + \|\nabla v\|_{0,p}^{p-1})\|\nabla w\|_{0,p}. \end{aligned} \tag{5.57}$$

Let u_H be the unique solution in V_H of

$$\int_\Omega a_{hom}(\nabla u_H + \nabla \bar{u}) \cdot \nabla v_H = 0 \qquad \forall\, v_H \in V_H, \tag{5.58}$$

and recall that u_H^h is the minimizer of (5.23), which thus satisfies the Euler-Lagrange equation

$$\int_\Omega a_h(\nabla u_H^h + \nabla \bar{u}) \cdot \nabla v_H = 0 \qquad \forall\, v_H \in V_H. \tag{5.59}$$

Taking $v = u_H^h + \bar{u}$, inequality (5.57) reads

$$\begin{aligned}\left|\int_\Omega \Big(a_{hom}(\nabla(u_H^h + \bar{u})) - a_h(\nabla(u_H^h + \bar{u}))\Big) \cdot \nabla w\right| \\ \leq Ch^{s\alpha\lambda/p}(1 + \|\nabla(u_H^h + \bar{u})\|_{0,p}^{p-1})\|\nabla w\|_{0,p}.\end{aligned}$$

The same arguments as in the proof of (5.54) show

$$\begin{aligned}\left|\int_\Omega \Big(a_{hom}(\nabla(u_H^h + \bar{u})) - a_h(\nabla(u_H^h + \bar{u}))\Big) \cdot \nabla w\right| \\ \leq Ch^{s\alpha\lambda/p}(1 + \|\nabla \bar{u}\|_{0,p}^{p-1})\|\nabla w\|_{0,p}.\end{aligned} \tag{5.60}$$

As u_H solves (5.58), u_H^h solves (5.59) and $u_H - u_H^h \in V_H$ is an admissible test function for both problems,

$$\int_\Omega \Big(a_{hom}(\nabla(u_H + \bar{u})) - a_h(\nabla(u_H^h + \bar{u}))\Big) \cdot \nabla(u_H - u_H^h) = 0. \tag{5.61}$$

The monotonicity (5.10) of a_{hom} given by Theorem 30 implies

$$\left|\int_\Omega \Big(a_{hom}(\nabla(u_H + \bar{u})) - a_{hom}(\nabla(u_H^h + \bar{u}))\Big) \cdot \nabla(u_H - u_H^h)\right| \geq c\|\nabla u_H - \nabla u_H^h\|_{0,p}^\beta,$$

whereas inequalities (5.60) and (5.61) give

$$\begin{aligned}&\left|\int_\Omega \Big(a_{hom}(\nabla(u_H + \bar{u})) - a_{hom}(\nabla(u_H^h + \bar{u}))\Big) \cdot \nabla(u_H - u_H^h)\right| \\ &\leq \left|\int_\Omega \Big(a_{hom}(\nabla(u_H + \bar{u})) - a_h(\nabla(u_H^h + \bar{u}))\Big) \cdot \nabla(u_H - u_H^h)\right| \\ &\qquad + \left|\int_\Omega \Big(a_h(\nabla(u_H^h + \bar{u})) - a_{hom}(\nabla(u_H^h + \bar{u}))\Big) \cdot \nabla(u_H - u_H^h)\right| \\ &\leq Ch^{s\alpha\lambda/p}(1 + \|\nabla \bar{u}\|_{0,p}^{p-1})\|\nabla u_H - \nabla u_H^h\|_{0,p}.\end{aligned}$$

The Poincaré inequality shows that there exists a constant C_1 depending only on c, C, α, β, $|\Omega|$, $\bar{u}$ and p, such that,

$$\|u_H - u_H^h\|_{1,p} \leq C_1 h^{\alpha\lambda s/(p(\beta-1))}. \tag{5.62}$$

As a_{hom} satisfies (5.9) with $\gamma = \dfrac{\alpha}{\beta - \alpha}$ instead of α, a variant of Lemma 5.4 implies the existence of a constant C_2 such that $\|u - u_H\|_{1,p} \leq C_2 H^{\frac{\lambda}{p(\beta-\alpha)}}$. The latter inequality, combined with (5.62) for $\lambda = 2$, implies

$$\|u - u_H^h\|_{1,p} \leq C_1 h^{\frac{2(\alpha+1)}{p\beta}\frac{\alpha}{\beta-1}} + C_2 H^{\frac{2}{p(\beta-\alpha)}}.$$

□

This result is also true for scalar monotone equations as the Laplace equation, for which $p = 2$, $\alpha = 1$ and $\beta = 2$. In this case, $\dfrac{2(\alpha+1)}{p\beta}\dfrac{\alpha}{\beta-1} = 1$ and $\dfrac{2}{p(\beta-\alpha)} = 1$. This estimate is however not optimal and something special occurs in the linear case, as noticed by Abdulle in [1]: the optimal estimate is given by

$$\|u - u_H^h\|_{1,2} \leq C_1 h^2 + C_2 H.$$

We do not prove this estimate and refer the reader to [1] for the original proof. The same phenomenon occurs in Chapter 6 (estimation (6.21) for the linear case) for which the argument is detailed in Section 6.2.5. In the general setting of monotone operators for which the continuity assumption (5.9) is only assumed for $\alpha = 0$, the previous analysis is no longer valid. However the following more general result holds.

Theorem 36 *With the notation of Theorem 35, assume Hypotheses 1 with $\alpha \geq 0$ in (5.9) and $p \geq \beta \geq 2$ in (5.10). Then there exist positive constants C_1 and C_2 independent of h and H, such that*

$$\|u - u_H^h\|_{1,p} \leq C_1 h^{\frac{2}{p}\frac{1}{p(p-\alpha)}} + C_2 H^{\frac{2}{p}\frac{\beta-\alpha}{p(\beta-\alpha)-\alpha}}. \tag{5.63}$$

Proof

Let us follow the proof of Theorem 35 and focus on W instead of a. As the continuity property (5.9) implies

$$\begin{aligned} |W(y,\xi_1) - W(y,\xi_2)| &= \left| \int_0^1 a(y, \xi_1 + t(\xi_2 - \xi_1)) \cdot (\xi_2 - \xi_1) \mathrm{d}\, t \right| \\ &\leq C|\xi_2 - \xi_1|(1 + |\xi_1|^{p-1} + |\xi_2 - \xi_1|^{p-1}) \\ &\leq C|\xi_2 - \xi_1|(1 + |\xi_1|^{p-1} + |\xi_2|^{p-1}), \end{aligned}$$

(5.57) can be replaced by

$$\left| \int_\Omega W_{hom}(\nabla v) - W^h(\nabla v) \right| \leq C h^{s\lambda/p}(1 + \|\nabla v\|_{0,p}^{p-1}), \tag{5.64}$$

(5.60) by

$$\left| \int_\Omega W_{hom}(\nabla(u_H^h + \bar{u})) - W^h(\nabla(u_H^h + \bar{u})) \right| \leq C h^{s\lambda/p}(1 + \|\nabla \bar{u}\|_{0,p}^{p-1}), \tag{5.65}$$

and (5.61) by

$$\left| \int_\Omega W_{hom}(\nabla(u_H + \bar{u})) - W^h(\nabla(u_H^h + \bar{u})) \right| \leq C h^{s\lambda/p}(1 + \|\nabla \bar{u}\|_{0,p}^{p-1}). \tag{5.66}$$

Inequality (5.66) is a direct consequence of the control of $W_{hom} - W^h$ close to the minima on V_H of W_{hom} and W^h:

$$\begin{aligned} \left| \inf_v \left\{ \int_\Omega W_{hom}(\nabla(v + \bar{u})) \right\} - \int_\Omega W^h(\nabla(u_H + \bar{u})) \right| &\leq C h^{s\lambda/p}(1 + \|\nabla \bar{u}\|_{0,p}^{p-1}), \\ \left| \inf_v \left\{ \int_\Omega W^h(\nabla(v + \bar{u})) \right\} - \int_\Omega W_{hom}(\nabla(u_H^h + \bar{u})) \right| &\leq C h^{s\lambda/p}(1 + \|\nabla \bar{u}\|_{0,p}^{p-1}). \end{aligned}$$

The following consequence of the monotonicity (5.10) of a_{hom} (Theorem 1 in [169]) allows to conclude: since u_H is a minimizer of $\displaystyle\int_\Omega W_{hom}(\nabla(\bar{u} + \cdot))$ on the convex set V_H and $u_H^h \in V_H$,

$$\left| \int_\Omega W_{hom}(\nabla(u_H + \bar{u})) - W_{hom}(\nabla(u_H^h + \bar{u})) \right| \geq c\|\nabla u_H - \nabla u_H^h\|_{0,p}^{\beta}. \tag{5.67}$$

Formula (5.63) is then obtained by combining (5.64), (5.65), (5.66) and (5.67).

□

Depending on how $\dfrac{1}{\beta}$ compares with $\dfrac{\alpha}{\beta - 1}$, either formula (5.47) or formula (5.63) gives a better estimate. However Theorem 36 is more general. The worst case is $\alpha = 0$ and $\beta = p$, and then $h \simeq H^p$ yields the optimal error in (5.63).

When dealing with more particular energies such as energies with p-structure (namely polynomial of degree p in ξ, with $\alpha = 1$ and $\beta = 2$), Ebmeyer and Liu have proved in [70], using the interpolation theory in Nikolskij spaces, the following optimal error estimate

Lemma 5.5. *[70] Assume Hypotheses 1 and properties (5.9) and (5.10) with $a(x,\cdot)$ polynomial of degree $p-1$ for all $x \in (0,1)^3$, $\alpha = 1$ and $\beta = 2$. Let $A \in \mathcal{M}_3(\mathbb{R})$, $u_A \in W^{1,p}_{\#}((0,1)^3,\mathbb{R}^3)$ be the solution of (5.42) and u_A^h be the solution of (5.43) in a P1-finite element subspace of $W^{1,p}_{\#}((0,1)^3,\mathbb{R}^3)$. Then there exists a constant $C > 0$ independent of h such that*

$$\|u_A - u_A^h\|_{1,p} \leq Ch.$$

Using Lemma 5.5, one can improve the error estimate of Theorem 35 (for the p-structure energy, $\beta = 2$, $\alpha = 1$), obtaining

$$\|u - u_H^h\|_{1,p} \leq C_1 h^{\frac{1}{p-1}} + C_2 H.$$

For the general quasiconvex case, we are not able to have a similar result. We may however suppose that the optimal meshsize h for the cell problem given the meshsize H for the homogenized problem could depend on p, the order of the growth condition (5.1).

5.4 Numerical method

In this section, a direct approach to numerical homogenization in the framework of nonlinear elasticity is introduced. The numerical resolution of (5.3)-(5.22) is directly tackled by solving (5.23).

The method is presented in the convenient case of zero body force and Dirichlet boundary conditions. The numerical tests of Section 5.6 are also performed in this setting. However the method adapts straightforwardly to more general body forces and boundary conditions provided classical adaptations of the energy density and of the variational spaces.

5.4.1 Presentation

The numerical analysis performed above makes use of a ball $B(u,\bar{r})$ where the minimizer u of (5.22) is isolated. The minimizer u, and consequently the ball $B(u,\bar{r})$, being unknown in practice, the numerical approach consists in considering, instead of (5.23), the problem

$$\inf \left\{ J^{N,h}(v+\bar{u})dx \,|\, v \in V_{\Omega,H} \right\}, \tag{5.68}$$

for N and h fixed, using the notation (5.20)-(5.21).

This minimum value is attained since $J^{N,h}$ is continuous on $W^{1,p}(\Omega)$, $J^{N,h}(v) \to \infty$ when $\|v\|_{1,p} \to \infty$ and $V_{\Omega,H}$ is a finite dimensional space.

In the remainder of Section 5.4.1 the energy density $W^{N,h}$ defined by (5.20) is supposed to be twice continuously differentiable. In this case, if u is a minimizer of (5.68) then u satisfies the Euler-Lagrange equation in the following weak form: for all $v \in V_{\Omega,H}$,

$$\int_\Omega \frac{\partial W^{N,h}}{\partial \xi}(\nabla(u+\bar{u})) \cdot \nabla v = 0. \tag{5.69}$$

The nonlinear equation (5.69) is solved by an iterative Newton-Raphson method. Knowing u^n at step n, the associated linearized problem at step $n+1$ reads: find u^{n+1} such that for all $v \in V_{\Omega,H}$,

$$\int_\Omega \Big(\frac{\partial W^{N,h}}{\partial \xi}(\nabla(\bar{u}+u^n)) + \frac{\partial^2 W^{N,h}}{\partial \xi^2}(\nabla(\bar{u}+u^n)) \cdot (\nabla u^{n+1} - \nabla u^n) \Big) \cdot \nabla v = 0, \tag{5.70}$$

and iterate until convergence.

To perform the Newton-Raphson method, an explicit expression of the stress tensor $\dfrac{\partial W^{N,h}}{\partial \xi}$ and the stiffness matrix $\dfrac{\partial^2 W^{N,h}}{\partial \xi^2}$ is needed. This is the matter of the following section.

5.4.2 Computation of the stress tensor and the stiffness matrix

This section aims at introducing two quantities (Theorem 38) that can actually be computed and that may be identified as the stress tensor and the stiffness matrix of the homogenized constitutive relation in some simple cases (Theorem 39). The validity of this identification is discussed in the general case at the end of this section.

Let

$$I : \mathcal{M}_3(\mathbb{R}) \times W^{1,p}_{\#}((0,N)^3) \to \mathbb{R}$$
$$(\xi, \phi) \mapsto \int_{(0,N)^3} W(y, \xi + \nabla\phi(y))dy,$$

and $\{\psi_i\}_i$ be a basis of $V_{N,h}$.

The following hypotheses are made so that I be regular.

Hypotheses 2 *The function $W(y,\cdot)$ is three times continuously differentiable on $\mathcal{M}_3(\mathbb{R})$ and satisfies (5.1) and the following growth properties*

$$\max\left\{|\frac{\partial W}{\partial \xi}(x,\xi)|, |\frac{\partial^2 W}{\partial \xi^2}(x,\xi)|, |\frac{\partial^3 W}{\partial \xi^3}(x,\xi)|\right\} \leq C(1+|\xi|^p) \tag{5.71}$$

In addition $V_{N,h} \subset W^{1,\infty}((0,N)^3)$.

Let us first study the differentiability of I.

Lemma 5.6. *If W and $V_{N,h}$ satisfy Hypotheses 2, then $I \in C^3(\mathcal{M}_3(\mathbb{R}) \times V_{N,h}, \mathbb{R})$.*

Proof

This proof is classical (see [125] e.g.) and is only sketched for the first derivative.

As $\nabla\psi \in L^\infty((0,N)^3) = L^1((0,N)^3)'$ and $\dfrac{\partial W}{\partial \xi}(y,\cdot)$ sends $L^p((0,N)^3)$ on$L^1((0,N)^3)$ thanks to (5.71), Lemma 5.2 implies that for all $\psi \in V_{N,h}$ and $\zeta \in \mathcal{M}_3(\mathbb{R})$,

$$\chi \mapsto \int_{(0,N)^3} \frac{\partial W}{\partial \xi}(y, \zeta + \nabla\chi) \cdot \nabla\psi$$

is continuous on $V_{N,h}$.

Next, for all $\psi \in V_{N,h}$ and $\zeta \in \mathcal{M}_3(\mathbb{R})$, $\sigma_{\psi,\zeta}$ is defined by

$$\sigma_{\psi,\zeta} : (t,y) \mapsto \frac{1}{t}\Big(W(y, \zeta + \nabla\chi(y) + t\nabla\psi(y)) - W(y, \zeta + \nabla\chi(y))\Big).$$

The Fréchet derivative of W at $\zeta + \nabla\chi(y)$ in the direction $\nabla\psi(y)$ is given by

$$\lim_{t\to 0} \int_{(0,N)^3} \sigma_{\psi,\zeta}(t,y)dy.$$

Pointwise, $\lim\limits_{t\to 0} \sigma_{\psi,\zeta}(y,t) = \dfrac{\partial W}{\partial \xi}(y, \zeta + \nabla\chi(y)) \cdot \nabla\psi(y)$.
As $W(y,\cdot)$ is C^1, for all $t \in (0,1)$ there exists $\theta \in (0,1)$ such that

$$\sigma_{\psi,\zeta}(y,t) = \frac{\partial W}{\partial \xi}(y, \zeta + \nabla\chi(y) + \theta\nabla\psi(y)) \cdot \nabla\psi(y).$$

Using (5.71), $\sigma_{\psi,\zeta}(y,t)$ is uniformly dominated in t by the integrable function

$$(y,t) \mapsto C(1 + (|\nabla\chi(y)| + |\nabla\psi(y)| + |\zeta|)^p)\|\psi\|_{1,\infty}.$$

The Lebesgue dominated convergence theorem shows

$$\frac{\partial I}{\partial \phi}(\zeta,\chi)\cdot\psi = \int_{(0,N)^3} \frac{\partial W}{\partial \xi}(y,\zeta+\nabla\chi)\cdot\nabla\psi. \tag{5.72}$$

Similarly,

$$\frac{\partial I}{\partial \xi}(\zeta,\chi) = \int_{(0,N)^3} \frac{\partial W}{\partial \xi}(y,\zeta+\nabla\chi)\cdot Id, \tag{5.73}$$

where Id is the fourth order identity tensor. As the right hand sides are continuous in (5.72) and (5.73), I is C^1 on $\mathcal{M}_3(\mathbb{R})\times V_{N,h}$.

Repeating the same arguments, I is proved to be three times continuously differentiable.

□

Definition 19 *For all $A\in\mathcal{M}_3(\mathbb{R})$, $\phi\in V_{N,h}$ is said to be a local minimizer of $I(A,\cdot)$ on $V_{N,h}$ if there exists $r>0$ such that for all $\psi\in B(\phi,r)\cap V_{N,h}$, $I(A,\phi)\le I(A,\psi)$.*
A local minimizer ϕ is global if for all $\psi\in V_{N,h}$, $I(A,\phi)\le I(A,\psi)$.
A minimizer ϕ on $V_{N,h}$ is isolated if there exists $\rho>0$ and if for all $\psi\in B(\phi,\rho)\cap V_{N,h}$ such that $\psi\neq\phi$, $I(A,\phi)<I(A,\psi)$.

Hypotheses 3 *Given $A\in\mathcal{M}_3(\mathbb{R})$, there exists a minimizer ϕ of $I(A,\cdot)$ on $V_{N,h}$, satisfying*

- *ϕ is an isolated global minimizer on $V_{N,h}$*
- *the Hessian matrix $\left(\int_{(0,N)^3}\nabla\psi_i(y)^T\cdot\frac{\partial^2 W}{\partial\xi^2}(y,A+\nabla\phi(y))\cdot\nabla\psi_j(y)dy\right)_{i,j}$ is positive definite.*

Theorem 37 *Let W and $V_{N,h}$ satisfy Hypotheses 2 and $(A,\phi)\in\mathcal{M}_3(\mathbb{R})\times V_{N,h}$ satisfy Hypotheses 3, then there exist two open balls $B_A\subset\mathcal{M}_3(\mathbb{R})$ and $B_\phi\subset V_{N,h}$, and there exists a function $g_\phi\in C^2(B_A,B_\phi)$, such that for all $\xi\in B_A$, $g_\phi(\xi)$ is an isolated local minimizer of $I(\xi,\cdot)$ on $V_{N,h}$. In addition, for $\{e_i\}_{1\le i\le 9}$ a basis of $\mathcal{M}_3(\mathbb{R})$,*

$$\frac{\partial\nabla g_\phi(\xi)}{\partial\xi}_{|\xi=A}\cdot e_i = \nabla v_i,$$

where v_i is the solution in $V_{N,h}$ of

$$\int_{(0,N)^3}\left(\frac{\partial^2 W(y,\xi)}{\partial\xi^2}_{|\xi=A+\nabla\phi(y)}\cdot(e_i+\nabla v_i)\right)\cdot\nabla\psi = 0 \quad \forall\ \psi\in V_{N,h}; \tag{5.74}$$

and

$$\frac{\partial^2\nabla g_\phi(\xi)}{\partial\xi^2}_{|\xi=A} : e_j\otimes e_i = \nabla w_{ij},$$

where w_{ij} is the solution in $V_{N,h}$ of

$$\begin{aligned}\int_{(0,N)^3}\Bigg(&\frac{\partial^2 W(y,\xi)}{\partial\xi^2}_{|\xi=A+\nabla\phi(y)}\cdot\nabla w_{ij}+\\ &\frac{\partial^3 W(y,\xi)}{\partial\xi^3}_{|\xi=A+\nabla\phi(y)}\cdot(e_i+\nabla v_i)\cdot(e_j+\nabla v_j)\Bigg)\cdot\nabla\psi = 0 \quad \forall\ \psi\in V_{N,h}.\end{aligned} \tag{5.75}$$

Proof

In this proof, ν and $\{\Psi_i\}_{1\le i\le\nu}$ denote the dimension and a basis of $V_{N,h}$. Theorem 37 is a direct application of the implicit function theorem to

$$\begin{aligned}\pi : \mathcal{M}_3(\mathbb{R})\times V_{N,h} &\to \mathbb{R}^\nu\\ (\zeta,\chi) &\mapsto \pi(\zeta,\chi),\end{aligned}$$

where

$$\pi(\zeta,\chi) = \begin{pmatrix} \displaystyle\int_{(0,N)^3} \frac{\partial W}{\partial \xi}(y,\zeta+\nabla\chi(y))\cdot\nabla\psi_1(y)dy \\ \vdots \\ \displaystyle\int_{(0,N)^3} \frac{\partial W}{\partial \xi}(y,\zeta+\nabla\chi(y))\cdot\nabla\psi_\nu(y)dy \end{pmatrix}.$$

As I is C^3, $\pi \in C^2(\mathcal{M}_3(\mathbb{R}) \times V_{N,h}, \mathbb{R}^\nu)$ and by definition of ϕ, $\pi(A,\phi) = 0$. As $\dfrac{\partial \pi}{\partial \chi}(A,\phi)$ is invertible since it is the Hessian matrix of Hypotheses 3, the implicit function theorem shows that there exist two open balls $\tilde{B}_A \ni A$ and $B_\phi \ni \phi$ and a function $g_\phi : \tilde{B}_A \to B_\phi$ such that for all $\zeta \in \tilde{B}_A$, $g_\phi(\zeta)$ is the unique solution in B_ϕ of $\pi(\zeta,\cdot) = 0$. In addition, g_ϕ is twice continuously differentiable.

As $\dfrac{\partial \pi}{\partial \chi}(\zeta, g_\phi(\zeta))$ is continuous and $\dfrac{\partial \pi}{\partial \chi}(A,\phi)$ is positive definite, there exists a non empty open ball $B_A \subset \tilde{B}_A$ such that for all $\zeta \in B_A$, $\dfrac{\partial \pi}{\partial \chi}(\zeta, g_\phi(\zeta))$ is also positive definite, which implies that $g_\phi(\zeta)$ is an isolated local minimizer of $I(\zeta,\cdot)$ on $V_{N,h}$.

Equations (5.74) and (5.75) are then obtained by differentiating once and twice respectively $\zeta \mapsto \pi(\zeta, g_\phi(\zeta))$ at $\zeta = A$.

□

Theorem 38 *Assume that W satisfies Hypotheses 2 and that (A,ϕ) satisfies Hypotheses 3, then, with the notation of Theorem 37, the function $\xi \mapsto I(\xi, g_\phi(\xi))$ is twice continuously differentiable on B_A at $\xi = A$ and its derivatives are given by*

$$\frac{\mathrm{d}}{\mathrm{d}\,\xi} I(\xi, g_\phi(\xi))\Big|_{\xi=A} = \int_{(0,N)^3} \frac{\partial W(y,\xi)}{\partial \xi}\Big|_{\xi=A+\nabla\phi(y)} dy, \tag{5.76}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\,\xi^2} I(\xi, g_\phi(\xi))\Big|_{\xi=A} = \tag{5.77}$$

$$\int_{(0,N)^3} \left(Id + \frac{\partial \nabla g_\phi(\xi)}{\partial \xi}\Big|_{\xi=A} \right)^T \cdot \frac{\partial^2 W(y,\xi)}{\partial \xi^2}\Big|_{\xi=A+\nabla\phi(y)} \cdot \left(Id + \frac{\partial \nabla g_\phi(\xi)}{\partial \xi}\Big|_{\xi=A} \right) dy,$$

where Id is the fourth oder identity tensor.

Proof

The function $\xi \mapsto I(\xi, g_\phi(\xi))$ is twice continuously differentiable on B_A at $\xi = A$ as the composition of two differentiable functions. A direct calculus shows

$$\frac{\mathrm{d}}{\mathrm{d}\,\xi} I(\xi, g_\phi(\xi))_{|\xi=A} = \int_{(0,N)^3} \frac{\partial W(y,\xi)}{\partial \xi}\Big|_{\xi=A+\nabla\phi(y)} \cdot \left(Id + \frac{\partial \nabla g_\phi(\xi)}{\partial \xi}\Big|_{\xi=A}(y) \right) dy.$$

As ϕ satisfies $\pi(A,\phi) = 0$ and $\dfrac{\partial \nabla g_\phi(\xi)}{\partial \xi} \in V_{N,h}$,

$$\int_{(0,N)^3} \frac{\partial W(y,\xi)}{\partial \xi}\Big|_{\xi=A+\nabla\phi(y)} \cdot \frac{\partial \nabla g_\phi(\xi)}{\partial \xi}\Big|_{\xi=A}(y) dy = 0,$$

which proves (5.76).

An analogous calculus leads to formula (5.77).

□

In general, whereas $W^{N,h}(\xi)$ is only defined by (5.20), $I(\xi, g_\phi(\xi))$ is the only quantity that can be computed. If the global minimizer defining $W^{N,h}(\xi)$ is unique and depends continuously on ξ then $W^{N,h}(\xi) = I(\xi, g_\phi(\xi))$. This is indeed the case for strictly convex energy densities as stated in

Theorem 39 *In addition to Hypotheses 2 and 3, assume, with the notation of Theorem 37, that $W(y, \cdot)$ is strictly convex for almost every y. Then, forall $A \in \mathcal{M}_3(\mathbb{R})$, the minimizer ϕ of $I(A, \cdot)$ on $V_{N,h}$ is unique, $W^{N,h}$ is twice differentiable, $W^{N,h}(\xi) = I(\xi, g_\phi(\xi))$ and its derivatives are given by the right hand sides of (5.76) and (5.77) respectively.*

Proof

Thanks to strict convexity and Hypotheses 3, the minimizer ϕ is unique and the Hessian is positive definite for all couples $(A, \phi(A))$. The function g_ϕ does not depend on ϕ and is denoted by g. It is defined on $\mathcal{M}_3(\mathbb{R})$ and for all $\xi \in \mathcal{M}_3(\mathbb{R})$, $g(\xi)$ is the unique global minimizer of $I(\xi, \cdot)$ on $V_{N,h}$, which implies $W^{N,h}(\xi) = I(\xi, g(\xi))$.

□

When dealing with nonconvex energy densities, the simple analysis performed above does not apply. We however use the derivatives of $I(\xi, g_\phi(\xi))$ in practice in order to compute the stress tensor and the stiffness matrix for the homogenized constitutive law. The assumption that $I(\xi, g(\xi))$ is a global minimizer is strong since its validity cannot be infered a posteriori. If the Newton algorithm converges, we have found a critical point of a numerical energy, that is expected to be close to the homogenized energy.

Following the work of Geymonat, Müller and Triantafyllidis in [94], this section can be rewritten in a variational setting. More precise results can be obtained assuming the exclusion of discontinuous bifurcations in the minimization of the cell-problems and making other assumptions hard to verify in practice. Our presentation is restricted to what the algorithm can actually perform and is therefore limited to local minimizers in general. If the computed minimizer happens to be global, then the results of [94] (Section 5.2) apply and justify the numerical approach.

5.4.3 Implementation of the algorithm in a nonlinear elasticity software

The direct approach to numerical homogenization presented here can be used in a nonlinear elasticity solver by using the right hand sides of (5.76) and (5.77) as derivatives for the stress tensor and the stiffness matrix in (5.70). This method has the important advantage not to modify the structure of the existing solver.

This method has been implemented within a classical finite element code (Modulef, INRIA, see [177] and [137]). The call of an analytical formula giving the stress tensor and the stiffness matrix at each Gauss point has been replaced by a subroutine solving itself a nonlinear elasticity cell problem (5.20) and providing the main program with (5.76) and (5.77). The global structure of the code remains therefore unchanged. Any sophisticated technique already used in the code directly adapts without modification: mixed finite elements, augmented Lagrangians, arc-length continuation and parallelization (see [177] and [125]). Numerical tests are reported on in the last section.

The major part of the computational cost comes from the computation of the homogenized constitutive law, as opposed to a classical nonlinear elasticity problem for which this is a simple evaluation of an analytical formula. For the computation of this homogenized constitutive relation itself, the main cost comes from solving the cell-problem (5.20). The resolution of linear systems is performed by direct inversions, such as Cholesky factorization, because of the large condition number and the lack of efficient preconditionners for nonlinear elasticity problems. Once the cell-problem is solved, the computation of (5.76) and (5.77) is obtained by solving a linear system with nine different right hand sides. This linear system is indeed the same as in the last iteration of the Newton algorithm solving the cell-problem, it is therefore already factorized.

On a PC with 2GB of memory, a three dimensional elasticity problem with 40 000 degrees of freedom can be solved without domain decomposition methods, which means for Q2-finite elements a mesh with 12 nodes per dimension in the cell-problem. In that sense, the cell-problems are a limiting factor. On the other hand the global CPU time does not vary too much with respect to the number of degrees of freedom of the macroscopic problem provided efficient domain decomposition methods and parallel computing for the macroscopic problem are used.

A simple way to reduce the cost of computation of the homogenized constitutive law is not to recompute it at each step if the strain gradient has not changed too much and to use the solution at the previous step or at a neighboring Gauss point as an initial guess in the cell-problem. Going further in this direction, another possibility would be to precompute and tabulate the homogenized constitutive relation for a wide number of strain gradients, in the spirit of the numerical practice for combustion problems. The latter issue has not been addressed in this work. However it is worth noticing that the convergence property of the Newton algorithm is very sensitive to the approximation of the stiffness matrix, which can be an obstacle for this kind of approach.

5.5 Alternative method: multiscale finite elements (MsFEM)

It is interesting to relate the direct method with more elaborate approaches, such as the multiscale finite element method.

5.5.1 Description of the method

We refer to the work of Hou, Efendiev and coauthors ([105], [72], [155]) for the detailed description of the MsFEM in the linear and nonlinear settings.

This method has primarily been designed in [105] to solve efficiently the linear elliptic equation arising in porous media flows in heterogeneous materials. It has been then extended from the linear case to the monotone case in [72] and [155]. This method is proved to converge in the periodic setting. It has also been used in a more general context and turns out to be quite efficient in the cases reported on by the authors.

Basically, the MsFEM is a Galerkin method for which the solution is searched in a specific space associated to the elliptic operator. Its convergence is then proved thanks to the homogenization theory. Conversely, in problem (5.23), a classical Galerkin space (classical finite elements) is used but the original elliptic operator is approximated by its homogenized operator. Although the methods seem to be different at first sight, they turn out to be identical under some hypotheses.

In the setting of periodic homogenization, Hou and coworkers exploit the periodicity of the operator to compute the multiscale finite element manifold on one periodic cell instead of one element (triangle or tetrahedra), which drastically reduces the size of the problems. In this case, the MsFEM exactly consists in solving (5.23) as shown in the next paragraph and coincides with the direct approach. This observation allows to apply Theorem 35, which thus provides with some insight in the choice of the meshsize of the fine triangulation used in the MsMFEM. In their analysis, Hou and coauthors have focused on the resonance error linked to the boundary condition used to build the multiscale finite elements. They have not addressed the question of the influence on the global error of the approximation by a Galerkin method of the multiscale map itself. This has been answered by Allaire in [8] in the linear case. The result may indeed be different in the nonlinear setting as shown by Theorem 35.

5.5.2 Comparison of the two methods in the periodic setting

The analysis of the MsFEM requires a corrector result. In order to compare the two methods a monotone operator is considered in the setting of periodic homogenization of Section 5.2.4. The notation of Theorems 30 and 31 is used.

We suppose that the macroscopic mesh perfectly fits to the underlying periodic structure so that there is no mismatch (triangular mesh with equilateral triangles and periodic structure with equilateral triangle in 2D e.g.). This technical requirement allows to use a single periodic cell in order to compute the multiscale map (i.d. basis functions in the linear case), as pointed out in [72]. In this case, there is no resonance error due to the boundary conditions.

Let us give some details on the computation of the multiscale finite element map in the P1-Lagrange case. The triangulation $\mathcal{T}_H$ and the finite element space V_H introduced in Theorem 35 are used. Given an element $T \in \mathcal{T}_H$ and a function $u \in V_H$, the associated multiscale function w is defined on T by

$$w(x) = u(x) + \epsilon \nabla v_{u_T}(\frac{x}{\epsilon}),$$

where v_{u_T} is the solution in $W^{1,p}_{\#}((0,1)^3)$ of

$$-\mathrm{div}\,\big(a(y, \nabla u_T + \nabla v_{u_T}(y))\big) = 0 \quad \text{in } (0,1)^3, \tag{5.78}$$

and ∇u_T is the gradient of u, which is constant on T since $u \in V_H$. As $u \in V_H$ implies $M_\epsilon \nabla u = \nabla u$, (5.78) is exactly (5.11) and the multiscale function w is thus exactly the sum of a classical part, the finite element function u, and of its associated corrector given by Theorem 31. In order to compare the formulation (5.7), completed by the corrector, to the MsFEM, we just have to compare the classical parts of the solutions.

Consider the multiscale finite element solution w_{MsFEM} of problem (5.6) in the sense of a Petrov-Galerkin formulation, see [72]. By definition of the MsFEM, for all P1-Lagrange shape functions $\{u_j\}_j$ on $\mathcal{T}_H$, we have with obvious notation:

$$\begin{aligned}
&\int_\Omega a(\frac{x}{\epsilon}, \nabla w_{MsFEM}(x)) \cdot \nabla u_j(x) \\
&= \sum_T \int_T a\big(\frac{x}{\epsilon}, \nabla u_{MsFEM}(x) + \epsilon \nabla v_{u_T}(\frac{x}{\epsilon})\big) \cdot (\nabla u_j)_T \\
&= \sum_T \int_T a_{hom}\Big((\nabla u_{MsFEM})_T\Big) \cdot (\nabla u_j)_T \\
&= \int_\Omega a_{hom}\Big(\nabla u_{MsFEM}(x)\Big) \cdot \nabla u_j(x).
\end{aligned}$$

Therefore, w_{MsFEM} is the MsFEM solution to (5.6) if and only if u_{MsFEM} is the (classical) finite element solution u_H to (5.7).

In the previous analysis, (5.78) is solved exactly. Suppose from now on that (5.78) is solved on V_h as in Theorem 35. Denoting by w^h_{MsFEM} the approximate multiscale solution, the same calculation as above shows

$$\int_\Omega a(\frac{x}{\epsilon}, \nabla w^h_{MsFEM}(x)) \cdot \nabla u_j(x) = \int_\Omega a_h\Big(\nabla u^h_{MsFEM}(x)\Big) \cdot \nabla u_j(x),$$

where a_h is given by (5.49). Thus, $u^h_{MsFEM} = u^h_H$, where u^h_H is the solution of (5.23) in the particular case of Theorem 35. Therefore the error analysis of Section 5.3.3 can be applied to the MsFEM, providing an a priori indication for the order of magnitude of the meshsize h of the fine triangulation in function of the meshsize H of the rough triangulation and of the continuity and ity properties (5.9) and (5.10) of the operator a.

5.6 Numerical tests

This section is dedicated to numerical tests in nonlinear elasticity. The numerical tests of the first paragraph confirm the simple analysis presented above for convex energy densities. In the subsequent paragraphs, some issues of theoretical and practical interest are investigated with the use of numerical experiments:

- buckling in the cell-problem;
- instability of the homogenized energy;
- application of the method to a wider class of energy densities.

5.6.1 The convex case

Problem (5.22) is considered with the following energy density:

$$\begin{aligned} W : (0,1)^3 \times \mathcal{M}_3(\mathbb{R}) &\to \mathbb{R} \\ (y,\xi) &\mapsto \gamma_1(y)|\xi|^4 + \gamma_2(y)|\xi|^2, \end{aligned}$$

where $\gamma_1, \gamma_2 \geq 1$ is Lipschitz and 1-periodic on $\mathbb{R}^3$. Theorem 35 and Lemma 5.4 apply with $p = 4$, $\alpha = 1$, $\beta = 2$ so that the error estimate (5.47) reads

$$\|u - u_H^h\|_{1,p} \leq C_1 h^{1/2} + C_2 H^{1/2}. \tag{5.79}$$

In this case, the algorithm indeed converges, the result does not depend on the number of periodic cells N considered. The numerical tests performed so far seem to show that the rate of convergence (5.79) is not sharp.

5.6.2 Buckling of the cell-problem in the standard case

In the convex case, the infimum in (5.5) is attained on one periodic cell, for $N = 1$. In [145], S. Müller gives an example in two dimensions for which the infimum in (5.5) is strictly smaller than the infimum on one single periodic cell. This example relies on the mechanical concept of buckling of a rigid bar in compression: there is a bifurcation and the equilibrium state with the lowest energy breaks the symmetry of the problem.

In three dimensions, the corresponding energy density reads

$$\begin{aligned} \tilde{W} : \mathcal{M}_3(\mathbb{R}) &\to \mathbb{R} \\ \xi &\mapsto |\xi|^4 + h(\det(\xi)), \end{aligned}$$

where h is given by

$$\begin{aligned} h : \mathbb{R} &\to \mathbb{R} \\ r &\mapsto \frac{12(1+a)^2}{r+a} - 12(1+a) - 9 && \text{if } r > 0, \\ & \quad \frac{12(1+a)^2}{r+a} - 12(1+a) - 9 - \frac{12(1+a)^2}{a^2} r && \text{if } r \leq 0, \end{aligned}$$

with $a \in]0, 1/2[$.

In the numerical tests, $a = 0.25$ and the following energy density has been used in the periodic cell $(0,1)^3$: $W(y,\xi) = C(y)\tilde{W}(\xi)$, where $C(y) = 0.01$ if $y \in (0,1/2) \times (0,1)^2$ and $C(y) = 10$ if $y \in (1/2,1) \times (0,1)^2$. This models a layered material, whose energy density satisfies the hypotheses of Theorem 28.

The cell-problem (5.20) has been solved for

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}$$

and different numbers of periodic cells. The results are collected in Table 5.2. The energies of the solutions and the associated stress tensors are displayed versus the periods of the solutions. Even if the minimizers are not approximated accurately enough, the energies reported on are meaningful upper bounds. As the periods of the solutions increase, the energies of the solutions decrease

and the material relaxes its stress. Therefore, several periodic cells have to be taken into account to reach the limit in (5.5). This is not an easy task since tracking bifurcations is quite hard in practice with a Newton method. In addition, even in the cases presented above, for which the form of the bifurcation is intuitive, the solution is very sensitive to the initial guess. This makes the automation of the procedure quite tricky. More precisely, the numerical approximations have been obtained using $Q2$ isoparametric finite elements, and with the initial guesses and meshes reported on in Table 5.3.

Period	1[a]	3	5	9	13	17	23
Energy	5.819	5.388	4.319	3.736	3.587	3.530	3.495
Ratio[b]	0	-7.4%	-25.8%	-35.8%	-38.3%	-39.3%	-39.9%
Stress tensor[c]	-0.0623 -9.25 -36.8	-0.0461 -4.30 -16.1	-0.0541 -1.78 -6.41	-0.0621 -0.693 -2.45	-0.0620 -0.426 -1.50	-0.0597 -0.319 -1.12	-0.0553 -0.246 -0.867

Table 5.2. Numerical tests on S. Müller's example

[a] reference
[b] difference with the reference energy
[c] this tensor turns out to be diagonal

Period	2	13	23
ddl/periodic cell	192	192	192
Initial guess [a]	$d_x = \sin(\pi \frac{z}{2})$	$d_x = 18 \sin(\pi \frac{z}{18})^2$	$d_x = 23.16 \sin(\pi \frac{z}{23})^2$ [b]
Number of iterations of the Newton Alg.	8	7	13

Table 5.3. Number of periodic cells, initial guesses and degrees of freedom

[a] displacement in the direction x in function of z, $d_y = d_z = 0$
[b] this initial guess is extremely sensitive

It is quite clear that the number of degrees of freedom per periodic cell (192) is not big enough to claim the convergence of the approximations. To check the influence of this parameter on the result, we present in Table 5.4 the approximations of the stress tensor for one and two periodic cells and 192, 1536, and 5184 degrees of freedom per periodic cell. Amazingly, for one periodic cell, the finite element approximation seems very accurate. For two periodic cells, as far as one can say, the coarse mesh captures relatively well the order of magnitude.

Period	1			2		
dof	192	1536	5184	192	1536	5184
Stress tensor	-0.0623 -9.25 -36.8	-0.0623 -9.25 -36.8	-0.0623 -9.25 -36.8	-0.0606 -8.95 -35.6	-0.0549 -7.86 -30.8	-0.0544 -7.75 -30.4
Number of iterations of the Newton Alg.	3	3	3	8	5	6

Table 5.4. Numerical tests on S. Müller's example

We have also checked on an example the influence of the buckling of the cell-problem on the solution of the macroscopic problem itself. it is simple enough to allow us to find automatically the buckling in the cell-problems and complex enough not to have a trivial solution. The result shows that the minimizers of the numerical homogenized energies are also very different, even

when the test is quite simple. The resolution of the macroscopic problem does not simplify or reduce the influence of buckling of the cell-problem. This test has been performed for 1-periodic (no buckling) and 3-periodic solutions (buckling) of the cell-problems, with 192 dof per periodic cell. The norm of the difference between the two macroscopic solutions (one finite element : 27 degrees of freedom) is reported on in Table 5.5. Qualitatively both macroscopic solutions respect the anisotropy of the heterogenous material in the $(0x)$ direction. However, more the solutions of the cell-problems get relaxed, smaller the macroscopic deformation is.

$\|u_H^{1,h}\|_{1,4}$	$\|u_H^{3,h} - u_H^{1,h}\|_{1,4}/\|u_H^{1,h}\|_{1,4}$
0.2620	14.12%

Table 5.5. Influence of N on $u_H^{N,h}$

The way to choose the number N of cells for computing the homogenized properties of a nonconvex energy is not clear. Either there is no such phenomenon as buckling and one periodic cell is enough, or several cells have to be considered. In the latter case, the numerical practice is complex since local minimizers of the cell-problem strongly depend on the initial guess. Without an a priori knowledge of the behavior of the minimizers (as opposed to the case dealt with in the present section), the global algorithm cannot be used in practice. This a priori knowledge is problem-dependent and can only be obtained by a systematic study of the cell-problem at stake.

5.6.3 Shear band instabilities

In [94], Geymonat et. al. have studied the stability of homogenized energy densities and have suggested that, under some hypotheses, the homogenized material can develop *shear band instabilities*, that is no resistance of the material in one shear direction at least.

These shear band instabilities are linked to a loss of strict rank-one convexity of the homogenized energy density: there exist two vectors $a, b \in \mathbb{R}^3$ such that for all $A \in \mathcal{M}_3(\mathbb{R})$ the function $\mathbb{R} \to \mathbb{R}, t \mapsto W_{hom}(A + ta \otimes b)$ is convex but not strictly convex.

We have performed numerical tests and have obtained several corresponding shear bands, strongly depending on the mesh. The problem considered is of type (5.68) and is posed on a cube submitted to:

- $u(x) = -0.2\, x_3\, e_3$ on the faces $x_3 = 0$ and $x_3 = 1$;
- homogeneous natural boundary conditions elsewhere.

The cell-problem has been posed on one single periodic cell, and numerically solved with 192 degrees of freedom. The macroscopic deformed solutions showing the shear bands are plotted on Figure 5.1. Due to geometric incompatibilities of the shear direction with the (very coarse) mesh, the system is invertible. The convergence of the Newton algorithm is however very slow (around 18 iterations), which shows that the energy landscape is very flat.

This shear band instability is the cause of two major difficulties: the approximation result of Section 5.3 is not valid any more since the minimizer is not isolated and the Newton algorithm fails to converge. Mechanically speaking this property of the homogenized energy density is an artefact due to the homogenization procedure: the real layered material with $\epsilon > 0$ is strictly rank-one convex and does not have shear band instabilities at any scale. Therefore, the homogenized energy has non-mechanical minimizers, which makes the numerical practice impossible. In order to be able to compute minimizers and to recover an approximation result we can stabilize the approximate energy density by adding a small strictly rank-one convex perturbation.

This stabilization procedure, which is naive and may certainly be improved in many ways, can also be seen as a filtering procedure which allows to get rid of a range of meaningless minimizers. In numerical tests, stabilizing in such a way is not always sufficient to guarantee the convergence

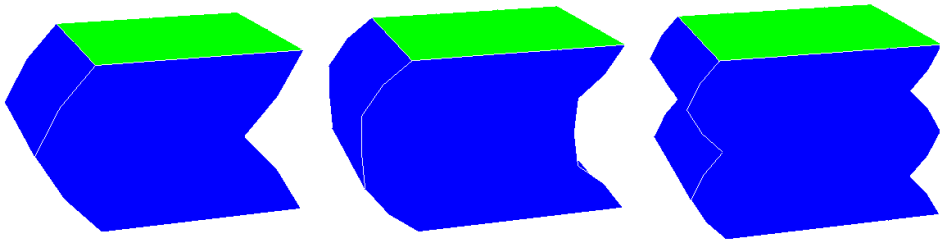

Fig. 5.1. 2, 3 and 4 shear bands

of the Newton algorithm as the macroscopic mesh gets finer. We are indeed limited by the ratio between the mesh of the cell-problem and the mesh of the macroscopic domain to go further in the numerical study.

Buckling is clearly a limit of the numerical approach developped throughout the paper. We have thus tried to determine numerically, for specific periodic cells, the occurence of buckling in the cell-problem for a given range of loads A. With the constitutive law of S. Müller and a ratio of 1000, as for $C(y)$, we have not been able to make cell-problems buckle within a wide range of loads for two simple geometries: a cubic inclusion in a matrix and a three-dimensional chessboard. This is no proof that we have reached a solution with the lowest energy but it allows us to deal with numerically stable cases.

5.6.4 Tests on a wider class of energies

Cases of practical interest do not usually satisfy the growth condition (5.1) and the homogenization formula (5.4) has not been proved in this framework. We have however tested the numerical method in such a case. The basic example of a polyconvex energy density dealt with models an ideal rubber foam, that is a material made of a rubber matrix with air bubbles of a few microns at a given concentration. Several constitutive laws have been proposed in the literature to model rubber foams. They are however more likely to give a coarse description than provide with quantitative results ([11], [61]).

The material considered is obtained by the periodic replication of a unit cell. This cell is composed of a rubber matrix and a bubble of air. The rubber matrix is a cube and the bubble is supposed spherical as in Figure 5.2. The proportion of air ranges from 5% to 15% in the numerical experiments.

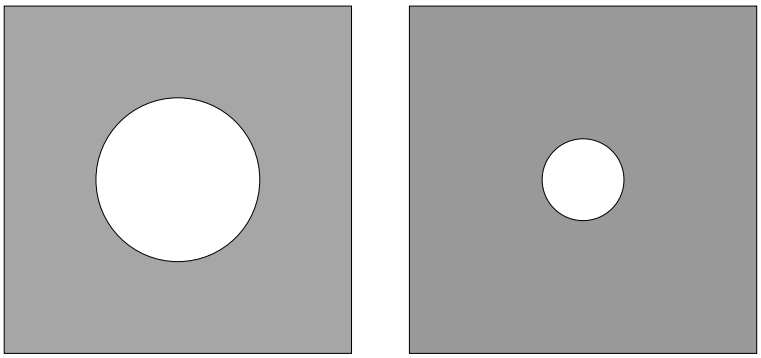

Fig. 5.2. Two examples of unit cell (in 2D)

A classical constitutive law for rubber-like materials is the Ciarlet-Geymonat constitutive law. Its stored energy function W is polyconvex and depends on the three invariants of the strain tensor ∇u, it is characterized by three positive constants C_1, C_2, a, and is given by

$$W(F) = C_1(I_1 - 3) + C_2(I_2 - 3) + a(I_3 - 1) - (C_1 + 2C_2 + a)lnI_3,$$

where $I_1 = \mathrm{Tr}(C)$, $I_2 = 1/2(\mathrm{Tr}(C)^2 - \mathrm{Tr}(C^2))$ and $I_3 = \det(C)$, with $C =^T (Id + \nabla u) \cdot (Id + \nabla u)$. The term $ln(I_3)$ does not satisfy (5.1).

The numerical values of the above constants are typically

$$\begin{cases} C_1 = 0.5\, MPa, \\ C_2 = 0.0056\, MPa. \end{cases}$$

W is compressible for finite a, typically of the order of C_1. In the limit $a \to \infty$ the material becomes incompressible. Quasi-incompressible materials are materials with a finite but quite important a. In the quasi-incompressible case, numerical difficulties arise (locking) if the volumetric part of W (that is the part depending on I_3) is not treated correctly ([125]). In Modulef, this difficulty is overcome thanks to a mixed formulation ([177]).

We have carried out some numerical tests with such cell-problems and constitutive relations. Among these were the Rivlin cube test, a test of compression and a test of extension of a simple cube (Figure 5.3). The pictures display one eighth of the deformed cubes of a rubber foam made of a matrix with an Ogden law and a hole (with an energy taken to be 10^{-3} times smaller than the one of the matrix). Symmetric relations have been imposed on the bottom and on two side faces of the cubes. The cell-problem has been solved with 648 degrees of freedom. In the cases under investigation, the algorithm is quite stable for a wide range of loads and has not encountered the convergence difficulties linked to the loss of strict rank-one convexity. The geometry has been chosen in order to allow us to use a unique periodic cell in the cell-problem. Therefore we have not numerically investigated the influence of the number of periodic cells for Ogden laws. The aim of the tests is to check the feasability of the approach; the behavior of the solution when H and h go to zero has not been addressed in the present work.

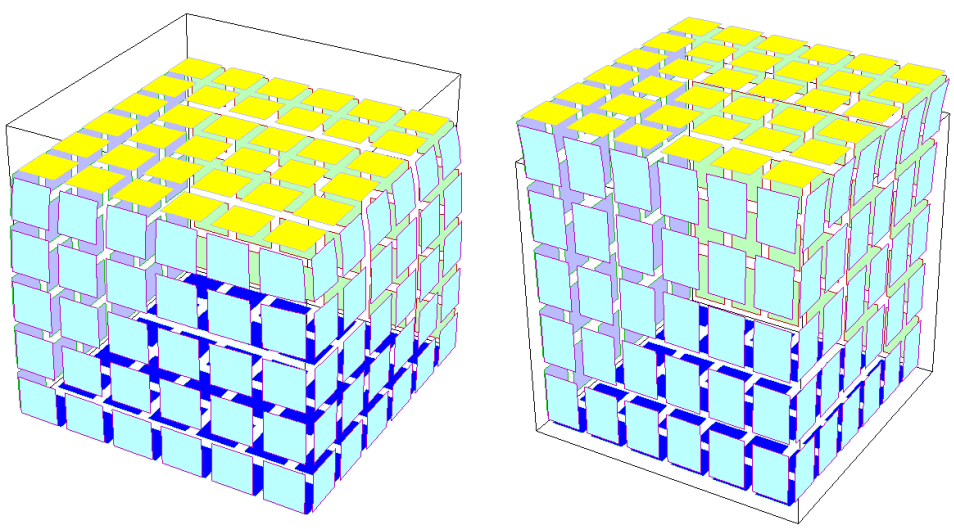

Fig. 5.3. Compression and extension (15%) of a porous rubber

6

An analytical framework for the numerical homogenization of monotone elliptic operators and quasiconvex energies

Summary. A number of methods have been proposed in recent years to perform the numerical homogenization of (possibly nonlinear) elliptic operators. These methods are usually defined at the discrete level. Most of them compute a numerical operator, close, in a sense to be made precise, to the homogenized elliptic operator for the problem. The purpose of the present work is to clarify the construction of this operator in the convex case by interpreting the method at the continuous level and to extend it to the quasiconvex setting. The discretization of this new operator may be performed in several ways, recovering and generalizing a variety of methods such as the multiscale finite element method (MsFEM) or the heterogeneous multiscale method (HMM). In addition to the above, we introduce an original and general numerical corrector in the convex case.

6.1 Setting of the problem and statement of the main results

For the numerical homogenization of elliptic problems, three major issues can be identified: the design of a numerical method, the convergence of the method and possibly the existence of a corrector, and the derivation of error estimates. The present paper focuses on the second issue in an abstract way which allows us to cover several methods proposed in the literature (multiscale finite element method (MsFEM) and heterogeneous multiscale method (HMM)) in new and important settings. Depending on the applications, two types of assumptions are usually made: a space assumption (periodicity, stationary process, ...) and a structure assumption (scalar- or vector-valued, linearity, monotonicity, ...). One cannot expect the three questions to be answered positively with any assumptions. We address the second issue in the wide class of convex and quasiconvex energies, and the third issue in the more restricted class of strictly convex energies. Error estimates are made explicit in the periodic case. Our main achievements are a numerical corrector result in the convex case without space assumption, a complete error analysis in the periodic and strictly convex case, and a convergence result in a general vector-valued case (nonlinear elasticity).

For the sake of simplicity and in order to relate the present work to existing approaches, we first consider a scalar elliptic PDE that is the Euler–Lagrange equation of a problem of minimization of an energy. The energy density is assumed to be convex and to vary at a scale small with respect to the size of the domain, which makes direct numerical simulations impossible to perform in practice. Our purpose is to introduce an averaged energy density which does not vary as much as the original energy density and prove that the associated minimization problem is a correct approximation of the original minimization problem in the sense of Γ-convergence.

Although strict convexity ensures the existence and uniqueness of the solution to the Euler–Lagrange equation and allows us to use a more direct approach (e.g., G-convergence), our arguments are based on variational principles. Most of the existing works (see, e.g., [72], [74], and [69]) actually treat the PDE. Our approach allows us to deal with the case of nonlinear elasticity, as will be seen. The first section is dedicated to the introduction of the averaged energy density in the framework of the homogenization theory for convex energies. Theorem 40 states the Γ-convergence of the averaged energy to the homogenized energy and Theorem 41 introduces a *new*

general corrector result, which describes the fine scale features of the solution of the original minimization problem *without assumption on the heterogeneities*. This corrector is called a numerical corrector in reference to [74], where it was first derived in the stochastic and stationary case. In Propositions 5 and 6, we also introduce some error estimates. Section 6.2 is dedicated to the proofs of the main results.

Besides the analysis of convex problems, a generalization of this approach is introduced in section 6.1.3 to deal with *quasiconvex energy densities*. Definition 29 and Theorem 42 are the natural counterparts of Definition 22 and Theorem 40 for quasiconvex energy densities. The proof of Theorem 40 is adapted to the quasiconvex case in section 6.2.3.

Finally, section 6.3 relates the present work to some well-known numerical methods such as the MsFEM and the HMM to which the present analysis applies both in the convex and quasiconvex settings.

We assume the reader is familiar with the basic properties of the Γ-convergence theory. Should the need arise, [36] provides with a good introduction and [56] gives a more systematic study of the subject. For consistency, let us recall some notation and properties of the Γ-convergence in Sobolev spaces. In what follows, Ω denotes an open bounded subset of $\mathbb{R}^n$ $(n \geq 1)$, $W^{1,p}(\Omega)$ denotes the Sobolev space for $p \geq 1$, and p' denotes the conjugate exponent defined by $\frac{1}{p} + \frac{1}{p'} = 1$. We will also make use of the notation $k \wedge k' = \max(k, k')$ for all $k, k' \in \mathbb{R}$.

Definition 20 *Let $I_\epsilon : W^{1,p}(\mathbb{R}^n) \to \mathbb{R}$ be a family of functions. We say that I_ϵ $\Gamma(w - W^{1,p})$-converges (resp., $\Gamma(W^{1,p})$-converges) to $I : W^{1,p}(\mathbb{R}^n) \to \mathbb{R}$ on Ω if and only if the two following properties are satisfied.*

(i) *Liminf inequality: for every $u \in W^{1,p}(\Omega)$ and every sequence u_ϵ such that $u_\epsilon \rightharpoonup u$ (resp., $u_\epsilon \to u$) in $W^{1,p}(\Omega)$,*

$$I(u) \leq \liminf_{\epsilon} I_\epsilon(u_\epsilon).$$

(ii) *Recovery sequence: for every $u \in W^{1,p}(\Omega)$ there exists a sequence $\bar{u}_\epsilon$ such that $\bar{u}_\epsilon \rightharpoonup u$ (resp., $\bar{u}_\epsilon \to u$) in $W^{1,p}(\Omega)$ and*

$$\limsup_{\epsilon} I_\epsilon(\bar{u}_\epsilon) \leq I(u).$$

Definition 20 is also referred to as the sequential Γ-convergence since it is stated using the convergence of sequences. We refer the reader to [36, section 1.4] for other definitions, which are equivalent in the present context.

The Γ-convergence implies the convergence of minima and minimizers of functions, as stated in the following.

Lemma 6.1. *Let $I_\epsilon : W^{1,p}(\mathbb{R}^n) \to \mathbb{R}$ be a family of functions that $\Gamma(w - W^{1,p})$-converges to I on Ω. If I_ϵ is lower semicontinuous for the weak topology of $W^{1,p}(\mathbb{R}^n)$ and equicoercive in the following sense,*

$$\exists \quad c > 0 \quad \text{for all } v \in W^{1,p}(\Omega), \text{ for all } \epsilon > 0, \quad c\|\nabla v\|^p_{L^p(\Omega)} \leq I_\epsilon(v),$$

then for every $u_0 \in W^{1,p}(\Omega)$

$$\lim_{\epsilon \to 0} \left(\inf\{I_\epsilon(v + u_0), v \in W_0^{1,p}(\Omega)\} \right) = \inf\{I(v + u_0), v \in W_0^{1,p}(\Omega)\}$$

and for every sequence u_ϵ of minimizers of $\inf\{I_\epsilon(v+u_0), v \in W_0^{1,p}(\Omega)\}$ there exists a subsequence (not relabeled) and a minimizer u of $\inf\{I(v + u_0), v \in W_0^{1,p}(\Omega)\}$ such that $u_\epsilon \rightharpoonup u$ in $W^{1,p}(\Omega)$.

In the following, Γ denotes the $\Gamma(L^p)$-convergence, which is equivalent to the $\Gamma(w - W^{1,p})$ for equicoercive functions. For all open bounded subset $\mathcal{O}$ of $\mathbb{R}^n$ and $u \in L^1(\mathcal{O})$, we denote

$$\langle u \rangle_{\mathcal{O}} = \frac{1}{\mathcal{L}^n(\mathcal{O})} \int_{\mathcal{O}} u(x)dx,$$

where $\mathcal{L}^n$ stands for the Lebesgue measure on $\mathbb{R}^n$.

6.1.1 Homogenization of convex energy densities

Let us first recall classical homogenization results for convex energy densities (see [38], [56], and [155]).

Definition 21 *A function $W : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a Carathéodory function if for every $\xi \in \mathbb{R}^n$, $W(\cdot, \xi)$ is measurable and if for almost all $x \in \mathbb{R}^n$, $W(x, \cdot)$ is continuous.*

A Carathéodory function on $\mathbb{R}^n \times \mathbb{R}^n$ is equivalent to a Borel function on $\mathbb{R}^n \times \mathbb{R}^n$.

Lemma 6.2 (see [56, Theorem 20.4]). *Let $W_\epsilon : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty)$ be a set of functions satisfying the following conditions:*

- *$H1$: W_ϵ is a Carathéodory function.*
- *$H2$: for almost every $x \in \mathbb{R}^n$, $W_\epsilon(x, \cdot)$ is convex.*
- *$H3$: there exist $0 < c \leq C$ and $p \geq 1$ such that*

$$c|\xi|^p \leq W_\epsilon(x, \xi) \leq C(1 + |\xi|^p)$$

for almost all $x \in \mathbb{R}^n$ and for all $\xi \in \mathbb{R}^n$.

Consider Ω a bounded open subset of $\mathbb{R}^n$ and set for all $\epsilon > 0$,

$$I_\epsilon(u) = \int_\Omega W_\epsilon(x, \nabla u(x))dx \tag{6.1}$$

for all $u \in W^{1,p}(\Omega, \mathbb{R})$. Then, up to extraction, there exists a function W_{hom} satisfying $H1$, $H2$, and $H3$, such that we have

$$\Gamma(L^p) - \lim_{\epsilon \to 0} I_\epsilon(u) = \int_\Omega W_{hom}(x, \nabla u(x))dx,$$

for all $u \in W^{1,p}(\Omega, \mathbb{R})$.

Lemma 6.3 (see [38, Theorem 23.2 and Remark 23.5]). *In addition to $H1$, $H2$, and $H3$, let us assume that $p \geq 2$, that*

- *$H4$: $W_\epsilon(x, \cdot)$ is continuously differentiable for almost all $x \in \Omega$ and $a_\epsilon(\cdot, 0) = \frac{\partial W_\epsilon}{\partial \xi}(\cdot, 0)$ is bounded,*

and that the following monotonicity and continuity properties hold:

$$\begin{aligned} &\exists\ 0 \leq \alpha \leq p-1, C > 0 \quad | \quad \text{for almost all } x \in \mathbb{R}^n, \text{ for all } \xi_1, \xi_2 \in \mathbb{R}^n, \\ &\qquad |a_\epsilon(x, \xi_1) - a_\epsilon(x, \xi_2)| \leq C(1 + |\xi_1| + |\xi_2|)^{p-1-\alpha}|\xi_1 - \xi_2|^\alpha, \end{aligned} \tag{6.2}$$

$$\begin{aligned} &\exists\ 2 \leq \beta < +\infty, c > 0 \quad | \quad \text{for almost all } x \in \mathbb{R}^n, \text{ for all } \xi_1, \xi_2 \in \mathbb{R}^n, \\ &\qquad (a_\epsilon(x, \xi_1) - a_\epsilon(x, \xi_2), \xi_1 - \xi_2) \geq c(1 + |\xi_1| + |\xi_2|)^{p-\beta}|\xi_1 - \xi_2|^\beta. \end{aligned} \tag{6.3}$$

Then, given $f \in L^{p'}(\Omega)$, the solution $u_\epsilon \in W_0^{1,p}(\Omega)$ to

$$-\text{div}\ (a_\epsilon(x, \nabla u_\epsilon)) = f$$

weakly converges in $W_0^{1,p}(\Omega)$, up to extraction, to the solution $u \in W_0^{1,p}(\Omega)$ to

$$-\text{div}\ (a_{hom}(x, \nabla u)) = f,$$

where $a_{hom} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is related to W_{hom} defined in Lemma 6.2 by

$$a_{hom} = \frac{\partial W_{hom}}{\partial \xi}.$$

In addition a_{hom} satisfies (6.3) with the same coefficient c and exponent β, and (6.2) with the same coefficient C and the exponent $\alpha/(\beta - \alpha)$ instead of α.

6.1.2 Main results

Let (W_ϵ) be a family of energy densities satisfying the assumptions $H1$, $H2$, and $H3$. The problem we consider is

$$\inf\left\{\int_\Omega W_\epsilon(x,\nabla u), u \in W^{1,p}(\Omega) + BC\right\}. \tag{6.4}$$

By $W^{1,p}(\Omega) + BC$ in (6.4), we mean any subspace of $W^{1,p}(\Omega)$ associated with usual boundary conditions. In particular, we will consider

(1) $\phi + W_0^{1,p}(\Omega)$ for any $\phi \in W^{1,p}(\Omega)$,
(2) $\{u \in W^{1,p}(\Omega) \,|\, u(x) = \xi \cdot x + v(x), v \in W_\#^{1,p}(\Omega)\}$ for any $\xi \in \mathbb{R}^n$, with $W_\#^{1,p}(\Omega) = \{v_{|\Omega} \,|\, v \in W_{loc}^{1,p}(\mathbb{R}^n), v$ is Ω-periodic$\}$ if $\mathbb{R}^n$ can be obtained by the periodic replication of Ω,
(3) $\{u \in W^{1,p}(\Omega) \,|\, \langle \nabla u\rangle_\Omega = \xi\}$ for any $\xi \in \mathbb{R}^n$.

These three boundary conditions indeed do not influence the energy density of the Γ-limit of (6.4), as briefly recalled in the appendix. This is why we do not make them specific in what follows.

Up to extraction, problem (6.4) Γ-converges to

$$\inf\left\{\int_\Omega W_{hom}(x,\nabla u), u \in W^{1,p}(\Omega) + BC\right\}. \tag{6.5}$$

The density W_{hom} is not explicitly known. Lemma 6.2 provides us only with an existence result.

For brevity, we denote by

$$I_\epsilon(u) = \int_\Omega W_\epsilon(x,\nabla u),$$
$$I_{hom}(u) = \int_\Omega W_{hom}(x,\nabla u)$$

for $u \in W^{1,p}(\Omega)$.

We now introduce a notion of energy averages on balls.

Definition 22 *For any $\eta > 0$, denoting by $B(x,\eta)$ the ball of radius η centered at point $x \in \mathbb{R}^n$, we define the energy density*

$$W_{\eta,\epsilon}(x,\xi) = \inf\left\{\langle W_\epsilon(\cdot,\nabla v(\cdot))\rangle_{B(x,\eta)} \,|\, v \in W^{1,p}(B(x,\eta)), \langle \nabla v\rangle_{B(x,\eta)} = \xi\right\} \tag{6.6}$$

from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ and the associated energy functional

$$I_{\eta,\epsilon}(u) = \int_\Omega W_{\eta,\epsilon}(x,\nabla u) \quad \text{for all } u \in W^{1,p}(\Omega).$$

Remark **10** We have used balls $B(x,\eta)$ for defining averaged energies. All the results presented throughout this work hold for generic open neighborhoods $N(x,\eta)$ with a Lipschitz boundary and satisfying that for every $x \in \Omega$, there exist $0 < c \le C$ such that for every $\eta > 0$, $c|B(x,\eta)| \le |N(x,\eta)| \le C|B(x,\eta)|$.

Let us introduce the concept of "equi-isolated minimizers" before stating the first main result.

Definition 23 *Given a family of energy functionals I_N on the metric space (V,d), we say that a family u_N of minimizers of I_N on (V,d) is equi-isolated if there exists a ball $B \subset V$ such that $u_N \in B$ and u_N is the unique minimizer of I_N on B for every $N > 0$.*

Our first result is the following.

Theorem 40 *For $p > 1$, the energy densities $W_{\eta,\epsilon}$ satisfy $H1$, $H2$, and $H3$, and the energy $I_{\eta,\epsilon}$ $\Gamma(L^p)$-and $\Gamma(W^{1,p})$-converges to I_{hom} as ϵ and η go to 0. Therefore, for any sequence $u_{\eta,\epsilon}$ of minimizers of $\inf\{I_{\eta,\epsilon}(v) \,|\, v \in W^{1,p}(\Omega,\mathbb{R}) + BC\}$, there exists a minimizer u_{hom} of $\inf\{I_{hom}(v) \,|\, v \in W^{1,p}(\Omega,\mathbb{R}) + BC\}$ such that*

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0} u_{\eta,\epsilon} = u_{hom} \quad \text{weakly in } W^{1,p}(\Omega,\mathbb{R}), \tag{6.7}$$

up to extraction.

Correspondingly, for any minimizer u_{hom} of $\inf\{I_{hom}(v) \,|\, v \in W^{1,p}(\Omega,\mathbb{R}) + BC\}$ there exists a sequence $u_{\eta,\epsilon}$ of minimizers of $\inf\{I_{\eta,\epsilon}(v) \,|\, v \in W^{1,p}(\Omega,\mathbb{R}) + BC\}$ such that

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0} u_{\eta,\epsilon} = u_{hom} \quad \text{strongly in } W^{1,p}(\Omega,\mathbb{R}). \tag{6.8}$$

In addition, if $u_{\eta,\epsilon}$ is a family of equi-isolated minimizers in the sense of Definition 23, then u_{hom} is also isolated and (6.8) holds.

In particular, if $W_\epsilon(x,\cdot)$ is strictly convex for almost every $x \in \mathbb{R}^n$, the unique sequence of minimizers $u_{\eta,\epsilon}$ strongly converges in $W^{1,p}(\Omega,\mathbb{R})$ to the unique minimizer u_{hom}.

Remark **11** The order of the limits in (6.7) is important and cannot be changed in general.

We can also define a set of numerical correctors to approximate ∇u_ϵ in $L^p(\Omega)$.

Definition 24 *Let $\{Q_{H,i}\}_{i\in[\![1,I_H]\!]}$ be a partition of Ω in disjoint subdomains of diameter of order H. We define a family (M_H) of approximations of identity on $L^p(\Omega)$ associated with $Q_{H,i}$: for every $w \in L^p(\Omega)$ and $H > 0$,*

$$M_H(w) = \sum_{i=1}^{I_H} \langle w\rangle_{Q_{H,i}} 1_{Q_{H,i}}.$$

Keeping the notation of Theorem 40, we define the numerical correctors $v^{H,i}_{\eta,\epsilon}$ for a strictly convex energy density as the unique minimizers of

$$\inf\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla v) \,|\, v \in W^{1,p}(Q_{H,i}), \langle\nabla v\rangle_{Q_{H,i}} = \langle\nabla u_{\eta,\epsilon}\rangle_{Q_{H,i}}\right\}. \tag{6.9}$$

In particular, $\langle\nabla v^{H,i}_{\eta,\epsilon}\rangle_{Q_{H,i}} = M_H(\nabla u_{\eta,\epsilon})_{|Q_{H,i}}$.

Theorem 41 *Assume $p \geq 2$, $H1$, $H2$, $H3$, $H4$, (6.2), and (6.3) with $\beta \leq p$. We keep the notation of Lemma 6.3, Theorem 40, and Definition 24. We have*

$$\lim_{\eta,H\to 0}\lim_{\epsilon\to 0}\left\|\nabla u_\epsilon - \sum_{i=1}^{I_H}\nabla v^{H,i}_{\eta,\epsilon} 1_{Q_{H,i}}\right\|_{L^p(\Omega)} = 0. \tag{6.10}$$

A similar result has been obtained by Murat in [147].

Remark **12** The order of the limits in H and η in (6.10) is not important, and we may take, e.g., $H = \eta \to 0$. However, we have to first let ϵ go to zero.

Remark **13** Theorem 40 holds if $W_{\eta,\epsilon}(x,\xi)$ is replaced by

$$\inf\left\{\langle W_\epsilon(\cdot,\nabla v(\cdot))\rangle_{B(x,\eta)} \,|\, v \in W^{1,p}(B(x,\eta)), v(y) = \xi\cdot y \text{ on } \partial B(x,\eta)\right\}. \tag{6.11}$$

Theorem 41 also holds if $v^{H,i}_{\eta,\epsilon}$ is replaced by the unique solution in $W^{1,p}(Q_{H,i})$ of

$$\inf\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla v) \,|\, v(x) = M_H(\nabla u_{\eta,\epsilon})_{|Q_{H,i}}\cdot x \text{ on } \partial Q_{H,i}\right\}. \tag{6.12}$$

The proofs of section 6.2 can easily be adapted to these cases.

Remark **14** If the neighborhood $N(x,\eta)$ is such that $\mathbb{R}^n$ can be obtained by the periodic replication of $N(x,\eta)$, then Theorem 40 holds with

$$W_{\eta,\epsilon}(x,\xi) = \inf\left\{\langle W_\epsilon(\cdot,\xi+\nabla v(\cdot))\rangle_{N(x,\eta)} \mid v \in W^{1,p}_{\#}(N(x,\eta))\right\}. \tag{6.13}$$

In the following proposition, we introduce a notion of local error and derive a global error estimate under the assumption of monotonicity.

Proposition 5 *We keep the notation of Lemma* 6.3 *and Theorem* 40*. For all* $\xi \in \mathbb{R}^n$*,* $\eta > 0$*, and* $\epsilon > 0$*, and for all* $x \in \Omega$*, let us denote by*

$$err^0_{\eta,\epsilon}(x,\xi) = |W_{\eta,\epsilon}(x,\xi) - W_{hom}(x,\xi)|,$$

$$err^1_{\eta,\epsilon}(x,\xi) = |a_{\eta,\epsilon}(x,\xi) - a_{hom}(x,\xi)|.$$

Assume $p \geq 2$*, H*1*, H*2*, H*3*, H*4*,* (6.2)*, and* (6.3)*; then there exists* $C_1 > 0$ *independent of* ϵ *and* η *such that for all Dirichlet boundary conditions* $\bar{u} \in W^{1,p}(\Omega)$ *of problem* (6.5)

$$\|u_{\eta,\epsilon} - u_{hom}\|_{1,p} \leq C_1(1+\|\bar{u}\|_{1,p})^{\frac{\beta\wedge p-p}{\beta}} \left(\int_\Omega err^0_{\eta,\epsilon}(x,\nabla u_{\eta,\epsilon}) + err^0_{\eta,\epsilon}(x,\nabla u_{hom})\right)^{\frac{1}{\beta\wedge p}}, \tag{6.14}$$

$$\|u_{\eta,\epsilon} - u_{hom}\|_{1,p} \leq C_1(1+\|\bar{u}\|_{1,p})^{\frac{\beta\wedge p-p}{\beta-1}} \left(\int_\Omega err^1_{\eta,\epsilon}(x,\nabla u_{hom})^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}\frac{1}{\beta\wedge p-1}}. \tag{6.15}$$

If there exist two functions $g_0, g_1 : \Omega \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ *such that*

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0}\sup_{x\in\Omega} g_i(x,\eta,\epsilon) = 0 \quad \textit{for } i = 0,1$$

and such that for all $\xi \in \mathbb{R}^n$*,* $\epsilon > 0$*,* $\eta > 0$*, and* $x \in \Omega$*,*

$$err^0_{\eta,\epsilon}(x,\xi) \leq g_0(x,\eta,\epsilon)(1+|\xi|^p), \tag{6.16}$$

$$err^1_{\eta,\epsilon}(x,\xi) \leq g_1(x,\eta,\epsilon)(1+|\xi|^{p-1}), \tag{6.17}$$

then the right-hand sides of (6.14) *and* (6.15) *can be made independent of the solutions* $u_{\eta,\epsilon}$ *and* u_{hom} *as follows:*

$$\|u_{\eta,\epsilon} - u_{hom}\|_{1,p} \leq C_1 \left(\sup_{x\in\Omega} g_0(x,\eta,\epsilon)\right)^{\frac{1}{\beta\wedge p}} (1+\|\bar{u}\|_{1,p}), \tag{6.18}$$

$$\|u_{\eta,\epsilon} - u_{hom}\|_{1,p} \leq C_1 \left(\sup_{x\in\Omega} g_1(x,\eta,\epsilon)\right)^{\frac{1}{\beta\wedge p-1}} (1+\|\bar{u}\|_{1,p}). \tag{6.19}$$

In particular, if $W_\epsilon(x,\cdot) = W(\frac{x}{\epsilon},\cdot)$*,* $y \mapsto W(y,\cdot)$ *is a* 1*-periodic function, and if* $W_{\eta,\epsilon}$ *is given by* (6.11) *on a cube (still denoted by* $B(x,\eta)$*) instead of a ball, then assumptions* (6.16) *and* (6.17) *hold, and there exists* C_2 *depending only on* $\bar{u}$ *such that* (6.18) *reads*

$$\|u_{\eta,\epsilon} - u_{hom}\|_{1,p} \leq C_2 \left(\frac{\epsilon}{\eta}\right)^{\frac{1}{\beta\wedge p}}. \tag{6.20}$$

In the specific case of a quadratic energy ($p = 2$*,* $\alpha = 1$*, and* $\beta = 2$*),* (6.19) *gives the improved error estimate*

$$\|u_{\eta,\epsilon} - u_{hom}\|_{1,2} \leq C_2 \left(\frac{\epsilon}{\eta}\right), \tag{6.21}$$

which is sharp.

Remark **15** The drawback of estimates (6.14) and (6.15) is that both sides of the inequalities depend on $u_{\eta,\epsilon}$ or u_{hom}. Assumptions (6.16) and (6.17) are natural: if the errors in (ϵ,η) and in ξ are decoupled, then the exponents p and $p-1$ may appear since the solutions belong to $W^{1,p}(\Omega)$; this is the case for periodic energy densities. It may be noticed that hypothesis (6.17) is a generalization for $p \geq 2$ of the assumption on $e(HMM)$ in [68, Theorem 1.1].

Remark **16** The improved error estimate (6.21) relies on specific properties of quadratic energies, namely, that $W(x,\xi)$ and $|a(x,\xi)||\xi|$ are of the same order and that the class of quadratic energies is closed by homogenization (the homogenized energy is still quadratic). This is very particular and cannot be generalized to other monotone operators.

In this last proposition, we give a partial result regarding the sharpness of the numerical correctors. We indeed compare the numerical corrector of Theorem 41 obtained with formula (6.12) to a corrector that would fit correctly to the periodic pattern and would be defined using the true solution u_{hom} of the homogenized problem, in the spirit of the classical corrector result of [58].

Proposition 6 *In the case of a periodic energy density satisfying $p \geq 2$, $H1$, $H2$, $H3$, $H4$, (6.2), and (6.3), let us consider the corrector functions defined by (6.12) with $H \simeq \eta$. Let $p^{H,i}(\xi)$ denote the restriction on $Q_{H,i}$ of the periodization of the rescaled classical corrector function $x \mapsto \epsilon\tilde{p}_\xi\left(\frac{x}{\epsilon}\right)$, where $y \mapsto \tilde{p}_\xi(y)$ is introduced in [58] as the minimizer of*

$$\inf\left\{\int_{(0,1)^n} W(y,\xi+\nabla p(y)), p \in W^{1,p}_{\#}((0,1)^n)\right\}. \tag{6.22}$$

Let us denote by $\xi_{hom,i} = \langle \nabla u_{hom}\rangle_{Q_{H,i}}$ for all i, and assume that $\{u_{\eta,\epsilon}\}$ is a bounded sequence in $W^{1,q}(\Omega)$ for some $q \geq p$; then there exist $C_3 > 0$ and $C_4 > 0$ independent of ϵ and η such that

$$\begin{aligned}
&\left\|\sum_{i=1}^{I_H}\left(\nabla v^{H,i}_{\eta,\epsilon} - (\xi_{hom,i} + \nabla p^{H,i}(\xi_{hom,i}))\right)1_{Q_{H,i}}\right\|_{L^p(\Omega)} \\
&\qquad \leq C_3\left(\frac{\epsilon}{\eta}\right)^{\frac{1}{\beta\wedge p}}\eta^{-\frac{n}{q}\frac{\beta\wedge p-p}{\beta}} + C_4\left(\|u_{\eta,\epsilon}-u_{hom}\|_{1,p}\right)^{\frac{1}{\beta\wedge p-\alpha}}\eta^{-\frac{n}{q}\frac{\beta\wedge p-\alpha-1}{\beta\wedge p-\alpha}} \\
&\qquad \leq C_3\left(\frac{\epsilon}{\eta}\right)^{\frac{1}{\beta\wedge p}}\eta^{-\frac{n}{q}\frac{\beta\wedge p-p}{\beta}} + C_4\left(\frac{\epsilon}{\eta}\right)^{\frac{1}{\beta\wedge p}\frac{1}{\beta\wedge p-\alpha}}\eta^{-\frac{n}{q}\frac{\beta\wedge p-\alpha-1}{\beta\wedge p-\alpha}}.
\end{aligned} \tag{6.23}$$

In the case of a quadratic energy ($p=2$, $\alpha=1$, and $\beta=2$), using (6.21), we recover the result obtained in [1] and [69],

$$\left\|\sum_{i=1}^{I_H}\left(\nabla v^{H,i}_{\eta,\epsilon} - (\xi_{hom,i} + \nabla p^{H,i}(\xi_{hom,i}))\right)1_{Q_{H,i}}\right\|_{L^2(\Omega)} \leq C_3\left(\frac{\epsilon}{\eta}\right)^{1/2} + C_4\left(\frac{\epsilon}{\eta}\right), \tag{6.24}$$

which is sharp.

