Multi-scale Finite Element Method for incompressible flows in heterogeneous media : Implementation and Convergence analysis.

Loïc Balazi Atchy Nillama

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2024IPPAX053

Thèse de doctorat

ÉCOLE POLYTECHNIQUE

IP PARIS

Multi-scale Finite Element Method for incompressible flows in heterogeneous media: Implementation and Convergence analysis

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École polytechnique

École doctorale n°574 École doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat: Mathématiques appliquées

Thèse présentée et soutenue à Gif-sur-Yvette, le 16 septembre 2024, par

LOÏC BALAZI

Composition du Jury :

Alexandre Ern Professeur, École nationale des ponts et chaussées	Président
Frédéric Legoll Professeur, École nationale des ponts et chaussées	Rapporteur
Alexei Lozinski Professeur, Université de Franche-Comté	Rapporteur
Sonia Fliss Professeure, ENSTA Paris	Examinatrice
Daniel Peterseim Professeur, Universität Augsburg	Examinateur
Maria Adela Puscas Ingénieure de Recherche, CEA	Examinatrice
Grégoire Allaire Professeur, École polytechnique	Directeur de thèse
Pascal Omnes Directeur de Recherche, CEA	Directeur de thèse

Abstract

This thesis is concerned with the application of a Multi-scale Finite Element Method (MsFEM) to solve incompressible flows in multi-scale media. Indeed, simulating the flow in a multi-scale media with numerous obstacles, such as nuclear reactor cores, is a highly challenging endeavour. In order to accurately capture the finest scales of the flow, it is necessary to use a very fine mesh. However, this often leads to intractable simulations due to the lack of computational resources. To address this limitation, this thesis develops an enriched non-conforming MsFEM to solve viscous incompressible flows in heterogeneous media, based on the classical non-conforming Crouzeix–Raviart finite element method with high-order weighting functions. The MsFEM employs a coarse mesh on which new basis functions are defined. These functions are not the classical polynomial basis functions of finite elements, but rather solve fluid mechanics equations on the elements of the coarse mesh. These functions are themselves numerically approximated on a fine mesh taking into account all the geometric details, which gives the multi-scale aspect of this method. A theoretical investigation of the proposed MsFEM is conducted at both the continuous and discrete levels. Firstly, the well-posedness of the discrete local problems involved in the MsFEM was demonstrated using new families of finite elements. To achieve this, a novel non-conforming finite element family in three dimensions on tetrahedra was developed. Furthermore, the first error estimate for the approximation of the Stokes problem in periodic perforated media using this MSFEM is derived, demonstrating its convergence. This is based on homogenization theory of the Stokes problem in periodic domains and on usual finite element theory. At the numerical level, the MsFEM to solve the Stokes and the Oseen problems in two and three dimensions is implemented in a massively parallel framework in FreeFEM. Furthermore, a methodology to solve the Navier–Stokes problem is provided.

Keywords: *Multi-scale Finite Element Method, Crouzeix–Raviart element, Navier–Stokes equations, Heterogeneous media, Numerical Analysis, High Performance Computing.*

Résumé

Cette thèse porte sur l'application d'une méthode d'éléments finis multi-échelles (MsFEM) pour résoudre les écoulements incompressibles dans des milieux multi-échelles. En effet, la simulation de l'écoulement dans un milieu multi-échelle comportant de nombreux obstacles, tel que le cœur d'un réacteur nucléaire, est un défi de taille. Afin de capturer avec précision les échelles les plus fines de l'écoulement, il est nécessaire d'utiliser un maillage très fin. Cependant, cela conduit souvent à des simulations difficiles à réaliser en raison du manque de ressources informatiques. Pour remédier à cette limitation, cette thèse développe une MsFEM non-conforme enrichie pour résoudre les écoulements visqueux incompressibles dans des milieux hétérogènes, basée sur la méthode classique des éléments finis non conformes de Crouzeix–Raviart avec des fonctions de poids d'ordre élevé. La MsFEM utilise un maillage grossier sur lequel de nouvelles fonctions de base sont définies. Ces fonctions ne sont pas les fonctions de base polynomiales classiques des éléments finis, mais résolvent les équations de la mécanique des fluides sur les éléments du maillage grossier. Ces fonctions sont elles-mêmes approximées numériquement sur un maillage fin en tenant compte de tous les détails géométriques, ce qui confère à cette méthode son aspect multi-échelle. Une étude théorique de la MsFEM proposée est menée aux niveaux continu et discret. Tout d'abord, le caractère bien posé des problèmes locaux discrets impliqués dans la MsFEM a été démontré à l'aide de nouvelles familles d'éléments finis. Pour ce faire, une nouvelle famille d'éléments finis non conformes en trois dimensions sur les tétraèdres a été développée. En outre, la première estimation d'erreur pour l'approximation du problème de Stokes dans des milieux perforés périodiques à l'aide de cette MSFEM est établie, démontrant sa convergence. Cette estimation est basée sur la théorie de l'homogénéisation du problème de Stokes dans les domaines périodiques et sur la théorie usuelle des éléments finis. Au niveau numérique, la MsFEM pour résoudre les problèmes de Stokes et d'Oseen en deux et trois dimensions a été implémenté dans un cadre massivement parallèle dans FreeFEM. En outre, une méthodologie pour résoudre le problème de Navier–Stokes est fournie.

Mots clés : *Méthode des Éléments Finis Multi-échelles, Élément de Crouzeix–Raviart, Équations de Navier–Stokes, Milieux hétérogènes, Analyse Numérique, Calcul Haute Performance.*

Acknowledgements

C'est le temps que tu as perdu pour ta rose qui fait ta rose si importante.

Antoine de Saint-Exupery
Le petit prince

Ce manuscrit, fruit de trois années de travail, ne reflète qu'une partie de l'expérience vécue. L'essentiel réside dans les moments partagés : les discussions devant un tableau blanc, les rires, les doutes, les joies, et les rencontres inoubliables. Je tiens donc à remercier ici toutes les personnes qui ont rendu ces trois années mémorables.

Je tiens tout d'abord à exprimer ma profonde gratitude à mes encadrants, Grégoire et Pascal. Pascal, je te remercie sincèrement de m'avoir offert l'opportunité de réaliser cette thèse et pour la confiance que tu m'as témoignée tout au long de ce parcours. Je te remercie pour ta bienveillance et toute la liberté que tu m'as laissée. Tes conseils avisés ont été d'une grande valeur, et je t'en suis profondément reconnaissant. Grégoire, je te remercie pour ta patience, ton aide précieuse et ta disponibilité constante. Ton implication indéfectible tout au long de la thèse a été un véritable pilier, que ce soit à travers les heures de calculs partagées devant un tableau blanc, les nombreuses idées que tu as apportées ou de tes nombreux conseils.

Je tiens ensuite à exprimer ma reconnaissance aux membres de mon jury. Tout d'abord, un grand merci à Frédéric et Alexei d'avoir accepté de consacrer une partie de votre été à rapporter ma thèse. Je vous suis également reconnaissant pour l'intérêt que vous avez manifesté à l'égard de mes travaux tout au long de ces trois années à chaque fois que nous avons eu l'occasion de nous rencontrer. C'est un honneur que ma thèse, qui a débuté avec vous lors d'un workshop à Besançon, se termine également en votre présence. Daniel, I would like to express my sincere gratitude to you for accepting to be part of my committee and for giving me the opportunity to continue my research journey in Augsburg. I am confident that I will have an enriching and enjoyable time working with you in Augsburg. Je remercie également Adela, avec qui ce fut un plaisir de discuter tout au long de ces trois années de thèse, Sonia Fliss et Alexandre Ern d'avoir accepté de faire partie de mon jury.

Je souhaite ensuite exprimer ma profonde gratitude à ceux sans qui cette thèse n'aurait pas été la même : Professeur Frédéric Hecht, Pierre Jolivet et Afeef Badri. Frédéric, je te remercie sincèrement pour ton aide précieuse dans l'implémentation des nouveaux éléments finis dans

FreeFEM. Ta gentillesse, ta patience et ta grande disponibilité ont été inestimables tout au long de ce projet. Je tiens également à te remercier pour ton accueil chaleureux lors des journées FreeFEM à Jussieu. Pierre, je te suis très reconnaissant pour m'avoir transmis ton expertise sur FreeFEM, PETSc ainsi que sur les méthodes de préconditionnement. Ton aide a été cruciale pour avancer dans mes travaux. Afeef, merci pour ton aide précieuse sur FreeFEM ainsi que pour le travail que tu as accompli pour l'installation de FreeFEM sur les supercalculateurs Orcus et Topaze. Ta contribution a été essentielle au bon déroulement de mes simulations, et je t'en suis très reconnaissant. Enfin, je n'oublie pas toute l'équipe SALOME, dont l'aide a été primordiale tout au long de ma thèse.

Je souhaite également remercier mes chefs de laboratoire, Maria Giovanna, et de service, Erwan, pour la grande liberté qu'ils m'ont accordée tout au long de ma thèse, que ce soit pour les missions, les vacations ou le temps passé au CMAP. Leur confiance m'a permis de travailler dans un cadre flexible et enrichissant. Je tiens aussi à exprimer ma gratitude à toutes les assistantes et tous les assistants, aussi bien au CEA qu'au CMAP, pour leur précieuse aide administrative tout au long de ce parcours.

Viennent ensuite les remerciements plus personnels.

Je souhaite tout particulièrement remercier Nathalie, ma "petite sœur" de thèse. Ces trois années n'auraient jamais été les mêmes sans toi. Je te remercie pour nos conférences ensemble, pour toutes les personnes que j'ai rencontrées grâce à toi et bien sûr pour mes heures (parfois mes journées) passées dans ton bureau. Merci aussi pour tes cours improvisés de libanais (Rah ehkiki lebnene shi nhar, wa3ed!) et même de photographie ! Nos discussions, nos sorties, les châteaux que nous avons visités, et tous les moments que j'ai pu partager avec toi (sans oublier les heures passées devant un tableau blanc à calculer des intégrales), restent parmi mes plus beaux souvenirs de thèse. Nathalie, da3mek, mousenadtek, rouhek lhelwe w woujoudek ken elon ahamiye kbire bi najeh hayde lthèse. Choukran 3a kel chi, mn jouwet albe! Rah tdalle ajmal hada lta2et fi.

Manon, je te remercie pour ton aide précieuse, tout soutien et tes encouragements tout au long de ces trois années de thèse (et pour avoir réussi à maintenir une température tropicale dans le bureau !). Tu as été bien plus qu'une collègue, tu es devenue une véritable amie. Merci pour tout !

Jiayi, ce fut un véritable plaisir de commencer cette thèse à tes côtés. Je tiens à te remercier pour ta grande gentillesse et pour toutes les découvertes gastronomiques que tu nous as fait partager.

Sanae, toujours prête à défendre ses convictions, qui m'a accompagné au LCAN et au CMAP (et la seule qui n'a jamais quitté la pièce 53), et dont les discussions plus ou moins philosophiques ont animé le bureau, c'était un plaisir de partager le bureau avec toi. Je te souhaite une excellente troisième année de thèse, pleine de succès !

Ensuite, je tiens à remercier mes “petits frères” de thèse, Andjib, Paulin, Richard, Valentin. Richard, merci pour toutes nos discussions et pour les mathématiques que tu m’as fait découvrir. Valentin, merci d’avoir pris le temps de relire ma thèse et pour tous les bons moments passés ensemble.

Je souhaite aussi remercier tous mes amis de CEA avec qui j’ai partagé ces trois ans de thèse : Andrew, Antonin, Anthony, Ayoub, Capucine, Chaïma, Chaïmaa, Clément, Élie, Étienne, Gabriel R., Gabriel S., Grégoire, Linkai, Mathis, Matteo, Matthieu, Maxence, Mayssa, Noa, Oumar, Sébastien, Téo, William, ainsi que tous les autres doctorants et stagiaires que j’ai pu rencontrer durant ces trois années.

Ensuite je tiens à remercier tous mes autres amis du CMAP pour tous les moments passés ensemble, tout d’abord toutes les personnes avec qui j’ai pu partager le bureau 2010, Adam, Ali B., Benjamin, Constantin, Kang, Mahmoud, Manon, Marius, Matthias, Zoubaïr, Yushan, celles des soirées jeux improvisés, Antoine, Armand, Guillaume, Raphaël, et enfin tous ceux que j’ai pu rencontrer durant ces trois années Ali A., Ana, Arthur, Clément, Corentin, Cristoph, Emmanouil, Grégoire, Jessie, Jules, Leila, Louis, Luce, Madeleine, Margherita, Marta, Maxime, Mehdi, Oskar, Orso, Quentin, Thomas, Ward, et pleins d’autres encore.

Je tiens également à remercier mes précieux amis de La Réunion, Adrien, Fabien et Radjah, pour votre soutien indéfectible, malgré la distance. Malgré mon absence prolongée, notre amitié reste intacte, et chaque retour me rappelle à quel point vous comptez pour moi.

Je remercie aussi tous les amis que j’ai eu la chance de garder tout au long de mes études. Je ne vais pas nommer chacun d’entre vous de peur d’en oublier, mais vous vous reconnaîtrez. Votre soutien et votre amitié ont été des éléments essentiels dans cette aventure, et je suis profondément reconnaissant de vous avoir à mes côtés.

Enfin, je tiens à remercier mes parents et ma petite sœur pour leur soutien indéfectible tout au long de ma scolarité et durant ces trois années de thèse. Votre encouragement et votre amour ont été des piliers essentiels dans cette aventure. Cette thèse vous est dédiée, car sans vous, rien de tout cela n’aurait été possible.

> Le bonheur est accessible, il suffit
> simplement de trouver sa vocation
> profonde, et de se donner à ce qu’on aime
> avec un abandon total de soi.
>
> *Romain Gary*
> *La promesse de l’aube*

Contents

List of Figures

List of Tables

Publications based on the work in this thesis

[29] L. Balazi, G. Allaire, P. Jolivet, and P. Omnes. Inf-sup stable non-conforming finite elements on tetrahedra with second and third order accuracy. https://hal.science/hal-04541809, April 2024.

[30] L. Balazi, G. Allaire, and P. Omnes. Sharp convergence rates for the homogenization of the Stokes equations in a perforated domain. *Discrete and Continuous Dynamical Systems - B*, 2024.

L. Balazi, G. Allaire, and P. Omnes. Discrete analysis of a high-order Multi-scale Finite Element Method (MsFEM) for Stokes flows in heterogeneous media. *(in preparation)*

Talk in conferences and seminars

During the preparation of this thesis, I have presented my research work in the following conferences and seminars.

- **International Conference on Scientific Computation and Differential Equations (SCICADE) 2024**, National University of Singapore, July 2024.
- **European Finite Element Fair (EFEF) 2024**, University College London, June 2024.
- **Numerical Analysis Congress (CANUM) 2024**, île de Ré, May 2024 (Organiser of a mini-symposium).
- **Analysis seminar (invited by Prof. Daniel Peterseim)**, Universität Augsburg, April 2024.
- **Séminaire Modélisation et Calcul Scientifique**, LAGA, Institut Galilée, Université Sorbonne Paris Nord, March 2024.
- **STMF PhD students seminar**, CEA Saclay, March 2024.
- **Seminar CEA-SMAI/GAMNI on Computational Fluids Dynamics**, Institut Henri Poincaré (IHP) - Paris, January 2024.
- **15th FreeFEM Days**, LJLL, Sorbonne Université, December 2023.
- **Congrès des Jeunes Chercheurs en Mathématiques et Applications (CJC-MA) 2023**, CentraleSupélec, September 2023.
- **European Conference on Numerical Mathematics and Advanced Applications (ENUMATH) 2023**, IST Lisbon, September 2023.
- **Days of PhD students of 2nd year**, CMAP, Ecole polytechnique, May 2023.
- **European Finite Element Fair (EFEF) 2023**, University of Twente, May 2023.
- **STMF PhD students seminar**, CEA Saclay, November 2022.
- **CMAP PhD students seminar**, Ecole polytechnique, October 2022.

I have presented a poster in the following conferences.

- **Foundations of Computational Mathematics (FOCM) 2023**, Workshop Multiresolution and Adaptivity in Numerical PDEs, Sorbonne Université, June 2023.
- **Seminar CEA-SMAI/GAMNI on Computational Fluids Dynamics**, Institut Henri Poincaré (IHP) - Paris, January 2023.
- **New trends in complex flows**, Institut Henri Poincaré (IHP) - Paris, September 2022.
- **Numerical Analysis Congress (CANUM) 2022**, Evian-les Bains, June 2022.

I have participated to the following summer schools.

- **CEA-EDF-INRIA Numerical Analysis summer school**, "Robust polytopal methods for computational mechanics", EDF Lab Paris-Saclay, Palaiseau, June 26 – June 30, 2023.
- **CEA-EDF-INRIA Numerical Analysis summer school**, "Error certifications on numerical simulations", EDF Lab Paris-Saclay, Palaiseau, June 27 – July 1st, 2022.

Part I

Introduction

Chapter 1

Introduction of the thesis

Contents

1.1 Motivation

1.1.1 Multi-scale media

This thesis is concerned with the development of computational methods for incompressible flows in multi-scale media. A multi-scale medium is a medium with either significant characteristics, such as diffusion coefficient or thermal conductivity, or its geometry itself, which vary at several different scales (magnitudes) of space or present an oscillating nature. The latter with a multi-scale geometry can be found in many areas. We can think, for example, of the soil, which are made up of materials of different sizes or a dense city, where buildings, cars or other installations

create obstacles of different sizes to the air or water flow. In this thesis, we are interested in this latter category of multi-scale media. In the literature, they can be referred to by various names: porous media in the geosciences, congested domains or heterogeneous domains in the nuclear field, or even perforated domains.

Indeed, at the French Alternative Energies and Atomic Energy Commission (CEA), what interests us is to compute the flow in the nuclear reactor cores, such as in a Reactor Pressure Vessel (RPV) of a Pressurized Water Reactor (PWR), which are definitely multi-scale media with obstacles whose sizes vary across more than four orders of magnitude and whose scales of interest spread over several orders of magnitude. To get an idea of the scales involved, we give, in what follows, some numerical values. The overall length of the RPV is about 14 meters. The reactor core is about 4 to 5 meters in diameter and consists of between 157 and 200 fuel assemblies (see Figure 1.1). The fuel assembly is about 4 to 5 meters long and consists of fuel rods or pins generally bundled in a square array of 14×14 to 17×17. The thickness of the mixing vanes in the spacer mesh is about 1 centimeter. To summarise, the RPV contains therefore various scales and the global scale is about 10^4 times larger than the local scale.

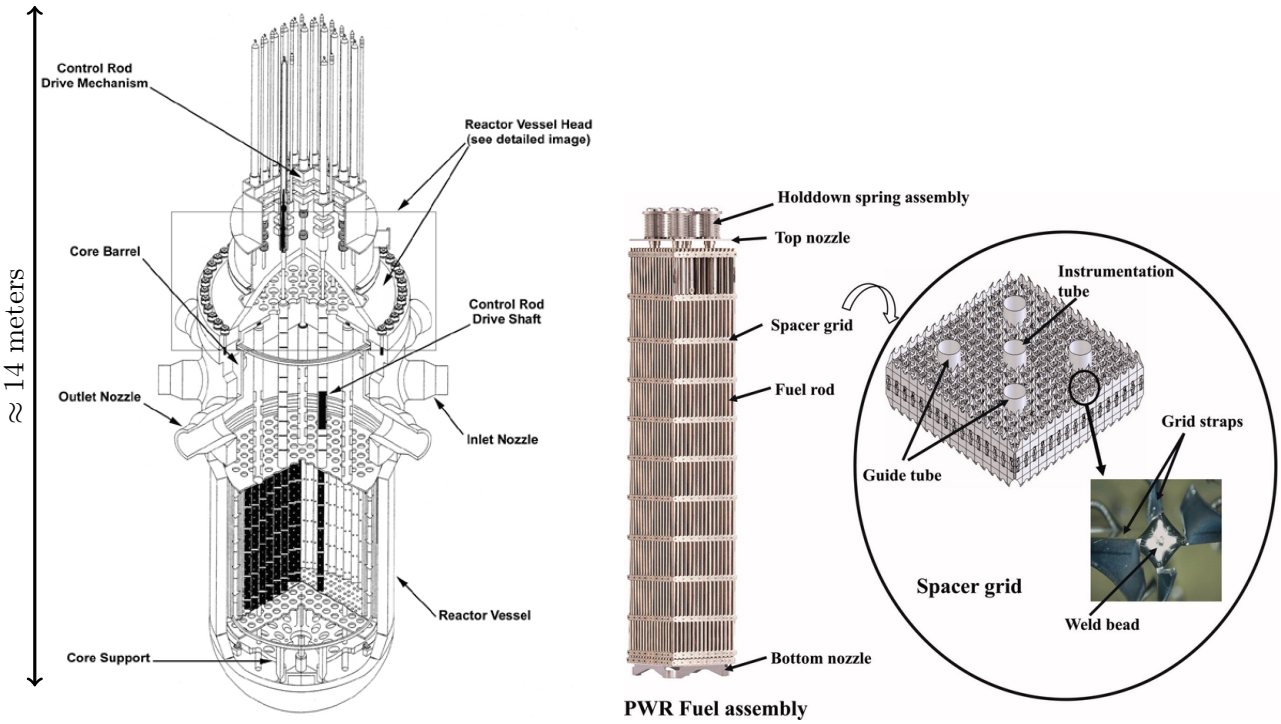


Figure 1.1: Schematic description of a PWR (left) and a fuel assemby (right) [1].