Let us discuss the results of Propositions 5 and 6. The error estimates (6.14), (6.15), (6.18), and (6.19) are quite abstract, and are related to those in [68] for $p=2$. Estimates (6.23) and (6.24) cannot be improved in general (using formula (6.12)); however, oversampling techniques can reduce drastically the first term of this error, known as the resonance error. In the linear case, e.g., one can obtain

$$\left\|\sum_{i=1}^{I_H}\left(\nabla \tilde{v}^{H,i}_{\eta,\epsilon} - (\xi_{hom,i} + \nabla p^{H,i}(\xi_{hom,i}))\right)1_{Q_{H,i}}\right\|_{L^2(\Omega)} \leq C_3\left(\frac{\epsilon}{\eta}\right) + C_4\left(\frac{\epsilon}{\eta}\right),$$

where $\tilde{v}^{H,i}_{\eta,\epsilon}$ denotes the modified correctors. We refer the reader to the literature for this issue (mainly [105], [107], and [69]).

Error estimate (6.23) needs to be discussed further since the regularity of solutions plays a significant role. If we consider a power-law material (W is a function of the space variable times a polynomial of degree p in ξ, $\beta = 2$, and $\alpha = 1$) on a convex set Ω, then under very weak hypotheses, $\{u_{\eta,\epsilon}\}$ is bounded in $W^{1+\frac{2}{p}-\tau,p}(\Omega)$ for all $\tau > 0$ (see [71], [169]). Therefore, in two dimensions, the Sobolev embedding theorem implies that $\{\nabla u_{\eta,\epsilon}\}$ is bounded in $L^q(\Omega)$ for all $q < \infty$, which shows that (6.23) does depend only on $\frac{\epsilon}{\eta}$ in practice. In three dimensions, however, the Sobolev embedding theorem implies only that $\{\nabla u_{\eta,\epsilon}\}$ is bounded in $L^{3p-\tau}(\Omega)$ for all $\tau > 0$, and $\frac{\epsilon}{\eta} \to 0$ is not enough to prove the convergence of the numerical corrector. This may be striking. In the proof, this comes from the nonlinearity of the problem: if you consider a nonlinear boundary value problem with two different boundary values, then the difference between the two solutions is in general bigger than the difference between the two boundary values, as quantified in Lemma 6.16.

The sharpness of (6.23) is an important issue. Indeed, as we will discuss in section 6.3, the second term of (6.23) cannot be reduced by oversampling methods and may dominate the first term, which could reduce the efficiency of oversampling techniques in nonlinear cases. Even if $\{u_{\eta,\epsilon}\}$ is bounded in $W^{1,\infty}(\Omega)$, the second term of (6.23) dominates the first term at the limit as soon as $p > 2$.

The result of Proposition 6 is, however, partial since it does not quantify the difference between the numerical corrector and the theoretical corrector (defined on one periodic cell). A control of $\|u_{\eta,\epsilon} - u_{hom}\|_{1,\infty}$ or some Lipschitz continuity of ∇u_{hom} would allow us to obtain such error estimates (see [68] for the linear case); however, $\nabla u_{hom} \notin L^\infty(\Omega)$ in general.

From a theoretical point of view, $u_{\eta,\epsilon}$ has the advantage of strongly converging to u_{hom} in $W^{1,p}(\Omega)$, whereas u_ϵ converges only weakly. This has an important consequence on the practical numerical computation of $u_{\eta,\epsilon}$ and u_ϵ. As $u_{\eta,\epsilon}$ strongly converges to u_{hom}, the gradient of $u_{\eta,\epsilon}$ does not oscillate at order 1 at the period ϵ when η and ϵ are sufficiently small if ∇u_{hom} does not oscillate at small scales. This may allow us to take a meshsize larger than ϵ to approximate $u_{\eta,\epsilon}$ with a finite element method. This is not the case for u_ϵ. In fact, state of the art multiscale methods for elliptic equations usually compute numerical approximations of $u_{\eta,\epsilon}$, as will be seen in section 6.3.

6.1.3 Extension to the quasiconvex case

Let us briefly recall the corresponding version of Lemma 6.2 for quasiconvex energy densities.

Definition 25 *Given $p \geq 1$, $n \geq 1$, and $d \geq 1$, a function $W : \mathbb{R}^{n\times d} \to [0, +\infty]$ is $W^{1,p}$-*quasiconvex *(or simply* quasiconvex *in what follows) if for all $A \in \mathbb{R}^{n\times d}$, there exists an open bounded subset E of $\mathbb{R}^n$ with $\mathcal{L}^n(\partial E) = 0$ such that*

$$W(A) = \min\left\{\frac{1}{\mathcal{L}^n(E)}\int_E W(A + \nabla\phi(x))dx \,|\, \phi \in W_0^{1,p}(E;\mathbb{R}^d)\right\}.$$

Remark **17** If a quasiconvex function is locally bounded, then it is rank-one convex; that is, for every rank-one matrix $\xi \in \mathbb{R}^{n\times d}$ and for every $\zeta \in \mathbb{R}^{n\times d}$, the function $\mathbb{R} \ni t \mapsto W(\zeta + t\xi)$ is convex.

The characteristics of quasiconvexity is to ensure the weak-lower semicontinuity of integral functionals.

Lemma 6.4 (see [38, Theorem 5.16]). *If $1 \leq p < \infty$, and $W : \mathbb{R}^{n\times d} \to \mathbb{R}$ is a quasiconvex function satisfying*

$$0 \leq W(A) \leq C(1 + |A|^p) \quad \text{for all } A \in \mathbb{R}^{n\times d},$$

then the functional $J : u \mapsto \int_\Omega W(\nabla u)$ is weakly lower semicontinuous on $W^{1,p}(\Omega)$.

Definition 26 (see [36, section 12.1]) *Given a continuous function $f : \mathbb{R}^{n\times d} \to \mathbb{R}, \xi \mapsto f(\xi)$ that satisfies* (6.25)*, its quasiconvex envelope Qf is defined as the greatest quasiconvex function lower than or equal to f. In particular, for all $\xi \in \mathbb{R}^{n\times d}$, it satisfies*

$$Qf(\xi) = \inf\left\{\int_{(0,1)^n} f(\xi + \nabla u(y))dy \,|\, u \in W_0^{1,p}((0,1)^n, \mathbb{R}^d)\right\}$$
$$= \inf\left\{\int_{(0,1)^n} f(\xi + \nabla u(y))dy \,|\, u \in W_\#^{1,p}((0,1)^n, \mathbb{R}^d)\right\}.$$

Definition 27 *Let $(x, A) \mapsto W(x, A)$ be a Carathéodory function defined on $\mathbb{R}^n \times \mathbb{R}^{n\times d}$, for which there exist an integer $p \geq 1$ and positive constants c and C, such that for almost all $x \in \mathbb{R}^n$ and for all $A \in \mathbb{R}^{n\times d}$,*

$$c|A|^p \leq W(x, A) \leq C(1 + |A|^p). \tag{6.25}$$

The function W is then said to satisfy a standard growth condition *(of order p).*

Definition 28 *The function $W : \mathbb{R}^n \times \mathbb{R}^{n\times d} \to \mathbb{R}, (x, A) \mapsto W(x, A)$ is a* standard energy density *if W is a quasiconvex Carathéodory function, that is,*

- *$W(\cdot, \cdot)$ is measurable in its first variable and continuous in its second variable,*
- *$W(x, \cdot)$ is quasiconvex for almost every $x \in \mathbb{R}^n$,*

and if W satisfies (6.25).

We are now in position to recall the homogenization result for quasiconvex energy densities.

Lemma 6.5 (see [38, Theorem 12.5]). *Let $W_\epsilon : \mathbb{R}^n \times \mathbb{R}^{n\times d} \to [0, +\infty)$ be a set of standard energy densities satisfying the growth condition* (6.25) *of order $p > 1$ uniformly in ϵ. For any bounded open subset Ω of $\mathbb{R}^n$, for all $u \in W^{1,p}(\Omega, \mathbb{R}^d)$ and $\epsilon > 0$, we set*

$$I_\epsilon(u) = \int_\Omega W_\epsilon(x, \nabla u(x))dx.$$

Then, there exists a homogenized standard energy density $W_{hom} : \Omega \times \mathbb{R}^{n\times d} \to [0, +\infty)$ satisfying (6.25) *and such that, up to extraction, $\Gamma(L^p) - \lim_{\epsilon\to 0} I_\epsilon = I_{hom}$ on $W^{1,p}(\Omega, \mathbb{R}^d)$, where $I_{hom}(u) = \int_\Omega W_{hom(x,\nabla u(x))dx}$.*

Contrary to the convex case, quasiconvexity is not preserved by the averaging of Definition 22. Therefore we have to use quasiconvex envelopes to obtain results corresponding to Theorem 40.

Definition 29 *For $\eta > 0$, let us denote by $C(x, \eta)$ the hypercube of $\mathbb{R}^n$ centered in $x \in \mathbb{R}^n$ and of length η. We then define the averaged energy density by*

$$\mathcal{W}_{\eta,\epsilon}(x, \xi) = \inf\left\{\langle W_\epsilon(\cdot, \xi + \nabla v(\cdot))\rangle_{C(x,\eta)} \,|\, v \in W_\#^{1,p}(C(x, \eta), \mathbb{R}^d)\right\} \tag{6.26}$$

from $\mathbb{R}^n \times \mathbb{R}^{n\times d}$ to $\mathbb{R}$ and the energy functional associated with its quasiconvex envelope $Q\mathcal{W}_{\eta,\epsilon}$

$$I_{\eta,\epsilon}(u) = \int_\Omega Q\mathcal{W}_{\eta,\epsilon}(x, \nabla u) \quad \text{for all } u \in W^{1,p}(\Omega, \mathbb{R}^d).$$

Remark **18** For $d = 1$ or $n = 1$, quasiconvexification reduces to convexification and formula (6.26) is equivalent to formula (6.13).

Theorem 42 *For $p > 1$, the energy densities $Q\mathcal{W}_{\eta,\epsilon}$ are standard energy densities satisfying* (6.25) *and $I_{\eta,\epsilon}$ $\Gamma(L^p)$- and $\Gamma(W^{1,p})$-converges to I_{hom} as ϵ and η go to* 0*. Therefore, for any sequence $u_{\eta,\epsilon}$ of minimizers of $\inf\{I_{\eta,\epsilon}(v) \,|\, v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$, there exists a minimizer u_{hom} of $\inf\{I_{hom}(v) \,|\, v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ such that*

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0} u_{\eta,\epsilon} = u_{hom} \quad \text{weakly in } W^{1,p}(\Omega, \mathbb{R}^d), \tag{6.27}$$

up to extraction.

Correspondingly, for any minimizer u_{hom} *of* $\inf\{I_{hom}(v) \,|\, v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$*, there exists a sequence* $u_{\eta,\epsilon}$ *of minimizers of* $\inf\{I_{\eta,\epsilon}(v) \,|\, v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ *such that*

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0} u_{\eta,\epsilon} = u_{hom} \quad \textit{strongly in } W^{1,p}(\Omega, \mathbb{R}^d). \tag{6.28}$$

In addition, if $u_{\eta,\epsilon}$ *is a family of equi-isolated minimizers in the sense of Definition* 23*, then* u_{hom} *is also isolated and* (6.28) *holds.*

Theorem 42 is abstract since quasiconvexification is not explicit in general and very hard to compute in practice. However, an alternative consists in considering the energies $\mathcal{W}_{\eta,\epsilon}$ on a finite dimensional space first and passing to the limit on the dimension in a second step, as stated in the following.

Theorem 43 *Let consider a set of finite dimensional subspaces of* $W^{1,p}(\Omega, \mathbb{R}^d)$ *and an associated equicontinuous family of projectors* (V_H, P_H)*, such that* $V_{H^1} \subset V_{H^2}$ *for* $0 \le H^2 \le H^1$*, and* $\overline{\cup_H V_H}^{1,p} = W^{1,p}(\Omega)$*. Let also introduce the following integral functionals on* $W^{1,p}(\Omega, \mathbb{R}^d)$*:*

$$I^H_{\eta,\epsilon}(u) = \int_\Omega \mathcal{W}_{\eta,\epsilon}(x, \nabla P_H u). \tag{6.29}$$

Under the assumptions of Theorem 42*,* $I^H_{\eta,\epsilon}$ $\Gamma(W^{1,p})$*-converges to* I_{hom} *as* ϵ*,* η*, and* H *go to* 0*.*

There exist minimizers $u^H_{\eta,\epsilon} \in V_H$ *of* $\inf\{I^H_{\eta,\epsilon}(v) \,|\, v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$*, and for any such sequence there exists a minimizer* u_{hom} *of* $\inf\{I_{hom}(v) \,|\, v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ *such that*

$$\lim_{H\to 0}\lim_{\eta\to 0}\lim_{\epsilon\to 0} u^H_{\eta,\epsilon} = u_{hom} \quad \textit{weakly in } W^{1,p}(\Omega, \mathbb{R}^d), \tag{6.30}$$

up to extraction.

Correspondingly, for any minimizer u_{hom} *of* $\inf\{I_{hom}(v) \,|\, v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$*, there exists a sequence* $u^H_{\eta,\epsilon}$ *of* $\inf\{I^H_{\eta,\epsilon}(v) \,|\, v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ *in* V_H *such that*

$$\lim_{H\to 0}\lim_{\eta\to 0}\lim_{\epsilon\to 0} u^H_{\eta,\epsilon} = u_{hom} \quad \textit{strongly in } W^{1,p}(\Omega, \mathbb{R}^d). \tag{6.31}$$

In addition, if $u^H_{\eta,\epsilon}$ *is a family of equi-isolated minimizers in an extended sense (for which* B *should be replaced by* $B \cap V_H$ *in Definition* 23*), then* (6.31) *holds.*

The idea behind Theorem 43 is that the compactness of minimizers is due to the finite dimension of the minimization space at finite ϵ and η and to quasiconvexity at the limit $H \to 0$.

Remark **19** If W is convex, then, for all $\xi \in \mathbb{R}^{n\times d}$,

$$\begin{aligned}\inf\left\{\int_{(0,1)^d} W(\xi + \nabla u), u \in W^{1,p}_{\#}((0,1)^d, \mathbb{R}^n)\right\} \\ = \inf\left\{\int_{(0,1)^d} W(\xi + \nabla u), u \in W^{1,p}((0,1)^d, \mathbb{R}^n), < \nabla u >_{(0,1)^d} = 0\right\}.\end{aligned}$$

This equality does not hold for quasiconvex energy densities, as can be easily seen on the polyconvex function $\xi \mapsto |\det(\xi) - 1|$ on $\mathbb{R}^{2\times 2}$ at point $\xi = 0$. Therefore the averaged energy density of Definition 22 is not a suitable definition in the quasiconvex case, for which periodic or Dirichlet boundary conditions have to be considered.

Remark **20** The convergence of the minimizers for quasiconvex energy densities in Theorem 42 is weak only in $W^{1,p}$, as opposed to the strictly convex case. This limitation is not technical, and the hypothesis of "equi-isolated minimizers" (which cannot be checked in practice) is in a way optimal. A counterexample is given by shear band instabilities for a homogenized energy which is not strictly rank-one convex, as considered in [94] and numerically investigated in Chaper 5.

Remark **21** The extension of Theorem 41 to the quasiconvex case is an issue of interest. However, in the quasiconvex case (as in the nonstrictly convex case), the minimization problem (6.9) does not characterize $v_{\eta,\epsilon}^{H,i}$ since the minimizers may be nonunique. This difficulty also arises in the classical theory of periodic homogenization of quasiconvex energy densities, for which there is still no corrector result.

Remark **22** Proposition 5 has no counterpart for quasiconvex energy densities since no simple structure (which could help deriving such estimates) is known for quasiconvexity.

6.2 Proof of the main results

For the sake of illustration, we first consider in section 6.2.1 the *one-dimensional linear* version of (6.4). In this simple case, all the computations may be performed analytically. We thus get some useful insight into the interest of $u_{\eta,\epsilon}$ and on the ingredients needed for the proof of Theorem 40, which is in turn performed in section 6.2.2 and adapted to the quasiconvex case in section 6.2.3. The proofs of Theorem 41 and Propositions 5 and 6 are then, respectively, the purposes of sections 6.2.4, 6.2.5, and 6.2.6.

6.2.1 The one-dimensional linear case

In the one-dimensional linear case, problem (6.4) reads

$$\inf\left\{\int_0^1 W_\epsilon(x,u'(x))dx, u\in H^1(0,1), u(0)=0, u(1)=1\right\}, \tag{6.32}$$

where $W_\epsilon(x,\xi)=\frac{1}{2}a_\epsilon(x)\xi^2$, and a_ϵ is a family of functions in $L^\infty(0,1)$ such that

$$0<c\le a_\epsilon(x)\le C<+\infty$$

for almost every $x\in\mathbb{R}$. To fix the ideas, we have made the boundary conditions specific in (6.32). This choice is arbitrary and plays no essential role.

The unique minimizer of (6.32) is

$$u_\epsilon(x)=C_\epsilon\int_0^x\frac{1}{a_\epsilon(y)}dy,$$

with $C_\epsilon=\left(\int_0^1\frac{1}{a_\epsilon(y)}\right)^{-1}$.

Extracting a subsequence, we may assume that $\frac{1}{a_\epsilon}$ weakly-* converges to some $\frac{1}{b^*}$ in L^∞. Consequently, C_ϵ converges to $C=\left(\int_0^1\frac{1}{b^*(x)}dx\right)^{-1}$ in $\mathbb{R}$. Thus u_ϵ weakly converges in $H^1(0,1)$ to $u_{hom}:x\mapsto C\int_0^x\frac{1}{b^*(y)}dy$.

In this one-dimensional case, the definition (6.6) of an averaged energy density reads

$$W_{\eta,\epsilon}(x,\xi)=\frac{1}{2\eta}\inf\left\{\int_{x-\eta}^{x+\eta}W_\epsilon(y,v'(y))dy, v\in H^1(x-\eta,x+\eta),\right.$$
$$\left.\frac{1}{2\eta}\int_{x-\eta}^{x+\eta}v'(y)dy=\xi\right\}.$$

Straightforward calculations give the explicit form

$$W_{\eta,\epsilon}(x,\xi)=\frac{1}{2}\left\langle\frac{1}{a_\epsilon(\cdot)}\right\rangle_{(x-\eta,x+\eta)}^{-1}\xi^2.$$

It follows that the minimizer of

$$\inf\left\{\int_0^1 W_{\eta,\epsilon}(x,u'(x))dx, u\in H^1(0,1), u(0)=0, u(1)=1\right\}$$

is the function

$$u_{\eta,\epsilon}(x)=C_{\eta,\epsilon}\int_0^x\left\langle\frac{1}{a_\epsilon(\cdot)}\right\rangle_{(y-\eta,y+\eta)}dy, \tag{6.33}$$

where $C_{\eta,\epsilon}=\left(\int_0^1\langle\frac{1}{a_\epsilon(\cdot)}\rangle_{z-\eta,z+\eta}dz\right)^{-1}$. Let $C_\eta=\left(\int_0^1\langle\frac{1}{b^*(\cdot)}\rangle_{(z-\eta,z+\eta)}dz\right)^{-1}$. When η is kept fixed, $u_{\eta,\epsilon}$ strongly converges in $H^1(0,1)$ (for the previous extraction in ϵ) to $u_{\eta,hom}(x)=C_\eta\int_0^x\langle\frac{1}{b^*(\cdot)}\rangle_{(y-\eta,y+\eta)}dy$ by the dominated convergence theorem. Let now η go to zero. For every Lebesgue point y of $\frac{1}{b^*}\in L^1(0,1)$ (thus almost everywhere on $(0,1)$), $\langle\frac{1}{b^*(\cdot)}\rangle_{(y-\eta,y+\eta)}\to\frac{1}{b^*(y)}$. The dominated convergence theorem then shows that $u_{\eta,hom}\to u_{hom}$ in $H^1(0,1)$. Consequently,

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0}u_{\eta,\epsilon}=u_{hom}\text{ in }H^1(0,1). \tag{6.34}$$

Since the convergence obtained is strong in $H^1(0,1)$, the energy $\int_0^1 W_{\eta,\epsilon}(x,u'_{\eta,\epsilon})$ also converges to $\int_0^1 W_{hom}(x,u'_{hom})$ in $\mathbb{R}$.

The convergence of $u_{\eta,\epsilon}$ to u_{hom} is strong, in contrast to that of u_ϵ.

Remark **23** Remark 11 can be advantageously related to formula (6.33), where we can see that the compactness of translations allows us to mix the limits in ϵ and η, as pointed out in [74] and illustrated below.

Let us make the above one-dimensional problem even more specific by considering the example of an operator of the form $a_\epsilon(x)=a(\frac{x}{\epsilon})$ where $a(\cdot)$ is 1-periodic on $\mathbb{R}$. In this case,

$$\left\langle\frac{1}{a_\epsilon(\cdot)}\right\rangle_{(z-\eta,z+\eta)}=\frac{1}{2\eta}\int_{z-\eta}^{z+\eta}\frac{1}{a(y/\epsilon)}dy=\frac{\epsilon}{2\eta}\int_{\frac{z-\eta}{\epsilon}}^{\frac{z+\eta}{\epsilon}}\frac{1}{a(y)}dy=\left\langle\frac{1}{a(\cdot)}\right\rangle_{(0,1)}+O\left(\frac{\epsilon}{\eta}\right).$$

This shows that $\lim_{\epsilon\to 0}u_{\eta(\epsilon),\epsilon}=u_{hom}$ in $H^1(0,1)$ if $\lim_{\epsilon\to 0}\eta(\epsilon)=0$ and $\lim_{\epsilon\to 0}\frac{\epsilon}{\eta(\epsilon)}=0$, which is more precise than (6.34). The same property holds for stochastic homogenization, for which we refer the reader to the preliminaries and section 4 of [74].

In the proof of Theorem 40 for the one-dimensional case, the main ingredient is the pointwise convergence of integrands due to the averaging of weakly converging functions on balls. Since the expression of the minimizer in (6.33) is analytical, the conclusion is achieved by using the dominated convergence theorem. This specific expression is linked to the dimension and to the linearity. When dealing with the multidimensional case, no such analytical formula holds for the minimizer. Following the line of the proof for the one-dimensional case, we can focus on the Green formula for the solution and use the abstract G-convergence theory. In the present work, however, we focus on the minimum instead of the minimizer and use Γ-convergence arguments to link the convergence of the energies to the convergence of the minimizers. This approach illustrates that in a way the modified energy of Definition 22 is a relaxed energy, in the spirit of the homogenization of multiple integrals dealt with in [38].

6.2.2 Proof of Theorem 40

The following three lemmata relate the pointwise convergence of energy densities to the Γ-convergence of the associated energy functionals.

Lemma 6.6 (see [56, Proposition 5.11]). *Let $X=\mathbb{R}^n$, $B_R=B(x_0,R)$ for $x_0\in X$ and $R>0$, and let $F:B_R\to\mathbb{R}$ be a* convex *function. Suppose that $\sup_{x\in B_R}F(x)=M<+\infty$. Consequently, $\inf_{x\in B_R}F(x)=m>-\infty$. Let $0<r<R$ and $K=(M-m)/(R-r)$. Then*

$$|F(x)-F(y)|\le K\|x-y\| \tag{6.35}$$

for every x, y in the closure $\bar{B}_r$ of B_r.

Lemma 6.6 is a classical result of convex analysis. The reader is referred to [56] for a proof. It is used in the appendix to prove the first of the following lemmata, which are crucial for what follows, and also stated and proved in [56].

Lemma 6.7 (see [56, Proposition 5.14]). *Let $\tilde{W}_\epsilon : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a set of Borel functions satisfying the growth condition*

$$0 \leq \tilde{W}_\epsilon(\cdot, \xi) \leq C(|\xi|^p + 1) \tag{6.36}$$

and such that for almost every $x \in \mathbb{R}^n$, $\tilde{W}_\epsilon(x, \cdot)$ is convex *on $\mathbb{R}^n$. Let us assume that there exists an open bounded subset ω of $\mathbb{R}^n$, such that for all $\xi \in \mathbb{R}^n$, $\tilde{W}_\epsilon(\cdot, \xi)$ converges pointwise almost everywhere on ω to a function $\tilde{W}(\cdot, \xi)$. If $\tilde{W}$ is a Borel function that satisfies* (6.36) *and for almost every $x \in \omega$, $\tilde{W}(x, \cdot)$ is* convex *on $\mathbb{R}^n$, then $\tilde{I}_\epsilon : u \mapsto \int_\omega \tilde{W}_\epsilon(x, \nabla u)dx$ $\Gamma(w - W^{1,p})$-converges to $\tilde{I} : u \mapsto \int_\omega \tilde{W}(x, \nabla u)dx$ on $W^{1,p}(\omega)$.*

Lemma 6.8 (see [56, Theorem 5.9]). *Let (X, d) be a metric space. Let (I_ϵ) be a sequence of functionals from X to $\mathbb{R}$. If (I_ϵ) is equilower semicontinuous on (X, d), then I_ϵ $\Gamma(d)$-converges to I in X if and only if I_ϵ converges to I pointwise in X.*

Lemma 6.8 uses the notion of Γ-convergence in metric spaces. Provided the right extension of Definition 20 (see, e.g., [36, section 1.4]), its proof is rather direct. We refer the reader to [56, Proposition 5.9] for details.

Finally, we recall a particular case of [56, Theorem 7.19] which relates the convergence of minimizers to the Γ-convergence for noncoercive functionals. To this aim, let us introduce the following notions.

Definition 30 *Let I be a functional from the metric space (X, d) to $\mathbb{R}$. We denote by $M(I)$ the possibly empty set of all the minimizers of I on X. Let now (I_ϵ) be a sequence of functionals from (X, d) to $\mathbb{R}$. We then denote by $K - \lim_{\epsilon \to 0} M(I_\epsilon)$ the possibly empty set of the limits of all the sequences of minimizers $u_\epsilon \in M(I_\epsilon)$ in (X, d).*

We refer the reader to [56, sections 4 and 7] for details on these notions and on the following.

Lemma 6.9 (see [56, Theorem 7.19]). *Assume that (I_ϵ) is a sequence of functionals which $\Gamma(d)$-converges on the metric space (X, d) to a functional I that is not identically $+\infty$. If $\lim_{\epsilon \to 0} \inf_X I_\epsilon = \inf_X I$ and the infima are attained on X (that is, $M(I_\epsilon) \neq \emptyset$), then $M(I) = K - \lim_{\epsilon \to 0} M(I_\epsilon)$. In particular, for any minimizer $u \in M(I)$, there exists a sequence $u_\epsilon \in M(I_\epsilon)$ such that $\lim_{\epsilon \to 0} u_\epsilon = u$ in (X, d).*

The sketch of the proof of Theorem 40 is the following. We study separately the limits in ϵ and in η. The arguments are, however, the same. First, using Lemma 6.7 and the pointwise convergence of the integrand, we prove the $\Gamma(L^p)$-convergence of the sequence of functionals. Since these functionals are equicoercive in the weak topology of $W^{1,p}(\Omega)$, Lemma 6.1 implies the convergence of the infima, and the existence and the weak convergence of the minimizers. In addition, the hypotheses of Lemma 6.8 are also satisfied on $(W^{1,p}(\Omega), \| \|_{W^{1,p}(\Omega)})$, and thus the sequence of functionals also $\Gamma(W^{1,p})$-converges to the same limit. Finally, we can apply Lemma 6.9 and deduce the strong convergence of the minimizers in the strictly convex case.

Limit $\epsilon \to 0$

Up to extracting a subsequence in ϵ (not relabeled), Lemma 6.2, applied to W_ϵ on the open bounded subset $\cup_{x \in \Omega} B(x, \eta)$, implies that there exists an energy density W_{hom} whose associated energy I_{hom} is the Γ-limit of I_ϵ on $W^{1,p}(\cup_{x \in \Omega} B(x, \eta))$. The locality of Γ-convergence (that is, the energy density of the Γ-limit does not depend on the domain of integration), the irrelevance of boundary conditions for the Γ-convergence (section 6.5.1 of the appendix) and the equicoercivity (Lemma 6.1) imply that

$$\begin{aligned}\lim_{\epsilon\to 0}\left(\inf\left\{\int_\omega W_\epsilon(y,\nabla v)\,|\,v\in W^{1,p}(\omega),\ \langle\nabla v\rangle_\omega=\xi\right\}\right)\\ =\inf\left\{\int_\omega W_{hom}(y,\nabla v)\,|\,v\in W^{1,p}(\omega),\ \langle\nabla v\rangle_\omega=\xi\right\}\end{aligned}\tag{6.37}$$

for every open subset $\omega\subset\cup_{x\in\Omega}B(x,\eta)$ and for all $\xi\in\mathbb{R}^n$. Equality (6.37) with $\omega=B(x,\eta)$ reads

$$\lim_{\epsilon\to 0}W_{\eta,\epsilon}(x,\xi)=W_{\eta,hom}(x,\xi),\tag{6.38}$$

for every $\xi\in\mathbb{R}^n$, where

$$\begin{aligned}W_{\eta,hom}(x,\xi)=\inf\big\{\langle W_{hom}(\cdot,\nabla v(\cdot))\rangle_{B(x,\eta)}\,|\\ v\in W^{1,p}(B(x,\eta)),\langle\nabla v(\cdot)\rangle_{B(x,\eta)}=\xi\big\}.\end{aligned}$$

The energy densities $W_{\eta,\epsilon}(\cdot,\xi)$ and $W_{\eta,hom}(\cdot,\xi)$ are measurable for all $\xi\in\mathbb{R}^n$ as the limits of the following measurable functions, $x\mapsto\int_{B(x,\eta)}W_\epsilon(y,\nabla u_n(y))$ and $x\mapsto\int_{B(x,\eta)}W_{hom}(y,\nabla v_n(y))$, where u_n and v_n are minimizing sequences of these integrals on the set $\{v\in W^{1,p}(\Omega)\,|\,\langle\nabla v\rangle_{B(x,\eta)}=\xi\}$. To prove that $W_{\eta,\epsilon}$ and $W_{\eta,hom}$ are Carathéodory functions, it remains to prove that for almost every $x\in\Omega$, $W_{\eta,\epsilon}(x,\cdot)$ and $W_{\eta,hom}(x,\cdot)$ are continuous on $\mathbb{R}^n$, which is actually a consequence of convexity (property $H2$).

We now show that the energy densities $W_{\eta,\epsilon}$ and $W_{\eta,hom}$ satisfy the properties $H2$ and $H3$. For almost every $x\in\Omega$ and for all $\xi\in\mathbb{R}^n$,

$$\begin{aligned}W_{\eta,\epsilon}(x,\xi)&=\inf\big\{\langle W_\epsilon(\cdot,\xi+\nabla v(\cdot))\rangle_{B(x,\eta)}\,|\,\langle\nabla v\rangle=0\big\}\\ &\le\langle W_\epsilon(\cdot,\xi)\rangle_{B(x,\eta)}\\ &\le C(1+|\xi|^p)\qquad\text{as } W_\epsilon \text{ satisfies } H3.\end{aligned}$$

Letting v_ξ denote a minimizer of $\inf\big\{\langle W_\epsilon(\cdot,\xi+\nabla v(\cdot))\rangle_{B(x,\eta)}\,|\,\langle\nabla v\rangle_{B(x,\eta)}=0\big\}$, we then have

$$\begin{aligned}W_{\eta,\epsilon}(x,\xi)&\ge c\,\langle|\xi+\nabla v_\xi(\cdot)|^p\rangle_{B(x,\eta)}\qquad\text{as } W_\epsilon \text{ satisfies } H3\\ &\ge c\inf\big\{\langle|\xi+\nabla v(\cdot)|^p\rangle_{B(x,\eta)}\,|\,\langle\nabla v\rangle=0\big\}\\ &=c\,|\xi|^p.\end{aligned}$$

Consequently, $W_{\eta,\epsilon}$ satisfies $H3$. The same calculations hold for $W_{\eta,hom}$. The convexity of $W_{\eta,\epsilon}(x,\cdot)$ and of $W_{\eta,hom}(x,\cdot)$ is a consequence of the following calculation. Let ξ, $\zeta\in\mathbb{R}^n$, $\lambda\in(0,1)$, and v_1, $v_2\in W^{1,p}(\Omega)$ be such that $\langle\nabla v_1\rangle=\langle\nabla v_2\rangle=0$. As W_ϵ is convex,

$$\begin{aligned}&\langle W_\epsilon(\cdot,\lambda(\xi+\nabla v_1(\cdot))+(1-\lambda)(\zeta+\nabla v_2(\cdot)))\rangle_{B(x,\eta)}\\ &\qquad\le\lambda\langle W_\epsilon(\cdot,\xi+\nabla v_1(\cdot))\rangle_{B(x,\eta)}+(1-\lambda)\langle W_\epsilon(\cdot,\zeta+\nabla v_2(\cdot))\rangle_{B(x,\eta)}.\end{aligned}$$

Let now $\tilde v_1$ and $\tilde v_2$ be, respectively, minimizers of the first and second terms of the right-hand side above. We have

$$\begin{aligned}&\langle W_\epsilon(\cdot,\lambda(\xi+\nabla\tilde v_1(\cdot))+(1-\lambda)(\zeta+\nabla\tilde v_2(\cdot)))\rangle_{B(x,\eta)}\\ &\qquad\le\lambda W_{\eta,\epsilon}(x,\xi)+(1-\lambda)W_{\eta,\epsilon}(x,\zeta).\end{aligned}$$

And consequently

$$W_{\eta,\epsilon}(x,\lambda\xi+(1-\lambda)\zeta)\le\lambda W_{\eta,\epsilon}(x,\xi)+(1-\lambda)W_{\eta,\epsilon}(x,\zeta).$$

It is worth noticing that if W_ϵ is strictly convex, then $W_{\eta,\epsilon}$ is also strictly convex.

Since $W_{\eta,\epsilon}$ and $W_{\eta,hom}$ satisfy $H1$, $H2$, and $H3$, we can use Lemma 6.7 and the pointwise convergence (6.38) to prove that

$$\Gamma(L^p)-\lim_{\epsilon\to 0}I_{\eta,\epsilon}=I_{\eta,hom}$$

on $W^{1,p}(\Omega)$, where $I_{\eta,hom} = u \mapsto \int_\Omega W_{\eta,hom}(x, \nabla u)$. Using Lemma 6.1, $H3$ also implies the convergence of the infima

$$\lim_{\epsilon\to 0} \inf \left\{ I_{\eta,\epsilon}(u), u \in W^{1,p}(\Omega) + BC \right\} = \inf \left\{ I_{\eta,hom}(u), u \in W^{1,p}(\Omega) + BC \right\}$$

and the weak convergence in $W^{1,p}(\Omega)$ of any corresponding sequence of minimizers $u_{\eta,\epsilon}$ to some minimizer $u_{\eta,hom}$, up to extraction.

In addition, by the application of the dominated convergence theorem, for every $u \in W^{1,p}(\Omega)$, $\int_\Omega W_{\eta,\epsilon}(x, \nabla u)$ converges to $\int_\Omega W_{\eta,hom}(x, \nabla u)$ as ϵ goes to 0. As $(W_{\eta,\epsilon})$ is convex and satisfies $H3$ for all ϵ and η, the associated energy functionals are equicontinuous on $W^{1,p}(\Omega)$. Thus Lemma 6.8 shows that $I_{\eta,\epsilon}$ $\Gamma(W^{1,p})$-converges to $I_{\eta,hom}$.

Lemma 6.9 shows that for every minimizer $u_{\eta,hom}$ there exists a sequence of minimizers $u_{\eta,\epsilon}$ such that $u_{\eta,\epsilon} \to u_{\eta,hom}$ in $W^{1,p}(\Omega)$. In addition, if W_ϵ is strictly convex, then $W_{\eta,\epsilon}$ is also strictly convex and there exists a unique sequence of minimizers $u_{\eta,\epsilon}$. As $M(I_{\eta,hom})$ is not empty, $(u_{\eta,\epsilon})$ strongly converges to some $u_{\eta,hom} \in M(I_{\eta,hom})$ (without other extraction) in $W^{1,p}(\Omega)$ as $\epsilon \to 0$, and the minimizer $u_{\eta,hom}$ is unique.

If a sequence of minimizers $u_{\eta,\epsilon}$ happens to be equi-isolated on $B \subset W^{1,p}(\Omega)$ in the sense of Definition 23, then the previous argument holds applying Lemma 6.9 on $(B, \|\cdot\|_{1,p})$ since the $\Gamma(W^{1,p})$-convergence on $W^{1,p}(\Omega)$ implies the $\Gamma(W^{1,p})$-convergence on $(B, \|\cdot\|_{1,p})$. We thus obtain the strong convergence of the sequence.

Limit $\eta \to 0$

For every $\xi \in \mathbb{R}^n$, let us first determine the pointwise limit on Ω as η goes to 0 of

$$\begin{aligned} W_{\eta,hom}(x,\xi) &= \inf \left\{ \langle W_{hom}(\cdot, \nabla v(\cdot)) \rangle_{B(x,\eta)} \mid v \in W^{1,p}(B(x,\eta)), \langle \nabla v \rangle_{B(x,\eta)} = \xi \right\} \\ &= \inf \Bigg\{ \int_{B(0,1)} W_{hom}(x + \eta y, \nabla v(y)), v \in W^{1,p}(B(0,1)), \\ &\qquad \langle \nabla v \rangle_{B(0,1)} = \xi \Bigg\} \cdot \frac{1}{\mathcal{L}^n(B(0,1))}. \end{aligned} \tag{6.39}$$

To this aim, let us denote by $\tilde{W}_x^\eta(y,\xi) = W_{hom}(x+\eta y, \xi)$ for almost every $x \in \Omega$, $y \in B(0,1)$, and $\xi \in \mathbb{R}^n$. Lemma 6.2 implies that the energy densities $\tilde{W}_x^\eta(\cdot,\cdot)$ and $W_{hom}(x,\cdot)$ satisfy $H1$, $H2$, and $H3$. In addition, for all $\xi \in \mathbb{R}^n$, every Lebesgue point $x \in \Omega$ (and consequently almost everywhere on Ω) of $W_{hom}(\cdot,\xi) \in L^1(\Omega, \mathbb{R})$, and almost every $y \in B(0,1)$,

$$\lim_{\eta\to 0} W_{hom}(x + \eta y, \xi) = W_{hom}(x, \xi).$$

We now apply Lemma 6.7 to obtain the $\Gamma(L^p)$-convergence of the associated integral functionals. Property $H3$ then implies the convergence of the infima (6.39) by the application of Lemma 6.1. For every $\xi \in \mathbb{R}^n$, this proves the following pointwise convergence almost everywhere on Ω:

$$\lim_{\eta\to 0} W_{\eta,hom}(x,\xi) = W_{hom}(x,\xi). \tag{6.40}$$

Lemma 6.2 and the same arguments as for $W_{\eta,\epsilon}$ show that $W_{\eta,hom}$ and W_{hom} also satisfy $H1$, $H2$, and $H3$. By the application of Lemma 6.7, the pointwise convergence (6.40) implies the $\Gamma(L^p)$-convergence of $I_{\eta,hom}$ to I_{hom} on $W^{1,p}(\Omega)$ as η goes to 0. Consequently, minimizers $u_{\eta,hom}$ converge weakly in $W^{1,p}(\Omega)$ to some minimizer u_{hom}, up to extraction, by Lemma 6.1. Using Lemmata 6.8 and 6.9 as for the limit $\epsilon \to 0$, we obtain the three last statements of Theorem 40.

6.2.3 Proof of Theorem 42

The proof of Theorem 42 follows exactly along the lines of the proof of Theorem 40. We indeed have the following.

Remark **24** The conclusion of Lemma 6.7 is unchanged if we replace the hypothesis of convexity on $\mathbb{R}^n$ by the hypothesis of quasiconvexity on $\mathbb{R}^{n\times d}$.

This version of Lemma 6.7 is more general and is the one proved in the appendix. For the sake of clarity, we will still refer to Lemma 6.7, even in the quasiconvex setting, for which it should be understood in the sense of Remark 24.

In contrast to the proof of Theorem 40, the hypotheses of Lemmata 6.7 and 6.8 are more technical to check in the quasiconvex case, especially the pointwise convergence of the integrands. To do that, we will make use of the following results of the calculus of variations: a characterization of quasiconvex hulls and a decomposition lemma.

Lemma 6.10 (see [2, III.7]). *Let f be a Carathéodory functional satisfying the growth condition* (6.25) *for $p > 1$ on $\mathbb{R}^n \times \mathbb{R}^{n\times d}$. Let Qf denote the quasiconvex envelope of f. Then for all Ω open bounded subset of $\mathbb{R}^n$ and for all $u \in W^{1,p}(\Omega, \mathbb{R}^d)$, there exists a sequence $\{\phi_k\}_k \in W^{1,p}(\Omega, \mathbb{R}^d)$ such that $\phi_k \rightharpoonup u$ in $W^{1,p}(\Omega, \mathbb{R}^d)$ and*

$$\int_\Omega Qf(x, \nabla u) = \lim_{k\to\infty} \int_\Omega f(x, \nabla\phi_k).$$

Lemma 6.11 (see [86]). *Let $p > 1$ and assume that $\partial\Omega$ is Lipschitz. Let $u_k \rightharpoonup v_0$ in $W^{1,p}(\Omega, \mathbb{R}^d)$. Then there exist a subsequence u_{k_l} of u_k and a sequence $v_l \in W^{1,\infty}(\mathbb{R}^n, \mathbb{R}^d)$ such that*

(i) *$v_l \rightharpoonup v_0$ in $W^{1,p}(\Omega, \mathbb{R}^d)$,*
(ii) *$v_l = v_0$ in a neighborhood of $\partial\Omega$,*
(iii) *$\{\nabla v_l\}_l$ is p-equi-integrable, that is, for all $\rho > 0$, there exists $\delta > 0$ such that for all measurable subset $A \subset \Omega$, $\sup_{l\in\mathbb{N}} \int_A |\nabla v_l|^p dx < \rho$ whenever $\mathcal{L}^n(A) < \delta$,*
(iv) *$\lim_{l\to\infty} \mathcal{L}^n(\{x \in \Omega : v_l(x) \neq u_{k_l}(x)$ or $\nabla v_l(x) \neq \nabla u_{k_l}(x)\}) = 0$.*

We finally introduce a lemma that relates the pointwise convergence of Lipschitz Carathéodory functions to the pointwise convergence of their quasiconvex envelopes. The proof of this lemma is postponed until the appendix.

Lemma 6.12. *Let f and f_ϵ be Carathéodory functions satisfying* (6.25) *on $\mathbb{R}^n \times \mathbb{R}^{n\times d}$. We assume that for all $R > 0$, there exists $K > 0$ such that for almost every $x \in \mathbb{R}^n$ and for all $\epsilon > 0$, $f_\epsilon(x, \cdot)$ and $f(x, \cdot)$ are K-Lipschitz on $B(0, R) = \{\xi \in \mathbb{R}^{n\times d} : |\xi| \leq R\}$. If $f_\epsilon(x, \xi)$ converges to $f(x, \xi)$ for almost every $x \in \mathcal{O}$ (open bounded subset of $\mathbb{R}^n$ with Lipschitz boundary $\partial\mathcal{O}$) and for all $\xi \in \mathbb{R}^{n\times d}$, then $Qf_\epsilon(x, \xi)$ converges to $Qf(x, \xi)$ for almost every $x \in \mathcal{O}$ and for all $\xi \in \mathbb{R}^{n\times d}$.*

We are now in position to prove Theorems 42 and 43. We treat only the limit $\epsilon \to 0$, the one for η being essentially the same as for ϵ provided the same adaptations as in section 6.2.2.

Proof of Theorem 42. We first introduce the following energy density defined for almost every $x \in \mathbb{R}^n$ and for all $\xi \in \mathbb{R}^{n\times d}$:

$$\mathcal{W}_{\eta,hom}(x, \xi) = \inf\Big\{\langle W_{hom}(y, \xi + \nabla v)\rangle_{C(x,\eta)} \,|\, v \in W^{1,p}_{\#}(C(x, \eta), \mathbb{R}^d)\Big\}. \tag{6.41}$$

Arguing as for the previous section, we can prove that $\mathcal{W}_{\eta,\epsilon}$ and $\mathcal{W}_{\eta,hom}$ satisfy $H2$ and $H3$. Thus $Q\mathcal{W}_{\eta,\epsilon}$ and $Q\mathcal{W}_{\eta,hom}$ are standard energy densities satisfying (6.25) with the same coefficients as for W_ϵ. Provided that

$$\lim_{\epsilon\to 0} Q\mathcal{W}_{\eta,\epsilon}(x, \xi) = Q\mathcal{W}_{\eta,hom}(x, \xi) \tag{6.42}$$

for almost every $x \in \mathbb{R}^n$ and for all $\xi \in \mathbb{R}^{n\times d}$, Lemma 6.7 implies the $\Gamma(L^p)$-convergence of $I_{\eta,\epsilon}$ to $I_{\eta,hom}$ on $W^{1,p}(\Omega, \mathbb{R}^d)$, defined by

$$I_{\eta,hom}(u) = \int_\Omega Q\mathcal{W}_{\eta,\mathrm{hom}}(x, \nabla u).$$

Let us prove the pointwise convergence (6.42). The energy densities $\mathcal{W}_{\eta,\epsilon}$ and $\mathcal{W}_{\eta,hom}$ actually satisfy the hypotheses of Lemma 6.12.

Arguing as in section 6.2.2, we have that

- $\lim_{\epsilon\to 0} \mathcal{W}_{\eta,\epsilon}(x,\xi) = \mathcal{W}_{\eta,hom}(x,\xi)$ for almost every $x \in \mathbb{R}^n$ and for all $\xi \in \mathbb{R}^{n\times d}$,
- $\mathcal{W}_{\eta,\epsilon}$ and $\mathcal{W}_{\eta,hom}$ satisfy (6.25).

It remains to prove that $\mathcal{W}_{\eta,\epsilon}(x,\cdot)$ and $\mathcal{W}_{\eta,hom}(x,\cdot)$ are equilocally Lipschitz (and thus continuous) on $\mathbb{R}^{n\times d}$ to fulfill the assumptions of Lemma 6.12.

Let us recall that standard energy densities W are locally Lipschitz in the following sense: there exists $\mathcal{C} > 0$ depending only on p, c, and C in (6.25) such that for almost every $x \in \mathbb{R}^n$ and for all $\xi_1, \xi_2 \in \mathbb{R}^{n\times d}$,

$$|W(x,\xi_1) - W(x,\xi_2)| \le \mathcal{C}(1 + |\xi_1|^{p-1} + |\xi_2|^{p-1})|\xi_1 - \xi_2|. \tag{6.43}$$

Convex functions such that (6.25) holds satisfy inequality (6.43) using Lemma 6.6. It can be proved for rank-one convex functions (and thus for quasiconvex functions) by introducing a decomposition of $\xi_1 - \xi_2$ in a sum of rank-one matrices.

For every fixed $x \in \mathbb{R}^n$, let u_{ξ_1} and u_{ξ_2} be minimizers of (6.26) for $\xi = \xi_1$ and ξ_2, respectively. The following four inequalities hold:

$$\begin{aligned}
&\mathcal{W}_{\eta,\epsilon}(x,\xi_1) - \frac{1}{\eta^n}\int_{C(x,\eta)} W_\epsilon(y, \nabla u_{\xi_2} + \xi_1) \le 0, \qquad (6.44)\\
&\mathcal{W}_{\eta,\epsilon}(x,\xi_2) - \frac{1}{\eta^n}\int_{C(x,\eta)} W_\epsilon(y, \nabla u_{\xi_1} + \xi_2) \le 0,\\
&\left|\mathcal{W}_{\eta,\epsilon}(x,\xi_1) - \frac{1}{\eta^n}\int_{C(x,\eta)} W_\epsilon(y, \nabla u_{\xi_1} + \xi_2)\right| \le C(1 + |\xi_1|^{p-1} + |\xi_2|^{p-1})|\xi_1 - \xi_2|,\\
&\left|\mathcal{W}_{\eta,\epsilon}(x,\xi_2) - \frac{1}{\eta^n}\int_{C(x,\eta)} W_\epsilon(y, \nabla u_{\xi_2} + \xi_1)\right| \le C(1 + |\xi_1|^{p-1} + |\xi_2|^{p-1})|\xi_1 - \xi_2|.
\end{aligned}$$

The first two inequalities of (6.44) are direct consequences of the definitions of u_{ξ_1} and u_{ξ_2}. The last two inequalities are obtained by integrating (6.43) over $C(x,\eta)$ with $W = W_\epsilon$ and noting that (6.25) implies $\int_{C(x,\eta)} |\nabla u_{\xi_1}|^p \le p(C+1)(1+|\xi_1|^p)\eta^n$ by the triangle inequality (see the proof of (6.61) in section 6.2.4 for details). The combination of these four inequalities shows that $\mathcal{W}_{\eta,\epsilon}$ satisfies (6.43) for some $\mathcal{C} > 0$ which is independent of ϵ and η. The same result and proof hold for $\mathcal{W}_{\eta,hom}$.

Lemma 6.12 then implies the pointwise convergence (6.42), and we obtain the $\Gamma(L^p)$-convergence of $I_{\eta,\epsilon}$ to $I_{\eta,hom}$ by applying Lemma 6.7. The functionals $Q\mathcal{W}_{\eta,\epsilon}$ and $Q\mathcal{W}_{\eta,hom}$ are equilocally bounded by (6.25), and thus equilocally Lipschitz and equicontinuous on $W^{1,p}(\Omega,\mathbb{R}^d)$. Therefore, as in section 6.2.2, Lemma 6.8 proves the $\Gamma(W^{1,p})$-convergence of these energies. From Lemma 6.9 we then deduce the existence of strong converging sequences of minimizers, and the strong convergence of sequences of equi-isolated minimizers if they exist.

Proof of Theorem 43. The proof of Theorem 43 is based on two arguments. At fixed H, the energies are equilocally Lipschitz because of the equicontinuity of the projections P_H. Therefore the pointwise convergence on $W^{1,p}(\Omega)$ of $I^H_{\eta,\epsilon}$ to I^H_{hom}, which is a consequence of the dominated convergence theorem, implies the $\Gamma(W^{1,p})$-convergence of the energy functionals at fixed H when ϵ and η go to 0.

In addition, the infimum $\inf\{I^H_{\eta,\epsilon}(v), v \in W^{1,p}(\Omega,\mathbb{R}^d) + BC\}$ is attained on V_H. It suffices indeed to consider the projection of any minimizing sequence in $W^{1,p}(\Omega,\mathbb{R}^d)$ on V_H. This new sequence is still a minimizing sequence since the energy is not changed by the projection and it is bounded in the finite dimensional space V_H. Therefore it converges, up to extraction, in $(V_H, ||\cdot||_{1,p})$. The continuity of the energy (it is locally Lipschitz) then allows us to pass to the limit and proves the existence of minimizers $u^H_{\eta,\epsilon} \in V_H$ of $\inf\{I^H_{\eta,\epsilon}(v), v \in W^{1,p}(\Omega,\mathbb{R}^d) + BC\}$. A direct argument also shows the convergence of the infima

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0} \inf\{I^H_{\eta,\epsilon}(v), v \in W^{1,p}(\Omega,\mathbb{R}^d) + BC\} = \inf\{I^H_{hom}(v), v \in W^{1,p}(\Omega,\mathbb{R}^d) + BC\}.$$

Let us detail this argument by considering a sequence $u^H_{\eta,\epsilon} \in V_H$ of minimizers of

$$\inf\{I^H_{\eta,\epsilon}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}.$$

This sequence is bounded in $W^{1,p}(\Omega, \mathbb{R}^d)$ using the growth condition (6.25) and thus compact in $W^{1,p}(\Omega)$ since V_H is finite dimensional. Therefore there exists $u^H_{hom} \in V_H$ such that $\lim_{\eta\to 0} \lim_{\epsilon\to 0} u^H_{\eta,\epsilon} = u^H_{hom}$ in $W^{1,p}(\Omega)$. We then have

$$|I^H_{\eta,\epsilon}(u^H_{\eta,\epsilon}) - I^H_{hom}(u^H_{hom})| \le |I^H_{\eta,\epsilon}(u^H_{\eta,\epsilon}) - I^H_{\eta,\epsilon}(u^H_{hom})| + |I^H_{\eta,\epsilon}(u^H_{hom}) - I^H_{hom}(u^H_{hom})|.$$

The second term of the right-hand side goes to 0 thanks to the pointwise convergence of the energy as ϵ and η go to 0, whereas the first term goes to 0 thanks to the equilocal Lipschitz property of $I^H_{\eta,\epsilon}$:

$$|I^H_{\eta,\epsilon}(u^H_{\eta,\epsilon}) - I^H_{\eta,\epsilon}(u^H_{hom})| \le C(1 + \|\nabla u^H_{\eta,\epsilon}\|_p^{p-1} + \|\nabla u^H_{hom}\|_p^{p-1})\|\nabla u^H_{\eta,\epsilon} - \nabla u^H_{hom}\|_p,$$

which also vanishes since $u^H_{\eta,\epsilon}$ converges to u^H_{hom} in $W^{1,p}(\Omega, \mathbb{R}^d)$. It remains to prove that u^H_{hom} is a minimizer of $\inf\{I^H_{hom}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$. For all $v \in W^{1,p}(\Omega, \mathbb{R}^d)$, for all η and ϵ, $I^H_{\eta,\epsilon}(u^H_{\eta,\epsilon}) \le I^H_{\eta,\epsilon}(v)$. At the limit, we obtain $I^H_{hom}(u^H_{hom}) \le I^H_{hom}(v)$, which proves the statement since v is arbitrary.

At fixed H, we have proved the $\Gamma(W^{1,p})$-convergence, when ϵ and η go to 0, of $I^H_{\eta,\epsilon}$ to I^H_{hom} on $W^{1,p}(\Omega, \mathbb{R}^d)$, and the existence and the strong convergence of any sequence of minimizers $u^H_{\eta,\epsilon} \in V_H$ of $\inf\{I^H_{\eta,\epsilon}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ to some minimizer $u^H_{hom} \in V_H$ of $\inf\{I^H_{hom}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$, up to extraction.

The limit $H \to 0$ can be dealt with as follows. The sequence u^H_{hom} is bounded in $W^{1,p}(\Omega)$ using the growth condition (6.25). Let us extract a subsequence (not relabeled) weakly converging to some $u_{hom} \in W^{1,p}(\Omega)$, and prove that u_{hom} is a minimizer of $\inf\{I_{hom}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ and that $I^H_{hom}(u^H_{hom}) \to I_{hom}(u_{hom})$. The quasiconvexity of W_{hom} and (6.25) imply the lower semicontinuity of I_{hom} for the weak topology of $W^{1,p}(\Omega)$, which shows

$$I_{hom}(u_{hom}) \le \lim_{H\to 0} I_{hom}(u^H_{hom}).$$

Next, as a consequence of the continuity of I_{hom} for the strong topology of $W^{1,p}(\Omega)$, for all $v \in W^{1,p}(\Omega) + BC$ and for all $\rho > 0$ there exist $H > 0$ and $v_H \in V_H + BC$ such that $I_{hom}(v) \ge I_{hom}(v_H) - \rho \ge I_{hom}(u^H_{hom}) - \rho \ge \lim_{H\to 0} I_{hom}(u^H_{hom}) - \rho$, which exists as a decreasing sequence of positive numbers. Thus u_{hom} is a minimizer of I_{hom} on $W^{1,p}(\Omega) + BC$.

The convergence of the infima is also obtained by this continuity argument. The pointwise convergence of the sequence I^H_{hom} and its equilocal Lipschitz property allow us to use Lemma 6.8 and show the $\Gamma(W^{1,p})$-convergence of I^H_{hom} to I_{hom} on $W^{1,p}(\Omega)$. Using the convergence of the infima, we can then apply Lemma 6.9 and prove that, with obvious notation, for any minimizer $u_{hom} \in W^{1,p}(\Omega)$ there exists a sequence of minimizers $u^H_{hom} \in W^{1,p}(\Omega)$ such that $\lim_{H\to 0} u^H_{hom} = u_{hom}$ in $W^{1,p}(\Omega)$. These minimizers may not belong to any V_H. Moreover, due to the projection P_H in the energy, the minimizers of $\inf\{I^H_{hom}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ are never isolated; therefore the argument used to prove the strong convergence in Theorem 42 has to be slightly modified to apply here.

To this aim, let us consider a sequence of minimizers $u^H_{hom} \in V_H$ which is equi-isolated in the following extended sense: there exists a ball $B \subset W^{1,p}(\Omega)$ such that u^H_{hom} is the unique minimizer of $\inf\{I^H_{hom}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ on $B \cap V_H$. Let u_{hom} be a weak limit of this sequence. Because of the weakly lower semicontinuity of the norm, $u_{hom} \in B$. In addition, there exists another sequence $v^H_{hom} \in W^{1,p}(\Omega)$ of minimizers of $\inf\{I^H_{hom}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ such that $v^H_{hom} \to u_{hom}$ in $W^{1,p}(\Omega)$. Because of this strong convergence, it is not restrictive to suppose $v^H_{hom} \in B$, and $P_H(v^H_{hom}) \in B$ as well. Since $P_H(v^H_{hom}) \in B$ is a minimizer of $\inf\{I^H_{hom}(v), v \in W^{1,p}(\Omega, \mathbb{R}^d) + BC\}$ and the sequence u^H_{hom} is equi-isolated in the sense above, $u^H_{hom} = P_H(v^H_{hom})$.

By the triangle inequality we then have $\|u^H_{hom} - u_{hom}\|_{1,p} \le \|P_H(v^H_{hom} - u_{hom})\|_{1,p} + \|P_H(u_{hom}) - u_{hom}\|_{1,p}$. The first term of the right-hand side goes to 0 because of the equicontinuity of the family

of projections (P_H) and the strong convergence of v^H_{hom} to u_{hom}, whereas the second term goes to 0 by definition of the family of spaces V_H. This shows the strong convergence of u^H_{hom} and concludes the proof of the theorem.

6.2.4 Proof of Theorem 41

The proof of Theorem 41 extensively uses the following consequences of the properties of equicontinuity (6.2) and equimonotonicity (6.3) of the elliptic operators.

Lemma 6.13. *We keep the notation of Lemma* 6.3*. Let B be a bounded open subset of $\mathbb{R}^n$. For almost all $x \in B$ and for all $\xi_1, \xi_2 \in \mathbb{R}^n$, property* (6.2) *implies*

$$|W_\epsilon(x,\xi_1) - W_\epsilon(x,\xi_2)| \le C|\xi_1 - \xi_2|(1 + |\xi_1|^{p-1} + |\xi_2|^{p-1}). \tag{6.45}$$

For all convex subset $K \subset W^{1,p}(B)$, if u minimizes $\inf\left\{\int_B W_\epsilon(x,\nabla v) \,|\, v \in K\right\}$*, then property* (6.3) *implies*

$$\left(\mathcal{L}^n(B)^{\frac{1}{p}} + \|u\|_{1,p} + \|v\|_{1,p}\right)^{\frac{p(\beta\wedge p - p)}{\beta\wedge p}} \left|\int_B (W_\epsilon(\nabla u) - W_\epsilon(\nabla v))\right|^{\frac{p}{\beta\wedge p}} \ge c\|\nabla u - \nabla v\|^p_{0,p} \tag{6.46}$$

for all $v \in K$ and $\beta > 0$.

The proof of Lemma 6.13 is classical (see [169]) and postponed until the appendix.

Let us introduce

$$u_\epsilon = \text{Argmin}\left\{\int_\Omega (W_\epsilon(x,\nabla u) - fu) \,|\, u \in W^{1,p}(\Omega) + BC\right\}, \tag{6.47}$$

$$u_{hom} = \text{Argmin}\left\{\int_\Omega (W_{hom}(x,\nabla u) - fu) \,|\, u \in W^{1,p}(\Omega) + BC\right\}, \tag{6.48}$$

$$u_{\eta,\epsilon} = \text{Argmin}\left\{\int_\Omega (W_{\eta,\epsilon}(x,\nabla u) - fu) \,|\, u \in W^{1,p}(\Omega) + BC\right\}, \tag{6.49}$$

$$u_{\eta,hom} = \text{Argmin}\left\{\int_\Omega (W_{\eta,hom}(x,\nabla u) - fu) \,|\, u \in W^{1,p}(\Omega) + BC\right\}, \tag{6.50}$$

$$u^{H,i}_\epsilon = \text{Argmin}\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla u) \,|\, \langle\nabla u\rangle_{Q_{H,i}} = \langle\nabla u_\epsilon\rangle_{Q_{H,i}}\right\}, \tag{6.51}$$

$$v^{H,i}_{\eta,\epsilon} = \text{Argmin}\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla u) \,|\, \langle\nabla u\rangle_{Q_{H,i}} = \langle\nabla u_{\eta,\epsilon}\rangle_{Q_{H,i}}\right\}, \tag{6.52}$$

$$u^{H,i}_{hom} = \text{Argmin}\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla u) \,|\, \langle\nabla u\rangle_{Q_{H,i}} = \langle\nabla u_{hom}\rangle_{Q_{H,i}}\right\}. \tag{6.53}$$

Let us denote by I_H the cardinal of the partition of Ω. We have

$$\begin{aligned}\int_\Omega \left|\nabla u_\epsilon - \sum_{i=1}^{I_H} \nabla v^{H,i}_{\eta,\epsilon} 1_{Q_{H,i}}\right|^p &= \sum_i^{I_H} \int_{Q_{H,i}} \left|\nabla u_\epsilon - \nabla v^{H,i}_{\eta,\epsilon}\right|^p \\ &\le p\sum_{i=1}^{I_H}\left(\int_{Q_{H,i}} \left|\nabla u_\epsilon - \nabla u^{H,i}_\epsilon\right|^p + \int_{Q_{H,i}} \left|\nabla u^{H,i}_\epsilon - \nabla v^{H,i}_{\eta,\epsilon}\right|^p\right) \\ &= p\sum_{i=1}^{I_H}(A^{\epsilon,H}_i + B^{\epsilon,H,\eta}_i).\end{aligned} \tag{6.54}$$

We now separately examine the limit $(H, \eta, \epsilon) \to 0$ in the two sums $\sum_{i=1}^{I_H} A_i^{\epsilon,H}$ and $\sum_{i=1}^{I_H} B_i^{\epsilon,H,\eta}$. We show that both terms vanish in the limit in the following sense:

$$\lim_{H\to 0} \limsup_{\epsilon\to 0} \sum_{i=1}^{I_H} A_i^{\epsilon,H} = 0, \tag{6.55}$$

$$\lim_{\eta\to 0} \limsup_{\epsilon\to 0} \sum_{i=1}^{I_H} B_i^{\epsilon,H,\eta} = 0 \quad \text{uniformly in } H. \tag{6.56}$$

Limit of $\sum_{i=1}^{I_H} A_i^{\epsilon,H}$

Since $u_\epsilon^{H,i}$ is defined as a minimizer on the convex set

$$K = \left\{ u \in W^{1,p}(Q_{H,i}) \,|\, \langle \nabla u \rangle_{Q_{H,i}} = \langle \nabla u_\epsilon \rangle_{Q_{H,i}} \right\},$$

and $\beta \le p$ by assumption, property (6.46) implies

$$A_i^{\epsilon,H} \le \frac{1}{c} \left| \int_{Q_{H,i}} W_\epsilon(x, \nabla u_\epsilon) - W_\epsilon(x, \nabla u_\epsilon^{H,i}) \right|. \tag{6.57}$$

In view of Lemmata 6.1 and 6.2,

$$\lim_{\epsilon\to 0} \int_{Q_{H,i}} W_\epsilon(x, \nabla u_\epsilon) = \int_{Q_{H,i}} W_{hom}(x, \nabla u_{hom}). \tag{6.58}$$

Actually, the locality of Γ-convergence and the liminf inequality imply

$$\lim_{\epsilon\to 0} \int_{\Omega - Q_{H,i}} W_\epsilon(x, \nabla u_\epsilon) \ge \int_{\Omega - Q_{H,i}} W_{hom}(x, \nabla u_{hom}).$$

Thus, $\lim_{\epsilon\to 0} \int_{Q_{H,i}} W_\epsilon(x, \nabla u_\epsilon) > \int_{Q_{H,i}} W_{hom}(x, \nabla u_{hom})$ would contradict the convergence of the infima $\lim_{\epsilon\to 0} \int_\Omega W_\epsilon(x, \nabla u_\epsilon) = \int_\Omega W_{hom}(x, \nabla u_{hom})$, which shows (6.58).

Let us prove that

$$\lim_{\epsilon\to 0} \int_{Q_{H,i}} W_\epsilon(x, \nabla u_\epsilon^{H,i}) = \inf \left\{ \int_{Q_{H,i}} W_{hom}(x, \nabla u) \,|\, \langle \nabla u \rangle_{Q_{H,i}} = \langle \nabla u_{hom} \rangle_{Q_{H,i}} \right\}. \tag{6.59}$$

We have $\langle \nabla u_\epsilon \rangle_{Q_{H,i}} = \langle \nabla u_{hom} \rangle_{Q_{H,i}} + \chi_{H,i}(\epsilon)$, where $\chi_{H,i}$ is a function satisfying $\lim_{\epsilon\to 0} |\chi_{H,i}(\epsilon)| = 0$ since ∇u_ϵ weakly converges to ∇u_{hom} in $L^p(\Omega)$.

In view of (6.51), we have

$$\int_{Q_{H,i}} W_\epsilon(x, \nabla u_{hom}^{H,i} + \chi_{H,i}(\epsilon)) \ge \int_{Q_{H,i}} W_\epsilon(x, \nabla u_\epsilon^{H,i}) \tag{6.60}$$

since $\langle \nabla u_{hom} + \chi_{H,i}(\epsilon) \rangle_{Q_{H,i}} = \langle \nabla u_\epsilon \rangle_{Q_{H,i}}$.

Next we show

$$\begin{aligned} &\left| \int_{Q_{H,i}} W_\epsilon(x, \nabla u_{hom}^{H,i} + \chi_{H,i}(\epsilon)) - W_\epsilon(x, \nabla u_{hom}^{H,i}) \right| \\ &\qquad \le \mathcal{C} |Q_{H,i}| |\chi_{H,i}(\epsilon)| (1 + |\langle \nabla u_{hom} \rangle_{Q_{H,i}}|^{p-1} + |\chi_{H,i}(\epsilon)|^{p-1}). \end{aligned} \tag{6.61}$$

To this end, let us start from (6.45):

$$
\begin{aligned}
&|W_\epsilon(x, \nabla u^{H,i}_{hom} + \chi_{H,i}(\epsilon)) - W_\epsilon(x, \nabla u^{H,i}_{hom})| \\
&\quad \le \mathcal{C}|\chi_{H,i}(\epsilon)|(1 + |\nabla u^{H,i}_{hom} + \chi_{H,i}(\epsilon)|^{p-1} + |\nabla u^{H,i}_{hom}|^{p-1}) \\
&\quad \le \mathcal{C}|\chi_{H,i}(\epsilon)|(1 + |\chi_{H,i}(\epsilon)|^{p-1} + |\nabla u^{H,i}_{hom}|^{p-1}),
\end{aligned}
$$

where $\mathcal{C}$ denotes various constants depending only on c and C in $H3$. Integrating on $Q_{H,i}$ yields

$$
\begin{aligned}
&\int_{Q_{H,i}} |W_\epsilon(x, \nabla u^{H,i}_{hom} + \chi_{H,i}(\epsilon)) - W_\epsilon(x, \nabla u^{H,i}_{hom})| \\
&\quad \le \int_{Q_{H,i}} \mathcal{C}|\chi_{H,i}(\epsilon)|(1 + |\chi_{H,i}(\epsilon)|^{p-1} + |\nabla u^{H,i}_{hom}|^{p-1}) \\
&\quad \le \mathcal{C}|\chi_{H,i}(\epsilon)| \left(|Q_{H,i}|(1 + |\chi_{H,i}(\epsilon)|^{p-1}) + \int_{Q_{H,i}} |\nabla u^{H,i}_{hom}|^{p-1} \right).
\end{aligned} \tag{6.62}
$$

Successively using Hölder's inequality, the minoration in $H3$, the definition (6.53), and the majoration in $H3$, we have

$$
\begin{aligned}
\int_{Q_{H,i}} |\nabla u^{H,i}_{hom}|^{p-1} &\le \left(\int_{Q_{H,i}} |\nabla u^{H,i}_{hom}|^p \right)^{\frac{p-1}{p}} \left(\int_{Q_{H,i}} 1 \right)^{\frac{1}{p}} \\
&= |Q_{H,i}|^{\frac{1}{p}} \left(\int_{Q_{H,i}} |\nabla u^{H,i}_{hom}|^p \right)^{\frac{p-1}{p}} \\
&\le \mathcal{C}|Q_{H,i}|^{\frac{1}{p}} \left(\int_{Q_{H,i}} W_\epsilon(x, \nabla u^{H,i}_{hom}(x)) \right)^{\frac{p-1}{p}} \\
&\le \mathcal{C}|Q_{H,i}|^{\frac{1}{p}} \left(\int_{Q_{H,i}} W_\epsilon(x, \langle \nabla u^{H,i}_{hom} \rangle_{Q_{H,i}}) \right)^{\frac{p-1}{p}} \\
&\le \mathcal{C}|Q_{H,i}|^{\frac{1}{p}} \left(|Q_{H,i}|(1 + |\langle \nabla u^{H,i}_{hom} \rangle_{Q_{H,i}}|^p) \right)^{\frac{p-1}{p}} \\
&\le \mathcal{C}|Q_{H,i}| \left(1 + |\langle \nabla u^{H,i}_{hom} \rangle_{Q_{H,i}}|^{p-1} \right).
\end{aligned} \tag{6.63}
$$

Inserting (6.63) into (6.62) shows (6.61).

Then we remark, in view of (6.53), that

$$
\int_{Q_{H,i}} W_\epsilon(x, \nabla u^{H,i}_{hom}) \le \int_{Q_{H,i}} W_\epsilon(x, \nabla u^{H,i}_\epsilon - \chi_{H,i}(\epsilon)) \tag{6.64}
$$

since $\langle \nabla u_\epsilon - \chi_{H,i}(\epsilon) \rangle_{Q_{H,i}} = \langle \nabla u_{hom} \rangle_{Q_{H,i}}$.

Arguing as above for (6.61), we obtain

$$
\begin{aligned}
&\left| \int_{Q_{H,i}} W_\epsilon(x, \nabla u^{H,i}_\epsilon - \chi_{H,i}(\epsilon)) - W_\epsilon(x, \nabla u^{H,i}_\epsilon) \right| \\
&\qquad \le C|Q_{H,i}||\chi_{H,i}(\epsilon)|(1 + |\langle \nabla u_\epsilon \rangle_{Q_{H,i}}|^{p-1} + |\chi_{H,i}(\epsilon)|^{p-1}).
\end{aligned} \tag{6.65}
$$

The combination of the four inequalities (6.60), (6.61), (6.64), and (6.65) yields

$$
\begin{aligned}
&\left| \int_{Q_{H,i}} W_\epsilon(x, \nabla u^{H,i}_{hom}) - W_\epsilon(x, \nabla u^{H,i}_\epsilon) \right| \\
&\quad \le \mathcal{C}|Q_{H,i}||\chi_{H,i}(\epsilon)|(1 + |\langle \nabla u_{hom} \rangle_{Q_{H,i}}|^{p-1} \\
&\qquad\qquad + |\langle \nabla u_\epsilon \rangle_{Q_{H,i}}|^{p-1} + |\chi_{H,i}(\epsilon)|^{p-1}).
\end{aligned} \tag{6.66}
$$

Letting ϵ go to zero in (6.66) proves inequality (6.59) since Lemmata 6.1 and 6.2 imply that

$$\lim_{\epsilon\to 0}\left(\inf\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla u)\,|\,\langle\nabla u\rangle_{Q_{H,i}} = \langle\nabla u_{hom}\rangle_{Q_{H,i}}\right\}\right)$$
$$= \inf\left\{\int_{Q_{H,i}} W_{hom}(x,\nabla u)\,|\,\langle\nabla u\rangle_{Q_{H,i}} = \langle\nabla u_{hom}\rangle_{Q_{H,i}}\right\}.$$

The combination of (6.57), (6.58), and (6.59) then shows

$$0 \le \limsup_{\epsilon\to 0} A_i^{\epsilon,H} \le \mathcal{C}\left|\int_{Q_{H,i}} W_{hom}(x,\nabla u_{hom}) - W_{hom}(x,\nabla u_{hom}^{H,i})\right|. \tag{6.67}$$

We thus obtain

$$\begin{aligned}\limsup_{\epsilon\to 0}\sum_{i=1}^{I_H} A_i^{\epsilon,H} \le \mathcal{C}\sum_{i=1}^{I_H}\Bigg|&\int_{Q_{H,i}} W_{hom}(x,\nabla u_{hom})\\ &- \inf\left\{\int_{Q_{H,i}} W_{hom}(x,\nabla u)\,|\,\langle\nabla u\rangle_{Q_{H,i}} = \langle\nabla u_{hom}\rangle_{Q_{H,i}}\right\}\Bigg|.\end{aligned} \tag{6.68}$$

It remains to prove that the right-hand side of (6.68) tends to 0 when H goes to 0.

Let $\bar{W}^H$ be the energy density defined by

$$\bar{W}^H(x,\xi) = \sum_i^{I_H}\inf\left\{\frac{1}{|Q_{H,i}|}\int_{Q_{H,i}} W_{hom}(x,\nabla u)\,|\,\langle\nabla u\rangle = \xi\right\} 1_{Q_{H,i}}(x).$$

Arguing as in the proof of (6.40), for almost all $x \in \Omega$ and for all $\xi \in \mathbb{R}^n$, it can be shown that $\lim_{H\to 0}\bar{W}^H(x,\xi) = W_{hom}(x,\xi)$. Since every argument of the absolute values in the right-hand side of (6.68) is positive, inequality (6.68) can be rewritten as

$$\begin{aligned}\limsup_{\epsilon\to 0}\sum_{i=1}^{I_H} A_i^{\epsilon,H} \le \mathcal{C}&\left|\int_\Omega \bar{W}^H(x,M_H(\nabla u_{hom})) - \bar{W}^H(x,\nabla u_{hom})\right|\\ &+\mathcal{C}\left|\int_\Omega \bar{W}^H(x,\nabla u_{hom}) - W_{hom}(x,\nabla u_{hom})\right|.\end{aligned} \tag{6.69}$$

Let us denote by Rhs_1^H and Rhs_2^H, respectively, the first and second terms of the right-hand side of (6.69). Since W_{hom} satisfies properties (6.2) and $H3$, we can argue as in the proof of (6.61) and obtain

$$|\bar{W}^H(x,\xi_1) - \bar{W}^H(x,\xi_2)| \le \mathcal{C}|\xi_1-\xi_2|(1+|\xi_1|^{p-1}+|\xi_2|^{p-1}),$$

which is independent of H. Thus, using Hölder's inequality, Rhs_1^H is dominated by

$$\mathcal{C}\|M_H(\nabla u_{hom}) - \nabla u_{hom}\|_{L^p(\Omega)}(1+\|M_H(\nabla u_{hom})\|_{L^p(\Omega)}^{p-1} + \|\nabla u_{hom}\|_{L^p(\Omega)}^{p-1})$$

which converges to 0 as H goes to 0 since $\lim_{H\to 0} M_H(\nabla u_{hom}) = \nabla u_{hom}$ in $L^p(\Omega)$.

Next, the dominated convergence theorem implies that Rhs_2^H also tends to 0 as H goes to 0 since the integrand pointwise converges to 0 and is dominated independently of H using $H3$.

This finally shows (6.55). It may be noticed that the assumption $\beta \le p$ allows us to obtain the global estimate (6.69) on Ω, starting from the local estimates (6.68) on $Q_{H,i}$. This would not be the case for $\beta > p$.

Limit of $\sum_{i=1}^{I_H} B_i^{\epsilon,H,\eta}$

Remark first that $\langle\nabla u_\epsilon\rangle_{Q_{H,i}} = \langle\nabla u_{hom}\rangle_{Q_{H,i}} + \chi_{H,i}(\epsilon)$ and $\langle\nabla u_{\eta,\epsilon}\rangle_{Q_{H,i}} = \langle\nabla u_{hom}\rangle_{Q_{H,i}} + \phi_{H,i}(\epsilon,\eta)$, where $\chi_{H,i}$ and $\phi_{H,i}$ are functions satisfying $\lim_{\epsilon\to 0}\chi_{H,i}(\epsilon) = 0$ and $\lim_{\eta\to 0}\lim_{\epsilon\to 0}\phi_{H,i}(\epsilon,\eta) = 0$. Consequently, arguing as for (6.61), we obtain

$$B_i^{\epsilon,H,\eta} \leq \mathcal{C}|Q_{H,i}|(|\chi_{H,i}(\epsilon)| + |\phi_{H,i}(\epsilon,\eta)|) \cdot (1 + |\langle \nabla u_{\eta,\epsilon}\rangle_{Q_{H,i}}|^{p-1} + |\langle \nabla u_\epsilon\rangle_{Q_{H,i}}|^{p-1} + |\chi_{H,i}(\epsilon)|^{p-1} + |\phi_{H,i}(\epsilon,\eta)|^{p-1}). \tag{6.70}$$

Letting ϵ go to 0 in (6.70) yields

$$\limsup_{\epsilon\to 0} B_i^{\epsilon,H,\eta} \leq \mathcal{C}|Q_{H,i}||\langle \nabla u_{\eta,hom}\rangle_{Q_{H,i}} - \langle \nabla u_{hom}\rangle_{Q_{H,i}}| \cdot (1 + |\langle \nabla u_{\eta,hom}\rangle_{Q_{H,i}}|^{p-1} + |\langle \nabla u_{hom}\rangle_{Q_{H,i}}|^{p-1}). \tag{6.71}$$

The summation of (6.71) for i from 1 to I_H exactly reads

$$\limsup_{\epsilon\to 0} \sum_{i=1}^{I_H} B_i^{\epsilon,H,\eta} \leq \mathcal{C} \int_\Omega |M_H(\nabla u_{\eta,hom} - \nabla u_{hom})| \cdot (1 + |M_H(\nabla u_{\eta,hom})|^{p-1} + |M_H(\nabla u_{hom})|^{p-1}). \tag{6.72}$$

Using Hölder's inequality, we obtain

$$\limsup_{\epsilon\to 0} \sum_{i=1}^{I_H} B_i^{\epsilon,H,\eta} \leq \mathcal{C}\|M_H(\nabla u_{\eta,hom} - \nabla u_{hom})\|_{L^p(\Omega)} \cdot \left(\int_\Omega 1 + |M_H(\nabla u_{\eta,hom})|^p + |M_H(\nabla u_{hom})|^p\right)^{\frac{p-1}{p}}. \tag{6.73}$$

Let us now prove that the right-hand side of (6.73) converges to 0 uniformly in H when η goes to 0. Using Hölder's inequality, for every $w \in L^p(\Omega)$ and every $H > 0$, we have

$$\begin{aligned}
\int_\Omega |M_H(w)|^p &= \sum_i \mathcal{L}^n(Q_{H,i}) \left| \frac{1}{\mathcal{L}^n(Q_{H,i})} \int_{Q_{H,i}} w \right|^p \\
&\leq \sum_i \mathcal{L}^n(Q_{H,i}) \left(\frac{1}{\mathcal{L}^n(Q_{H,i})} \left(\int_{Q_{H,i}} |w|^p \right)^{1/p} \left(\int_{Q_{H,i}} 1 \right)^{(p-1)/p} \right)^p \\
&= \sum_i \int_{Q_{H,i}} |w|^p \\
&= \|w\|^p_{L^p(\Omega)}.
\end{aligned}$$

This calculation shows that (M_H) is an equicontinuous family of operators on $L^p(\Omega)$. By the application of Theorem 40, $u_{\eta,hom}$ converges strongly to u_{hom} in $W^{1,p}(\Omega)$. This strong convergence and the equicontinuity of M_H show that there exists $\mathcal{D} > 0$ such that

$$\left(\int_\Omega 1 + |M_H(\nabla u_{\eta,hom})|^p + |M_H(\nabla u_{hom})|^p\right)^{\frac{p-1}{p}} \leq \mathcal{D}$$

for every η and H, and that

$$\lim_{\eta\to 0} \|M_H(\nabla u_{\eta,hom} - \nabla u_{hom})\|_{L^p(\Omega)} = 0$$

uniformly in H, which proves (6.56).

The combination of (6.54), (6.55), and (6.56) concludes the proof of Theorem 41.

6.2.5 Proof of Proposition 5

Proof of (6.18). We first prove (6.14), from which we will deduce (6.18). Let us start from

$$\begin{aligned}\left|\int_\Omega W_{hom}(x,\nabla u_{hom}) - W_{\eta,\epsilon}(x,\nabla u_{hom})\right| &\le \int_\Omega err^0_{\eta,\epsilon}(x,\nabla u_{hom}),\\ \left|\int_\Omega W_{hom}(x,\nabla u_{\eta,\epsilon}) - W_{\eta,\epsilon}(x,\nabla u_{\eta,\epsilon})\right| &\le \int_\Omega err^0_{\eta,\epsilon}(x,\nabla u_{\eta,\epsilon}).\end{aligned} \tag{6.74}$$

Since $|\inf\{f_1(w), w \in W\} - f_2(w_1)| \le \tau_1$ and $|\inf\{f_2(w), w \in W\} - f_1(w_2)| \le \tau_2$ imply $|f_1(w_1) - f_2(w_2)| \le \tau_1 + \tau_2$ if $f_1(w_1) = \inf\{f_1(w), w \in W\}$ and $f_2(w_2) = \inf\{f_2(w), w \in W\}$, inequalities (6.74) yield

$$\left|\int_\Omega W_{hom}(x,\nabla u_{hom}) - W_{\eta,\epsilon}(x,\nabla u_{\eta,\epsilon})\right| \le \int_\Omega err^0_{\eta,\epsilon}(x,\nabla u_{hom}) + err^0_{\eta,\epsilon}(x,\nabla u_{\eta,\epsilon}). \tag{6.75}$$

Using the triangle inequality, (6.74), and (6.75), we then have

$$\left|\int_\Omega W_{hom}(x,\nabla u_{hom}) - W_{hom}(x,\nabla u_{\eta,\epsilon})\right| \le \int_\Omega err^0_{\eta,\epsilon}(x,\nabla u_{hom}) + 2err^0_{\eta,\epsilon}(x,\nabla u_{\eta,\epsilon}). \tag{6.76}$$

Finally, (6.14) is a consequence of Lemma 6.13 applied to W_{hom}, u_{hom}, and $u_{\eta,\epsilon}$ and of the Poincaré–Wirtinger inequality. In addition, if (6.16) holds, one can use Hölder's inequality and bound $\|\nabla u_{\eta,\epsilon}\|_{0,p}$ and $\|\nabla u_{hom}\|_{0,p}$ by $C(1 + \|\nabla \bar{u}\|_{0,p})$ using $H3$, which allows us to conclude and prove (6.18).

Proof of (6.19). Let us now focus on the operator instead of the energy and start from

$$\int_\Omega a_{hom}(x,\nabla u_{hom}) \cdot \nabla v = \int_\Omega a_{\eta,\epsilon}(x,\nabla u_{\eta,\epsilon}) \cdot \nabla v \tag{6.77}$$

for all $v \in W^{1,p}_0(\Omega)$ (this equality holds with a nonzero right-hand side in the elliptic PDE). Therefore,

$$\begin{aligned}&\int_\Omega (a_{\eta,\epsilon}(x,\nabla u_{\eta,\epsilon}) - a_{\eta,\epsilon}(x,\nabla u_{hom})) \cdot (\nabla u_{\eta,\epsilon} - \nabla u_{hom})\\ &\qquad = \int_\Omega (a_{hom}(x,\nabla u_{hom}) - a_{\eta,\epsilon}(x,\nabla u_{hom})) \cdot (\nabla u_{\eta,\epsilon} - \nabla u_{hom})\end{aligned} \tag{6.78}$$

since $u_{\eta,\epsilon} - u_{hom} \in W^{1,p}_0(\Omega)$. Using the following lemma, properties (6.2) and (6.3), and the Poincaré–Wirtinger inequality, we obtain (6.15).

Lemma 6.14 (see [38, Lemma 23.9]). *Let v_1 and v_2 be two functions belonging to $L^p(\Omega)$. Then*

$$\begin{aligned}\|\nabla v_1 - \nabla v_2\|_{0,p} \le c &\left(\int_\Omega |\nabla v_1 - \nabla v_2|^\beta (1 + |\nabla v_1| + |\nabla v_2|)^{p-\beta\wedge p}\right)^{\frac{1}{\beta\wedge p}}\\ &\cdot \left(\mathcal{L}^n(\Omega)^{\frac{1}{p}} + \|\nabla v_1\|_{0,p} + \|\nabla v_2\|_{0,p}\right)^{\frac{\beta\wedge p - p}{\beta}}\end{aligned} \tag{6.79}$$

with $c > 0$ depending only on n and p.

In addition, if (6.17) holds, Hölder's inequality and $H3$ allow us to conclude.

Proof of (6.20). In order to estimate the global error, we first control $err^0_{\eta,\epsilon}(x,\xi)$.

Lemma 6.15. *Let $W_\epsilon(x,\xi) = W(\frac{x}{\epsilon}, x)$, where W is 1-periodic and satisfies the hypotheses $H1$, $H2$, and $H3$; then there exist $C > 0$ and $N \in \mathbb{N}$ such that for all $x \in \Omega$, $\xi \in \mathbb{R}^n$, and for all $\eta > 0$ and $\epsilon > 0$ with $\frac{\eta}{\epsilon} \ge N$, we have*

$$|W_{\eta,\epsilon}(x,\xi) - W_{hom}(\xi)| \le C\frac{\epsilon}{\eta}(1 + |\xi|^p) \tag{6.80}$$

(let us recall that W_{hom} does not depend on the space variable in periodic homogenization).

Remark **25** As will be clear in the proof, (6.80) can be replaced by

$$|W_{\eta,\epsilon}(x,\xi) - W_{hom}(\xi)| \leq C\frac{\epsilon}{\eta}|\xi|^2 \tag{6.81}$$

if the energy density is quadratic.

Let us prove Lemma 6.15 in two steps by suitably bounding $W_{\eta,\epsilon}(x,\xi)$ from above and from below.

Upper bound. In order to dominate $W_{\eta,\epsilon}(x,\xi)$ we introduce a specific test function $u \in W^{1,p}(N(x,\eta))$ such that $u(x) = \xi \cdot x$ on $\partial N(x,\eta)$. Before doing this, let us fix some geometric notations. We consider the hypercube $N(x,\eta)$ and decompose it into three domains B_M, D_M, and D_η. The domain B_M is a cube containing M^n periodic cells, where $M = \left[\frac{\eta}{\epsilon}\right] - 2$, and such that $dist(B_M, \partial N(x,\eta)) \geq \epsilon$. The domain D_M is a square crown around B_M made of one layer of periodic cells and such that $dist(B_M, \partial N(x,\eta)) \leq \epsilon$. Finally, D_η is the possible empty set defined by $N(x,\eta) - B_M \cup D_M$. A sketch of B_M, D_M, and D_η is given in Figure 6.1.

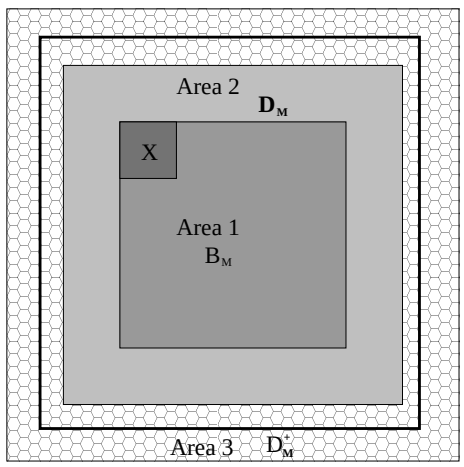

Fig. 6.1. Geometric notation: $\partial N(x,\eta)$ is the thick black line, the periodic cell is $\mathbf{X}$, B_M is Area 1, D_M is Area 2, and D_M^+ is Area 3.

The function u we consider is defined as follows:

- $u_{|B_M}(y) = v_{B_M}(y) + \xi \cdot y$, where v_{B_M} is the unique solution of

$$\inf\left\{\int_{B_M} W_\epsilon(y, \xi + \nabla v), v \in W^{1,p}_{\#}(B_M)\right\}.$$

- $u_{|D_M}(y) = v_0(y) + \xi \cdot y$, where $v_0 \in W^{1,p}(D_M)$ satisfies $\mathrm{Tr}(v_0, D_M \cap B_M) = \mathrm{Tr}(v_{B_M}, D_M \cap B_M)$ and $\mathrm{Tr}(v_0, \partial D_M - D_M \cap B_M) = 0$.
- $u_{|D_\eta}(y) = \xi \cdot y$ if D_η is not empty.

By construction, $u \in W^{1,p}(N(x,\eta))$ and $u(y) = \xi \cdot y$ on $\partial N(x,\eta)$. Therefore

$$\frac{1}{\eta^n}\int_{N(x,\eta)} W_\epsilon(y, \nabla u) \geq W_{\eta,\epsilon}(x,\xi). \tag{6.82}$$

We claim that there exists $C > 0$, such that for $\eta \geq 3\epsilon$ and for all $x \in \Omega$, $\xi \in \mathbb{R}^n$, we have

$$\frac{1}{\eta^n}\int_{N(x,\eta)} W_\epsilon(y, \nabla u) \leq \frac{(M\epsilon)^n}{\eta^n} W_{hom}(\xi) + C\frac{\epsilon}{\eta}(1 + |\xi|^p),$$

where $M = \left[\frac{\eta}{\epsilon}\right] - 2$, which also reads, using (6.82),

$$W_{\eta,\epsilon}(x,\xi) \leq W_{hom}(\xi) + C\frac{\epsilon}{\eta}(1+|\xi|^p) \tag{6.83}$$

since $W_{hom}(\xi) \geq 0$ and $\frac{M\epsilon}{\eta} \leq 1$. The first part of (6.83) is a consequence of the following equality obtained by a uniqueness argument:

$$\inf\left\{\int_{B_M} W_\epsilon(y,\xi+\nabla v), v \in W^{1,p}_{\#}(B_M)\right\} = M^n W_{hom}(\xi).$$

The contribution of $u_{|D_\eta}$ to the energy is controlled by $\frac{1}{\eta^n}2^n\epsilon\eta^{n-1}C(1+|\xi|^p)$ by using $H3$ and estimating the measure of D_η. The last contribution, which is due to $u_{|D_M}$, can also be bounded using $H3$. To this aim we consider smooth functions χ_M on the rescaled crowns $\tilde{D}_M$ (that is of thickness 1 and not ϵ) such that $\chi_M(y) = 1$ if $dist(y, \partial\{\tilde{D}_M\}_{int}) \leq 1/4$, $\chi(y) = 0$ if $dist(y, \partial\{\tilde{D}_M\}_{ext}) \leq 1/4$, $\|\chi_M\|_{L^\infty} \leq 1$, and $\|\nabla\chi_M\|_{L^\infty} \leq 4$, where $\partial\{\tilde{D}_M\}_{int}$ is the inner boundary of the crown and $\partial\{\tilde{D}_M\}_{ext}$ is the outer boundary of the crown. Such a family of functions clearly exists (it is enough to check for $M = 3$). One can then construct an explicit function v_0 using χ_M, the periodic replication on $\tilde{D}_M$ of $u_{\#}(\xi)$ which is the minimizer of (6.22), and the scaling $L^p((0,1)^n) \ni w \mapsto sc(w) = \epsilon w\left(\frac{\cdot}{\epsilon}\right) \in L^p((0,\epsilon)^n)$. With such a function v_0, we can estimate the contribution of $u_{|D_M}$ by

$$\int_{D_M} W_\epsilon(y,\nabla u_{|D_M}) \leq C2^n\epsilon\eta^{n-1}(1+|\xi|^p + \|u_{\#}(\xi)\|^p_{W^{1,p}((0,1)^n)}).$$

The growth condition $H3$ can be used to bound $\|\nabla u_{\#}(\xi)\|^p_{L^p((0,1)^n)}$ by $C(1+|\xi|^p)$, and also $\|u_{\#}(\xi)\|^p_{W^{1,p}((0,1)^n)}$ using the Poincaré–Wirtinger inequality. We thus have

$$\int_{D_M} W_\epsilon(y,\nabla u_{|D_M}) \leq C\epsilon\eta^{n-1}(1+|\xi|^p).$$

For a quadratic energy, we can skip the constant 1 in $H3$ and therefore also in (6.83).
Lower bound. To obtain a lower bound on $W_{\eta,\epsilon}(x,\xi)$ we will proceed in the same way as above, but we will complete the cube instead of considering a smaller cube B_M. To this aim, we introduce $B_{M+2} = B_M \cup D_M$ and consider D^+_M, which is defined as the crown around B_{M+2} made of one layer of periodic cells. These sets satisfy the following properties: $D_\eta \subset D^+_M$, $N(x,\eta) \subset B_{M+2} \cup D^+_M$ and $\mathcal{L}^n(B_{M+2} \cup D^+_M - N(x,\eta)) \leq C\eta^{n-1}\epsilon$.

Let $w \in W^{1,p}(N(x,\eta))$ be the unique minimizer of

$$\inf\left\{\int_{N(x,\eta)} W_\epsilon(y,\nabla v), v_{|\partial N(x,\eta)}(y) = \xi\cdot y\right\}.$$

We introduce the following test function $u \in W^{1,p}(B_{M+2} \cup D^+_M)$ defined by

- $u_{|N(x,\eta)} = w$,
- $u_{|B_{M+2}\cup D^+_M - N(x,\eta)}(y) = \xi\cdot y$.

By construction,

$$W_{\eta,\epsilon}(x,\xi) = \frac{1}{\eta^n}\left(\int_{B_{M+2}\cup D^+_M} W_\epsilon(y,\nabla u) - \int_{B_{M+2}\cup D^+_M - N(x,\eta)} W_\epsilon(y,\xi)\right).$$

Thus

$$
\begin{aligned}
W_{\eta,\epsilon}(x,\xi) &\geq \frac{1}{\eta^n}\inf\left\{\int_{B_{M+2}\cup D_M^+} W_\epsilon(y,\nabla v), v_{|\partial B_{M+2}\cup D_M^+}(y) = \xi\cdot y\right\} \\
&\qquad - C\frac{\epsilon}{\eta}(1+|\xi|^p) \\
&\geq \frac{1}{\eta^n}\inf\left\{\int_{B_{M+2}\cup D_M^+} W_\epsilon(y,\xi+\nabla v), v\in W^{1,p}_{\#}(B_{M+2}\cup D_M^+)\right\} \\
&\qquad - C\frac{\epsilon}{\eta}(1+|\xi|^p) \\
&\geq W_{hom}(\xi) + c\frac{\epsilon}{\eta}W_{hom}(\xi) - C\frac{\epsilon}{\eta}(1+|\xi|^p).
\end{aligned}
\tag{6.84}
$$

The lower and upper bounds (6.83) and (6.84) then imply (6.80), which proves Lemma 6.15.

Finally, the combination of Lemma 6.15 and (6.18) shows (6.20).

Proof of (6.21). This error estimate is a direct consequence of (6.19) once we have noticed the following fact: if there exists $\theta > 0$ such that for all ξ and x,

$$
|W_{\eta,\epsilon}(x,\xi) - W_{hom}(x,\xi)| \leq \theta|\xi|^2, \tag{6.85}
$$

then

$$
|a_{\eta,\epsilon}(x,\xi) - a_{hom}(x,\xi)| \leq \theta|\xi|. \tag{6.86}
$$

This comes from the very structure of $W_{\eta,\epsilon}$ and W_{hom}: for all $x\in\Omega$, there exist $A_{\eta,\epsilon}(x)\in\mathcal{M}_n(\mathbb{R})$ and $A_{hom}\in\mathcal{M}_n(\mathbb{R})$ such that $W_{\eta,\epsilon}(x,\xi) = \xi^T A_{\eta,\epsilon}(x)\xi$ and $W_{hom}(x,\xi) = \xi^T A_{hom}\xi$. Therefore (6.85) implies that $|A_{\eta,\epsilon}(x) - A_{hom}| \leq \theta$, which shows (6.86). Combining this observation with (6.81), and using (6.19) with $p=2$, $\beta=2$, and $\alpha=1$, we obtain (6.21).

6.2.6 Proof of Proposition 6

The error is decomposed into two terms. The first term is linked to the mismatch between boundary conditions (namely periodic and Dirichlet) and to the fact that $\frac{\eta}{\epsilon}$ may not be an integer. This error could be cancelled if we used periodic boundary conditions on a domain fitting the periodic pattern. This is also the aim of oversampling methods to reduce this first term. The second term comes from the approximation of $\xi_{hom,i}$ by $\langle\nabla u_{\eta,\epsilon}\rangle_{|Q_{H,i}}$.

First term

As shown in the proof of the upper and lower bounds for (6.80), the energy difference for different boundary conditions (periodic and Dirichlet) on a domain $B(x,\eta)$ is of order $\epsilon\eta^{n-1}(1+|\xi|^p)$. Let us consider an extension $\bar{Q}_{H,i}$ of $Q_{H,i}$ such that $Q_{H,i}\subset\bar{Q}_{H,i}$, $\bar{Q}_{H,i}$ is a union of periodic cells and $\mathcal{L}^n(\bar{Q}_{H,i} - Q_{H,i}) \leq 2^n\epsilon\eta^{n-1}$. Let $\bar{v}^{H,i}_{\eta,\epsilon}\in W^{1,p}(\bar{Q}_{H,i})$ be a function whose restriction on $Q_{H,i}$ is $v^{H,i}_{\eta,\epsilon}$, as was u to w in the proof for the lower bound of (6.80). Let us also consider the function $v^{\#,i}_{\eta,\epsilon}\in W^{1,p}(\bar{Q}_{H,i})$ defined as the minimizer of

$$
\inf\left\{\int_{\bar{Q}_{H,i}} W_{\eta,\epsilon}(y,\nabla v), v(y) = \langle\nabla u_{\eta,\epsilon}\rangle_{|Q_{H,i}}\cdot y + v_{\#}(y), v_{\#}\in W^{1,p}_{\#}(\bar{Q}_{H,i})\right\}.
$$

The first term of (6.23) is $\|\nabla v^{H,i}_{\eta,\epsilon} - \nabla v^{\#,i}_{\eta,\epsilon}\|_{L^p(Q_{H,i})} \leq \|\nabla v^{H,i}_{\eta,\epsilon} - \nabla v^{\#,i}_{\eta,\epsilon}\|_{L^p(\bar{Q}_{H,i})}$. It can be controlled using Lemma 6.13 since $v^{\#,i}_{\eta,\epsilon}$ is defined as a minimizer on a convex set. We treat only the case $\beta\geq p$, the converse case being similar. With $B=\bar{Q}_{H,i}$, (6.46) reads

$$
\begin{aligned}
\|\nabla v^{H,i}_{\eta,\epsilon} - \nabla v^{\#,i}_{\eta,\epsilon}\|^p_{L^p(Q_{H,i})} &\leq \left|\int_{\bar{Q}_{H,i}} W_{\eta,\epsilon}(\nabla v^{\#,i}_{\eta,\epsilon}) - W_{\eta,\epsilon}(\nabla v^{H,i}_{\eta,\epsilon})\right|^{\frac{p}{\beta}} \\
&\qquad\cdot\left(\mathcal{L}^n(\bar{Q}_{H,i})^{\frac{1}{p}} + \|\nabla v^{H,i}_{\eta,\epsilon}\|_{L^p(\bar{Q}_{H,i})} + \|\nabla v^{\#,i}_{\epsilon}\|_{L^p(\bar{Q}_{H,i})}\right)^{\frac{p(\beta-p)}{\beta}}.
\end{aligned}
\tag{6.87}
$$

As $\mathcal{L}^n(\bar{Q}_{H,i} - Q_{H,i}) \leq c\eta^{n-1}\epsilon$, the first term of (6.87) can be bounded by

$$C(\epsilon\eta^{n-1})^{\frac{p}{\beta}} \frac{1}{\mathcal{L}^n(Q_{H,i})^{\frac{p}{\beta}}} \frac{\mathcal{L}^n(\Omega)}{\mathcal{L}^n(Q_{H,i})} \frac{\mathcal{L}^n(Q_{H,i})}{\mathcal{L}^n(\Omega)} \left(\int_{Q_{H,i}} (1 + |\nabla u_{\eta,\epsilon}|)^p \right)^{\frac{p}{\beta}},$$

using Hölder's inequality. Summing the contributions of all $Q_{H,i}$, using the concavity of $z \mapsto z^{\frac{p}{\beta}}$ and replacing $\mathcal{L}^n(Q_{H,i})$ by η^n, we obtain the following upper bound for the contribution of the first terms on the whole Ω:

$$C \left(\frac{\epsilon}{\eta} \right)^{\frac{p}{\beta}} \eta^{-n} \left(\eta^n \int_\Omega (1 + |\nabla u_{\eta,\epsilon}|)^p \right)^{\frac{p}{\beta}}. \tag{6.88}$$

As the following chain of inequalities holds,

$$\begin{aligned} \|\nabla v_{\eta,\epsilon}^{H,i}\|^p_{L^p(\bar{Q}_{H,i})} &\leq C\mathcal{L}^n(\bar{Q}_{H,i})(1 + \langle \nabla u_{\eta,\epsilon} \rangle^p_{Q_{H,i}}) \\ &\leq C \frac{\mathcal{L}^n(\bar{Q}_{H,i})}{\mathcal{L}^n(Q_{H,i})} \left(\mathcal{L}^n(Q_{H,i}) + \|\nabla u_{\eta,\epsilon}\|^p_{L^p(Q_{H,i})} \right) \\ &\leq C(\mathcal{L}^n(Q_{H,i}) + \|\nabla u_{\eta,\epsilon}\|^p_{L^p(Q_{H,i})}), \end{aligned}$$

if $\{u_{\eta,\epsilon}\}$ is bounded in $W^{1,q}(\Omega)$ for $q \geq p$, the second term of (6.87) is lower than or equal to

$$\eta^{n\frac{q-p}{q}\frac{\beta-p}{\beta}} \left(\mathcal{L}^n(\Omega)^{\frac{1}{q}} + \|\nabla u_{\eta,\epsilon}\|_{L^q(\Omega)} + \|\nabla u_{hom}\|_{L^q(\Omega)} \right)^{\frac{p(\beta-p)}{\beta}},$$

using Hölder's inequality. This contribution is constant and factorized in front of (6.88). Including the L^q norms in the constant C_3, we obtain the first term of (6.23).

Second term

This term is more interesting and corresponds to the error made by approximating ∇u_{hom} by $\nabla u_{\eta,\epsilon}$ in the minimization problem (6.22), for which periodic boundary conditions are used. If the numerical corrector is modified (e.g., by an oversampling method), this second term still remains. We thus have to estimate $\|\tilde{p}(\xi_{hom,i}) - \tilde{p}(\langle \nabla u_{\eta,\epsilon} \rangle_{|Q_{H,i}})\|_{W^{1,p}((0,1)^n)}$, which can be done using the following lemma.

Lemma 6.16 (see [38, Lemma 23.12]). *Let $\xi_1, \xi_2 \in \mathbb{R}^n$, and $\tilde{p}(\xi_1)$ and $\tilde{p}(\xi_2)$, be the solutions of* (6.22)*, respectively, for $\xi = \xi_1$ and $\xi = \xi_2$. Then the following estimate holds:*

$$\|\tilde{p}(\xi_1) - \tilde{p}(\xi_2)\|^p_{W^{1,p}((0,1)^n)} \leq c(1 + |\xi_1|^p + |\xi_2|^p)^{\frac{\beta\wedge p-\alpha-1}{\beta\wedge p-\alpha}} |\xi_1 - \xi_2|^{\frac{p}{\beta\wedge p-\alpha}}. \tag{6.89}$$

This lemma is proved in [38, p. 234] for $\beta \geq p$, which is the more subtle case. For $\beta \leq p$, the adaptation of the proof is straightforward. We will consider $\beta \geq p$, without loss of generality.

Let us apply Lemma 6.16 for $\xi_1 = \xi_{hom,i}$ and $\xi_2 = \langle \nabla u_{\eta,\epsilon} \rangle_{|Q_{H,i}}$. Using Hölder's inequality, we have

$$\begin{aligned} \|\tilde{p}(\xi_1) - \tilde{p}(\xi_2)\|^p_{W^{1,p}((0,1)^n)} \leq \frac{c}{\mathcal{L}^n(Q_{H,i})} &\left(\int_{Q_{H,i}} |\nabla u_{\eta,\epsilon} - \nabla u_{hom}|^p \right)^{\frac{1}{\beta-\alpha}} \\ &\cdot (\mathcal{L}^n(Q_{H,i}) + \|\nabla u_{\eta,\epsilon}\|^p_{L^p(Q_{H,i})} + \|\nabla u_{hom}\|^p_{L^p(Q_{H,i})})^{\frac{\beta-\alpha-1}{\beta-\alpha}}. \end{aligned} \tag{6.90}$$

Thus, passing from $\tilde{p}$ to p by a scaling argument and integrating over $Q_{H,i}$ yields

$$\begin{aligned} \|p(\xi_1) - p(\xi_2)\|^p_{W^{1,p}(Q_{H,i})} \leq c &\left(\int_{Q_{H,i}} |\nabla u_{\eta,\epsilon} - \nabla u_{hom}|^p \right)^{\frac{1}{\beta-\alpha}} \\ &\cdot (\mathcal{L}^n(Q_{H,i}) + \|\nabla u_{\eta,\epsilon}\|^p_{L^p(Q_{H,i})} + \|\nabla u_{hom}\|^p_{L^p(Q_{H,i})})^{\frac{\beta-\alpha-1}{\beta-\alpha}}. \end{aligned} \tag{6.91}$$

If $\{u_{\eta,\epsilon}\}$ is bounded in $W^{1,q}(\Omega)$ for $q \geq p$, then the second term of (6.91) is controlled by

$$\eta^{n\frac{q-p}{q}\frac{\beta-\alpha-1}{\beta-\alpha}}(\mathcal{L}^n(\Omega)^{\frac{p}{q}} + \|\nabla u_{\eta,\epsilon}\|^p_{L^q(\Omega)} + \|\nabla u_{hom}\|^p_{L^q(\Omega)})^{\frac{\beta-\alpha-1}{\beta-\alpha}},$$

which is constant. Adding the contributions of (6.91) for all $Q_{H,i}$ and using the concavity of $z \mapsto z^{\frac{1}{\beta-\alpha}}$, we finally obtain

$$\begin{aligned}\sum_i \|p(\xi_{hom,i}) - p(\langle\nabla u_{\eta,\epsilon}\rangle_{|Q_{H,i}})\|^p_{W^{1,p}(Q_{H,i})} \\ \leq c\eta^{n\frac{q-p}{q}\frac{\beta-\alpha-1}{\beta-\alpha}}(\mathcal{L}^n(\Omega)^{\frac{p}{q}} + \|\nabla u_{\eta,\epsilon}\|^p_{L^q(\Omega)} + \|\nabla u_{hom}\|^p_{L^q(\Omega)})^{\frac{\beta-\alpha-1}{\beta-\alpha}} \\ \cdot \frac{\mathcal{L}^n(\Omega)}{\eta^n}\left(\sum_i \frac{\eta^n}{\mathcal{L}^n(\Omega)}\int_{Q_{H,i}} |\nabla u_{\eta,\epsilon} - \nabla u_{hom}|^p\right)^{\frac{1}{\beta-\alpha}},\end{aligned} \tag{6.92}$$

which gives the second term of (6.23) and concludes the proof.

6.3 Relation to some existing numerical approaches

The analytical developments of the previous sections may serve as a theoretical framework for both the heterogeneous multiscale method (HMM) and the multiscale finite element method (MsFEM), when applied to the homogenization of elliptic operators. Indeed, as will be shown in sections 6.3.1 and 6.3.2, respectively, the HMM and the MsFEM for elliptic equations can both basically be rewritten as approximations by a quadrature rule of the energy functional

$$\begin{aligned}I_{\eta,\epsilon}(u) = \int_\Omega \inf\big\{\langle W_\epsilon(\cdot,\nabla v(\cdot))\rangle_{B(x,\eta)} \mid \\ v \in W^{1,p}(B(x,\eta)), \langle\nabla v\rangle_{B(x,\eta)} = \nabla u(x)\big\}\, dx\end{aligned}$$

introduced in Definition 22. If computed with the same η, the MsFEM and the HMM are basically two different approximations of the same continuous problem, which can explain why the two methods give similar results, as illustrated in some examples in [136], even with general heterogeneities for which neither the limit in η nor the limit in ϵ is attained. The complexities of these numerical methods may, however, be different, and we refer the reader to [136] and [72] for details.

A third existing methodology can be related to the theoretical framework developed in sections 6.1 and 6.2, but it will not be detailed as much as the above two here. It is commonly used in mechanics, when dealing with linear composite materials. The method consists in introducing a small volume (namely a cube), called the representative volume element (RVE) and which represents the material at the macroscopic scale. The macroscopic stress-strain relation $\sigma(x) = A(x)\cdot\nabla u(x)$ of the material can then be approximated by the averaged response $\langle\sigma(y)\rangle_{RVE}$ of the RVE under the homogeneous displacement $\nabla u(x)\cdot y$ of its boundary. In the present paper, the domain $B(x,\eta)$ plays the role of the RVE. As already mentionned, several boundary conditions can be associated with $W_{\eta,\epsilon}$: at least periodic boundary conditions and Dirichlet boundary conditions. Other formulations can be developed using the Legendre transform of the energy and periodic or Dirichlet boundary conditions. In the limit ϵ and η go to zero, and all these energy densities converge to the same homogenized energy density. A huge body of literature in mechanics is devoted to the choice of η and of the boundary conditions that best fit the behavior of the original material (see [154] for a review).

It may be stressed that the full analysis of the two following methods has already been done in the linear case. The linear case is dealt with in the present work only to highlight the fact that some properties, which definitely imply stronger results than the results proved throughout this paper, are specific to the linear case and do not have simple generalizations. We refer the reader to Table 6.1 where *some* references for the already known results are gathered.

6.3.1 HMM

We first assume that W_ϵ is a strictly convex energy density and describe the method within this context, for which minimizing the energy and solving the Euler–Lagrange equation are equivalent. The case of quasiconvex energies is the object of the end of this section.

If we assume that $W_{\eta,\epsilon}$ is perfectly known, then the FE approximation of

$$\inf\left\{\int_\Omega W_{\eta,\epsilon}(x,\nabla u) - fu \,|\, u \in W^{1,p}(\Omega) + BC\right\} \tag{6.93}$$

coincides with the method described in [68, section 3] as an application of the HMM methodology to elliptic operators. Indeed, problem (6.93) is approximated at the discrete level by

$$\inf\left\{\sum_{i=1}^{N_{mesh}}\sum_{j=1}^{N_{GP}} q_j\big(W_{\eta,\epsilon}(x_{ij},\nabla u_H(x_{ij})) - f(x_{ij})u_H(x_{ij})\big), u_H \in V_H + BC\right\}, \tag{6.94}$$

where N_{mesh} is the number of mesh elements, N_{GP} is the number of Gauss points per element, x_{ij} are the Gauss points, q_j are the weights, and V_H is an FE space. Then the computation of the FE minimizer of (6.94) requires only evaluations of derivatives of $W_{\eta,\epsilon}(x_{ij},\xi)$ for particular ξ at Gauss points x_{ij}. The first derivatives are needed for writing the Euler–Lagrange equation, the second derivatives if the latter is nonlinear, and a Newton-type algorithm is used for the solution procedure. This is detailed in Chapter 5, where an explicit formula is given for periodic homogenization (the cell problem is posed on $(0,1)^n$). The setting of Chapter 5 can easily be generalized replacing the periodic cell $(0,1)^n$ by $B(x,\eta)$.

The details of this general procedure may depend on the specific choice of η, of the shape of $B(x,\eta)$ (see Remark 10), and of the boundary conditions in the definition (6.6) of $W_{\eta,\epsilon}$ (see Remark 13). Many variants of the method follow. It is to be emphasized that the macroscopic meshsize H is a priori independent of η, which can be exploited in the case of scale separation.

The works [136] and [1] report some numerical experiments, where $B(x,\eta)$ is replaced by a cube and for which better numerical results are obtained using periodic boundary conditions instead of Dirichlet boundary conditions, even for the nonperiodic problems tested.

Let us go back to the case for which $W_{\eta,\epsilon}$ is not analytically known but also numerically evaluated. We define an FE approximation of $W_{\eta,\epsilon}(x,\xi)$ by

$$W^h_{\eta,\epsilon}(x,\xi) = \inf\big\{\langle W_\epsilon(y,\xi+\nabla v_h)\rangle_{B(x,\eta)} \,|\, v_h \in V_h\big\}, \tag{6.95}$$

where V_h is an FE subspace of $W^{1,p}(B(x,\eta))$ such that $\langle \nabla v_h\rangle_{B(x,\eta)} = 0$ for all $v_h \in V_h$. On the theoretical side, the difference between the minimizer u_{hom} of the exact homogenized problem (6.5) and the FE approximation $u^{h,\epsilon}_{\eta,H}$, defined as the minimizer of

$$\inf\left\{\int_\Omega W^h_{\eta,\epsilon}(x,\nabla u_H) - fu_H \,|\, u_H \in V_H + BC\right\}, \tag{6.96}$$

may be decomposed into three components. Let us denote by u^ϵ_η the minimizer of (6.93) and by $u^\epsilon_{\eta,H}$ the minimizer of

$$\inf\left\{\int_\Omega W_{\eta,\epsilon}(x,\nabla u_H) - fu_H \,|\, u_H \in V_H + BC\right\}. \tag{6.97}$$

Denoting by $err1(\eta,\epsilon) = \|u_{hom} - u^\epsilon_\eta\|_{W^{1,p}(\Omega)}$, by $err2(\eta,\epsilon,H) = \|u^\epsilon_\eta - u^\epsilon_{\eta,H}\|_{W^{1,p}(\Omega)}$, and by $err3(\eta,\epsilon,H,h) = \|u^\epsilon_{\eta,H} - u^{h,\epsilon}_{\eta,H}\|_{W^{1,p}(\Omega)}$, we then have by the triangle inequality

$$\begin{aligned}\|u_{hom} - u^{h,\epsilon}_{\eta,H}\|_{W^{1,p}(\Omega)} &\le \|u_{hom} - u^\epsilon_\eta\|_{W^{1,p}(\Omega)} + \|u^\epsilon_\eta - u^\epsilon_{\eta,H}\|_{W^{1,p}(\Omega)} \\ &\quad + \|u^\epsilon_{\eta,H} - u^{h,\epsilon}_{\eta,H}\|_{W^{1,p}(\Omega)} \\ &= err1 + err2 + err3.\end{aligned} \tag{6.98}$$

In [68], [136], and [69], Weinan E and collaborators have studied $\|u_{hom} - u^{\epsilon}_{\eta,H}\|_{W^{1,p}(\Omega)}$ instead of $err1$ and $err2$ in the framework of the HMM. The error analysis has been performed for elliptic operators linear in the gradient, that is, operators of the form $u \mapsto a(x,u)\nabla u$. Under this assumption, this error has been expressed in terms of another error, called $e(HMM)$, that is a measure of the difference between the homogenized operator and the averaged and discretized operator. The error has then been made explicit in several norms (L^2, H^1, and $W^{1,\infty}$) for the periodic case and for the stochastic and stationary case.

In the present work, we have chosen to make a distinction between $err1$ and $err2$. The second component $err2$ is the difference between the minimizer of (6.93) and its numerical FE approximation, the functional $W_{\eta,\epsilon}$ being considered as explicitly known. This analysis is indeed more classical and can be performed without any assumption on the heterogeneities.

The third component $err3$ is the error due to the approximation of the energy density $W_{\eta,\epsilon}$ itself. In contrast to the difference between W_{hom} and $W_{\eta,\epsilon}$, which cannot be estimated in the general case, the error between $W_{\eta,\epsilon}$ and its numerical approximation $W^h_{\eta,\epsilon}$ can be estimated. This is also the case for the difference between the minimizers $u^{\epsilon}_{\eta,H}$ and $u^{h,\epsilon}_{\eta,H}$. In the linear case, the analysis has been made in [1]. For the nonlinear case, the contribution to $err3$ of the approximation of $W_{\eta,\epsilon}$ by $W^h_{\eta,\epsilon}$ has been analyzed in Chapter 5, where it is shown that the difference between $W_{\eta,\epsilon}(x,\xi)$ and $W^h_{\eta,\epsilon}(x,\xi)$ can be magnified by the nonlinearity in $err3$.

Except in the specific case of linear periodic homogenization, for which the fully discrete HMM is analyzed in [1], a complete analysis of $\|u_{\epsilon} - u^{h,\epsilon}_{\eta,H}\|_{L^p(\Omega)}$ seems out of reach, especially without spatial assumptions.

Let us highlight the contribution of the present work in this setting. Theorem 40 shows that $\lim_{\eta\to 0}\lim_{\epsilon\to 0} err1(\eta,\epsilon) = 0$ in the strictly convex case. In addition, Theorem 41 shows that the reconstructed solution proposed in [69] is also a numerical corrector for general heterogeneities and monotone operators. No further estimate can be derived without making more specific hypotheses on the dependence of W_{ϵ} upon the space variable. Yet, for periodic homogenization of monotone operators we almost have a full error analysis, as stated in the following proposition.

Proposition 7 *We keep the notation of Lemma* 6.3 *and Theorem* 40 *and consider a periodic energy density* $W_{\epsilon}(x,\cdot) = W(\frac{x}{\epsilon},\cdot)$. *Assume* $p \geq 2$, $H1$, $H2$, $H3$, $H4$, (6.2), *and* (6.3) *with a p-structure (thus* $\alpha = 1$, $\beta = 2$*). Assume in addition that* Ω *is a convex polygon and that some technical Lipschitz hypotheses on* W_{ϵ} *hold (see* [*Hypotheses* 1, Chapter 5]*). Then* $W_{\eta,\epsilon}$ *satisfies the same technical assumptions. Let us denote by* $u^{h,\epsilon}_{\eta,H}$ *the minimizer of* (6.96), *where* V_h *and* V_H *are* $P1$*-conformal FE spaces, respectively, associated with triangulations of order* $h\eta$ *and* $H diam(\Omega)$ *(*$diam(\Omega)$ *standing for the diameter of* Ω*). Then there exist six constants independent of* H, h, η, *and* ϵ *such that*

$$\|u^{h,\epsilon}_{\eta,H} - u_{hom}\|_{1,p} \leq C_1\left(\frac{\epsilon}{\eta}\right)^{\frac{1}{p}} + C_2 h^{\frac{1}{p-1}} + C_3 H. \tag{6.99}$$

The error estimate for the discrete numerical correctors $v^{H,h,i}_{\eta,\epsilon}$ *associated with Lemma* 7 *and computed on the fine scale* $h\eta$ *now reads, provided* $H \geq \eta$,

$$\begin{aligned}&\left\|\sum_{i=1}^{I_H}\Big(\nabla v^{H,i}_{\eta,\epsilon} - (\xi_{hom,i} + \nabla p^{H,i}(\xi_{hom,i}))\Big)1_{Q_{H,i}}\right\|_{0,p}\\ &\quad\leq C_{4,q}\left[C_1\left(\frac{\epsilon}{\eta}\right)^{\frac{1}{p}} + C_2 h^{\frac{1}{p-1}} + C_3 H\right]^{\frac{1}{p-1}}\eta^{-\frac{n}{q}\frac{p-2}{p-1}} + C_6\left(\frac{\epsilon}{\eta}\right)^{\frac{1}{p}} + C_5 h^{\frac{1}{p-1}},\end{aligned} \tag{6.100}$$

which holds for all $q < \infty$ *in two dimensions, and for all* $q < 3p$ *in three dimensions,* $C_{4,q}$ *depending on the constant of continuity in the Sobolev embedding theorem.*

Remark **26** For the linear case ($p = 2$), in [1] Abdulle has derived the following sharp error estimates:

$$\begin{aligned}
&\|u_{\eta,H}^{h,\epsilon} - u_{hom}\|_{1,2} \leq C_1 \left(\frac{\epsilon}{\eta}\right) + C_2 h^2 + C_3 H, \\
&\left\| \sum_{i=1}^{I_H} \left(\nabla v_{\eta,\epsilon}^{H,i} - (\xi_{hom,i} + \nabla p^{H,i}(\xi_{hom,i})) \right) 1_{Q_{H,i}} \right\|_{0,2} \\
&\qquad \leq C_4 \left[C_1 \left(\frac{\epsilon}{\eta}\right) + C_2 h^2 + C_3 H \right] + C_6 \left(\frac{\epsilon}{\eta}\right)^{\frac{1}{2}} + C_5 h.
\end{aligned} \tag{6.101}$$

The proof in [1] uses the symmetry of the linear operator. This argument can also be used in the proof of Proposition 7 to refine the error estimate in the linear case.

Remark **27** The hypotheses of Proposition 7 are rarely fulfilled in practice since [Hypotheses 1, Chapter 5] imposes the spatial continuity of the energy (and thus of the heterogeneities). For composite materials, W_ϵ does not satisfy this assumption. However, W_{hom} and $W_{\eta,\epsilon}$ do so. We can then prove the following weaker error estimate in the case of a composite of power-law materials (that is with p-structure):

$$\begin{aligned}
&\|u_{\eta,H}^{h,\epsilon} - u_{hom}\|_{1,p} \leq C_1 \left(\frac{\epsilon}{\eta}\right)^{\frac{1}{p}} + C_2 h^{\frac{1}{p(p-1)^2}} + C_3 H, \\
&\left\| \sum_{i=1}^{I_H} \left(\nabla v_{\eta,\epsilon}^{H,i} - (\xi_{hom,i} + \nabla p^{H,i}(\xi_{hom,i})) \right) 1_{Q_{H,i}} \right\|_{0,p} \\
&\qquad \leq C_{4,q} \left[C_1 \left(\frac{\epsilon}{\eta}\right)^{\frac{1}{p}} + C_2 h^{\frac{1}{p(p-1)^2}} + C_3 H \right]^{\frac{1}{p-1}} \eta^{-\frac{n}{q}\frac{p-2}{p-1}} + C_6 \left(\frac{\epsilon}{\eta}\right)^{\frac{1}{p}} + C_5 h^{\frac{1}{p(p-1)^2}},
\end{aligned} \tag{6.102}$$

instead of (6.99) and (6.100), with the same notation as in Proposition 7.

The proof of these error estimates is sketched in the appendix. It basically amounts to combining regularity estimates and monotonicity, as in Chapter 5. We also make use of optimal convergence rates obtained in [70] using the interpolation theory in Nikolskij spaces.

The sharpness of the error estimates of Proposition 7 and Remark 27 has not been investigated in the present work.

Let us now make some remarks on the quasiconvex case. The numerical practice for such energy densities has been addressed in Chapter 5 in the setting of periodic homogenization. The method and results therein hold mutatis mutandis for nonperiodic homogenization using the framework of section 6.1.3. From a theoretical point of view, Theorem 43 is weaker than Theorem 40 since we have proved only the weak convergence of minimizers: $u_\eta^{h,\epsilon} \rightharpoonup u_{hom}$ in $W^{1,p}(\Omega)$. In addition, there exist examples for which this convergence cannot be strong and that are linked to the possible loss of strict rank-one convexity by homogenization (see [94] and Chapter 5).

6.3.2 MsFEM

Let $\{Q_{H,i}\}_i$ be a triangulation of Ω of size H. At each mesh element $Q_{H,i}$ we associate a point $x_i \in Q_{H,i}$. Given $u \in W^{1,p}(\Omega)$, we can approximate the integral $I_{\eta,\epsilon}(u)$ by

$$\sum_i |Q_{H,i}| W_{\eta,\epsilon}(x_i, \nabla u(x_i)). \tag{6.103}$$

To obtain the MsFEM starting from (6.103), we have to make specific the value of η, the boundary conditions in (6.6), and the FE space V_H. The specificity of the MsFEM with respect to the HMM is the link between η and H: η is taken to be H. The energy density used in the MsFEM is defined at x_i by

$$\inf \left\{ \frac{1}{|Q_{H,i}|} \int_{Q_{H,i}} W_\epsilon(y, \nabla v) \,|\, v(y) = \xi \cdot y \text{ on } \partial Q_{H,i} \right\} \tag{6.104}$$

and denoted by $W_{H,\epsilon}^{MsFEM}(x_i,\xi)$ in the following. Taking V_H as the space of $P1$-finite elements on the triangulation $\{Q_{H,i}\}_i$, we obtain the discrete version of the MsFEM:

$$\inf\left\{\sum_{i=1}^{N_{mesh}}|Q_{H,i}|W_{H,\epsilon}^{MsFEM}(x_i,\nabla u_H(x_i)) - \sum_{i=1}^{N_{mesh}}\sum_{j=1}^{N_{GP}} q_j f(x_{ij})u_H(x_{ij}), u_H\in V_H+BC\right\}, \tag{6.105}$$

where the second term of the energy has been integrated by a quadrature rule associated with the triangulation. Problem (6.105) is to be compared to problem (6.94).

To prove the convergence of the MsFEM, Theorem 40 is not sufficient since both the FE space V_H and the numerical energy density $W_{H,\epsilon}^{MsFEM}$ depend on H. However, an easy adaptation of the arguments of the proof of Theorem 40 allows us to conclude. The rest of this section is devoted to such a proof in the general quasiconvex case, which has not been addressed by the authors of the method in their series of papers (e.g., [72], [74], [76]). Before we get to this, let us mention that some error estimates have been derived in [72] by Efendiev, Hou, and Ginting in a nonlinear setting for the specific case of periodic homogenization of monotone operators, for which they have obtained for $\beta=p$

$$\|u_H-u_{hom}\|_{1,p}\le C\left(\frac{\epsilon}{H}\right)^{\frac{\alpha}{p(p-1)(p-\alpha)}}+C\left(\frac{\epsilon}{H}\right)^{\frac{1}{p-1}}+CH^{\frac{1}{p-1}}, \tag{6.106}$$

where u_H corresponds to an FE approximation of $u_{\eta,\epsilon}$ in the notation of section 6.1. For $\alpha=p-1$, (6.20) is equivalent to the first two terms of (6.106).

The following proof makes use of arguments that have been developed in other parts of the present work. They will only be sketched here. The new arguments are mainly linked to the convergence of the infima. We have also chosen to treat together the convergence in ϵ and H in the presentation.

Adaptation of the proof of Theorem 42. The original nonlinear MsFEM (that is, without the quadrature rule) can indeed be rewritten in the form

$$\inf\left\{\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla u_H)-fu_H\,|\,u_H\in V_H+BC\right\}, \tag{6.107}$$

if we extend the definition of $W_{H,\epsilon}^{MsFEM}$ on Ω by

$$W_{H,\epsilon}^{MsFEM}(x,\xi)=\sum_i W_{H,\epsilon}^{MsFEM}(x_i,\xi)1_{Q_{H,i}}(x),$$

since ∇u_H is constant on each $Q_{H,i}$.

Let denote by (P_H) an equicontinuous family of projectors from $W^{1,p}(\Omega)$ onto $(V_H,\|\cdot\|_{1,p})$, a family of $\mathcal{P}1$-FE spaces. In particular, there exists $C>0$ such that for all $u\in W^{1,p}(\Omega)$, $\lim_{H\to 0}\|u-P_H(u)\|_{1,p}=0$ and $\|P_H(u)\|_{1,p}\le C\|u\|_{1,p}$.

The same arguments as for the proof of Theorem 42 show that the family of energy functionals $I_{H,\epsilon}^{MsFEM}:u\mapsto\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla P_H(u))-fP_H(u)$ is equilocally Lipschitz since (P_H) is equicontinuous on $W^{1,p}(\Omega)$. It converges pointwise on $W^{1,p}(\Omega)$ to the functional $\tilde{I}_{hom}:u\mapsto\int_\Omega W_{hom}(x,\nabla u)-fu$ as ϵ and H go to 0, by application of the dominated convergence theorem. Thus $I_{H,\epsilon}^{MsFEM}$ $\Gamma(W^{1,p})$-converges to $\tilde{I}_{hom}$ by Lemma 6.8.

Next we prove the convergence of infima

$$\begin{aligned}\lim_{H\to 0}\lim_{\epsilon\to 0}&\left(\inf\left\{\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla P_H(u))-fP_H(u)|u\in W^{1,p}(\Omega)+BC\right\}\right)\\&=\inf\left\{\int_\Omega W_{hom}(x,\nabla u)-fu\,|\,u\in W^{1,p}(\Omega)+BC\right\}.\end{aligned} \tag{6.108}$$

Denoting by $\tilde{V}_H = \{v \in W^{1,p}(\Omega) \,|\, v_{|\partial Q_{H,i}}$ is linear for every $i\}$, (6.104) implies

$$\inf\left\{\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla u) - fu \,|\, u \in V_H + BC\right\}$$
$$= \inf\left\{\int_\Omega W_\epsilon(x,\nabla u) - fP_H(u) \,|\, u \in \tilde{V}_H + BC\right\}.$$

Consequently,

$$\inf\left\{\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla u) - fu \,|\, u \in V_H + BC\right\}$$
$$\geq \inf\left\{\int_\Omega W_\epsilon(x,\nabla u) - fP_H(u) \,|\, u \in W^{1,p}(\Omega) + BC\right\},$$

which implies

$$\begin{aligned}\liminf_{H\to 0}\liminf_{\epsilon\to 0}\left(\inf\left\{\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla u) - fu \,|\, u \in V_H + BC\right\}\right)\\ \geq \inf\left\{\int_\Omega W_{hom}(x,\nabla u) - fu \,|\, u \in W^{1,p}(\Omega) + BC\right\},\end{aligned} \tag{6.109}$$

since $u \mapsto P_H(u)$ is a continuous perturbation with respect to the topology of the $\Gamma(L^p)$-convergence (the liminf and the limsup inequalities still hold with the perturbation).

Conversely, we have, for any minimizer u_{hom} of the homogenized problem,

$$\begin{aligned}&\limsup_{H\to 0}\limsup_{\epsilon\to 0}\left(\inf\left\{\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla P_H(u)) - fP_H(u) | u \in W^{1,p}(\Omega) + BC\right\}\right)\\ &\leq \limsup_{H\to 0}\limsup_{\epsilon\to 0}\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla P_H(u_{hom})) - fP_H(u_{hom})\\ &= \inf\left\{\int_\Omega W_{hom}(x,\nabla u) - fu \,|\, u \in W^{1,p}(\Omega) + BC\right\}.\end{aligned} \tag{6.110}$$

The combination of (6.109) and (6.110) shows (6.108).

From this point of the proof, we have to distinguish two cases: the strictly convex case and the general case.

In the strictly convex case, the same arguments as for the proof of the strong convergence of the equi-isolated minimizers in Theorem 43 directly show the uniqueness of the minimizer for the homogenized energy and the strong convergence of the unique sequence of minimizers $u_{H,\epsilon} \in V_H + BC$ of

$$\inf\left\{\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla P_H(u)) - fP_H(u) \,|\, u \in W^{1,p}(\Omega) + BC\right\} \tag{6.111}$$

to u_{hom} when ϵ and H go to 0.

For the general quasiconvex case, let us consider a sequence of minimizers $u_{H,\epsilon} \in V_H + BC$ of (6.111) and prove that it weakly converges, up to extraction, to some minimizer u_{hom} of the homogenized problem. We skip the convergence in ϵ since the difficulty does not lie there. It remains to prove, with obvious notation, that the weak limit of $u_{H,hom} \in V_H + BC$, still denoted by u_{hom}, is a minimizer of the homogenized problem. We already know that

$$\lim_{H\to 0} I_{H,hom}^{MsFEM}(u_{H,hom}) = \inf\{\tilde{I}_{hom}(u), u \in W^{1,p}(\Omega) + BC\}. \tag{6.112}$$

We have

$$I_{H,hom}^{MsFEM}(u_{H,hom}) - \tilde{I}(u_{hom}) = I_{H,hom}^{MsFEM}(u_{H,hom}) - \tilde{I}(u_{H,hom}) + \tilde{I}(u_{H,hom}) - \tilde{I}(u_{hom}).$$

Provided that

$$\lim_{H\to 0}\left(I_{H,hom}^{MsFEM}(u_{H,hom}) - \tilde{I}(u_{H,hom})\right) = 0, \tag{6.113}$$

the weakly lower semicontinuity of $\tilde{I}_{hom}$ implies that

$$\liminf_{H\to 0} I_{H,hom}^{MsFEM}(u_{H,hom}) \geq \tilde{I}(u_{hom}),$$

which, combined with (6.112), shows that u_{hom} is a minimizer of the homogenized problem.

Actually, we do not exactly prove (6.113) for $u_{H,hom}$ but for a p-equi-integrable sequence $v_{H,hom}$ still weakly converging to u_{hom} and given by Lemma 6.11. Such a proof for a p-equi-integrable sequence is detailed twice in the appendix (sections 6.5.2 and 6.5.3) and not recalled here. We thus have $\liminf_{H\to 0} I_{H,hom}^{MsFEM}(v_{H,hom}) \geq \tilde{I}(u_{hom})$. In addition, outside a set B_H such that $\lim_{H\to 0}\mathcal{L}^n(B_H) = 0$ given by Lemma 6.11, $\nabla v_{H,hom}$ and $\nabla u_{H,hom}$ coincide, and thus

$$\begin{aligned}\int_{\Omega - B_H} W_{H,hom}^{MsFEM}(x, \nabla v_{H,hom}) &= \int_{\Omega - B_H} W_{H,hom}^{MsFEM}(x, \nabla u_{H,hom}) \\ &\leq \int_{\Omega} W_{H,hom}^{MsFEM}(x, \nabla u_{H,hom}).\end{aligned}$$

The p-equi-integrability of $v_{H,hom}$ and the growth condition (6.25) imply that

$$\limsup_{H\to 0}\int_{\Omega - B_H} W_{H,hom}^{MsFEM}(x, \nabla v_{H,hom}) = \limsup_{H\to 0}\int_{\Omega} W_{H,hom}^{MsFEM}(x, \nabla v_{H,hom}).$$

Thus, $\limsup_{H\to 0}\int_\Omega W_{H,hom}^{MsFEM}(x, \nabla v_{H,hom}) \leq \lim_{H\to 0}\int_\Omega W_{H,hom}^{MsFEM}(x, \nabla u_{H,hom})$ and

$$\begin{aligned}\tilde{I}_{hom}(u_{hom}) &\leq \liminf_{H\to 0} I_{H,hom}^{MsFEM}(v_{H,hom}) \\ &\leq \lim_{H\to 0}\int_\Omega W_{H,hom}^{MsFEM}(x, \nabla u_{H,hom}) \\ &= \inf\{\tilde{I}_{hom}(u), u \in W^{1,p}(\Omega) + BC\},\end{aligned}$$

which proves that u_{hom} is a minimizer of the homogenized functional.

As for Theorem 43, we can also prove the strong convergence of the equi-isolated minimizers. This concludes the extension of the proof of Theorem 42 to the MsFEM.

From a practical point of view, one may also introduce a parameter h to approximate the multiscale finite elements, namely the minimizer of problem (6.104). Doing so, we have an extra parameter h related to the size of the fine mesh $\tilde{h}$ of $Q_{H,i}$ by $\tilde{h} = Hh$. One can then perform a numerical analysis in the spirit of Section 6.3.1, in function of ϵ, H and h, following the analysis of Section 5.5 in the periodic case.

Several refinements of the MsFEM have been developed, both to reduce the cost [72] and the resonance errors of the original method presented above [105], [106], [107]. They have not been analyzed in the present framework, and we refer the reader to the specific literature for details on the possible improvements of the method.

In the analysis performed above, the proofs of the convergence for the MsFEM and the HMM are basically the same. Both proofs use the same arguments, and both methods can be interpreted as (different) approximations of the same averaged or relaxed energy from a variational viewpoint.

This analysis does not give any information on the implementation issues. The structure of the codes for HMM and MsFEM are very different, and the complexity may also be very different if everything is done naively according to the present (and naive) description of both methods. There is a huge body of literature dedicated to these important issues. We refer the reader to Table 6.1 and to the bibliography.

6.3.3 Summary of some numerical analysis results

In the two previous sections, we have compared the averaged energies introduced in section 6.1 to some existing numerical methods. For the sake of completeness and in order to clarify the interest of the present work besides the alternative proofs given for some well-known results, we have gathered in Table 6.1 some questions related to numerical homogenization in several settings and provided some references where each problem is addressed and (partially) solved (X stands for the present work). This table is far from being exhaustive and aims only at giving some possible entries to the literature.

	Linear operator	Monotone operator	Quasiconvex energy
Periodic			
Convergence of the method	[105], [106], [68], [1],X	[72],X	X
Error estimate $\|\cdot - u_{hom}\|_{1,p}$	[105], [106], [68], [1],X	[72],X	
Error estimate on the corrector without oversampling	[106], [68], [1],X	X	
Error estimate on the corrector with oversampling	[106], [107], [68]	X	
Stochastic			
Convergence of the method	[74], [76], [68]	[74], [76]	
Error estimate $\|\cdot - u_{hom}\|_{1,p}$	[68]		
Convergence of the corrector	[74], [76], [68]	[74], [76]	
Error estimate on the corrector (with oversampling)	[68]		
General heterogeneities			
Convergence of the method	[74],X	[74],X	X
Abstract error estimate $\|\cdot - u_{hom}\|_{1,p}$	[68],X	X	
Convergence of the corrector	X	X	
Abstract error estimate on the corrector (with oversampling)	[68]		

Table 6.1. Partial summary of some numerical analysis results available to date.

Regarding the issue related to numerical correctors and oversampling methods, the error estimate (6.23) is of interest. Except for $\beta \wedge p - \alpha = 1$ (which are the only cases numerically investigated, typically the linear case), the second term is greater than the first term in (6.23). However, only the first term can be reduced by oversampling methods. Although it has not been numerically confirmed yet, the use of oversampling methods could therefore be less efficient in nonlinear cases.

6.4 Conclusion

The setting of the present work is the minimization of an energy which depends on the gradient of a function. The prototypical example is hyperelasticity, for which the mechanical energy density of the material depends only on the gradient of deformation. More specifically we have considered an energy density that varies a lot spatially, which makes direct numerical simulations rather

impossible to perform in practice. We have then introduced an alternative energy density, which is a kind of averaged, effective, or relaxed energy density. This energy density should have the advantage not to vary as much as the original one and is therefore easier to simulate numerically. These analytical developments have allowed us to recover some convergence results for the MsFEM in the convex case and to prove the corresponding results and derive error estimates for the HMM, without spatial assumptions. We have also extended these convergence results to *nonlinear elasticity*, by considering quasiconvex energy densities. Finally, we have proved that the usual reconstruction procedure to recover the fine scale features of the solution is indeed valid under classical monotonicity hypotheses for *general heterogeneities* and not only for periodic or stochastic and stationary operators.

6.5 Appendix

Besides a remark on the effect of boundary conditions on the Γ-limit, the present appendix contains, for consistency, the proof of Lemma 6.7 in the quasiconvex case, which is essentially the same of the one for the convex case in [56, Theorem 5.14], and the proof of the classical inequalities of Lemma 6.13. We also prove Lemma 6.12.

6.5.1 Γ-limit and boundary conditions

To prove that the Γ-limit on the spaces associated with (1), (2), and (3) is the same as the Γ-limit on $W^{1,p}(\Omega)$, it suffices to exhibit a recovery sequence which belongs to the variational space associated with the boundary conditions considered since Lemma 6.2 implies the liminf inequality is valid on the whole space $W^{1,p}(\Omega)$. We refer the reader to [38, Proposition 11.7], where this issue is addressed for (1). The case (2) is a direct adaptation of the proof for (1) since as constructed in [38, Proposition 11.7], the recovery sequence can indeed by periodized if the limit u is periodic. The case (3) can be dealt with as follows: given any $u \in \{v \in W^{1,p}(\Omega) \,|\, \langle \nabla v \rangle_\Omega = \xi\}$, there exists a sequence (u_ϵ) such that $u_\epsilon \rightharpoonup u$ in $W^{1,p}(\Omega)$ and $\lim_{\epsilon\to 0} \int_\Omega W_\epsilon(x, \nabla u_\epsilon) = \int_\Omega W_{hom}(x, \nabla u)$ by the definition of Γ-convergence. Let us then consider the following sequence: $\tilde{u}_\epsilon(x) = u_\epsilon(x) + (\xi - \langle \nabla u_\epsilon \rangle_\Omega) \cdot x$. This sequence belongs to $\{v \in W^{1,p}(\Omega) \,|\, \langle \nabla v \rangle_\Omega = \xi\}$ and weakly converges to u in $W^{1,p}(\Omega)$. In addition, as a consequence of Lemma 6.6, $\|\nabla u_\epsilon - \nabla \tilde{u}_\epsilon\|_{L^p(\Omega)} = |\xi - \langle \nabla u_\epsilon \rangle_\Omega| \mathcal{L}^n(\Omega) \to 0$ implies that

$$\lim_{\epsilon\to 0} \left| \int_\Omega W_\tau(x, \nabla u_\epsilon) - \int_\Omega W_\tau(x, \nabla \tilde{u}_\epsilon) \right| = 0$$

uniformly in τ. Therefore,

$$\lim_{\epsilon\to 0} \int_\Omega W_\epsilon(x, \nabla \tilde{u}_\epsilon) = \int_\Omega W_{hom}(x, \nabla u),$$

which proves that $(\tilde{u}_\epsilon)$ is a recovery sequence in $\{v \in W^{1,p}(\Omega) \,|\, \langle \nabla v \rangle_\Omega = \xi\}$.

6.5.2 Proof of Lemma 6.7 (see [56, Theorem 5.14])

This proof requires some basic properties of Γ-convergence, which can be read in the first chapter of [36]. We recall that Lemma 6.6 also holds for rank-one functions on $\mathbb{R}^{n\times d}$ with a different constant K since any two matrices of $\mathbb{R}^{n\times d}$ are connected by at most nd rank-one matrices.

By the dominated convergence theorem, $\lim_{\epsilon\to 0} \int_\omega \tilde{W}_\epsilon(x, \nabla u) dx = \int_\omega \tilde{W}(x, \nabla u) dx$ for every $u \in W^{1,p}(\omega)$. Thus, $\Gamma - \limsup_{\epsilon\to 0} \int_\omega \tilde{W}_\epsilon(x, \nabla u) dx \le \int_\omega \tilde{W}(x, \nabla u) dx$. The conclusion is achieved if we prove that

$$\int_\omega \tilde{W}(x, \nabla u) dx \le \Gamma - \liminf_{\epsilon\to 0} \int_\omega \tilde{W}_\epsilon(x, \nabla u) dx. \tag{6.114}$$

Let us fix $u \in W^{1,p}(\Omega)$. By the absolute continuity of the integral for every $\theta > 0$ there exists $\delta > 0$ such that

$$\int_A (|\nabla u|^p + 1)dx < \theta \tag{6.115}$$

for every mesurable subset A of ω with $\mathcal{L}^N(A) < \delta$. Moreover, there exists $R > 0$ such that $\mathcal{L}^N(\{|\nabla u| \geq R\}) < \delta$.

Let $K = C((R+1)^p + 1)$ and let $\xi_1, \ldots, \xi_m$ be points in the ball $B(0,R)$ such that

$$B(0,R) \subseteq \bigcup_{i=0}^{m} B(\xi_i, \theta/K). \tag{6.116}$$

By the Egorov theorem, the sequences $(\tilde{W}_\epsilon(\cdot,\xi_i))$ converge to $\tilde{W}(\cdot,\xi_i)$ quasi-uniformly on ω. Therefore, there exist a measurable subset A of ω, with $\mathcal{L}^N(A) < \delta$, and a constant k such that $|\tilde{W}_\epsilon(x,\xi_i) - \tilde{W}(x,\xi_i)| \leq \theta$ for every $x \in \omega \setminus A$, $i = 1, \ldots, m$, and $\epsilon \leq 1/k$. By (6.116) and by Lemma 6.6, we obtain

$$|\tilde{W}_\epsilon(x,\xi) - \tilde{W}(x,\xi)| < 3\theta \tag{6.117}$$

for every $x \in \omega \setminus A$, for every $\xi \in B(0,R)$, and $\epsilon \leq 1/k$.

Let $B = A \cup \{|\nabla u| \geq R\}$, let $g : \omega \times \mathbb{R}^n \to \mathbb{R}$ be the function defined by

$$g(x,\xi) = \begin{cases} \tilde{W}(x,\xi) & \text{if } x \notin B, \\ 0 & \text{if } x \in B, \end{cases}$$

and let $G : W^{1,p}(\omega) \to \mathbb{R}$ be the corresponding integral functional defined by

$$G(u) = \int_\omega g(x, \nabla u)dx.$$

If $c = 3\mathcal{L}^N(\omega)$, (6.117) implies $\int_\omega \tilde{W}_\epsilon(x,\nabla u)dx + c\theta \geq G(u)$ for every $\epsilon \leq 1/k$. As G is lower semicontinuous for the weak topology of $W^{1,p}(\omega)$, we conclude that

$$\left(\Gamma - \liminf_{\epsilon\to 0} \int_\omega \tilde{W}_\epsilon(x,\nabla u)dx\right) + c\theta \geq G(u).$$

Since $\int_\omega \tilde{W}(x,\nabla u)dx \leq G(u) + c_1 \int_B (|\nabla u|^p + 1)dx$, from (6.115) we get

$$\int_\omega \tilde{W}(x,\nabla u)dx \leq G(u) + 2c_1\theta \leq \left(\Gamma - \liminf_{\epsilon\to 0} \int_\omega \tilde{W}_\epsilon(x,\nabla u)dx\right) + (c + 2c_1)\theta,$$

so that (6.114) can be obtained by taking the limit as θ tends to 0.

6.5.3 Proof of Lemma 6.12

We divide the proof into two steps. For almost every $x \in \mathcal{O}$ and for all $\xi \in \mathbb{R}^{n\times d}$, we first prove

$$\limsup_{\epsilon\to 0} Qf_\epsilon(x,\xi) \leq Qf(x,\xi). \tag{6.118}$$

Then we prove the converse inequality

$$\liminf_{\epsilon\to 0} Qf_\epsilon(x,\xi) \geq Qf(x,\xi). \tag{6.119}$$

Proof of inequality (6.118). For all $\epsilon > 0$, for almost every $x \in \mathcal{O}$, and for all $\xi \in \mathbb{R}^{n\times d}$, we have

$$\begin{aligned} Qf_\epsilon(x,\xi) &= \inf\left\{\int_{(0,1)^n} f_\epsilon(x,\xi+\nabla v(y))dy \,|\, v\in W_0^{1,p}((0,1)^n,\mathbb{R}^d)\right\} \\ &\le \int_{(0,1)^n} f_\epsilon(x,\xi+\nabla u(y))dy \end{aligned} \tag{6.120}$$

for every $u\in W_0^{1,p}((0,1)^n,\mathbb{R}^d)$. Using the growth condition (6.25) and the dominated convergence theorem, we have

$$\limsup_{\epsilon\to 0} Qf_\epsilon(x,\xi) \le \int_{(0,1)^n} f(x,\xi+\nabla u(y))dy. \tag{6.121}$$

Since (6.121) holds for every $u\in W_0^{1,p}((0,1)^n,\mathbb{R}^d)$, we obtain

$$\limsup_{\epsilon\to 0} Qf_\epsilon(x,\xi) \le \inf\left\{\int_{(0,1)^n} f(x,\xi+\nabla v) \,|\, v\in W_0^{1,p}((0,1)^n,\mathbb{R}^d)\right\} = Qf(x,\xi).$$

Proof of inequality (6.119). Let us prove first that for all $\epsilon>0$, for all $\xi\in\mathbb{R}^{n\times d}$, and for all Lipschitz open subsets ω of $\mathcal{O}$, there exists a sequence $\{\phi_k^\epsilon\}_k\in W^{1,p}(\omega,\mathbb{R}^d)$ such that

$$\int_\omega Qf_\epsilon(x,\xi) = \lim_{k\to\infty}\int_\omega f_\epsilon(x,\nabla\phi_k^\epsilon),$$

$\phi_k^\epsilon \rightharpoonup u_\xi$ in $W^{1,p}(\omega,\mathbb{R}^d)$, with $u_\xi(x)=\xi\cdot x$ on ω, and satisfying the following properties:

(a) there exists $\mathcal{C}\in\mathbb{R}$ such that for all $\epsilon>0$ and for all k, $\|\nabla\phi_k^\epsilon\|_{L^p(\omega)}\le \mathcal{C}(1+|\xi|^p)$;
(b) for all ϵ and k, $\|\phi_k^\epsilon - u_\xi\|_{L^p(\omega)}\le \frac{1}{k}$.

To this aim, we show that the sequence ϕ_k^ϵ given by Lemma 6.10 satisfies (a) and (b), up to extraction. Property (b) is a direct consequence of the convergence of ϕ_k^ϵ to u_ξ in $L^p(\omega,\mathbb{R}^d)$. Using the growth condition (6.25), we have $\int_\omega Qf_\epsilon(x,\xi)\le\int_\omega f_\epsilon(x,\xi)\le C(1+|\xi|^p)\mathcal{L}^n(\omega)$ and $\int_\omega f_\epsilon(x,\nabla\phi_k^\epsilon)\ge c\|\nabla\phi_k^\epsilon\|^p_{L^p(\omega)}$, where the constants c and C do not depend on ϵ, which implies property (a) for some $\mathcal{C}\in\mathbb{R}$.

As Qf_ϵ satisfies (6.25) uniformly in ϵ (the lower bound in (6.25) is a convex function lower than or equal to f_ϵ, and thus it is also lower than or equal to its quasiconvex envelope), there exists a subsequence ϵ_j and $L\in\mathbb{R}$ such that $\lim_{j\to\infty}\int_\omega Qf_{\epsilon_j}(x,\xi)=L$. Therefore, $\lim_{j\to\infty}\lim_{k\to\infty}\int_\omega f_{\epsilon_j}(x,\nabla\phi_k^{\epsilon_j})$ exists, and a diagonal extraction argument shows there exists an extraction function π such that $\lim_{k\to\infty}\pi(k)=+\infty$ and

$$\lim_{k\to\infty}\int_\omega f_{\epsilon_{\pi(k)}}(x,\nabla\phi_k^{\epsilon_{\pi(k)}}) = L.$$

As the sequence $\phi_k^{\epsilon_{\pi(k)}}$ satisfies properties (a) and (b), it is bounded in $W^{1,p}(\omega,\mathbb{R}^d)$ and satisfies $\lim_{k\to\infty}\phi_k^{\epsilon_{\pi(k)}}=u_\xi$ in $L^p(\omega,\mathbb{R}^d)$. Up to a further extraction, we may suppose that $\phi_k^{\epsilon_{\pi(k)}}$ converges weakly to u_ξ in $W^{1,p}(\omega,\mathbb{R}^d)$ by the uniqueness of the limit in the sense of distributions.

Applying now Lemma 6.11, we obtain the existence of a sequence $\chi_k\in W^{1,p}(\omega,\mathbb{R}^d)$ such that $\nabla\chi_k$ is p-equi-integrable, and χ_k weakly converges to u_ξ in $W^{1,p}(\omega,\mathbb{R}^d)$ and satisfies

$$\lim_{k\to\infty}\mathcal{L}^N(\{x\in\omega:\nabla\chi_k(x)\ne\nabla\phi_k^{\epsilon_{\pi(k)}}(x)\})=0, \tag{6.122}$$

up to a further extraction.

Since $f_\epsilon\ge 0$,

$$\begin{aligned} \lim_{k\to\infty}\int_\omega f_{\epsilon_{\pi(k)}}(x,\nabla\phi_k^{\epsilon_{\pi(k)}}) &\ge \limsup_{k\to\infty}\int_{\{x\in\omega:\nabla\chi_k(x)=\nabla\phi_k^{\epsilon_{\pi(k)}}(x)\}} f_{\epsilon_{\pi(k)}}(x,\nabla\chi_k) \\ &\ge \limsup_{k\to\infty}\int_\omega f_{\epsilon_{\pi(k)}}(x,\nabla\chi_k), \end{aligned} \tag{6.123}$$

the last inequality being a consequence of the p-equi-integrability of $\nabla\chi_k$, the growth condition (6.25), and (6.122).

Next we prove that

$$\lim_{\epsilon\to 0}\int_\omega f_\epsilon(x,\nabla\chi_k)=\int_\omega f(x,\nabla\chi_k) \tag{6.124}$$

uniformly in k. From (6.124), we will deduce

$$L\geq \limsup_{k\to\infty}\int_\omega f_{\epsilon_{\pi(k)}}(x,\nabla\chi_k)\geq \liminf_{k\to\infty}\int_\omega f(x,\nabla\chi_k)\geq \int_\omega Qf(x,\xi) \tag{6.125}$$

since $v\mapsto \int_\omega Qf(x,\nabla v)$ is the lower semicontinuous envelope of $v\mapsto\int_\omega f(x,\nabla v)$ for the weak topology of $W^{1,p}(\omega,\mathbb{R}^d)$ and $\chi_k\rightharpoonup u_\xi$ in $W^{1,p}(\omega,\mathbb{R}^d)$.

The proof of (6.124) follows the first part of the proof of Lemma 6.7, and it relies on the p-equi-integrability of $\nabla\chi_k$ and on the fact that (f_ϵ) are equi-Lipschitz functions.

The p-equi-integrability of $\nabla\chi_k$ implies that for all $\theta>0$ there exists $\delta>0$ such that for all $k>0$, $\int_B(|\nabla\chi_k|^p+1)dx<\theta$ for every measurable subset B of ω with $\mathcal{L}^N(B)<\delta$. Moreover, as $\nabla\chi_k$ is a bounded sequence in $L^p(\omega)$, there also exists $R>0$ such that for all $k>0$, $\mathcal{L}^N\{x\in\omega:|\nabla\chi_k(x)\geq R|\}<\delta$.

By assumption, there exists $K>0$ such that for all $\epsilon>0$, f_ϵ and f are K-Lipschitz on $B(0,R)$ for almost every $x\in\omega$. Let $\xi_1,\dots,\xi_m$ be points in the ball $B(0,R)$ such that

$$B(0,R)\subseteq\bigcup_{i=0}^m B(\xi_i,\theta/K). \tag{6.126}$$

By the Egorov theorem, the sequences $f_\epsilon(\cdot,\xi_i)$ converge to $f(\cdot,\xi_i)$ quasi-uniformly on ω. Therefore, there exist a measurable set A of ω, with $\mathcal{L}^N(A)<\delta$, and an integer κ such that $|f_\epsilon(x,\xi_i)-f(x,\xi_i)|\leq\theta$ for every $x\in\omega\setminus A$, $i=1,\dots,m$, and $\epsilon\leq 1/\kappa$. By (6.126) and by the K-Lipschitz properties of f_ϵ and f, we obtain

$$|f_\epsilon(x,\xi)-f(x,\xi)|<3\theta \tag{6.127}$$

for every $x\in\omega\setminus A$, for every $\xi\in B(0,R)$, and $\epsilon\leq 1/\kappa$.

The dominated convergence theorem, inequality (6.127), the majoration in (6.25) with the constant C, and the definition of A and R imply that for all $\theta>0$, there exists $\kappa\in\mathbb{N}$ such that for all $\epsilon<1/\kappa$ and $k>0$,

$$\int_\omega|f_\epsilon(x,\nabla\chi_k)-f(x,\nabla\chi_k)|\leq(3\mathcal{L}^N(\omega)+2C)\theta, \tag{6.128}$$

which proves (6.124).

Finally, as (6.125) holds for any converging subsequence of $\int_\omega Qf_{\epsilon_j}(x,\xi)$ and for any Lipschitz open subset ω of $\mathcal{O}$, we obtain (6.119). This concludes the proof of Lemma 6.12.

6.5.4 Proof of Lemma 6.13

This proof is based on elementary calculus:

$$\begin{aligned}|W(x,\xi_1)-W(x,\xi_2)|&=\left|\int_0^1 a(x,\xi_1+t(\xi_2-\xi_1))\cdot(\xi_2-\xi_1)\mathrm{d}\,t\right|\\&\leq C|\xi_2-\xi_1|(1+|\xi_1|^{p-1}+|\xi_2-\xi_1|^{p-1})\\&\leq C|\xi_2-\xi_1|(1+|\xi_1|^{p-1}+|\xi_2|^{p-1})\end{aligned}$$

for all $\alpha\geq 0$ in (6.2).

We now prove inequality (6.46). Let u minimize $\inf\left\{\int_\Omega W(x,\nabla v)\,|\,v\in K\right\}$. Since for every $v\in K$, $\int_\Omega a(x,\nabla u)\cdot(\nabla v-\nabla u)\geq 0$ (6.46) is a consequence of the following inequality:

$$\begin{aligned}\int_\Omega W(x,\nabla v) - \int_\Omega W(x,\nabla u) - \int_\Omega a(x,\nabla u)\cdot(\nabla v - \nabla u)\\ \geq c\|\nabla v - \nabla u\|^{\beta\wedge p}_{L^p(\Omega)}\left(\mathcal{L}^n(\Omega)^{\frac{1}{p}} + \|\nabla u\|_{L^p(\Omega)} + \|\nabla v\|_{L^p(\Omega)}\right)^{p-\beta\wedge p}.\end{aligned} \tag{6.129}$$

For every $v \in K$, let us introduce the function

$$g : [0,1] \to \mathbb{R}, t \mapsto g(t) = \int_\Omega a(x, \nabla u + t(\nabla v - \nabla u)).$$

This function is real, differential, and convex. As

$$g'(t) = \int_\Omega a(x, \nabla u + t(\nabla v - \nabla u))\cdot(\nabla v - \nabla u),$$

we have, using, respectively, (6.3) and Lemma 6.14,

$$\begin{aligned}&\int_\Omega W(x,\nabla v) - \int_\Omega W(x,\nabla u) - \int_\Omega a(x,\nabla u)\cdot(\nabla v - \nabla u)\\ &= g(1) - g(0) - g'(0)\\ &= \int_0^1 (g'(t) - g'(0))\mathrm{d}\,t\\ &= \int_0^1\int_\Omega (a(x,\nabla u + t(\nabla v - \nabla u)) - a(x,\nabla u))\cdot t(\nabla v - \nabla u)\frac{\mathrm{d}\,t}{t}\\ &\geq c\int_0^1\int_\Omega (1 + |\nabla u + t(\nabla v - \nabla u)| + |\nabla u|)^{p-\beta\wedge p}|\nabla v - \nabla u|^{\beta\wedge p}t^{\beta\wedge p-1}\mathrm{d}\,t\\ &\geq c\int_0^1\int_\Omega (1 + t|\nabla u| + t|\nabla v|)^{p-\beta\wedge p}|\nabla v - \nabla u|^{\beta\wedge p}t^{\beta\wedge p-1}\mathrm{d}\,t\\ &\geq \frac{c}{4(\beta\wedge p)}\int_\Omega (1 + |\nabla u| + |\nabla v|)^{p-\beta\wedge p}|\nabla v - \nabla u|^{\beta\wedge p}\\ &\geq c\|\nabla v - \nabla u\|^{\beta\wedge p}_{L^p(\Omega)}\left(\mathcal{L}^n(\Omega)^{\frac{1}{p}} + \|\nabla u\|_{L^p(\Omega)} + \|\nabla v\|_{L^p(\Omega)}\right)^{p-\beta\wedge p}\end{aligned}$$

since property (6.3) also implies $(a(x,\xi_1) - a(x,\xi_2))\cdot(\xi_1 - \xi_2) \geq c|\xi_1 - \xi_2|^p$ for $\beta \leq p$. This proves (6.129) and consequently (6.46).

6.5.5 Sketch of proof of Proposition 6

We refer the reader to the numerical analysis of Chapter 5 for the details of the proof. We focus here only on the differences with Chapter 5.

For p-structures energy densities with regular spatial dependence, Ebmeyer and Liu [70] have proved the following error estimates for conforming $P1$-finite elements on a convex domain Ω. With obvious notation,

$$\|u - u_h\|_{1,p} \leq Ch.$$

Inserting this error estimate into the proof of [Theorem 35, Chapter 5], we obtain the second term of (6.99). As W_{hom} is homogeneous in space, it also satisfies the assumptions made in [70]. This gives the third term of (6.99). Error estimate (6.100) is based on the same arguments, noticing that $W_{\eta,\epsilon}$ also satisfies the assumptions of [70] since the continuity is preserved (and reinforced) by averaging. Using the regularity estimate of [71], we obtain that $\{u_{\eta,\epsilon}\}$ is bounded in $W^{1+\frac{2}{p}-\tau,p}(\Omega)$

for all $\tau > 0$ which explains the origin and the use of the Sobolev embedding theorem, and the dependence on the dimension.

In the case of a composite material, W_ϵ does not satisfy the assumptions of [70] since it is not regular enough spatially. However, Savaré has proved in [169] that the solution $u_\epsilon \in W^{1+\frac{1}{p}-\tau,p}(\Omega)$ for all $\tau > 0$; therefore one can still use Lemma 5.4 of Chapter 5 and obtain the second term of the first inequality in (6.102). The third term is a consequence of the spatial independence of W_{hom}, and the other terms more generally a consequence of the regularity obtained by averaging the energy density, fulfilling therefore the assumptions of [70].