A typical modelling of such a multi-scale medium is to consider a domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ perforated by a set of periodic obstacles B^ε, where $\varepsilon > 0$ is the period (or the length-scale of the heterogeneities). It is important to note that this domain can be also non-periodic. However, for further quantitative results (see for example Chapter 6 and Chapter 9), especially for homogenization, the periodic nature is one of the possible framework. Quantitative homogenization results can also be obtained in the stochastic framework but with additional hypothesis such as ergodicity and stationarity [22]. However, this setting is out of the scope of this thesis.

An illustrative example of a perforated domain is given in Figures 1.2 and 1.3 respectively

in two and three dimensions. The domain Ω is divided into a fixed solid part B^ε and its complementary fluid part Ω^ε, i.e. $\Omega^\varepsilon = \Omega \setminus B^\varepsilon$.

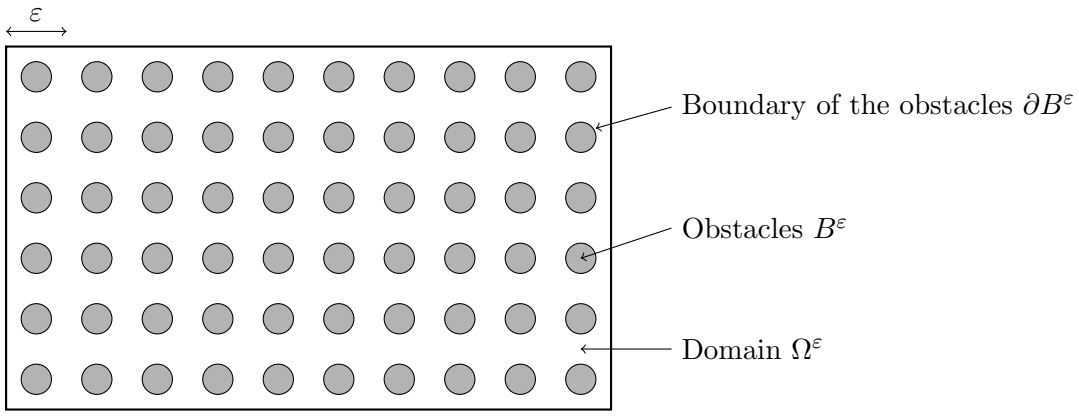


Figure 1.2: Two-dimensional porous domain Ω^ε, obtained from a domain Ω perforated by a set of solid obstacles B^ε.

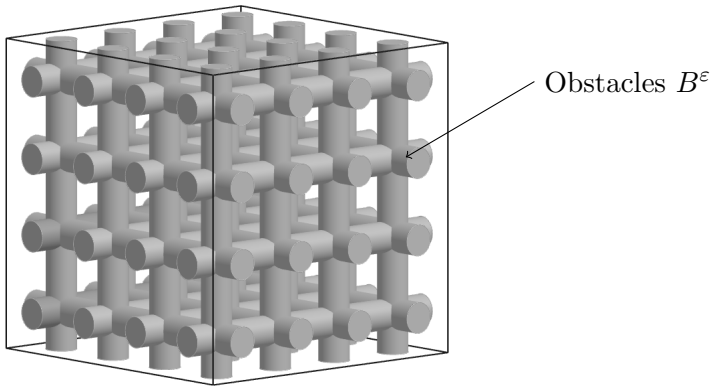


Figure 1.3: Three-dimensional porous domain Ω^ε, obtained from a domain Ω perforated by a set of solid obstacles B^ε.

1.1.2 Flow modelling

Once the multi-scale medium given, the flow problem in this medium has to be modelled. In the case of Pressurised Water Reactors (PWRs), the water in the core reaches approximately 325°C and is maintained at a high pressure of approximately 150 bar in order to prevent it from transforming into steam. Under nominal conditions, the Reynolds number in the core is of the order of 1×10^5, which suggests that the flow is fully turbulent [2, 59]. In this work, we restrict our attention to incompressible, single-phase problems as a preliminary step before considering more complex applications. Furthermore, we do not consider turbulent flows, as this would necessitate the introduction of a turbulence model, such as the $k - \varepsilon$ model proposed by Hanjalić and Launder [89] or the $k - \omega$ model proposed by Wilcox [167]. This is beyond the scope of this thesis. We note the velocity $\boldsymbol{u}_\varepsilon$ and the pressure p_ε to recall that we solve the flow in a multi-scale domain with heterogeneities length-scale (or the period) of size ε. An overview of different flow problems is given below.

The Navier–Stokes equations The incompressible Navier–Stokes equations can be used to model fluid flows in the domain Ω^ε. The steady-state incompressible Navier–Stokes problem with homogeneous Dirichlet boundary conditions is to find the velocity $\boldsymbol{u}_\varepsilon : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\left\{ \begin{array}{rcll} -\nu\Delta \boldsymbol{u}_\varepsilon + (\boldsymbol{u}_\varepsilon \cdot \boldsymbol{\nabla})\boldsymbol{u}_\varepsilon + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{in } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u}_\varepsilon &=& 0 & \text{in } \Omega^\varepsilon, \\ \boldsymbol{u}_\varepsilon &=& \boldsymbol{0} & \text{on } \partial\Omega^\varepsilon, \end{array} \right. \tag{1.1}$$

with $\boldsymbol{f}$ the applied force and $\nu > 0$ the viscosity.

The Oseen equations The Oseen approximation is a partial linearisation of the Navier–Stokes equations. The non-linear term $(\boldsymbol{u}_\varepsilon \cdot \boldsymbol{\nabla})\boldsymbol{u}_\varepsilon$ is replaced by the linear term $(\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u}_\varepsilon$, with $\boldsymbol{U_0}$ a known velocity, called the Oseen velocity. The steady state Oseen problem with homogeneous Dirichlet boundary conditions is to find the velocity $\boldsymbol{u}_\varepsilon : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\left\{ \begin{array}{rcll} -\nu\Delta \boldsymbol{u}_\varepsilon + (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u}_\varepsilon + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{in } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u}_\varepsilon &=& 0 & \text{in } \Omega^\varepsilon, \\ \boldsymbol{u}_\varepsilon &=& \boldsymbol{0} & \text{on } \partial\Omega^\varepsilon, \end{array} \right. \tag{1.2}$$

with $\boldsymbol{U_0}$ the Oseen velocity, $\boldsymbol{f}$ the applied force and $\nu > 0$ the viscosity.

In both above problems, the non-linear convective term $(\boldsymbol{u}_\varepsilon \cdot \boldsymbol{\nabla})\boldsymbol{u}_\varepsilon$ for the Navier–Stokes problem, and the linear convective term $(\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u}_\varepsilon$ for the Oseen problem, increase the difficulty in mathematical analysis and numerical solutions of these problems. The predominance of the convective term can be predicted by the Reynolds number, noted Re, a dimensionless number, defined as the ratio between the inertial forces and the viscous forces in the flow, i.e.

$$Re = \frac{U_c L}{\nu},$$

where U_c is a characteristic velocity magnitude, ν the viscosity and L a characteristic length.

The Stokes equations The Reynolds number of flows in porous media is usually small and the inertial forces can be neglected compared to the viscous forces. This simplification leads to the Stokes problem. The steady state Stokes problem with homogeneous Dirichlet boundary conditions is to find the velocity $\boldsymbol{u}_\varepsilon : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\left\{ \begin{array}{rcll} -\nu\Delta \boldsymbol{u}_\varepsilon + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{in } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u}_\varepsilon &=& 0 & \text{in } \Omega^\varepsilon, \\ \boldsymbol{u}_\varepsilon &=& \boldsymbol{0} & \text{on } \partial\Omega^\varepsilon, \end{array} \right. \tag{1.3}$$

with $\boldsymbol{f}$ the applied force and $\nu > 0$ the viscosity.

The first approach, to study the flow in a multi-scale medium, is the direct computation of one of the flow problems mentioned above (the Stokes, the Oseen or the Navier–Stokes problems). However, simulating the flow in a multi-scale medium with many obstacles is very challenging.

Indeed, in such a medium, the flow is influenced by phenomena which occur at the finest scales. Then, to capture the finest scales of the flow, one needs to use a very fine mesh which often leads to intractable simulations due to the lack of computational resources. At the CEA, the computational fluid dynamics (CFD) studies of the nuclear reactor cores are conducted using the open-source software TrioCFD [45]. It is a software developed for about 20 years in the Nuclear Energy Division of the CEA. It is an object-oriented, massively parallel code implemented in C++, which has been developed for use in a variety of scientific and industrial studies and research applications. The code has been designed to treat efficiently a range of physical problems, including turbulent flows, fluid/solid coupling, multiphase flows and flows in porous media [32, 153, 147]. An overview of the TrioCFD code can be found in [16, 17].