7

An analytical framework for numerical homogenization - Part II: windowing and oversampling

Summary. In Chapter 6, we have introduced an analytical framework to study the convergence properties of some numerical homogenization methods for elliptic problems. In the applications however, these methods are coupled with windowing or oversampling techniques. In the present work, the author addresses this issue within the latter framework and proves the convergence of the methods with windowing, for convex and quasiconvex energies, in the context of general heterogeneities. This analysis provides us with an interesting variational interpretation of the Petrov-Galerkin formulation of the nonconforming multiscale finite element method for periodic problems.

7.1 Introduction

The goal of this chapter is to continue the analysis of multiscale methods for the numerical homogenization of elliptic equations initiated in Chapter 5 and Chapter 6, where the convergence and the numerical analysis of some numerical homogenization methods are addressed under quite general hypotheses on the heterogeneities (general spatial dependence of the operator) and on the nature of the operator (convex or quasiconvex associated energy density). In practice however, these methods are usually combined with more sophisticated techniques such as windowing and oversampling. Windowing techniques basically amounts to imposing boundary conditions further from the region of interest to minimize their effects. It gives rise to oversampling techniques for the MsFEM and is also used in the HMM. It is implicitly used or refered to in [8], [105], [106], [72], [74], [76], and theoretically analyzed in [108], [73] and [69] in the linear and periodic or stochastic cases, for which error estimates are also provided ([73, 108] for the periodic case and [69] for the stochastic case). In the present work, we focus on proving the convergence of some numerical homogenization methods with windowing under the same general hypotheses as in Chapter 6, concerning both the approximation of the homogenized solution and the corrector whenever the notion is well-established. The chapter is organized as follows. In the first Section, we very briefly recall the context, the main results of Chapter 6 and two numerical methods. Then we discuss the issue of windowing in the periodic case, before addressing it in a more general setting in Section 7.4. We also give a variational interpretation of the nonconforming Petrov-Galerkin formulation of the multiscale finite element method that allows us to prove its convergence in a general setting. Some arguments and proofs are only sketched whenever they are mainly based on the corresponding ones in Chapter 6. We refer the reader to Section 6.1 and the references therein for further details on notations and useful results that may not be extensively recalled in Section 7.2.

7.2 Numerical homogenization methods

Let us recall the analytical framework introduced in Chapter 6, and two numerical methods to which the analysis applies: the Heterogeneous Multiscale Method (HMM) and the Multiscale Finite Element Method (MsFEM).

7.2.1 Main notations

Let us introduce the major notations used in the sequel. Given a function v defined on an open set $\mathcal{O}$, we denote by

$$\langle v \rangle_{\mathcal{O}} = \frac{1}{|\mathcal{O}|} \int_{\mathcal{O}} v.$$

For a metric space (V, d), we say that a functional $F_\epsilon : V \to \mathbb{R}$ $\Gamma(d)$-converges to a functional $F : V \to \mathbb{R}$ if for all $v \in V$

$$F(v) = \inf_{\{v_\epsilon\} \in V^{\mathbb{N}}, d(v_\epsilon - v) \to 0} \{\liminf_{\epsilon \to 0} F_\epsilon(v_\epsilon)\} = \inf_{\{v_\epsilon\} \in V^{\mathbb{N}}, d(v_\epsilon - v) \to 0} \{\limsup_{\epsilon \to 0} F_\epsilon(v_\epsilon)\}$$

and we denote it by $F = \Gamma(d) - \lim_{\epsilon \to 0} F_\epsilon$. We refer the reader to Section 6.1 and the references therein for some useful properties of Γ-convergence. The metric space we will use to state the Γ-convergence results are either Sobolev spaces or Lebesgue spaces. In particular we will denote by

$$W^{1,p}_{\#}(Q) = \{v_{|Q}, v \in W^{1,p}_{loc}(\mathbb{R}^n), v(x) = v(x + Q) \text{ for almost every } x \in \mathbb{R}^n\},$$

where Q is a hypercube.

We also adopt generic conventions for the following symbols:

- Ω: open bounded domain of $\mathbb{R}^n$;
- x: generic point in Ω (or more generally in $\mathbb{R}^n$);
- $C(x, \eta)$: cube of length $\eta > 0$ centered at point x;
- y: generic point in $C(x, \eta)$;
- u: (with various indices) function in $W^{1,p}(\Omega)$;
- v: (with various indices) function in $W^{1,p}(C(x, \eta))$, for some $x \in \Omega$ and $\eta > 0$.

7.2.2 Minimization problem

In what follows, we consider minimization problems, or the associated Euler-Lagrange equations whenever the two approaches are equivalent, e.g. for monotone operators. The problem under investigation is

$$\inf \left\{ \int_\Omega W_\epsilon(x, \nabla u) dx, u \in W^{1,p}(\Omega, \mathbb{R}^d) + BC \right\}, \tag{7.1}$$

where W_ϵ is a family of energy densities and BC denotes classical boundary conditions (let say Dirichlet or mixed Dirichlet-Neumann boundary conditions, see page 98, Chapter 6) in weak form.

Denoting by $I_\epsilon : W^{1,p}(\Omega) \to \mathbb{R}, u \mapsto I_\epsilon(u) = \int_\Omega W_\epsilon(x, \nabla u) dx$, we want to study the behaviour of I_ϵ when ϵ vanishes and to numerically approximate the minimizers of I_ϵ on sets of prescribed boundary conditions.

Under the following sets of hypotheses on W_ϵ, both issues can be answered positively:

The convex case

- H1: W_ϵ is a Carathéodory function;
- H2: for almost every $x \in \mathbb{R}^n$, $W_\epsilon(x, \cdot)$ is convex on $\mathbb{R}^{d \times n}$;
- H3: there exist $0 < c \leq C$ and $p \geq 1$ such that

$$c|\xi|^p - 1 \leq W_\epsilon(x, \xi) \leq C(1 + |\xi|^p)$$

for almost all $x \in \mathbb{R}^n$ and for all $\xi \in \mathbb{R}^n$.

The quasiconvex case: $n > 1$ and $d > 1$

- H1;
- H4: for almost every $x \in \mathbb{R}^n$, $W_\epsilon(x, \cdot)$ is quasiconvex on $\mathbb{R}^{d\times n}$;
- H3.

Assumption H4 generalizes Assumption H2. An energy density satisfying H1-H4-H3 will be refered to as a standard energy density. The direct method of the calculus of variations shows that the minimization problem (7.1) has at least one solution $u_\epsilon \in W^{1,p}(\Omega, \mathbb{R}^d)$.

In addition, there exist a standard energy density W_{hom} and the associated energy functional $I_{hom} : W^{1,p}(\Omega) \to \mathbb{R}, u \mapsto I_{hom}(u) = \int_\Omega W_{hom}(x, \nabla u)dx$, such that, up to extraction, $I_{hom} = \Gamma(L^p) - \lim_{\epsilon\to 0} I_\epsilon$. For every sequence of minimizers u_ϵ, there exists a minimizer u_{hom} of I_{hom} on the same set of prescribed boundary conditions such that $u_\epsilon \rightharpoonup u_{hom}$ in $W^{1,p}(\Omega)$. In what follows we will consider Γ-converging energies, without loss of generality up to extraction.

The aim of the following section is to recall the definition of an averaged energy density $I_{\eta,\epsilon}$ that approximates I_{hom} and u_{hom} in the sense of Γ-convergence.

7.2.3 Averaged energy densities

For convex standard energy densities W_ϵ, we set the following

Definition 31 *For any $\eta > 0$, denoting by $B(x, \eta)$ the ball of radius η centered at point $x \in \mathbb{R}^n$, we define the energy density*

$$W_{\eta,\epsilon}(x, \xi) = \inf \left\{ \langle W_\epsilon(\cdot, \nabla v(\cdot))\rangle_{B(x,\eta)} \,|\, v \in W^{1,p}(B(x, \eta)), \langle \nabla v\rangle_{B(x,\eta)} = \xi \right\} \tag{7.2}$$

from $\mathbb{R}^n \times \mathbb{R}^{d\times n}$ to $\mathbb{R}$ and the associated energy functional

$$I_{\eta,\epsilon}(u) = \int_\Omega W_{\eta,\epsilon}(x, \nabla u) \quad \text{for all } u \in W^{1,p}(\Omega).$$

whereas for quasiconvex standard energy densities, we set

Definition 32 *For $\eta > 0$, let us denote by $C(x, \eta)$ the hypercube of $\mathbb{R}^n$ centered in $x \in \mathbb{R}^n$ and of length η. We then define the averaged energy density by*

$$\mathcal{W}_{\eta,\epsilon}(x, \xi) = \inf \left\{ \langle W_\epsilon(\cdot, \xi + \nabla v(\cdot))\rangle_{C(x,\eta)} \,|\, v \in W^{1,p}_{\#}(C(x, \eta), \mathbb{R}^d) \right\} \tag{7.3}$$

from $\mathbb{R}^n \times \mathbb{R}^{d\times n}$ to $\mathbb{R}$ and the energy functional associated with its quasiconvex envelope $Q\mathcal{W}_{\eta,\epsilon}$

$$I_{\eta,\epsilon}(u) = \int_\Omega Q\mathcal{W}_{\eta,\epsilon}(x, \nabla u) \quad \text{for all } u \in W^{1,p}(\Omega, \mathbb{R}^d).$$

We then have the following convergence theorem for strictly convex energy densities

Theorem 44 *Let W_ϵ satisfy H1, H2 (strictly), and H3 uniformly for $p > 1$, then the energy densities $W_{\eta,\epsilon}$ also satisfy H1, H2 (strictly), and H3, and the energy $I_{\eta,\epsilon}$ $\Gamma(L^p)$-and $\Gamma(W^{1,p})$-converges to I_{hom} as ϵ and η go to 0. Therefore, the unique sequence $u_{\eta,\epsilon}$ of minimizers of $I_{\eta,\epsilon}$ on $W^{1,p}(\Omega, \mathbb{R}) + BC$ strongly converges in $W^{1,p}(\Omega, \mathbb{R}^n)$ to the unique minimizer u_{hom} of I_{hom} on $W^{1,p}(\Omega, \mathbb{R}) + BC$.*

and the corresponding one for quasiconvex energy densities

Theorem 45 *Let W_ϵ satisfy H1, H4, and H3 uniformly for $p > 1$, then the energy densities $Q\mathcal{W}_{\eta,\epsilon}$ are standard energy densities and $I_{\eta,\epsilon}$ $\Gamma(L^p)$- and $\Gamma(W^{1,p})$-converges to I_{hom} as ϵ and η go to 0. Therefore, for any sequence $u_{\eta,\epsilon}$ of minimizers of $I_{\eta,\epsilon}$ on $W^{1,p}(\Omega, \mathbb{R}^d) + BC$, there exists a minimizer u_{hom} of I_{hom} on $W^{1,p}(\Omega, \mathbb{R}^d) + BC$ such that*

$$\lim_{\eta\to 0} \lim_{\epsilon\to 0} u_{\eta,\epsilon} = u_{hom} \quad \text{weakly in } W^{1,p}(\Omega, \mathbb{R}^d), \tag{7.4}$$

up to extraction.

Remark **28** *The trial space $W^{1,p}_{\#}(C(x,\eta),\mathbb{R}^d)$ can be replaced by $W^{1,p}_0(C(x,\eta),\mathbb{R}^d)$ without changing the convergence results of Theorems 44 and 45.*

Remark **29** *The order of the limits in (7.4) is important and cannot be changed in general.*

7.2.4 Numerical corrector and fine scale features

The approximation of the homogenized solution u_{hom} in $W^{1,p}(\Omega)$ is not enough to approximate u_ϵ in $W^{1,p}(\Omega)$ since u_ϵ only converges weakly to u_{hom} in $W^{1,p}(\Omega)$. To this aim, numerical correctors have been widely introduced and used to approximate ∇u_ϵ in $L^p(\Omega)$, which describes the fine scale features of the solution. Their convergence properties have been analyzed for general heterogeneities and monotone operators in Chapter 6.

Definition 33 *Let $\{Q_{H,i}\}_{i\in[\![1,I_H]\!]}$ be a partition of Ω in disjoint subdomains of diameter of order H. Keeping the notation of Theorem 44, for all i, we define the numerical correctors $v^{H,i}_{\eta,\epsilon}$ for a strictly convex energy density as the unique minimizers (up to a constant) of*

$$\inf\left\{\int_{Q_{H,i}} W_\epsilon(x,\nabla v)\,|\, v\in W^{1,p}(Q_{H,i}), \langle\nabla v\rangle_{Q_{H,i}} = \langle\nabla u_{\eta,\epsilon}\rangle_{Q_{H,i}}\right\}. \tag{7.5}$$

The following convergence result holds (Theorem 41, Chapter 6):

Theorem 46 *In addition to $H1$, $H2$, and $H3$, let us assume that $p\geq 2$, that $W_\epsilon(x,\cdot)$ is continuously differentiable for almost all $x\in\Omega$ and $a_\epsilon(\cdot,0)=\frac{\partial W_\epsilon}{\partial\xi}(\cdot,0)$ is bounded, and that the following monotonicity and continuity properties hold:*

$$\begin{aligned}&\exists\ 0\leq\alpha\leq p-1, C>0 \quad | \quad \textit{for almost all } x\in\mathbb{R}^n,\ \textit{for all } \xi_1,\xi_2\in\mathbb{R}^n,\\ &\qquad |a_\epsilon(x,\xi_1)-a_\epsilon(x,\xi_2)|\leq C(1+|\xi_1|+|\xi_2|)^{p-1-\alpha}|\xi_1-\xi_2|^\alpha,\end{aligned} \tag{7.6}$$

$$\begin{aligned}&\exists\ 2\leq\beta<+\infty, c>0 \quad | \quad \textit{for almost all } x\in\mathbb{R}^n,\ \textit{for all } \xi_1,\xi_2\in\mathbb{R}^n,\\ &\qquad (a_\epsilon(x,\xi_1)-a_\epsilon(x,\xi_2),\xi_1-\xi_2)\geq c(1+|\xi_1|+|\xi_2|)^{p-\beta}|\xi_1-\xi_2|^\beta.\end{aligned} \tag{7.7}$$

Then, denoting by u_ϵ the unique minimizer of I_ϵ on $W^{1,p}(\Omega)+BC$, we have

$$\lim_{\eta,H\to 0}\lim_{\epsilon\to 0}\left\|\nabla u_\epsilon-\sum_{i=1}^{I_H}\nabla v^{H,i}_{\eta,\epsilon}1_{Q_{H,i}}\right\|_{L^p(\Omega)}=0. \tag{7.8}$$

Let us briefly discuss the link between the original Tartar's correctors and the numerical correctors of Definition 33. We first recall the definition of Tartar's correctors [149] in the linear case and make some comments about the numerical interest of such a result.

Let $A_{\text{hom}}\in L^\infty(\Omega,\mathcal{M}_n(\mathbb{R}))$ be the homogenized matrix of an H-converging sequence A_ϵ (see [149] e.g.), and u_{hom} be the solution of the homogenized problem $\inf\{\int_\Omega\nabla u\cdot A_{hom}\nabla u-fu, u\in H^1(\Omega)+BC\}$. The convergence of u_ϵ to u_{hom} is only weak in $H^1(\Omega)$. The corrector matrices are designed to approximate the gradient of u_ϵ by a function depending linearly on ∇u_{hom}. Given compactly supported subsets $\omega\subset\subset\omega_1\subset\subset\Omega$, a corrector matrix $P_\epsilon\in H^1(\omega)^n$ is defined by its entries $P_\epsilon\cdot e_j=P^j_\epsilon\in H^1(\omega)$, where P^j_ϵ is the restriction on ω of the unique solution of

$$\inf\left\{\int_{\omega_1}\nabla v_\epsilon\cdot A_\epsilon\nabla v_\epsilon-e_j\cdot A_{hom}\nabla v_\epsilon, v_\epsilon(y)=e_j\cdot y+w_\epsilon, w_\epsilon\in H^1_0(\omega_1)\right\}, \tag{7.9}$$

for e_j describing the canonical basis of $\mathbb{R}^n$. A corrector for u_ϵ is then given on ω by

$$C_\epsilon=\sum_j(\nabla u_{hom}\cdot e_j)\nabla P^j_\epsilon.$$

As a sum of products of two L^2-functions, $C_\epsilon \in L^1(\omega)^n$, and we have ([149, Theorem 3 pp. 39])

$$\lim_{\epsilon \to 0} \|C_\epsilon - \nabla u_\epsilon\|_{L^1(\omega)} = 0.$$

It is worth noticing that correctors are not gradient fields in general. In addition, correctors are not unique.

From a computational point of view, Tartar's correctors are too abstract since the precise knowledge of A_{hom} is required to calculate the correctors, whereas A_{hom} is in principle unknown. However, if A_{hom} is constant, then (7.9) turns out to be solvable in practice (the term depending on A_{hom} is constant in the energy). As pointed out by Allaire and Brizzi in [8], the simplest approximation of A_{hom} is the piecewise constant approximation.

The numerical corrector amounts to taking $\omega = \omega_1 = Q_{H,i}$ and approximating A_{hom} by $\langle A_{\eta,\epsilon} \rangle_{Q_{H,i}}$. It should be noticed that ω is not compactly supported in ω_1. In addition to the convergence in ϵ (and η), there is an error linked to H and the piecewise constant approximation of A_{hom}. Up to an error which depends on H, the numerical corrector may be interpreted as an approximation of some Tartar's corrector on $Q_{H,i}$. In addition, Theorem 46 shows that the convergence of the numerical corrector holds in $L^2(\Omega)$. Imposing $\omega \subset\subset \omega_1$ in the definition of a numerical corrector is a way to introduce windowing, as it will be seen in Section 7.4.

7.2.5 HMM

The application of the Heterogeneous Multiscale Method to elliptic problems introduced in [68] can be interpreted as the minimization of a discretization of $I_{\eta,\epsilon}$ plus a lower order term $f \in L^{p'}(\Omega)$ on a finite element basis, which reads

$$\inf \left\{ \sum_{i=1}^{I_H} \sum_{j=1}^{N_{GP}} q_j \big(W_{\eta,\epsilon}(x_{ij}, \nabla u_H(x_{ij})) - f(x_{ij}) u_H(x_{ij}) \big), u_H \in V_H + BC \right\}, \tag{7.10}$$

where I_H is the number of mesh elements, N_{GP} is the number of Gauss points per element, x_{ij} are the Gauss points, q_j are the weights, and V_H is a FE space. Then the computation of the FE minimizer of (7.10) requires only evaluations of derivatives of $W_{\eta,\epsilon}(x_{ij}, \xi)$ for particular ξ at Gauss points x_{ij}. We refer the reader to [68] and Chapter 6 for details on the method and its analysis.

7.2.6 MsFEM

The Multiscale Finite Element Method may also be interpreted as the minimization of a (different) discretization of $I_{\eta,\epsilon}$ on a $P1$-finite element space V_H associated with a triangulation $\{Q_{H,i}\}_i$ of Ω, namely:

$$\begin{aligned} \inf \Bigg\{ \sum_{i=1}^{I_H} |Q_{H,i}| W_{H,\epsilon}^{MsFEM}(x_i, \nabla u_H(x_i)) \\ - \sum_{i=1}^{I_H} \sum_{j=1}^{N_{GP}} q_j f(x_{ij}) u_H(x_{ij}), u_H \in V_H + BC \Bigg\}, \end{aligned} \tag{7.11}$$

where x_i typically denotes the center of $Q_{H,i}$, the second term of the energy has been integrated by a quadrature rule on Gauss points x_{ij}, and

$$W_{H,\epsilon}^{MsFEM}(x_i, \xi) = \inf \left\{ \frac{1}{|Q_{H,i}|} \int_{Q_{H,i}} W_\epsilon(y, \nabla v) \,|\, v(y) = \xi \cdot y \text{ on } \partial Q_{H,i} \right\}, \tag{7.12}$$

which is a particular energy density of type $W_{\eta,\epsilon}$.

In the analysis of the MsFEM in Section 6.3.2, we have extensively used the following rewriting of the problem:

$$\inf\left\{\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla u_H) - f u_H \,|\, u_H \in V_H + BC\right\}, \tag{7.13}$$

extending the definition of $W_{H,\epsilon}^{MsFEM}$ on Ω by

$$W_{H,\epsilon}^{MsFEM}(x,\xi) = \sum_i W_{H,\epsilon}^{MsFEM}(x_i,\xi) 1_{Q_{H,i}}(x),$$

since ∇u_H is constant on each $Q_{H,i}$. With this formulation, estimates on the energy are easy to obtain due to the inclusion $V_H \subset W^{1,p}(\Omega)$.

To relate this formulation with the original formulation of the MsFEM, it is enough to notice that for any $u_H \in V_H$ one can define the restriction of the associated multiscale finite element $u_{H,\epsilon}^{MsFEM}$ (at least in the monotone case) on each mesh element $Q_{H,i}$ as the unique solution of

$$\inf\left\{\frac{1}{|Q_{H,i}|}\int_{Q_{H,i}} W_\epsilon(y,\nabla v) \,|\, v(y) = \nabla u_H \cdot y \text{ on } \partial Q_{H,i}\right\}.$$

We then have

$$\int_\Omega W_{H,\epsilon}^{MsFEM}(x,\nabla u_H) = \int_\Omega W_\epsilon(x,\nabla u_{H,\epsilon}^{MsFEM}).$$

Other numerical methods can be recast within this analytical framework, such as the residual-free bubbles finite element method introduced in [168], where the duality of points of view $(W_{H,\epsilon}^{MsFEM}, V_H)$ and $(W_\epsilon, \{u_{H,\epsilon}^{MsFEM}\})$ is pointed out.

7.3 Windowing in the periodic case

As a matter of fact, windowing is often used even if not always mentioned. The present terminology is borrowed from the mechanical community. Applied to the MsFEM, windowing is refered to as oversampling. Combined with HMM, it has no particular name and we will make use of the mechanical terminology.

7.3.1 Setting of the problem

Since the homogenized equation is explicitly known when $W_\epsilon(\cdot,\cdot) = W(\frac{\cdot}{\epsilon},\cdot)$ and W is 1-periodic in space, the application of the MsFEM or the HMM strategies to this case allows us to perform a numerical analysis of the methods.

Doing so ([108] for MsFEM and [69, (1.8')] for HMM e.g.), the error between the numerical approximation and the solution of the homogenized problem is proved to exhibit some term called the cell resonance and boundary layer error. This error is linked to two phenomena:

- when $C(x,\eta)$ is not a multiple of the periodic cell on the one hand (cell resonance), and
- when linear Dirichlet boundary conditions are used on the other hand (the cell problem in periodic homogenization is completed by periodic boundary conditions).

These phenomena are the sources of a boundary layer, which gives the following error in the linear case between the solutions of the homogenized and averaged problems

$$\|u_{hom} - u_{\eta,\epsilon}\|_{H^1(\Omega)} \le C\frac{\epsilon}{\eta} \tag{7.14}$$

(the same estimate holds for the error between the homogenized and averaged coefficients of the linear operator). For the error on the correctors, it yields:

$$\left\| C_\epsilon - \sum_{i=1}^{I_H} \nabla v_{\eta,\epsilon}^{H,i} 1_{Q_{H,i}} \right\|_{L^2(\Omega)} \le C\sqrt{\frac{\epsilon}{\eta}}, \tag{7.15}$$

where C_ϵ denotes now the two-scale corrector for periodic homogenization (see [7, Thm 9.9] for a definition of the corrector and [73, 108] for the error estimate). We refer to Proposition 5 for the corresponding estimates in the monotone case. The aim of windowing is twofold: improve the convergence rate, and if not, at least, improve the prefactor, by reducing both sources of error.

In the linear periodic case, windowing restores a convergence of order $\frac{\epsilon}{\eta}$ for the $L^2(\Omega)$ norm of the corrector and reduces the prefactor multiplying the error of order $\frac{\epsilon}{\eta}$ for the homogenized coefficients ([69] and [105] e.g.). The strategy consists in introducing bigger volume elements $C(x, \eta+\zeta)$, where $\zeta = \zeta(\epsilon, \eta) > 0$ (hence the term *windowing*), and using the information only on $C(x, \eta)$ to avoid the boundary layer of presumed order ζ. The application of windowing is different for the MsFEM and the HMM. In particular, for the MsFEM, several choices (mainly depending on the relative weights for the construction of the MsFEM basis, see Remark 34 for one variant) are possible. One of them has been analyzed in great details in [108] in the linear periodic case. The mathematical formulation of windowing is introduced in the following section.

7.3.2 Mathematical formulation

Let $\zeta = \zeta(\epsilon, \eta) \in \mathbb{R}_+$. We define a 'windowed' energy density

$$W^{win}_{\eta,\epsilon,\zeta}(x, \xi) = \langle W_\epsilon(y, \xi + \nabla v^{win}_{\eta,\epsilon,\zeta}(y))\rangle_{C(x,\eta)}, \tag{7.16}$$

where $v^{win}_{\eta,\epsilon,\zeta}$ is the restriction on $C(x, \eta)$ of the solution $\tilde{v}^{win}_{\eta,\epsilon,\zeta}$ of the following minimization problem posed on $C(x, \eta+\zeta)$

$$\inf\left\{\langle W_\epsilon(\cdot, \xi + \nabla v(\cdot))\rangle_{C(x,\eta+\zeta)} \,|\, v \in W^{1,p}_0(C(x, \eta+\zeta))\right\}. \tag{7.17}$$

We can naturally extend this definition to balls $B(x, \eta)$ and other boundary conditions ($\langle\nabla v\rangle = \xi$, periodic boundary conditions). Since the arguments and results are the same, we will focus on Dirichlet boundary conditions in what follows.

For the MsFEM with oversampling, we will adopt the following notations:

$$\int_\Omega W^{MsFEM}_{H,\epsilon,over}(x, \nabla w_H) = \sum_i \int_{Q_{H,i}} W_\epsilon(x, \nabla w^{MsFEM}_{H,\epsilon,over\,|Q_{H,i}}), \tag{7.18}$$

where $w^{MsFEM}_{H,\epsilon,over\,|Q_{H,i}}$ is the restriction on $Q_{H,i}$ of a solution of

$$\inf\left\{\int_{Q_{H+\zeta,i}} W_\epsilon(y, \nabla v) \,|\, v(y) = \nabla w_H \cdot y \text{ on } \partial Q_{H+\zeta,i}\right\} \tag{7.19}$$

and $Q_{H+\zeta,i}$ is an extension of $Q_{H,i}$ such that $\zeta \le d(\partial Q_{H,i}, \partial Q_{H+\zeta,i}) \le 2\zeta$. In what follows, we will also make use of the oscillating part of $\nabla w^{MsFEM}_{H,\epsilon,over\,|Q_{H,i}}$, namely

$$\nabla w^{H,i}_{\epsilon,over}(y) = \nabla w^{MsFEM}_{H,\epsilon,over\,|Q_{H,i}} - \langle\nabla w^{MsFEM}_{H,\epsilon,over\,|Q_{H,i}}\rangle_{|Q_{H,i}}. \tag{7.20}$$

We will not make a full error analysis of the cancellations that occur due to windowing, and we refer the reader to [108] for a deep analysis in the linear case. We will however make two remarks that may explain why numerical errors are reduced by the use of windowing methods. The aim of the present work is to prove the convergence of numerical homogenization methods with windowing for general energy densities and general heterogeneities. In this setting, we are not able to exhibit error estimates, but we will prove convergence results and relate windowing methods to Tartar's correctors in homogenization.

7.3.3 A remark on boundary conditions

The mismatch between boundary conditions may be understood as follows. Let us consider a cubic domain $C(x,\eta)$ with $\eta/\epsilon \in \mathbb{N}$. The domain $C(x,\eta)$ is then exactly obtained by the concatenation of a given number of periodic cells. If periodic boundary conditions were used on $C(x,\eta)$, $W_{\eta,\epsilon}$ would exactly be W_{hom} (since (7.3) would exactly be the cell-problem). If Dirichlet boundary conditions are used, far from the boundary, the solution 'tends to be' periodic. Close to the boundary the solution is very different from the periodic solution as illustrated in [8, Fig. 3]. In order to reduce the error, it is then natural to use the solution on periodic cells contained in $C(x,\eta)$ that are far from the boundary $\partial C(x,\eta)$. We refer the reader to the example of the half space dealt with in [22]. This remark is of great interest for the reconstruction of the fine scales features which highly suffers from this boundary layer.

The technique of windowing is unlikely to change the rate of convergence for the homogenized coefficients, as it can be easily seen in the periodic one-dimensional case (a direct calculation shows that the error still scales like $\frac{\epsilon}{\eta}$ and not better). However, not taking into account the boundary layer may improve the prefactor of the error. For the two-dimensional numerical examples considered in [108] and [73], namely a heterogeneous Laplace equation of type $-\text{div}\, A_\epsilon(x)\nabla u = 0$, with a ϵ-periodic operator A_ϵ defined by

$$A_\epsilon(x) = \left(\frac{2+1.8\sin(2\pi x/\epsilon)}{2+1.8\cos(2\pi y/\epsilon)} + \frac{2+\sin(2\pi y/\epsilon)}{2+1.8\cos(2\pi x/\epsilon)}\right)\text{Id}, \tag{7.21}$$

the associated energy density is $W_\epsilon(x,\xi) = \frac{1}{2}A_\epsilon(x)\xi\cdot\xi$. To be more precise Tables 7.1 and 7.2 report on some simple numerical tests that show the significant effect of windowing in the illustrative case (7.21). In Table 7.1 the error between the homogenized coefficients and the approximated homogenized coefficients is reported on, using Dirichlet boundary conditions on an increasing number of periodic cells. The first approximated operator is obtained without windowing, whereas the second one is obtained by considering only the periodic cell which is at the center of the domain of computation. The convergence rate is clearly proportional to the inverse of the number of periodic cells per dimension in both cases, however the prefactor is four times smaller in the second case.

Number of periodic cells per dimension	without windowing			with windowing		
	error	rate of convergence	prefactor (rate=1)	error	rate of convergence	prefactor (rate=1)
1	0.157	-	0.157	0.157	-	0.157
2	0.0845	0.895	0.169	0.0210	2.90	0.0420
4	0.0433	0.963	0.173	0.0118	0.835	0.0471
8	0.0219	0.983	0.175	0.00597	0.979	0.0478
12	0.0146	1.01	0.175	0.00397	1.00	0.0476
16	0.0110	0.965	0.176	0.00299	0.985	0.0478
20	0.00876	1.03	0.175	0.00239	1.00	0.0478

Table 7.1. Error on the approximated homogenized coefficients (performed with [88, FreeFEM] on a Laplace operator ($-\text{div}\, A_\epsilon\nabla$) with $P2$-finite elements on the cube $]0,1[^2$, with 100 elements per periodic cell).

For the reconstruction of the fine scale features of the solution u_ϵ, windowing reduces the L^2-norm of the error between the numerical corrector and the two-scale corrector (7.15), whose rate of convergence with respect to $\frac{\epsilon}{\eta}$ passes from $\frac{1}{2}$ to 1 using windowing, as illustrated in Table 7.2. This issue is dealt with theoretically in [108] and [69]. Its proof is based on the two-scale expansion of the solution and will not be detailed here since it cannot be generalized to other heterogeneities. In Proposition 6, the interest of windowing for nonlinear operators is addressed in terms of error

contributions: the error made on the homogenized energy could be greater than the error due to the boundary layers of the corrector. Therefore the effect of windowing on the correctors may not change the order of the error for the fine scales. In this case however, the prefactor of the error can still be reduced by windowing.

Number of periodic cells per dimension	without windowing			with windowing		
	error	rate of convergence	prefactor (rate=0.5)	error	rate of convergence	prefactor (rate=1)
1	0.210	-	0.210	0.210	-	0.210
2	0.156	0.425	0.221	0.0116	0.893	0.0232
4	0.113	0.468	0.226	0.00361	1.684	0.0144
8	0.0808	0.484	0.229	0.00181	0.988	0.0145
12	0.0662	0.491	0.229	0.00121	1.00	0.0145
16	0.0574	0.496	0.230	0.000910	0.992	0.0146
20	0.0515	0.492	0.230	0.000726	1.02	0.0145

Table 7.2. L^2-norm of the error on the corrector (performed with [88, FreeFEM] on a Laplace operator ($-\text{div } A_\epsilon \nabla$) with $P2$-finite elements on the cube $]0,1[^2$, with 100 elements per periodic cell).

7.3.4 A remark on the volume element $C(x,\eta)$

Assuming that the mismatch due to boundary conditions is reduced, one still has to deal with another souce of error: the mismatch between $C(x,\eta)$ and the periodic cell. The domain $C(x,\eta)$ may not be exactly a multiple of the periodic cell. Therefore the mean of the energy on a periodic cell with a given periodic function does not coincide with the mean on $C(x,\eta)$ of the energy with the same periodic function. This error is more subtle than the previous one and of the same order, namely $\frac{\epsilon}{\eta}$ in (7.14) and for the the homogenized coefficients in the linear periodic case. This source of error is of a lower order for the Petrov-Galerkin formulation of the MsFEM [108], as will be discussed in the following subsection. This solution is intimately linked to the particular formulation (or discretization of $I_{\eta,\epsilon}$ in other terms) of the MsFEM. It does not apply to the HMM for example. In the latter case, something else has to be done.

Remark **30** *In the case for which $C(x,\eta)$ is not exactly a multiple of the periodic cell, for both periodic and Dirichlet boundary conditions, numerical tests exhibit the same behaviour as in Tables 7.1 and 7.2.*

7.3.5 Interpretation of the MsFEM in Petrov-Galerkin formulation

Let us first introduce some notations. Given a minimization problem

$$\inf_{w\in V} \int_\Omega W(x,\nabla w) - fw,$$

its associated Euler-Lagrange equation

$$-\text{div } \partial_\xi W(x,\nabla w) = f,$$

and two finite dimensional spaces V_H^1 and V_H^2, we consider a discrete formulation: find $u_H \in V_H^1$ such that for all $w_H \in V_H^2$,

$$\int_\Omega \partial_\xi W(x,\nabla u_H)\nabla w_H = \int_\Omega f w_H.$$

We 'abusively' say that the formulation is:

- conforming if $V_H^1 \subset V$, and nonconforming if $V_H^1 \not\subset V$;
- Galerkin if $V_H^1 = V_H^2$, and Petrov-Galerkin if $V_H^1 \neq V_H^2$.

We also have to detail some kind of 'generalized' variational formulation for the nonlinear MsFEM method. Let V_H denote a classical $P1$-finite element space. For all $w_H \in V_H$, there exists a function $w_{H,\epsilon}^{MsFEM} \in L^p(\Omega)$ such that $w_{H,\epsilon}^{MsFEM}{}_{|Q_{H,i}} \in W^{1,p}(Q_{H,i})$ for all i, and

$$\int_\Omega W_{H,\epsilon}^{MsFEM}(x, \nabla w_H) = \int_\Omega W_\epsilon\left(x, \nabla w_{H,\epsilon}^{MsFEM}\right), \tag{7.22}$$

where $\nabla w_{H,\epsilon}^{MsFEM}$ abusively denotes $\sum_i \nabla w_{H,\epsilon}^{MsFEM}{}_{|Q_{H,i}} 1_{|Q_{H,i}}$. The correspondence (7.22) between w_H and $w_{H,\epsilon}^{MsFEM}$ defines a (nonlinear) mapping from V_H to $\oplus_i W^{1,p}(Q_{H,i})$, as introduced in [74]. The nonlinear mapping provides us with a relationship of duality between the points of view $(W_{H,\epsilon}^{MsFEM}, V_H)$ and $(W_\epsilon, \{w_{H,\epsilon}^{MsFEM}\})$.

Without oversampling, the mapping takes values in $W^{1,p}(\Omega)$, whereas with oversampling the restrictions of $w_{H,\epsilon,over}^{MsFEM}$ belong to $W^{1,p}(Q_{H,i})$ but $w_{H,\epsilon,over}^{MsFEM} \notin W^{1,p}(\Omega)$. The 'generalized' variational formulation for the MsFEM then reads: find $u_H \in V_H$ such that for all $w_H \in V_H$,

$$\int_\Omega \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon}^{MsFEM}) \nabla w_{H,\epsilon}^{MsFEM} = \int_\Omega f w_{H,\epsilon}^{MsFEM}, \tag{7.23}$$

where $u_{H,\epsilon}^{MsFEM}$ and $w_{H,\epsilon}^{MsFEM}$ are related to u_H and w_H by the nonlinear mapping (7.22). One can also define the following formulation: find $u_H \in V_H$ such that for all $w_H \in V_H$,

$$\int_\Omega \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon}^{MsFEM}) \nabla w_H = \int_\Omega f w_H. \tag{7.24}$$

According to the above definitions, Formulation (7.23) is a Galerkin formulation, which is conforming for the classical MsFEM and nonconforming for the MsFEM with oversampling. Formulation (7.24) is a Petrov-Galerkin formulation, which is also conforming for the classical MsFEM and nonconforming for the MsFEM with oversampling.

In Section 7.3.4, we have defined a 'geometric error'. The MsFEM does not suffer from this kind of error mainly because $\eta = H$ in the formulation. The proportion of each material at scale ϵ is globally conserved in $W_{H,\epsilon}^{MsFEM}$ due to (7.22). In other terms, if W_ϵ is the energy density of a periodic composition of two materials A and B, the ratio of A and B in W_ϵ is exactly preserved in $W_{H,\epsilon}^{MsFEM}$ whereas it is only preserved up to an error of order $\frac{\epsilon}{\eta}$ for a generic averaged energy density $W_{\eta,\epsilon}$. Recalling the short discussion of Section 7.3.4, each periodic cell of the material Ω is exactly accounted for once in the MsFEM: if only half of a given periodic cell belongs to some $Q_{H,i}$ then the other half belongs to some other $Q_{H,j}$. For a generic $W_{\eta,\epsilon}$, this may not be true. In [108], the nonconforming Galerkin MsFEM of [73] is shown to still exhibit a cell resonance error for linear problems that can be reduced using a Petrov-Galerkin method. In the remainder of this section, we give a simple argument that may explain why the Petrov-Galerkin version of the nonconforming MsFEM is better in general than the Galerkin version. And we show how the nonconforming Petrov-Galerkin MsFEM may be recast within the analytic framework of Section 7.2. In particular, without a Petrov-Galerkin formulation, there is no equality corresponding to (7.22) if oversampling is used (actually there is no variational interpretation, see Remark 32 hereafter).

To do so, let us study some basic properties of Formulations (7.23) and (7.24) for both classical and 'oversampled' MsFEM. We first consider the classical MsFEM. In this case, the following calculation,

$$\begin{aligned}
\int_\Omega \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon}^{MsFEM}) \nabla w_{H,\epsilon}^{MsFEM} &= \sum_i \int_{Q_{H,i}} \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon}^{MsFEM}) \nabla w_{H,\epsilon}^{MsFEM} \\
&= \sum_i \int_{Q_{H,i}} \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon}^{MsFEM})) \langle \nabla w_H \rangle_i \\
&= \int_\Omega \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon}^{MsFEM}) \nabla w_H,
\end{aligned}$$

that holds due to the Euler-Lagrange equation associated with (7.12), shows that the differential terms of the variational Formulations (7.23) and (7.24) indeed coincide.

On the contrary, for the 'oversampled' MsFEM, the differential terms of the variational Formulations (7.23) and (7.24) do not coincide. The Euler-Lagrange equation of (7.19) is not defined on $Q_{H,i}$ anymore but on a larger domain $Q_{H+\zeta,i}$ which prevents us from writing the decomposition as a sum of Euler-Lagrange equations on $Q_{H,i}$, as it is done in the previous calculation. Thus, for the 'oversampled' method,

$$\int_\Omega \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon,over}^{MsFEM}) \nabla w_{H,\epsilon,over}^{MsFEM} \neq \int_\Omega \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon,over}^{MsFEM}) \nabla w_H.$$

We actually have instead

$$\begin{aligned}
&\int_\Omega \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon,over}^{MsFEM}) \nabla w_{H,\epsilon,over}^{MsFEM} \\
&\qquad = \sum_i \int_{Q_{H,i}} \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon,over}^{MsFEM}) \nabla w_{H,\epsilon,over}^{MsFEM} \\
&\qquad = \sum_i \int_{Q_{H,i}} \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon,over}^{MsFEM}))(\langle \nabla w_H \rangle_i + \nabla w_{\epsilon,over}^{H,i}) \\
&\qquad = \sum_i \int_{Q_{H,i}} \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon,over}^{MsFEM}) \langle \nabla w_H \rangle_i \\
&\qquad\qquad - \sum_i \int_{Q_{H+\zeta,i} \setminus Q_{H,i}} \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon,over}^{MsFEM}) \nabla w_{\epsilon,over}^{H,i},
\end{aligned} \tag{7.25}$$

formally using the Euler-Lagrange equation associated with (7.19):

$$\int_{Q_{H+\zeta,i}} \partial_\xi W_\epsilon(y, \nabla u_{H,\epsilon,over}^{MsFEM}) \cdot \nabla w_{\epsilon,over}^{H,i} = 0,$$

where $\nabla w_{\epsilon,over}^{H,i}$ is given by (7.20). The difference in the differential operator between the Petrov-Galerkin and the Galerkin formulations of the 'oversampled' MsFEM (which are both nonconforming) is therefore given by

$$\sum_i \int_{Q_{H+\zeta,i} \setminus Q_{H,i}} \partial_\xi W_\epsilon(x, \nabla u_{H,\epsilon}^{MsFEM}) \nabla w_{\epsilon,over}^{H,i}. \tag{7.26}$$

It is now time to come back to the origin of windowing methods. Windowing aims at getting rid of the effects of the boundary layer on $Q_{H,i}$ by computing the minimum (7.17) on a domain $Q_{H+\zeta,i}$ of diameter of order $H+\zeta$ and by only considering the restriction of the associated solution $w_{\epsilon,over}^{H,i} \in W^{1,p}(Q_{H+\zeta,i})$ on $Q_{H,i}$. Our simple calculation shows that the nonconforming Galerkin MsFEM implicitly takes into account the term (7.26). This term involves the restriction of the multiscale finite element on $Q_{H+\zeta,i} \setminus Q_{H,i}$, which contains a part of the boundary layer that was supposed to be cancelled by the windowing method. Erasing the very last term of (7.25) and going backwards, we recover the nonconforming Petrov-Galerkin formulation of the MsFEM, which truly avoids the boundary layer.

To conclude this section, let us show that the nonconforming Petrov-Galerkin formulation of the MsFEM is equivalent to the variational formulation recalled in Section 7.2 combined with the windowing introduced in Section 7.3.2. Using the results of Section 5.4 to switch the derivation with respect to ξ and the minimization (7.17) for convex energies with invertible Hessians (and proceeding formally otherwise), one may write the nonconforming Petrov-Galerkin MsFEM as follows,

$$\begin{aligned}\int_\Omega \partial_\xi W_\epsilon(x,\nabla u^{MsFEM}_{H,\epsilon,over})\nabla w_H &= \sum_i \int_{Q_{H,i}} \partial_\xi W_\epsilon(x,\nabla u^{MsFEM}_{H,\epsilon,over}))\langle\nabla w_H\rangle_i \\ &= \sum_i \int_{Q_{H,i}} \partial_\xi W^{MsFEM}_{H,\epsilon,over}(x,\langle\nabla u_H\rangle_i)\langle\nabla w_H\rangle_i \\ &= \int_\Omega \partial_\xi W^{MsFEM}_{H,\epsilon,over}(x,\nabla u_H)\nabla w_H.\end{aligned}$$

In Section 7.4, we will prove the convergence of the Petrov-Galerkin formulation of the 'oversampled' MsFEM for rather general elliptic operators and general heterogeneities using the variational formulation of Section 7.2.

Remark **31** *Let us stress the fact that the previous calculation shows that the nonconforming Petrov-Galerkin formulation of the MsFEM (which yields discontinuous tests functions) for W_ϵ can be equivalently seen as a Galerkin method for $W^{MsFEM}_{H,\epsilon,over}$. This explains why variational methods can be used on $W^{MsFEM}_{H,\epsilon,over}$ to prove the convergence of the PG-MsFEM for W_ϵ.*

Remark **32** *The nonconforming Galerkin formulation of the MsFEM cannot be recast within the framework of Section 7.2.*

7.4 Windowing for general heterogeneities

In this section, we first define the windowing method for general heterogeneities by making precise the dependence of the windowing upon the parameter ϵ and the characteristic lengthscale η. Provided a right scaling, we then prove the convergence of numerical homogenization methods with windowing, within the framework of Section 7.2.

There are numerical evidence that show the practical interest of windowing for nonperiodic problems. There is also another motivation that is related to Tartar's corrector. In Section 7.2.4, the numerical corrector has been related to Tartar's corrector provided an approximation depending on H and provided $\omega = \omega_1$. In Tartar's original work however, the correctors are proved to exist using $\omega \subset\subset \omega_1$, which is windowing in the present language. Numerical correctors with windowing are therefore approximations of Tartar's correctors that may seem more natural than the numerical correctors of Section 7.2.4. The use of windowing allows us to recover all the diversity of the original Tartar's correctors.

7.4.1 Scaling of the windowing

The aim of windowing is to reduce the mismatch between the free oscillations of an unconstrained solution at fixed ϵ and the boundary conditions on domains $C(x,\eta)$. A major assumption concerns the convergence of the energies I_ϵ to a homogenized energy I_{hom} which is supposed not to exhibit oscillations at small scales. The windowing for general heterogeneities should match the scales of the oscillations. Therefore it has to vanish with η, but it may also already vanish with ϵ. We set

Definition 34 *Let $\zeta : \mathbb{R}_+ \to \mathbb{R}_+$. For all $x \in \Omega$, $\eta > 0$ and $\epsilon < \eta$, ζ defines an ϵ-admissible windowing domain $C(x,\eta+\zeta(\epsilon))$ if*

$$\lim_{\epsilon\to 0}\zeta(\epsilon) = 0,$$

and an η-admissible windowing domain $C(x,\eta+\zeta(\eta))$ if

$$\lim_{\eta\to 0}\frac{\zeta(\eta)}{\eta} = 0.$$

If ϵ measures the typical "size" of the heterogeneities, then ζ should satisfy $\lim_{\epsilon\to 0}\zeta(\epsilon) = 0$ and a property of the type $\lim_{\epsilon\to 0}\frac{\zeta(\epsilon)}{\epsilon} = +\infty$ in order to see the effect of windowing. The prototypical

example is given by the linear periodic case in Section 7.3.3 for which the boundary layer is of order ϵ (thus any $\zeta(\epsilon) = \epsilon^\alpha$ with $1 > \alpha > 0$ is enough).

The above heuristics is related to geometric properties of the heterogeneities (in particular the period, or the correlation length). Hence, ϵ should ideally be a typical lengthscale whereas it appears as a simple parameter (which has no geometric meaning) for the assumption I_ϵ $\Gamma(L^p)$-converges to I_{hom}. In Definition 34, we thus only consider windowings that are 'stable under' a change of parametrization.

In the following subsections, we prove that the use of admissible windowings does not affect the convergence of the numerical homogenization method, meaning that the method also converges using windowing.

Remark **33** *The windowings introduced in Definition 34 are two extreme cases. One can also introduce particular windowings depending both on ϵ and η and providing us with suitable regimes for given applications. They can also be seen as particular cases of the η-admissible windowing.*

7.4.2 Convergence results

In this section, we prove the convergence of numerical homogenization with η-admissible windowing, which also implies the convergence with ϵ-admissible windowing. We first address the convergence of a continuous 'windowed' energy density whose FE-discretization leads to the HMM. We then show the convergence of two versions of the nonconforming Petrov-Galerkin MsFEM.

Convergence at the continuous level

The 'windowed' continuous energy density is given by the following

Definition 35 *Let ζ be an η-admissible windowing and W_ϵ satisfy $H1$, $H4$, and $H3$. For all $\eta, \epsilon > 0$, the associated 'windowed' energy density is defined by*

$$\mathcal{W}^{win}_{\eta,\epsilon}(x,\xi) = \langle W_\epsilon(y, \xi + \nabla v^{win}_{\eta,\epsilon})\rangle_{C(x,\eta)}$$

where $v^{win}_{\eta,\epsilon}$ is the restriction on $C(x,\eta)$ of $\tilde{v}^{win}_{\eta,\epsilon}$, solution of (7.17) with $\zeta = \zeta(\eta)$.

We then have the following two convergence results.

Theorem 47 *Let W_ϵ satisfy $H1$, $H2$ (strictly), and $H3$ uniformly for $p > 1$, and ζ be an η-admissible windowing, then the energy densities $C\mathcal{W}^{win}_{\eta,\epsilon}$ also satisfy $H1$, $H2$ and $H3$ for η small enough, and the energy $I^{win}_{\eta,\epsilon} : v \in W^{1,p}(\Omega) \mapsto \int_\Omega C\mathcal{W}^{win}_{\eta,\epsilon}(x, \nabla v)$ $\Gamma(L^p)$-and $\Gamma(W^{1,p})$-converges to I_{hom} as ϵ and η go to 0, where $C\mathcal{W}$ denotes the convex envelop of $\mathcal{W}$. Therefore, any sequence $u^{win}_{\eta,\epsilon}$ of minimizers of $\inf\{I^{win}_{\eta,\epsilon}(v) \,|\, v \in W^{1,p}(\Omega,\mathbb{R}) + BC\}$ strongly converges to the unique minimizer u_{hom} of $\inf\{I_{hom}(v) \,|\, v \in W^{1,p}(\Omega,\mathbb{R}) + BC\}$ in $W^{1,p}(\Omega,\mathbb{R}^n)$.*

Theorem 48 *Let W_ϵ satisfy $H1$, $H4$, and $H3$ uniformly for $p > 1$, and ζ be an η-admissible windowing, then the energy densities $Q\mathcal{W}^{win}_{\eta,\epsilon}$ are standard energy densities and $I^{win}_{\eta,\epsilon} : v \in W^{1,p}(\Omega) \mapsto \int_\Omega Q\mathcal{W}^{win}_{\eta,\epsilon}(x, \nabla v)$ $\Gamma(L^p)$- and $\Gamma(W^{1,p})$-converges to I_{hom} as ϵ and η go to 0. Therefore, for any sequence $u^{win}_{\eta,\epsilon}$ of minimizers of $\inf\{I^{win}_{\eta,\epsilon}(v) \,|\, v \in W^{1,p}(\Omega,\mathbb{R}^d) + BC\}$, there exists a minimizer u_{hom} of $\inf\{I_{hom}(v) \,|\, v \in W^{1,p}(\Omega,\mathbb{R}^d) + BC\}$ such that*

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0} u^{win}_{\eta,\epsilon} = u_{hom} \quad \text{weakly in } W^{1,p}(\Omega,\mathbb{R}^d), \tag{7.27}$$

up to extraction.

In practice, one does not need to convexify $\mathcal{W}^{win}_{\eta,\epsilon}$ since the minimum is searched in a finite dimensional subspace of $W^{1,p}(\Omega)$, the strict convexity being recovered at the limit $\epsilon \to 0$ for ϵ-windowings and $\eta \to 0$ for η-windowings, in the spirit of Theorem 43 for the quasiconvex case.

Theorems 47 and 48 imply the convergence of the HMM with windowing in the general case.

Proof of Theorems 47 and 48.
We divide the proof in two steps. We first introduce an averaged energy density for which the strategy used to prove Theorems 40 and 42 holds. We then show the 'windowed' energy density to be uniformly close to this averaged energy as η goes to zero.

Let us introduce the averaged energy density

$$\tilde{\mathcal{W}}^{win}_{\eta,\epsilon}(x,\xi) = \langle W_\epsilon(y,\xi+\nabla\tilde{v}^{win}_{\eta,\epsilon})\rangle_{C(x,\eta+\zeta(\eta))}.$$

This energy density is of type (7.3) (cf. Remark 28) up to denoting by $\tilde{\eta} = \eta + \zeta(\eta)$. Thus Theorems 44 and 45 apply and we denote by $\tilde{I}^{win}_{\eta,\epsilon}$ the associated energy functional.

Let us prove now that the $\Gamma(L^p)$ and $\Gamma(W^{1,p})$-convergence of $\tilde{I}^{win}_{\eta,\epsilon}$ and $I^{win}_{\eta,\epsilon}$ are equivalent. Due to Meyers' regularity estimate, H1, H4, H3 and possibly a convolution argument (see [134] and [38, Theorem C.2]), there exist $\alpha > 0$ and $c > 0$, independent of η and ϵ such that

$$\|\tilde{v}^{win}_{\eta,\epsilon}\|_{W^{1,p+\alpha}(C(x,\eta+\zeta(\eta)))} \le c\|\tilde{v}^{win}_{\eta,\epsilon}\|_{W^{1,p}(C(x,\eta+\zeta(\eta)))}. \tag{7.28}$$

The coefficients (exponent α and prefactor c) appearing in the Meyers' estimate do only depend on the growth conditions and on the domain (see [96] and [95, Theorem 3.1 and Remark 3.5]). Let us prove that they do not depend on η either. Up to introducing the scaling

$$W^{1,p}_{\#}((0,1)^n) \ni v \mapsto \bar{v}(\cdot) = (\eta+\zeta(\eta))v\left(\frac{\cdot}{\eta+\zeta(\eta)}\right) \in W^{1,p}_{\#}((0,\eta+\zeta(\eta))^n),$$

we have

$$\int_{(0,1)^n} W(x,\nabla_x v(x))dx = \frac{1}{(\eta+\zeta(\eta))^n}\int_{(0,\eta+\zeta(\eta))^n} W\left(\frac{y}{\eta+\zeta(\eta)},\nabla_y\bar{v}(y)\right)dy$$

for any standard energy density. Let c_1 denote the coefficient provided by Meyers' theorem on the domain $(0,1)^n$ and by c_2 the constant of the Poincaré-Wirtinger inequality. Let $v \in W^{1,p}_{\#}(0,1)^n$ be a minimizer of the associated energy on a given set. We have for $\eta+\zeta(\eta) \le 1$:

$$\begin{aligned}
\|\bar{v}\|^{p+\alpha}_{W^{1,p+\alpha}((0,\eta+\zeta(\eta))^n)} &= (\eta+\zeta(\eta))^{n+p+\alpha}\|v\|^{p+\alpha}_{L^{p+\alpha}((0,1)^n)} + (\eta+\zeta(\eta))^n\|\nabla v\|^{p+\alpha}_{L^{p+\alpha}((0,1)^n)}\\
&\le (\eta+\zeta(\eta))^n\|v\|^{p+\alpha}_{W^{1,p+\alpha}((0,1)^n)}\\
&\le (\eta+\zeta(\eta))^n c_1^{p+\alpha}\|v\|^{p+\alpha}_{W^{1,p}((0,1)^n)}\\
&\le (\eta+\zeta(\eta))^n c_1^{p+\alpha}(1+c_2)^{p+\alpha}\|\nabla v\|^{p+\alpha}_{L^p((0,1)^n)}\\
&\le c_1^{p+\alpha}(1+c_2)^{p+\alpha}\|\bar{v}\|^{p+\alpha}_{W^{1,p}((0,\eta+\zeta(\eta))^n)},
\end{aligned}$$

which shows that (7.28) holds with $c = c_1(1+c_2)$.

Using the growth condition H3 on W_ϵ, (7.17) and (7.28) we obtain

$$\|\tilde{v}^{win}_{\eta,\epsilon}\|^p_{W^{1,p+\alpha}(C(x,\eta+\zeta(\eta)))} \le c(\eta+\zeta(\eta))^n(1+|\xi|^p). \tag{7.29}$$

The application of Hölder inequality yields

$$\|\tilde{v}^{win}_{\eta,\epsilon}\|^p_{W^{1,p}(C(x,\eta+\zeta(\eta))\setminus C(x,\eta))} \le C[\eta^{n-1}\zeta(\eta)]^{\frac{\alpha}{p+\alpha}}(\|\tilde{v}^{win}_{\eta,\epsilon}\|^p_{W^{1,p+\alpha}(C(x,\eta+\zeta(\eta)))})^{\frac{p}{p+\alpha}}$$

which implies

$$\|\tilde{v}^{win}_{\eta,\epsilon}\|^p_{W^{1,p}(C(x,\eta+\zeta(\eta))\setminus C(x,\eta)))} \le C(\eta+\zeta(\eta))^{n\frac{p}{p+\alpha}}[\eta^{n-1}\zeta(\eta)]^{\frac{\alpha}{p+\alpha}}(1+|\xi|^p)$$

using (7.29) and noticing that $(1+|\xi|^p)^{\frac{p}{p+\alpha}} \le (1+|\xi|^p)$. We finally deduce

$$|\tilde{\mathcal{W}}^{win}_{\eta,\epsilon}(x,\xi) - \mathcal{W}^{win}_{\eta,\epsilon}(x,\xi)| \le C\left(\left[\frac{\zeta(\eta)}{\eta}\right]^{\frac{\alpha}{p+\alpha}} + \frac{\zeta(\eta)}{\eta}\right)(1+|\xi|^p), \tag{7.30}$$

using the well known quasi-uniform Lipschitz property of rank-one convex functions (see Formula (6.43) e.g.), and noticing $\frac{\eta}{\eta+\zeta} = \eta(1 - \frac{\zeta}{\eta} + o(\frac{\zeta}{\eta}))$ and

$$\frac{(\eta+\zeta(\eta))^{n\frac{p}{p+\alpha}}}{(\eta+\zeta(\eta))^n}[\eta^{n-1}\zeta(\eta)]^{\frac{\alpha}{p+\alpha}} = (\eta+\zeta(\eta))^{-n\frac{\alpha}{p+\alpha}}[\eta^{n-1}\zeta(\eta)]^{\frac{\alpha}{p+\alpha}} \leq \left[\frac{\zeta(\eta)}{\eta}\right]^{\frac{\alpha}{p+\alpha}}.$$

In particular, (7.30) implies that $\mathcal{W}^{win}_{\eta,\epsilon}$ satisfies H3 with a modified but strictly positive constant c for η small enough (since $\tilde{\mathcal{W}}^{win}_{\eta,\epsilon}$ does), which ensures the existence of minimizers for the relaxed problem.

The dominated convergence theorem then allows us to prove the uniform convergence to zero of $\tilde{I}^{win}_{\eta,\epsilon} - I^{win}_{\eta,\epsilon}$ on any bounded subset of $W^{1,p}(\Omega)$ as η goes to zero. This is enough to ensure the equivalence of the Γ-convergences of the energy functionals, as briefly recalled below (see [36] or [56] for classical definitions related to Γ-convergence).

Let us first notice that the energies are only finite on $W^{1,p}(\Omega)$. For all $w \in W^{1,p}(\Omega)$ and all sequence $w_{\eta,\epsilon} \in W^{1,p}(\Omega)$ such that $\lim_{\eta\to 0}\lim_{\epsilon\to 0} w_{\eta,\epsilon} = w$ in $L^p(\Omega)$, either

$$\lim_{\eta\to 0}\lim_{\epsilon\to 0}\tilde{I}^{win}_{\eta,\epsilon}(w_{\eta,\epsilon}) = \lim_{\eta\to 0}\lim_{\epsilon\to 0} I^{win}_{\eta,\epsilon}(w_{\eta,\epsilon}) = +\infty$$

and the Γ-liminf inequality trivially holds, or the sequence $\{w_{\epsilon,\eta}\}$ is bounded in $W^{1,p}(\Omega)$. In the latter case, the sequence belongs to a set on which the convergence of $\tilde{I}^{win}_{\eta,\epsilon} - I^{win}_{\eta,\epsilon}$ to zero is uniform. Thus, $\lim_{\eta\to 0}\lim_{\epsilon\to 0}\tilde{I}^{win}_{\eta,\epsilon}(w_{\eta,\epsilon}) - I^{win}_{\eta,\epsilon}(w_{\eta,\epsilon}) = 0$ and the $\Gamma(L^p)$-liminf (resp. limsup) of $I^{win}_{\eta,\epsilon}$ and $\tilde{I}^{win}_{\eta,\epsilon}$ coincide. Therefore they have the same $\Gamma(L^p)$-limit. The same reasoning holds for the $\Gamma(W^{1,p})$-convergence.

As a consequence, the Γ-convergence results obtained for $\tilde{I}^{win}_{\eta,\epsilon}$ hold for $I^{win}_{\eta,\epsilon}$, which concludes the proof of Theorems 47 and 48.

□

Convergence of the nonconforming Petrov-Galerkin MsFEM

Definition 36 *Let ζ be an η-admissible windowing and W_ϵ satisfy H1, H4, and H3. Let $\{Q_{H,i}\}_i$ be a triangulation of Ω. For all $\eta, \epsilon > 0$, the associated MsFEM 'oversampled' energy density is defined by*

$$W^{MsFEM}_{H,\epsilon,over}(x,\xi) = \sum_i \langle W_\epsilon(y, \xi + \nabla v^{H,i}_{\epsilon,over})\rangle_{Q_{H,i}} 1_{Q_{H,i}}(x)$$

where $v^{H,i}_{\epsilon,over}$ is a solution of (7.19).

The associated result is then the following.

Theorem 49 *Let W_ϵ satisfy H1, H4 (resp. H2 strictly), and H3 uniformly for $p > 1$, and ζ be an η-admissible windowing. Let V_H be the space of P1-finite elements on the regular triangulation $\{Q_{H,i}\}_i$. The 'oversampled' Petrov-Galerkin MsFEM reads:*

$$\inf\left\{\sum_{i=1}^{I_H}|Q_{H,i}|W^{MsFEM}_{H,\epsilon,over}(x_i,\nabla u_H(x_i)) - \sum_{i=1}^{I_H}\sum_{j=1}^{N_{GP}} q_j f(x_{ij})u_H(x_{ij}), u_H \in V_H + BC\right\}, \tag{7.31}$$

where the second term $f \in L^{p'}(\Omega)$ of the energy has been integrated by a quadrature rule associated with the triangulation. Then any sequence of solutions $\{u^{over}_{\epsilon,H}\}$ to (7.31) *converges weakly in $W^{1,p}(\Omega)$ up to extraction (resp. converges strongly in $W^{1,p}(\Omega)$) to a minimizer (resp. the unique minimizer) of $w \mapsto I_{hom}(w) - \int_\Omega fw$ on $W^{1,p}(\Omega) + BC$.*

Proof of Theorem 49.

Let us divide the proof in three steps. First we introduce an averaged energy density whose associated energy functional $\Gamma(W^{1,p})$-converges to the homogenized energy. We then prove the convergence of the associated infima to the infimum of the homogenized energy, following the proof in Section 6.3.2. This implies the results of Theorem 49 for this averaged energy. We finally apply the argument of uniform convergence used in the proof of Theorems 47 and 48.

Let us consider the following averaged energy density

$$\tilde{W}^{MsFEM}_{H,\epsilon,over}(x,\xi) = \sum_i \langle W_\epsilon(y,\xi+\nabla v^{H,i}_{\epsilon,over})\rangle_{Q_{H+\zeta(H),i}} 1_{Q_{H,i}}(x).$$

Let P_H be an equicontinuous family of projectors from $W^{1,p}(\Omega)$ to V_H such that for all $w \in W^{1,p}(\Omega)$, $\lim_{H\to 0} \|P_H w - w\|_{W^{1,p}(\Omega)} = 0$. We then associate with $\tilde{W}^{MsFEM}_{H,\epsilon,over}$ an energy functional $\tilde{I}^{MsFEM}_{H,\epsilon,over} : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\tilde{I}^{MsFEM}_{H,\epsilon,over}(w) = \int_\Omega \tilde{W}^{MsFEM}_{H,\epsilon,over}(x,\nabla P_H w). \tag{7.32}$$

This family of energy functionals is equicontinuous on $W^{1,p}(\Omega)$ (see page 127, Chapter 6) and converges pointwise on $W^{1,p}(\Omega)$ to I_{hom} as ϵ and H vanish. Thus Lemma 6.7 (or [56, Theorem 5.9]) implies the $\Gamma(W^{1,p})$-convergence of $\tilde{I}^{MsFEM}_{H,\epsilon,over}$ to I_{hom}. It remains to prove the convergence of the infima of $\tilde{I}^{MsFEM}_{H,\epsilon,over}$ to the infimum of the homogenized energy to obtain the thesis of Theorem 49 for the family $\tilde{I}^{MsFEM}_{H,\epsilon,over}$. We only treat in detail the new argument (based on Meyer's estimates) with respect to Section 6.3.2. It is enough to prove that $\tilde{I}^{MsFEM}_{H,\epsilon,over} - I^{MsFEM}_{H,\epsilon}$ converges uniformly to zero on bounded subsets of $W^{1,p}(\Omega)$. To this aim, we can apply Meyers' estimates on each mesh element $Q_{H,i}$. The exponent α in (7.28) may however depend on H. Due to the regularity of the mesh this is not the case and there exists $\bar{\alpha}$ independent of H such that Meyers' estimate holds on every $Q_{H,i}$ with exponent $\bar{\alpha}$.

It suffices to introduce a linear transformation $T_{H,i}$ which maps the reference mesh element Q onto $Q_{H,i}$. Let then denote by (λ_k) the eigenvalues of $T_{H,i}$. Up to a change of variable using $T^{-1}_{H,i}$ and an isotropic dilatation by a factor $\sqrt[n]{\det T^{-1}_{H,i}}$:

$$W^{1,p}_0(Q_{H,i}) \ni \bar{v} \mapsto v(\cdot) = \left(\sqrt[n]{\det T^{-1}_{H,i}}\right) \bar{v}\left(T_{H,i}\cdot\right) \in W^{1,p}_0(Q),$$

we have

$$\int_{Q_{H,i}} W(y,\nabla_y \bar{v}(y))dy = \det T_{H,i} \int_Q W\left(T_{H,i}x, \frac{T^{-1}_{H,i}}{\sqrt[n]{\det T^{-1}_{H,i}}}\nabla_x v(x)\right) dx$$

for any standard energy density. The 'effective' energy density

$$W^{H,i}(x,\xi) = W\left(T_{H,i}x, \frac{T^{-1}_{H,i}}{\sqrt[n]{\det T^{-1}_{H,i}}}\xi\right)$$

on Q satisfies a growth condition of order p with constants only depending on c, C and the quotients $\{\frac{\lambda_k}{\sqrt[n]{\det T_{H,i}}}\}_{k\in[\![1,n]\!]}$. These quotients are bounded from above and below uniformly in H and i by definition of the regularity of the mesh. Therefore, there exists $\bar{\alpha}$ associated with the reference mesh element and this growth condition, such that Meyers' estimate holds on all $Q_{H,i}$. The strategy used in the proof of Theorem 47 then shows the uniform convergence of $\tilde{I}^{MsFEM}_{H,\epsilon,over} - I^{over}_{H,\epsilon}$ to zero on bounded subsets of $W^{1,p}(\Omega)$, which implies the convergence of the infimum of $\tilde{I}^{MsFEM}_{H,\epsilon,over}$ to the limit of the infima of $I^{MsFEM}_{H,\epsilon}$, which is exactly the infimum of I_{hom}

as proved in Chaper 6, pp. 127-130. The results of Theorem 49 then hold for the energy density $\tilde{W}^{MsFEM}_{H,\epsilon,over}$.

As for the proof of Theorem 47 we use Meyers' estimate once more to obtain the uniform convergence of $\tilde{I}^{MsFEM}_{H,\epsilon,over} - I^{MsFEM}_{H,\epsilon,over}$ to zero on bounded subsets of $W^{1,p}(\Omega)$, proving Theorem 49. □

Remark **34** *In [108], the 'oversampled' energy density is defined in a slightly different way. In the same spirit, one can replace $v^{over,i}_{H,\epsilon}$ by $\bar{v}^{over,i}_{H,\epsilon}(y) = v^{over,i}_{H,\epsilon}(y) - \langle \nabla v^{over,i}_{H,\epsilon} \rangle_{Q_{H,i}} \cdot y$. In particular, this formulation satisfies Hill's lemma, which is widely used in Mechanics. The present proof easily adapts since*

$$\lim_{H\to 0} \lim_{\epsilon\to 0} \langle \nabla v^{over,i}_{H,\epsilon} \rangle_{Q_{H,i}} = 0$$

and $W_\epsilon(y,\cdot)$ is uniformly (in space) Lipschitz-continuous.

7.4.3 Fine scale reconstruction

We now extend the numerical corrector of Definition 24 to the case of windowing. In the linear periodic case, windowing improves the approximation a lot since a great part of the error is located in a boundary layer of order ϵ. For general heterogeneities we are not able to show that the approximation is better. Even if it were, in view of Section 6.3.2, it is not clear whether the order of the global error is reduced. It also seems delicate to generalize the estimates derived in [73, 108] to monotone operators in a periodic setting since the analysis of the correctors is far less complete in this case than for linear problems.

However we prove that the numerical corrector associated with the windowing method has the same general convergence properties as the numerical corrector without windowing. The interest of windowing then relies on the possible reduction of the prefactor term in the error. Its efficiency is illustrated numerically in [75, p. 67] for a linear stochastic case.

Definition 37 *Let $\{Q_{H,i}\}_{i\in[\![1,I_H]\!]}$ be as in Definition 33. Keeping the notation of Theorem 47, we define the numerical correctors $v^{H,i}_{\eta,\epsilon,win}$ for a strictly convex energy density as the restriction on $Q_{H,i}$ of the unique minimizers (up to a constant) of*

$$\inf\left\{ \int_{Q_{H+\zeta(\eta),i}} W_\epsilon(x,\nabla v) \,|\, v \in W^{1,p}(Q_{H+\zeta(H),i}), v(y) = \langle \nabla u^{win}_{\eta,\epsilon} \rangle_{Q_{H,i}} \cdot y \text{ on } \partial Q_{H+\zeta(H),i} \right\}, \tag{7.33}$$

where $Q_{H+\zeta(H),i}$ is the concatenation of $Q_{H,i}$ and of a crown of width $\zeta(H)$.

We then have the corresponding convergence result of Theorem 46:

Theorem 50 *In addition to $H1$, $H2$, and $H3$, let us assume that $p \geq 2$, that $W_\epsilon(x,\cdot)$ is continuously differentiable for almost all $x \in \Omega$ and $a_\epsilon(\cdot,0) = \frac{\partial W_\epsilon}{\partial \xi}(\cdot,0)$ is bounded, and satisfies the monotonicity and continuity properties (7.7) and (7.6). Then, denoting by u_ϵ the unique minimizer of I_ϵ on $W^{1,p}(\Omega) + BC$, we have*

$$\lim_{\eta\sim H\to 0} \lim_{\epsilon\to 0} \left\| \nabla u_\epsilon - \sum_{i=1}^{I_H} \nabla v^{H,i}_{\eta,\epsilon,win} 1_{Q_{H,i}} \right\|_{L^p(\Omega)} = 0. \tag{7.34}$$

Proof of Theorem 50.
The convergence of the numerical correctors is a direct consequence of Theorems 47 and 41

The proof in Section 6.2.4 is based on two arguments: the strong convergence of $u_{\eta,\epsilon}$ to u_{hom} in $W^{1,p}(\Omega)$ and a passage from local estimates on $Q_{H,i}$ to a global estimate on Ω. The first argument holds for the 'windowed' method due to Theorem 47. The local estimates are now obtained on $Q_{H+\zeta(H),i}$ and also imply a global estimate on Ω since $\lim_{H\to 0} \frac{\zeta(H)}{H} = 0$. All the details of Section 6.2.4 adapt straightforwardly to the present case.

□

Remark **35** *In Definition 7.33, one can replace the Dirichlet boundary conditions by periodic boundary conditions.*

Remark **36** *A corollary of Theorem 50 shows that the family* $\{v_{\eta,\epsilon}^{win,i}\}$ *associated with Definition 36 and formulation* (7.31) *form also a corrector, which completes the convergence result of the non-conforming Petrov-Galerkin formulation of the MsFEM in the monotone case.*

7.5 Conclusion

In numerical homogenization, the choice of the boundary conditions for the problem at the micro scale in order to speed up the convergence of the numerical homogenization process is a difficult issue. It has been discussed a lot in the literature: e.g. in [136] for the community of *applied mathematics* and [154] for the community of *mechanics*. An alternative issue is given by windowing, whose aim is precisely to minimize the effect of the boundary conditions of the micro scale problem. In the classical periodic and stochastic cases, windowing has been proved to give better theoretical and numerical results, independently of the boundary conditions used. In the present work, we have extended the convergence results of Chapter 6 to the case of windowing. This has allowed us to prove the convergence of advanced numerical methods such as HMM with windowing and the nonconforming Petrov-Galerkin formulation of MsFEM in a general setting. To sum up, numerical homogenization methods with windowing do indeed converge. In addition, windowing may improve the convergence of the numerical methods in two ways. Concerning the approximation of the homogenized energy, windowing does not improve the convergence rate in general but may improve the prefactor. For the numerical corrector however, both the convergence rate and the prefactor may be improved.

Part III

Discrete to continuum limits

8

Exact bounds for the effective behaviour of a 'discrete' polycrystal

Summary. In a recent paper by Braides and Francfort, the problem of the characterization of the overall properties of lattice energies describing networks with arbitrary mixtures of two types of linear conductors has been addressed in a two-dimensional setting. In this paper we investigate the connection between that discrete optimization process and the theory of bounds for mixtures of continuum energies, for which the choice of the relationships between the different phases of the mixture is unusual and leads to remarkably simple results in terms of G-closure.

8.1 Introduction

In a paper by Braides and Francfort [39], the problem of the characterization of the overall properties of lattice energies describing networks with arbitrary mixtures of two types of linear conductors has been addressed. In a two-dimensional setting, with the notation of Γ-convergence, that problem translates into the limit analysis as $h \to 0$ of discrete energies of the type

$$E_h(u) = \frac{1}{2} \sum_{(i,j)\in\mathcal{N}_h} c_{ij}^h (u_i - u_j)^2 \qquad u : \mathcal{N}_h \to \mathbb{R},$$

where $\mathcal{N}_h$ denotes the set of *nearest neighbours* (i.e., pairs (i,j) such that $|i-j| = h$) on a portion of a square lattice $h\mathbb{Z}^2 \cap \Omega$, Ω being a regular open subset of $\mathbb{R}^2$, and

$$c_{ij}^h \in \{\alpha, \beta\}$$

are chosen arbitrarily. It should be noted that in the extreme cases $c_{ij}^h \equiv \alpha$ and $c_{ij}^h \equiv \beta$ the functionals above are simply discretizations of the Dirichlet integral with the corresponding coefficient. Since the analogue of the 'localization principle' holds in this discrete setting, the Γ-limit problem translates in a sort of 'discrete G-closure problem', where bounds on a matrix given by a 'discrete homogenization formula' must be exhibited.

The results in [39] show that all possible limits of E_h have the form

$$F(u) = \int_\Omega \langle A(x)\nabla u, \nabla u\rangle \, dx \qquad u \in H^1(\Omega),$$

with $A(x)$ belonging to a set of matrices determined only by the local limit proportion $\theta(x)$ at x of 'α-connections'. Such sets can be compared with those obtained by limits of mixtures of isotropic energies; i.e., of functionals of the form

$$E_\epsilon(u) = \int_\Omega \big(\alpha\chi_{\Omega_\epsilon} + \beta(1-\chi_{\Omega_\epsilon})\big)|\nabla u|^2 \, dx \qquad u \in H^1(\Omega),$$

with Ω_ϵ arbitrary measurable subsets of Ω (χ_{Ω_ϵ} denotes the corresponding characteristic function). These limits have the same form F with $A(x)$ now belonging to a set whose form depends on the

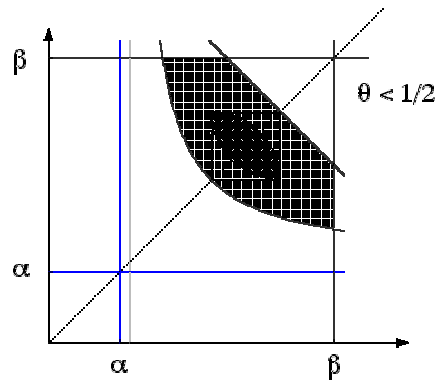

Fig. 8.1. Comparison of the bounds for discrete and isotropic continuous materials

limit volume fraction $\theta(x)$ of the α-phase; i.e., on the value at x of the weak*-limit of χ_{Ω_ϵ} (see Murat and Tartar [148, 149] and Cherkaev and Lurie [131]). In Figure 8.1 the two problems are compared in terms of the possible eigenvalues of $A(x)$ for some $\theta(x) < 1/2$: the shaded sets represent all possible pairs of eigenvalues in the two cases, the smaller set being the 'continuous' one (see Tartar [171, 172]). This comparison shows that the discrete energies cannot be simply interpreted as a mixture of the two 'extremal' isotropic energies.

The discrete functionals can be heuristically interpreted as 'anisotropic' continuous ones in the following way, pictured in Figure 8.2. We introduce a triangulation with underlying lattice $h\mathbb{Z}^2$,

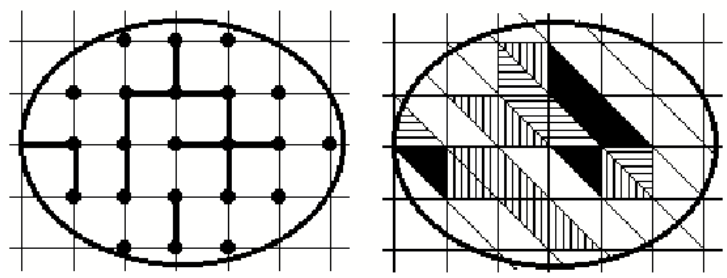

Fig. 8.2. a conducting network and the corresponding continuous coefficients

and to every triangle T with a horizontal side corresponding to a connection with value $c^h_{ij} = c_h$ and a vertical side corresponding to a connection with $c^h_{ij} = c_v$ we associate the matrix

$$\widetilde{A}_h(x) = \begin{pmatrix} c_h & 0 \\ 0 & c_v \end{pmatrix} \qquad \text{for } x \in T.$$

In Figure 8.2, on the left, 'thick connections' correspond to the coefficient β (the other ones to α), while on the right on black triangles we take $\widetilde{A}_h(x) = \beta I$, on white triangles $\widetilde{A}_h(x) = \alpha I$, and on the remaining triangles the anisotropic conductivities. Note that $\widetilde{A}_h$ takes only four possible values. Chosing $\tilde{u}_h \in H^1(\Omega)$ as the piecewise affine interpolation of u_h defined on $\mathcal{N}_h$, we have

$$E_h(u_h) = \widetilde{E}_h(\tilde{u}_h) + o(1),$$

(the last term deriving from an asymptotically negligible boundary term; see *e.g.* [4]) where

$$\widetilde{E}_h(u) = \int_\Omega \langle \widetilde{A}_h(x)\nabla u, \nabla u\rangle \, dx \qquad u \in H^1(\Omega).$$

Note that the infimum of $\widetilde{E}_h$ on $H^1(\Omega)$ is in general strictly lower than the corresponding one for E_h provided compatible boundary conditions are given. On the other hand, any continuous configuration can be approximated by a discrete configuration (let simply think of the discretization of the Dirichlet integral). Therefore, the Γ-closure of the discrete energies E_h contains the Γ-closure of the continuous energies $\widetilde{E}_h$. Whether this inclusion is strict or not is unclear. However, we will show that for polycristalline structures the two closures coincide.

Continuous energies as $\widetilde{E}_h$ can also be thought of as being a polycrystalline mixture of *three* conducting materials by regarding one anisotropic value of $\widetilde{A}_h$ as the rotation of the other one, so that we may rewrite

$$\widetilde{E}_h(u) = \int_\Omega \langle A_h(x) R_h(x) \nabla u(x), R_h(x) \nabla u(x) \rangle dx, \qquad u \in H^1(\Omega),$$

with $A_h(x)$ taking one of the values

$$\begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}, \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}, \tag{8.1}$$

and the rotation matrix $R_h(x)$ being either the identity or the rotation by ninety degrees. A more natural setting for the continuous energy is obtained by removing the geometric restrictions due to the lattice symmetries, and determine all the possible conductivity tensors that can be obtained as possible Γ-limits of energies

$$F_\epsilon(u) = \int_\Omega \langle A_\epsilon(x) R_\epsilon(x) \nabla u(x), R_\epsilon(x) \nabla u(x) \rangle dx, \tag{8.2}$$

with A_ϵ and R_ϵ families of matrices, with A_ϵ taking the three values above, and R_ϵ arbitrary rotations. If one adds the restriction that R_ϵ be piecewise-constant, then we may interpret this continuous energy as a limit of discrete polycrystals, that is a polycrystal obtained by a discrete to continuum process. This discrete polycrystalline material is composed of grains of diameter much larger than the lattice spacing h, at whose interior we consider a network structure whose lattice orientation is described by the rotation matrix $R_\epsilon(x)$, as depicted in Figure 8.3 (the black and grey subdomains represent isotropic parts of the energy whereas the other subdomains correspond to underlying lattices oriented according to the arrows). In this setting, the characterization of the Γ-closure of such discrete polycrystals is complete since the additional invariance by rotation of the Γ-limit allows us to focus on diagonal matrices, which have been fully described in [38].