For illustration purposes, let us define a mesh $\mathcal{T}_H$ of a polyhedral periodic perforated domain Ω^ε as a decomposition of Ω^ε into polyhedra T, each of diameter at most H. We define the classical velocity space $V = H_0^1(\Omega^\varepsilon)^d$ and pressure space $M = L_0^2(\Omega^\varepsilon) = \{q \in L^2(\Omega^\varepsilon),\ s.t.\ \int_{\Omega^\varepsilon} q = 0\}$ equipped respectively with the $|\cdot|_1$ H^1 semi-norm and the $\|\cdot\|_0$ L^2 norm. For non-conforming approaches, we introduce the broken Sobolev space as

$$H^1(\mathcal{T}_H) = \left\{ \boldsymbol{u} \in L^2(\Omega^\varepsilon)^d \text{ such that } \boldsymbol{u}|_T \in H^1(T)^d,\ \forall T \in \mathcal{T}_H \right\}.$$

We also define the H^1-broken norm $|\boldsymbol{u}|_{H,1}$ in the H^1 broken space as

$$|\boldsymbol{u}|_{H,1} = \left(\sum_{T \in \mathcal{T}_H} |\boldsymbol{u}|^2_{H^1(T)} \right)^{\frac{1}{2}}.$$

Then, we can think about solving the Stokes equations (1.3) in this domain Ω^ε. Assume that the Stokes problem is discretized with the Crouzeix–Raviart Finite Element [61], then, it is well-known that the error between the exact solution $(\boldsymbol{u_\varepsilon}, p_\varepsilon)$ and its finite element approximation $(\boldsymbol{u_H}, p_H)$ is given by

$$\frac{1}{|\boldsymbol{u_\varepsilon}|_{H^1(\Omega^\varepsilon)}} \left(|\boldsymbol{u_\varepsilon} - \boldsymbol{u_H}|_{H,1} + \|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \right) \leq CH \underbrace{\left(|\boldsymbol{u_\varepsilon}|_{H^2(\Omega^\varepsilon)} + |p_\varepsilon|_{H^1(\Omega^\varepsilon)} \right) \frac{1}{|\boldsymbol{u_\varepsilon}|_{H^1(\Omega^\varepsilon)}}}_{\beta_\varepsilon},$$

with C a constant independent of the mesh. Now, it is known that in the periodic case (see Chapter 6 for more details), functionally, i.e. for the function and its derivatives of any order, we have $\boldsymbol{u_\varepsilon} \approx \varepsilon^2 \boldsymbol{u}(x, \frac{x}{\varepsilon})$ and therefore the term β_ε is of order ε^{-1} (since at each derivation of $\boldsymbol{u}(x, \frac{x}{\varepsilon})$, a factor $1/\varepsilon$ comes out). Consequently, obtaining a good approximation of the solution requires a mesh size H much smaller that ε. Since ε can be very small in practice, satisfying this condition is often impossible. For example, to capture the finest scales of the flow in the physical RPV presented above, a typical mesh to discretize this domain can contain hundreds of millions of elements. It is clear that direct simulations on such a mesh require a massive amount of memory and computing time, which can exceed today's computing power. Even if such a computation is possible, it cannot be done on a regular basis, but only in exceptional cases, as the CFD calculations of turbulent flows around PWR mixing grids with more than one billion

degrees of freedom done in TrioCFD [32].

In order to overcome these difficulties, a popular class of numerical methods, based on the strategy of averaging or "upscaling", has been developed. It corresponds to the process to approximate fine scale models by a coarse scale models. In whats follows, we present two methods of upscaling, the Representative Elementary Volume (REV) and homogenization.

1.1.3 Approaches based on averaging strategy or upscaling

Representative Elementary Volume (REV) REV is a first method to "upscale" or average fine scale models [149]. The idea of this approach is the following [102]. Let u be a real valued function on a domain Ω which describes certain physical quantities with rapid spatial oscillations. To smoothen this function, one considers local averages of the form

$$\langle u \rangle(x) = \int_{V(x)} u(y) \mathrm{d}y,$$

where $V(x)$ is a small neighborhood of the point x of the size of a REV. If the oscillations of u reflect the behaviour of the physical quantity in question on a "microscale", the averaged function $\langle u \rangle$ is supposed to describe its properties on a larger scale, i.e. on a "macroscale". Several works [56, 120] applied this strategy to derive coarse scale model of the flow in porous media.

Homogenization Homogenization is another mathematical method that allows to "upscale" or average differential equations [102]. The idea of homogenization is to consider a family of functions u_ε (where $\varepsilon > 0$ is a spatial (length) scale parameter). And then, ones have to determine the limit

$$u = \lim_{\varepsilon \to 0} u_\varepsilon, \tag{1.4}$$

and considering this limit as the result of the upscaling procedure. The objective of the homogenization procedure is to determine the differential equations that the limit u satisfies and proving that (1.4) holds. Applied to incompressible flows, it is well-known that the homogenization of the Stokes equations (1.3), i.e. finding the limit system satisfied by the limit of $(\boldsymbol{u_\varepsilon}, p_\varepsilon)$ as ε goes to zero [154, 160, 9], leads to Darcy's law for the velocity $\boldsymbol{u}$ and the homogenized pressure p^*,

$$\left\{ \begin{array}{rcll} \boldsymbol{u} & = & \frac{1}{\nu} A^* (\boldsymbol{f} - \boldsymbol{\nabla} p^*) & \text{in } \Omega, \\ \operatorname{div} \boldsymbol{u} & = & 0 & \text{in } \Omega, \\ \boldsymbol{u} \cdot \boldsymbol{n} & = & 0 & \text{on } \partial\Omega, \end{array} \right.$$

where A^* is a constant permeability tensor (see Chapter 6 for more details). This method is often combined with upscaling method in which the fine scales properties of the problem are averaged in order to find effective macroscopic parameters, which are then introduced in the coarse scale model defined by the homogenization theory. For example, in case of the Darcy problem, this method consists in finding an effective permeability from local permeabilities. This method has been applied for several problems in the literature to derive coarse scale flow problems in highly heterogeneous media or composite materials [169, 51, 110, 53, 77, 109].

The drawback of these two mentioned approaches is that only the average behaviour of the flow is captured. Indeed, they cannot capture the flow behaviour at the finest scales, for example the oscillating or fluctuating features, unless correctors are used.

To overcome this limitation, another approach, inspired by these ideas, has been developed to attempt to resolve scales below the coarse mesh scale. Instead of computing the effective properties as in upscaling approaches, the fine scale features are incorporated in the coarse scale computation via a multi-scale approach. Most of the methods using this approach consist in solving fine scale problems and then, using these solutions in a coarse scale problem. A review of some multi-scale methods is done in Section 1.2. In this thesis, we focus on one of them, which is the **Multi-scale Finite Element Method (MsFEM)**, since it is the continuation of previous works on MsFEM for incompressible flows [111, 78, 79]. In Section 1.3, we give the main ideas of MsFEM and we illustrate this approach through a simple example in Section 1.4. Then, in Section 1.5, we give an overview of the development of MsFEM in the literature.

1.2 An overview of different multi-scale methods

In this section, we give an overview of different multi-scale methods developed in the literature.

1.2.1 The Variational Multi-scale Method (VMS)

The Variational Multi-scale Method was first introduced by [106, 108] as a general procedure for deriving numerical methods capable of dealing with multi-scale phenomena. This method is based on the decomposition of the solution, $u = \overline{u} + \tilde{u}$, where $\overline{u}$ is the resolved part or coarse scale and $\tilde{u}$ is the unresolved part or fine scale. Then, it consists in solving $\overline{u}$ numerically and attempting to determine $\tilde{u}$ analytically, eliminating it from the problem for $\overline{u}$.

1.2.2 Localized Orthogonal Decomposition (LOD)

The Localized Orthogonal Decomposition (LOD) was first introduced in [97, 98], in which the authors construct a local generalized finite element basis for elliptic problems with heterogeneous and highly varying coefficients. The basis functions are solutions of local problems on vertex patches. The advantage of this method is that the local problems are localized with a fast decrease allowing to restrict the computational domain of the local problems on a small patch. This method has been applied successfully in the literature for example for Helmhotz-type problem [140], for wave equation [6] or for performing numerical homogenization [95, 13].

1.2.3 Heterogeneous Multi-scale Method (HMM)

The Heterogeneous Multi-scale Method (HMM) was first introduced in [166, 67]. This method provides a general methodology for the efficient numerical computation of multi-scale and multi-physics problem, involving a macroscopic model U and a microscopic model u. These models are defined on different domains and are linked together with reconstruction and compression operators. The aim of this method is to approximate the macroscopic state of the system on a macroscopic grid that resolves the large scale of the problem, and estimate the necessary missing

data for the macroscale scheme from the microscopic scheme. This method involves therefore two main components: an overall macroscopic scheme for U and an estimate of the missing data from the microscopic model for u. This method is qualified as heterogeneous to emphasize the fact that different physical models and numerical techniques are used at different scales and on different grids. This method has been applied successfully in the literature for different problems such as diffusion problems [7], homogenization problems [68, 66] or for modelling complex fluids [150].

1.2.4 Multi-scale Finite Volume Method (MsFVM)

The Multi-scale Finite Volume Method (MsFVM) was first introduced by [112] for sub-surface flow simulations and extended in [113] for multiphase flows and transport in porous media. This method is based on the construction of two set of basis functions: the first one is used to build transmissibilities for the coarse scale problem and the second one is used to reconstruct the fine scale velocity field from the coarse solution. These basis functions which ensure the mass balance at both scales are computed on a dual grid. This method was mainly applied for subsurface flow and reservoir simulations [126, 168, 114, 137, 87, 127, 88].

1.2.5 Multi-scale Hybrid Mixed Method (MHM Method)

The Multi-scale Hybrid Mixed Method (MHM Method) was first introduced by [91] for the Darcy problem and by [19] for elliptic problems. This method aims to incorporate multiple scales into the construction of basis functions. This method is based on a mixed-formulation: it relaxes the continuity of the primal normal component through the action of Lagrange multipliers, while ensuring the strong continuity of the normal component of the flux. The local problems consist in finding basis functions that approximate the space of the Lagrange multipliers. This method has been applied for advection-reaction dominated problems with heterogeneous coefficients [92] or for second order elliptic problems with rough periodic coefficients [143]. It has been used as well, for solving in heterogeneous media, the Stokes and Brinkman problems [20], the Maxwell equations [119], the Helmholtz equation [47] or the Oseen problem [18].

1.2.6 Other multi-scale approaches

There are several other methods in addition to the ones mentioned above. We can cite for example the Mortar Multi-scale Finite Element Methods [85] used for solving Stokes and Darcy flows or the Multi-scale Discontinuous Galerkin Method [73, 72] used for second order elliptic problems.

1.3 Introduction to Multi-scale Finite Element Method (MsFEM)

The main idea of the Multi-scale Finite Element Method (MsFEM) is to approximate a global problem on a coarse mesh $\mathcal{T}_H$, which does not take into account the geometric details (see for example Figure 1.4b), using a Galerkin approximation. However, the basis functions are

no longer the classical Lagrange polynomial basis functions of finite elements, but solve local problems on the elements T of the coarse mesh. These functions are themselves numerically approximated on a fine mesh $\mathcal{T}_h(T)$ taking into account all the geometric details (see for example Figure 1.4c), which gives the multi-scale aspect of this method.

So far, we have not made any assumptions about the shape of the elements of the coarse mesh. These elements may be of different types (triangles, squares or polygons in two dimensions, and tetrahedra, rectangles or polyhedra in three dimensions). However, for convenience, the elements of the coarse mesh are chosen to be triangles and tetrahedra (respectively in two and three dimensions) in the different illustrations and in the numerical simulations.

Two key points should be considered at this stage. The first point is that these local basis functions do not depend on the boundary conditions and the applied forces of the global problem. Consequently, these basis functions can be reused to solve the global problem but with different boundary conditions and applied forces, which makes this method very relevant for parametric studies. The second point is that the basis functions are entirely local, i.e. the basis functions associated with a given coarse element is independent of those of other coarse elements. This point makes the computation of the local basis functions very efficient, since their computations can be done entirely in parallel, i.e. without any communication between the different elements of the coarse mesh.

This method consists of two main steps, described below.

The offline step The first step, called the offline step, corresponds to the computation of the local basis functions in each coarse element of the coarse mesh, by solving local problems of type $\mathcal{A}_T(\phi, v) = \mathcal{F}_T(v)$. During this step, matrices of the global problem are assembled. This step is considered as the expensive stage since it requires to solve, possibly in parallel, several local problems on fine meshes.

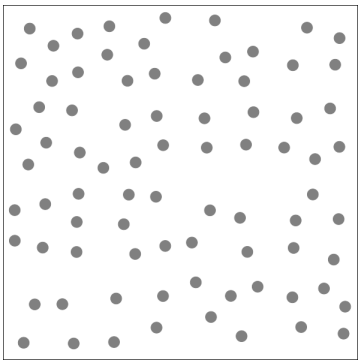
The online step Once the basis functions computed, the second step, called the online step, consists in assembling the right-hand side, applying the boundary conditions and solving the global problem of type $\mathcal{A}_{\mathcal{T}_H}(u, v) = \mathcal{F}_{\mathcal{T}_H}(v)$ on the coarse mesh. This step is considered as the cheap stage, since it implies to solve a problem on a coarse mesh which contains few elements. If the global problem needs to be recomputed, but with different boundary conditions and loads, only the online step, i.e. the cheap one, needs to be re-computed.

A third step consists in reconstructing the solutions on the fines meshes and post-processing the results.

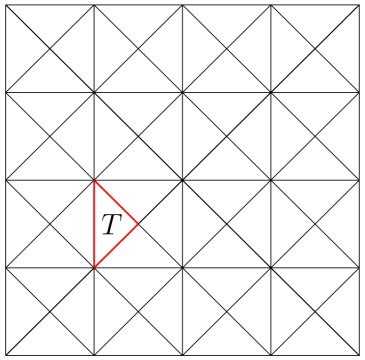
To summarise, the MsFEM procedure reads as follows.

Algorithm 1.1 (MsFEM procedure)**.**

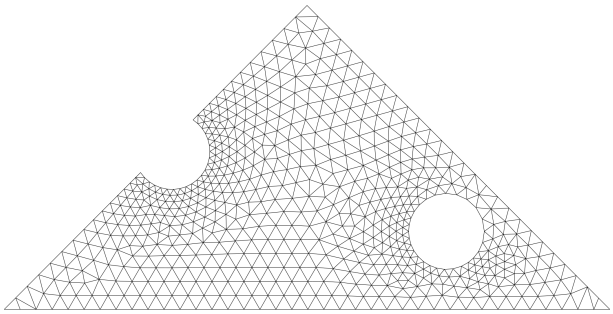
- Partition the domain into a set of coarse elements, which forms the coarse mesh (see Figure 1.4b).
- Offline stage. For each coarse element (**In parallel**):
  - Partition the element into a fine mesh (see Figure 1.4c).
  - Construct multi-scale basis functions by solving local problems on this fine mesh.
  - Compute matrices locally and assemble global matrices.
- Online stage. Apply boundary conditions, assemble the right-hand side and solve the global problem on the coarse mesh.
- For each coarse element (**In parallel**):
  - Reconstruct fine-scale solutions on the fine meshes.
  - Post-process the solutions.



(a) Heterogeneous domain Ω^ε.



(b) Coarse mesh $\mathcal{T}_H$ and element T.



(c) Fine mesh $\mathcal{T}_h(T)$.

Figure 1.4: Description of MsFEM procedure.

For the sake of comparison, the fine-scale solutions can be reconstructed on the whole fine mesh, called "reference mesh", given in Figure 1.5.

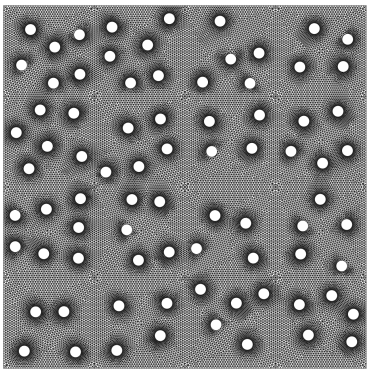


Figure 1.5: Reference mesh $\mathcal{T}_h(\Omega^\varepsilon)$.

In what follows, for the sake of understanding, we illustrate the MsFEM through a simple example: a diffusion problem with oscillating diffusion coefficient in one dimension.

1.4 A simple illustrative example in one dimension: an oscillating diffusion problem

Let I be the interval $[0,1] \subset \mathbb{R}$ and $f \in L^2(I)$ be a given function. We consider the following diffusion problem with homogeneous Dirichlet boundary conditions.

$$\begin{aligned} -\frac{\mathrm{d}}{\mathrm{d}x}\left(A_\varepsilon(x)\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}x}(x)\right) &= f(x) \text{ in } I, \\ u_\varepsilon(0) &= 0, \\ u_\varepsilon(1) &= 0. \end{aligned} \tag{1.5}$$

We consider the following oscillating diffusion coefficient A_ε,

$$A_\varepsilon(x) = \frac{1}{2+\cos\left(2\pi(x/\varepsilon)\right)},$$

with $\varepsilon > 0$ a small parameter. The oscillations of A_ε are presented in Figure 1.6 for $\varepsilon = 1\times 10^{-1}$.

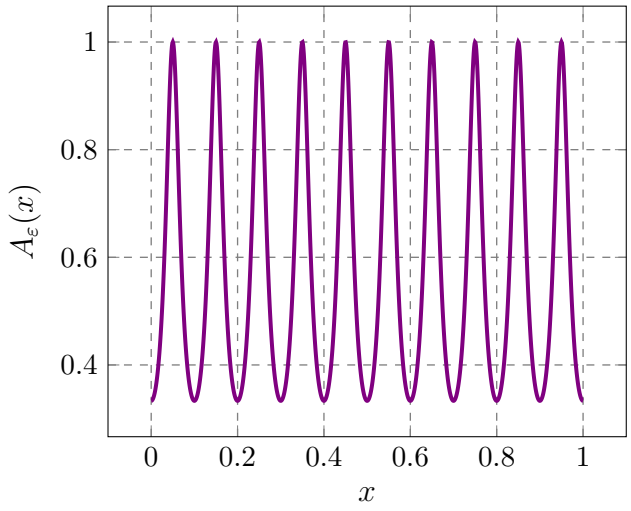


Figure 1.6: Diffusion coefficient A_ε, $\varepsilon = 1 \times 10^{-1}$.

For the particular right-hand side $f = 1$, the exact solution to the problem (1.5) is given by

$$u_\varepsilon(x) = x - x^2 + \frac{\varepsilon}{2\pi}\left(\frac{1}{2} - x\right)\sin\left(2\pi\frac{x}{\varepsilon}\right) + \frac{\varepsilon^2}{4\pi^2}\left(1 - \cos\left(2\pi\frac{x}{\varepsilon}\right)\right).$$

The weak formulation of problem (1.5) is to find $u_\varepsilon \in H_0^1(I)$ such that, for all $v \in H_0^1(I)$,

$$a(u_\varepsilon, v) = \int_I f v \mathrm{d}x,$$

with

$$a(u_\varepsilon, v) = \int_I A_\varepsilon(x)\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}x}\frac{\mathrm{d}v}{\mathrm{d}x}\mathrm{d}x.$$

Let $N \in \mathbb{N}^*$ be a given integer. Let the segment I be divided into $N+1$ sub-intervals $I_i = [x_i, x_{i+1}]$ of size $H = \frac{1}{N+1}$ with $i \in [\![0, N]\!]$. Let us suppose that the exact solution to (1.4) is approached by the solution to the following discrete problem. Find $u_H \in V_H$ such that, for all $v_H \in V_H$,

$$a(u_H, v_H) = \int_I f v_H \mathrm{d}x, \tag{1.6}$$

where V_H, the approximation space, is a subset of $H_0^1(I)$ to be defined.

Approximation with piecewise affine functions First we propose to solve (1.6) with linear piecewise functions. The approximation space is the $\mathbb{P}_1$ Lagrange space, defined as

$$V_H = \left\{\varphi \in \mathcal{C}^0(I) \ s.t. \ \varphi = 0 \text{ on } \partial I, \ \varphi|_{I_i} \in \mathbb{P}_1(I_i), \ \forall i \in [\![0, N]\!]\right\},$$

where $\mathbb{P}_1(I_i)$ denotes the space of polynomials of total degree at most 1 in I_i. The basis of this space in one dimension is shown in Figure 1.7.

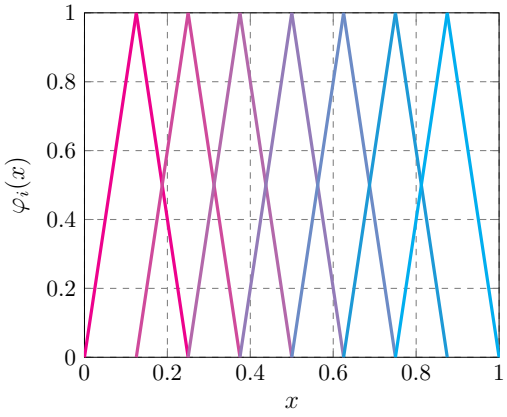


Figure 1.7: $\mathbb{P}_1$ basis functions φ_i, $H = 1/8$.