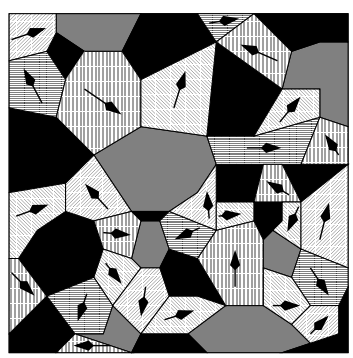

Fig. 8.3. Polycrystal

The scope of the present paper is twofold: to prove that the bounds obtained by considering piecewise-constants rotations R_ϵ still hold for general rotations, and that these bounds - which are optimal in the discrete setting - are also attained in this continuous setting. This amounts to proving that the Γ-closures of discrete and continuous polycrystals are the same. In what follows,

these two types of polycrystal will be refered to as 'discrete' since the Γ-closures are related to the very special relationship between the phases, that is inherited from the discrete description.

From a modeling viewpoint, these energies can equivalently be thought of as describing a bilayered polycrystal made of two types of layers (namely $\mathcal{A}$ and $\mathcal{B}$). Each layer has a conductivity tensor $C\nu \otimes \nu$, where $\nu \in S^1(\mathbb{R}^2)$, $C = \alpha$ for type $\mathcal{A}$ and $C = \beta$ for type $\mathcal{B}$, with $\beta > \alpha > 0$. Each simple crystal is a bilayer $C_1\nu_1 \otimes \nu_1 + C_2\nu_2 \otimes \nu_2$ with $C_i \in \{\alpha, \beta\}$ obtained by the superposition of two simple layers in orthogonal directions $\nu_1 \perp \nu_2$, such that the conductivity tensor of the crystal is of the form above (8.1). The association of rotated simple crystals gives rise to the polycrystal under investigation.

It is convenient to rewrite the energies F_ϵ in (8.2) in a different form, as

$$\begin{aligned} F_\epsilon(u) = \alpha \int_{E_1^\epsilon} |\nabla u|^2\, dx + \beta \int_{\Omega \setminus E_2^\epsilon} |\nabla u|^2\, dx \\ + \int_{E_2^\epsilon \setminus E_1^\epsilon} \left\langle \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} R_\epsilon(x)\nabla u(x), R_\epsilon(x)\nabla u(x) \right\rangle dx, \end{aligned} \tag{8.3}$$

where $E_1^\epsilon \subset E_2^\epsilon$. The set E_1^ϵ is the part of the polycrystal with isotropic conductivity α, $\Omega \setminus E_2^\epsilon$ is the part of the polycrystal with isotropic conductivity β, and $E_2^\epsilon \setminus E_1^\epsilon$ is where each single crystal has an anistropic conductivity of type $\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$, up to a rotation.

In order to describe the limit of F_ϵ, up to the extraction of a subsequence, we may suppose that $\chi_{E_1^\epsilon} \rightharpoonup \theta_1$ and $\chi_{E_2^\epsilon} \rightharpoonup \theta_2$ weakly* in $L^\infty(\Omega)$, and we define $\theta = \dfrac{\theta_1 + \theta_2}{2}$. In the notation above the value $\theta(x)$ represents the fixed limit local proportion of layers of type $\mathcal{A}$ at the point x, and is the analog of the local proportion of α-connections in the discrete setting. With the classical notation of G-closure, we would denote by θ_α the volume fraction of isotropic conductivity α, by θ_β the volume fraction of isotropic conductivity β and by $\theta_{\alpha,\beta} = 1 - \theta_\alpha - \theta_\beta$ the volume fraction of anisotropic conductivity. The problem we are interested in is not the classical G-closure of a three phase mixture, which is a very tough issue. Actually our discrete setting imposes an additional relationship between θ_α, θ_α and $\theta_{\alpha,\beta}$, namely $\theta_\alpha + \dfrac{\theta_{\alpha,\beta}}{2} = \theta$, or equivalently $\theta_\beta + \dfrac{\theta_{\alpha,\beta}}{2} = 1 - \theta$, where θ is fixed by the local proportion of α-connections. Our aim is then to characterize the Γ-limits for all values of $\theta \in [0, 1]$. Surprisingly, this additional relationship trivializes the associated G-closure problem, which can be interpreted as a 'constrained' three-phase mixture G-closure problem.

The paper is organized as follows. In the second section, we give a rigorous formulation of the result that we prove in two steps: first we derive bounds on the eigenvalues of $A^0(x)$, then we show these bounds to be optimal. We perform a derivation of these bounds following a purely variational approach to highlight how our parameter θ intervenes in their computations, so that the optimality of those bounds is simply obtained by making the previous reasoning sharp. This proof is direct and self-contained, and can be compared with the one in the discrete setting in [39]. In the third section, we generalize the result to the uniaxial three-dimensional case. The bounds we get in the continuous case are indeed the 'trivial bounds' usually obtained in the continuum theory of bounds (for a general exposition of results connected to the computation of bounds for the effective properties of composites we refer to the monograph by Milton [135]), which may be surprising at the first glance since we deal with a three-phase mixture. The last section is dedicated to the reformulation and interpretation of the present results in terms of a constrained G-closure problem, where we recall the more classical proof of the trivial bounds and generalize the results.

8.2 Main result and derivation of bounds in two dimensions

8.2.1 Optimal bounds

Having set the notation in the Introduction, we may describe the Γ-limits of functionals F_ϵ as follows.

Theorem 51 *Let F_ϵ be defined by* (8.3)*, let $\chi_{E_1^\epsilon} \rightharpoonup \theta_1$ and $\chi_{E_2^\epsilon} \rightharpoonup \theta_2$ weakly* in $L^\infty(\Omega)$, and let $\theta = \dfrac{\theta_1+\theta_2}{2}$. Up to the extraction of a (further) subsequence F_ϵ Γ-converges to an energy of the form*

$$F^0(u) = \int_\Omega \langle A^0(x)\nabla u, \nabla u\rangle\, dx \tag{8.4}$$

on $H^1(\Omega)$, where for almost all $x \in \Omega$ we have $A^0(x) \in \mathcal{H}^2_d(\theta(x))$, and for fixed $\theta \in [0,1]$ the set $\mathcal{H}^2_d(\theta)$ is defined as the set of symmetric 2×2 matrices whose eigenvalues λ_1, λ_2 satisfy the bounds

$$\alpha \le \lambda_1, \lambda_2 \le \beta, \qquad \lambda_1+\lambda_2 \le 2\left(\theta\alpha + (1-\theta)\beta\right), \qquad \frac{1}{\lambda_1}+\frac{1}{\lambda_2} \le 2\left(\frac{\theta}{\alpha}+\frac{1-\theta}{\beta}\right) \tag{8.5}$$

Conversely, for all measurable θ and A^0 such that $A^0(x) \in \mathcal{H}^2_d(\theta(x))$, there exist E_i^ϵ as above such that F_ϵ Γ-converges to F^0 in (8.4)*.*

Theorem 51 states that the effective overall conductivity of the polycrystal can be characterized in terms of the local proportion of layers of type $\mathcal{A}$. As motivated in the Introduction, these bounds correspond to the bounds obtained for a square conducting lattice in [39] for diagonal matrices, or for a discrete polycrystal. With this observation in mind we have added the subscript 'd' for *discrete* in the notation above. In terms of those bounds it must be noted that the optimal micro-geometries correspond to lattices locally oriented in the directions of the eigenvectors of the matrix $A^0(x)$.

For fixed $\theta \in [0,1]$ we will denote by $\mathcal{H}^2_{\rm hom}(\theta)$ the set of all *homogenized matrices* obtained by the homogenization of a periodic polycrystal as above with underlying volume fraction θ of $\mathcal{A}$-layers; i.e., of all matrices B that satisfy the equality

$$\begin{aligned}\langle B\xi,\xi\rangle = \inf\Bigg\{&\alpha\int_{E_1}|\xi+\nabla u|^2\,dy + \beta\int_{(0,1)^2\setminus E_2}|\xi+\nabla u|^2\,dy\\ &+\int_{E_2\setminus E_1}\left\langle\begin{pmatrix}\alpha & 0\\ 0 & \beta\end{pmatrix}R(y)(\xi+\nabla u(y)), R(y)(\xi+\nabla u(y))\right\rangle dy : u\in H^1_\#((0,1)^2)\Bigg\},\end{aligned} \tag{8.6}$$

for some rotation $R(y)$, and subsets $E_1 \subset E_2 \subset (0,1)^2$, with $|E_1|+|E_2| = 2\theta$. Here and after $H^1_\#((0,1)^2)$ denotes the space of $H^1_{\rm loc}(\mathbb{R}^2)$ functions that are 1-periodic in the coordinate directions. Note that the energy $\int_\Omega \langle B\nabla u, \nabla u\rangle\, dx$ corresponding to such B is a particular Γ-limit of a family of the type (8.3), with $E_i^\epsilon = \epsilon E_i$ (E_i are extended by 1-periodicity) and $R_\epsilon(y) = R(y/\epsilon)$. The triplet (E_1, E_2, R) is called the (*micro-*)*geometry* of the corresponding periodic polycrystal.

First, note that by a general compactness result for the Γ-convergence of quadratic forms ([38], [36], [56], [37] Theorem 4.13) we obtain that, up to subsequences, F_ϵ Γ-converges to some F^0 represented as in (8.4) for some $A^0(x)$. By the well known *localization principle* (see [37] Section 5.4.1 for a simple version, or [164] for a slightly more general one), we have that for almost every $x \in \Omega$ and for all $\xi \in \mathbb{R}^2$,

$$A^0(x) \in \mathcal{H}^2_{\rm hom}(\theta(x)+o(1)) + o(1); \tag{8.7}$$

i.e., for every $\eta > 0$ there exist R, E_1, and E_2, with $||E_1|+|E_2|-2\theta(x)| < \eta$, such that if B is as above then $|B - A^0(x)| \le o(1)$ as $\eta \to 0$.

In the rest of the section we will prove that $\mathcal{H}^2_{\rm hom}(\theta) = \mathcal{H}^2_d(\theta)$. From this inequality, the continuity of the bounds in (8.5), and the closure of $\mathcal{H}^2_d(\theta)$, it then follows that $A^0(x) \in \mathcal{H}^2_d(\theta(x)) =$

$\mathcal{H}^2_{\rm hom}(\theta(x))$ for a.a. $x \in \Omega$ and the first part of Theorem 51. We will therefore fix E_1, E_2, and R as in (8.6), and introduce a rotation angle ϕ. so that we canwrite R as

$$R(y) = \begin{pmatrix} \cos(\phi(y)) & \sin(\phi(y)) \\ -\sin(\phi(y)) & \cos(\phi(y)) \end{pmatrix}.$$

Note that the set $\mathcal{H}^2_{\rm hom}(\theta)$ is invariant by multiplication by rotations. Since it is also composed of symmetric matrices, this set is consequently characterized by the set of all possible eigenvalues of such matrices. The bounds are therefore derived in the plane of the eigenvalues (λ_1, λ_2) of matrices in $\mathcal{H}^2_{\rm hom}(\theta)$.

8.2.2 Simple bounds

Proposition 8.1. *For $i = 1, 2$ we have $\alpha \le \lambda_i \le \beta$.*

These easy bounds follow immediately by the pointwise estimate $\alpha\,{\rm Id} \le A_\epsilon(y) \le \beta\,{\rm Id}$.

8.2.3 Bounds from arithmetic means

Proposition 8.2. *We have*

$$\lambda_1 + \lambda_2 \le 2\left(\alpha\theta + (1-\theta)\beta\right). \tag{8.8}$$

Proof. Take $\xi = e_i$ and the test function $u = 0$ in (8.6). We then have for $i = 1, 2$

$$\lambda_i \le \alpha|E_1| + \beta(1 - |E_2|) + \int_{E_2 \setminus E_1} \left\langle \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} R(y)e_i, R(y)e_i \right\rangle dy \tag{8.9}$$

In addition, since $\{e_1, e_2\}$ is an orthonormal basis of $\mathbb{R}^2$ and $R(y)$ is a rotation, for almost every $y \in E_2 \setminus E_1$ we have

$$\left\langle \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} R(y)e_1, R(y)e_1 \right\rangle + \left\langle \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} R(y)e_2, R(y)e_2 \right\rangle = \alpha + \beta. \tag{8.10}$$

Indeed the first term of (8.10) reads $\alpha\cos^2\phi + \beta\sin^2\phi$ and the second term $\beta\cos^2\phi + \alpha\sin^2\phi$. The combination of (8.9) and (8.10) gives

$$\lambda_1 + \lambda_2 \le 2\alpha|E_1| + 2\beta(1 - |E_2|) + (|E_2| - |E_1|)(\alpha + \beta), \tag{8.11}$$

which yields (8.8) using $\theta = \dfrac{|E_2| + |E_1|}{2}$.

8.2.4 Bounds from harmonic means

Proposition 8.3. *We have*

$$\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \le 2\left(\frac{\theta}{\alpha} + \frac{1-\theta}{\beta}\right). \tag{8.12}$$

The rest of the section is devoted to proving this bound, relying on two arguments. The first one is a slicing argument, which is based on the simple inequality $|\nabla u|^2 \ge (\partial_i u)^2$ and allows to reduce to one-dimensional problems. The second tool is some convexity inequality.

Slicing argument

We first prove the following inequalities:

$$\begin{aligned}
\lambda_1 \geq \int_0^1 \Bigg(& \frac{1}{\alpha}|E_1^1(y_2)| + \frac{1}{\beta}(1-|E_2^1(y_2)|) \\
& + \int_{E_2^1(y_2)\setminus E_1^1(y_2)} \frac{\alpha \sin^2\phi(y) + \beta\cos^2\phi(y)}{\alpha\beta} dy_1 \Bigg)^{-1} dy_2, \\
\lambda_2 \geq \int_0^1 \Bigg(& \frac{1}{\alpha}|E_1^2(y_1)| + \frac{1}{\beta}(1-|E_2^2(y_1)|) \\
& + \int_{E_2^2(y_1)\setminus E_1^2(y_1)} \frac{\beta \sin^2\phi(y) + \alpha\cos^2\phi(y)}{\alpha\beta} dy_2 \Bigg)^{-1} dy_1,
\end{aligned} \tag{8.13}$$

where

$$\begin{aligned}
E_1^1(y_2) &= \{y_1 \in (0,1), \chi_{E_1}(y_1,y_2) = 1\}, & E_2^1(y_2) &= \{y_1 \in (0,1), \chi_{E_2}(y_1,y_2) = 1\}, \\
E_1^2(y_1) &= \{y_2 \in (0,1), \chi_{E_1}(y_1,y_2) = 1\}, & E_2^2(y_1) &= \{y_2 \in (0,1), \chi_{E_2}(y_1,y_2) = 1\},
\end{aligned}$$

and χ_E denotes the *characteristic function* of the set E. These sets satisfy

$$\int_0^1 |E_1^1(y_2)|dy_2 = \int_0^1 |E_1^2(y_1)|dy_1 = \theta_1 \text{ and } \int_0^1 |E_2^1(y_2)|dy_2 = \int_0^1 |E_2^2(y_1)|dy_1 = \theta_2.$$

To get (8.13), we first prove the following result.

Proposition 8.4. *For all $(u_1,u_2) \in \mathbb{R}^2$ and $\phi \in (0,2\pi)$, we have*

$$\left\langle \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} R(\phi)(u_1e_1+u_2e_2), R(\phi)(u_1e_1+u_2e_2) \right\rangle \geq u_1^2 \frac{\alpha\beta}{\alpha\sin^2\phi + \beta\cos^2\phi}, \tag{8.14}$$

$$\left\langle \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} R(\phi)(u_1e_1+u_2e_2), R(\phi)(u_1e_1+u_2e_2) \right\rangle \geq u_2^2 \frac{\alpha\beta}{\beta\sin^2\phi + \alpha\cos^2\phi}, \tag{8.15}$$

where $\{e_1,e_2\}$ is the canonical basis of $\mathbb{R}^2$.

Proof. It is enough to prove (8.14). Denote by $\chi_\phi(u_1,u_2)$ the left-hand side of (8.14). A direct calculation shows that

$$\chi_\phi(u_1,u_2) = \alpha(u_1\cos\phi - u_2\sin\phi)^2 + \beta(u_1\sin\phi + u_2\cos\phi)^2.$$

At fixed u_1, $v \mapsto \chi_\phi(u_1,v)$ is quadratic and convex. This function attains its minimum on $\mathbb{R}$, at

$$v_2 = u_1 \frac{\alpha-\beta}{\alpha\sin^2\phi + \beta\cos^2\phi} \sin\phi\cos\phi.$$

We then have

$$\begin{aligned}
\chi_\phi(u_1,v_2) &= u_1^2 \left(\alpha\cos^2\phi + \beta\sin^2\phi - \frac{(\beta-\alpha)^2\sin^2\phi\cos^2\phi}{\alpha\sin^2\phi+\beta\cos^2\phi} \right) \\
&= u_1^2 \frac{\alpha\beta\left(\cos^2\phi+\sin^2\phi\right)^2}{\alpha\sin^2\phi+\beta\cos^2\phi} \\
&= u_1^2 \frac{\alpha\beta}{\alpha\sin^2\phi+\beta\cos^2\phi}
\end{aligned}$$

as desired.

We now integrate in the e_2 direction. If u is such that $\nabla u = e_1 + \nabla w$, $w \in H^1_\#((0,1))$, we then have

$$
\begin{aligned}
\lambda_1 = \int_0^1 \Bigg(& \alpha \int_{E_1^1(y_2)} |\nabla u|^2 dy_1 + \beta \int_{(0,1)\setminus E_2^1(y_2)} |\nabla u|^2 dy_1 \\
& + \int_{E_2^1(y_2)\setminus E_1^1(y_2)} \left\langle \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} R(y)\nabla u, R(y)\nabla u \right\rangle dy_1 \Bigg) dy_2 \\
\geq \int_0^1 \Bigg(& \alpha \int_{E_1^1(y_2)} \partial_1 u^2 dy_1 + \beta \int_{(0,1)\setminus E_2^1(y_2)} \partial_1 u^2 dy_1 \\
& + \int_{E_2^1(y_2)\setminus E_1^1(y_2)} dy_1 \frac{\partial_1 u^2 \alpha\beta}{\alpha \sin^2 \phi(y) + \beta \cos^2 \phi(y)} \Bigg) dy_2 \\
\geq \int_0^1 \inf \Bigg\{ & \alpha \int_{E_1^1(y_2)} v'(y_1)^2 dy_1 + \beta \int_{(0,1)\setminus E_2^1(y_2)} v'(y_1)^2 dy_1 \\
& + \int_{E_2^1(y_2)\setminus E_1^1(y_2)} \frac{v'(y_1)^2 \alpha\beta}{\alpha \sin^2 \phi(y) + \beta \cos^2 \phi(y)} dy_1 : v \in H_0^1((0,1)) \Bigg\} dy_2.
\end{aligned}
\tag{8.16}
$$

The infimum in the integral of (8.16) can be explicitly computed by using the following well-known result.

Proposition 8.5. *Let $f \in L^\infty((0,1))$ be such that* $\inf f > 0$*, then*

$$
\inf \left\{ \int_0^1 f(y) v'(y)^2 : v \in H^1(0,1), v(0) = 0, v(1) = 1 \right\} = \left(\int_0^1 \frac{dy}{f(y)} \right)^{-1} \tag{8.17}
$$

Inserting formula (8.17) in (8.16), we obtain

$$
\begin{aligned}
\lambda_1 \geq \int_0^1 \Bigg(& \frac{1}{\alpha} |E_1^1(y_2)| + \frac{1}{\beta}(1 - |E_2^1(y_2)|) \\
& + \int_{E_2^1(y_2)\setminus E_1^1(y_2)} \frac{\alpha \sin^2 \phi(y) + \beta \cos^2 \phi(y)}{\alpha\beta} dy_1 \Bigg)^{-1} dy_2.
\end{aligned}
\tag{8.18}
$$

The same proof holds for λ_2.

To conclude the proof of the bounds, we use some convexity properties.

A convexity inequality

Proposition 8.6. *Let $b > a > 0, c > 0, \gamma > 0$ and $f \in L^1((a,b), \mathbb{R}^+)$, then*

$$
\int_a^b \frac{dy}{c + f(y)} \geq \frac{(b-a)^2}{\displaystyle\int_a^b (c + f(y)) dy} \tag{8.19}
$$

This result just amounts to stating that the harmonic mean is lower than the arithmetic mean, and it is immediately obtained by Hölder's inequality.

Applying Proposition 8.6 to (8.13), for $a = 0$, $b = 1$ and $c = 1/\alpha$, we obtain

$$
\begin{aligned}
\lambda_1 &\geq \left[\int_0^1 \left(\frac{1}{\alpha}|E_1^1(y_2)| + \frac{1}{\beta}(1-|E_2^1(y_2)|)\right.\right. \\
&\qquad \left.\left. + \int_{E_2^1(y_2)\setminus E_1^1(y_2)} \frac{\alpha\sin^2\phi(y)+\beta\cos^2\phi(y)}{\alpha\beta}dy_1\right)dy_2\right]^{-1} \\
&= \left(\frac{\theta_1}{\alpha} + \frac{1-\theta_2}{\beta} + \int_{E_2\setminus E_1} \frac{\alpha\sin^2\phi(y)+\beta\cos^2\phi(y)}{\alpha\beta}dy\right)^{-1} \\
\lambda_2 &\geq \left[\int_0^1 \left(\frac{1}{\alpha}|E_1^2(y_1)| + \frac{1}{\beta}(1-|E_2^2(y_1)|)\right.\right. \\
&\qquad \left.\left. + \int_{E_2^2(y_1)\setminus E_1^2(y_1)} \frac{\beta\sin^2\phi(y)+\alpha\cos^2\phi(y)}{\alpha\beta}dy_2\right)dy_1\right]^{-1} \\
&= \left(\frac{\theta_1}{\alpha} + \frac{1-\theta_2}{\beta} + \int_{E_2\setminus E_1} \frac{\beta\sin^2\phi(y)+\alpha\cos^2\phi(y)}{\alpha\beta}dy\right)^{-1}
\end{aligned}
\tag{8.20}
$$

Therefore we have

$$
\begin{aligned}
\frac{1}{\lambda_1} + \frac{1}{\lambda_2} &\leq 2\frac{\theta_1}{\alpha} + 2\frac{1-\theta_2}{\beta} + \int_{E_2\setminus E_1}\left(\frac{\alpha\cos^2\phi+\beta\sin^2\phi}{\alpha\beta}\right. \\
&\qquad \left. + \frac{\beta\cos^2\phi+\alpha\sin^2\phi}{\alpha\beta}\right)dy \\
&= 2\frac{\theta_1}{\alpha} + 2\frac{1-\theta_2}{\beta} + (\theta_2-\theta_1)\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)
\end{aligned}
$$

which exactly reads as (8.12).

8.3 Optimality of the bounds

We now prove the last statement of Theorem 51. By approximation with piecewise-constant functions, it is sufficient to consider the case of θ and A^0 constant; i.e., it will be sufficient to show that for all symmetric A^0 with eigenvalues satisfying (8.5) we find a suitable geometry such that each $\langle A^0\xi,\xi\rangle$ is represented as a minimum problem as in (8.6). By rotational invariance, it suffices to consider diagonal A^0. Furthermore, it is not restrictive to treat the case of eigenvalues belonging to the boundary of the set defined by (8.6) only (the 'extremal' cases). In fact, using a well-known "lamination formula" (see [171,172] Proposition 3) we may easily construct geometries for a generic diagonal A^0 using those 'extremal geometries'.

We will consider polycrystals $(0,1)^2$ defined by $A(y) = \begin{pmatrix} a_1(y) & 0 \\ 0 & a_2(y)\end{pmatrix}$ for $y \in (0,1)^2$, and the corresponding homogenized A^0, defined by

$$
\left\langle A^0\xi,\xi\right\rangle = \inf\left\{\int_{(0,1)^2} \left\langle A(y)(\xi+\nabla u(y)),\xi+\nabla u(y)\right\rangle, u \in H^1_\#((0,1)^2)\right\} \tag{8.21}
$$

for all $\xi \in \mathbb{R}^2$. Note that in terms of the micro-geometry of the polycrystal the rotation R is either the identity or a rotation of ninety degrees.

8.3.1 Arithmetic bound

Let $\theta_1 \in (0,1)$ and $\theta_2 \in (0,1)$. We define $A(y)$ by $a_1(y) = \alpha$ if $y_2 \leq \theta_1$, $a_1(y) = \beta$ if $y_2 > \theta_1$, and $a_2(y) = \alpha$ if $y_1 \leq \theta_2$, $a_2(y) = \beta$ if $y_1 > \theta_2$.

We will then prove that $A^0 = \begin{pmatrix} \bar{a}(\theta_1) & 0 \\ 0 & \bar{a}(\theta_2)\end{pmatrix}$, where $\bar{a}(\lambda) = \alpha\lambda + \beta(1-\lambda)$.

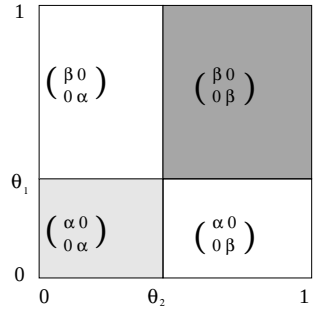

Fig. 8.4. Optimal geometry for the arithmetic bound

Testing with $u = 0$ in (8.21), we obtain $\langle A^0\xi, \xi\rangle \le \bar{a}(\theta_1)\xi_1^2 + \bar{a}(\theta_2)\xi_2^2$. To prove the converse inequality we use the slicing method, obtaining

$$\int_{(0,1)^2} \langle A(y)(\xi + \nabla u(y)), \xi + \nabla u(y)\rangle = \int_0^1 \left(\int_0^{\theta_1} \alpha(\xi_1 + \partial_1 u(y_1, y_2))^2 dy_2 \right) dy_1$$
$$+ \int_0^1 \left(\int_{\theta_1}^1 \beta(\xi_1 + \partial_1 u(y_1, y_2))^2 dy_2 \right) dy_1 + \int_0^1 \left(\int_0^{\theta_2} \alpha(\xi_2 + \partial_2 u(y_1, y_2))^2 dy_1 \right) dy_2$$
$$+ \int_0^1 \left(\int_{\theta_2}^1 \beta(\xi_2 + \partial_2 u(y_1, y_2))^2 dy_1 \right) dy_2. \tag{8.22}$$

Switching the order of the integrals and minimizing each term of the right-hand side of (8.22) separately yield the desired inequality, and thus the equality

$$\langle A^0\xi, \xi\rangle = \bar{a}(\theta_1)\xi_1^2 + \bar{a}(\theta_2)\xi_2^2.$$

If $\theta_2 + \theta_1 = 2\theta$ then $\bar{a}(\theta_1) + \bar{a}(\theta_2) = \alpha\theta + \beta(1 - \theta)$, which is the arithmetic bound.

8.3.2 Harmonic bound

For the harmonic bound, let us consider $a_1(y) = \alpha$ if $y_1 \le \theta_1$, $a_1(y) = \beta$ if $y_1 > \theta_1$, and $a_2(y) = \alpha$ if $y_2 \le \theta_2$, $a_2(y) = \beta$ if $y_2 > \theta_2$.

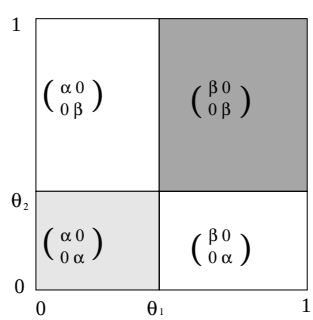

Fig. 8.5. Optimal geometry for the harmonic bound

We will then prove that $A^0 = \begin{pmatrix} \underline{a}(\theta_1) & 0 \\ 0 & \underline{a}(\theta_2) \end{pmatrix}$, where $\underline{a}(\lambda) = \left(\dfrac{\lambda}{\alpha} + \dfrac{1-\lambda}{\beta} \right)^{-1}$.

Let $u_1 : (0,1) \mapsto \mathbb{R}$ be the one-dimensional minimizer of

$$\inf\left\{\int_0^{\theta_1}\alpha(\xi_1+v'(z))^2dz+\int_{\theta_1}^1\beta(\xi_1+v'(z))^2dz : v(0)=v(1)=0\right\}=\underline{a}(\theta_1)\xi_1^2,$$

and $u_2:(0,1)\mapsto\mathbb{R}$ be the one-dimensional minimizer of

$$\inf\left\{\int_0^{\theta_2}\alpha(\xi_2+v'(z))^2dz+\int_{\theta_2}^1\beta(\xi_2+v'(z))^2dz : v(0)=v(1)=0\right\}=\underline{a}(\theta_2)\xi_2^2.$$

Let $u(y)=u_1(y_1)+u_2(y_2)$ be a test function in (8.21). We then have

$$\left\langle A^0\xi,\xi\right\rangle\le\underline{a}(\theta_1)\xi_1^2+\underline{a}(\theta_2)\xi_2^2.$$

Once more we prove the converse inequality using the slicing argument.

$$\begin{aligned}\int_{(0,1)^2}&\langle A(y)(\xi+\nabla u(y)),\xi+\nabla u(y)\rangle=\\&\int_0^1\left(\int_0^{\theta_1}\alpha(\xi_1+\partial_1u(y_1,y_2))^2dy_1+\int_{\theta_1}^1\beta(\xi_1+\partial_1u(y_1,y_2))^2dy_1\right)dy_2\\&+\int_0^1\left(\int_0^{\theta_2}\alpha(\xi_2+\partial_2u(y_1,y_2))^2dy_2+\int_{\theta_2}^1\beta(\xi_2+\partial_2u(y_1,y_2))^2dy_2\right)dy_1\end{aligned}\tag{8.23}$$

We obtain the desired inequality by optimizing pointwise the integrands of the right handside of (8.23).

If $\theta_2+\theta_1=2\theta$ then $\dfrac{1}{\underline{a}(\theta_1)}+\dfrac{1}{\underline{a}(\theta_2)}=\dfrac{\theta}{\alpha}+\dfrac{1-\theta}{\beta}$, which is the harmonic bound.

8.3.3 Optimal bounds

The simple bounds $\lambda_i=\alpha,\beta$ can always be reached by layering. Depending on θ, the shape of the optimal bounds may vary (see Figure 5 in [39], and the details therein). Once the arithmetic and harmonic bounds are shown to be attained, the convex set delimited by these bounds and the simple bounds can also be attained by layering.

8.4 Extension to higher dimension

Theorem 51 has a natural counterpart in higher dimension. We only consider the physically meaningful three-dimensional case, and especially the uniaxial basic crystal. In this case we deal with the integral functionals

$$\begin{aligned}F_\epsilon(u)=\alpha\int_{E_1^\epsilon}|\nabla u|^2+\beta\int_{\Omega\setminus E_2^\epsilon}|\nabla u|^2&\\+\int_{E_3^\epsilon}\left\langle\begin{pmatrix}\alpha&0&0\\0&\beta&0\\0&0&\beta\end{pmatrix}R_\epsilon(x)\nabla u(x),R_\epsilon(x)\nabla u(x)\right\rangle dx&\\+\int_{(E_2^\epsilon\setminus E_1^\epsilon)\setminus E_3^\epsilon}\left\langle\begin{pmatrix}\beta&0&0\\0&\alpha&0\\0&0&\alpha\end{pmatrix}R_\epsilon(x)\nabla u(x),R_\epsilon(x)\nabla u(x)\right\rangle dx,&\end{aligned}\tag{8.24}$$

where $E_1^\epsilon\subset E_2^\epsilon$, $E_3^\epsilon\subset E_2^\epsilon\setminus E_1^\epsilon$ and R_ϵ is a rotation.

As in the two-dimensional case, in order to describe the Γ-limit of F_ϵ, up to the extraction of a subsequence, we may suppose that $\chi_{E_1^\epsilon}\rightharpoonup\theta_1$, $\chi_{E_2^\epsilon}\rightharpoonup\theta_2$, and $\chi_{E_3^\epsilon}\rightharpoonup\theta_3$ weakly* in $L^\infty(\Omega)$, and we define $\theta=\dfrac{\theta_1+2\theta_2-\theta_3}{3}$, again the value $\theta(x)$ representing the fixed limit local proportion of layers of type $\mathcal{A}$ at the point x.

Theorem 52 *Let F_ϵ be defined by (8.24), let $\chi_{E_i^\epsilon} \rightharpoonup \theta_i$ weakly* in $L^\infty(\Omega)$, and let $\theta = \dfrac{\theta_1 + 2\theta_2 - \theta_3}{3}$. Up to the extraction of a (further) subsequence F_ϵ Γ-converges to an energy of the form*

$$F^0(u) = \int_\Omega \langle A^0(x)\nabla u, \nabla u\rangle dx \tag{8.25}$$

on $H^1(\Omega)$, where for almost all $x \in \Omega$ we have $A^0(x) \in \mathcal{H}_d^3(\theta(x))$, where for fixed $\theta \in [0,1]$ the set $\mathcal{H}_d^3(\theta)$ is defined as the set of symmetric 3×3 matrices whose eigenvalues $\lambda_1, \lambda_2, \lambda_3$ satisfy the bounds

$$\alpha \le \lambda_1, \lambda_2, \lambda_3 \le \beta, \qquad \lambda_1 + \lambda_2 + \lambda_3 \le 3\bar{a}(\theta), \qquad \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} \le \frac{3}{\underline{a}(\theta)} \tag{8.26}$$

Conversely, for all measurable θ and A^0 such that $A^0(x) \in \mathcal{H}_d^3(\theta(x))$, there exist E_i^ϵ as above such that F_ϵ Γ-converges to F^0 in (8.25).

We now adapt Sections 8.2 and 8.3 to prove Theorem 52.

Since the anisotropic conductivity matrices only have two eigenvalues α and β, in order to diagonalize the matrix it suffices to determine the eigenvector associated to the eigenvalue of multiplicity 1, the orthogonal subspace being an eigensubspace associated to the other eigenvalue. Therefore, we only need two angles for the rotations which can be written as

$$R(\phi,\psi) = \begin{pmatrix} \sin\psi & \cos\phi\cos\psi & \sin\phi\cos\psi \\ \cos\psi & -\cos\phi\sin\psi & -\sin\phi\sin\psi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix}.$$

The calculations are more involved in the three-dimensional case. We leave the details to the interested reader and only give the intermediate results.

8.4.1 Bounds from arithmetic means

Direct calculations give

$$\left\langle \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{pmatrix} R(\phi,\psi)e_1, R(\phi,\psi)e_1 \right\rangle = \alpha\sin^2\psi + \beta\cos^2\psi$$

$$\left\langle \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{pmatrix} R(\phi,\psi)e_2, R(\phi,\psi)e_2 \right\rangle = \alpha\cos^2\phi\cos^2\psi + \beta(\cos^2\phi\sin^2\psi + \sin^2\phi)$$

$$\left\langle \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{pmatrix} R(\phi,\psi)e_3, R(\phi,\psi)e_3 \right\rangle = \alpha\sin^2\phi\cos^2\psi + \beta(\sin^2\phi\sin^2\psi + \cos^2\phi),$$

whose sum is $\alpha + 2\beta$. The calculation for the other anisotropic conductivity matrix yields $\beta + 2\alpha$. Therefore,

$$\begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 &\le 3\theta_1\alpha + 3(1-\theta_2)\beta + \theta_3(\alpha + 2\beta) + (\theta_2 - \theta_1 - \theta_3)(2\alpha + \beta) \\ &= 3(\alpha\theta + (1-\theta)\beta), \end{aligned}$$

which is the desired arithmetic bound.

8.4.2 Bounds from harmonic means

For convenience, we will make use of c for cos, s for sin, and $v = (v_1, v_2, v_3)$ in the sequel. We set

$$H(v_1,v_2,v_3) := \left\langle \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{pmatrix} R(\phi,\psi)v, R(\phi,\psi)v \right\rangle$$

$$= \alpha(v_1\,\mathrm{s}_\psi + v_2\,\mathrm{c}_\phi\,\mathrm{c}_\psi + v_3\,\mathrm{s}_\phi\,\mathrm{c}_\phi)^2 + \beta(v_1\,\mathrm{c}_\psi - v_2\,\mathrm{c}_\phi\,\mathrm{s}_\psi - v_3\,\mathrm{s}_\phi\,\mathrm{s}_\psi)^2 + \beta(-v_2\,\mathrm{s}_\phi + v_3\,\mathrm{c}_\phi)^2$$

As in the two-dimensional case, we optimize H on $\mathbb{R}^2$, one variable being fixed. These minimizers exist since H is a symmetric positive definite quadratic form. The first derivatives of H with respect to v_1, v_2, v_3 read as

$$\begin{aligned}
\partial_1 H(v_1,v_2,v_3) &= v_1(\alpha\,\mathrm{s}_\psi^2 + \beta\,\mathrm{c}_\psi^2) + v_2\,\mathrm{c}_\phi\,\mathrm{c}_\psi\,\mathrm{s}_\psi(\alpha-\beta) + v_3\,\mathrm{s}_\phi\,\mathrm{c}_\psi\,\mathrm{s}_\psi(\alpha-\beta) \\
\partial_2 H(v_1,v_2,v_3) &= v_1(\alpha-\beta)\,\mathrm{c}_\phi\,\mathrm{c}_\psi\,\mathrm{s}_\psi + v_2(\alpha\,\mathrm{c}_\phi^2\,\mathrm{c}_\psi^2 + \beta\,\mathrm{c}_\phi^2\,\mathrm{c}_\psi^2 + \beta\,\mathrm{s}_\phi^2) \\
&\quad + v_3(\alpha-\beta)\,\mathrm{c}_\psi^2\,\mathrm{c}_\phi\,\mathrm{s}_\phi \\
\partial_3 H(v_1,v_2,v_3) &= v_1(\alpha-\beta)\,\mathrm{s}_\phi\,\mathrm{c}_\psi\,\mathrm{s}_\psi + v_2(\alpha-\beta)\,\mathrm{c}_\psi^2\,\mathrm{c}_\phi\,\mathrm{s}_\phi + v_3(\alpha\,\mathrm{s}_\phi^2\,\mathrm{c}_\psi^2 \\
&\quad + \beta\,\mathrm{s}_\phi^2\,\mathrm{s}_\psi^2 + \beta\,\mathrm{c}_\phi^2)
\end{aligned}$$

We treat three cases, at fixed u_1, u_2 and u_3, respectively.

Case 1

We minimize $H(u_1,\cdot,\cdot)$ on $\mathbb{R}^2$. This is equivalent to solving

$$\begin{cases} \partial_2 H(u_1,v_2,v_3) = 0 \\ \partial_3 H(u_1,v_2,v_3) = 0. \end{cases}$$

The solution is given by

$$\begin{aligned}
v_2 &= u_1(\beta-\alpha)\frac{\mathrm{c}_\phi\,\mathrm{c}_\psi\,\mathrm{s}_\psi}{\alpha\,\mathrm{c}_\psi^2 + \beta\,\mathrm{s}_\psi^2} \\
v_3 &= u_1(\beta-\alpha)\frac{\mathrm{s}_\phi\,\mathrm{c}_\psi\,\mathrm{s}_\psi}{\alpha\,\mathrm{c}_\psi^2 + \beta\,\mathrm{s}_\psi^2}.
\end{aligned}$$

The minimum then reads

$$\begin{aligned}
\inf_{w_2,w_3} H(u_1,w_2,w_3) &= H(u_1,v_2,v_3) \\
&= \alpha\beta\frac{u_1^2}{(\alpha\,\mathrm{c}_\psi^2 + \beta\,\mathrm{s}_\psi^2)^2}(\beta\,\mathrm{s}_\psi^2 + \alpha\,\mathrm{c}_\psi^2) \\
&= u_1^2\frac{\alpha\beta}{\alpha\,\mathrm{c}_\psi^2 + \beta\,\mathrm{s}_\psi^2}.
\end{aligned} \tag{8.27}$$

Case 2

We minimize $H(\cdot,u_2,\cdot)$ on $\mathbb{R}^2$. This is equivalent to solving

$$\begin{cases} \partial_1 H(v_1,u_2,v_3) = 0 \\ \partial_3 H(v_1,u_2,v_3) = 0. \end{cases}$$

The solution is given by

$$\begin{aligned}
v_1 &= u_2(\beta-\alpha)\frac{\mathrm{c}_\psi\,\mathrm{c}_\phi\,\mathrm{s}_\psi}{\mathrm{c}_\phi^2(\alpha\,\mathrm{s}_\psi^2 + \beta\,\mathrm{c}_\psi^2) + \alpha\,\mathrm{s}_\phi^2} \\
v_3 &= u_2(\beta-\alpha)\frac{\mathrm{s}_\phi\,\mathrm{c}_\phi\,\mathrm{c}_\psi^2}{\mathrm{c}_\phi^2(\alpha\,\mathrm{s}_\psi^2 + \beta\,\mathrm{c}_\psi^2) + \alpha\,\mathrm{s}_\phi^2}.
\end{aligned}$$

The minimum then reads

$$\begin{aligned}\inf_{w_1,w_3} H(w_1,u_2,w_3) &= H(v_1,u_2,v_3)\\ &= \alpha\beta\frac{u_2^2}{(\,\mathrm{c}_\phi^2(\alpha\,\mathrm{s}_\psi^2+\beta\,\mathrm{c}_\psi^2)+\alpha\,\mathrm{s}_\phi^2)^2}(\beta\,\mathrm{c}_\phi^2\,\mathrm{c}_\psi^2+\alpha\,\mathrm{s}_\psi^2\,\mathrm{c}_\phi^2+\alpha\,\mathrm{s}_\phi^2)\\ &= u_2^2\frac{\alpha\beta}{\mathrm{c}_\phi^2(\alpha\,\mathrm{s}_\psi^2+\beta\,\mathrm{c}_\psi^2)+\alpha\,\mathrm{s}_\phi^2}. \end{aligned} \tag{8.28}$$

Case 3

Case 3 is the same as Case 2 up to switching c_ϕ and s_ϕ. Thus

$$\inf_{w_1,w_2} H(w_1,w_2,u_3) = u_3^2\frac{\alpha\beta}{\mathrm{s}_\phi^2(\alpha\,\mathrm{s}_\psi^2+\beta\,\mathrm{s}_\psi^2)+\alpha\,\mathrm{c}_\phi^2}. \tag{8.29}$$

Slicing argument and convexity inequalities

In the three-dimensional case, the slicing argument and inequality (8.28) imply, with obvious notation,

$$\begin{aligned}\lambda_2 = \int_{(0,1)^2}\Bigg(&\alpha\int_{E_1^2(y_1,y_3)}|\nabla u|^2dy_2+\beta\int_{(0,1)^2\setminus E_2^2(y_1,y_3)}|\nabla u|^2dy_2\\ &+\int_{\left[E_2^2(y_1,y_3)\setminus E_1^2(y_1,y_3)\right]\setminus E_3^2(y_1,y_3)}\left\langle\begin{pmatrix}\beta&0&0\\0&\alpha&0\\0&0&\alpha\end{pmatrix}R(y)\nabla u,R(y)\nabla u\right\rangle dy_2\\ &+\int_{E_3^2(y_1,y_3)}\left\langle\begin{pmatrix}\alpha&0&0\\0&\beta&0\\0&0&\beta\end{pmatrix}R(y)\nabla u,R(y)\nabla u\right\rangle dy_2\Bigg)\,dy_1dy_3\\ \geq\int_{(0,1)^2}\inf\Bigg\{&\alpha\int_{E_1^2(y_1,y_3)}v'(y_2)^2dy_2+\beta\int_{(0,1)^2\setminus E_2^2(y_1,y_3)}v'(y_2)^2dy_2\\ &+\int_{\left[E_2^2(y_1,y_3)\setminus E_1^2(y_1,y_3)\right]\setminus E_3^2(y_1,y_3)}\frac{v'(y_2)^2\alpha\beta}{\mathrm{c}_\phi^2(\beta\,\mathrm{s}_\psi^2+\alpha\,\mathrm{c}_\psi^2)+\beta\,\mathrm{s}_\phi^2}dy_2\\ &+\int_{E_3^2(y_1,y_3)}\frac{v'(y_2)^2\alpha\beta}{\mathrm{c}_\phi^2(\alpha\,\mathrm{s}_\psi^2+\beta\,\mathrm{c}_\psi^2)+\alpha\,\mathrm{s}_\phi^2}dy_2,\, v\in H_0^1((0,1))\Bigg\}\,dy_1dy_3\end{aligned}$$

Similar bounds can be derived for λ_1 and λ_3.

Next, we can use Proposition 8.5. In addition, Proposition 8.6 also holds in any dimension. Therefore

$$
\begin{aligned}
\frac{1}{\theta_1}+\frac{1}{\theta_2}+\frac{1}{\theta_3} &\le 3\theta_1\frac{1}{\alpha}+3(1-\theta_2)\frac{1}{\beta} \\
&\quad+\int_{E_3}\left(\frac{\alpha\,\mathrm{c}_\psi^2+\beta\,\mathrm{s}_\psi^2}{\alpha\beta}+\frac{\mathrm{c}_\phi^2(\alpha\,\mathrm{s}_\psi^2+\beta\,\mathrm{c}_\psi^2)+\alpha\,\mathrm{s}_\phi^2}{\alpha\beta}\right.\\
&\qquad\left.+\frac{\mathrm{s}_\phi^2(\alpha\,\mathrm{s}_\psi^2+\beta\,\mathrm{s}_\psi^2)+\alpha\,\mathrm{c}_\phi^2}{\alpha\beta}\right)dy \\
&\quad+\int_{(E_2\setminus E_1)\setminus E_3}\left(\frac{\beta\,\mathrm{c}_\psi^2+\alpha\,\mathrm{s}_\psi^2}{\alpha\beta}+\frac{\mathrm{c}_\phi^2(\beta\,\mathrm{s}_\psi^2+\alpha\,\mathrm{c}_\psi^2)+\beta\,\mathrm{s}_\phi^2}{\alpha\beta}\right.\\
&\qquad\left.+\frac{\mathrm{s}_\phi^2(\beta\,\mathrm{s}_\psi^2+\alpha\,\mathrm{s}_\psi^2)+\beta\,\mathrm{c}_\phi^2}{\alpha\beta}\right)dy \\
&= 3\theta_1\frac{1}{\alpha}+3(1-\theta_2)\frac{1}{\beta}+\theta_3\left(\frac{1}{\alpha}+\frac{2}{\beta}\right)+(\theta_2-\theta_1-\theta_3)\left(\frac{1}{\beta}+\frac{2}{\alpha}\right)\\
&= 3\left(\frac{\theta}{\alpha}+\frac{1-\theta}{\beta}\right),
\end{aligned}
$$

which is the desired harmonic bound.

8.4.3 Optimality of the bounds

Using the natural extensions of the two-dimensional test functions of Section 8.3 to three dimensions, one can show the harmonic and arithmetic bounds to be attained. The entire convex set is also attained by layering.

8.5 Interpretation in terms of G-closure

In this last section, we briefly recall a more classic way to derive the arithmetic and harmonic bounds, also known as the Voigt and Reuss bounds. We then interpret our results in terms of quasiconvex envelope and attainment of infima following the seminal papers of Kohn and Strang [116], based on the vectorization of the problem by Tartar [148,149], and further studied by Allaire and Francfort [9].

8.5.1 Derivation of the trivial bounds

Let $n \in \mathbb{N}$, $\xi \in \mathcal{M}_n(\mathbb{R})$ and consider

$$
H_0^1((0,1)^n,\mathbb{R}^n) \ni u \mapsto F(\xi,\nabla u) = \int_{(0,1)^n} \langle A(y)(\xi+\nabla u),\xi+\nabla u\rangle\,dy
$$

The arithmetic bound is a consequence of

$$
\inf\{F(\xi,\nabla u) : u \in H_0^1((0,1)^n)\} \le F(\xi,0)
$$

for $\xi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The derivation of the harmonic bound is related to the same minimization on an enlarged space, namely on

$$
L^2_{\#}((0,1)^n)^{n\times n} = \left\{ v \in L^2((0,1)^n)^{n\times n}\,,\ \int_{(0,1)^n} v\,dy = 0 \right\}.
$$

In this case, a direct computation shows

$$\inf\{F(\xi,\nabla u) : u \in H_0^1((0,1)^n\} \geq \inf\{F(\xi,v) : v \in L^2_{\#}((0,1)^n)^{n\times n}\}$$
$$= \Big\langle \Big(\int_{(0,1)^n} A^{-1}\Big)^{-1}\xi,\xi\Big\rangle.$$

This inequality has an interpretation in terms of quadratic forms since the left-hand side is a quadratic form (quadratic forms are closed by Γ-convergence). Taking the inverse on both sides and reversing the inequality yields the harmonic bound for $\xi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

This way of deriving the trivial bounds is more straightforward and holds for any dimension. However it does not immediately suggest the form of optimal configurations, as opposed to the present 'pointwise' approach.

8.5.2 Interpretation in terms of quasiconvexification

Following the approach initiated in [116], we may look for the harmonic bound as the minimum of a functional depending on the geometry of the different phases of the material in $(0,1)^n$.

Let χ_{E_1} and χ_{E_2} be the characteristic functions associated to the two isotropic phases $A_1 = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}$ and $A_2 = \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix}$. And let A_3 be the anisotropic phase. For $\xi \in \mathcal{M}_n(\mathbb{R})$, the objective is to minimize the functional

$$J(\chi_1,\chi_2) = \int_{(0,1)^n} \big\langle (\chi_1 A_1 + \chi_2 A_2 + (1-\chi_1-\chi_2)A_3)\,(\xi + \nabla u_{\chi_1,\chi_2}), \xi + \nabla u_{\chi_1,\chi_2}\big\rangle dy$$

on the space of admissible characteristic functions

$$\Xi = \Big\{ (\chi_1,\chi_2) \in L^\infty((0,1)^n;\{0,1\}^2) : \chi_1\chi_2 = 0, \int_{(0,1)^n} \big(\chi_1 + \frac{1}{2}(1-\chi_1-\chi_2)\big)dy = \theta,$$
$$\int_{(0,1)^n} \big(\chi_2 + \frac{1}{2}(1-\chi_1-\chi_2)\big)dy = 1-\theta \Big\},$$

$\theta \in (0,1)$ being fixed. Here u_{χ_1,χ_2} realizes the minimum of $F(\xi,\nabla(\cdot))$ on $H_0^1((0,1)^n)$, with $A(y) = \chi_1 A_1 + \chi_2 A_2 + (1-\chi_1-\chi_2)A_3$.

We then have

$$\inf_{\Xi} J(\chi_1,\chi_2) = \inf_{\Xi}\inf_{u\in H_\xi} \int_{(0,1)^n} \big\langle (\chi_1 A_1 + \chi_2 A_2 + (1-\chi_1-\chi_2)A_3)\,\nabla u, \nabla u\big\rangle dy$$
$$= \inf_{u\in H_\xi} \int_{(0,1)^n} \inf_{\chi_1,\chi_2=0,1;\chi_1\chi_2=0} \Big[\big\langle (\chi_1 A_1 + \chi_2 A_2$$
$$+(1-\chi_1-\chi_2)A_3)\nabla u, \nabla u\big\rangle + l_1(\chi_1 + \frac{1}{2}(1-\chi_1-\chi_2))$$
$$+l_2(\chi_2 + \frac{1}{2}(1-\chi_1-\chi_2))\Big]dy,$$

where $H_\xi = \{u \in H^1((0,1)^n,\mathbb{R}^n) : u(x) = \xi\cdot x \text{ on } \partial(0,1)^n\}$ and l_1, l_2 are Lagrange multipliers to impose the volume averages of χ_1 and χ_2.

Therefore, one can interpret the computation of the optimal bounds on the trace of the homogenized matrix as the computation of an infimum of a vector-valued integral functional. In order to apply the direct method of the calculus of variations, one may first compute the quasiconvex envelope of the integrand. The computation of a quasiconvex envelope is rarely explicit and this is the main reason why it may seems surprising at a first glance that we achieve this difficult task with such very simple arguments.

Actually, the bounds we have obtained being the trivial bounds, the quasiconvex envelope coincides with the convex envelope and

$$\inf_{\Xi} J(\chi_1,\chi_2) = \inf_{v\in V_\xi}\int_{(0,1)^n}\inf_{\chi_1,\chi_2=0,1;\chi_1\chi_2=0}\Big[\big\langle(\chi_1A_1+\chi_2A_2+(1-\chi_1-\chi_2)A_3)v,v\big\rangle$$
$$+l_1(\chi_1+\tfrac{1}{2}(1-\chi_1-\chi_2))+l_2(\chi_2+\tfrac{1}{2}(1-\chi_1-\chi_2))\Big]dy,$$

where $V_\xi = \{v \in L^2_\#((0,1)^{n\times n} : \int_{(0,1)^n} v = \xi\}$. There exists a family of matrix fields for which this infimum is attained and which saturate the harmonic bound. The associated minimizers $v \in V_\xi$ turn out to belong to H_ξ. The matrices are represented on Figure 8.5 with varying θ_1 and θ_2 such that $\theta_1+\theta_2 = 2\theta$.

Turning now to the arithmetic bounds, there also exists a family of matrix fields that saturate this inequality. These matrices are represented on Figure 8.4, with varying θ_1 and θ_2 such that $\theta_1+\theta_2 = 2\theta$. In particular, they are divergence free.

8.5.3 Extensions to any number of conductivities and dimensions

The interpretation in terms of quasiconvexification and the generality of the Voigt and Reuss bounds allow us to state a general version of Theorem 51.

Theorem 53 *Let $\{\alpha_1,\dots,\alpha_N\}$ be N ordered conductivities. We denote by $\mathcal{A} = \{A_{\beta_1,\dots,\beta_n}\}$ the set of diagonal matrices of order n such that $A_{\beta_1,\dots,\beta_n}(i,i) = \alpha_{\beta_i}$. Let $A : (0,1)^n \to \mathbb{R}^n$ be such that, up to a rotation and for almost all $x \in (0,1)^n$, $A(x) \in \mathcal{A}$, we set $A = \sum_{\beta_1,\dots,\beta_n=1,N}\chi_{\beta_1,\dots,\beta_n}R(x)^TA_{\beta_1,\dots,\beta_n}R(x)$, where $\chi_{\beta_1,\dots,\beta_n}$ are characteristic functions and $R(x)$ are rotations. We denote by $\theta_{\beta_1,\dots,\beta_n} = \int_{(0,1)^n}\chi_{\beta_1,\dots,\beta_n}$. Let now fix to $\theta_{\alpha_i} \in (0,1)$ the proportions of the conductivities α_i. We have $\sum_{i=1,N}\theta_{\alpha_i} = 1$. With the following relationships between the phases,*

$$\begin{cases}\theta_{1,\dots,1}+\displaystyle\sum_{\beta_n\neq 1}\frac{1}{n}\theta_{1,\dots,1,\beta_n}+\cdots+\sum_{\beta_2,\dots,\beta_n\neq 1}\frac{n-1}{n}\theta_{1,\beta_2,\dots,\beta_n}=\theta_{\alpha_1}\\ \dots\\ \theta_{N,\dots,N}+\displaystyle\sum_{\beta_n\neq N}\frac{1}{n}\theta_{N,\dots,N,\beta_n}+\cdots+\sum_{\beta_2,\dots,\beta_n\neq N}\frac{n-1}{n}\theta_{N,\beta_2,\dots,\beta_n}=\theta_{\alpha_N}\end{cases},$$

the G-closure is exactly given by the Voigt and Reuss bounds, that is the convex intersection of the following curves in the plane of the eigenvalues $\lambda_1,\dots,\lambda_n$:

$$\begin{cases}\alpha_1\le\lambda_i\le\alpha_N\\ \displaystyle\sum_{i=1,n}\lambda_i\le n\sum_{j=1,N}\theta_{\alpha_j}\alpha_j\\ \displaystyle\sum_{i=1,n}\frac{1}{\lambda_i}\le n\sum_{j=1,N}\theta_{\alpha_j}\frac{1}{\alpha_j}.\end{cases}$$

An example of optimal configuration is given for the arithmetic bounds in Figure 8.6. The rotation of $\pi/2$ of this configuration gives an optimal configuration for the harmonic bound.

41	42	43	44
31	32	33	34
21	22	23	24
11	12	13	14

Fig. 8.6. Optimal geometry for the arithmetic bound in 2 dimensions for 4 conductivities

It should be emphasized that these rank-one laminate configurations correspond to the general class introduced in [97, 170] for multiphase materials.

The exact bounds are plotted on Figure 8.7 for a three layer-type composite $\alpha = 1$, $\beta = 2$, $\gamma = 3$ with the proportions $\theta_\alpha = \theta_\beta = \theta_\gamma = 1/3$.

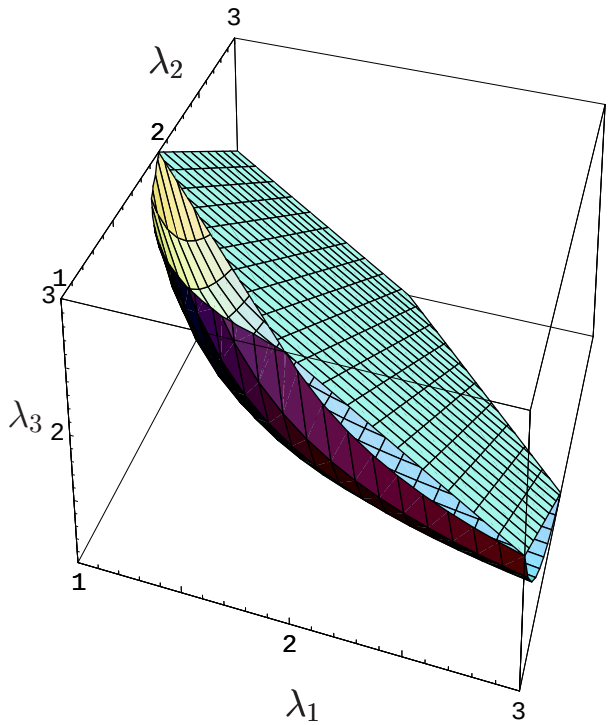

Fig. 8.7. Exact bounds in 3 dimensions (note that $1 \leq \lambda_i \leq 3$)

9

Variational description and homogenization of bulk energies for bounded and unbounded spin systems

Summary. We study the asymptotic behaviour of a general class of discrete energies defined on functions $u : \alpha \in \epsilon\mathbb{Z}^N \cap \Omega \mapsto u(\alpha) \in \mathbb{R}^m$ of the form $E_\epsilon(u) = \sum_{\alpha,\beta\in\epsilon\mathbb{Z}^N\cap\Omega} \epsilon^N g_\epsilon(\alpha, \beta, u(\alpha), u(\beta))$, as the mesh size ϵ goes to 0. We prove that under general assumptions, that cover the case of bounded and unbounded spin system in the thermodynamic limit, the variational limit of E_ϵ has the form $E(u) = \int_\Omega g(x, u(x))dx$. The case of homogenization and that of non-pairwise interacting systems (e.g. multiple-exchange spin-systems) are also discussed.

9.1 Introduction

Both in the applied mathematical and physical literature, there is much interest in the origin of pattern formation at the mesoscopic scale. It is believed that the competition between short range and long range interactions is responsible for many of the observed patterns in physical systems. On one side continuous descriptions provide a successful interpretation of pattern formation in terms of non attainment of infima (austenite/martensite phase transformations, micromagnetics in thin films, two wells problems etc., see [15, 114] and [55, 62, 115, 146] for reviews). At the other side of the spectrum, statistical mechanics aims at predicting such patterns starting from discrete systems of particles in interaction. In general, the problem can be stated as follows. Given an integer M, let Λ_M denote $\mathbb{Z}^N \cap [0, M]^N$, the intersection of the lattice $\mathbb{Z}^N$ with a cube of side M. A configuration of a discrete system on $\mathbb{Z}^N$ is a function $\underline{u} : \mathbb{Z}^N \to \mathbb{R}^m, x \mapsto \underline{u}(x)$. An energy for discrete systems can be written as

$$H_M(\underline{u}) = \sum_{x\neq y\in\Lambda_M} g(x, y, \underline{u}(x), \underline{u}(y)).$$

According to the range of $\underline{u}$ and the choice of g (regarding the typical distance of the interactions, e.g.), we may recover many different models for spin systems, crystals, foams, polymers ... To study the macroscopic behaviour of such systems, one can characterize the thermodynamic limits of their free energies for general values of the temperature. In general, not much is known on the fine properties of the Gibbs states (such as pattern formation). At small temperature however, a good insight may consist in characterizing the ground states of the system at the bulk limit, namely:

$$\lim_{M\to\infty} \frac{1}{M^N} \inf\{H_M(\underline{u}), \text{boundary conditions}\}.$$

There is actually a complete equivalence between letting the domain invade $\mathbb{R}^N$ (in the sense of Van Hove, e.g.) and taking the bulk limit on the one hand (as it is usually done in statistical mechanics [166]), and considering a fixed domain and letting the lattice spacing go to zero on the other hand. Information on ground states at the bulk limit can then be recast in terms of information on the recovery sequences for the Γ-convergence of discrete systems. This point of view amounts to writing $\frac{1}{M^N} H_N(\underline{u})$ as

$$E_\epsilon(u) = \sum_{\alpha_1,\alpha_2\in\epsilon\mathbb{Z}^N\cap[0,1]^N} \epsilon^N g_\epsilon(\alpha_1,\alpha_2,u(\alpha_1),u(\alpha_2)), \tag{9.1}$$

where $\epsilon = \frac{1}{M}$, $u(\alpha) = \underline{u}(\frac{\alpha}{\epsilon})$ and $g_\epsilon(\alpha_1,\alpha_2,u(\alpha_1),u(\alpha_2)) = g(\frac{\alpha_1}{\epsilon},\frac{\alpha_2}{\epsilon},\underline{u}(\frac{\alpha_1}{\epsilon}),\underline{u}(\frac{\alpha_2}{\epsilon}))$. The problem now translates in the computation of the minimum of $E_\epsilon(u)$ as ϵ goes to zero.

Within this framework, Ising type energies have been studied in [3, 5], respectively for $u \in \{-1,+1\}$ and $u \in \{v \in \mathbb{R}^m, |v| = 1\}$. In both cases, only purely ferromagnetic or purely antiferromagnetic interactions have been considered. Thus the existence of a bulk limit is trivial and fine properties of minimizers appear at a successive scale (interface or vortex-type phase transitions). Having in mind other models for spin systems, more complex energies are to be considered. Very recently, Giuliani, Lebowitz and Lieb [98] have addressed the characterization of ground states of a spin system mixing both short range ferromagnetic and long range antiferromagnetic interactions. In such cases, the existence and the form of the bulk limit are not clear a priori. In particular, the limit cannot always be written as the integral of a local function (see [35]). The aim of the present paper is to find a wide class of energies of type (9.1) for which the bulk limit can be written as

$$E(u) = \int_{(0,1)^N} g(x,u(x))dx, \tag{9.2}$$

in terms of Γ-convergence.

To perform our analysis, it is useful to make a change of variables and rewrite the energies (9.1) as

$$E_\epsilon(u) = \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha,\alpha+\epsilon\xi\in(0,1)^N} \epsilon^N f_\epsilon^\xi(\alpha,u(\alpha),u(\alpha+\epsilon\xi)). \tag{9.3}$$

We distinguish whether the range of u is bounded (or even a in a finite set) or not. The first case models classical spin systems, whereas the second one is usually referred to as the unbounded spin system case, which has been studied by Lebowitz and Presutti in [129] from the statistical mechanics point of view. We make two types of hypotheses on f_ϵ^ξ, namely growth conditions that ensure the limit functional to be finite on L^p (for $1 < p < \infty$) or on L^∞, and a decay assumption on the range of the interactions that ensures the finiteness of the discrete energy a priori and the locality of the limit functional. Under this set of hypotheses we are able to prove a compactness theorem asserting that, up to a subsequence, E_ϵ Γ-converges to a functional of type (9.2). We also give a homogenization result when $f_\epsilon^\xi(\cdot,u,v) = f^\xi(\frac{\cdot}{\epsilon},u,v)$ and $f^\xi(\cdot,u,v)$ is a periodic function.

All the results proved for energies of the form (9.3) hold true in the more general setting of non-pairwise-interactions energies. This case cover the model of Heisenberg spin systems with multiple-exchange spin interactions. For this type of models we provide an example which gives us the opportunity to show how the limit energy-density may depend on the geometric frustration of the spin system on different lattices.

The article is organized as follows: in Section 9.2, we introduce our main notation and recall some integral representation results. In Section 9.3, we state and prove our main results for pairwise interaction energies. We then derive the convergence of infimum problems in Section 9.4 and homogenization results in Section 9.5. Section 9.6 is devoted to the study of the example of Giuliani, Lebowitz and Lieb, whereas in the last Section we extend our results to non-pairwise interaction energies.

9.2 Notation and preliminary results

In what follows $\Omega \subset \mathbb{R}^N$ denotes a bounded open set with $\mathcal{L}^N(\partial\Omega) = 0$. Let $\mathcal{A}(\Omega)$ be the class of all open bounded subsets of Ω and by $\mathcal{A}'(\Omega)$ the class of all open bounded subsets $U \subset \Omega$ such that $\mathcal{L}^N(\partial U) = 0$. For all $B \subset \mathbb{R}^N$ we define $\mathbb{Z}_\epsilon(B) = \epsilon\mathbb{Z}^N \cap B$ and, for any $\xi \in \mathbb{Z}^N$,

$R_\epsilon^\xi(B) = \{\alpha \in \epsilon\mathbb{Z}^N : \ \alpha, \ \alpha + \epsilon\xi \in B\}$. In the remainder of the article, α implicitly depends on ϵ and belongs to $\epsilon\mathbb{Z}^N$, whereas $\beta, \xi \in \mathbb{Z}^N$.

We will make use of the following integral representation theorem on Lebesgue spaces by Buttazzo and Dal Maso [46] for functionals defined on pairs function-sets:

Theorem 9.1. (Integral representation) *Let $p \in [1, \infty[$, and let $F : L^p(\Omega, \mathbb{R}^m) \times \mathcal{A}(\Omega) \to [0, +\infty]$ be a functional satisfying:*

(i) F is local on $\mathcal{A}(\Omega)$; i.e. $\forall u, v \in L^p(\Omega, \mathbb{R}^m)$ and $\forall B \in \mathcal{A}(\Omega)$, $u = v$ a.e. on $B \Rightarrow F(u, B) = F(v, B)$;

(ii) F is additive on $\mathcal{A}(\Omega)$; i.e. $\forall u \in L^p(\Omega, \mathbb{R}^m)$, and $\forall B_1, B_2 \in \mathcal{A}(\Omega) : \ B_1 \cap B_2 = \emptyset \Rightarrow F(u, B_1 \cup B_2) = F(u, B_1) + F(u, B_2)$;

(iii) there exists $u_0 \in L^p(\Omega, \mathbb{R}^m)$ such that $F(u_0, \cdot)$ is the restriction of a Borel measure on $\mathcal{A}(\Omega)$ which is absolutely continuous with respect to (w.r.t.) the Lebesgue measure,

(iv) the functional $F(\cdot, \Omega)$ is lower-semicontinuous (l.s.c.) with respect to the strong convergence of $L^p(\Omega, \mathbb{R}^m)$,

then there exists a unique positive normal integrand f such that

$$F(u, B) = \int_B f(x, u(x))dx,$$

for all $u \in L^p(\Omega, \mathbb{R}^m)$ and $B \in \mathcal{A}(\Omega)$. In addition,

(v) if $F(\cdot, \Omega)$ is l.s.c. with respect to the weak convergence of $L^p(\Omega, \mathbb{R}^m)$ then f is a convex integrand,

(vi) if $F(\cdot, \Omega)$ is continuous with respect to the strong convergence of $L^p(\Omega, \mathbb{R}^m)$, then f is a Carathéodory function satisfying a growth condition of order p; i.e. there exist two constant $c, C \in \mathbb{R}_*^+$ and two functions $d, D \in L^1(\Omega, \mathbb{R}^m)$ such that

$$c|t|^p - d(x) \leq f(x, z) \leq C|t|^p + D(x) \qquad \text{for all } z \in \mathbb{R}^m \text{ and } x \in \Omega.$$

9.3 Compactness and integral representation results for spin systems

In this section we define the class of energies that we mainly consider in the prsent paper, i.e. pairwise-interaction energies. For this class of energies we prove a compactness and integral representation result asserting that, any sequence belonging to this family has a Γ-convergent subsequence whose Γ-limit is an integral functional.

9.3.1 Pairwise-interaction energies

Given $\Omega \subset \mathbb{R}^N$ and $\epsilon > 0$, the energy of a pairwise-interacting spin system with spin variable $u \in \mathbb{R}^m$ and energy-density $g_\epsilon : (\epsilon\mathbb{Z}^N \cap \Omega)^2 \times \mathbb{R}^{2m} \to \mathbb{R}$ on the lattice $\epsilon\mathbb{Z}^N \cap \Omega$ is given by the functional $E_\epsilon : \mathbb{R}^m \to (-\infty, +\infty)$:

$$E_\epsilon(u) = \sum_{\alpha_1, \alpha_1 \in \mathbb{Z}_\epsilon(\Omega)} \epsilon^N g_\epsilon(\alpha_1, \alpha_1, u(\alpha_1), u(\alpha_1)).$$

As it is well known, we explicitly observe that there is no loss of generality in considering the interactions symmetric. This symmetry condition is expressed by the formula $g_\epsilon(\alpha_1, \alpha_1, u, v) = g_\epsilon(\alpha_1, \alpha_1, v, u)$. Note that, otherwise, one could deal with

$$\tilde{g}_\epsilon(\alpha_1, \alpha_1, u, v) = \frac{1}{2}(g_\epsilon(\alpha_1, \alpha_1, v, u) + g_\epsilon(\alpha_1, \alpha_1, u, v)).$$

In the following we find it useful to rewrite the energy by a change of variable. Given $\xi \in \mathbb{Z}^N$ we define:

$$g_\epsilon(\alpha, \alpha + \epsilon\xi, u, v) = f_\epsilon^\xi(\alpha, u, v)$$

and then we have

$$E_\epsilon(u) = \sum_{\xi \in \mathbb{Z}^N} \sum_{\alpha \in R_\epsilon^\xi(\Omega)} \epsilon^N f_\epsilon^\xi(\alpha, u(\alpha), u(\alpha + \epsilon\xi)).$$

Note that, in the present variables, the symmetry condition reads $f_\epsilon^\xi(\alpha, u, v) = f_\epsilon^{-\xi}(\alpha + \epsilon\xi, v, u)$. Set, for any $k \in \mathbb{N}$,

$$C_\epsilon^k(\Omega) = \{u : \mathbb{R}^N \to \mathbb{R}^k \,:\, u \text{ constant on } \alpha + [0, \epsilon)^N \text{ for any } \alpha \in \mathbb{Z}_\epsilon(\Omega)\}.$$

we may identify any function $u : \mathbb{Z}_\epsilon(\Omega) \to \mathbb{R}^k$ as a piecewise-constant function belonging to $C_\epsilon^k(\Omega)$ and then define the family of energies E_ϵ on this subset of $L^p(\Omega, \mathbb{R}^m)$. One may extend such energies on the whole $L^p(\Omega, \mathbb{R}^m)$ and define a family of functionals $F_\epsilon : L^p(\Omega, \mathbb{R}^m) \to (-\infty, +\infty]$ by

$$F_\epsilon(u) = \begin{cases} \displaystyle\sum_{\xi \in \mathbb{Z}^N} \sum_{\alpha \in R_\epsilon^\xi(\Omega)} \epsilon^N f_\epsilon^\xi(\alpha, u(\alpha), u(\alpha + \epsilon\xi)) & \text{if } u \in C_\epsilon^m(\Omega) \\ \\ +\infty & \text{otherwise,} \end{cases} \tag{9.4}$$

where $f_\epsilon^\xi : \epsilon\mathbb{Z}^N \cap \Omega \times \mathbb{R}^{2m} \to \mathbb{R}$ are given functions.

The sets of hypotheses we deal with depend on whether we consider the case $1 < p < \infty$ or $p = \infty$.

Remark 9.2. Note that pairwise-interaction energies do not provide with the most general frame we can deal with. Nevertheless, we think the simpler presentation adopted here will help clarify the proofs. The very same arguments, combined with a more complex structure of the discrete system itself, also allow us to treat non pairwise interaction energies, as will be discussed in Section 9.7.

9.3.2 Case $1 < p < \infty$

For $1 < p < \infty$ let us make the following set of hypotheses on the family of functions f_ϵ^ξ:

(H1) **Coercivity hypothesis.** For all ξ, ϵ and α, there exist $c_{\epsilon,\alpha}^\xi \geq 0$ and $d_\epsilon^\xi \in C_\epsilon^1(\Omega)$, $d_\epsilon^\xi(\alpha) \geq 0$ such that

$$f_\epsilon^\xi(\alpha, u, v) \geq c_{\epsilon,\alpha}^\xi(|u|^p + |v|^p) - d_\epsilon^\xi(\alpha) \quad \text{for all } (u, v) \in \mathbb{R}^{2m},$$

$$\lim_{R \to \infty} \liminf_{\epsilon \to 0} \inf_{\alpha \in \epsilon\mathbb{Z}^N \cap \Omega} \sum_{|\xi| \leq R} c_{\epsilon,\alpha}^\xi \geq c > 0$$

and the function $d_\epsilon \in C_\epsilon^1(\Omega)$ defined by $d_\epsilon(\alpha) = \displaystyle\sum_\xi d_\epsilon^\xi(\alpha)$ weakly converges to d in $L^1(\Omega)$.

(H2) **Growth hypothesis.** For all ξ, ϵ and α, there exist $C_{\epsilon,\alpha}^\xi \geq 0$ and $D_\epsilon^\xi \in C_\epsilon^1(\Omega)$, $D_\epsilon^\xi(\alpha) \geq 0$ such that

$$f_\epsilon^\xi(\alpha, u, v) \leq C_{\epsilon,\alpha}^\xi(|u|^p + |v|^p) + D_\epsilon^\xi(\alpha) \text{ for all } (u, v) \in \mathbb{R}^{2m},$$

$$\limsup_{\epsilon \to 0} \sup_\alpha \sum_{\xi \in \mathbb{Z}^N} C_{\epsilon,\alpha}^\xi \leq C < \infty$$

and the function $D_\epsilon \in C_\epsilon^1(\Omega)$ defined by $D_\epsilon(\alpha) = \displaystyle\sum_\xi D_\epsilon^\xi(\alpha)$ weakly converges to D in $L^1(\Omega)$.

(H3) **Decay hypothesis.** For all $\delta > 0$, there exists $M_\delta > 0$ such that

$$\limsup_{\epsilon\to 0} \sup_{\alpha} \sum_{|\xi|\geq M_\delta} C^{\xi}_{\epsilon,\alpha} \leq \delta.$$

As will be made precise hereafter, hypotheses (H1)-(H2) ensure that any Γ-limit of a subsequence of E_ϵ is defined on $L^p(\Omega)$. Hypothesis (H3) provides a control on the long-range interactions which implies the locality of the limit functional.

Theorem 9.3. *Let F_ϵ be as in (9.3), and $\{f^\xi_\epsilon\}$ satisfy hypotheses (H1), (H2) and (H3). Then, for every sequence converging to zero, there exists a subsequence (ϵ_j) and a Carathéodory function $f : \Omega \times \mathbb{R}^m \to \mathbb{R}$ convex in the second variable and satisfying the following growth condition of order p*

$$c|y|^p - d(x) \leq f(x,y) \leq C|y|^p + D(x) \qquad \text{for all } y \in \mathbb{R}^m \text{ and } x \in \Omega, \tag{9.5}$$

such that $(F_{\epsilon_j}(\cdot))$ Γ-converges with respect to the weak convergence of $L^p(\Omega,\mathbb{R}^m)$ ($\Gamma(w-L^p)$-converges) to the functional defined by

$$\int_\Omega f(x,u(x))dx, \tag{9.6}$$

for all $u \in L^p(\Omega,\mathbb{R}^m)$.

9.3.3 Case $p = \infty$

Let $p = \infty$, and $K \subset \mathbb{R}^m$ be a bounded set. In this case, let us consider the following set of hypotheses on f^ξ_ϵ.

(H4) For all ξ, ϵ and α, $f^\xi_\epsilon(\alpha,u,v) = +\infty$ if $(u,v) \notin K^2$,
(H5) For all ξ, ϵ and α, there exists $C^\xi_{\epsilon,\alpha} \geq 0$ such that

$$|f^\xi_\epsilon(\alpha,u,v)| \leq C^\xi_{\epsilon,\alpha} \quad \text{for all } (u,v) \in K^2,$$

$$\limsup_{\epsilon\to 0} \sup_{\alpha} \sum_{\xi\in\mathbb{Z}^N} C^\xi_{\epsilon,\alpha} < \infty,$$

(H6) for all $\delta > 0$, there exists $M_\delta > 0$ such that

$$\limsup_{\epsilon\to 0^+} \sup_{\alpha} \sum_{|\xi|\geq M_\delta} C^\xi_{\epsilon,\alpha} \leq \delta.$$

Remark 9.4. Hypotheses (H5) and (H6) do not imply $\limsup_\epsilon \sum_\xi \sup_\alpha C^\xi_{\epsilon,\alpha} < \infty$. To check this, let us take for instance $C^{\alpha/\epsilon}_{\epsilon,\alpha} = \frac{1}{\left|\frac{\alpha}{\epsilon}\right|+1}$ and $C^\xi_{\epsilon,\alpha} = 0$ for $\xi \neq \frac{\alpha}{\epsilon}$.

Theorem 9.5. *Let F_ϵ be as in (9.3), and $\{f^\xi_\epsilon\}$ satisfy hypotheses (H4), (H5) and (H6). Then, for every sequence converging to zero, there exists a subsequence (ϵ_j) and a Carathéodory function $f : \Omega \times \overline{K} \to \mathbb{R}$ convex in the second variable such that $(F_{\epsilon_j}(\cdot))$ Γ-converges with respect to the weak *-convergence of $L^\infty(\Omega,\mathbb{R})$ ($\Gamma(w*-L^\infty)$-converges) to the functional defined by*

$$\begin{cases} \displaystyle\int_\Omega f(x,u(x))dx & \text{if } u \in L^\infty(\Omega,\overline{K}) \\ +\infty & \text{otherwise,} \end{cases} \tag{9.7}$$

where $\overline{K}$ is the convex hull of K in $\mathbb{R}^m$.

We now briefly discuss the optimality of hypothesis (H5) on two simple examples.

Example 9.6. In this example we show that if we weaken assumption (H5) by only assuming that

$$\limsup_{\epsilon} |\sum_{\xi} f_\epsilon^\xi(\alpha, u, v)| < \infty \quad \forall \alpha \in R_\epsilon^\xi(\Omega),\ (u, v) \in K^2,$$

then the Γ-limit may go to $-\infty$ at some point. Let us consider a one-dimensional discrete energy of the form (9.3) with energy density given by:

$$f_\epsilon^\xi(\alpha, u, v) = \begin{cases} \frac{(-1)^{|\xi|+1}}{|\xi|+1} uv & \text{if } u, v \in \{-1, 1\}, \\ +\infty & \text{if } u, v \notin \{-1, 1\}. \end{cases}$$

For $\Omega = (0, 1)$ and $\epsilon = \frac{1}{n}$, the energy of the system for $u : \frac{1}{n}\mathbb{Z} \cap (0, 1) \to \{-1, 1\}$ can thus be written as

$$F_n(u) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k+1} \sum_{i=0}^{n-k} \frac{1}{n} u(\frac{i}{n}) u(\frac{i+k}{n}).$$

Set $u_n(\frac{i}{n}) = (-1)^i$, we have that $u_n \rightharpoonup^* 0$ in $L^\infty((0, 1))$, and

$$\lim_n F_n(u_n) = -\sum_{k=1}^{\infty} \frac{1}{k+1} = -\infty.$$

Hence $\Gamma\text{-}\lim_n F_n(0) = -\infty$. However, $\Gamma\text{-}\lim_n F_n$ is not identically $-\infty$. Indeed it can be easily proved that

$$\Gamma - \lim_n F_n(1) = \lim_n F_n(1) = \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k+1}.$$

In the following two examples we consider other two cases in which (H5) is not satisfied. In the first case, we may still have compactness if some additional condition holds, whereas in the second case the energy is not bounded.

Example 9.7. In this example we weaken assumption (H5) by assuming that $C_\epsilon^{e_1}$ goes to infinity as $\epsilon \to 0$. Let us consider a one-dimensional nearest-neighbors spin system on $(0, 1)$, with a spin field taking values in $K = \{-1, 0, 1\}$. For $u : \epsilon\mathbb{Z} \cap (0, 1) \to K$, let the energy of the system be of the form

$$F_\epsilon(u) = \sum_{\alpha \in \epsilon\mathbb{Z} \cap (0,1)} \epsilon f_\epsilon(u(\alpha), u(\alpha + \epsilon)), \tag{9.8}$$

where the pair potential $f_\epsilon(u, v) : K^2 \to \mathbb{R}^+$ is such that $f_\epsilon(u, v) = f_\epsilon(v, u)$ and is given by

$$f_\epsilon(\alpha, u, v) = \begin{cases} \dfrac{1}{\epsilon} & \text{if } (u, v) = (0, 1) \\ 1 & \text{otherwise.} \end{cases} \tag{9.9}$$

This energy does not satisfy (H5) since $f_\epsilon(0, 1) \to \infty$. However, any $u \in L^\infty((0, 1), [-1, 1])$ can be approximated in the w^*-topology of L^∞ by a sequence $u_\epsilon : \epsilon\mathbb{Z} \cap (0, 1) \to \{-1, 0, 1\}$ such that $(u_\epsilon(\alpha), u_\epsilon(\alpha + \epsilon)) \neq (0, 1)$ for all $\alpha \in \epsilon\mathbb{Z} \cap (0, 1)$. This suggests us that, if in the definition of f_ϵ we replace $\frac{1}{\epsilon^\gamma}$ by any $C \geq \max\{f_\epsilon(u, v), (u, v) \neq (0, 1)\}$, the modified energy satisfies assumption $(H5)$ and has the same Γ-limit of the original one.

Let us consider the case when in (9.8) the energy density in (9.9) is replaced by

$$f_\epsilon(u, v) = \begin{cases} \frac{1}{\epsilon} & \text{if } (u, v) \in \{(0, 1), (-1, 1)\} \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Let us now consider the piecewise constant function $u_k(x) = -1$ for $x < 1/k$, and $u_k(x) = 1$ for $x \geq 1/k$. For all $u_\epsilon \rightharpoonup^* u_k$, we have $F_\epsilon(u_\epsilon) \geq 1 + \frac{1}{2} + O(\epsilon) = \frac{3}{2} + O(\epsilon)$. This can be easily seen by minimizing pointwise the energy and noticing that we need at least one jump from 0 to 1 or from -1 to 1 to approximate u_k. Thus, if the $\Gamma - \lim_\epsilon F_\epsilon =: F$ exists, it satisfies $F(u_k) \geq \frac{3}{2}$. We also have that $F(-1) = F(1) = \frac{1}{2}$. Let us suppose now that F admits an integral representation of the type $F(v) = \int_0^1 f(x, v(x))dx$. As $f \geq 0$, $F(u_k) = \int_0^{1/k} f(x,-1) + \int_{1/k}^1 f(x,1) \leq \int_0^1 f(x,-1) + \int_0^1 f(x,1) = F(-1)+F(1) = 1$, which contradicts $F(u_k) \geq \frac{3}{2}$. Therefore the integral representation does not hold.

If $f_\epsilon(0,1) = f_\epsilon(-1,1) = \frac{1}{\epsilon^2}$, we cannot even find sequences of equi-bounded energies converging to u_k. Therefore the Γ-limit is $+\infty$.

9.3.4 Proof in L^p, $1 < p < \infty$

In the proofs, we implicitly take $m = 1$, since the arguments do not depend on the dimension (the problem is scalar as opposed to vectorial as in [4]). Let us recall that if (F_ϵ) is a family of functionals indexed by $\epsilon > 0$, we may define the *lower and upper Γ-limits* by

$$F'(u) = \Gamma\text{-}\liminf_{\epsilon\to 0^+} F_\epsilon(u) = \inf\{\liminf_{\epsilon\to 0^+} F_\epsilon(u_\epsilon) : \ u_\epsilon \to u\},$$
$$F''(u) = \Gamma\text{-}\limsup_{\epsilon\to 0^+} F_\epsilon(u) = \inf\{\limsup_{\epsilon\to 0^+} F_\epsilon(u_\epsilon) : \ u_\epsilon \to u\},$$

respectively. The functional F_ϵ is said to Γ-converge to F as $\epsilon \to 0^+$ if and only if $F'(u) = F''(u) = F(u)$. Note that the functions F' and F'' are lower semicontinuous (we refer to [36] and [56] for definition and properties of Γ-convergence).

The structure of the proof is the following. We first prove a growth condition on the Γ-liminf and limsup in order to apply a compactness result that ensures the existence of a Γ-limit. We then prove some properties of the Γ-limit to obtain an integral representation on $L^p(\Omega)$.

Proposition 9.8. *Let $A \in \mathcal{A}'(\Omega)$, and $\{f_\epsilon^\xi\}$ satisfy (H1). If $u \in L^p(\Omega)$ such that $F'(u, A) < \infty$ then*

$$F'(u, A) \geq c\left(\|u\|^p_{L^p(A)} - \|d\|_{L^1(A)}\right) \tag{9.10}$$

for some positive constant c independent of u and A.

Proof. -Let $\epsilon_n \to 0$, and let $u_n \rightharpoonup u$ in $L^p(\Omega)$ be such that $\liminf F_{\epsilon_n}(u_n, A) < \infty$. Let $A_\eta = \{x \in A | d(x, \partial A) > \eta\}$ for all $\eta > 0$. By the growth condition (H1), we have for $0 < \eta' < \eta$,

$$\begin{aligned}
&F_{\epsilon_n}(u_n, A)\\
&\geq \sum_{\alpha\in A_{\eta'}} \sum_{|\xi|\leq \eta/\epsilon_n} \epsilon_n^N c^\xi_{\epsilon_n,\alpha} |u_n(\alpha)|^p - \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R^\xi_{\epsilon_n}(A)} \epsilon_n^N d^\xi_{\epsilon_n}(\alpha)\\
&\geq \sum_{\alpha\in A_{\eta'}} \epsilon_n^N \inf_{\alpha\in\epsilon_n\mathbb{Z}^N\cap\Omega} \left(\sum_{|\xi|\leq\eta/\epsilon_n} c^\xi_{\epsilon_n,\alpha}\right) |u_n(\alpha)|^p - \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R^\xi_{\epsilon_n}(A)} \epsilon_n^N d^\xi_{\epsilon_n}(\alpha)\\
&\geq \sum_{\alpha\in A_{\eta'}} \epsilon_n^N \frac{c}{2} |u_n(\alpha)|^p - \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R^\xi_{\epsilon_n}(A)} \epsilon_n^N d^\xi_{\epsilon_n}(\alpha)\\
&\geq \frac{c}{2}\int_{A_\eta} |u_n(x)|^p dx - \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R^\xi_{\epsilon_n}(A)} \epsilon_n^N d^\xi_{\epsilon_n}(\alpha)
\end{aligned} \tag{9.11}$$

for ϵ_n small enough, and setting $u_n(y) = u_n(\alpha)$ for $y \in \alpha + [-\epsilon/2, \epsilon/2[^N$, and $u_n(y) = 0$ otherwise. Using now the lower semicontinuity of the norm for the weak convergence of L^p and (H1), we obtain

$$F'(u,A) \geq \frac{c}{2}\int_{A_\eta} |u(x)|^p dx - \int_A d(x)dx$$

Letting η go to zero, we obtain the thesis.

□

Proposition 9.9. *Let $A \in \mathcal{A}(\Omega)$, and $\left\{f_\epsilon^\xi\right\}$ satisfy (H2). If $u \in L^p(\Omega)$ then*

$$F''(u,A) \leq C\left(\|u\|^p_{L^p(A)} + \|D\|_{L^1(A)}\right) \tag{9.12}$$

for some positive constant C independent of u and A.

Proof. -Let $u \in C^0(\Omega)$ and let define u_n by $u_n(y) = u(\alpha)$ for all $y \in \alpha + [-\epsilon/2, \epsilon/2[^N$ if $(\alpha + [-\epsilon/2, \epsilon/2[^N \subset A$, and $u_n(y) = 0$ otherwise. We then have $u_n \to u$ in $L^p(A)$ and

$$\begin{aligned}
F_{\epsilon_n}(u_n, A) &\leq \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R^\xi_{\epsilon_n}(A)} \epsilon_n^N \left(C^\xi_{\epsilon_n,\alpha}(|u_n(\alpha)|^p + |u_n(\alpha+\epsilon_n\xi)|^p) + D^\xi_{\epsilon_n}(\alpha)\right) \\
&\leq 2 \sum_{\alpha\in\epsilon_n\mathbb{Z}^N\cap A} \sum_{\xi\in\mathbb{Z}^N} \epsilon_n^N C^\xi_{\epsilon_n,\alpha}|u_n(\alpha)|^p + \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R^\xi_{\epsilon_n}(A)} \epsilon_n^N D^\xi_{\epsilon_n}(\alpha) \\
&\leq C \sum_{\alpha\in\epsilon_n\mathbb{Z}^N\cap A} \epsilon_n^N |u_n(\alpha)|^p + \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R^\xi_{\epsilon_n}(A)} \epsilon_n^N D^\xi_{\epsilon_n}(\alpha) \\
&\leq C\int_A |u_n(x)|^p dx + \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R^\xi_{\epsilon_n}(A)} \epsilon_n^N D^\xi_{\epsilon_n}(\alpha),
\end{aligned}$$

due to the symmetry of the interactions. Letting ϵ_n go to zero, we obtain

$$F''(u,A) \leq C\left(\|u\|^p_{L^p(A)} + \|D\|_{L^1(A)}\right).$$

Using a density argument, we deduce the thesis for all $u \in L^p(\Omega)$.

□

To prove the existence of the Γ-limit for every regular open set $A \subset \Omega$ up to extraction, we use the following theorem with $X = L^p(A)$ and $\Psi(u) = c\left(\|u\|^p_{L^p(A)} - \|d\|_{L^1(A)}\right)$.

Theorem 9.10. *[56, Corollary 8.12] Assume that X is a Banach space with a separable dual. Let $\Psi : X \to \overline{\mathbb{R}}$ be a function satisfying*

$$\lim_{\|x\|_X\to\infty} \Psi(x) = \infty.$$

If $F_h \geq \Psi$ for every $h \in \mathbb{N}$, then there exists a subsequence of (F_h) which Γ-converges in the weak topology of X.

So far, we only know that there exists a Γ-limit for every regular open set $A \subset \Omega$ up to extraction. These Γ-limits may however be different. To use a diagonal extraction argument we first need to prove the inner-regularity of the Γ-limsup on $\mathcal{A}'(\Omega)$. Unfortunately, F'' could be non inner-regular, due to (H1). Nevertheless, setting $\tilde{f}_\epsilon^\xi(\alpha,u,v) = f_\epsilon^\xi(\alpha,u,v) + d_\epsilon^\xi(\alpha)$, we can define the associated energy $\tilde{F}_\epsilon$, its Γ-liminf $\tilde{F}'$ and its Γ-limsup $\tilde{F}''$. It can then be easily proved that $F''(u,A) = \tilde{F}''(u,A) - \int_A d$ and $F'(u,A) = \tilde{F}'(u,A) - \int_A d$. Proving a Γ-convergence result for F_ϵ or $\tilde{F}_\epsilon$ is thus equivalent. We will focus on $\tilde{F}_\epsilon$. The following proposition shows that $\tilde{F}''$ is inner-regular.

Proposition 9.11. *Let $\left\{f_\epsilon^\xi\right\}$ satisfy (H1)-(H3). If $u \in L^p(\Omega)$ and $A \in \mathcal{A}'(\Omega)$ then there holds*

$$\sup_{A'\subset\subset A} \tilde{F}''(u,A') = \tilde{F}''(u,A). \tag{9.13}$$

Proof. -Since $\tilde{f}^{\xi}_{\epsilon} \geq 0$, $\tilde{F}''(u,\cdot)$ is an increasing set function, and it suffices to prove that

$$\sup_{A' \subset\subset A} \tilde{F}''(u,A') \geq \tilde{F}''(u,A).$$

Given $\delta > 0$, there exists $A'' \subset\subset A$ such that

$$\|D\|_{L^1(A\setminus\overline{A''})} + \|u\|^p_{L^p(A\setminus\overline{A''})} \leq \delta.$$

Reasoning by approximation, we may find $v_\epsilon \in L^p(\Omega)$ such that v_ϵ weakly converges to u in $L^p(\Omega)$ and

$$\limsup_{\epsilon\to 0^+} \tilde{F}_\epsilon(v_\epsilon, A\setminus\overline{A''}) \leq C\left(\|D\|_{L^1(A\setminus\overline{A''})} + \|u\|^p_{L^p(A\setminus\overline{A''})}\right) \leq C\delta. \tag{9.14}$$

Let $A' \in \mathcal{A}(\Omega)$ be such that $A'' \subset\subset A' \subset\subset A$ and let $u_\epsilon \in L^p(\Omega)$ weakly converge to u in $L^p(\Omega)$, with

$$\limsup_{\epsilon\to 0^+} \tilde{F}_\epsilon(u_\epsilon, A') = \tilde{F}''(u,A').$$

Set

$$d := \mathrm{dist}(A'', A'^c)$$

and for any $M \in \mathbb{N}$ and $i \in \{1,\dots,M\}$ define

$$A_i = \{x \in A : \ \mathrm{dist}(x, A'') < i\frac{d}{M}\}.$$

Let φ_i be the indicative function of A_i. Then for any $i \in \{1,\dots,M\}$ consider the family of functions $w^i_\epsilon \in \mathcal{A}_\epsilon(\Omega)$ still weakly converging to u in $L^p(\Omega)$ defined by

$$w^i_\epsilon(\alpha) := \varphi_i(\alpha)u_\epsilon(\alpha) + (1-\varphi_i(\alpha))\, v_\epsilon(\alpha).$$

Let $i \in \{1,2,\dots,M-3\}$ be fixed. Given $\xi \in \mathbb{Z}^N$ and $\alpha \in R^\xi_\epsilon(A)$, then either $\alpha \in R^\xi_\epsilon(A_i)$, or $\alpha \in R^\xi_\epsilon(A\setminus\overline{A}_{i+1})$, or

$$]\alpha, \alpha+\epsilon\xi[\cap\left(\overline{A}_{i+1}\setminus A_i\right)\cap\overline{A'}^c \neq \emptyset.$$

Then, if we set

$$\left(\overline{A}_{i+1}\setminus A_i\right)^{\epsilon,\xi} := \{x = y + t\xi,\ |t| \leq \epsilon,\ y \in \overline{A}_{i+1}\setminus A_i\},$$
$$S^{\epsilon,\xi}_i := \left(\overline{A}_{i+1}\setminus A_i\right)^{\epsilon,\xi} \cap A,$$

we get

$$R^\xi_\epsilon(A) \subseteq R^\xi_\epsilon(A_i) \cup R^\xi_\epsilon(A\setminus\overline{A}_{i+1}) \cup R^\xi_\epsilon\left(S^{\epsilon,\xi}_i\right).$$

Let $M_\delta > 0$ be such that $\limsup_{\epsilon^+\to 0} \sum_{|\xi|>M_\delta} C^\xi_\epsilon < \delta$. Then, summing on $\xi \in \mathbb{Z}^N$, using (H2), (H3) and the previous decomposition we get

$$\begin{aligned}
\tilde{F}_\epsilon(w^i_\epsilon, A) \leq\ & \tilde{F}_\epsilon(u_\epsilon, A') + \tilde{F}_\epsilon(v_\epsilon, A\setminus\overline{A''}) \\
& + C\sum_{|\xi|\leq M_\delta} C^\xi_\epsilon \sum_{\alpha\in R^\xi_\epsilon\left(S^{\epsilon,\xi}_i\right)} \epsilon^N(|u_\epsilon(\alpha)|^p + |u_\epsilon(\alpha+\epsilon\xi)|^p \\
& \qquad + |v_\epsilon(\alpha)|^p + |v_\epsilon(\alpha+\epsilon\xi)|^p) \\
& + C\sum_{|\xi|>M_\delta} C^\xi_\epsilon \sum_{\alpha\in A} \epsilon^N(|u_\epsilon(\alpha)|^p + |u_\epsilon(\alpha+\epsilon\xi)|^p \\
& \qquad + |v_\epsilon(\alpha)|^p + |v_\epsilon(\alpha+\epsilon\xi)|^p) \\
\leq\ & \tilde{F}_\epsilon(u_\epsilon, A') + \tilde{F}_\epsilon(v_\epsilon, A\setminus\overline{A''}) \\
& + C\sum_{|\xi|\leq M_\delta} C^\xi_\epsilon \sum_{\alpha\in R^\xi_\epsilon\left(S^{\epsilon,\xi}_i\right)} \epsilon^N(|u_\epsilon(\alpha)|^p + |v_\epsilon(\alpha)|^p) \\
& + C\left(\sum_{|\xi|>M_\delta} C^\xi_\epsilon\right)(\|u_\epsilon\|^p_{L^p(A)} + \|v_\epsilon\|^p_{L^p(A)}),
\end{aligned}$$

by symmetry of the interactions. Note that, for ϵ small enough and $|\xi| \leq M_\delta$, we have that $R_\epsilon^\xi\left(S_i^{\epsilon,\xi}\right) \cap R_\epsilon^\xi\left(S_j^{\epsilon,\xi}\right) \neq \emptyset$ if and only if $|i-j| = 1$. Note also that $\cup_{i=1}^{M-3} R_\epsilon^\xi\left(S_i^{\epsilon,\xi}\right) \subseteq R_\epsilon^\xi\left(A' \setminus \overline{A''}\right)$. Thus, summing over $i \in \{1, 2, \cdots, M-3\}$, averaging and taking into account (9.14) and (H3), we get

$$\begin{aligned}\frac{1}{M-3}\sum_{i=1}^{M-3} \tilde{F}_\epsilon(w_\epsilon^i, A) \leq \tilde{F}_\epsilon(u_\epsilon, A') + C\delta \\ + \frac{1}{M-3} C \left(\sum_{|\xi| \leq M_\delta} C_\epsilon^\xi\right) (\|u_\epsilon\|_{L^p(\Omega)}^p + \|v_\epsilon\|_{L^p(\Omega)}^p) \\ + C\delta(\|u_\epsilon\|_{L^p(\Omega)}^p + \|v_\epsilon\|_{L^p(\Omega)}^p).\end{aligned} \tag{9.15}$$

For all M and ϵ we can choose $i(\epsilon) \in \{1, 2, \cdots, M-3\}$ such that

$$\tilde{F}_\epsilon(w_\epsilon^{i(\epsilon)}, A) \leq \frac{1}{M-3}\sum_{j=1}^{M-3} \tilde{F}_\epsilon(w_\epsilon^j, A).$$

Then, $w_\epsilon^{i(\epsilon)}$ still weakly converges to u in $L^p(\Omega)$. Therefore, letting ϵ go to zero, we obtain

$$\tilde{F}''(u, A) \leq \sup_{A' \subset\subset A} \tilde{F}''(u, A') + C\left(\frac{1}{M-3} + \delta\right).$$

Letting δ go to zero and M to infinity concludes the proof of the thesis.

□

By the compactness property of the Γ-convergence [38, Theorem 10.3] and by Proposition 9.11, for every sequence ϵ_j converging to zero, there exist a subsequence (ϵ_j) (not relabeled) and an increasing set function $\tilde{F} : L^p(\Omega) \times \mathcal{A}'(\Omega) \to \mathbb{R}^+, (u, A) \mapsto \tilde{F}(u, A)$, such that $F_{\epsilon_j}(\cdot, A)$ $\Gamma(w-L^p)$-converges to $\tilde{F}(\cdot, A)$ for all $A \in \mathcal{A}(\Omega)$.

It only remains to prove the integral representation formula. Let us introduce some further properties of the Γ-limit obtained.

Proposition 9.12. *Let $\left\{f_\epsilon^\xi\right\}$ satisfy (H1)-(H3). If $u \in L^p(\Omega)$ and $A, B \in \mathcal{A}'(\Omega)$ then there holds*

$$\tilde{F}''(u, A \cup B) \leq \tilde{F}''(u, A) + \tilde{F}''(u, B). \tag{9.16}$$

In addition, if $A \cap B = \emptyset$ then

$$\tilde{F}''(u, A \cup B) \geq \tilde{F}''(u, A) + \tilde{F}''(u, B). \tag{9.17}$$

Proof. -Using the same strategy as for Proposition 9.11, we may prove that for all $A', B' \in \mathcal{A}(\Omega)$ such that $A' \subset\subset A$ and $B' \subset\subset B$ we have

$$\tilde{F}''(u, A' \cup B') \leq \tilde{F}''(u, A) + \tilde{F}''(u, B). \tag{9.18}$$

Since for all $C \in \mathcal{A}(\Omega)$ such that $C \subset\subset A \cup B$ there exist $A', B' \in \mathcal{A}(\Omega)$ such that $A' \subset A, B' \subset B$ and $C \subset A' \cup B'$, Proposition 9.11 shows that (9.18) implies (9.16). In addition, $\tilde{F}_\epsilon(u, \cdot)$ is clearly superadditive, and so is $\tilde{F}''$ at the limit.

□

Proposition 9.13. *Let $\left\{f_\epsilon^\xi\right\}$ satisfy (H1)-(H3) and $\tilde{F}$ be a Γ-limit of $\tilde{F}_\epsilon$ for the weak convergence of $L^p(\Omega)$. Then $\tilde{F}$ is local: for all $A \in \mathcal{A}'(\Omega)$ and $u, v \in L^p(\Omega)$ such that $v = u$ almost everywhere on A, one has*

$$\tilde{F}(u, A) = \tilde{F}(v, A).$$

Proof. -Let u and $v \in L^p(\Omega)$ be such that $u_{|A} = v_{|A}$ almost everywhere on $A \in \mathcal{A}(\Omega)$. As $\tilde{F}(\cdot, A)$ is a Γ-limit of $\tilde{F}_\epsilon(\cdot, A)$, we have that for all $w_\epsilon \rightharpoonup v$ and $\tilde{w}_\epsilon \rightharpoonup u$ in $L^p(A)$, $\tilde{F}(u, A) \leq \liminf \tilde{F}_\epsilon(\tilde{w}_\epsilon, A)$ and $\tilde{F}(v, A) \leq \liminf \tilde{F}_\epsilon(w_\epsilon, A)$.

Let now u_ϵ and v_ϵ be recovery sequences for $\tilde{F}(u, A)$ and $\tilde{F}(v, A)$ in $L^p(A)$. As $u_{|A} = v_{|A}$ almost everywhere on A, one also has $v_\epsilon \rightharpoonup u_{|A}$ in $L^p(A)$ and $u_\epsilon \rightharpoonup v_{|A}$ in $L^p(A)$. Thus, $\tilde{F}(v, A) \leq \liminf \tilde{F}_\epsilon(u_\epsilon, A) = \tilde{F}(u, A)$ and $\tilde{F}(u, A) \leq \liminf \tilde{F}_\epsilon(v_\epsilon, A) = \tilde{F}(v, A)$, which shows the thesis.

□

The Γ-limit $\tilde{F}(u, \cdot)$ satisfies the De Giorgi-Letta criteria for all $u \in L^p(\Omega)$, therefore it is the restriction on $\mathcal{A}(\Omega)$ of a Borel measure, which, together with (9.12), implies hypothesis (iii) of Theorem 9.1. The other assumptions (i)-(ii)-(v) are also satisfied by $\tilde{F}$. As an application of Theorem 9.1, $\tilde{F}$ can be written as an integral:

$$\tilde{F}(u, U) = \int_U f(x, u(x))dx,$$

for all $u \in L^p(\Omega)$ and $U \in \mathcal{A}'(\Omega)$. In addition, the normal integrand f is convex in the second variable and non negative. Due to the growth condition (9.12), taking $u : \Omega \ni y \mapsto u(y) = \xi \in \mathbb{R}$, $U_\eta = B(x, \eta)$ and letting $\eta \to 0$, we obtain the growth condition (9.5) pointwise almost everywhere in Ω for the integrand f. Thus it satisfies a local Lipschitz condition and is continuous. The integrand is therefore a Carathéodory function.

This concludes the proof of Theorem 9.3 for $\tilde{F}_\epsilon$ and therefore also for F_ϵ.

9.3.5 Proof in L^∞

The proof of Theorem 9.5 is an easy adaptation of the proof of Theorem 9.3. We just have to slightly modify the growth conditions satisfied by the Γ-liminf and Γ-limsup. Then using the fact that the topologies of the $\Gamma(w - L^p)$-convergence for any $p \geq 1$ are equivalent on

$$V = L^\infty(\Omega, \overline{K}),$$

we may still apply Theorems 9.10 and 9.1.

Proposition 9.14. *Let $A \in \mathcal{A}'(\Omega)$ and $\left\{f_\epsilon^\xi\right\}$ satisfy (H4) and (H5). If $F'(u, A)$ is finite then $u \in L^\infty(A, \overline{K})$, and*

$$F'(u, A) \geq -c|A| \tag{9.19}$$

for some positive constant c independent of u and A.

Proof. -Let $\epsilon_n \to 0$ and $u_n \in L^\infty(\Omega)$ such that $\liminf F_{\epsilon_n}(u_n, A) < \infty$, and $u_n \rightharpoonup u$ in $L^\infty(A, \overline{K})$. By the growth condition H2, we have

$$\begin{aligned} F_{\epsilon_n}(u_n, A) &\geq \sum_{\xi \in \mathbb{Z}^N} \sum_{\alpha \in R_{\epsilon_n}^\xi(A)} -\epsilon_n^N c_{\epsilon_n, \alpha}^\xi \\ &\geq \sum_{\alpha \in A} -\epsilon_n^N \left(\sum_\xi c_{\epsilon_n, \alpha}^\xi \right) \\ &\geq -c|A|(1 + O(\epsilon_n)). \end{aligned}$$

□

Proposition 9.15. *Let $A \in \mathcal{A}(\Omega)$, and $\left\{f_\epsilon^\xi\right\}$ satisfy (H4) and (H5). If $u \in V$ then*

$$F''(u, A) \leq C|A| \tag{9.20}$$

for some positive constant C independent of u and A.

Proof. -Let $u \in C^0(\Omega)$ and let define u_n by $u_n(x) = u(\alpha)$ for all $x \in \alpha + [0, \epsilon_n[^N$. We then have $u_n \to u$ in $L^\infty(\Omega)$ and

$$\begin{aligned} F_{\epsilon_n}(u_n, A) &\leq \sum_{\xi \in \mathbb{Z}^N} \sum_{\alpha \in R^\xi_{\epsilon_n}(A)} \epsilon_n^N C^\xi_{\epsilon_n,\alpha} \\ &= \sum_{\xi \in \mathbb{Z}^N} \sum_{\alpha \in R^\xi_{\epsilon_n}(A)} \epsilon_n^N C^\xi_{\epsilon_n,\alpha} \\ &\leq C(|A| + O(\epsilon_n)). \end{aligned}$$

Letting ϵ_n go to zero, we obtain

$$F''(u, A) \leq C|A|.$$

Using a density argument, we deduce the thesis for all $u \in V$.

□

The conclusion is achieved by reasoning on V endowed with the weak convergence of $L^2(\Omega)$ e.g. and using the results of the previous section for the existence of the Γ-limit and the integral representation formula.

9.4 Minimum problems

In this section we derive a convergence result for minimum problems whenever the functionals are subjected to mean type constraints. Let us introduce the notion of discrete mean.

Definition 9.16. *For any A open subset of Ω, $\epsilon > 0$, and $u \in \mathcal{A}_\epsilon(A)$, we set*

$$\langle u \rangle_A^{d,\epsilon} = \frac{1}{\#(\epsilon\mathbb{Z}^N \cap A)} \sum_{\alpha \in \epsilon\mathbb{Z}^N \cap A} u(\alpha).$$

Let $z \in \mathbb{R}^m$, we define $F^z_\epsilon : L^p(\Omega) \times \mathcal{A}(\Omega)$ by

$$F^z_\epsilon(u, A) = \begin{cases} F_\epsilon(u, A) & \langle u \rangle_A^{\epsilon,d} = z \\ +\infty & \text{otherwise.} \end{cases} \tag{9.21}$$

The following theorem holds true.

Theorem 9.17. *Let $\{f^\xi_\epsilon\}$ satisfy hypotheses (H1)-(H3). Given a sequence of positive real numbers converging to 0, let (ϵ_j) and f be as in Theorem 9.3. For any $z \in \mathbb{R}^m$, let $F^z_{\epsilon_j}$ be as in (9.21). Then, for any $A \in \mathcal{A}(\Omega)$, $(F^z_{\epsilon_j}(\cdot, A))$ $\Gamma(w - L^p)$-converges to the functional $F^z : L^p(\Omega) \times \mathcal{A}(\Omega) \to (-\infty, +\infty]$ defined as*

$$F^z(u, A) = \begin{cases} \int_A f(x, u)\, dx & \langle u \rangle_A = z \\ +\infty & \textit{otherwise.} \end{cases}$$

Proof. -Let us first prove the lower bound inequality. Let (u_j) be a sequence of functions converging to u w.r.t. the weak convergence of $L^p(\Omega)$ such that

$$\liminf_j F^z_{\epsilon_j}(u_j, A) = \lim_j F^z_{\epsilon_j}(u_j, A) < +\infty.$$

Then, for any $A \in \mathcal{A}(\Omega)$, $\langle u_j \rangle_A^{\epsilon_j,d} = z$ and, by the equi-integrability of u_j we get that $\langle u \rangle_A = z$. Then the lower bound inequality follows by Theorem 9.3, observing that

$$F^z_{\epsilon_j}(u_j, A) \geq F_{\epsilon_j}(u_j, A).$$

To prove the upper bound inequality let us observe that, fixed $z \in \mathbb{R}^m$ and $u \in L^p(\Omega)$ such that $\langle u \rangle_A = z$, by using the same argument exploited in the proof of Proposition 9.11, for every $\delta > 0$ there exists $B \subset\subset A$ and a sequence of functions $u_j \to u$ weakly in $L^p(\Omega)$ such that

$$\limsup_j F_{\epsilon_j}(u_j, A) \leq F(u, A) + \delta,$$
$$\limsup_j \sum_{\alpha \in R^\xi_\epsilon(A)} \epsilon^N (|u_j(\alpha)|^p + D_\epsilon(\alpha)) \leq C\delta$$

for some constant $C > 0$. Set $z_j = \langle u_j \rangle_A^{\epsilon_j, d}$ and let B' be such that $B \subset\subset B' \subset\subset A$. We then define

$$v_j(\alpha) = \begin{cases} u_j(\alpha) & \alpha \in \epsilon_j \mathbb{Z}^N \cap B' \\ u_j(\alpha) + c_j & \alpha \in \epsilon_j \mathbb{Z}^N \cap (A \setminus \overline{B'}), \end{cases}$$

where $c_j = (z - z_j) \frac{\#(\epsilon_j \mathbb{Z}^N \cap A)}{\#(\epsilon_j \mathbb{Z}^N \cap (A \setminus \overline{B'})}$. Thanks to this definition we have that $v_j \to u$ weakly in $L^p(A)$, and $\langle v_j \rangle_A^{\epsilon_j, d} = z$. Observing that we have constructed v_j modifying u_j on a set where we the energy is controlled by δ, we conclude that

$$\limsup_j F^z_{\epsilon_j}(v_j, A) \leq F^z(u, A) + \delta.$$

By letting δ go to 0 we obtain the claim.

Remark 9.18. For all $\eta > 0$ set $A_\eta = \{x \in A | d(x, \partial A) > \eta\}$. If for every $R > 0$ we define

$$F^z_\epsilon(u, A) = \begin{cases} F_\epsilon(u, A) & \langle u \rangle_A^{\epsilon, d} = z \text{ and } u(\alpha) = z \text{ if } \alpha \in A \setminus A_{\epsilon R} \\ +\infty & \text{otherwise.} \end{cases} \tag{9.22}$$

the analogous convergence result holds.

By the equicoercivity of the energies F^z_ϵ and the properties of Γ-convergence we derive the following corollary

Corollary 9.19. *Under the hypotheses of Theorem 9.17, for any $z \in R^m$, $A \in \mathcal{A}$ and for R large enough,*

$$\lim_j \Big(\inf\{F_{\epsilon_j}(v, A) : \ \langle v \rangle_A^{\epsilon, d} = z \text{ and } v(\alpha) = z \text{ if } \alpha \in A \setminus A_{\epsilon R}\} \Big)$$
$$= \min\{F(v, A) : \ \langle v \rangle_A = z\}.$$

In addition, if $u_j \rightharpoonup u$ is such that

$$\lim_j F_{\epsilon_j}(u_j, A) = \lim_j \Big(\inf\{F_{\epsilon_j}(v, A) : \ \langle v \rangle_A^{\epsilon, d} = z \text{ and } v(\alpha) = z \text{ if } \alpha \in A \setminus A_{\epsilon R}\} \Big),$$

then $F(u) = \min\{F(v, A) : \ \langle v \rangle_A = z\}$.

Proof. -It suffices to observe that, by the coercivity assumption (H1), for R large enough, the minimizing sequence u_j is bounded in the L^p-norm. Then the conclusion follows by Theorem 9.17 and the properties of Γ-convergence.

□

In the L^∞ case, since the functions u take values in a set which will be relaxed in the limit procedure, one is forced to relax the condition on the discrete mean and consider, for all $z \in \mathbb{R}^m$ and $\rho > 0$, the functional $F^{z,\rho}_\epsilon : L^\infty(\Omega) \times \mathcal{A}(\Omega) \to \mathbb{R}$ given by

$$F^{z,\rho}_\epsilon(u, A) = \begin{cases} F_\epsilon(u, A) & \langle u \rangle_A^{\epsilon, d} \in \overline{B}(z, \rho) \\ +\infty & \text{otherwise,} \end{cases} \tag{9.23}$$

with F_ϵ as in (9.3). The following Γ-convergence result holds true.

Theorem 9.20. *Let $\{f_\epsilon^\xi\}$ satisfy hypotheses (H4)-(H6). Given a sequence of positive real numbers converging to* 0*, let (ϵ_j) and f be as in Theorem 9.5. Then, for any $z \in \mathbb{R}^m$, $\rho > 0$ and $A \in \mathcal{A}(\Omega)$, $(F_{\epsilon_j}^{z,\rho}(\cdot, A))$ $\Gamma(w * -L^\infty)$-converges to the functional $F^{z,\rho} : L^\infty(\Omega) \times \mathcal{A}(\Omega) \to (-\infty, +\infty]$ defined by*

$$F^{z,\rho}(u, A) = \begin{cases} \int_A f(x, u)\ dx & u \in L^\infty(A; \overline{K})\ \langle u \rangle_A \in \overline{B}(z, \rho) \\ +\infty & \textit{otherwise.} \end{cases} \tag{9.24}$$

Proof. -The lower bound inequality is a consequence of Theorem 9.5, observing that the constraint is closed under weak *-convergence. By density it is enough to prove the upper bound inequality for u such that $\langle u \rangle_A \in B(z, \rho)$. For such a u we conclude by noting that the optimizing sequence u_j for $F(u, A)$ satisfies the constraint $\langle u \rangle_A^{\epsilon_j, d} \in B(z, \rho)$ for j large enough.

□

By the properties of Γ-convergence, the previous theorem yields the following result for the convergence of minimum problems.

Corollary 9.21. *Under the hypotheses of Theorem 9.20, for any $z \in R^m$, $\rho > 0$ and $A \in \mathcal{A}(\Omega)$,*

$$\lim_j \Big(\inf\{F_{\epsilon_j}(v, A) :\ \langle v \rangle_A^{\epsilon_j, d} \in \overline{B}(z, \rho)\}\Big) = \min\{F(v, A) :\ \langle v \rangle_A \in \overline{B}(z, \rho)\}.$$

In addition, if $u_j \rightharpoonup u$ is such that

$$\lim_j F_{\epsilon_j}(u_j, A) = \lim_\rho \lim_j \Big(\inf\{F_{\epsilon_j}(v, A) :\ \langle v \rangle_A^{\epsilon_j, d} \in \overline{B}(z, \rho)\}\Big),$$

then $F(u, A) = \min\{F(v, A) :\ \langle v \rangle_A \in \overline{B}(z, \rho)\}$.

9.5 Homogenization

If the functions f_ϵ^ξ are obtained by scaling functions f^ξ periodic in the space variable, and satisfy some growth properties, then the Γ-limit does not depend on the extraction and is given by a homogenization cell formula.

9.5.1 Homogenization in L^p, $1 < p < \infty$

The general case

Let $k \in \mathbb{N}$ and for any $\xi \in \mathbb{Z}^N$, let $f^\xi : \mathbb{Z}^N \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ be such that $f^\xi(\cdot, u, v)$ is k-periodic for any $u, v \in \mathbb{R}^m$. We then set

$$f_\epsilon^\xi(\alpha, u, v) := f^\xi(\frac{\alpha}{\epsilon}, u, v). \tag{9.25}$$

In this case, hypotheses (H1)-(H3) read

(H7) For all β and ξ there exist $c^\xi \geq 0$ and $d^\xi \geq 0$ such that

$$f^\xi(\beta, u, v) \geq c^\xi(|u|^p + |v|^p) - d^\xi$$

for all $(u, v) \in \mathbb{R}^{2m}$, there exists $\bar{\xi} \in \mathbb{Z}^N$ with $c^{\bar{\xi}} > 0$, and $\sum_\xi d^\xi < \infty$.

(H8) For all β and ξ, there exists $C^\xi \geq 0$ such that

$$f^\xi(\beta, u, v) \leq C^\xi(|u|^p + |v|^p + 1)$$

for all $(u, v) \in \mathbb{R}^{2m}$, and $\sum_\xi C^\xi < \infty$.

In what follows, for the sake of clarity, we write $\langle u \rangle_A^d$ for $\langle u \rangle_A^{d,1}$. We have the following

Theorem 9.22. *Let $\{f_\epsilon^\xi\}$ satisfy (9.25), (H7) and (H8). Then the functional F_ϵ $\Gamma(w-L^p)$-converges to*

$$F(u)=\int_\Omega f_{hom}(u(x))dx$$

for all $u\in L^p(\Omega)$, where f_{hom} is given by the homogenization formula

$$f_{hom}(z)=\lim_{h\to+\infty}\frac{1}{h^N}\inf\left\{\sum_{\xi\in\mathbb{Z}^N}\sum_{\beta\in R_1^\xi(Q_h)}f^\xi(\beta,v(\beta),v(\beta+\xi)),\langle v\rangle_{Q_h}^d=z\right\} \tag{9.26}$$

and $Q_h=(0,h)^N$.

Proof. -Let (ϵ_n) be a sequence of positive numbers converging to 0. Then, by Theorem 9.3, we can extract a subsequence (not relabeled) such that $(F_{\epsilon_n}(\cdot,A))$ Γ-converges to a functional $F(\cdot,A)$ defined as in (9.6). The theorem is proved if we show that the energy density f does not depend on the space variable x and if $f\equiv f_{hom}$.

To prove the independence upon the space variable, it is enough to show that

$$F(z,B(x,\rho))=F(z,B(y,\rho))$$

for all $x,y\in\Omega$, $\rho>0$ and $z\in\mathbb{R}^m$. Using the inner regularity and by changing the roles of x and y, it suffices to have

$$F(z,B(x,\rho'))\le F(z,B(y,\rho)) \tag{9.27}$$

for all $\rho'<\rho$. Let $v_n\rightharpoonup z$ in $L^p(\Omega)$ be such that

$$\lim_n F_{\epsilon_n}(v_n,B(y,\rho))=F(z,B(y,\rho)).$$

Then set

$$u_n(\alpha)=\begin{cases} v_n\left(\alpha-\epsilon_n\left[\dfrac{x-y}{\epsilon_n}\right]_k\right) & \text{if } \alpha\in\epsilon_n\mathbb{Z}^N\cap B(x,\rho')\\ z & \text{otherwise}\end{cases}$$

Due to the periodicity (9.25), for n large enough, we have

$$F_{\epsilon_n}(u_n,B(x,\rho'))\le F_{\epsilon_n}(v_n,B(y,\rho)).$$

From this, we easily get (9.27) since $u_n\rightharpoonup z$.

The second step consists in proving that $f\equiv f_{hom}$. To this aim, we note that, since $f(\cdot)$ is a convex function, there holds

$$\begin{aligned} f(z)&=\frac{1}{r^N}\min\left\{\int_{Q_r}f(u)dx,\langle u\rangle_{Q_r}=z\right\}\\ &=\lim_n\frac{1}{r^N}\inf\left\{F_{\epsilon_n}(u,Q_r),\langle u\rangle_{Q_r}^{d,\epsilon_n}=z\right\}.\end{aligned} \tag{9.28}$$

The second equality is a consequence of the convergence of minima given in Corollary 9.19. Set $h_n=\left[\dfrac{r}{\epsilon_n}\right]+1$, then (9.28) holds with $\epsilon_n h_n$ instead of r. Eventually, through the change of variable: for $\alpha\in\epsilon\mathbb{Z}^N$,

$$\beta=\frac{\alpha}{\epsilon},\qquad v(\beta)=u(\epsilon\beta), \tag{9.29}$$

we get

$$f(z)=\lim_n\frac{1}{h_n^N}\inf\left\{\sum_{\xi\in\mathbb{Z}^N}\sum_{\beta\in R_1^\xi(Q_{h_n})}f^\xi(\beta,v(\beta),v(\beta+\xi)),\langle v\rangle_{Q_{h_n}}=z\right\}.$$

One then infers the thesis from the existence of $\lim_{n\to\infty} I(n,z)$, where

$$I(n,z) = \frac{1}{n^N} \inf \left\{ \sum_{\xi\in\mathbb{Z}^N} \sum_{\beta\in R_1^\xi(Q_n)} f^\xi(\beta, v(\beta), v(\beta+\xi)), \langle v\rangle_{Q_n} = z \right\}. \tag{9.30}$$

To prove the existence of this limit, let us first truncate the range of the interactions and define for any $R>0$,

$$F_1^R(u,Q_n) = \sum_{\xi\in\mathbb{Z}^N, |\xi|\le R} \sum_{\beta\in R_1^\xi(Q_n)} f^\xi(\beta, v(\beta), v(\beta+\xi)),$$

and

$$I^R(n,z) = \frac{1}{n^N} \inf \left\{ F_1^R(u,Q_n), \langle v\rangle_{Q_n} = z \right\}.$$

By the growth and decay hypotheses on the energy density, one can easily prove that

$$\lim_{R\to\infty} \sup_n |I^R(n,z) - I(n,z)| = 0. \tag{9.31}$$

Let us introduce for $n>R$,

$$I^{R,R}(n,z) = \frac{1}{n^N} \inf \left\{ F_1^R(u,Q_n), \langle v\rangle_{Q_n} = z, v(\beta) = z \ \forall \beta \in Q_n \setminus Q_{n-R} \right\}. \tag{9.32}$$

By using the same arguments of Theorem 9.17, thanks to Remark (9.18) and Corollary (9.19), for any sequence $n_h \to +\infty$ there exists a subsequence (not relabelled) such that

$$\lim_h I^R(n_h,z) = \lim_h I^{R,R}(n_h,z). \tag{9.33}$$

The existence of the limit (9.26) is then a consequence of the existence, for all $z\in\mathbb{R}^m$, of $\lim_{n\to\infty} I^{R,R}(n,z)$.

To prove the latter, let $n\in\mathbb{N}$ and v_n be a test function in the infimum problem (9.32) such that

$$\frac{1}{n^N} F_1^R(v_n,Q_n) \le I^{R,R}(n,z) + \frac{1}{n}.$$

For any $k>n$, we then define another test function u_k as follows:

$$u_k(\beta) = \begin{cases} v_n(\beta - n\cdot i) & \text{if } \beta \in n\cdot i + Q_n, \quad i\in\{0,\dots,[\frac{k}{n}]-1\}^N \\ z & \text{otherwise.} \end{cases}$$

By the growth hypotheses and the constancy of u_k near the boundary of Q_n, we get

$$\begin{aligned} I_k^{R,R}(z) &\le \frac{1}{k^N} F_1^R(u_k,Q_k) \le \left[\frac{k}{n}\right]^N \frac{1}{k^N} F_1^R(v_n,Q_n) \\ &\quad + C|z|^p \frac{1}{k^N} \left(k^N - \left[\frac{k}{n}\right]^N n^N + \left[\frac{k}{n}\right]^N \left((n+R)^N - (n-R)^N\right) \right) \\ &\le \left[\frac{k}{n}\right]^N \frac{n^N}{k^N} \left(I_n^{R,R}(z) + \frac{1}{n} \right) \\ &\quad + C|z|^p \frac{1}{k^N} \left(k^N - \left[\frac{k}{n}\right]^N n^N + \left[\frac{k}{n}\right]^N \left((n+R)^N - (n-R)^N\right) \right). \end{aligned}$$

Letting k tend to $+\infty$, we have

$$\limsup_k f_k^{R,R}(M) \le I_n^{R,R}(z) + \frac{1}{n} + C|z|^p \frac{1}{n^N} \left((n+R)^N - (n-R)^N\right)$$

Eventually, letting n tend to $+\infty$, we obtain

$$\limsup_k I_k^{R,R}(z) \le \liminf_n I_n^{R,R}(z),$$

which is the claim.

□

The convex case

In this subsection we prove that in the convex case the function f_{hom} can be obtained by a minimization problem posed on the single periodic cell $Q_k = [0,k)^N$. Set

$$I_k = \{0, \dots, k-1\}^N$$

and

Theorem 9.23. *Let (f_ϵ^ξ) satisfy the assumptions of Theorem 9.22 and in addition let $f_\epsilon^\xi(\alpha, u, v)$ be convex w.r.t. the couple (u, v) for all $\alpha \in \epsilon\mathbb{Z}^N$, $\epsilon > 0$ and $\xi \in \mathbb{Z}^N$. Then the conclusion of Theorem 9.22 holds with f_{hom} given by*

$$f_{hom}(z) = \frac{1}{k^N} \inf\Big\{ \sum_{\xi\in\mathbb{Z}^N} \sum_{\beta\in I_k} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right), \quad \langle v\rangle^d_{Q_k} = z \Big\},$$

for all $z \in \mathbb{R}^N$.

Proof. -Set

$$\overline{f}(z) = \frac{1}{k^N} \inf\Big\{ \sum_{\xi\in\mathbb{Z}^N} \sum_{\beta\in I_k} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right), \quad \langle v\rangle^d_{Q_k} = z \Big\}.$$

We first prove that

$$f_{hom}(z) \le \overline{f}(z). \tag{9.34}$$

With fixed $\delta > 0$, let v be such that $\langle v\rangle^d_{Q_k} = z$ and

$$\frac{1}{k^N} \sum_{\xi\in\mathbb{Z}^N} \sum_{\beta\in I_k} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right) \le \overline{f}(z) + \delta.$$

For $n \in \mathbb{N}$, let $I(n, z)$ be as in (9.30). Since in particular $\langle v\rangle^d_{Q_{n\cdot k}} = z$,

$$\begin{aligned} I(n\cdot k, z) &\le \frac{1}{n^N k^N} \sum_{\xi\in\mathbb{Z}^N} \sum_{\beta\in R_1^\xi(Q_{n\cdot k})} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right) \\ &\le \frac{1}{k^N} \sum_{\xi\in\mathbb{Z}^N} \sum_{\beta\in I_k} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right) \le \overline{f}(z) + \delta. \end{aligned}$$

Estimate (9.34) follows by letting n go to $+\infty$, thanks to the arbitrariness of δ. We now prove that

$$f_{hom}(z) \ge \overline{f}(z).$$

To this end we set

$$\overline{f}^R(z) = \frac{1}{k^N} \inf\Big\{ \sum_{|\xi|\le R} \sum_{\beta\in I_k} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right), \quad \langle v\rangle^d_{Q_k} = z \Big\},$$

and

$$f^R_{hom}(z) = \lim_{n\to+\infty} I^{R,R}(n, z)$$

where $I^{R,R}(n, z)$ is defined by (9.32). Using (9.31) and (9.33) we get

$$\lim_{R\to+\infty} f^R_{hom}(z) = f_{hom}(z).$$

Analogously one can show that

$$\lim_{R\to+\infty} \overline{f}^R(z) = \overline{f}(z).$$

Thus it suffices to prove that, for every $R > 0$,

$$f^R_{hom}(z) \geq \overline{f}^R(z). \tag{9.35}$$

For $n \in \mathbb{N},\ n \cdot k > R$, let v be such that $\langle v\rangle_{Q_{n\cdot k}} = z, v(\beta) = z\ \forall \beta \in Q_{n\cdot k} \setminus Q_{n\cdot k-R}$. Hence

$$\begin{aligned}
&\frac{1}{n^N k^N} \sum_{|\xi|\leq R} \sum_{\beta\in R^\xi_1(Q_{n\cdot k})} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right)\\
&= \frac{1}{n^N k^N} \sum_{|\xi|\leq R} \sum_{\beta\in I_{n\cdot k}} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right) - O(\frac{1}{n})\\
&= \frac{1}{k^N} \sum_{|\xi|\leq R} \sum_{\beta\in I_k} \frac{1}{n^N} \sum_{\gamma\in\{1,\dots,n\}^N} f^\xi\left(\beta, v(\beta + k\sum_{i=1}^N \gamma_i e_i), v(\beta + k\sum_{i=1}^N \gamma_i e_i + \xi)\right) - O(\frac{1}{n})\\
&\geq \frac{1}{k^N} \sum_{|\xi|\leq R} \sum_{\beta\in I_k} f^\xi\left(\beta, v_n(\beta), v_n(\beta+\xi)\right) - O(\frac{1}{n}),
\end{aligned} \tag{9.36}$$

where we have set

$$v_n(\beta) = \frac{1}{n^N} \sum_{\gamma\in\{1,\dots,n\}^N} v(\beta + k\sum_{i=1}^N \gamma_i e_i)$$

and the last inequality follows by the convexity hypothesis on f^ξ. Since $\langle v_n\rangle_{Q_k} = z$, by (9.36) and the definition of $\overline{f}^R(z)$, we get

$$\frac{1}{n^N k^N} \sum_{|\xi|\leq R} \sum_{\beta\in R^\xi_1(Q_{n\cdot k})} f^\xi\left(\beta, v(\beta), v(\beta+\xi)\right) \geq \overline{f}^R(z) - O(\frac{1}{n}).$$

Taking the infimum with respect to v and then letting n tend to $+\infty$, we obtain (9.35).

□

9.5.2 Homogenization in L^∞

Let f^ξ_ϵ be as in (9.25) where $f^\xi(\cdot, u, v)$ is k-periodic for any $u, v \in \mathbb{R}^m$. In this case hypotheses (H4)-(H6) read:

(H9) For all β and ξ, $f^\xi(\beta, u, v) = +\infty$ if $(u, v) \notin K^2$.
(H10)For all β and ξ, there exists $C^\xi \geq 0$ such that $|f^\xi(\beta, u, v)| \leq C^\xi$ for all $(u, v) \in K^2$, and $\sum_\xi C^\xi < \infty$.

The following theorem holds.

Theorem 9.24. *Let $\{f^\xi_\epsilon\}$ satisfy* (9.25)*, (H9) and (H10). Then F_ϵ $\Gamma(w*-L^\infty)$-converges to*

$$F(u) = \int_\Omega f_{hom}(u(x))dx$$

for all $u \in L^\infty(\Omega; \overline{K})$, where f_{hom} is given by the homogenization formula

$$f_{hom}(z) = \lim_{\rho\to 0} \lim_{h\to+\infty} \frac{1}{h^N} \inf\left\{ \sum_{\xi\in\mathbb{Z}^N} \sum_{\beta\in R^\xi_1(Q_h)} f^\xi(\beta, v(\beta), v(\beta+\xi)), \langle v\rangle^d_{Q_h} \in \overline{B}(z,\rho)\right\}. \tag{9.37}$$

Proof. -Let (ϵ_n) be a sequence of positive numbers converging to 0. Then, by Theorem 9.5, we can extract a subsequence (not relabeled) such that $(F_{\epsilon_n}(\cdot, A))$ Γ-converges to a functional $F(\cdot, A)$ defined as in (9.7). The theorem is proved if we show that the density function f does not depend on the space variable x and if $f \equiv f_{hom}$. The proof of the independence on the space variable proceeds as in the L^p case. In order to prove that $f \equiv f_{hom}$ we first observe that, by the convexity of f and Corollary 9.21, it holds

$$\begin{aligned} f(z) &= \lim_{\rho\to 0} \frac{1}{r^N} \min\left\{ \int_{Q_r} f(u)dx, \langle u\rangle_{Q_r} \in \overline{B}(z,\rho)\right\} \\ &= \lim_{\rho\to 0} \lim_n \frac{1}{r^N} \inf\left\{ F_{\epsilon_n}(u, Q_r), \langle u\rangle^{d,\epsilon_n}_{Q_r} \in \overline{B}(z,\rho)\right\}. \end{aligned} \tag{9.38}$$

Analogously to the L^p case we scale the problem as follows. Setting $h_n = \left[\dfrac{r}{\epsilon_n}\right] + 1$, through the change of variable: for all $\alpha \in \epsilon\mathbb{Z}^N$,

$$\beta = \frac{\alpha}{\epsilon}, \qquad v(\beta) = u(\epsilon\beta),$$

equality (9.38) becomes

$$f(z) = \lim_{\rho\to 0} \lim_{n\to+\infty} \frac{1}{h_n^N} \inf\left\{ \sum_{\xi\in\mathbb{Z}^N} \sum_{\beta\in R_1^\xi(Q_{h_n})} f^\xi(\beta, v(\beta), v(\beta+\xi)), \langle v\rangle^d_{Q_{h_n}} \in \overline{B}(z,\rho)\right\}.$$

The conclusion follows by proving the existence of the first limit in (9.37) for any $\rho > 0$. This can be done by repeating the construction used in the L^p case.

□

Remark 9.25. **(Bravais lattices)** The analysis we have made since now in the model case of a square lattice extends to the case of more general Bravais lattices.

9.6 Ferromagnetic-antiferromagnetic systems: existence of the bulk limit

In this section, we recall the model dealt with in [98] and we prove that it can be recast in the setting of Section 9.3. In particular, the associated family of energies satisfy the hypotheses of Theorem 9.24.

Given an integer M, let Λ_M denote $[-M, M[^d \cap \mathbb{Z}^N$. The energy of a Λ_M-periodic configuration $\underline{\sigma} : \Lambda_M \to \{-1, 1\}$ is given by

$$H_M(\underline{\sigma}) = -J \sum_{k=1}^N \sum_{i\in\Lambda_M} \sigma_i\sigma_{i+e_k} + \sum_{i,j\in\Lambda_M, i\neq j} \sigma_i J_p(j-i)\sigma_j, \tag{9.39}$$

where $J > 0$ (and if $i + e_k \notin \Lambda_M$ we assume $\sigma_{i+e_k} = \sigma_{i-2Me_k}$), and J_p is defined, for $p > 1$, by

$$J_p(j-i) = \sum_{k\in\mathbb{Z}^N} \frac{1}{|i-j+2kM|^p}.$$

The first term of (9.39) models the ferromagnetic interactions between nearest neighbors (with periodic conditions, which means that the whole space $\mathbb{Z}^N$ is covered with the periodic replication of Λ_M) and is called the 'exchange energy'. The second term models the antiferromagnetic interactions at long range (also with periodic boundary conditions). It is the 'dipolar energy'. Heuristically, short range interactions prefer uniform states (either of $+1$ or -1), and long range interactions favor alternating states $(+1, -1)$.

The problem of the variational convergence of $\frac{H_M(\sigma)}{M^N}$ as $M \to +\infty$ can be equivalently studied on a fixed domain $\Lambda = [-1,1)^N$. To this end we set $\epsilon = \frac{1}{M}$ and, for any $\underline{\sigma} : \Lambda_M \to \{-1,1\}$, $u(\alpha) := \underline{\sigma}(\frac{\alpha}{\epsilon})$ for all $\alpha \in \epsilon\mathbb{Z}^N \cap \Lambda$. Then, up to lower order terms, we can rewrite $\frac{H_M(\bar{\sigma})}{M^N}$ as follows:

$$F_\epsilon(u) = F_\epsilon^1(u) + F_\epsilon^2(u),$$

where

$$F_\epsilon^1(u) = -J\sum_{k=1}^N \sum_{\alpha\in R_\epsilon^{e_k}(\Lambda)} \epsilon^d u(\alpha)u(\alpha+\epsilon e_k) + \sum_{\alpha_1,\alpha_2\in\epsilon\mathbb{Z}^N\cap\Lambda:\ \alpha_1\neq\alpha_2} \epsilon^N\epsilon^p \frac{u(\alpha_1)u(\alpha_2)}{|\alpha_1-\alpha_2|^p},$$

and

$$\begin{aligned} F_\epsilon^2(u) &= \sum_{\alpha_1,\alpha_2\in\epsilon\mathbb{Z}^N\cap\Lambda:\ \alpha_1\neq\alpha_2} \epsilon^N \sum_{k\in\mathbb{Z}^N\setminus\{0\}} \epsilon^p \frac{u(\alpha_1)u(\alpha_2)}{|\alpha_1-\alpha_2+2k|^p} \\ &= \sum_{\alpha_1,\alpha_2\in\epsilon\mathbb{Z}^N\cap\Lambda:\ \alpha_1\neq\alpha_2} \epsilon^N (f_1^\epsilon(\alpha_1-\alpha_2, u(\alpha_1), u(\alpha_2)) \\ &\qquad\qquad + f_2^\epsilon(\alpha_1-\alpha_2, u(\alpha_1), u(\alpha_2))), \end{aligned}$$

where

$$\begin{aligned} f_1^\epsilon(z,u,v) &= \sum_{k\in\mathbb{Z}^N, |k|>\sqrt{N}+1} \frac{\epsilon^p uv}{|z+2k|^p}, \\ f_2^\epsilon(z,u,v) &= \sum_{0<|k|\le\sqrt{N}+1} \frac{\epsilon^p uv}{|z+2k|^p}. \end{aligned}$$

Let us prove that for $p > N$,

$$\lim_{\epsilon\to 0} F_\epsilon^2(u) = 0 \tag{9.40}$$

uniformly with respect to u. Once (9.40) is proved, we have

$$\Gamma - \lim_{\epsilon\to 0} F_\epsilon(u) = \Gamma - \lim_{\epsilon\to 0} F_\epsilon^1(u).$$

In addition, $F_\epsilon^1(u)$ can be rewritten as

$$F_\epsilon^1(u) = -J\sum_{k=1}^N \sum_{\alpha\in R_\epsilon^{e_k}(\Lambda)} \epsilon^N u(\alpha)u(\alpha+\epsilon e_k) + \sum_{\xi\in\mathbb{Z}^N} \sum_{\alpha\in R_\epsilon^{\xi}(\Lambda)} \epsilon^N \frac{u(\alpha)u(\alpha+\epsilon\xi)}{|\xi|^p}$$

and turns out to satisfy the hypotheses of Theorem 9.24 for $p > N$. This implies the integral representation property of its Γ-limit.

To prove (9.40) we first estimate the term in the energy with f_1^ϵ. Since $|\alpha_1-\alpha_2| < 2\sqrt{N}$ and $k > \sqrt{N}+1$, by applying the triangular inequality, we have that

$$|\alpha_1-\alpha_2+2k|^p \ge ||2k| - |\alpha_1-\alpha_2||^p \ge |2k|^p \left|1 - \frac{|\alpha_1-\alpha_2|}{|2k|}\right|^p \ge C|2k|^p.$$

Thus

$$|f_1^\epsilon(z,u,v)| \le C\epsilon^p \sum_{k\in\mathbb{Z}^N,\ |k|\neq 0} \frac{1}{|2k|^p} \le C\epsilon^p$$

for $p > N$, and

$$\begin{aligned}
&\sum_{\alpha_1,\alpha_2\in\epsilon\mathbb{Z}^N\cap\Lambda:\ \alpha_1\neq\alpha_2} \epsilon^N |f_1^\epsilon(\alpha_1-\alpha_2,u(\alpha_1),u(\alpha_2))| \\
&\le \epsilon^N \sum_{\alpha_1,\alpha_2\in\epsilon\mathbb{Z}^N\cap\Lambda:\ \alpha_1\neq\alpha_2} \epsilon^p \sum_{k\in\mathbb{Z}^N,\ |k|\neq 0} \frac{1}{|2k|^p} \\
&\le C\epsilon^{N+p}\epsilon^{-2N} = C\epsilon^{p-N}.
\end{aligned}$$

To estimate the term with f_2^ϵ one has to be more precise. Noting that $|\alpha_1-\alpha_2+2k|\ge\epsilon$ we collect the interactions according to a logarithmic scale in ϵ as follows:

$$\begin{aligned}
&\sum_{\alpha_1,\alpha_2\in\epsilon\mathbb{Z}^N\cap\Lambda:\ \alpha_1\neq\alpha_2} \epsilon^N |f_2^\epsilon(\alpha_1-\alpha_2,u(\alpha_1),u(\alpha_2))| \\
&\qquad \le \epsilon^{N+p} \sum_{0<|k|<\sqrt{N}+1} \sum_{i=0}^{M-1} \sum_{\alpha_1,\alpha_2\in I_i} \frac{1}{|\alpha_1-\alpha_2+2k|^p} \\
&\qquad\quad + \epsilon^{N+p} \sum_{0<|k|<\sqrt{N}+1} \sum_{|\alpha_1-\alpha_2+2k|\ge 1} \frac{1}{|\alpha_1-\alpha_2+2k|^p},
\end{aligned} \tag{9.41}$$

where, for $i\in\{0,1,\dots,M-1\}$, we have set

$$I_i = \{(\alpha_1,\alpha_2)\in(\epsilon\mathbb{Z}^N\cap\Lambda)^2:\ \epsilon^{\frac{i+1}{M}}\le|\alpha_1-\alpha_2+2k|<\epsilon^{\frac{i}{M}}\}.$$

Since $I_i\subset\tilde{I}_i := \{(\alpha_1,\alpha_2)\in(\epsilon\mathbb{Z}^N\cap\Lambda)^2:\ |\alpha_1-\alpha_2+2k|<\epsilon^{\frac{i}{M}}\}$, we have that

$$\#(I_i)\le\#(\tilde{I}_i)\le C\epsilon^{\frac{(N+1)i}{M}}\epsilon^{-2N}. \tag{9.42}$$

Set, for $\eta>0$ $I^\eta := \{\alpha_1,\alpha_2\in\epsilon\mathbb{Z}^N\cap\Lambda:\ |\alpha_1-\alpha_2+2k|\le\eta\}$. One can show $\#(I^\eta)\le C\left(\frac{\eta}{\epsilon^N}\right)\left(\frac{\eta^N}{\epsilon^N}\right)$. Since

$$\epsilon^{N+p} \sum_{0<|k|<\sqrt{N}+1} \sum_{|\alpha_1-\alpha_2+2k|\ge 1} \frac{1}{|\alpha_1-\alpha_2+2k|^p} \le C\epsilon^{p-N}$$

we only need to estimate the first term in the right hand side of (9.41) to conclude. Using (9.42), we have

$$\begin{aligned}
\epsilon^{N+p} \sum_{0<|k|<\sqrt{N}+1} \sum_{i=0}^{M-1} \sum_{\alpha_1,\alpha_2\in I_i} \frac{1}{|\alpha_1-\alpha_2+2k|^p} &\le C\epsilon^{N+p} \sum_{i=0}^{M-1} \frac{\#(I_i)}{\epsilon^{\frac{p(i+1)}{M}}} \\
&\le C\epsilon^{p-N} \sum_{i=0}^{M-1} \epsilon^{\frac{(N+1-p)i}{M}} \epsilon^{-\frac{p}{M}} \\
&= C\epsilon^{p-N-\frac{p}{M}} \sum_{i=0}^{M-1} \left(\epsilon^{\frac{(N+1-p)}{M}}\right)^i =: L(\epsilon,M).
\end{aligned} \tag{9.43}$$

If $p=N+1$ then

$$L(\epsilon,M)\le CM\epsilon^{1-\frac{N+1}{M}}$$

which converges to zero as $\epsilon\to 0$ provided M is chosen large enough. If $p\neq N+1$, let $q=\epsilon^{\frac{N+1-p}{M}}$. As $\sum_{i=0}^{M-1} q^i = \frac{1-q^M}{1-q}$, there holds

$$L(\epsilon,M)\le C\epsilon^{p-N-\frac{p}{M}}\left(\frac{1-\epsilon^{N+1-p}}{1-\epsilon^{N+1-pM}}\right).$$

It is then easy to verify that the last term converges to zero as $\epsilon\to 0$ for M large enough.

9.7 Non-pairwise-interaction energies

In this section we deal with more general discrete systems driven by non pairwise-interaction energies. For such systems, for any $u \in C^m_\epsilon(\Omega)$, the energy is given by

$$F_\epsilon(u) = \sum_{j=1}^{k} \sum_{\overline{\xi} \in \mathbb{Z}^{jN}} \sum_{\alpha \in R^{\overline{\xi}}_\epsilon(\Omega)} \epsilon^N f^{\overline{\xi}}_\epsilon(\alpha, u(\alpha), u(\alpha+\epsilon\xi_1), \dots, u(\alpha+\epsilon\xi_j)) \tag{9.44}$$

where $\overline{\xi} = (\xi_1, \xi_2, \dots, \xi_j) \in \mathbb{Z}^{jN}$ and

$$R^{\overline{\xi}}_\epsilon(\Omega) = \{\alpha \in \epsilon\mathbb{Z}^N : \ \alpha, \alpha+\epsilon\xi_1, \dots, \alpha+\epsilon\xi_j \in \mathbb{Z}_\epsilon(\Omega)\}.$$

It may be easily checked that all the arguments we have used so far to prove our results in the case of pairwise-interacting discrete systems can be exploited in order to treat more general systems driven by non pairwise-interaction energies of the form (9.44) provided that we modify assumptions (H1)-(H6) by substituting in each formula ξ by $\overline{\xi}$ and $|\xi|$ by $\|\overline{\xi}\|_\infty := \max_{i\in\{1,\dots,j\}} |\xi_i|$. Under this new set of hypotheses the analogue of all the Theorems we have stated so far hold true.

A particular case of non pairwise-interacting discrete system to which all the previous result apply, is provided by those Heisenberg spin systems driven by energies containing multiple-spin exchange terms, namely energies $F_\epsilon : L^\infty(\Omega; K) \to \mathbb{R}$ of the form

$$F_\epsilon(u) = \sum_{j=2}^{k} J_j \sum_{I(\alpha_1,\dots,\alpha_j)} \epsilon^N u(\alpha_1)u(\alpha_2)\dots u(\alpha_j), \tag{9.45}$$

where $K \in \mathbb{R}^m$ is a bounded set, $k \geq 3$ and for all $j \in \{1, \dots, k\}$, the constant J_j is also known as the exchange constant of the j-body nearest-neighbors interaction. Here and in the following $I(\alpha_1, \dots, \alpha_j)$ denotes a set of j-ples of points of the lattice subject to some constraints depending on the model. To better specify the set I in some cases of interest, let us introduce some additional notation. Let $\tilde{\mathbb{Z}}$ be a N-dimensional Bravais lattice and $V(\alpha)$ be its open unitary cell centered in α. Scaling by ϵ, we denote by $\tilde{\mathbb{Z}}_\epsilon = \epsilon\tilde{\mathbb{Z}}$, $\tilde{\mathbb{Z}}_\epsilon(\Omega) = \tilde{\mathbb{Z}}_\epsilon \cap \Omega$ and $V_\epsilon(\alpha) = \epsilon V(\alpha)$. For any given $\alpha \in \tilde{\mathbb{Z}}_\epsilon(\Omega)$, we set $\mathcal{B}_\epsilon(\alpha) := B_\epsilon(\alpha) \cap \tilde{\mathbb{Z}}_\epsilon(\Omega)$, the discrete ball of the lattice $\tilde{\mathbb{Z}}_\epsilon(\Omega)$, centered in α and with radius ϵ.

We say that the k-ple $(\alpha_1, \alpha_2, \dots, \alpha_k) \in (\tilde{\mathbb{Z}}_\epsilon(\Omega))^k$ is a k-body chain of nearest neighbors (or shortly a k-chain) if, for all $j \in \{1, 2, \dots, k\}$,

$$\#(\overline{\mathcal{B}}_\epsilon(\alpha_j) \cap \{\alpha_1, \alpha_2, \dots, \alpha_k\}) \in \{2, 3\}.$$

The k-ple $(\alpha_1, \alpha_2, \dots, \alpha_k) \in (\tilde{\mathbb{Z}}_\epsilon(\Omega))^k$ is a k-cycle of nearest neighbors (or shortly a k-cycle) if, for all $j \in \{1, 2, \dots, k\}$,

$$\#(\overline{\mathcal{B}}_\epsilon(\alpha_j) \cap \{\alpha_1, \alpha_2, \dots, \alpha_k\}) = 3.$$

A k-chain $(\alpha_1, \alpha_2, \dots, \alpha_k)$ is contained in a cell (or, in short, is a cell k-chain) of the lattice $\tilde{\mathbb{Z}}_\epsilon(\Omega)$ if

$$\max\{|\alpha_i - \alpha_j|, \ i, j \in \{1, 2, \dots, k\}\} \leq \mathrm{diam}(V_\epsilon(\alpha))$$

for some $\alpha \in \tilde{Z}_\epsilon$.

Depending on the values of the exchange constants and on the constraint specified by the definition of I, many different models can be obtained. In particular one may consider the cases

$$I(\alpha_1, \alpha_2, \dots, \alpha_k) := \{(\alpha_1, \alpha_2, \dots, \alpha_k) \in (\tilde{\mathbb{Z}}_\epsilon(\Omega))^k : \ (\alpha_1, \alpha_2, \dots, \alpha_k) \text{ is a } k\text{-chain}\},$$

and

$$I(\alpha_1, \alpha_2, \dots, \alpha_k) := \{(\alpha_1, \alpha_2, \dots, \alpha_k) \in (\tilde{\mathbb{Z}}_\epsilon(\Omega))^k : \ (\alpha_1, \alpha_2, \dots, \alpha_k) \text{ is a } k\text{-cycle}\}.$$

In the latter cases it is not easy to guess the explicit form of the energy density and many attempts have been done both from the analytical and the computational point of view to address the problem (see e.g. [20], [132], [103]). Let us stress that the homogenization results of Theorems 9.22 and 9.24 hold in the general case, providing with the existence of a local limit functional of integral type and an implicit asymptotic formula for its energy density.

We conclude this section with an example of a two-dimensional ferromagnetic model with 3-spin exchange energy in which it is possible to explicitly write the limit energy.

Example 9.26. Let $\Omega \subset \mathbb{R}^2$ and $K = \{-1, 1\}$. Then, for $k = 3$, $J_2 = 0$ (the case $J_2 \neq 0$ can be dealt with similarly) and $J_3 = -1$ the energy (9.45) reads

$$F_\epsilon(u) = -\sum_{I(\alpha_1,\alpha_2,\alpha_3)} \epsilon^2 u(\alpha_1)u(\alpha_2)u(\alpha_3). \tag{9.46}$$

In what follows we consider the special case of cell 3-chain interactions; i.e. we take

$$I(\alpha_1, \alpha_2, \alpha_3) := \{(\alpha_1, \alpha_2, \alpha_3) \in (\tilde{\mathbb{Z}}_\epsilon(\Omega))^3 : \ (\alpha_1, \alpha_2, \alpha_3) \text{ is a cell 3-chain}\}.$$

We distinguish two cases.
Case (i): triangular lattice. Let $\tilde{Z}$ be a regular triangular lattice, that is $\tilde{Z} = \xi_1\mathbb{Z} \oplus \xi_2\mathbb{Z}$ with $\xi_1 = (1, 0)$ and $\xi_2 = (\frac{1}{2}, \frac{\sqrt{3}}{2})$. The energies (9.46) can be extended on $L^\infty(\Omega; \{-1, 1\})$ by

$$F_\epsilon(u) = \begin{cases} -\sum\limits_{I(\alpha_1,\alpha_2,\alpha_3)} \epsilon^2 u(\alpha_1)u(\alpha_2)u(\alpha_3) & \text{if } u \in C^R_\epsilon(\Omega) \\ +\infty & \text{otherwise,} \end{cases} \tag{9.47}$$

where we have denoted by $C^R_\epsilon(\Omega)$ the set of piecewise-constant functions that take the same value on the rhombus with center α and two sides parallel to the generators of the lattice and of length ϵ.

To perform the Γ-convergence analysis we use an argument similar to the so-called dual lattice approach, already exploited for pairwise-interaction discrete systems in [3]. Let us introduce the following dual lattice

$$\tilde{\mathcal{Z}}_\epsilon(\Omega) := \left\{ \frac{\alpha_1 + \alpha_2 + \alpha_3}{3} : \ (\alpha_1, \alpha_2, \alpha_3) \text{ are the vertices of a cell of } \tilde{\mathbb{Z}}_\epsilon(\Omega) \right\}$$

and let

$$C^T_\epsilon(\Omega) := \{v : \mathbb{R}^2 \to \mathbb{R} : \ v(x) = v(\alpha) \ \forall x \in \{\alpha + V_\epsilon(\alpha)\}, \ \alpha \in \tilde{\mathcal{Z}}_\epsilon(\Omega)\}$$

be the set of all piecewise constant functions on the cells of the lattice $\tilde{\mathcal{Z}}_\epsilon(\Omega)$. With any function $u \in C^R_\epsilon(\Omega)$ let us associate the function $v \in C^T_\epsilon(\Omega)$ given by

$$v(\alpha) = \frac{u(\alpha_1) + u(\alpha_2) + u(\alpha_3)}{3}, \quad \text{where } \alpha = \frac{\alpha_1 + \alpha_2 + \alpha_3}{3}.$$

Note that $v \in \{-1, -\frac{1}{3}, \frac{1}{3}, 1\}$ and that if $u_\epsilon \rightharpoonup^* u$ in $L^\infty(\Omega)$, then $v_\epsilon \rightharpoonup^* u$ in $L^\infty(\Omega)$ where v_ϵ is the sequence of functions associated with u_ϵ. We have

$$F_\epsilon(u) = G_\epsilon(v)$$

where $G_\epsilon : L^\infty(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$G_\epsilon(v) := \begin{cases} \sum\limits_{\alpha \in \tilde{\mathcal{Z}}_\epsilon(\Omega)} \epsilon^2 g^T(v(\alpha)) & \text{if } v \in C^T_\epsilon(\Omega) \\ +\infty & \text{otherwise,} \end{cases}$$

and $g^T : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$, by

$$g^T(z) = \begin{cases} -1 & \text{if } z \in \{-\frac{1}{3}, 1\}, \\ +1 & \text{if } z \in \{-1, \frac{1}{3}\} \\ +\infty & \text{otherwise.} \end{cases}$$

Observe that the change of variables we made allows us to regard the non trivial case of a multiple-exchange spin-type energy in (9.47) as an energy of a non-interacting spin system in the sense of [3]. Then, according to [3, Theorem 1] the following result holds:

Theorem 9.27. *Let $F_\epsilon : L^\infty(\Omega; \{-1, 1\}) \to \mathbb{R} \cup \{+\infty\}$ be as in (9.47), then F_ϵ Γ-converges with respect to the w^*-topology of L^∞ to the functional $F^T : L^\infty(\Omega) \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$F^T(u) = \begin{cases} \int_\Omega \psi^T(u(x))\, dx & \textit{if } u \in L^\infty(\Omega; [-1, 1]) \\ +\infty & \textit{otherwise} \end{cases}$$

where $\psi^T : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ satisfies

$$\psi^T(z) = (g^T)^{**}(z) = \begin{cases} -3z - 2 & \textit{if } -1 \le z \le -\frac{1}{3} \\ -1 & \textit{if } -\frac{1}{3} \le z \le 1 \\ +\infty & \textit{otherwise.} \end{cases}$$

Case (ii): square lattice Let $\tilde{Z} = \mathbb{Z}^2$. In this case the energies (9.46) can be extended on $L^\infty(\Omega; \{-1, 1\})$ by

$$F_\epsilon(u) = \begin{cases} -\sum\limits_{I(\alpha_1, \alpha_2, \alpha_3)} \epsilon^2 u(\alpha_1) u(\alpha_2) u(\alpha_3) & \text{if } u \in C^1_\epsilon(\Omega) \\ +\infty & \text{otherwise.} \end{cases} \tag{9.48}$$

Arguing as before, let

$$\mathcal{Z}_\epsilon(\Omega) := \left\{ \frac{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}{4} : \ (\alpha_1, \alpha_2, \alpha_3, \alpha_4) \text{ are the vertices of a cell of } \mathbb{Z}_\epsilon(\Omega) \right\}$$

and set

$$C^S_\epsilon(\Omega) := \{v : \mathbb{R}^2 \to \mathbb{R} : \ v(x) = v(\alpha) \ \forall x \in \{\alpha + V_\epsilon(\alpha)\}, \ \alpha \in \mathbb{Z}_\epsilon(\Omega)\}.$$

For any function $u \in C^1_\epsilon(\Omega)$ let $v \in C^S_\epsilon(\Omega)$ be given by

$$v(\alpha) = \frac{u(\alpha_1) + u(\alpha_2) + u(\alpha_3) + u(\alpha_4)}{4}, \quad \text{where } \alpha = \frac{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}{4}.$$

Note that $v \in \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$ and that, as in the previous case, if $u_\epsilon \rightharpoonup^* u$ in $L^\infty(\Omega)$, then $v_\epsilon \rightharpoonup^* u$ in $L^\infty(\Omega)$, where v_ϵ is the sequence of functions associated with u_ϵ. There holds

$$F_\epsilon(u) = G_\epsilon(v)$$

where $G_\epsilon : L^\infty(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$G_\epsilon(v) := \begin{cases} \sum\limits_{\alpha \in \mathcal{Z}_\epsilon(\Omega)} \epsilon^2 g^S(v(\alpha)) & \text{if } v \in \tilde{C}^S_\epsilon(\Omega) \\ +\infty & \text{otherwise,} \end{cases}$$

and $g^S : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is the odd function

$$g^S(z) = \begin{cases} +4 & \text{if } z = -1, \\ -2 & \text{if } z = -\frac{1}{2}, \\ +\infty & \text{otherwise on } z < 0. \end{cases}$$

The following result holds:

Theorem 9.28. *Let $F_\epsilon : L^\infty(\Omega; \{-1,1\}) \to \mathbb{R} \cup \{+\infty\}$ be as in (9.48), then F_ϵ Γ-converges with respect to the w^*-topology of L^∞ to the functional $F^S : L^\infty(\Omega) \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$F^S(u) = \begin{cases} \int_\Omega \psi^S(u(x))\,dx & \text{if } u \in L^\infty(\Omega; [-1,1]) \\ +\infty & \text{otherwise} \end{cases}$$

where $\psi^S : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ satisfies

$$\psi^S(z) = (g^S)^{**}(z) = \begin{cases} -12z - 8 & \text{if } -1 \le z \le -\frac{1}{2} \\ -\frac{4}{3}z - \frac{8}{3} & \text{if } -\frac{1}{2} \le z \le 1 \\ +\infty & \text{otherwise.} \end{cases}$$

We remark that some features of the energy density obtained in the two cases are peculiar of the geometric frustration of the system. In fact the triangular case provides an example of non-frustrated system, while the square case is a frustrated spin system. Here the geometric frustration can be read in the fact that the triple of values $(-1, 1, 1)$ minimizes the energy density but cannot be repeated on the square lattice in order to be minimal on each cell of the lattice. The frustration is responsible for the fact that the minimum of the limit-energy density is non-degenerate, which implies that phase-transition phenomena do not take place at scale ϵ. On the contrary, in the triangular case e.g., the limit energy-density ψ^T has multiple minima.