The approximation of the oscillating problem (1.5) with affine piecewise functions is given in Figure 1.8. It can be easily seen that the $\mathbb{P}_1$ approximation of this problem fails to capture the oscillating solution, unless using a very fine mesh. To solve this problem, the integrals have been computed using the default quadrature formula in FreeFEM for one-dimensional integration, i.e. `qforder` $= 6$, which is exact on $\mathbb{P}_5$.

It should be noted that the phenomena observed in Figure 1.8 two causes: an insufficient precise integration of $A_\varepsilon(x)$ when H is large and an intrinsic aspect of the $\mathbb{P}_1$ approximation which fails to capture the oscillations [13].

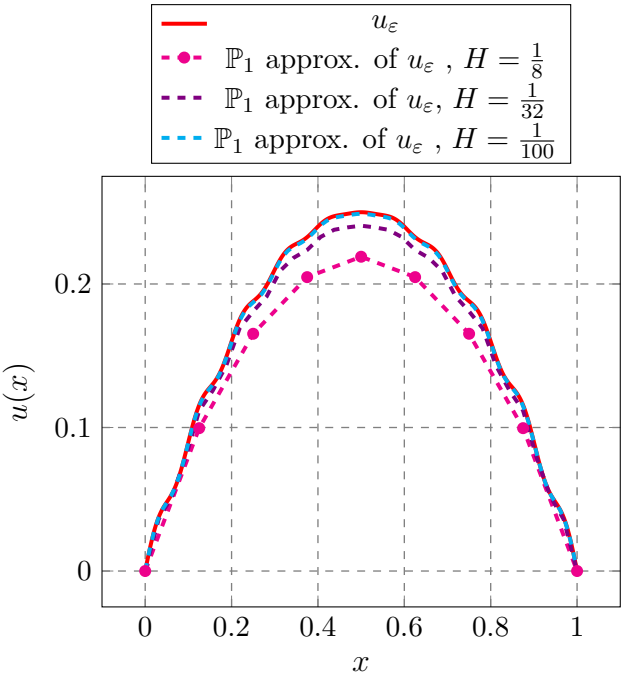


Figure 1.8: Finite element approximation of the oscillating diffusion problem.

Approximation with MsFEM approach Now, instead of considering an approximation space consisting in piecewise affine functions, we consider an approximation space composed of well-adapted functions to the oscillating problem. We define the new approximation space V_H as

$$V_H = \left\{ \varphi \in \mathcal{C}^0(I),\ -\frac{\mathrm{d}}{\mathrm{d}x}\left(A_\varepsilon(x)\frac{\mathrm{d}}{\mathrm{d}x}\varphi(x)\right)|_{I_i} = 0,\ \forall i \in [\![0, N]\!] \right\},$$

The basis of this space in a one-dimensional domain is shown in Figure 1.9. It can be seen that the basis functions encode the oscillating nature of the problem in themselves.

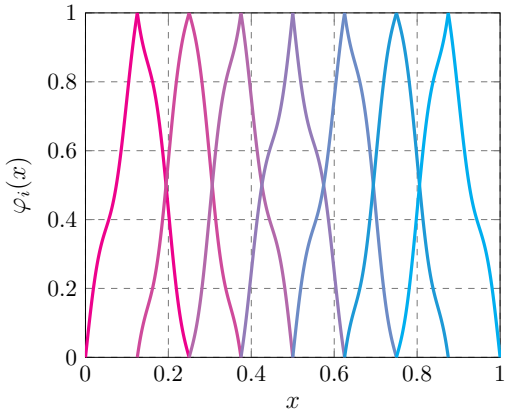


Figure 1.9: MsFEM basis functions, φ_i, $H = 1/8$.

The approximation of the oscillating problem (1.5) with the Multi-scale Finite Element

Method is given in Figure 1.10. It can be seen that, now, MsFEM allows to approximate the solution to problem (1.5) much better than the $\mathbb{P}_1$ Finite Element. In addition, the plot of the gradient of the multi-scale approximation in Figure 1.11 (computed element-wise since the gradients are not continuous at the interface of coarse element) allows to see that the multi-scale approach allows to capture the oscillations of the solution compared to the $\mathbb{P}_1$ approximation which does not see the oscillations.

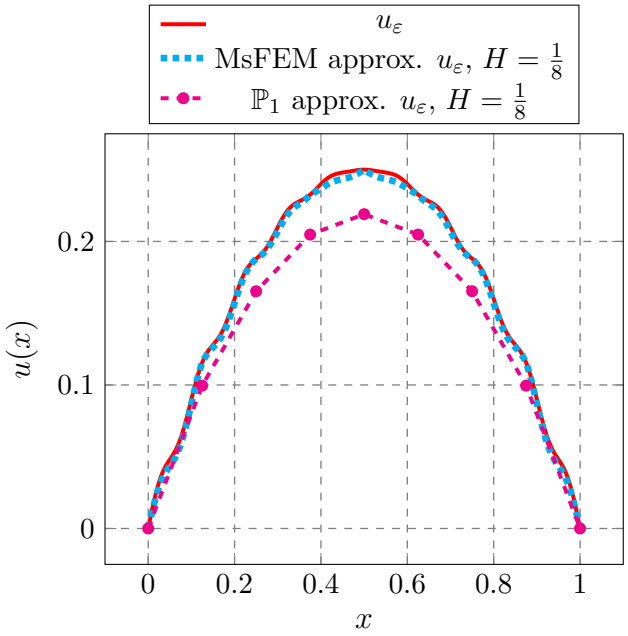


Figure 1.10: MsFEM approximation of the oscillating diffusion problem.

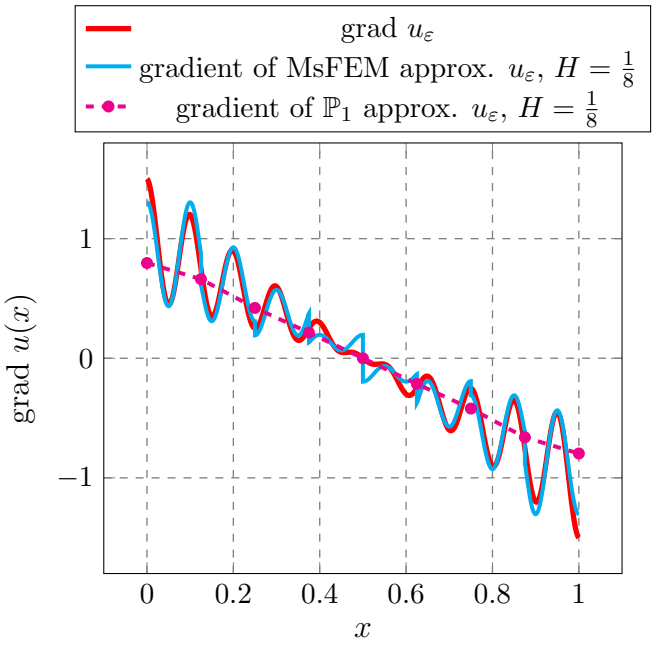


Figure 1.11: Gradient of MsFEM approximation of the oscillating diffusion problem.

The following section presents an overwiew of the development of Multi-scale Finite Element Methods (MsFEMs) in the literature.

1.5 An overview of Multi-scale Finite Element Methods (MsFEMs) in the literature

Given that the evolution of the MsFEM has been largely motivated by the improvement of the error estimate between the exact solution (denoted with the index ε) and the approximated solution (denoted with the index H), we include in loose terms the corresponding error estimates in this overview. In these estimates, H is the size of the coarse mesh which consists of coarse elements T, and C a generic constant that does not depend on H. It should be noted that the error estimates presented in this section are valid only for periodic coefficients, and obtained by assuming that the computation of the local basis functions is exact. For non-conforming approaches in case of perforated domains Ω^ε, we define the broken Sobolev space as

$$H^1(\mathcal{T}_H) = \left\{ \boldsymbol{u} \in L^2(\Omega^\varepsilon)^d \text{ such that } \boldsymbol{u}|_T \in H^1(T \cap \Omega^\varepsilon)^d, \ \forall T \in \mathcal{T}_H \right\},$$

and the H^1-broken norm $|\boldsymbol{u}|_{H,1}$ in the H^1 broken space as

$$|\boldsymbol{u}|_{H,1} = \left(\sum_{T \in \mathcal{T}_H} |\boldsymbol{u}|^2_{H^1(T \cap \Omega^\varepsilon)} \right)^{\frac{1}{2}}.$$

1.5.1 MsFEM for elliptic problems

In the context of the preliminary work that has been carried out by [24, 133], the original Multi-scale Finite Element Method was first introduced in [104] for elliptic problems in composite materials as well as in porous media and in [103] for elliptic problems with rapidly oscillating coefficients. In these first works, the following elliptic problem has been considered

$$-\operatorname{div}\left(A_\varepsilon \boldsymbol{\nabla} u_\varepsilon\right) = f \text{ in } \Omega, \qquad u_\varepsilon = 0 \text{ on } \partial\Omega, \tag{1.7}$$

with A_ε an oscillating diffusion coefficient. The local basis functions are computed using a similar elliptic problem in each coarse element. For their computations different boundary conditions have been chosen. Considering first linear boundary conditions on each coarse element, this method leads to the following error estimate

$$|u_\varepsilon - u_H|_1 \leq C \left(\sqrt{\varepsilon} + H + \sqrt{\frac{\varepsilon}{H}} \right), \tag{1.8}$$

where u_H is the solution obtained using the MsFEM. Oscillatory boundary conditions (boundary conditions that are given by the solution to the oscillatory ordinary differential equation obtained by taking the trace of the original equation on edges) have also been considered. It was shown that the same bound of the error (1.8) holds for this approach. Nevertheless, numerical examples showed that using oscillatory boundary conditions leads to significant improvement on the accuracy of the numerical results. However, it was shown that considering one of these two types of boundary conditions leads to a resonance between the mesh scale and the small scale in the physical solution. This is emphasized by the presence of the term $\sqrt{\frac{\varepsilon}{H}}$, called "resonance error", in the error estimate (1.8).

Then, to mitigate the cell resonance, due to the boundary layers and the choice of boundary conditions, [104, 71] proposed the so-called *oversampling* method. This method consists in solving the local problems on a domain bigger than the coarse element and then truncating the solution inside the coarse element. This method, which makes the solution less dependent of the boundary conditions chosen, leads to non-conforming problems since quantities are not continuous on the element borders. It was shown that using oversampling leads to the following error estimate

$$|u_\varepsilon - u_H|_{H,1} \leq C\left(\sqrt{\varepsilon} + H + \frac{\varepsilon}{H}\right). \tag{1.9}$$

From (1.9), it is easy to see that the resonance is mitigated, but still exists.