10

Mathematical derivation of a rubber-like stored energy functional

Rubber elasticity is usually "derived" using statistical physics arguments and entropy considerations on one polymeric chain. It is induced at the continuum scale assuming some affine dependence. The stress-strain relation may then be obtained by applying linear boundary conditions on the reference cell oriented according to the principal stretch or by averaging on the orientations. Starting from the same model for one polymeric chain, we adopt here an approach closer to that of Treloar concerning the derivation of the continuous energy. We only rely on the energy of a chain network and on minimization principles. The free energy of an elastomeric chain and the network being given, we rigorously perform a discrete to continuum limit to derive a continuous energy density. We also discuss the motivations of the variational approach and the mechanical properties of the energy density obtained.

10.1 Mesoscopic model

10.1.1 Free energy of a polymeric chain

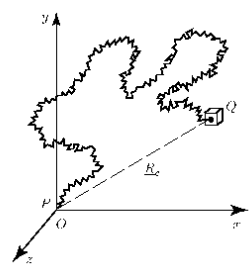

Fig. 10.1. Polymeric chain

Given a polymeric chain made of N rigid segments of length l (see Figures 10.1 and 10.2) at temperature β, with a chain density n, the free energy (of entropic origin) for a chain of length r_c reads

$$\tilde{W}_c(r_c) = \frac{n}{\beta} N \left(\frac{r_c}{Nl} \theta \left(\frac{r_c}{Nl} \right) + \log \frac{\theta \left(\frac{r_c}{Nl} \right)}{\sinh \theta \left(\frac{r_c}{Nl} \right)} \right) - \frac{c}{\beta},$$

where c is a constant and θ the inverse of the Langevin function $\mathcal{L}(\alpha) = \coth \alpha - \frac{1}{\alpha}$.

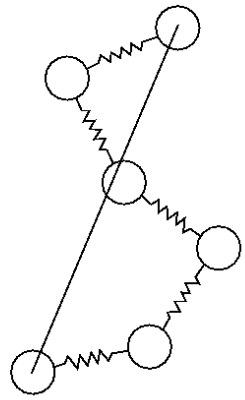

Fig. 10.2. Model of a polymeric chain

In particular, the energy is infinite as soon as $r_c > Nl$, the total length of the chain, which is already a modeling since chains may actually break. For discrete to continuum derivations, θ is usually replaced by the first terms of its series expansions:

$$\theta(r) = 3r + \frac{9}{5}r^3 + \frac{297}{175}r^5 + \frac{1539}{875}r^7 + O(r^9). \tag{10.1}$$

This approximation constitutes a relaxation of the constraint on the maximal length of the chains. The behavior of the series expansion at infinity corresponds to the classical coercivity assumption on hyperelastic materials at infinity in continuum mechanics. Replacing the inverse of the Langevin function by the first terms of a series expansion is a rather good modeling at high temperature (see [127]). Some alternative is discussed in Section 10.4.2. A surprising property of such an energy is $\tilde{W}_c(0) = 0$ and $\tilde{W}_c(1) > 0$. In particular the prefered configuration of a polymer chain satisfies $r_c = 0$. In order to model correctly the mesoscopic structure of a polymer network, we will also add some term depending on the volumetric changes.

In what follows, we deal with polymeric chain free energies depending only on the end-to-end vector r and satisfying the standard growth properties:

$$c|r|^p - 1 \leq W(r) \leq C(1 + |r|^p), \tag{10.2}$$

where c and C are positive constants and $p > 1$. This growth condition is clearly satisfied for $p = 7$ if we use the polynomial expansion (10.1) to approximate $\tilde{W}_c$ (the logarithm plays no role for these bounds). Given the free energy of one polymeric chain, we may introduce the mesoscopic modeling of a rubber-like polymer.

10.1.2 Mesoscopic modeling of a rubber-like polymer

According to [176], the mesoscopic structure of rubber can be seen as a network of crosslinked polymer chains, as it is illustrated on Figure 10.3 in the case of a porous polyamaleimide network (extracted from [112]). This network may be thought of as periodic or as the realization of a stochastic process, as this is often used in mechanics (see [153]). Such a modeling is sketched on Figure 10.4. The typical lengthscale of the network is denoted by $\epsilon = \sqrt{N}l$ and will tend to zero in the following section. A first part of the energy of the system is then given by the sum of the contributions of all the links.

In mechanics, the complex mesoscopic structure is usually bypassed and an *ad hoc* representative volume element (RVE) is defined in order to model the macroscopic behaviour of the polymeric network. One of the most popular models, the eight-chain model, has been introduced by Arruda and Boyce in [10], see also [11,23]. The RVE consists in eight chains that link the center of a cube

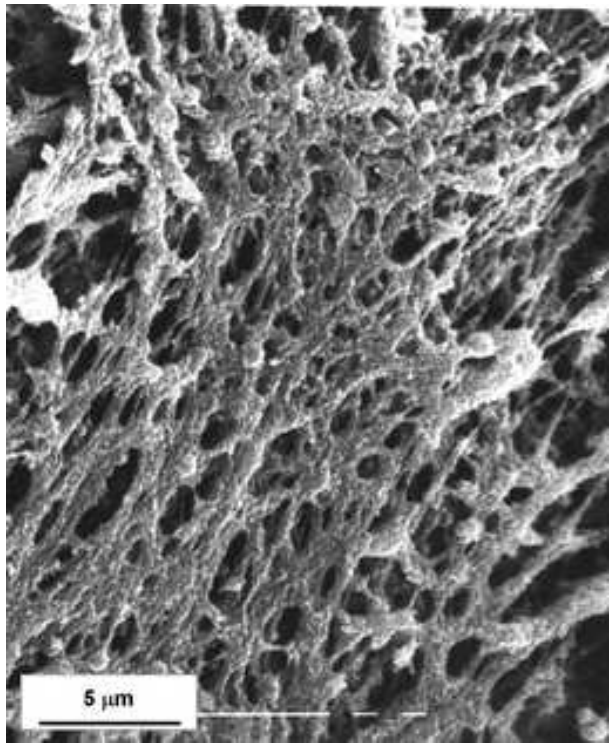

Fig. 10.3. 5 SEM micrograph of a crosslinked porous polymaleimide network

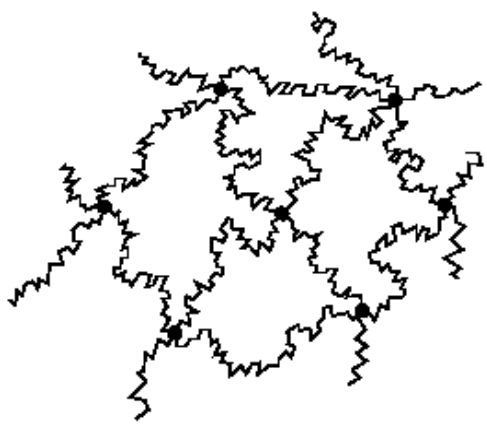

Fig. 10.4. Example of mesoscopic network

to the eight corners. The macroscopic energy associated to a deformation gradient ξ is then given by the energy of the RVE when each edge of the cube is aligned and deformed according to the principal stretches of ξ (and the displacement of the center is obtained by symmetry arguments). In particular, the isotropy of this model is strongly enforced since the RVE follows *by definition* the principal axes of ξ. To this energy is added an incompressible term, whose origin may be the Van der Waals forces. Surprinsingly, although the description is rough and may lack of physical motivation (as questioned by Treloar), the associated macroscopic behaviour is in very good agreement with experiments, even with only two adjustable parameters.

Another approach - geometrically more realistic - consists in using a statistical description of the deformed chains, in terms of length and orientations. To keep the model analytically

tractable, an additional assumption, denoted by *affine assumption*, imposes that the deformation of each chain corresponds exactly to the macroscopic deformation. Within this model, isotropy is an explicit consequence of the structure of the network *at large scales* and is not postulated from the very beginning on. The affine assumption is however unsatisfactory and this has lead Flory to partially relax it within the statistical approach. A counter example to the affine assumption has been given by Friesecke and Theil. In [90], they consider a periodic network, linearly deform its boundary and show that the absolute minimizers are not always affine. There is a partial relaxation in the network.

A rigorous derivation of rubber-like elasticity should not bypass the complex mesoscopic structure. However the derivation should not rely on the affine assumption either. In the present work, we keep the complex mesoscopic structure and we replace the affine assumption by a variational approach which allows for relaxation. Doing this, we face the following difficulty: the one chain energy $\tilde{W}_c$ does not allow the network structures encountered in rubber-like materials to be equilibrium configurations since the prefered configuration is a concentration of all the chains in one point. Physically, Van der Waals forces prevent the rubber chains from getting too close to one another. There are at least two ways to model the effect of such repulsive forces. One may directly incorporate this effect in the form of the pair potential by imposing that the stochastic network is a ground state of the energy. This amounts to chosing W such that $W(1) = 0$ and $W(r) > 0$ if $r \neq 1$. With such a property, one can deal with rather general energies and networks. Theorem 54 states our main result with this type of model.

Another way to deal with compressible effects consists in keeping unchanged the energy of a polymer chain and adding another energy term related to volumetric changes at the discrete level. From a modeling point of view, the difficulty of this approach comes from the definition of the volume in a discrete setting. A possible solution is provided with in Section 10.3.4 and relies on the use of regular Delaunay triangulations and piecewise affine interpolation.

The discrete to continuum derivation of incompressible materials is not easy to deal with mathematically. This issue is also partially addressed in Section 10.3.4. Imposing an *ad hoc* incompressibility term at the continuous scale is in contradiction with the principles of the present derivation: there may be an interplay at the mesoscopic scale between the different parts of the energy that is not taken into account by treating separately the polymer chain energies and the volumetric energy. Actually, this interplay is important: the macroscopic energy of real rubber-like materials may slightly depend on the second invariant of the strain gradient. This dependence can be reproduced neither by the eight-chain model, nor by the statistical approach since, by construction, they only depend on the first invariant (and possibly the third invariant if incompressibility is imposed at the continuous level). The latter dependence is mechanically interpreted as an effect of compressibility at the mesoscopic scale, which is automatically taken into account by the variational approach.

10.1.3 Stochastic networks and mesoscopic energies

As a starting point, we consider a particular case of the stochastic networks introduced by Blanc, Le Bris and Lions in [25, 26]. In what follows, $(\Omega, \mathcal{F}, P)$ denotes a given probability space.

Definition 38 *Let $\Lambda = \{x_i\}_{i\in\mathbb{Z}^d} \in \left(\mathbb{R}^d\right)^{\mathbb{Z}^d}$ be a set of points. We say that Λ is an admissible set of points if it satisfies the two following conditions:*

i. there exists $R > 0$ such that $\#\Lambda \cap B(x, R) > 0$ for all $x \in \mathbb{R}^d$;
ii. there exists $r > 0$ such that $d(x_i, \Lambda \setminus \{x_i\}) \geq r$ for all $i \in \mathbb{Z}^d$.

In particular, to each admissible set of points Λ one can associate a Delaunay triangulation $\mathcal{D}(\Lambda)$.

Definition 39 *A stochastic lattice $\mathcal{L} : \Omega \to \left(\mathbb{R}^d\right)^{\mathbb{Z}^d}$ is said to be admissible, if for P-almost every $\omega \in \Omega$, $\mathcal{L}(\omega)$ is an admissible set of points, and if the Delaunay triangulation is regular in the*

sense of the interpolation theory. Given a measure preserving group $\{\tau_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^{\mathbf{d}}}$ *acting on* Ω, $\mathcal{L}$ *is said to be stationary if*

$$\mathcal{L}(\tau_{\mathbf{k}}(\omega)) = \mathcal{L}(\omega) - \mathbf{k} \tag{10.3}$$

for all $\mathbf{k} \in \mathbb{Z}^{\mathbf{d}}$ *and ergodic if for any* $A \in \mathcal{F}$

$$(\tau_{\mathbf{k}}(A) = A) \quad \textit{implies} \quad P(A) = 0 \textit{ or } P(A) = 1.$$

Remark **37** *In two dimensions, any admissible set of points has a regular Delaunay triangulation. In three dimensions, the conditions i. and ii. are not enough to ensure the existence of a regular triangulation.*

To each realization ω of the stochastic lattice, we associate an energy which only depends on the realization.

Definition 40 *For all regular bounded open subset* A *of* $\mathbb{R}^d$*, and all* $u : \mathcal{L}(\omega) \to \mathbb{R}^n$*, we define the energy of the lattice by*

$$E(u, A)(\omega) = \sum_{x_i \in \mathcal{L}(\omega)\cap A} \sum_{\substack{x_j \neq x_i \in \mathcal{L}(\omega)\cap A \\ [x_i, x_j] \subset A}} J(x_j - x_i) f\left(\frac{u(x_j) - u(x_i)}{|x_j - x_i|}\right), \tag{10.4}$$

where J *and* f *are given functions that do not depend on* ω *and* A*.*

The precise forms of J and f will be prescribed in the following section. Let us define now the rescaled energy on a fixed domain.

Definition 41 *Let* D *be a fixed domain of* $\mathbb{R}^d$*. For all* $\epsilon > 0$*, and* $u : \epsilon\mathcal{L}(\omega) \to \mathbb{R}^n$*, we set*

$$E_\epsilon(u, D)(\omega) = \sum_{x_i \in \epsilon\mathcal{L}(\omega)\cap D} \epsilon^d \sum_{\substack{x_j \neq x_i \in \epsilon\mathcal{L}(\omega)\cap D \\ [x_i, x_j] \subset D}} J\left(\frac{x_j - x_i}{\epsilon}\right) f\left(\frac{u(x_j) - u(x_i)}{|x_j - x_i|}\right). \tag{10.5}$$

10.2 Derivation of a macroscopic model

10.2.1 Main result

Hypotheses 4 *We make the following assumptions on the energy (10.5)*

- *There exist* $p > 1$, $C_1, C_2 > 0$*, such that for all* $v \in \mathbb{R}^n$*,*

$$C_1|v|^p - 1 \leq f(v) \leq C_2|v|^p + 1 \tag{10.6}$$

- $J \geq 0$ *and* $\inf_{B(0,4R)} J(z) > 0$*,*
- $\int_{\mathbb{R}^d} J(z)dz < \infty$*.*

In particular, if $n = d$ and for all $w \in S^1(\mathbb{R}^d)$, $f(w) = 0$, then $u : x \mapsto x$ is a minimizer of the discrete energy and the lattice is a ground state of the system of interacting points. This may model in particular the ground state of a rubber-like polymer at the scale of the polymer chain.

The convergence result is expressed and proved in terms of Γ-convergence.

Theorem 54 *Let* D *be an open bounded subset of* $\mathbb{R}^d$*,* $\mathcal{L} : \Omega \to \left(\mathbb{R}^d\right)^{\mathbb{Z}^d}$ *be an admissible stochastic lattice with regular Delaunay triangulations, and let* f *and* J *satisfy Hypotheses 4. For* P*-almost all* $\omega \in \Omega$*,* $E_\epsilon(\cdot, D)(\omega)$ $\Gamma - L^p(D)$ *converges to* $E(\cdot, D)$ *defined on* $W^{1,p}(D)$ *by*

$$E(v,D) = \int_D W_{hom}(\nabla v(x))dx, \tag{10.7}$$

where W_{hom} is an homogeneous in space quasiconvex energy density satisfying a growth condition (10.6) of order $p > 1$, and given by the following asymptotic formula

$$W_{hom}(\xi) = \lim_{N\to\infty} \frac{1}{N^d} \int_\Omega \inf \big\{ E_1(u,(0,N)^d)(\omega), u(z) = \xi \cdot z \text{ for } z \in \mathcal{L}(\omega), \\ dist(z, \partial(0,N)^d) \le 2R \big\}\, dP(\omega), \tag{10.8}$$

In addition, infimum problems with prescribed boundary conditions also converge.

10.2.2 Sketch of the proof

In the present proof of Theorem 54, we need a piecewise affine interpolation. To this aim, we intoduce the following

Definition 42 *For all $\omega \in \Omega$, let $\mathcal{T} = \{T_i\}_{i\in\mathbb{Z}^d}$ be the set of the triangles of a regular Delaunay triangulation associated to $\epsilon\mathcal{L}(\omega)$. To each $u : \epsilon\mathcal{L}(\omega) \to \mathbb{R}^n$ we associate an admissible deformation (still denoted by u) defined as the continuous piecewise affine interpolation of u on $\mathcal{T}$.*

In particular, for all open bounded subset $A \subset \mathbb{R}^d$, and all $v \in \mathcal{C}_0^\infty(\mathbb{R}^d)$, the interpolations v_ϵ of v on the space of piecewise affine functions of $W^{1,\infty}(A)$ associated with $\epsilon\mathcal{L}(\omega)$ satisfy

$$\lim_{\epsilon\to 0} \|v_\epsilon - v\|_{W^{1,p}(A)} = 0,$$

for all $p \ge 1$.

In order to use these interpolation results, we need to consider regular triangulations. This is quite general in two dimensions, but not in three dimensions, for which this restriction on the triangulation has no simple characterization in terms of sets of points. The details of the following proof, as well as another approach to deal with more general three dimensional cases, will be given in [A2].

Individual compactness.

The first step of the proof consists in obtaining individual compactness. For all $\omega \in \Omega$, we prove that there exists a subsequence ϵ_n, such that $E_{\epsilon_n}(\cdot, D)(\omega)$ Γ-converges to some $E(\cdot, D)(\omega)$ on $W^{1,p}(D)$, for any open bounded regular subset $D \subset \mathbb{R}^d$. To this aim, one proceeds as in [4]. The estimates are however more delicate to get. Using a regular Delaunay triangulation, one can introduce nearest neighbors, use the lower bound on the pairwise interactions between them and obtain the coercivity of the Γ-liminf. Up to introducing a coarser periodic reference lattice, one can classify the interactions according to their range and re-use the results proved in [4] for the growth condition from above, and the subadditivity of the Γ-limsup. Using then a well-known compactness result of Γ-convergence, we obtain the existence of a Γ-limit up to extraction. This limit satisfies the De Giorgi-Letta criteria and can therefore be expressed in terms of an integral functional. The second part of the proof is devoted to show that $E(\cdot, D)(\omega)$ does not depend on ω and the associated energy density W_{hom} does not depend on the space variable.

Ergodicity of subadditive processes.

The result is achieved if we prove the existence of the limit (10.8) for the asymptotic formula. The proof relies on the subadditive ergodic theorem, and more precisely on the variant used in [59]. The energy $E_1(\cdot, D)$ is not subadditive since the interactions are not local. So is its infimum. One may however introduce a modified energy defined on the borelians of $\mathbb{R}^d$, which is subadditive and such that the asymptotic limits of type (10.8) coincide for both energies, the convergence of one implying the convergence of the other. Doing so, the ergodic theorem shows the existence of (10.8)

and therefore concludes the proof. Let us quickly define the modified energy in the case of a finite range $\mathbf{R}$ of interactions. Let $B \in \mathcal{B}(\mathbb{R}^d)$ be a bounded borelian set. We denote by $\overline{C}(B)$ the union of the finite number of closed unit cubes C centered in points of $\mathbb{Z}^d$ such that $C \cap B \neq \emptyset$, and by $C(B)$ the interior of $\overline{C}(B)$. For all $\alpha > 0$ and for all $\xi \in \mathcal{M}_d(\mathbb{R})$, we set

$$G^\xi(B)(\omega) = \inf\{E(u, C(B))(\omega), u(x_i) = \xi \cdot x \text{ if } \mathrm{d}(x_i, \partial C(B)) \leq \mathbf{R}\} + \alpha \,\mathrm{perim}(C(B)),$$

where perim denotes the perimeter (or $\mathcal{L}^{d-1}$ measure) of the set. For all ξ, there exists α big enough such that G^ξ is a subadditive set functional. In addition, for regular sets B,

$$\lim_{|B|\to\infty} \frac{1}{|B|} \alpha \,\mathrm{perim}(C(B)) = 0.$$

Therefore, [59, Prop. 1] implies the existence of

$$\lim_{|B|\to\infty} \frac{1}{|B|} G^\xi(B)(\omega)$$

and its independence upon ω, which concludes the proof.

10.3 Mechanical properties of the macroscopic model

This last section is dedicated to the properties of the continuous energy density obtained and to some rather direct generalizations of Theorem 54 to deal with realistic examples.

10.3.1 Objectivity

At the discrete level, due to its specific dependence upon finite differences, the energy of the discrete network is frame invariant for all $\omega \in \Omega$. This property is conserved by the discrete to continuum derivation, and the following theorem holds.

Theorem 55 *The energy density W_{hom} is frame invariant.*

10.3.2 Isotropy

Let us now suppose, as it is done in the statistical approach to discrete to continuum limits, that the network explores uniformly (in P) all the directions of the space $\mathbb{R}^d$. This assumption can be stated as follows

Definition 43 *Let Λ be an admissible set of points. For all $O \in SO_d(\mathbb{R})$, $O\Lambda$ is the admissible set of points defined by the rotation of Λ by O centered at the origin. An admissible stochastic network is said to be* ***isotropic*** *if it is rotation invariant in the following sense: for P-almost every $\omega \in \Omega$ and all $O \in SO_d(\mathbb{R})$, there exists $\omega_O \in \Omega$ such that $O\mathcal{L}(\omega) = \mathcal{L}(\omega_O)$ and if for all $B \in \mathcal{F}$ and $O \in SO_d(\mathbb{R})$,*

$$P(B) = P(B_O),$$

where $B_O = \{\omega_O, \omega \in B\}$.

The following theorem holds.

Theorem 56 *If the stochastic network is isotropic, then the energy density W_{hom} is isotropic.*

This result is a major difference with [4], where the homogenized energy density remembers the anisotropy of the underlying periodic network.

10.3.3 Behaviour at small strains

The behaviour of the obtained energy density at small strains is interesting since, close to the identity, any two times differentiable frame invariant energy density is equivalent to the Saint Venant-Kirchhoff energy density [52, Th. 4.5-1.].

Unfortunately, we are not able to prove the required regularity and the homogenized energy density is only known to be Lipschitz continuous (as a consequence of quasiconvexity and of the growth condition). This equivalence is only formal.

10.3.4 Compressibility issues

In this paragraph, we address two issues. First we show how to consider volumetric changes in a discrete setting, and second how to deal with incompressibility. The compressibility part of the energy is usually understood as a Van der Waals effect or as an influence of the solvant.

Given a realization of a stochastic network, we consider a regular Delaunay triangulation. Using Definition 42, for any deformation u of the network, one can consider the deformation of each of the triangle (or tetraedra) of the Delaunay triangulation and define the local Jacobian of the transformation by the piecewise constant function $\det(\nabla u)$, where u also denotes the admissible deformation associated to u. We thus add to the energy $E_\epsilon(\cdot, D)$ a volumetric part, obtaining

$$F_\epsilon(u, D)(\omega) = E_\epsilon(u, D) + \sum_{T \in \epsilon\mathcal{T}, T \subset D} |T| V(\det(\nabla u_{|T})), \tag{10.9}$$

for any admissible deformation u and open bounded regular set $D \subset \mathbb{R}^d$.

Remark **38** *It may be noticed that the Delaunay triangulation allows us to define nearest neighbors and to extend all our convergence results to the case of pair interaction potentials between nearest neighbors only.*

If $\mathcal{M}_d(\mathbb{R}) \ni \xi \mapsto V(\det \xi) \geq 0$ satisfies the growth condition (10.6) from above with the same p as J then the result of Theorem 54 still holds, as stated in the following theorem.

Theorem 57 *Let D be an open bounded subset of $\mathbb{R}^d$, $\mathcal{L} : \Omega \to \left(\mathbb{R}^d\right)^{\mathbb{Z}^d}$ be an admissible stochastic lattice with regular Delaunay triangulations, and let f and J satisfy Hypotheses 4. Let $V : \mathbb{R} \to \mathbb{R}^+$ be a continuous function such that $\mathcal{M}_d(\mathbb{R}) \ni \xi \mapsto V(\det \xi)$ satisfies a growth condition of order p from above. For all ω, let $\mathcal{T}(\omega)$ denote the regular Delaunay triangulation of $\mathcal{L}(\omega)$. The discrete energy is given by (10.9). For P-almost all $\omega \in \Omega$, $F_\epsilon(\cdot, D)(\omega)$ $\Gamma - L^p(D)$ converges to $F(\cdot, D)$ defined on $W^{1,p}(D)$ by*

$$F(v, D) = \int_D W_{hom}(\nabla v(x)) dx, \tag{10.10}$$

where W_{hom} is an homogeneous in space quasiconvex energy density satisfying a growth condition (10.6) of order p, and given by the following asymptotic formula

$$\begin{aligned} W_{hom}(\xi) = \lim_{N \to \infty} \frac{1}{N^d} \int_\Omega \inf \big\{ & F_1(u, (0, N)^d)(\omega), u(z) = \xi \cdot z \text{ for } z \in \mathcal{L}(\omega), \\ & dist(z, \partial(0, N)^d) \leq 2R \big\} \, dP(\omega), \end{aligned} \tag{10.11}$$

In addition, the energies being equi-coercive, the Γ-convergence implies the convergences of the infima and of the minimizers.

In more interesting cases, $\xi \mapsto V(\det \xi)$ does not satisfy (10.6) since the growth condition cannot model incompressible behaviours. From a mathematical point of view this is a very difficult task. Already for classical periodic homogenization in nonlinear elasticity, not much is known

concerning incompressibility issues. In the present case, one can however penalize incompressibility up to a given rate at the discrete level (in order to satisfy (10.6)), and apply Theorem 57. We can then recover incompressibility by passing to the limit again. The incompressible energy obtained is weakly lower semicontinuous and can be minimized. In what follows, we consider quasi-incompressible energies, as in the following theorem.

Theorem 58 *Let D be an open bounded subset of $\mathbb{R}^d$, $\mathcal{L} : \Omega \to \left(\mathbb{R}^d\right)^{\mathbb{Z}^d}$ be an admissible stochastic lattice with regular Delaunay triangulations, and let f and J satisfy Hypotheses 4. Let $V : \mathbb{R} \to \mathbb{R}^+ \cup \{+\infty\}$ be a convex function, continuous on $\mathbb{R}^+_*$, such that $V(w) = +\infty$ if $w < 0$ and $\lim_{w\to 0} V(w) = +\infty$, and such that $\xi \mapsto V(det\xi)$ satisfies a growth condition of order p from above if $det\xi \geq 1$. For any $\eta > 0$, we define $V^\eta : \mathbb{R} \to \mathbb{R}^+$ as follows:*

$$\begin{cases} V^\eta(w) = V(w) & \text{if } w > \eta \\ V^\eta(w) = V(\eta) & \text{if } w \leq \eta \end{cases}. \tag{10.12}$$

For all ω, let $\mathcal{T}(\omega)$ denote the regular Delaunay triangulation of $\mathcal{L}(\omega)$, and $F_{\epsilon,\eta}(\cdot, D)$ denote the energy associated to E_ϵ and V^η. For P-almost all $\omega \in \Omega$ and all $\eta > 0$, $F_{\epsilon,\eta}(\cdot, D)(\omega)$ $\Gamma(L^p)$-converges to $F_\eta(\cdot, D)$ defined on $W^{1,p}(D)$ by

$$F_\eta(v, D) = \int_D W_{hom,\eta}(\nabla v(x))dx, \tag{10.13}$$

where $W_{hom,\eta}$ is an homogeneous in space quasiconvex energy density satisfying a growth condition (10.6) of order p, and given by (10.11) with $F_{1,\eta}$ in place of F_1. In addition, $F_\eta(\cdot, D)$ $\Gamma(L^p)$-converges as $\eta \to 0$ to $F(\cdot, D)$ defined for all $v \in W^{1,p}(D)$ by

$$\begin{cases} F(v, D) = \displaystyle\int_D W_{hom}(\nabla v) & \text{if } v \in V \\ F(v, D) = +\infty & \text{if } v \notin V \end{cases}, \tag{10.14}$$

$V = \{v \in W^{1,p}(D), det(\nabla v) > 0 \quad a.e.\}$ and

$$W_{hom}(\xi) = \lim_{\eta\to 0} W_{hom,\eta}(\xi)$$

for all $\xi \in \mathcal{M}_d(\mathbb{R})$. In particular, F is sequentially lower-semicontinuous for the weak-topology of $W^{1,p}(D)$, W_{hom} is quasiconvex but does not satisfy a growth condition of type (10.6).

The first part of Theorem 58 is a direct application of Theorem 57 for $\eta > 0$. The last statement is a consequence of the four following arguments:

- The sequence of energy densities $W_{hom,\eta}$ is positive and increasing, therefore the energy $F_\eta(v, D)$ converges on $W^{1,p}(D)$ using Fatou's lemma. In addition it is bounded on V;
- The supremum of a family of lower-semicontinuous functionals being lower-semicontinuous, $F(\cdot, D)$ is lower-semicontinuous;
- By approximation and positiveness of the energy, F is infinite if $v \notin V$;
- Γ-convergence of monotone sequences is equivalent to pointwise convergence [37, Remark 2.12].

10.3.5 Physical setting

In this paragraph, we quickly explain how to relate the free energy of a polymeric chain $\tilde{W}_c$ to (10.5). Let us assume that any realization of the stochastic network describes interactions of nearest neighbours ($x_i - x_j$ being the end to end vector of the polymer chain). We further assume that a polymer chain is made of a given amount of segments of length l, and we set $r = l$ in Definition 38-*ii*. The maximum distance between two nearest neighbours is given by R, which is the maximum averaged length of the polymeric chains of the network. Assuming that the averaged length $|x_i - x_j|$ of a polymer chain is related to its number of segments N by $|x_i - x_j| = \sqrt{N}l$, the

number of segments of a polymer chain of averaged length $|x_i - x_j|$ is simply given by $N = \frac{|x_i - x_j|^2}{l^2}$. With the notation of the stochastic network and the modeling assumptions made above, one can write $\tilde{W}_c$ as

$$\tilde{W}_c(|u(x_i) - u(x_j)|, r) = \frac{1}{\beta}\frac{|x_i - x_j|^2}{r^2}\left(\frac{r|u(x_i) - u(x_j)|}{|x_i - x_j|^2}\theta\left(\frac{r|u(x_i) - u(x_j)|}{|x_i - x_j|^2}\right) + \log\frac{\theta\left(\frac{r|u(x_i)-u(x_j)|}{|x_i-x_j|^2}\right)}{\sinh\theta\left(\frac{r|u(x_i)-u(x_j)|}{|x_i-x_j|^2}\right)}\right) - \frac{c}{\beta}.$$

Inserting the first terms of the series expansion (10.1) up to the order $p - 1 > 0$ in $\tilde{W}_c$ we obtain an energy potential W_c^p. In order to perform the discrete to continuum derivation, we scale all the distances by ϵ and we set

$$E_\epsilon(u, D)(\omega) = \sum_{x_i \in \epsilon\mathcal{L}(\omega)\cap D} \epsilon^d \sum_{\substack{x_j \in \epsilon\mathcal{L}(\omega)\cap D \\ \{x_i, x_j\} \text{ n.n.}}} W_c^p(|u(x_i) - u(x_j)|, \epsilon r),$$

the energy of the fixed domain D. Due to the properties of the network, $R\epsilon \geq |x_i - x_j| \geq \epsilon r$ and the energy satisfies all the assumptions of Theorem 54. However the natural state is trivial and given by $u = 0$. In order to take into account the Van der Waals forces, it is necessary to add a volumetric term, as in (10.9). Applying then Theorems 57 and 58 we obtain a rigorous derivation of a microscopically based rubber-like energy density. In this case, it may be noticed that the realizations of the stochastic network are not natural states of the system. This may not be surprising since the Van der Waals forces prevent the polymer chains from being at rest. The polymer chains are 'pre-stressed' while the macroscopic rubber is at rest.

10.4 Numerical tests and further issues

10.4.1 Numerical validation of the model

In [27,31], Böl and Reese have considered a "finite element modelling" of rubber. Given a tetraedral mesh $\mathcal{T}_h$ of $A \subset \mathbb{R}^3$, they have defined an energy E_h on the $P1$-finite element space V_h associated with $\mathcal{T}_h$. For all $v_h \in V_h$, one can rewrite their energy as

$$E_h(v_h) = \int_A W_{comp}(\nabla v_h) + \sum_{T\in\mathcal{T}_h} |T| \sum_{e\in T} W_{chain}\left(\frac{|\nabla v_h \cdot e|}{e \cdot e}\right), \tag{10.15}$$

where $W_{comp}(\xi) = a\left((\det\xi)^2 - 2\log(\det\xi) - 1\right)$, $a > 0$, models the volumetric part of the energy, $\{e \in T\}$ denotes the set of edges of T, and W_{chain} is the free energy of a polymeric chain recalled in Section 10.1.1.

For h small enough, numerical tests have been carried out in [29] with prescribed boundary conditions on ∂A, such as uniaxial extension, compression, simple shear *etc.* The results have been compared to phenomelogical behaviours and are in good agreement with Ogden constitutive laws and experimental data, as illustrated on Figure 10.5.

The theoretical convergence properties of the model when $h \to 0$ have not been addressed so far.

Actually, up to replacing W_{comp} by W_{comp}^η as in (10.12), E_h is of the type considered in Theorem 57 for nearest neighbours interactions (see Remark 38). In particular, if one identifies $\mathcal{T}_h$ with $(\epsilon\mathcal{T})(\omega)_{|\epsilon=h} \cap A$ for some $\omega \in \Omega$, and if $\mathcal{T}$ satisfies the hypotheses of Theorem 54, then the finite element modelling of rubber introduced in [27] converges as $\epsilon \to 0$ to a hyperelastic model, whose properties are listed in Theorems 55-56-57.

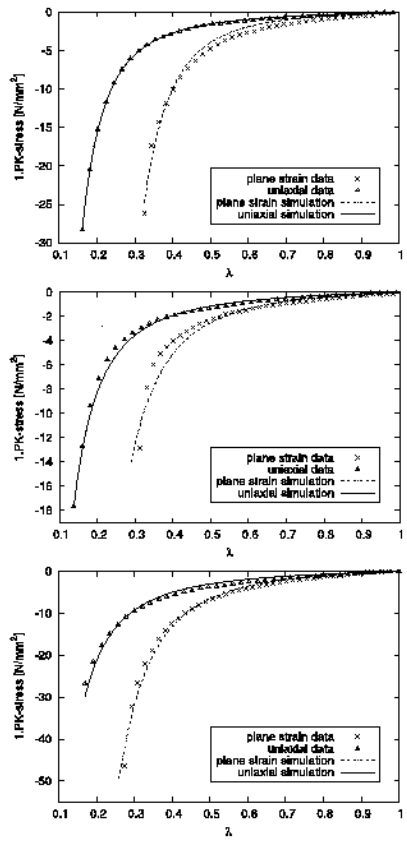

Fig. 10.5. Comparison between simulations and experiments: (a) silicon rubber, (b) gum rubber, (c) neoprene rubber, courtesy of M. Böl and S. Reese, extracted from [29]

Conversely, the numerical tests performed in [27] can be interpreted as the computations of numerical approximations of the homogenized energy density (10.11). These tests validate the finite element modelling and therefore acknowledge the assumptions made in Section 10.1 on the mesoscopic model, which is precisely our starting point.

To go further in the validation, one could investigate numerically the behaviour of the homogenized material. Combining the solution method described in Chapter 5 at the macroscopic level and the computation of E_h at the microscopic level introduced in [27], one could consider more interesting validation tests.

10.4.2 Comments and further developments

Discrete to continuum limits within a stochastic frame has been addressed first by Iosifescu, Licht and Michaille. In [110, 111] they have considered a discrete energy defined on a one-dimensional fixed periodic lattice with a stochastic pairwise interaction. This point of view is in a way a dual approach to the one adopted in the present work. The energy of stochastic lattices has been introduced by Blanc, Le Bris and Lions in [25, 26], and studied in terms of energies of quantum chemistry models. In particular, they have addressed the convergence of the ground state of the electronic cloud within the Born-Oppenheimer hypothesis.

Rubber-like materials do have more complex features than the ones obtained in the present work. These features mainly come from the polymer network level. The behaviour of the polymer chains are actually more complex. Some chains can break, some filler is usually introduced by

vulcanization, some phenomena like reptation may occur. Therefore, to reach more specific features of realistic materials, one needs to complexify the mesoscopic model, as it has been done by Bergström and Boyce in [23] starting from the eight-chain model, or by Böl and Reese in [28, 30]. In particular, one may think of the Muellins effect and more generally of the rupture of bonds. Actually, to account for the finite extensibility of the polymeric chains, one could introduce a probability of rupture, as it is done by Braides and Piatnitski in [43]. Easier to deal with mathematically than $W_{chain}(r_c) = \infty$ for $r_c > Nl$, this approach also seems closer to reality.

Part IV

Fluid-structure interaction problems

11

Domain decomposition based Newton methods for fluid-structure interaction problems

We review various fluid-structure algorithms based on domain decomposition techniques and we propose a new one. The standard methods used in fluid-structure interaction problems are generally "nonlinear on subdomains". We propose a scheme based on the principle "linearize first, then decompose". In other words we extend to fluid-structure problems domain decomposition techniques classically used in nonlinear elasticity.

11.1 Introduction

En guise d'introduction à cette partie, nous présentons rapidement les modèles : les équations de Navier-Stokes dans un domaine mobile, quelques spécificités numériques des modèles de coques élastiques et le principe du couplage de ces deux modèles. Cette section s'inspire des notes de cours de Fernández et Gerbeau sur l'interaction fluide-structure [79].

11.1.1 Equations de Navier-Stokes

Les équations de Navier-Stokes sont les équations qui régissent l'écoulement des fluides newtoniens, comme l'eau ou le sang dans les grosses artères. L'étude théorique d'existence et d'unicité des solutions de ces équations d'une part et les méthodes numériques d'autre part ne seront pas abordées ici. Ce paragraphe a pour but d'introduire le couplage fluide-structure en rappelant les équations de Navier-Stokes en formulation eulérienne et en formulation ALE (arbitraire Euler-Lagrange).

Soit $\Omega^{\mathrm{f}}(t)$ un domaine borné de $\mathbb{R}^3$. On considère un fluide newtonien incompressible sans force de volume. Les équations de Navier-Stokes incompressibles en formulation eulérienne conservative s'écrivent

$$\left\{\begin{aligned} \rho^{\mathrm{f}}\frac{\partial u}{\partial t} + \operatorname{div}\left(\rho^{\mathrm{f}} u\otimes u - \sigma^{\mathrm{f}}(u,p)\right) &= 0, \quad \text{dans} \quad \Omega^{\mathrm{f}}(t),\\ \operatorname{div} u &= 0, \quad \text{dans} \quad \Omega^{\mathrm{f}}(t). \end{aligned}\right. \tag{11.1}$$

Dans (11.1), ρ^{f} désigne la densité du fluide, u sa vitesse, p sa pression (multiplicateur de Lagrange de la contrainte d'incompressibilité), et $\sigma^{\mathrm{f}} = \sigma^{\mathrm{f}}(u,p) = -pI + 2\mu\epsilon(u) = -pI + \mu\left(\nabla u + \nabla u^T\right)$ est le tenseur des contraintes. Ce système d'équations doit être complété par des conditions aux limites sur u ou sur la contrainte normale $\sigma^{\mathrm{f}}\cdot\mathbf{n}$, par exemple.

Quand on travaille sur des domaines $\Omega^{\mathrm{f}}(t)$ mobiles, il est commode d'utiliser les dérivées en temps ALE (les dérivées en temps eulériennes peuvent être difficilement calculables dans des maillages mobiles) définies pour une quantité q par :

$$\frac{\partial q}{\partial t}_{|\hat{x}}(x_t,t) = \lim_{\delta t\to 0}\frac{q(x_{t+\delta t},t+\delta t) - q(x,t)}{\delta t}$$

où $t \mapsto x_t$ est un paramétrage en temps du domaine fluide mobile. En particulier, on note $\widehat{\mathcal{A}} : \hat{\Omega}^{\mathrm{f}} \times \mathbb{R}^+ \to \mathbb{R}^3, (\hat{x}, t) \mapsto x = \widehat{\mathcal{A}}(\hat{x}, t)$ un paramétrage tel que $\widehat{\mathcal{A}}(\hat{\Omega}^{\mathrm{f}}, t) = \Omega^{\mathrm{f}}(t)$ pour tout $t \geq 0$. La vitesse du volume de contrôle (ou vitesse ALE) est définie par

$$\hat{w}(\hat{x}, t) = \frac{\partial \widehat{\mathcal{A}}}{\partial t}(\hat{x}, t), \qquad \forall \hat{x} \in \hat{\Omega}^{\mathrm{f}}.$$

Un tel volume de contrôle est illustré Figure 11.1 pour le problème de l'interaction fluide-structure dans les vaisseaux sanguins.

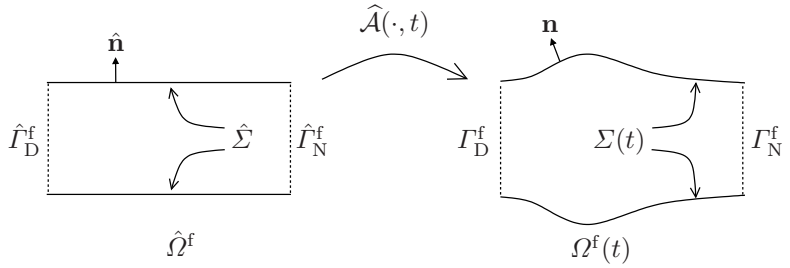

Fig. 11.1. Description du volume de contrôle via $\widehat{\mathcal{A}}$

La dérivée ALE s'écrit

$$\frac{\partial q}{\partial t}_{|\hat{x}} = w \cdot \nabla q + \frac{\partial q}{\partial t}.$$

On retrouve la dérivée eulérienne dès que $w = 0$, ou la description lagrangienne pour $w = u$. Les équations de Navier-Stokes en formulation ALE conservative sont données par

$$\begin{cases} \dfrac{\rho^{\mathrm{f}}}{J_{\widehat{\mathcal{A}}}} \dfrac{\partial J_{\widehat{\mathcal{A}}} u}{\partial t}_{|\hat{x}} + \operatorname{div}\left(\rho^{\mathrm{f}} u \otimes (u - w) - \sigma^{\mathrm{f}}(u, p)\right) = 0, & \text{dans} \quad \Omega^{\mathrm{f}}(t), \\ \operatorname{div} u = 0, & \text{dans} \quad \Omega^{\mathrm{f}}(t), \end{cases} \tag{11.2}$$

où $J_{\widehat{\mathcal{A}}} = \det(\nabla_{\hat{x}} \widehat{\mathcal{A}})$ (strictement positif par hypothèse : $\widehat{\mathcal{A}}$ est supposé injectif).

11.1.2 Théorie des coques

Dans ce chapitre, nous utiliserons une forme spéciale des équations de l'élasticité : un problème posé sur une sous-variété de codimension 1. Il s'agit de la recherche d'une surface d'énergie minimale dans un espace ambiant à trois dimensions. Cette théorie est appelée la théorie des coques.

Des travaux récents [89] ont clarifié rigoureusement la hiérarchie des modèles de coques par leur dérivarion à partir de l'élasticité tridimensionnelle. Les théories des coques peuvent être vues (dans de nombreux cas) comme la limite au sens de la Γ-convergence de l'élasticité tridimensionnelle quand une des dimensions du système tridimensionnel tend vers zéro (voir également [120, 122, 156, 157]).

La description précise des éléments finis de coque dépasse largement le cadre de cette introduction. On se contente ici de donner un aperçu très bref et très incomplet des enjeux ou difficultés dans la conception d'éléments finis de coque.

Tout d'abord, la cinématique d'une surface est bien plus complexe à décrire et fait intervenir des notations et concepts de géométrie différentielle (bases covariantes et contravariances, tenseurs métriques). Très grossièrement, on peut dire qu'une coque est constituée d'une surface moyenne (définie par une normale), d'une épaisseur et d'une hypothèse géométrique sur le déplacement

transverse dans l'épaisseur (constant, linéaire, quadratique etc.) et la condition de compatibilité associée sur la contrainte (un déplacement transverse constant implique un gradient de déformation transverse nul et donc un terme de contrainte nul $\sigma_{33} = 0$). Les deux comportements limites des coques minces sont le comportement membranaire (pas de variation de la normale) et le comportement en flexion pure (pas de déplacement dans le plan tangent). Il est difficile de concevoir un élément fini aussi efficace en membrane qu'en flexion pures. L'un des éléments finis de coque ne présentant pas de vérouillage numérique en flexion et cependant précis en membrane est un élément quadrangulaire (*i.d.* MITC4, $Q1$, ou MITC9, $Q2$, voir [17]). La conception d'un élément triangulaire avec les mêmes propriétés de convergence est une des directions de recherche actuelles [158]).

D'un point de vue complètement différent, les degrés de liberté associés aux coques minces peuvent paraître inappropriés à certaines modélisations mécaniques : les degrés sont portés sur la surface moyenne alors que les structures minces sont certes minces mais d'épaisseur non nulle. Le couplage cinématique avec le milieu environnant devient donc difficile, ou tout au moins "peu" précis. Cette problématique est illustrée par l'exemple de renforcement d'un pneu par une feuille métallique considéré dans [49] ou l'exemple de l'interaction fluide-structure. Dans ce cas, il paraît plus satisfaisant d'avoir une fine couche tridimensionnelle et de pouvoir considérer tous les degrés de liberté en déplacement sur chacune des surfaces. Avec une telle modélisation, il faudrait mailler très finement la couche ou risquer un vérouillage numérique important dû essentiellement au comportement en flexion. Tout comme l'approximation numérique de la quasi-incompressibilité emprunte la formulation développée pour les matériaux incompressibles, l'approximation numérique d'une couche fine est grandement améliorée en considérant le modèle de coque sous-jacent, avec une hypothèse géométrique sur le déplacement transverse (par exemple quadratique [50]) et surtout le traitement adéquat (par réinterpolation) des verouillages numériques (en membrane et flexion, mais aussi en *pinching* [51]). A partir des degrés de liberté des éléments finis tridimensionnels (uniquement $Q2$ isoparamétriques, HEXA3QCC dans [177]), il faut alors reconstruire les normales pour réinterpoler correctement selon la méthode MITC (*mixed interpolation of tensorial components*) [17, Sec. 7.2]. L'utilisation des coques épaisses (dites 3D) rend alors possible l'emploi de toutes les lois de comportement tridimensionnelles et plus seulement les lois de comportement de coque (qui imposent des relations de compatibilité entre la cinématique et la forme de la densité d'énergie) et un couplage clair avec des structures tridmensionnelles.

11.1.3 Couplage fluide-structure

Le couplage fluide-structure modélise l'interaction d'un fluide et d'un solide. Son champ d'application couvre aussi bien le mouvement d'un solide dans un liquide que l'écoulement d'un liquide dans une tuyère élastique. Dans le cas considéré par la suite, le problème mathématique consiste à coupler les équations de Navier-Stokes avec les équations de l'élastodynamique tridimensionnelle ou l'élastodynamique des coques.

Les résultats d'existence de solutions dans les cadres généraux et le sens mathématique du couplage sont des problèmes ouverts qui ne seront pas développés ici. Les résultats d'existence concernent soit l'aspect dynamique pour des problèmes linéaires, soit l'aspect statique en petites déformations pour les problèmes non linéaires. La problématique abordée dans cette thèse concerne uniquement les méthodes numériques de couplage de modèles.

Une présentation commode des équations du couplage consiste à écrire les équations de Navier-Stokes en formulation ALE (11.2) et les équations de l'élastodynamique en formulation lagrangienne (1.4). Le fluide et le solide interagissent à leur frontière commune $\hat{\Sigma}$ (en configuration de référence). L'interface est à l'équilibre en configuration déformée $\Sigma(t)$ si les vitesses (normales en particulier) du fluide et du solide sont les mêmes à l'interface et si les composantes normales des contraintes de Cauchy se compensent exactement. Le domaine de contrôle du fluide Ω^{f} se déforme selon le déplacement du solide $\hat{d}^{\mathrm{s}}$. La vitesse du paramétrage du domaine est donnée par la dérivée temporelle d'un opérateur d'extension. Les seconds membres sont donnés dans la configuration

déformée, ce sont des forces vives, ce qui nécessite leur transport dans la configuration de référence pour le solide. Les différents domaines sont schématisés Figure 11.2.

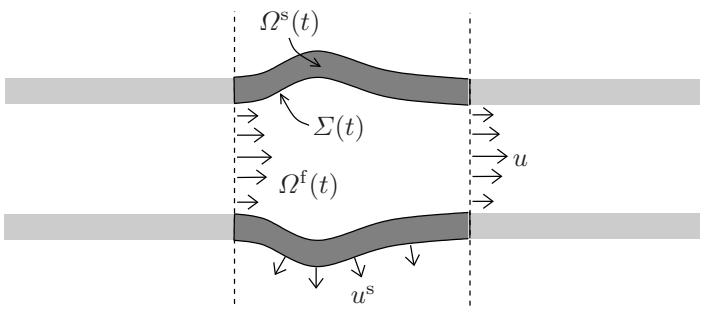

Fig. 11.2. Configuration géométrique (section 2D)

Le système d'équations est donné Section 11.3.

11.1.4 Aim of the study

In this chapter we review various numerical methods to treat the interaction between an incompressible fluid and an elastic structure, and we propose a new approach based on a Newton algorithm and domain decomposition methods.

Fluid-structure algorithms are too numerous to be reviewed exhaustively. A classification of the various approaches is not obvious either. To begin with, we can consider two groups of methods: the "strongly coupled" and the "loosely coupled" schemes. This distinction is quite clear since it corresponds to a precise property: those schemes which can ensure a well-balanced energy transfer between the fluid and the structure can be called "strongly coupled", the other ones are "loosely coupled". All the methods presented in this study are strongly coupled. Loosely coupled schemes, which are very powerful in many applications but can be unstable in others, are not considered here. We refer for example to [77, 162] for explicit coupling schemes and to [80, 81] for a semi-implicit coupling scheme.

We can then distinguish "monolithic" and "partitioned" schemes. For example, an *ad hoc* solver whose purpose is to solve simultaneously the fluid and the structure typically leads to a monolithic scheme (see for example [18,66,101,109,167,173,180]). On the other hand, coupling one fluid solver and one structure solver as black boxes clearly yields a partitioned scheme. Such a partitioned scheme can be strongly coupled as soon as sub-iterations are performed at each time step. The number of subiterations being very large in some application, acceleration techniques have been investigated in several articles: for example Le Tallec and Mouro [128] propose a steepest descent approach, Mok, Wall and Ramm [140] propose an Aitken acceleration which is based on the two previously computed solutions, and Vierendeels [178] a least-square method which uses several previously computed solutions.

It is well-known, in particular since the work by Le Tallec and Mouro [128] and more recently by Deparis *et al.* [64, 65], that fluid-structure problems can be tackled with domain decomposition approaches. Indeed, a fluid-structure problem can be viewed as a general continuum mechanics problem set on one domain which is split into a fluid part and a structure part. The fluid-structure coupling conditions then appear as the transmission conditions which ensure that the solution of the global problem is obtained by "sticking" the two sub-problem solutions. This point of view has been adopted in various studies, either with the so-called "Dirichlet-Neumann" algorithms (see for example [78, 91, 138]) or with "Neumann-Neumann" algorithms ([64, 65]).

All these methods have been devised following the rule "apply domain decomposition to the nonlinear global problem and then solve on each subdomain the nonlinear problems". On the contrary, in other fields – for example nonlinear elasticity [124] – domain decomposition is usually

applied with the rule "linearize first, then solve the tangent problem using domain decomposition". The purpose of this chapter is to propose a fluid-structure algorithm based on the last rule. The resulting algorithm can be viewed as a monolithic scheme in the sense that we apply a Newton algorithm to the global fluid-structure problem. But it is more conform to the practical implementation to consider it as a partitioned scheme since the fluid and the structure are solved with two different solvers, with their own schemes, and can be run in parallel. Contrarily to the methods following the first rule, these solvers are only used to solve the tangent problems and to evaluate nonlinear residuals. The use of two different solvers has well-known advantages (reusability of existing codes, flexible choice of the numerical methods adapted to each sub-problem, *etc.*). Compared to monolithic schemes presented in the literature [101, 109, 167], our approach may have another advantage: the use of domain decomposition methods to solve the tangent problem is expected to be more efficient than direct methods or iterative methods based on block-preconditioners.

In Section 11.2 we review some standard approaches to solve fluid-structure interaction problems, in particular those based on domain decomposition arguments, that use the so-called Steklov-Poincaré operators. In Section 11.3 we recall the fluid and solid models and we set the main notations. The time scheme is presented in Section 11.4. In Section 11.5 the new algorithm is introduced. It is difficult to anticipate the advantages in terms of efficiency of the proposed approach. Nevertheless we propose in Section 11.5.3 a simplified complexity analysis whose conclusion may be sum up as follows: the more expensive the structure problem an the more nonlinear the fluid are (let think of the Navier-Stokes equations but also of complex models for the fluid), the more competitive this new formulation is expected to be. In Sections 11.6 and 11.7 we present the 3D shell element used to model thin 3D structures, and report on some preliminary numerical results. More extensive simulations and comparisons with existing methods will be proposed in a forthcoming work [A2].

11.2 Classical solution methods

In this section we briefly review the existing algorithms for the numerical solution of the nonlinear system arising in the time discretization of the fluid-structure problem with an implicit coupling scheme. These methods are typically based on the application of a particular nonlinear iterative method to three different formulations of the nonlinear coupled system.

In general, the time discretization of a fluid-structure problem with an implicit coupling scheme leads to a coupled nonlinear problem of type: Find the interface displacement $\boldsymbol{\gamma}$, the fluid state $\boldsymbol{x}_f$ and the solid state $\boldsymbol{x}_s$ such that

$$\text{Formulation (I):} \qquad \begin{cases} \mathcal{F}\left(\boldsymbol{x}_f, \boldsymbol{\gamma}\right) = 0, \\ \mathcal{S}\left(\boldsymbol{x}_s, \boldsymbol{\gamma}\right) = 0, \\ \mathcal{I}\left(\boldsymbol{x}_f, \boldsymbol{x}_s\right) = 0. \end{cases} \tag{11.3}$$

Equations of $(11.3)_1$ and $(11.3)_2$ ensure the equilibrium of momentum when the fluid and the solid are subjected to an interface displacement $\boldsymbol{\gamma}$, whereas the last equation enforces the equilibrium of mechanical stresses at the interface.

Problem (11.3) can be reformulated in terms of $\boldsymbol{\gamma}$ by eliminating the fluid and solid unknowns $\boldsymbol{x}_f, \boldsymbol{x}_s$. This yields to the so-called Steklov-Poincaré formulation: Find the interface displacement $\boldsymbol{\gamma}$ such that,

$$\text{Formulation (II):} \qquad S_f(\boldsymbol{\gamma}) + S_s(\boldsymbol{\gamma}) = 0. \tag{11.4}$$

Here, S_f and S_s stand for the fluid and solid Steklov-Poincaré operators which can be defined as follows: for a given interface displacement $\boldsymbol{\gamma}$, $S_f(\boldsymbol{\gamma})$ gives the stress exerted by the fluid on the interface, and analogously for S_s. All these notations will be made precise below. In section 11.4.2, we shall describe the link between (11.3) and (11.4).

Finally, the composition of the inverse operator S_s^{-1} with (11.4) gives rise to the so-called Dirichlet-to-Neumann formulation:

$$\text{Formulation (III):} \qquad S_s^{-1}\big(-S_f(\boldsymbol{\gamma})\big) - \boldsymbol{\gamma} = 0. \tag{11.5}$$

Formally speaking, Formulations (II) and (III) are similar. Nevertheless, we prefer to distinguish them since they correspond to different approaches in the literature. The denominations "Dirichlet-Neumann formulation" and "Steklov-Poincaré formulation" are purely conventional (both of them clearly involve Steklov-Poincaré operators).

The three following paragraphs address a brief state-of-the-art on the iterative methods for the numerical solution of (11.3), (11.4) and (11.5).

11.2.1 Monolithic formulation

A common approach in the numerical solution of nonlinear systems, arising in implicit coupling, consists in applying a Newton based algorithm to the global formulation (11.3). This requires the repeated solution of a tangent (or approximated tangent) problem with the following block structure:

$$\begin{bmatrix} \mathrm{D}_{\boldsymbol{x}_f}\mathcal{F}(\boldsymbol{x}_f,\boldsymbol{\gamma}) & 0 & \mathrm{D}_{\gamma}\mathcal{F}(\boldsymbol{x}_f,\boldsymbol{\gamma}) \\ 0 & \mathrm{D}_{\boldsymbol{x}_s}\mathcal{S}(\boldsymbol{x}_s,\boldsymbol{\gamma}) & \mathrm{D}_{\gamma}\mathcal{S}(\boldsymbol{x}_s,\boldsymbol{\gamma}) \\ \mathrm{D}_{\boldsymbol{x}_f}\mathcal{I}(\boldsymbol{x}_f,\boldsymbol{x}_s) & \mathrm{D}_{\boldsymbol{x}_s}\mathcal{I}(\boldsymbol{x}_f,\boldsymbol{x}_s) & 0 \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x}_f \\ \delta\boldsymbol{x}_s \\ \delta\gamma \end{bmatrix} = - \begin{bmatrix} \mathcal{F}(\boldsymbol{x}_f,\boldsymbol{\gamma}) \\ \mathcal{S}(\boldsymbol{x}_s,\boldsymbol{\gamma}) \\ \mathcal{I}(\boldsymbol{x}_f,\boldsymbol{x}_s) \end{bmatrix}. \tag{11.6}$$

Newton algorithms based on the numerical solution of (11.6) in a *monolithic* fashion, *i.e.* using global direct or iterative methods, have been reported in [18,66,101,173,180]. It is worth noticing that such a monolithic approach makes difficult the use of separate solvers for the fluid and structure sub-problems. Alternatively, system (11.6) can be solved in a *partitioned* manner through a block-Gauss elimination of $\delta\boldsymbol{x}_f$, which leads to the so called block-Newton methods [82,83].

11.2.2 Dirichlet to Neumann formulations

Formulation (III) reduces problem (11.3) to the determination of a fixed point of the *Dirichlet-to-Neumann* operator $\boldsymbol{\gamma} \mapsto S_s^{-1}\big(-S_f(\boldsymbol{\gamma})\big)$. This motivates the use of fixed-point based methods [128,139,140,152]:

$$\boldsymbol{\gamma}^{k+1} = \omega^k S_s^{-1}\big(-S_f(\boldsymbol{\gamma}^k)\big) + (1-\omega^k)\boldsymbol{\gamma}^k, \tag{11.7}$$

with ω^k a given relaxation parameter which is chosen in order to enhance convergence [63,139,140]. Alternatively, one can use Newton based methods [84, 91] for a fast convergence towards the solution of (11.5). This requires the solution of a tangent problem of the type

$$(J(\boldsymbol{\gamma}^k) - \boldsymbol{I})\delta\gamma = -\big(S_s^{-1}\left(-S_f(\boldsymbol{\gamma}^k)\right) - \boldsymbol{\gamma}^k\big), \tag{11.8}$$

where $J(\boldsymbol{\gamma})$ stands for the Jacobian, or approximated Jacobian [91], of the composed operator $\boldsymbol{\gamma} \mapsto S_s^{-1}\big(-S_f(\boldsymbol{\gamma})\big)$. It is worth noticing that exact Jacobian computations require shape derivative calculus for the fluid [84]. Let us also stress the fact that these methods are naturally partitioned.

11.2.3 Symmetric Steklov-Poincaré formulation

The Dirichlet-Neumann formulations share a common feature: their implementation is purely sequential. The Steklov-Poincaré formulation (11.4) may allow to set up parallel algorithms to solve the interface equation.

Following the presentation of Deparis *et al.* [64], the nonlinear problem (11.4) can be solved through nonlinear Richardson iterations:

$$P(\boldsymbol{\gamma}^{k+1} - \boldsymbol{\gamma}^k) = \omega^k(-S_f(\boldsymbol{\gamma}^k) - S_s(\boldsymbol{\gamma}^k)), \tag{11.9}$$

for an appropriate choice of the preconditioner P, namely

$$P_k^{-1} = \alpha^k \left[S_f'(\boldsymbol{\gamma}^k)\right]^{-1} + (1-\alpha^k)\left[S_s'(\boldsymbol{\gamma}^k)\right]^{-1}, \tag{11.10}$$

where $\boldsymbol{\lambda} \mapsto S_f'(\beta)\cdot\boldsymbol{\lambda}$ is the differential of S_f at β, and $\left[S_f'(\beta)\right]^{-1}$ its inverse. This choice generalizes the standard preconditioners of linear domain decomposition methods (for which $S' = S$). If α_k is $0, 1$ or 0.5 we retrieve respectively Dirichlet-Neumann, Neumann-Dirichlet or Neumann-Neumann preconditioners. On the other hand, since equation (11.4) is nonlinear, one can apply a Newton method,

$$\big(S_f'(\boldsymbol{\gamma}^k) + S_s'(\boldsymbol{\gamma}^k)\big)(\boldsymbol{\gamma}^{k+1} - \boldsymbol{\gamma}^k) = -S_f(\boldsymbol{\gamma}^k) - S_s(\boldsymbol{\gamma}^k). \tag{11.11}$$

which corresponds to the nonlinear Richardson iteration (11.9) preconditioned with $P_k = S_f'(\boldsymbol{\gamma}^k) + S_s'(\boldsymbol{\gamma}^k)$. This linear equation can be solved, for example, by a GMRES algorithm, with or without preconditioning. For instance, in [64] the authors propose to use the preconditioners (11.10).

The Newton method applied to the Dirichlet-Neumann formulation is not equivalent to the Newton method applied to the Steklov formulation, since the roles played by the fluid and by the structure are not symmetric in the first approach whereas they are in the second. After linearization, one cannot compose (11.8) with S_s to retrieve (11.11). Finally (11.10) is not equivalent to (11.11) since in general $(A + B)^{-1} \neq A^{-1} + B^{-1}$.

The advantage of formulation (II) compared to formulation (III) is that the fluid and the structure sub-problems can be solved simultaneously and independently for the residual computation (right-hand sides of (11.9)) and the application of the preconditioner (S_f' and S_s') as soon as $\alpha \notin \{0, 1\}$. However, as we shall see in section 11.5.3, a simplified complexity analysis shows that the overall computational costs of both methods might be of the same order, for instance, whenever the cost of the fluid sub-problems solution is cheaper.

The formulations recalled in Sections 11.2.2 and 11.2.3 are first based on the coupling conditions, giving rise to a nonlinear equation on the interface, which involves nonlinear sub-problems. The algorithm we introduce in Section 11.5 first treats the nonlinearity of the whole problem through a Newton method, and uses a Steklov-Poincaré formulation on the tangent problems.

11.3 Mechanical setting

Let $\widehat{\Omega} = \widehat{\Omega}_f \cup \widehat{\Omega}_s$ be a reference configuration of the system, see Figure 11.3. We introduce the motion of the solid medium

$$\widehat{\boldsymbol{\varphi}}_s : \widehat{\Omega}_s \times \mathbb{R}^+ \longrightarrow \mathbb{R}^3.$$

The current configuration of the structure is then denoted by $\Omega_s(t) = \boldsymbol{\varphi}_s(\widehat{\Omega}_s, t)$. We introduce the deformation gradient $\widehat{\boldsymbol{F}}_s(\widehat{\boldsymbol{x}}, t) \stackrel{\text{def}}{=} \boldsymbol{\nabla}_{\widehat{\boldsymbol{x}}}\boldsymbol{\varphi}_s(\widehat{\boldsymbol{x}}, t)$, and its determinant $\widehat{J}_s(\widehat{\boldsymbol{x}}, t) \stackrel{\text{def}}{=} \det \widehat{\boldsymbol{F}}_s(\widehat{\boldsymbol{x}}, t)$. The displacement of the solid domain is given by $\widehat{\boldsymbol{d}}_s(\widehat{\boldsymbol{x}}, t) \stackrel{\text{def}}{=} \widehat{\boldsymbol{\varphi}}_s(\widehat{\boldsymbol{x}}, t) - \widehat{\boldsymbol{x}}$. The fluid domain $\Omega_f(t)$ is parametrized by the Arbitrary Lagrangian Eulerian ALE mapping (see [67], for instance),

$$\widehat{\mathcal{A}} : \widehat{\Omega}_f \times \mathbb{R}^+ \longrightarrow \mathbb{R}^3,$$

such that $\Omega_f(t) = \widehat{\mathcal{A}}(\widehat{\Omega}_f, t)$. In the sequel we will use the notation $\widehat{\mathcal{A}}_t \stackrel{\text{def}}{=} \widehat{\mathcal{A}}(\cdot, t)$ and the superscript $\widehat{\ }$ will be related to fields defined on the reference configuration $\widehat{\Omega}_f$ or $\widehat{\Omega}_s$. In addition, for a given Eulerian fluid quantity q (*i.e.* defined in $\Omega_f(t)$ for $t > 0$) we will denote its ALE description by $\widehat{q}$, as a field defined in $\widehat{\Omega}_f \times \mathbb{R}^+$ as

$$\widehat{q}(\widehat{\boldsymbol{x}}, t) = q\big(\widehat{\mathcal{A}}_t(\boldsymbol{x}), t\big), \quad \forall \boldsymbol{x} \in \widehat{\Omega}_f. \tag{11.12}$$

We introduce the deformation gradient of the fluid domain $\widehat{\boldsymbol{F}}_f(\widehat{\boldsymbol{x}}, t) \stackrel{\text{def}}{=} \boldsymbol{\nabla}_{\widehat{x}}\widehat{\mathcal{A}}(\widehat{\boldsymbol{x}}, t)$, and its determinant $\widehat{J}_f(\widehat{\boldsymbol{x}}, t) \stackrel{\text{def}}{=} \det \widehat{\boldsymbol{F}}_f(\widehat{\boldsymbol{x}}, t)$. The displacement of the fluid domain is given by $\widehat{\boldsymbol{d}}_f(\widehat{\boldsymbol{x}}, t) \stackrel{\text{def}}{=} \widehat{\mathcal{A}}(\widehat{\boldsymbol{x}}, t) - \widehat{\boldsymbol{x}}$ and its velocity by

$$\widehat{\boldsymbol{w}} \stackrel{\text{def}}{=} \frac{\partial \widehat{\mathcal{A}}}{\partial t}.$$

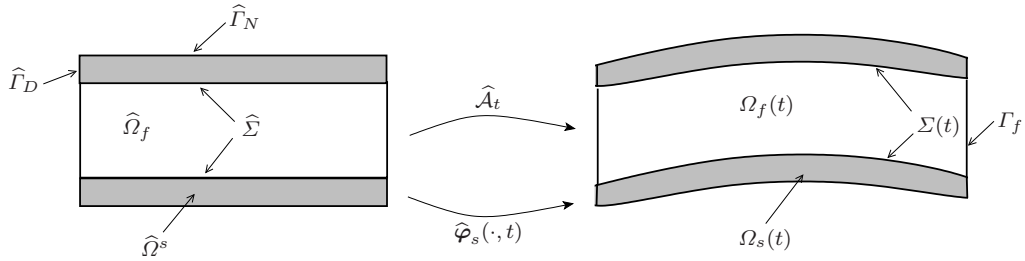

Fig. 11.3. Parametrization of the domains $\Omega_f(t)$ and $\Omega_s(t)$.

The fluid-structure interface, namely $\partial\Omega_f(t) \cap \partial\Omega_s(t)$ is denoted by $\Sigma(t)$, and $\Gamma_f = \partial\Omega_f(t) \backslash \Sigma(t)$ stands for the portion of the fluid boundary that is not shared with the boundary of the structure. The surface Γ_f is assumed to be independent of t. The boundary $\partial\widehat{\Omega}_s$ of the reference configuration for the structure is divided into three disjoint parts $\widehat{\Gamma}_D$, $\widehat{\Gamma}_N$ and $\widehat{\Sigma}$, with $\Sigma(t) = \widehat{\mathcal{A}}_t(\widehat{\Sigma})$. We denote by $\boldsymbol{n}$ the outward unit normal on the fluid boundary in the current configuration, and by $\widehat{\boldsymbol{n}}_s$ the outward unit normal on the reference structure boundary.

11.3.1 The coupled problem

We consider a Newtonian viscous, incompressible fluid with density ρ_f and dynamic viscosity μ. Its state is described by its Eulerian velocity $\boldsymbol{u}$ and pressure p. The constitutive law for the Cauchy stress tensor is given by the following expression:

$$\boldsymbol{\sigma}(\boldsymbol{u}, p) = -p\boldsymbol{I} + 2\mu\boldsymbol{\epsilon}(\boldsymbol{u}),$$

with $\boldsymbol{\epsilon}(\boldsymbol{u}) = \left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^{\mathrm{T}}\right]/2$. In absence of body forces, these unknowns satisfy the incompressible Navier-Stokes equations in an ALE formulation:

$$\begin{cases} \rho_f \dfrac{\partial \boldsymbol{u}}{\partial t}\Big|_{\widehat{\boldsymbol{x}}} + \rho_f(\boldsymbol{u} - \boldsymbol{w}) \cdot \boldsymbol{\nabla}\boldsymbol{u} - \operatorname{div}\big(2\mu\boldsymbol{\epsilon}(\boldsymbol{u})\big) + \boldsymbol{\nabla}p = 0, & \text{in} \quad \Omega_f(t), \\ \operatorname{div}\boldsymbol{u} = 0, & \text{in} \quad \Omega_f(t), \\ \boldsymbol{\sigma}(\boldsymbol{u}, p) \cdot \boldsymbol{n} = \mathbf{g}, & \text{on} \quad \Gamma_f, \end{cases} \tag{11.13}$$

where $\dfrac{\partial}{\partial t}\Big|_{\widehat{\boldsymbol{x}}}$ stands for the ALE time derivative, $\boldsymbol{w} \stackrel{\text{def}}{=} \widehat{\boldsymbol{w}} \circ \widehat{\mathcal{A}}_t^{-1}$, and $\mathbf{g}$ a given density of surface force.

The structure is supposed to be hyperelastic under large displacements and deformations. Its density is denoted by ρ_s. Its state is described by its displacement $\widehat{\boldsymbol{d}}_s$ and its first Piola-Kirchoff stress tensor $\widehat{\boldsymbol{T}}$. The latter is related to $\widehat{\boldsymbol{d}}_s$ as the gradient of an internal stored energy function $\mathcal{W}(\widehat{\boldsymbol{F}}_s)$. The choice of the internal stored energy will depend on the problem under consideration and will not change the setting of the fluid-structure problem. Assuming that the structure is clamped on Γ_D and under no body and surface forces, these unknowns are driven by the following elastodynamic equations

$$\begin{cases} \widehat{J}_s\rho_s \dfrac{\partial^2 \widehat{\boldsymbol{d}}_s}{\partial t^2} - \operatorname{div}_{\widehat{\boldsymbol{x}}}\widehat{\boldsymbol{T}} = \mathbf{0}, & \text{in} \quad \widehat{\Omega}_s, \\ \widehat{\boldsymbol{d}} = \mathbf{0}, & \text{on} \quad \widehat{\Gamma}_D, \\ \widehat{\boldsymbol{T}} \cdot \widehat{\boldsymbol{n}}_s = 0, & \text{on} \quad \widehat{\Gamma}_N. \end{cases} \tag{11.14}$$

The coupling between the solid and the fluid, namely equations (11.13) and (11.14), is realized through standard boundary conditions at the fluid-structure interface $\Sigma(t)$ that ensure the balance of the mechanical energy over the whole domain. This is achieved by imposing three interface conditions:

- A geometrical condition enforcing the matching between $\boldsymbol{\varphi}_s$ and $\widehat{\mathcal{A}}$ on the interface

$$\widehat{\boldsymbol{d}}_f = \widehat{\boldsymbol{d}}_s, \quad \text{on} \quad \widehat{\Sigma}. \tag{11.15}$$

  Inside $\widehat{\Omega}_f$, the fluid domain displacement $\widehat{\boldsymbol{d}}_f$ can be defined as an arbitrary L^2-extension of $\widehat{\boldsymbol{d}}_s$ over the domain $\widehat{\Omega}_f$, namely,

$$\widehat{\boldsymbol{d}}_f = \mathrm{Ext}(\widehat{\boldsymbol{d}}_{s|\widehat{\Sigma}}) \tag{11.16}$$

  (see Remark 39 below).
- A kinematic condition enforcing the continuity of the velocities at the interface

$$\boldsymbol{u} = \frac{\partial \widehat{\boldsymbol{d}}_s}{\partial t} \circ \widehat{\mathcal{A}}_t^{-1}, \quad \text{on} \quad \Sigma(t). \tag{11.17}$$

- And a kinetic condition imposing the stress continuity at the interface

$$\widehat{\boldsymbol{T}}\widehat{\boldsymbol{n}}_s = \widehat{J}_f \widehat{\boldsymbol{\sigma}(\boldsymbol{u}, p)} \widehat{\boldsymbol{F}}_f^{-T} \widehat{\boldsymbol{n}}_s, \quad \text{on} \quad \widehat{\Sigma}. \tag{11.18}$$

To summarize, the fluid-structure system involving an incompressible viscous fluid and a hyperelastic structure is described in terms of the unknowns $(\boldsymbol{u}, p, \widehat{\boldsymbol{d}}_f, \widehat{\boldsymbol{d}}_s)$ satisfying the coupled problem (11.13)-(11.18).

Remark **39** *In practice, we can choose as operator Ext a harmonic extension operator, by solving a Laplace equation*

$$\begin{cases} -\kappa \Delta \widehat{\boldsymbol{d}}_f = 0, & \text{on} \quad \widehat{\Omega}_f, \\ \widehat{\boldsymbol{d}}_f = \widehat{\boldsymbol{d}}_s, & \text{on} \quad \widehat{\Sigma}, \\ \widehat{\boldsymbol{d}}_f = \boldsymbol{0}, & \text{on} \quad \widehat{\Gamma}_f, \end{cases} \tag{11.19}$$

where $\kappa > 0$ is a given "diffusion" coefficient, that might depend on $\widehat{\boldsymbol{d}}_s$. Other alternative extension approaches can be found, for instance, in [19, 174].

Remark **40** *The combination of* (11.15) *and* (11.17) *enforces $\boldsymbol{u} = \boldsymbol{w}$ on $\Sigma(t)$. This requirement is not strictly necessary but simplifies the construction of the ALE map. In general we could replace* (11.16) *by $\boldsymbol{u} \cdot \boldsymbol{n} = \boldsymbol{w} \cdot \boldsymbol{n}$ on $\Sigma(t)$.*

Remark **41** *For simplicity, we have only prescribed Neumann boundary conditions in* (11.13). *In practice we may use Dirichlet conditions on some part of the boundary.*

11.3.2 Weak formulation

Problem (11.13)-(11.18) can be reformulated in a weak variational form using appropriate test functions, performing integrations by parts and taking into account the boundary and interface conditions.

In what follows, we will make explicit the dependence of $\Omega_f(t)$ and $\Sigma(t)$ on $\widehat{\boldsymbol{d}}_f$ by introducing the notations

$$\Omega_f(\widehat{\boldsymbol{d}}_f) \stackrel{\text{def}}{=} \Omega_f(t), \quad \Sigma(\widehat{\boldsymbol{d}}_f) \stackrel{\text{def}}{=} \Sigma(t).$$

Let $(\widehat{\boldsymbol{v}_f}, \widehat{q}) \in [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f)$, multiplying the fluid problem (11.13) by $(\boldsymbol{v}_f, q) = (\widehat{\boldsymbol{v}_f} \circ \widehat{\mathcal{A}}_t^{-1}, \widehat{q} \circ \widehat{\mathcal{A}}_t^{-1})$ integrating over $\Omega_f(\widehat{\boldsymbol{d}}_f)$ and after integrations by parts we get

$$\frac{\mathrm{d}}{\mathrm{dt}} \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \rho_f \boldsymbol{u} \cdot \boldsymbol{v}_f \, dx + \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \mathrm{div} \left[\rho_f \boldsymbol{u} \otimes \left(\boldsymbol{u} - \boldsymbol{w}(\widehat{\boldsymbol{d}_f}) \right) \right] \cdot \boldsymbol{v}_f \, dx + \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \boldsymbol{\sigma}(\boldsymbol{u}, p) : \nabla \boldsymbol{v}_f \, dx$$
$$- \int_{\Sigma(\widehat{\boldsymbol{d}_f})} \boldsymbol{\sigma}(\boldsymbol{u}, p) \cdot \boldsymbol{v}_f \cdot \boldsymbol{n} \, \mathrm{d}\boldsymbol{a} - \int_{\Gamma_{\mathrm{in-out}}} \boldsymbol{g} \cdot \boldsymbol{v}_f \, \mathrm{d}\boldsymbol{a} - \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} q \, \mathrm{div} \, \boldsymbol{u} \, dx = 0,$$

where

$$\boldsymbol{w}(\widehat{\boldsymbol{d}_f}) = \frac{\partial \widehat{\boldsymbol{d}_f}}{\partial t} \circ \widehat{\mathcal{A}}_t^{-1}.$$

For the structure, multiplying (11.14) by $\widehat{\boldsymbol{v}_s} \in [H^1_{\Gamma_D}(\widehat{\Omega}_s)]^3$, integrating over $\widehat{\Omega}_s$ and integrating by parts, one gets

$$\int_{\widehat{\Omega}_s} \rho_0 \frac{\partial^2 \widehat{\boldsymbol{d}_s}}{\partial t^2} \cdot \widehat{\boldsymbol{v}_s} \, \mathrm{d}\hat{\boldsymbol{x}} + \int_{\widehat{\Omega}_s} \frac{\partial W}{\partial F}(\mathrm{I} + \nabla \widehat{\boldsymbol{d}_s}) : \nabla \widehat{\boldsymbol{v}_s} \, \mathrm{d}\hat{\boldsymbol{x}} - \int_{\widehat{\Sigma}} \frac{\partial W}{\partial F}(\mathrm{I} + \nabla \widehat{\boldsymbol{d}_s}) \widehat{\boldsymbol{n}}_s \cdot \widehat{\boldsymbol{v}_s} \, \mathrm{d}\hat{\boldsymbol{a}} = 0,$$

where $\rho_0 = \widehat{J}_s \rho_s$. Therefore, taking into account the coupling condition (11.18), it follows that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \rho_f \boldsymbol{u} \cdot \boldsymbol{v}_f \, dx + \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \operatorname{div}\Big[\rho_f \boldsymbol{u} \otimes \Big(\boldsymbol{u} - \boldsymbol{w}(\widehat{\boldsymbol{d}_f})\Big)\Big] \cdot \boldsymbol{v}_f \, dx + \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \boldsymbol{\sigma}(\boldsymbol{u}, p) : \nabla \boldsymbol{v}_f \, dx \\ &- \int_{\Gamma_{\mathrm{in-out}}} \boldsymbol{g} \cdot \boldsymbol{v}_f \, \mathrm{d}\boldsymbol{a} - \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} q \operatorname{div} \boldsymbol{u} \, dx + \int_{\widehat{\Omega}_s} \rho_0 \frac{\partial^2 \widehat{\boldsymbol{d}_s}}{\partial t^2} \cdot \widehat{\boldsymbol{v}_s} \, \mathrm{d}\hat{\boldsymbol{x}} + \int_{\widehat{\Omega}_s} \frac{\partial W}{\partial F}(\mathrm{I} + \nabla \widehat{\boldsymbol{d}_s}) : \nabla \widehat{\boldsymbol{v}_s} \, \mathrm{d}\hat{\boldsymbol{x}} = 0, \end{aligned} \tag{11.20}$$

for all $(\widehat{\boldsymbol{v}_f}, \widehat{q}) \in [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f)$ and $\widehat{\boldsymbol{v}_s} \in [H^1_{\Gamma_D}(\widehat{\Omega}_s)]^3$ with $\widehat{\boldsymbol{v}_f} = \widehat{\boldsymbol{v}_s}$ on $\widehat{\Sigma}$. The weak form of the geometry coupling conditions (11.15) and (11.16) are rewritten in terms of the interface displacement $\boldsymbol{\gamma} \in [H^{\frac{1}{2}}(\widehat{\Sigma})]^3$ as

$$\int_{\widehat{\Omega}_f} \Big(\widehat{\boldsymbol{d}_f} - \mathrm{Ext}(\boldsymbol{\gamma})\Big) \cdot \widehat{\boldsymbol{\tau}} \, \mathrm{d}\hat{\boldsymbol{x}} + \int_{\widehat{\Sigma}} (\widehat{\boldsymbol{d}_s} - \boldsymbol{\gamma}) \cdot \widehat{\boldsymbol{\zeta}} \, \mathrm{d}\hat{\boldsymbol{a}} = 0, \tag{11.21}$$

for all $\widehat{\boldsymbol{\tau}} \in [L^2(\widehat{\Omega}_f)]^3$ and $\widehat{\boldsymbol{\zeta}} \in [L^2(\widehat{\Sigma})]^3$. Finally, the continuity of the velocities at the interface (11.17) is reformulated as

$$\int_{\widehat{\Sigma}} \Big(\widehat{\boldsymbol{u}} - \widehat{\boldsymbol{w}}(\widehat{\boldsymbol{d}_f})\Big) \cdot \widehat{\boldsymbol{\xi}} \, \mathrm{d}\hat{\boldsymbol{a}} = 0, \tag{11.22}$$

for all $\widehat{\boldsymbol{\xi}} \in [L^2(\widehat{\Sigma})]^3$.