Consequently, to remove the "cell resonance", in [105], the approach has been modified, introducing a Petrov–Galerkin formulation with non-conforming multi-scale trial functions and linear test functions. Using this new approach, the following error estimate has been reached,

$$|u_\varepsilon - u_H|_{H,1} \leq C\left(\sqrt{\varepsilon} + H + \varepsilon\right),$$

which does not involve the "resonance error". A complete study of these first methods can be found in [70].

In [12], Multi-scale Finite Element Method has been used to perform numerical homogenization of diffusion problems. The same elliptic problem (1.7) has been considered. One of the novelty of this work is the introduction of a composition rule, or change of variables, for the construction of finite element basis which allows a simple treatment of high-order finite element methods. An optimal fine mesh size h minimising the numerical error has also been exhibited. Inspired by [12], [101] developed a high-order Multi-scale Finite Element Method for elliptic problems with highly oscillating coefficients, but by constructing a more explicit multi-scale finite element space.

One of the last improvement in the field of Multi-scale Finite Element Methods was to consider a method in the vein of the classical Crouzeix–Raviart finite element method, i.e. a method in which the continuity through the faces of the coarse elements is enforced in a weak sense: only the mean value is continuous through the faces. Since boundary conditions are required on each coarse element, the MsFEM à la Crouzeix-Raviart allows natural boundary conditions on the faces of the coarse element. This method was first introduced in [121] for highly oscillatory elliptic problems (1.7). In order to solve this problem, the Galerkin approximation space has been constructed from ideas similar to those by Crouzeix and Raviart in their construction of a classical finite element space [61]. The following error estimate has been reached,

$$|u_\varepsilon - u_H|_{H,1} \leq CH\|f\|_{L^2(\Omega)} + C\left(\sqrt{\varepsilon} + H + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{\nabla} u^*\|_{C^1(\overline{\Omega})},$$

where u^* is the solution to the homogenized problem and with the constant C independent of H, ε, f and u^*. Although the error estimate was not improved compared with the MsFEM with linear boundary conditions on each coarse element, numerical experiments showed that this approach outperforms the classical one.

This work has been pursued in [122], in which the approach was adapted to address multi-scale problems on a perforated domain Ω^ε, considering the following problem. Find $u_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$, the solution to

$$-\Delta u_\varepsilon = f \text{ in } \Omega^\varepsilon, \quad u_\varepsilon = 0 \text{ on } \partial\Omega^\varepsilon.$$

The main novelty of this approach is the enrichment of the multi-scale finite element space using bubble functions. In the case of periodic perforations in two dimensions, with a sufficient regularity of the right-hand side f, the approximation u_H satisfies

$$|u_\varepsilon - u_H|_{H,1} \leq C\varepsilon \left(\sqrt{\varepsilon} + H + \sqrt{\frac{\varepsilon}{H}} \right) \|f\|_{H^2(\Omega)},$$

with a constant C independent of H, ε and f, but depending on the geometry of the mesh and other parameters of the problem.

Next, in [63], the application of MsFEM à la Crouzeix-Raviart has been extended to advection-diffusion problems with non-homogeneous boundary conditions. The following problem has been considered. Find $u_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$, the solution to

$$\begin{cases} -\operatorname{div}(A_\varepsilon \boldsymbol{\nabla} u_\varepsilon) + \boldsymbol{w} \cdot \boldsymbol{\nabla} u_\varepsilon &= f \quad \text{in } \Omega^\varepsilon, \\ u_\varepsilon &= 0 \quad \text{on } \partial B^\varepsilon \cap \Omega^\varepsilon, \\ u_\varepsilon &= g \quad \text{on } \partial\Omega \cap \Omega^\varepsilon, \end{cases}$$

with $\boldsymbol{w}$ the advection term. Through several numerical examples, it was shown that this particular method outperforms more conventional MsFEM approaches. The study of MsFEM for advection-diffusion problems has been the subject of numerous theses [141, 128, 33].

Recently, [123] developed another MsFEM approach for diffusion problems with oscillatory diffusion coefficient, in which the basis functions associated to faces and elements are enriched using Legendre polynomials. This approach allows to reduce the "resonance error" in the regime where the coarse mesh size H is of order of the small scale ε of the oscillations. Besides, MsFEM for elliptic distributed optimal control problem with rough coefficients has been investigated in [38, 39].

In addition other works [99, 46] deal with adaptative MsFEM by computing an a posteriori error estimate which depends on the coarse mesh size H and the fine mesh size h.

1.5.2 Extension of MsFEM to incompressible flows

The first use of MsFEM for incompressible flow was done in [136], in which the authors proposed an MsFEM à la Crouzeix–Raviart (i.e. only the mean value of the velocity is continuous through the faces of the coarse elements) for solving the Stokes problem (1.3). However, they reached an error estimate that was not entirely satisfactory since depending on the oscillating nature of the exact solution on the faces of the coarse elements.

Then, [111] generalized this method to arbitrary sets of weighting functions used to enforced continuity across the faces of coarse elements. They gave an error bound for a particular set of weighting functions in a periodic setting in two dimensions (the so-called CR3 MsFEM, in

which the three weighting functions on faces are $\boldsymbol{e_1}$, $\boldsymbol{e_2}$, the canonical basis of $\mathbb{R}^2$, and the normal component of a linear polynomial), which reads as follows

$$\begin{aligned}&|\boldsymbol{u_\varepsilon} - \boldsymbol{u_H}|_{H,1} + \varepsilon \|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \\ &\leq C\varepsilon \left[\left(H + \sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}} \right) \left(\|\boldsymbol{f}\|_{H^2(\Omega)\cap C^1(\overline{\Omega})} + \|p^*\|_{H^2(\Omega)} \right) \right],\end{aligned}$$

where p^* is the homogenized pressure and C is a constant depending only on the mesh regularity and the perforation pattern.

Next, [135] extended this approach for solving the Oseen problem (1.2) in heterogeneous media. However, the applications were limited for small Reynolds number. Recently [79] developed an enriched non-conforming MsFEM for solving the Stokes problem in heterogeneous media, which generalized the method proposed in [111], by considering high-order sets of weighting functions on the faces as well as in the elements, rather than just high-order sets of weighting functions on the faces.

1.5.3 Conclusion

In this section, several MsFEMs have been described. However, it is relevant to note that in all the works presented in this review, the numerical applications have been considered only in two-dimensional domains. Furthermore, the study of the different estimates (obtained by assuming that the computation of the local basis functions is exact) presented in this section is not straightforward. Indeed, the errors converge to zero as H and ε tend to zero together (the convergence of H to zero depending, nevertheless, on that of ε). However, for a given ε (fixed), the error between the exact solution and the multi-scale approximation is not expected to tend to zero as the coarse mesh size H decreases (in the regime nevertheless in which $H > \varepsilon$), but will be bounded by a quantity depending on ε. An optimal value of H that minimises the error can be found.

This thesis work is the continuation of the study led in this latter work [79]. Especially, this thesis focus on the discrete level by showing the well-posedness of the discrete local problems and by deriving an error estimate for the MsFEM approximations. We propose also an extension of this method for solving the Oseen problem. The main contributions of the thesis are listed in Section 1.6. The Multi-scale Finite Element Methods (MsFEMs) described in this thesis are intended to be incorporated into the software TrioCFD. However, for the sake of convenience, the preliminary numerical applications of this method are done in FreeFEM software [93], a more academic software, chosen for its flexibility to implement new finite elements and its ease of use for implementing and testing different methods. FreeFEM is an open-source parallel finite element software. It allows to solve problems involving partial differential equations arising from several branches of physics, such as fluid mechanics, fluid-structure interactions and many others. In FreeFEM, the problem to be solved is defined in terms of its variational formulation. FreeFEM is written in C++ and its language is a C++ idiom. For more complex applications, we also take advantage of the fact that FreeFEM interfaces different libraries such as the partitioner METIS [118] or the PETSc library [28].

1.6 Contributions of the thesis

The main contributions of this thesis can be divided in two parts. Indeed, the first contributions (**Contributions 1 to 3**) concern the development of preliminary tools, useful for the analysis of the Multi-scale Finite Element Method (MsFEM). However, it is important to note that these preliminary tools are not limited to the study of the MsFEM but can be re-employed in other contexts. **Contribution 1** concerns the derivation of a quantitative sharp convergence rate of the homogenization of the Stokes problem in a perforated domain. The main theorem of this contribution is one of the main ingredient of the derivation of an error bound for the MsFEM applied to the Stokes problem. **Contribution 2** concerns the introduction of a new family of non-conforming finite element in three dimensions on tetrahedra in the vein of the non-conforming finite element introduced in [132] in two dimensions on triangles. These families of non-conforming finite element, in two and three dimensions, allow to prove the discrete well-posedness of local problems arising from the MsFEM applied to the Stokes or the Oseen problem. Indeed, for these locals problems involving Lagrange multipliers, the well-posedness cannot be shown with classical pairs of finite elements. Consequently, for numerical purposes, **Contribution 3** concerns the implementation of the new non-conforming finite elements in FreeFEM source code. Afterwards, the other contributions (**Contributions 4 to 7**) concern more specifically the analysis of the MsFEM from a theoretical and numerical point of view.

Contribution 1 The first contribution is concerned with the derivation of a quantitative sharp convergence rate of the homogenization of the Stokes problem (1.3) in a periodic perforated domain. This contribution is presented in Chapter 6 and led to a publication [30]. It is well-known that the homogenized or effective equations for the Stokes system, in a periodic perforated domain, is Darcy's law for the velocity $\boldsymbol{u}$ and the homogenized pressure p^*,

$$\left\{\begin{array}{rcll} \boldsymbol{u} & = & \frac{1}{\nu}A^*(\boldsymbol{f}-\boldsymbol{\nabla}p^*) & \text{in } \Omega, \\ \operatorname{div}\boldsymbol{u} & = & 0 & \text{in } \Omega, \\ \boldsymbol{u}\cdot\boldsymbol{n} & = & 0 & \text{on } \partial\Omega. \end{array}\right.$$

where A^* is a constant permeability tensor (see Proposition 6.5). We establish a sharp convergence rate $O(\sqrt{\varepsilon})$ for the energy norm of the difference of the velocities and the pressures, where ε represents the size of the solid obstacles. This is achieved by using a two-scale asymptotic expansion of the Stokes equations and a new construction of a cut-off function which avoids the introduction of boundary layers. This $\sqrt{\varepsilon}$ error estimate has been first derived by [111] in a two-dimensional space, and was independently and further improved in [156] for any space dimension $d \geq 2$, both under the assumption of isolated solid obstacles (as in Figure 6.1). However, this assumption is not physically realistic in dimension $d \geq 3$. Therefore, we extend this $\sqrt{\varepsilon}$ error estimate to the case of connected solid obstacles in any dimension space $d \geq 2$.