Therefore, after summation of (11.20)-(11.22) we obtain the following global weak formulation of problem (11.13)-(11.18): Find $\widehat{\boldsymbol{u}} : \widehat{\Omega}_f \times \mathbb{R}^+ \to \mathbb{R}^3$, $\widehat{p} : \widehat{\Omega}_f \times \mathbb{R}^+ \to \mathbb{R}$, $\widehat{\boldsymbol{d}}_f : \widehat{\Omega}_f \times \mathbb{R}^+ \to \mathbb{R}^3$, $\widehat{\boldsymbol{d}}_s : \widehat{\Omega}_s \times \mathbb{R}^+ \to \mathbb{R}^3$ and $\boldsymbol{\gamma} : \widehat{\Sigma} \times \mathbb{R}^+ \to \mathbb{R}^3$ such that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \rho_f \boldsymbol{u} \cdot \boldsymbol{v}_f \, dx + \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \operatorname{div}\Big[\rho_f \boldsymbol{u} \otimes \Big(\boldsymbol{u} - \boldsymbol{w}(\widehat{\boldsymbol{d}_f})\Big)\Big] \cdot \boldsymbol{v}_f \, dx + \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} \boldsymbol{\sigma}(\boldsymbol{u}, p) : \nabla \boldsymbol{v}_f \, dx \\ &- \int_{\Gamma_{\mathrm{in-out}}} \boldsymbol{g} \cdot \boldsymbol{v}_f \, \mathrm{d}\boldsymbol{a} - \int_{\Omega_f(\widehat{\boldsymbol{d}_f})} q \operatorname{div} \boldsymbol{u} \, dx + \int_{\widehat{\Omega}_s} \rho_0 \frac{\partial^2 \widehat{\boldsymbol{d}_s}}{\partial t^2} \cdot \widehat{\boldsymbol{v}_s} \, \mathrm{d}\hat{\boldsymbol{x}} + \int_{\widehat{\Omega}_s} \frac{\partial W}{\partial F}(\mathrm{I} + \nabla \widehat{\boldsymbol{d}_s}) : \nabla \widehat{\boldsymbol{v}_s} \, \mathrm{d}\hat{\boldsymbol{x}} \\ &\quad + \int_{\widehat{\Omega}_f} \Big(\widehat{\boldsymbol{d}_f} - \mathrm{Ext}(\boldsymbol{\gamma})\Big) \cdot \widehat{\boldsymbol{\tau}} \, \mathrm{d}\hat{\boldsymbol{x}} + \int_{\widehat{\Sigma}} (\widehat{\boldsymbol{d}_s} - \boldsymbol{\gamma}) \cdot \widehat{\boldsymbol{\zeta}} \, \mathrm{d}\hat{\boldsymbol{a}} + \int_{\widehat{\Sigma}} \Big(\widehat{\boldsymbol{u}} - \widehat{\boldsymbol{w}}(\widehat{\boldsymbol{d}_f})\Big) \cdot \widehat{\boldsymbol{\xi}} \, \mathrm{d}\hat{\boldsymbol{a}} = 0, \end{aligned} \tag{11.23}$$

with $\boldsymbol{u} = \widehat{\boldsymbol{u}} \circ \widehat{\mathcal{A}}_t^{-1}$, $p = \widehat{p} \circ \widehat{\mathcal{A}}_t^{-1}$, and for all $(\widehat{\boldsymbol{v}_f}, \widehat{q}) \in [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f)$, $\boldsymbol{v}_s \in [H^1_{\Gamma_D}(\widehat{\Omega}_s)]^3$ with $\widehat{\boldsymbol{v}_f} = \widehat{\boldsymbol{v}_s}$ on $\widehat{\Sigma}$, $\widehat{\boldsymbol{\tau}} \in [L^2(\widehat{\Omega}_f)]^3$, $\widehat{\boldsymbol{\zeta}} \in [L^2(\widehat{\Sigma})]^3$ and $\widehat{\boldsymbol{\xi}} \in [L^2(\widehat{\Sigma})]^3$.

11.4 Semi-discretized weak formulation

In this section, the weak coupled formulation (11.23) is semi-discretized in time using an implicit coupling-scheme. The resulting nonlinear problem will be turned into an abstract form. This will allow us to introduce in the next section general nonlinear iterative solution methods.

11.4.1 Implicit coupling scheme

We use an implicit Euler scheme for the ALE Navier-Stokes equations, with a semi-implicit treatment of the nonlinear convective term. Furthermore we use a mid-point rule for the structural equation. Thus, given a time step $\delta t > 0$, for $n = 0, 1, \ldots$, the time semi-discretized coupled problem writes: Given $\left(\widehat{\boldsymbol{u}}^n, \widehat{p}^n, \widehat{\boldsymbol{d}_f}^n, \widehat{\boldsymbol{d}_s}^n, \boldsymbol{\gamma}^n\right)$, find

$$\left(\widehat{\boldsymbol{u}}^{n+1}, \widehat{p}^{n+1}, \widehat{\boldsymbol{d}_f}^{n+1}, \widehat{\boldsymbol{d}_s}^{n+1}, \boldsymbol{\gamma}^{n+1}\right) \in [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [H^1(\widehat{\Omega}_f)]^3 \times [H^1(\widehat{\Omega}_s)]^3 \times [H^{\frac{1}{2}}(\widehat{\Sigma})]^3,$$

such that

$$\begin{aligned}
&\frac{1}{\delta t}\int_{\Omega_f(\widehat{\boldsymbol{d}}_f^{n+1})} \rho_f \boldsymbol{u}^{n+1}\cdot \boldsymbol{v}_f\, dx - \frac{1}{\delta t}\int_{\Omega_f(\widehat{\boldsymbol{d}}_f^{n})} \rho_f \boldsymbol{u}^{n}\cdot \boldsymbol{v}_f\, dx + \int_{\Omega_f(\widehat{\boldsymbol{d}}_f^{n+1})} \boldsymbol{\sigma}(\boldsymbol{u}^{n+1}, p^{n+1}) : \nabla \boldsymbol{v}_f\, dx \\
&+ \int_{\Omega_f(\widehat{\boldsymbol{d}}_f^{n+1})} \operatorname{div}\left[\rho_f \boldsymbol{u}^{n+1}\otimes\left(\boldsymbol{u}^n - \boldsymbol{w}\left(\widehat{\boldsymbol{d}_f}^{n+1}\right)\right)\right]\cdot \boldsymbol{v}_f\, dx - \int_{\Gamma_{\text{in-out}}} \boldsymbol{g}^{n+1}\cdot \boldsymbol{v}_f\, \mathrm{d}\boldsymbol{a} \\
&- \int_{\Omega_f(\widehat{\boldsymbol{d}}_f^{n+1})} q \operatorname{div} \boldsymbol{u}^{n+1}\, dx + \int_{\widehat{\Omega}_f}\left(\widehat{\boldsymbol{d}}_f^{n+1} - \operatorname{Ext}\left(\boldsymbol{\gamma}^{n+1}\right)\right)\cdot \widehat{\boldsymbol{\tau}}\, \mathrm{d}\hat{\boldsymbol{x}} + \int_{\widehat{\Sigma}}\left(\widehat{\boldsymbol{u}}^{n+1} - \widehat{\boldsymbol{w}}\left(\widehat{\boldsymbol{d}}_f^{n+1}\right)\right)\cdot \widehat{\boldsymbol{\xi}}\, \mathrm{d}\hat{\boldsymbol{a}} \\
&+ \frac{2}{\delta t^2}\int_{\widehat{\Omega}_s} \rho_0 \widehat{\boldsymbol{d}}_s^{n+1}\cdot \widehat{\boldsymbol{v}_s}\, \mathrm{d}\hat{\boldsymbol{x}} - \frac{2}{\delta t^2}\int_{\widehat{\Omega}_s} \rho_0\left(\widehat{\boldsymbol{d}}_s^{n} + \Delta t \dot{\widehat{\boldsymbol{d}}}_s^{n}\right)\cdot \widehat{\boldsymbol{v}_s}\, \mathrm{d}\hat{\boldsymbol{x}} \\
&+ \int_{\widehat{\Omega}_s} \frac{\partial W}{\partial F}\left(I + \frac{1}{2}\nabla(\widehat{\boldsymbol{d}}_s^{n} + \widehat{\boldsymbol{d}}_s^{n+1})\right) : \nabla \widehat{\boldsymbol{v}_s}\, \mathrm{d}\hat{\boldsymbol{x}} + \int_{\widehat{\Sigma}}\left(\widehat{\boldsymbol{d}}_s^{n+1} - \boldsymbol{\gamma}^{n+1}\right)\cdot \widehat{\boldsymbol{\zeta}}\, \mathrm{d}\hat{\boldsymbol{a}} = 0,
\end{aligned} \tag{11.24}$$

for all $(\widehat{\boldsymbol{v}_f}, \widehat{q}, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\tau}}, \widehat{\boldsymbol{\zeta}}, \widehat{\boldsymbol{v}_s}) \in [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [L^2(\widehat{\Sigma})]^3 \times [L^2(\widehat{\Omega}_f)]^3 \times [L^2(\widehat{\Sigma})]^3 \times [H^1_{\Gamma_D}(\widehat{\Omega}_s)]^3$ such that $\widehat{\boldsymbol{v}_f} = \widehat{\boldsymbol{v}_s}$ on $\widehat{\Sigma}$, and with $\boldsymbol{u}^n = \widehat{\boldsymbol{u}}^n \circ (\boldsymbol{I} + \widehat{\boldsymbol{d}_f}^n)^{-1}$ (analogously for p^n) and $\dot{\widehat{\boldsymbol{d}}}_s^{n+1} = \frac{2}{\delta t}\left(\widehat{\boldsymbol{d}}_s^{n+1} - \widehat{\boldsymbol{d}}_s^{n}\right) - \dot{\widehat{\boldsymbol{d}}}_s^{n}$.

11.4.2 Abstract formulations

Problem (11.24) can be rewritten in a more compact form in terms of the fluid, solid and interface state operators. This is the aim of the following paragraphs.

Based on the discrete weak formulation (11.24) we introduce the fluid operator

$$\begin{aligned}
\mathcal{F} : [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [H^1(\widehat{\Omega}_f)]^3 \times [H^{\frac{1}{2}}(\widehat{\Sigma})]^3 \\
\longrightarrow \left([H^1_{\widehat{\Sigma}}(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [L^2(\widehat{\Sigma})]^3 \times [L^2(\widehat{\Omega}_f)]^3\right)',
\end{aligned}$$

defined by

$$\begin{aligned}
\left\langle \mathcal{F}\left(\widehat{\boldsymbol{u}}, \widehat{p}, \widehat{\boldsymbol{d}}_f, \boldsymbol{\gamma}\right), \left(\widehat{\boldsymbol{v}_f}, \widehat{q}, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\tau}}\right)\right\rangle =& \frac{1}{\Delta t}\int_{\Omega_f(\widehat{\boldsymbol{d}}_f)} \rho_f \boldsymbol{u}\cdot \boldsymbol{v}_f\, dx - \frac{1}{\Delta t}\int_{\Omega_f(\widehat{\boldsymbol{d}}_f^n)} \rho_f \boldsymbol{u}^n\cdot \boldsymbol{v}_f\, dx \\
&+ \int_{\Omega_f(\widehat{\boldsymbol{d}}_f)} \operatorname{div}\left[\rho_f \boldsymbol{u}\otimes\left(\boldsymbol{u}^n - \boldsymbol{w}\left(\widehat{\boldsymbol{d}}_f\right)\right)\right]\cdot \boldsymbol{v}_f\, dx \\
&+ \int_{\Omega_F(\widehat{\boldsymbol{d}}_f)} \boldsymbol{\sigma}(\boldsymbol{u}, p) : \nabla \boldsymbol{v}_f\, dx - \int_{\Gamma_{\text{in-out}}(\widehat{\boldsymbol{d}}_f)} \boldsymbol{g}^{n+1}\cdot \boldsymbol{v}_f\, \mathrm{d}\boldsymbol{a} \\
&- \int_{\Omega_f(\widehat{\boldsymbol{d}}_f)} q \operatorname{div} \boldsymbol{u}\, dx + \int_{\widehat{\Sigma}}\left(\widehat{\boldsymbol{u}} - \widehat{\boldsymbol{w}}\left(\widehat{\boldsymbol{d}}_f\right)\right)\cdot \widehat{\boldsymbol{\xi}}\, \mathrm{d}\hat{\boldsymbol{a}} \\
&+ \int_{\widehat{\Omega}_f}\left(\widehat{\boldsymbol{d}}_f - \operatorname{Ext}(\boldsymbol{\gamma})\right)\cdot \widehat{\boldsymbol{\tau}}\, \mathrm{d}\hat{\boldsymbol{x}},
\end{aligned} \tag{11.25}$$

for all $(\widehat{\boldsymbol{v}_f}, \widehat{q}, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\tau}}) \in [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [L^2(\widehat{\Sigma})]^3 \times [L^2(\widehat{\Omega}_f)]^3$.

Analogously, from (11.24), the solid operator

$$\mathcal{S} : [H^1(\widehat{\Omega}_s)]^3 \times [H^{\frac{1}{2}}(\widehat{\Sigma})]^3 \longrightarrow ([H^1_{\Gamma_D \cup \widehat{\Sigma}}(\widehat{\Omega}_s)]^3 \times [L^2(\widehat{\Sigma})]^3)',$$

is given by

$$\begin{aligned} \left\langle \mathcal{S}(\widehat{\boldsymbol{d}}_s, \boldsymbol{\gamma}), (\widehat{\boldsymbol{v}_s}, \widehat{\boldsymbol{\zeta}}) \right\rangle =& \frac{2}{\delta t^2} \int_{\widehat{\Omega}_s} \rho_0 \widehat{\boldsymbol{d}}_s \cdot \boldsymbol{v}_s \, \mathrm{d}\hat{\boldsymbol{x}} - \frac{2}{\delta t^2} \int_{\widehat{\Omega}_s} \rho_0 \left(\widehat{\boldsymbol{d}}_s^n + \delta t \dot{\widehat{\boldsymbol{d}}}_s^n \right) \cdot \boldsymbol{v}_s \, \mathrm{d}\hat{\boldsymbol{x}} \\ &+ \int_{\widehat{\Omega}_s} \frac{\partial W}{\partial F} \left(I + \frac{1}{2} \nabla \left(\widehat{\boldsymbol{d}}_s^n + \widehat{\boldsymbol{d}}_s \right) \right) : \nabla \widehat{\boldsymbol{v}_s} \, \mathrm{d}\hat{\boldsymbol{x}} + \int_{\widehat{\Sigma}} (\widehat{\boldsymbol{d}}_s - \boldsymbol{\gamma}) \cdot \widehat{\boldsymbol{\zeta}} \, \mathrm{d}\hat{\boldsymbol{a}}, \end{aligned} \tag{11.26}$$

for all $(\widehat{\boldsymbol{v}_s}, \widehat{\boldsymbol{\zeta}}) \in [H^1_{\Gamma_D}(\widehat{\Omega}_s)]^3 \times [L^2(\widehat{\Sigma})]^3$.

Finally, let $\mathcal{L}_f : [H^{\frac{1}{2}}(\widehat{\Sigma})]^3 \to [H^1_{\Gamma_{\text{in-out}}}(\widehat{\Omega}_f)]^3$ and $\mathcal{L}_s : [H^{\frac{1}{2}}(\widehat{\Sigma})]^3 \to [H^1_{\partial\widehat{\Omega}_s \setminus \widehat{\Sigma}}(\widehat{\Omega}_s)]^3$ be two given continuous linear lift operators. The interface operator

$$\mathcal{I} : [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [H^1(\widehat{\Omega}_f)]^3 \times [H^1(\widehat{\Omega}_s)]^3 \longrightarrow [H^{-\frac{1}{2}}(\widehat{\Sigma})]^3,$$

is then defined by

$$\left\langle \mathcal{I}\left(\widehat{\boldsymbol{u}}, \widehat{p}, \widehat{\boldsymbol{d}}_f, \widehat{\boldsymbol{d}}_s \right), \boldsymbol{\mu} \right\rangle = \left\langle \mathcal{F}(\widehat{\boldsymbol{u}}, \widehat{p}, \widehat{\boldsymbol{d}}_f, \boldsymbol{\gamma}), (\mathcal{L}_f \boldsymbol{\mu}, 0, \mathbf{0}, \mathbf{0}) \right\rangle + \left\langle \mathcal{S}(\widehat{\boldsymbol{d}}_s, \boldsymbol{\gamma}), (\mathcal{L}_s \boldsymbol{\mu}, \mathbf{0}) \right\rangle, \tag{11.27}$$

for all $\boldsymbol{\mu} \in [H^{\frac{1}{2}}(\widehat{\Sigma})]^3$.

Remark **42** *The interface operator does not depend on* $\boldsymbol{\gamma}$ *since, due to the choice of the test functions, the terms involving* $\boldsymbol{\gamma}$ *vanishe in the right-hand side of* (11.27).

According to the above definitions, problem (11.24) is equivalent to

$$\text{Formulation (I):} \qquad \begin{cases} \mathcal{F}\left(\widehat{\boldsymbol{u}}^{n+1}, \widehat{p}^{n+1}, \widehat{\boldsymbol{d}}_f^{n+1}, \boldsymbol{\gamma}^{n+1} \right) = 0, \\ \mathcal{S}\left(\widehat{\boldsymbol{d}}_s^{n+1}, \boldsymbol{\gamma}^{n+1} \right) = 0, \\ \mathcal{I}\left(\widehat{\boldsymbol{u}}^{n+1}, \widehat{p}^{n+1}, \widehat{\boldsymbol{d}}_f^{n+1}, \widehat{\boldsymbol{d}}_s^{n+1} \right) = 0. \end{cases} \tag{11.28}$$

11.4.3 Steklov-Poincaré operators

In order to describe partitioned methods for the numerical solution of (11.24), we now introduce the nonlinear fluid and solid Steklov-Poincaré operators.

The nonlinear fluid Steklov-Poincaré operator

$$S_f : [H^{\frac{1}{2}}(\widehat{\Sigma})]^3 \longrightarrow [H^{-\frac{1}{2}}(\widehat{\Sigma})]^3,$$

is defined by

$$\langle S_f(\boldsymbol{\gamma}), \boldsymbol{\mu} \rangle = \left\langle \mathcal{I}\big(\widehat{\boldsymbol{u}}(\boldsymbol{\gamma}), \widehat{p}(\boldsymbol{\gamma}), \widehat{\boldsymbol{d}}_f(\boldsymbol{\gamma}), \mathbf{0}\big), \boldsymbol{\mu} \right\rangle,$$

for all $\boldsymbol{\gamma}, \boldsymbol{\mu} \in [H^{\frac{1}{2}}(\hat{\Sigma})]^3$, where $\big(\widehat{\boldsymbol{u}}(\boldsymbol{\gamma}), \widehat{p}(\boldsymbol{\gamma}), \widehat{\boldsymbol{d}}_f(\boldsymbol{\gamma})\big)$ is the solution of the Dirichlet fluid problem:

$$\mathcal{F}\left(\widehat{\boldsymbol{u}}(\boldsymbol{\gamma}), \widehat{p}(\boldsymbol{\gamma}), \widehat{\boldsymbol{d}_f}(\boldsymbol{\gamma}), \boldsymbol{\gamma} \right) = 0.$$

In an analogous way, we introduce the nonlinear solid Steklov-Poincaré operator

$$S_s : [H^{\frac{1}{2}}(\hat{\Sigma})]^3 \longrightarrow [H^{-\frac{1}{2}}(\hat{\Sigma})]^3,$$

given by

$$\langle S_s(\boldsymbol{\gamma}), \boldsymbol{\mu}\rangle = \left\langle \mathcal{I}\big(\mathbf{0}, 0, \mathbf{0}, \widehat{\boldsymbol{d}}_s(\boldsymbol{\gamma})\big), \boldsymbol{\mu}\right\rangle,$$

for all $\boldsymbol{\gamma}, \boldsymbol{\mu} \in [H^{\frac{1}{2}}(\hat{\Sigma})]^3$ and where $\widehat{\boldsymbol{d}_s}(\boldsymbol{\gamma})$ is the solution of the Dirichlet solid problem:

$$\mathcal{S}\left(\widehat{\boldsymbol{d}}_s(\boldsymbol{\gamma}), \boldsymbol{\gamma}\right) = 0.$$

From the above definitions, it follows that problem (11.24) (or (11.28)) is equivalent to

$$\text{Formulation (II):} \qquad S_f(\boldsymbol{\gamma}) + S_s(\boldsymbol{\gamma}) = 0. \tag{11.29}$$

The composition of (11.29) with the inverse operators S_s^{-1} gives rise to the Dirichlet-to-Neumann formulation, namely

$$\text{Formulation (III):} \qquad S_s^{-1}\big(-S_f(\boldsymbol{\gamma})\big) - \boldsymbol{\gamma} = 0. \tag{11.30}$$

We could also consider the Neumann-to-Dirichlet formulation

$$S_f^{-1}\big(-S_s(\boldsymbol{\gamma})\big) - \boldsymbol{\gamma} = 0$$

by composing (11.29) with S_f^{-1}. Nevertheless it is rarely used in practice and it is known to lead to poor algorithms in some cases as pointed out in [48].

11.5 A partitioned Newton method

In what follows, we skip the upper script n since the time step is fixed. The method presented here consists in solving (11.28) by a Newton method: given an initial guess $(\widehat{\boldsymbol{u}}_0, \widehat{p}_0, \widehat{\boldsymbol{d}_f}^0, \widehat{\boldsymbol{d}_s}^0, \boldsymbol{\gamma}_0)$, the algorithm reads

1. Evaluate the nonlinear residual of problem (11.28).
2. Solve the tangent problem (see (11.34) below) by a domain decomposition method.
3. Update solution: $\left(\widehat{\boldsymbol{u}}, \widehat{p}, \widehat{\boldsymbol{d}}_f, \widehat{\boldsymbol{d}}_s, \boldsymbol{\gamma}\right) \leftarrow \left(\widehat{\boldsymbol{u}}, \widehat{p}, \widehat{\boldsymbol{d}}_f, \widehat{\boldsymbol{d}}_s, \boldsymbol{\gamma}\right) + \left(\delta\widehat{\boldsymbol{u}}, \delta\widehat{p}, \delta\widehat{\boldsymbol{d}}_f, \delta\widehat{\boldsymbol{d}}_s, \delta\boldsymbol{\gamma}\right)$.
4. repeat until convergence.

Compared to the known fluid-structure algorithms presented in Section 11.2.3, this partitioned Newton method amounts to switching the domain decomposition and the linearization in the resolution of the coupled problem. We provide the tangent problem and detail the domain decomposition algorithm in the following sections.

11.5.1 Weak state operators derivatives

In this section, we present the differentiation of the fluid, structure and interface operators of Section 11.4.2 with respect to their arguments. This derivation uses shape derivative calculus for the differentiation of integral terms with respect to their supports. We refer the reader to [84] where this issue is addressed.

The linearized fluid operator at state $(\widehat{\boldsymbol{u}}, \widehat{p}, \widehat{\boldsymbol{d}}_f, \boldsymbol{\gamma}) \in [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [H^1(\widehat{\Omega}_f)]^3 \times [H^{\frac{1}{2}}(\widehat{\Sigma})]^3$ is denoted by

$$\begin{aligned}\mathrm{D}\,\mathcal{F}(\widehat{\boldsymbol{u}}, \widehat{p}, \widehat{\boldsymbol{d}}_f, \boldsymbol{\gamma}) : {} & [H^1(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [H^1(\widehat{\Omega}_f)]^3 \times [H^{\frac{1}{2}}(\widehat{\Sigma})]^3 \\ & \longrightarrow \left([H^1_{\widehat{\Sigma}}(\widehat{\Omega}_f)]^3 \times L^2(\widehat{\Omega}_f) \times [L^2(\widehat{\Sigma})]^3 \times [L^2(\widehat{\Omega}_f)]^3\right)',\end{aligned}$$

and is given by

$$
\begin{aligned}
&\langle \mathrm{D}\,\mathcal{F}(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\boldsymbol{\gamma})\cdot(\delta\widehat{\boldsymbol{u}},\delta\widehat{p},\delta\widehat{\boldsymbol{d}_f},\delta\boldsymbol{\gamma}),(\widehat{\boldsymbol{v}_f},\widehat{q},\widehat{\boldsymbol{\xi}},\widehat{\boldsymbol{\tau}})\rangle\\
&=\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}\operatorname{div}\left[\rho_f\delta\boldsymbol{u}\otimes(\boldsymbol{u}^n-\boldsymbol{w}(\widehat{\boldsymbol{d}_f}))\right]\cdot\boldsymbol{v}_f\,dx+\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}\boldsymbol{\sigma}(\delta\boldsymbol{u},\delta p):\nabla\boldsymbol{v}_f\,dx\\
&-\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}q\operatorname{div}\delta\boldsymbol{u}\,dx+\frac{1}{\Delta t}\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}(\operatorname{div}\delta\widehat{\boldsymbol{d}_f})\rho_f\boldsymbol{u}\cdot\boldsymbol{v}_f\,dx\\
&+\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}\operatorname{div}\left\{\rho_f\boldsymbol{u}\otimes(\boldsymbol{u}^n-\boldsymbol{w}(\widehat{\boldsymbol{d}_f}))\left[\mathrm{I}\operatorname{div}\delta\widehat{\boldsymbol{d}_f}-(\nabla\delta\widehat{\boldsymbol{d}_f})^{\mathrm{T}}\right]\right\}\cdot\boldsymbol{v}_f\,dx\\
&-\frac{1}{\Delta t}\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}\operatorname{div}(\rho_f\boldsymbol{u}\otimes\delta\widehat{\boldsymbol{d}_f})\cdot\boldsymbol{v}_f\,dx+\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}\boldsymbol{\sigma}(\boldsymbol{u},p)\left[\mathrm{I}\operatorname{div}\delta\widehat{\boldsymbol{d}_f}-(\nabla\delta\widehat{\boldsymbol{d}_f})^{\mathrm{T}}\right]:\nabla\boldsymbol{v}_f\,dx\\
&-\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}\mu\left[\nabla\boldsymbol{u}\nabla\delta\widehat{\boldsymbol{d}_f}+(\nabla\delta\widehat{\boldsymbol{d}_f})^{\mathrm{T}}(\nabla\boldsymbol{u})^{\mathrm{T}}\right]:\nabla\boldsymbol{v}_f\,dx\\
&-\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}q\operatorname{div}\left\{\boldsymbol{u}\left[\mathrm{I}\operatorname{div}\delta\widehat{\boldsymbol{d}_f}-(\nabla\delta\widehat{\boldsymbol{d}_f})^{\mathrm{T}}\right]\right\}\,dx+\int_{\widehat{\Sigma}}\left(\delta\widehat{\boldsymbol{u}}-\frac{\delta\widehat{\boldsymbol{d}_f}}{\Delta t}\right)\cdot\widehat{\boldsymbol{\xi}}\,\mathrm{d}\hat{\boldsymbol{a}}\\
&+\frac{\rho}{\Delta t}\int_{\Omega_F(\widehat{\boldsymbol{d}_f})}\delta\boldsymbol{u}\cdot\boldsymbol{v}_f\,dx+\int_{\widehat{\Omega}_F}(\delta\widehat{\boldsymbol{d}_f}-\operatorname{Ext}(\delta\boldsymbol{\gamma}))\cdot\widehat{\boldsymbol{\tau}}\,\mathrm{d}\hat{\boldsymbol{x}}
\end{aligned}
\tag{11.31}
$$

for all $(\widehat{\boldsymbol{v}_f},\widehat{q},\widehat{\boldsymbol{\xi}},\widehat{\boldsymbol{\tau}})\in[H^1(\widehat{\Omega}_f)]^3\times L^2(\widehat{\Omega}_f)\times[L^2(\widehat{\Sigma})]^3\times[L^2(\widehat{\Omega}_f)]^3$.

The linearized solid operator at state $(\widehat{\boldsymbol{d}}_s,\boldsymbol{\gamma})\in[H^1_{\Gamma_D}(\widehat{\Omega}_s)]^3\times[L^2(\widehat{\Sigma})]^3$

$$
\mathrm{D}\,\mathcal{S}(\widehat{\boldsymbol{d}}_s,\boldsymbol{\gamma}):[H^1_{\Gamma_D}(\widehat{\Omega}_s)]^3\times[H^{\frac{1}{2}}(\widehat{\Sigma})]^3\longrightarrow([H^1_{\Gamma_D\cup\widehat{\Sigma}}(\widehat{\Omega}_s)]^3\times[L^2(\widehat{\Sigma})]^3)',
$$

is given by

$$
\begin{aligned}
\langle \mathrm{D}\,\mathcal{S}(\widehat{\boldsymbol{d}}_s,\boldsymbol{\gamma})\cdot(\delta\widehat{\boldsymbol{d}_s},\delta\boldsymbol{\gamma}),(\widehat{\boldsymbol{v}_s},\widehat{\boldsymbol{\zeta}})\rangle=&\frac{2}{(\Delta t)^2}\int_{\widehat{\Omega}_S}\rho_0\delta\widehat{\boldsymbol{d}_s}\cdot\boldsymbol{v}_s\,\mathrm{d}\hat{\boldsymbol{x}}\\
&+\frac{1}{2}\int_{\widehat{\Omega}_S}\nabla\delta\widehat{\boldsymbol{d}_s}:\left(\frac{\partial^2W}{\partial F^2}(I+\nabla\widehat{\boldsymbol{d}_s})\right):\nabla\boldsymbol{v}_s\,\mathrm{d}\hat{\boldsymbol{x}}\\
&+\int_{\widehat{\Sigma}}(\delta\widehat{\boldsymbol{d}_s}-\delta\boldsymbol{\gamma})\cdot\widehat{\boldsymbol{\zeta}}\,\mathrm{d}\hat{\boldsymbol{a}},
\end{aligned}
\tag{11.32}
$$

for all $(\widehat{\boldsymbol{v}_s},\widehat{\boldsymbol{\zeta}})\in[H^1_{\Gamma_D}(\widehat{\Omega}_s)]^3\times[L^2(\widehat{\Sigma})]^3$.

We finally introduce a linearized interface operator at state $(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\widehat{\boldsymbol{d}}_s)$

$$
\mathrm{D}\,\mathcal{I}(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\widehat{\boldsymbol{d}}_s):[H^1(\widehat{\Omega}_f)]^3\times L^2(\widehat{\Omega}_f)\times[H^1(\widehat{\Omega}_f)]^3\times[H^1(\widehat{\Omega}_s)]^3\longrightarrow[H^{-\frac{1}{2}}(\widehat{\Sigma})]^3,
$$

defined by

$$
\begin{aligned}
\left\langle \mathrm{D}\,\mathcal{I}(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\widehat{\boldsymbol{d}}_s)\cdot\left(\delta\widehat{\boldsymbol{u}},\delta\widehat{p},\delta\widehat{\boldsymbol{d}}_f,\delta\widehat{\boldsymbol{d}}_s\right),\boldsymbol{\mu}\right\rangle=&\left\langle \mathrm{D}\,\mathcal{F}\left(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\boldsymbol{0}\right)\cdot\left(\delta\widehat{\boldsymbol{u}},\delta\widehat{p},\delta\widehat{\boldsymbol{d}}_f,\boldsymbol{0}\right),(\mathcal{L}_f\boldsymbol{\mu},0,\boldsymbol{0},\boldsymbol{0})\right\rangle\\
&+\left\langle \mathrm{D}\,\mathcal{S}\left(\widehat{\boldsymbol{d}}_s,\boldsymbol{0}\right)\left(\delta\widehat{\boldsymbol{d}}_s,\boldsymbol{0}\right),(\mathcal{L}_s\boldsymbol{\mu},\boldsymbol{0})\right\rangle,
\end{aligned}
\tag{11.33}
$$

for all $\boldsymbol{\mu}\in[H^{\frac{1}{2}}(\widehat{\Sigma})]^3$.

In terms of the above defined operators, the tangent problem associated to (11.28) reads

$$
\left\{
\begin{aligned}
\mathrm{D}\,\mathcal{F}\left(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\boldsymbol{\gamma}\right)\cdot\left(\delta\widehat{\boldsymbol{u}},\delta\widehat{p},\delta\widehat{\boldsymbol{d}}_f,\delta\boldsymbol{\gamma}\right)&=-\mathcal{F}\left(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\boldsymbol{\gamma}\right),\\
\mathrm{D}\,\mathcal{S}\left(\widehat{\boldsymbol{d}}_s,\boldsymbol{\gamma}\right)\cdot\left(\delta\widehat{\boldsymbol{d}}_s,\delta\boldsymbol{\gamma}\right)&=-\mathcal{S}\left(\widehat{\boldsymbol{d}}_s,\boldsymbol{\gamma}\right),\\
\mathrm{D}\,\mathcal{I}\left(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\widehat{\boldsymbol{d}}_s\right)\cdot\left(\delta\widehat{\boldsymbol{u}},\delta\widehat{p},\delta\widehat{\boldsymbol{d}}_f,\delta\widehat{\boldsymbol{d}}_s\right)&=-\mathcal{I}\left(\widehat{\boldsymbol{u}},\widehat{p},\widehat{\boldsymbol{d}}_f,\widehat{\boldsymbol{d}}_s\right).
\end{aligned}
\right.
\tag{11.34}
$$

Once the linear fluid, solid and interface operators $\mathrm{D}\mathcal{F}$, $\mathrm{D}\mathcal{S}$ and $\mathrm{D}\mathcal{I}$ are defined, we can introduce the linear Steklov-Poincaré operators $S_{F,l}$ and $S_{S,l}$ using the formula of Section 11.4.3 with the linearized operators instead of the nonlinear operators. It may be noticed that the linear Steklov-Poincaré operators are different from the linearization of the nonlinear Steklov operators of Section 11.4.3.

11.5.2 Domain decomposition method

In this section, we briefly describe the domain decomposition method used to solve the linear problems introduced above, with a Dirichlet-Neumann preconditioner (by the structure). The algorithm reads:

1. Evaluate the Newton residual (right hand side of the linear problem)
2. Initialization of the Domain Decomposition method: lifting of the external load and boundary conditions (zero on the interface)

Solid	Fluid
Receive zero from master	Receive zero from master
Matrix construction Matrix factorization (Dirichlet) Forward backward substitution	Computation of the preconditionner (in the fluid subdomain) GMRES
Send linear residual to master	Send linear residual to master

3. Preconditioning

Solid
Receive residual from master
if first preconditioning: Matrix factorization (Neumann) Forward backward substitution
Send linear displacement to master

4. Residual evaluation

Fluid
Receive displacement from master
GMRES
Send linear residual to master

5. End of Domain Decomposition

Solid	Fluid
Receive displacement from master	Receive displacement from master
Forward backward substitution (with external BC[a] and displacement on the interface)	GMRES (with external BC and velocity[b] on the interface)

[a] Boundary conditions
[b] note that $\delta u = \frac{1}{\Delta t}\delta d$ on the interface

As will be discussed in a forthcoming work (see [A2]), the domain decomposition method is part of the Newton algorithm, and needs not be solved very accurately since the test criterium is driven by the Newton residual.

11.5.3 Complexity analysis

Let us make a formal complexity analysis to have a rough hint on the cost of the Steklov type, Dirichlet to Neumann formulation based, and partitioned Newton type methods. We make the following assumptions: the fluid to be solved at each time step is linear (e.g. semi-implicit Euler for Navier-Stokes), the structure problem is solved by a Newton algorithm and the linearized structure problems by direct methods. We only take into account the factorization for the resolution of the structure sub-problem and consider the matrices as already factorized when dealing with linear domain decomposition methods.

In the following analysis we assume that the number of Newton iterations Ne_{FSI} for the global problem does not depend on the formulation used: (I), (II) or (III). We denote by Ne_s the number of iterations for a Newton algorithm in the structure problem. The number of GMRES iterations G is assumed not to depend on the algorithm if optimal preconditioners (let say Dirichlet-Neumann) are used. These simplifications allow us to compare the cost of the different methods. In the sequel Cr and Fa denote respectively the cost of the construction and factorization a matrix in the solid, Fl_1 the resolution cost per time step of the fluid problem, and Fl_2 the resolution cost for a tangent fluid problem. The estimations of costs for the three types of methods are gathered in Table 11.1 both for a sequential and a parallel implementation when possible. For the parallel implementation, we have assumed that $Fa + Cr \geq Fl$ and $Fl \geq Fa$.

Method	(III) Newton on DtoN-formulation	(II) NtoD preconditioned Newton on Steklov	(I) NtoD preconditioned partitioned Newton
Sequential	$Ne_{FSI}\,[(Ne_s+1)(Fa+Cr)$ $+Fl_1+GFl_2]$	$Ne_{FSI}\,[(Ne_s+1)(Fa+Cr)$ $+Fa+GFl_2+Fl_1]$	$Ne_{FSI}\,[2Fa+Cr$ $+GFl_2+Fl_1]$
Parallel	-	$Ne_{FSI}\,[(Ne_s+1)(Fa+Cr)$ $+Fa+GFl_2]$	$Ne_{FSI}\,[2Fa+Cr$ $+GFl_2]$

Table 11.1. Estimation of cost

Let us comment Table 11.1. For the sequential implementation, the estimations for the method (I) and (II) only differ by the factorization cost of a solid tangent matrix, which is rather small with respect to the whole cost. This is in agreement with the tests performed in [64] where method (II) is shown to be roughly equivalent to method (I) in terms of cost. If our estimation is still valid, the method (III) should be at least as efficient as the first two, especially if the structure is nonlinear and expensive. On the contrary, if $Fl \geq Fa + Cr$ then the parallel implementations of methods (II) and (III) seem to be completely equivalent in terms of cost, which is only determined by the fluid. For the parallel implementation, the cost reduction strongly depends on the number of GMRES iterations, and the method (III) still seems to compete with method (II).

The condition $Fa + Cr \geq Fl$ is almost never satisfied if classical shell elements are used. However, this condition may be satisfied when 3D shell elements are used to model more realistic constitutive laws for the structure (see [A2]). Let consider for instance a mesh with 38000 nodes in the fluid (let say 150000 degrees of freedom). For MITC4 shell elements, we then have 3300 nodes and 16500 degrees of freedom. Preliminary tests show that in this case, with the same computer, $Fl \simeq 45s$, $Fa \simeq 0.7s$ and $Cr \simeq 1.7s$. Let now consider 3D shell elements (hexahedra, 27 nodes per element) on the same mesh. The number of nodes for the structure increases from 3300 to 22100, and the number of degrees of freedom from 16500 to 66300. The costs for the solid are now $Fa \simeq 13s$ and $Cr \simeq 50s$. We are thus in the situation $Cr + Fa \geq Fl$ and $Fl \geq Fa$.

11.6 Description of a 3D shell for fluid-structure interaction

A general structural model of the blood flow with complex and realistic geometries has to be three-dimensional and handle large displacements.

Given that the wall of the blood vessels is thin, it is convenient to use shell elements; they accurately describe its geometry. All finite elements adopted in our simulations are general shell elements. Previously, Gerbeau *et. al.* have used the MITC4 elements [91,92] with a 3D constitutive law for which the transversal stress is null and a kinematikal constraint is needed to make the model compatible with a Reissner-Mindlin shell model. This restricts the choice of the energy.

In addition to the MITC4 element we consider here 3D-shell elements [49–51]. These elements appear as standard three-dimensional elements. Thus it is very easy to couple them to other three-dimensional formulations through the nodes on the faces. The element considered here, called MI3D, uses standard 3D $Q2$ shape functions. The advantage of a quadratic approximation in the shell's thickness is that it is possible to deal with standard 3D energies such as generalized Hook or any hyperelastic stored energy defined by using the Cauchy-Green tensor's invariants.

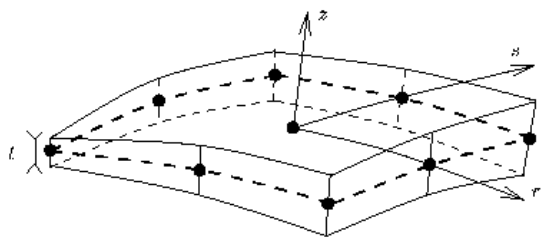

Fig. 11.4. 3D shell element

In order to be able to apply MITC techniques to stabilize the fomulation, it is necessary to compute the first and second derivatives of the stored energy with respect to the Green-Lagrange tensor, defined hereafter, in local coordinates (r, s, z), as it is usually done in shell element (see Figure 11.4):

$$e_{ij}(\mathbf{U}) = \frac{1}{2}(\mathbf{g}_i \cdot \mathbf{U}_{,j} + \mathbf{g}_j \cdot \mathbf{U}_{,i} + \mathbf{U}_{,i} \cdot \mathbf{U}_{,j}), \tag{11.35}$$

where $\mathbf{g}_i$ is a covariant basis.

The first and second order infinitesimal variations are given by:

$$\delta e_{ij} = \frac{1}{2}(\mathbf{g}_i \cdot \delta\mathbf{U}_{,j} + \mathbf{g}_j \cdot \delta\mathbf{U}_{,i} + \mathbf{U}_{,i} \cdot \delta\mathbf{U}_{,j} + \mathbf{U}_{,j} \cdot \delta\mathbf{U}_{,i}),$$

$$d\delta e_{ij} = \frac{1}{2}(d\mathbf{U}_{,i} \cdot \delta\mathbf{U}_{,j} + d\mathbf{U}_{,j} \cdot \delta\mathbf{U}_{,i}).$$

At each time step, in the Newmark algorithm a nonlinear problem has to be solved. The bilinear form appearing in this algorithm is the following:

$$A = A^L + A^{NL},$$

with

$$A^L(d\mathbf{U}, \delta\mathbf{U}) = \int_\Omega \frac{\partial^2 \mathcal{W}}{\partial e_{ij} \partial e_{kl}} de_{kl} \delta e_{ij} \, dV, \tag{11.36}$$

$$A^{NL}(d\mathbf{U}, \delta\mathbf{U}) = \int_\Omega \frac{\partial \mathcal{W}}{\partial e_{ij}} d\delta e_{ij} \, dV, \tag{11.37}$$

and the corresponding non-linear right hand side

$$F^{NL}(\delta\mathbf{U}) = \int_\Omega \frac{\partial \mathcal{W}}{\partial e_{ij}} \delta e_{ij} \, dV. \tag{11.38}$$

In practice, the values of the deformation are not directly computed by (11.35), but are re-interpolated at the tying points defined by MITC methods. The first and second order infinitesimal

variations in (11.36)–(11.38) have to be re-interpolated using the same rules in order to obtain a consistent tangent problem.

Both the MITC4 and the MI3D elements can be employed in actual computations. The MITC4 with 4 nodes and 5 degrees of freedom per node has 20 dof per element, the MI3D with 27 nodes and 3 dof per node has 81 dof per element. The MI3D is indeed more expensive than the MITC4 but it is more convenient for realistic models of the arteries which consider the different physiological layers (the intima, media and adventitia).

11.7 Preliminary numerical results

As a first benchmark test (see [87]), we consider a simplified version of the coupled problem on a simple geometry, and we briefly compare the results with the nonlinear Domain Decomposition method (II). Let Ω be a cylinder of radius $R_0 = 0.5cm$ and of length $L = 5cm$. We use a coarse mesh, made of 4800 tetraedra (3916 dof) for the fluid, and 40 hexaedra (1584 dof) for the structure. We set $\Delta t = 10^{-4}s$. The structure is modeled by 3D-shell elements (see [49–51]) with a Saint Venant-Kirchhoff constitutive law and the physical parameters $E = 3 \cdot 10^6 dynes/cm^2$, $\nu = 0.3$ and $\rho_s = 1.2g/cm^3$. The thickness of the shell is $h = 0.1cm$. In particular we use a $Q2$-finite element with 27 nodes combined with a MITC interpolation rule in the thin direction of the hexaedra. These elements allow us to easily use three dimensional constitutive laws while keeping a reasonable cost (only one layer of elements is necessary). The fluid we consider is driven by the Navier-Stokes equation on a fixed domain (in particular we skip the ALE formulation in this simple test), with $\mu = 0.03poise$ and $\rho_f = 1.g/cm^2$. Initially, the fluid is at rest and an over pressure of $1.3332 \cdot 10^4 dynes/cm^2$ ($10mmHg$) has been imposed at the inlet for $0.005s$. As expected, a propagation of the pressure wave is observed and comparable with more classical shell elements.

On Figure 11.5, we plot the deformation of the structure at times 4, 8 and $13ms$.

The same benchmark test has been solved by using method (II) (with the same tolerance for the Newton algorithms and GMRES). The comparison between both methods for a Dirichlet-Neumann preconditioner is reported on Table 11.2. So far, the two methods seem to compete equally in terms of cost.

Method	(I)	(II)
overall CPU time	74 min	81 min
Average number of GMRES iterations	8	9
Average number of Newton iterations	3.3	1.7

Table 11.2. Preliminary numerical results, 150 time steps on the coarse mesh

11.8 Conclusion

We have proposed a Newton algorithm for fluid-structure problems. The starting point of the method is the same as for the so-called monolithic approaches since we consider the global fluid-structure equations, but the tangent problem is solved with domain decomposition techniques. The resulting method is therefore partitioned: it is based on two different solvers for the fluid and the structures and can be parallelized. A simplified complexity analysis shows in which cases the proposed method can be competitive. Further numerical simulations and comparison with existing methods will be presented in a forthcoming work [A2].

Besides blood flow in arteries, this numerical method may significantly improve the efficiency of fluid-structure interaction computations when the structure involves a multiscale modeling, such

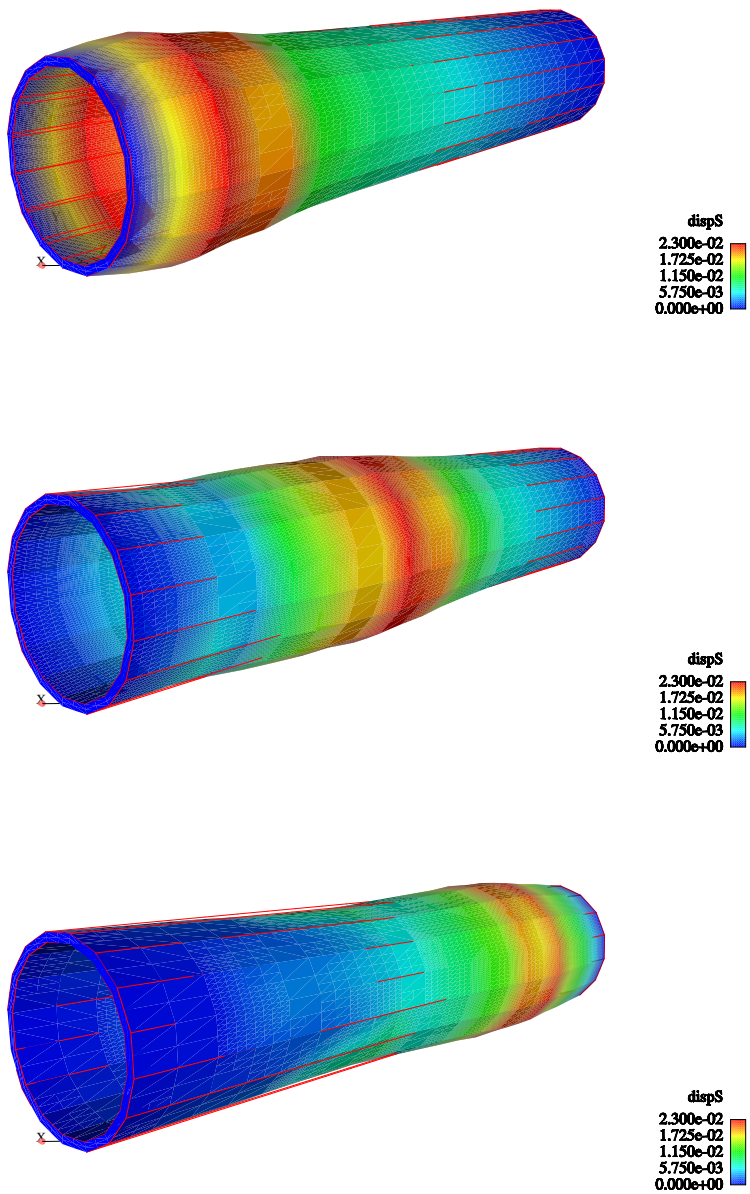

Fig. 11.5. Deformation of the structure (magnified by 3) at time 4, 8 and 13ms. Note that the structure is made of one layer of 3D-shell elements.

as the homogenization model of Section 5. In this case, the resolution of the structure sub-problem may be so much more expensive than the fluid that one only sees the cost of the structure in the coupled problem. Reducing the number of resolutions of the structure problem (or more precisely the number of stiffness matrix constructions) allows then to drastically reduce the overall cost of the fluid-homogenized structure problem.

Bibliographie

[1] A. Abdulle. On a priori error analysis of fully discrete heterogeneous multiscale FEM. *Multiscale Model. Simul.*, 4:447–459, 2005.

[2] E. Acerbi and N. Fusco. Semicontinuity problems in the calculus of variations. *Arch. Rational Mech. Anal.*, 86(2):125–145, 1984.

[3] R. Alicandro, A. Braides, and M. Cicalese. Phase and anti-phase boundaries in binary discrete systems: a variational viewpoint. *Netw. Heterog. Media.*, 1:85–107, 2006.

[4] R. Alicandro and M. Cicalese. A general integral representation result for the continuum limits of discrete energies with superlinear growth. *SIAM J. Math. Anal.*, 36(1):1–37, 2004.

[5] R. Alicandro and M. Cicalese. In preparation.

[6] G. Allaire. Homogenization and two-scale convergence. *SIAM J. Math. Anal.*, 23:1482–1518, 1992.

[7] G. Allaire. *Shape optimization by the homogenization method*, volume 146 of *Applied mathematical sciences*. Springer-Verlag, New York, 2000.

[8] G. Allaire and R. Brizzi. A multiscale finite element method for numerical homogenization. *Multiscale Model. Simul.*, 4:790–812, 2005.

[9] G. Allaire and G. Francfort. Existence of minimizers for non-quasiconvex functionals arising in optimal design. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 15:301–339, 1998.

[10] E.M. Arruda and M.C. Boyce. A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. *Journal of the Mechanics and Physics of Solids*, 41:389–412, 1993.

[11] E.M. Arruda and M.C. Boyce. Constitutive models of rubber elasticity : a review. *Rubber Chemistry and Technology*, 72:504–523, 2000.

[12] J.-F. Babadjian, M. Baía, and P. Santos. Characterization of two-scale young-measures and applications to homogenization. submitted.

[13] M. Baía and I. Fonseca. The limit behavior of a family of variational multiscale problems. *Indiana Univ. Math. J.*, to appear.

[14] J.M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Arch. Rat. Mech. Anal.*, 63:337–403, 1977.

[15] J.M. Ball. A version of the fundamental theorem for Young measures. In *PDEs and continuum models of phase transitions (Nice 1988)*, volume 344 of *Lecture Notes in Phys.*, pages 207–215. Springer, Berlin, 1980.

[16] M. Barchiesi. Loss of polyconvexity by homogenization: a new example. *Calc. Var. Partial Differential Equations*, 2007.

[17] K.J. Bathe and D. Chapelle. *The finite element analysis of shells—fundamentals*. Computational Fluid and Solid Mechanics. Springer-Verlag, Berlin, 2003.

[18] K.J. Bathe and H. Zhang. Finite element developments for general fluid flows with structural interactions. *Int. J. Num. Meth. Engng.*, 2004.

[19] J.T. Batina. Unsteady Euler airfoil solutions using unstructured dynamic meshes. *AIAA J.*, 28(8):1381–1388, 1990.

[20] R.J. Baxter. *Exactly solved models in statistical mechanics*. Academic Press, Inc., London, 1982.

[21] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1978.

[22] A. Bensoussan, J.L. Lions, and G. Papanicolaou. Boundary layer analysis in homogenization of diffusion equations with dirichlet conditions in the half space. Wiley, 1976.

[23] J.S. Bergström and M.C. Boyce. Mechanical behavior of particle filled elastomers. *Rubber Chemistry and Technology*, (72):633–656, 1999.

[24] X. Blanc, C. Le Bris, and P.L. Lions. From molecular models to continuum mechanics. *Arch. Rational Mech. Anal.*, 164(4):341–381, 2002.

[25] X. Blanc, C. Le Bris, and P.L. Lions. Du discret au continu pour des réseaux aléatoires d'atomes. *C. R. Acad. Sci. Paris, Série I*, 342:627–633, 2006.

[26] X. Blanc, C. Le Bris, and P.L. Lions. The energy of some microscopic stochastic lattices. *Arch. Rational Mech. Anal.*, to appear.

[27] M. Böl and S. Reese. Finite element modelling of polymer networks based on chain statistics. In J. Busfield and A. Muhr, editors, *Constitutive models for rubber III (Proceedings of the third European conference on constitutive models for rubber)*, pages 203–211, Lisse, 2003. A.A. Balkema.

[28] M. Böl and S. Reese. On a micromechanically-based finite element simulation of the viscoelasic and damage behaviour of rubber-like polymers. In J. Busfield and A. Muhr, editors, *Constitutive models for rubber III (Proceedings of the third European conference on constitutive models for rubber)*, pages 213–220, Lisse, 2003. A.A. Balkema.

[29] M. Böl and S. Reese. Finite element modelling of rubber-like materials - a comparison between simulation and experiment. *Journal of Materials Sc.*, 40:5933–5939, 2005.

[30] M. Böl and S. Reese. New method for simulation of Mullins effect using finite element method. *Plastics, Rubber and Composites*, 34(8):343–348, 2005.

[31] M. Böl and S. Reese. Finite element modelling of rubber-like polymers based on chain statistics. *Int. J. Sol. Struc.*, 43:2–26, 2006.

[32] G. Bouchitte and I. Fragalà. Homogenization of thin structures by two-scale method with respect to measures. *SIAM J. Math. Anal.*, 32:1198–1226, 2001.

[33] A. Braides. Homogenization of some almost periodic functionals. *Rend. Accad. Naz. Sci. XL*, 103:261–281, 1985.

[34] A. Braides. Loss of polyconvexity by homogenization. *Arch. Rat. Mech. Anal.*, 127:183–190, 1994.

[35] A. Braides. On local variational limits of discrete systems. *Comm. Contemporary Math.*, 2:285–297, 2000.

[36] A. Braides. *Γ-convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and Its Applications*. Oxford University Press, 2002.

[37] A. Braides. A handbook of Γ-convergence. volume 3 of *Handbook of Differential Equations: Stationary Partial Differential Equations*, pages 101–213. Elsevier, Amsterdam, 2006.

[38] A. Braides and A. Defranceschi. *Homogenization of Multiple Integrals*, volume 12 of *Oxford Lecture Series in Mathematics and Its Applications*. Oxford University Press, 1998.

[39] A. Braides and G.A. Francfort. Bounds on the effective behaviour of a square conducting lattice. *Proc. R. Soc. Lond. A*, 460:1755–1769, 2004.

[40] A. Braides and M.S. Gelli. Continuum limits of discrete systems without convexity hypotheses. *Math. Mech. Solids*, 6:395–414, 2002.

[41] A. Braides and M.S. Gelli. Limits of discrete systems with long range interactions. *J. Convex Anal.*, 9:363–299, 2002.

[42] A. Braides, M.S. Gelli, and M. Sigalotti. The passage from non-convex discrete systems to variational problems in Sobolev spaces: the one-dimensional case. *Proc. Steklov Inst. Math.*, 236:395–414, 2002.

[43] A. Braides and A. Piatnitski. Overall properties of a discrete membrane with randomly distributed defects. 2006. preprint CVGMT, *http://cvgmt.sns.it/papers/brapia04/*.

[44] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*. New York - Berlin, 1991.

[45] F. Brezzi, D. Marini, and E. Süli. Residual-free bubbles for advection-diffusion problems: the general error analysis. *Numerische Mathematik*, 85(1), 2000.

[46] G. Buttazzo. *Integral representation theory for some classes of local functions*, volume 244 of *Pitman Res. Notes Math. Ser.* Longman Sci. Tech., 1992.

[47] C. Carstensen and Dolzmann G. An a priori error estimate for finite element discretizations in nonlinear elasticity for polyconvex materials under small loads. *Numerische Mathematik*, 97:67–80, 2004.

[48] P. Causin, J.-F. Gerbeau, and F. Nobile. Added-mass effect in the design of partitioned algorithms for fluid-structure problems. *Comp. Meth. Appl. Mech. Engng.*, 194(42–44):4506–4527, 2005.

[49] D. Chapelle and A. Ferent. Modeling of the inclusion of a reinforcing sheet within a 3D medium. *Math. Models Methods Appl. Sci.*, 13(4):573–595, 2003.

[50] D. Chapelle, A. Ferent, and K.J. Bathe. 3D-shell elements and their underlying mathematical model. *Math. Models Methods Appl. Sci.*, 14(1):105–142, 2004.

[51] D. Chapelle, A. Ferent, and P. Le Tallec. The treatment of "pinching locking" in 3D-shell elements. *M2AN Math. Model. Numer. Anal.*, 37(1):143–158, 2003.

[52] P.G. Ciarlet. *Mathematical elasticity. Vol. I*, volume 20 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1988.

[53] D. Cioranescu, A. Damlamian, and R. De Arcangelis. Homogenization of quasiconvex integrals via the periodic unfolding method. *SIAM J. Math. Anal.*, 36(5):1435–1453, 2006.

[54] D. Cioranescu and P. Donato. *An introduction to homogenization*, volume 17 of *Oxford Lecture Series in Mathematics and Its Applications*. Oxford University Press, 1999.

[55] S. Conti, A. De Simone, S. Müller, and F. Otto. Multiscale modeling of materials—the role of analysis. In *Trends in nonlinear analysis*, pages 375–408, Berlin, 2003. Springer.

[56] G. Dal Maso. *An Introduction to Γ-Convergence*. Birkhäuser Boston, Boston, MA, 1993.

[57] G. Dal Maso, A. De Simone, M.G. Mora, and M. Morini. Time-dependent systems of generalized young measures. *Netw. Heterogeneous Media*, 2(1):1–36, 2007.

[58] G. Dal Maso and A. Defranceschi. Correctors for homogenization of monotone operators. *Differential Integral Equations*, 3:1151–1166, 1990.

[59] G. Dal Maso and L. Modica. Nonlinear stochastic homogenization and ergodic theory. *J. Reine Angew. Math.*, 368:28–42, 1986.

[60] G. Dal Maso, M. Negri, and D. Percivale. Linearized elasticity as Γ-limit of finite elasticity. *Set-Valued Anal.*, 10(12):165–183, 2002.

[61] M. Danielsson, D.M. Parks, and M.C. Boyce. Constitutive modeling of porous hyperelastic materials. *Mechanics of Materials*, 36:347–358, 2004.

[62] A. De Simone, R.V. Kohn, S. Müller, and F. Otto. Magnetic microstructures—a paradigm of multiscale problems. In *ICIAM 99 (Edinburgh)*, pages 175–190, Oxford, 2000. Oxford Univ. Press.

[63] S. Deparis. *Numerical Analysis of Axisymmetric Flows and Methods for Fluid-Structure Interaction Arising in Blood Flow Simulation*. PhD thesis, EPFL, Switzerland, 2004.

[64] S. Deparis, M. Discacciati, G. Fourestey, and A. Quarteroni. Fluid-structure algorithms based on Steklov-Poincaré operators. *Comput. Methods Appl. Mech. Engrg.*, 195(41-43):5797–5812, 2006.

[65] S. Deparis, M. Discacciati, and A. Quarteroni. A domain decomposition framework for fluid-structure interaction problems. In *Proceedings of the Third International Conference on Computational Fluid Dynamics (ICCFD3)*, 2004.

[66] W. Dettmer and D. Perić. A computational framework for fluid-structure interaction: Finite element formulation and applications. *Comp. Meth. Appl. Mech. Engrg.*, 195(41-43):5754–5779, 2006.

[67] J. Donéa, S. Giuliani, and J. P. Halleux. An arbitrary Lagrangian-Eulerian finite element method for transient dynamic fluid-structure interactions. *Comp. Meth. Appl. Mech. Engng.*, pages 689–723, 1982.

[68] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden. Heterogeneous multiscale methods: A review. *Commun. Comput. Phys.*, 2:367–450, 2007.

[69] W. E, P.B. Ming, and P.W. Zhang. Analysis of the heterogeneous multiscale method for elliptic homogenization problems. *J. Amer. Math. Soc.*, 18:121–156, 2005.

[70] C. Ebmeyer and W.B. Liu. Quasi-norm interpolation error estimates for the piecewise linear finite element approximation of p-Laplacian problems. *Numer. Math.*, 100:223–258, 2005.

[71] C. Ebmeyer, W.B. Liu, and M. Steinhauer. Global regularity in fractional order sobolev spaces for the p-Laplace equation on polyhedral domains. *Zeit. Anal. Anwend.*, 24:353–374, 2005.

[72] Y.R. Efendiev, T.Y. Hou, and V. Ginting. Multiscale finite element methods for nonlinear problems and their applications. *Com. Math. Sc.*, 2(4):553–589, 2004.

[73] Y.R. Efendiev, T.Y. Hou, and X.H. Wu. Convergence of a nonconforming multiscale finite element method. *SIAM J. Num. Anal.*, 37:888–910, 2000.

[74] Y.R. Efendiev and A. Pankov. Numerical homogenization of monotone elliptic operators. *Multiscale Model. Simul.*, 2(1):62–79, 2003.

[75] Y.R. Efendiev and A. Pankov. Numerical homogenization and correctors for nonlinear elliptic equations. *SIAM J. Appl Math.*, 65(1):43–68, 2004.

[76] Y.R. Efendiev and A. Pankov. Numerical homogenization of nonlinear random parabolic operators. *Multiscale Model. Simul.*, 2:237–268, 2004.

[77] C. Farhat, K. van der Zee, and Ph. Geuzaine. Provably second-order time-accurate loosely-coupled solution algorithms for transient nonlinear aeroelasticity. *Comp. Meth. Appl. Mech. Engng.*, 195(17-18):1973–2001, 2006.

[78] M. A. Fernández and M. Moubachir. Numerical simulation of fluid-structure systems via Newton's method with exact Jacobians. In P Neittaanmäki, T. Rossi, K. Majava, and O. Pironneau, editors, *4th European Congress on Computational Methods in Applied Sciences and Engineering*, volume 1, Jyväskylä, Finland, July 2004.

[79] M.A. Fernández and J.-F. Gerbeau. Méthodes numériques en interaction fluide-structure. Notes de cours de master M2, université Paris 6.

[80] M.A. Fernández, J.-F. Gerbeau, and C. Grandmont. A projection algorithm for fluid-structure interaction problems with strong added-mass effect. *C. R. Acad. Sci. Paris, Math.*, 342:279–284, 2006.

[81] M.A. Fernández, J.-F. Gerbeau, and C. Grandmont. A projection semi-implicit scheme for the coupling of an elastic structure with an incompressible fluid. Int. J. Num. Meth. Engng. in press, 2006.

[82] M.A. Fernández and M. Moubachir. An exact block-Newton algorithm for solving fluid-structure interaction problems. *C. R. Math. Acad. Sci. Paris*, 336(8):681–686, 2003.

[83] M.A. Fernández and M. Moubachir. An exact block-newton algorithm for the solution of implicit time discretized coupled systems involved in fluid-structure interaction problems. In K.J. Bathe, editor, *Second M.I.T. Conference on Computational Fluid and Solid Mechanics*, pages 1337–1341. Elsevier, 2003.

[84] M.A. Fernández and M. Moubachir. A Newton method using exact Jacobians for solving fluid-structure coupling. *Comp. & Struct.*, 83:127–142, 2005.

[85] F. Féyel. Multiscale FE^2 elastoviscoplastic analysis of composite structures. *Comp. Mat. Sci.*, 16:344–354, 1999.

[86] I. Fonseca, S. Müller, and P. Pedregal. Analysis of concentration and oscillations effects generated by gradients. *SIAM J. Math. Anal.*, 29:736–756, 1998.

[87] L. Formaggia, J.-F. Gerbeau, F. Nobile, and A. Quarteroni. On the coupling of 3D and 1D Navier-Stokes equations for flow problems in compliant vessels. *Comp. Meth. Appl. Mech. Engrg.*, 191(6-7):561–582, 2001.

[88] FreeFEM. http://www.freefem.org/.

[89] G. Friesecke, R.D. James, and S. Müller. A hierarchy of plate models derived from nonlinear elasticity by gamma-convergence. *Arch. Rat. Mech. Anal.*, 180:183–236, 2006.

[90] G. Friesecke and F. Theil. Validity and failure of the Cauchy-Born rule. *J. Nonlin. Sci.*, 12:445–478, 2002.

[91] J.-F. Gerbeau and M. Vidrascu. A quasi-Newton algorithm based on a reduced model for fluid-structure interactions problems in blood flows. *Math. Model. Num. Anal.*, 37(4):631–648, 2003.

[92] J.-F. Gerbeau, M. Vidrascu, and P. Frey. Fluid-structure interaction in blood flows on geometries based on medical imaging. *Comp. & Struct.*, 83(2-3):155–165, 2005.

[93] J.F. Gerbeau, P. Hauret, P. Le Tallec, and M. Vidrascu. Fluid structure interaction problems in large deformation. *C. R. Acad. Sci. Paris, Série II*, 333(12):91O–922, 2005.

[94] G. Geymonat, S. Müller, and N. Triantafyllidis. Homogenization of nonlinearly elastic materials, microscopic bifurcation and macroscopic loss of rank-one convexity. *Arch. Rat. Mech. Anal.*, 122:231–290, 1993.

[95] M. Giaquinta. *Multiple integrals in the calculus of variations and nonlinear elliptic systems*, volume 105 of *Annals of Mathematics Studies*. Princeton University Press, 1983.

[96] M. Giaquinta and E. Giusti. On the regularity of minimizers of variational integrals. *Acta Math.*, 148:31–46, 1982.

[97] L.V. Gibiansky and O. Sigmund. Multiphase composites with extremal bulk modulus. *J. Mech. Phys. Solids*, 48:461–498, 2000.

[98] A. Giuliani, J.L. Lebowitz, and E.H. Lieb. Ising models with long-range dipolar and short range ferromagnetic interactions. *Phys. Rev. B*, 74:064420, 2006.

[99] Z. Hashin and S. Shtrikman. A variational approach to the theory of the elastic behaviour of multiphase materials. *J. Mech. Phys. Solids*, VII:127–140, 1963.

[100] P. Hauret and P. Le Tallec. Energy-controlling time integration methods for nonlinear elastodynamics and low-velocity impact. *Comp. Meth. Appl. Mech. Eng.*, 195:4890–4916, 2006.

[101] M. Heil. An efficient solver for the fully coupled solution of large-displacement fluid-structure interaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193(1-2):1–23, 2004.

[102] R. Hill. A self-consistent mechanics of composite materials. *J. Mech. Phys. Solids*, 12(4):213, 1965.

[103] T. Horiguchi. Ising models with two-spin interactions and three-spin interactions on a square lattice. *Phys. A*, 136(1):109–123, 1986.

[104] T.Y. Hou. Numerical approximations to multiscale solutions in partial differential equations. In Craig A.W. Blowey, J.F. and T. Shardlow, editors, *Frontier in Numerical Analysis*, pages 241–302. Springer Publications, 2003.

[105] T.Y. Hou and X.H Wu. A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys.*, 134:169–189, 1997.

[106] T.Y. Hou, X.H. Wu, and Z.Q. Cai. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. *Math. Comput.*, 68:913–943, 1999.

[107] T.Y. Hou, X.H. Wu, and Y. Zhang. Removing the cell resonance error in the multiscale finite element method via a petrov–galerkin formulation. *Commun. Math. Sci.*, 2:185–205, 2004.

[108] T.Y. Hou, X.H. Wu, and Y. Zhang. Removing the cell resonance error in the multiscale finite element method via a Petrov-Galerkin formulation. *Comm. in Math. Sci.*, 2(2):185–205, 2004.

[109] B. Hübner, E. Walhorn, and D. Dinkle. A monolithic approach to fluid-structure interaction using space-time finite elements. *Comp. Meth. Appl. Mech. Engng.*, 193:2087–2104, 2004.

[110] O. Iosifescu, C. Licht, and G. Michaille. Variational limit of a one-dimensional discrete and statistically homogeneous system of material points. *C. R. Math. Acad. Sci. Paris*, 32:575–580, 2001.

[111] O. Iosifescu, C. Licht, and G. Michaille. Variational limit of a one-dimensional discrete and statistically homogeneous system of material points. *Asymptot. Anal.*, 28:309–329, 2001.

[112] P. Jannasch. Porous polymaleimide networks. *J. Mater. Chem.*, 11:2303–2306, 2001.

[113] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, 1994.

[114] R.V. Kohn. The relaxation of a double-well energy. *Contin. Mech. Thermodyn.*, 3(3):193–236, 1991.

[115] R.V. Kohn. Energy-driven pattern formation. In *Proceedings of ICM2006*, to appear.
[116] R.V. Kohn and G. Strang. Optimal design and relaxation of variational problems, I, II and III. *Comm. Pure Appl. Math.*, 39, 1986.
[117] M.A. Kranosl'skii. *Topological Methods in the Theory of Nonlinear Integral Equations*. Pergamon Press, New York, 1964.
[118] M. Laroussi, K. Sab, and A. Alaoui. Foam mechanics: nonlinear response of an elastic 3D-periodic microstructure. *Int. J. of Sol. and Str.*, 39:3599–3623, 2002.
[119] C. Le Bris. *Systèmes multi-échelles - Modélisation et simulation*, volume 47 of *Mathématiques & Applications*. Springer-Verlag, Berlin, 2005.
[120] H. Le Dret and A. Raoult. The nonlinear membrane model as variational limit of nonlinear three-dimensional elasticity. *J. Math. Pures Appl.*, 74(6):549–578, 1995.
[121] H. Le Dret and A. Raoult. The quasiconvex envelope of the Saint Venant-Kirchhoff stored energy function. *Proc. Roy. Soc. Edinburgh Sect. A*, 125(6):1179–1192, 1995.
[122] H. Le Dret and A. Raoult. The membrane shell model in nonlinear elasticity: a variational asymptotic derivation. *J. Nonlinear Science*, 6:59–84, 1996.
[123] P. Le Tallec. Existence and approximation results for nonlinear mixed problems: application to incompressible finite elasticity. *Numer. Math.*, 38(3):365–382, 1981/82.
[124] P. Le Tallec. Domain decomposition methods in computational mechanics. In *Computational Mechanics Advances, Vol. 1, no.2*, pages 123–217. North Holland, 1994.
[125] P. Le Tallec. Numerical methods for nonlinear three-dimensional elasticity. In *Handbook of numerical analysis, Vol. III*, pages 465–622. North-Holland, 1994.
[126] P. Le Tallec. *Introduction à la dynamique des structures*. Ellipses, 2000.
[127] P. Le Tallec. *Mécanique des Milieux Continus*. Editions de l'École polytechnique, 2006.
[128] P. Le Tallec and J. Mouro. Fluid structure interaction with large structural displacements. *Comput. Meth. Appl. Mech. Engrg.*, 190:3039–3067, 2001.
[129] J.L. Lebowitz and E. Presutti. Statistical mechanics of systems of unbounded spins. *Com. Math. Phys.*, 50 and 78:195–218 and 151, 1976 and 1980.
[130] J.-L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaires*. Dunod, 1969.
[131] K. Lurie and A.V. Cherkaev. Exact estimates of conductivity of mixtures composed of two materials taken in prescribed proportion (plane problem). *Doklady Akademii Nauk SSSR*, 264:1129–1130, 1982.
[132] A. Malakis. Two- and three-spin triangular ising model: variational approximations. *J. Stat. Phys.*, 27(1):1–17, 1982.
[133] A.-M. Matache and C. Schwab. Two-scale FEM for homogenization problems. *M2AN*, 36:537–572, 2002.
[134] N. Meyers and A. Elcrat. Some results on regularity for solutions of nonlinear elliptic systems and quasiregular functions. *Duke Math.*, 42:121–136, 1975.
[135] G.W. Milton. *The Theory of Composites*, volume 6 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2002.
[136] P.B. Ming and X.Y. Yue. Numerical methods for multiscale elliptic problems. *J. Comput. Phys.*, 214:421–445, 2006.
[137] Modulef. http://www-rocq.inria.fr/modulef/english.html.
[138] D. P. Mok and W. A. Wall. Partitioned analysis schemes for the transient interaction of incompressible flows and nonlinear flexible structures. In K. Schweizerhof W.A. Wall, K.U. Bletzinger, editor, *Trends in computational structural mechanics*, Barcelona, 2001. CIMNE.
[139] D. P. Mok, W. A. Wall, and E. Ramm. Partitioned analysis approach for the transient, coupled response of viscous fluids and flexible structures. In W. Wunderlich, editor, *Proceedings of the European Conference on Computational Mechanics*. ECCM'99, TU Munich, 1999.
[140] D. P. Mok, W. A. Wall, and E. Ramm. Accelerated iterative substructuring schemes for instationary fluid-structure interaction. In K.J. Bathe, editor, *Computational Fluid and Solid Mechanics*, pages 1325–1328. Elsevier, 2001.

[141] C.B. Morrey. Quasiconvexity and the semicontinuity of multiple integrals. *Pacific J. Math.*, 2:25–53, 1952.

[142] S. Moskow and M. Vogelius. First-order corrections to the homogenised eigenvalues of a periodic composite medium. a convergence proof. *Proc. Roy. Soc. Edinburgh Sect. A*, 125(6):1263–1299, 1997.

[143] H. Moulinec and P. Suquet. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Comput. Methods Appl. Mech. Engrg.*, 157:69–94, 1998.

[144] J. Mouro. *Interactions fluide structure en grands déplacements. Résolution numérique et application aux composants hydrauliques automobiles.* PhD thesis, Ecole Polytechnique, 1994.

[145] S. Müller. Homogenization of nonconvex integral functionals and cellular elastic materials. *Arch. Rat. Anal. Mech.*, 99:189–212, 1987.

[146] S. Müller. Variational models for microstructure and phase transitions. In *Calculus of variations and geometric evolution problems (Cetraro, 1996)*, volume 1713 of *Lecture Notes in Math.*, pages 85–210. Springer, Berlin, 1999.

[147] F. Murat. Unpublished.

[148] F. Murat. H-convergence. Séminaire d'Analyse fonctionnelle et numérique, Univ. Alger, multigraphié, 1978.

[149] F. Murat and L. Tartar. H-convergence. In A.V. Cherkaev and R.V. Kohn, editors, *Topics in the Mathematical Modelling of Composites Materials*, volume 31 of *Progress in nonlinear differential equations and their applications*, pages 21–44. Birkhäuser, 1997.

[150] G. Nguetseng. A general convergence result for a functional related to the theory of homogenization. *SIAM J. Math. Anal.*, 20:608–629, 1989.

[151] J.A. Nitsche. On Korn's second inequality. *RAIRO Analyse numérique*, 15:237–248, 1981.

[152] F. Nobile. *Numerical approximation of fluid-structure interaction problems with application to haemodynamics.* PhD thesis, EPFL, Switzerland, 2001.

[153] M. Ostoja-Starzewski. Lattice models in micromechanics. *Appl. Mech. Rev.*, 55(1):35–60, 2002.

[154] M. Ostoja-Starzewski. Material spatial randomness—from statistical to representative volume element. *Prob. Eng. Mech.*, 21:112–132, 2006.

[155] A. Pankov. *G-convergence and Homogenization of Nonlinear Partial Differential Operators.* Kluwer Academic Publishers, Dortrecht, The Netherlands, 1997.

[156] O. Pantz. Une justification partielle du modèle de plaque en flexion par Γ-convergence. *C. R. Acad. Sci. Paris S´er. I*, 332:587–592, 2001.

[157] O. Pantz. On the justification of the nonlinear inextensional plate model. *Arch. Rational Mech. Anal.*, 167:179–209, 2003.

[158] I. Paris. *Robustesse des éléments finis triangulaires de coque.* PhD thesis, université de Paris VI and INRIA, 2006.

[159] G.A. Pavliotis and A.M. Stuart. *Multiscale methods: averaging and homogenization.* 2007. Lecture notes given at MSRI Berkeley - April, 2007.

[160] P. Pedregal. On the numerical analysis of non-convex variational problems. *Numer. Math.*, 74:325–336, 1996.

[161] P. Pedregal. *Parametrized measures and variational principles.* Birkhäuser, 1997.

[162] S. Piperno, C. Farhat, and B. Larrouturou. Partitioned procedures for the transient solution of coupled aeroelastic problems. Part I: Model problem, theory and two-dimensional application. *Comp. Meth. Appl. Mech. Engrg.*, 124:79–112, 1995.

[163] A. Quarteroni and A. Valli. *Domain decomposition methods for partial differential equations.* Oxford University Press, 2000.

[164] U. Raitums. On the local representation of G-closure. *Arch. Ration. Mech. Anal.*, 158:213–234, 2001.

[165] A. Raoult. Nonpolyconvexity of the stored energy function of a Saint Venant-Kirchhoff material. *Apl. Mat.*, 31(6):417–419, 1986.

[166] D. Ruelle. *Statistical mechanics. Rigorous results.* World Scientific Publishing Co., Inc., River Edge, NJ; Imperial College Press, London, 1999. Reprint of the 1989 edition.

[167] S. Rugonyi and K.J. Bathe. On finite element analysis of fluid flows coupled with structural interaction. *CMES - Comp. Modeling Eng. Sci.*, 2(2):195–212, 2001.

[168] G. Sangalli. Capturing small scales in elliptic problems using a residual-free bubbles finite element method. *Multiscale Model. Simul.*, 1(3):485–503, 2003.

[169] G. Savaré. Regularity results for elliptic equations in Lipschitz domains. *Journal of Functional Analysis*, 152:176–201, 1998.

[170] O. Sigmund. A new class of extremal composites. *J. Mech. Phys. Solids*, 48:397–428, 2000.

[171] L. Tartar. Estimations de coefficients homogénéisés. In *Computing methods in applied sciences and engineering (Proc. Third Internat. Sympos., Versailles, 1977)*, volume 704 of *Lecture Notes in Math.*, pages 364–373, Berlin, 1979. Springer.

[172] L. Tartar. Estimations fines de coefficients homogénéisés. In P. Krée, editor, *Ennio De Giorgi Colloquium*, volume 125 of *Reseach Notes in Mathematics*, pages 168–187, London, 1985. Pitman.

[173] T.E. Tezduyar. Finite element methods for fluid dynamics with moving boundaries and interfaces. *Arch. Comput. Methods Engrg.*, 8:83–130, 2001.

[174] P.D. Thomas and C.K. Lombard. Geometric conservation law and its application to flow computations on moving grids. *AIAA J.*, 17(10):1030–1037, 1979.

[175] A. Toselli and O. Widlung. *Domain Decomposition Methods - Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer, 2005.

[176] L.R.G Treloar. *The Physics of Rubber Elasticity.* Oxford at the Clarendon Press, Oxford, 1949.

[177] M. Vidrascu. Solution of non-linear elasticity problems using the *continu* software. *Inria Report Research (http://www.inria.fr/rrrt/rr-4128.html)*, 2001.

[178] J. Vierendeels. Implicit coupling of partioned fluid-structure interaction solvers using reduced order models. In M. Schäfer H.J. Bungartz, editor, *Fluid-Structure interaction, Modelling, Simulation, Optimization*, pages 1–18. Springer, 2006.

[179] V. Šverák. New examples of quasiconvex functions. *Arch. Rational Mech. Anal.*, 119:293–300, 1992.

[180] H. Zhang, X. Zhang, S. Ji, G. Guo, Y. Ledezma, N. Elabbasi, and H. deCougny. Recent development of fluid-structure interaction capabilities in the ADINA system. *Computers & Structures*, 81(8-11):1071–1085, 2003.

[181] K. Zhang. Energy minimizers in nonlinear elastostatics and the implicit function theorem. *Arch. Rat. Mech. Anal.*, 114:95–117, 1991.