Before stating the main result (in loose terms, see Theorem 6.9 for a precise statement) we need to introduce some notations. Actually the Darcy velocity $\boldsymbol{u}$ does not see the solid obstacles and thus cannot be a good approximation of the original velocity $\boldsymbol{u}_\varepsilon$. It must be corrected by

introducing a better approximation which is

$$\boldsymbol{u_2}\left(x,\frac{x}{\varepsilon}\right)=\frac{1}{\nu}\sum_{i=1}^{d}\boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\left(f_i(x)-\frac{\partial p^*}{\partial x_i}(x)\right),$$

where the local velocities $\boldsymbol{\omega_i}$ are solutions of cell problems (6.6). Actually the link between $\boldsymbol{u_2}$ and the Darcy velocity $\boldsymbol{u}$ is that $\boldsymbol{u}(x)$ is the average of $\boldsymbol{u_2}(x,y)$ with respect to the periodic variable y. The main novelty is that our analysis applies for the physically relevant case of a porous medium where both the fluid and solid parts are connected. In loose terms, this contribution reads as follows.

Theorem 1.2. Let $\boldsymbol{u_\varepsilon}$, p_ε be the solution to the Stokes problem (1.3) and $\boldsymbol{u_2}$, p^*, their homogenized approximations. Assuming that $\boldsymbol{f}$ is smooth enough, there exists a constant C, independent of ε, such that

$$\frac{\|p_\varepsilon-p^*\|_{L^2(\Omega^\varepsilon)}}{\|p^*\|_{L^2(\Omega^\varepsilon)}}\leq C\varepsilon^{\frac{1}{2}},$$

$$\frac{|\boldsymbol{u_\varepsilon}-\varepsilon^2\boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)}}{|\varepsilon^2\boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)}}\leq C\varepsilon^{\frac{1}{2}},$$

$$\frac{\|\boldsymbol{u_\varepsilon}-\varepsilon^2\boldsymbol{u_2}\|_{L^2(\Omega^\varepsilon)}}{\|\varepsilon^2\boldsymbol{u_2}\|_{L^2(\Omega^\varepsilon)}}\leq C\varepsilon^{\frac{1}{2}}.$$

Theorem 1.2 is stated in terms of relative errors since $\varepsilon^2\boldsymbol{u_2}$ (and thus $\boldsymbol{u_\varepsilon}$) is small. This theorem is used in the derivation of an error estimate for the MsFEM approximation of the Stokes problem in a periodic perforated domain (see Section 9.2).

Contribution 2 The second contribution is the development of a family of scalar non-conforming finite elements of accuracy order two and three with respect to the H^1-norm on tetrahedra. This contribution is presented in Chapter 5 and led to a publication [29]. Indeed, for the class of MsFEM developed in this thesis, the local problems involve polynomial divergence and Lagrange multipliers, and therefore cannot be solved using classical pairs of finite elements. In two dimensions, the non-conforming finite elements introduced by [132] (and presented in Chapter 4) associated with discontinuous piecewise polynomial pressures of order n allow to solve these local problems. However, in three dimensions, no existing finite element pairs led to a discrete problem whose well-posedness could be proved, except the ones developed here. Therefore, in the vein of [132], we introduce a family of scalar non-conforming finite elements of order two and three with respect to the H^1-norm on tetrahedra. Their vector-valued version generates, together with a discontinuous pressure approximation of order one and two respectively, an inf-sup stable finite element pair of convergence order two and three for the Stokes problem in energy norm. The scalar (local) degrees of freedom are defined by moments on faces F_α and in tetrahedra K as

$$N_j^{F_\alpha}(v)=\int_{F_\alpha}vL_j^{F_\alpha}\qquad j\geq 0,\ \alpha=1,2,3,4,$$

$$N_j^K(v) = \int_K v M_j^K \qquad j \geq 0,$$

where $(L_j^{F_\alpha})_j$ define an arbitrary basis of $\mathbb{P}_n(F_\alpha)$ of dimension $(n+1)(n+2)/2$, and $(M_j^K)_j$ define an arbitrary basis of $\mathbb{P}_{n-1}(K)$ of dimension $n(n+1)(n+2)/6$. For each element K in the mesh $\mathcal{T}_h$ and any integer $n \geq 1$, we define the set of degrees of freedom $\mathcal{N}_{n+1}(K)$ by

$$\begin{aligned} \mathcal{N}_{n+1}(K) := &\left\{ N_j^{F_\alpha} \qquad 1 \leq j \leq \frac{(n+1)(n+2)}{2},\ \alpha = 1,2,3,4 \right\} \\ &\cup \left\{ N_j^K \qquad 1 \leq j \leq \frac{n(n+1)(n+2)}{6} \right\}. \end{aligned} \tag{1.10}$$

Following [132], we enrich the local space $\mathbb{P}_{n+1}(K)$ with a proper subspace of $\mathbb{P}_{n+2}(K)$, denoted by $\Sigma_{n+2}(K) \subset \mathbb{P}_{n+2}(K)$, with a trivial intersection with $\mathbb{P}_{n+1}(K)$. Thus, the enriched space $V_{n+1}(K)$ is given by

$$V_{n+1}(K) = \mathbb{P}_{n+1}(K) \oplus \Sigma_{n+2}(K).$$

Such spaces V_{n+1} have been found only for the cases $n = 1$ and $n = 2$. Then, the main lemma of this contribution reads as follows.

Lemma 1.3. Let K be a tetrahedron with barycentric coordinates λ_1, λ_2, λ_3, λ_4. Then, the finite element spaces

$$V_2 = \mathbb{P}_2 + \text{span}\{\lambda_1\lambda_2^2,\ \lambda_1\lambda_3^2,\ \lambda_2\lambda_3^2\},$$

and

$$V_3 = \mathbb{P}_3 + \text{span}\{\lambda_1^3\lambda_2,\ \lambda_2^3\lambda_3,\ \lambda_3^3\lambda_4,\ \lambda_4^3\lambda_1,\ \lambda_2^3\lambda_1,\ \lambda_1^3\lambda_4,\ \lambda_4^3\lambda_3,\ \lambda_3^3\lambda_2\},$$

are unisolvent, respectively, with respect to the set of degrees of freedom $\mathcal{N}_2$ and $\mathcal{N}_3$ described in (1.10).

It should be noted that the finite element spaces V_2 and V_3 described in Lemma 1.3 are not unique (see for example the list of all suitable basis V_3 in Appendix A.2). Furthermore, for the case $n = 1$, we recover the finite element proposed by [50].

Contribution 3 As explained above, the discrete well-posedness of the local problems involved in MsFEM is shown with non-classical finite element. Indeed, in two dimensions we made use of the family of finite elements introduced in [132] and presented in Chapter 4. In three dimensions, we made use of the family of finite elements introduced in Chapter 5 and in [29]. In both cases, these finite elements are associated with discontinuous polynomial pressure approximation of order n. Consequently, for numerical purposes, i.e. the computations of the local basis functions, these new finite elements have been implemented in FreeFEM [93] source code. First, we implemented the non-conforming finite element of accuracy order three, introduced in [132], in FreeFEM source code, and it is known under the name *P3pnc*[1]. The non-conforming

[1] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P3pnc.cpp

finite element of accuracy order two, known under the name $P2pnc$[2] was already implemented. For numerical applications in three dimensions, we implemented the two non-conforming finite elements, of order two and three discussed in Chapter 5 and in [29]. They are known respectively under the name $P2pnc3d$[3] and $P3pnc3d$[4]. **It is pertinent to highlight that as well as being useful for this thesis, these non-conforming finite elements can now be used by any FreeFEM users to solve physical problems.**

Contribution 4 We propose the first error estimate for the approximation of the Stokes problem in a periodic perforated domain using the MsFEM developed in this thesis, which allows to show the convergence of the proposed method. The derivation of the error estimate is presented in Section 9.2 and is based on the Strang Lemma, which is the classical lemma for error estimation for non-conforming finite element method, and the quantitative homogenization convergence introduced in Chapter 6. This error estimate is obtained by assuming that the computation of the local basis functions is exact. The methodology used is inspired by the error derivation done in [111]. The main result reads in loose term (see Theorem 9.27 for a precise statement) as follows.

Theorem 1.4. Let $\boldsymbol{f}$ be sufficiently smooth. The following error bound holds between the solution $(\boldsymbol{u_\varepsilon}, p_\varepsilon)$ to the Stokes problem (1.3) in perforated periodic domain and its MsFEM approximation $(\boldsymbol{u_H}, p_H)$.

$$\begin{aligned}&\varepsilon^{-1}|\boldsymbol{u_\varepsilon} - \boldsymbol{u_H}|_{H,1} + \|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \\ &\leq C\Big(H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)} + H^n|p^*|_{H^{n+1}(\Omega)} + \left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\Big)\end{aligned}$$

where the constant C is independent of ε, H and $\boldsymbol{f}$.

More precisely, at H fixed, the term $p_\varepsilon - p_H$ is $O(1)$ as ε goes to zero. Besides, recalling that in a periodic domain, we have $\boldsymbol{u_\varepsilon} \approx \varepsilon^2 \boldsymbol{u}(x, \frac{x}{\varepsilon})$ (and that at each derivation of $\boldsymbol{u}(x, \frac{x}{\varepsilon})$, a factor $1/\varepsilon$ comes out), then, the above error estimates for the velocity and pressure essentially says that the relative errors are of the order $\left(H^n + \sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)$. In this sum, the term $\sqrt{\varepsilon}$, which comes from the boundary layers in the homogenization process, is negligible. Indeed, the dominant term is the classical "resonance error" $\sqrt{\frac{\varepsilon}{H}}$, which says that the coarse mesh size H should be much larger than ε to have a good approximation. However, the presence of the term H^n allows to choose an H big enough without significantly reducing the accuracy of the approximation, which shows the interest of considering high-order methods. The optimal value

[2] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P2pnc.cpp

[3] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P2pnc_3d.cpp

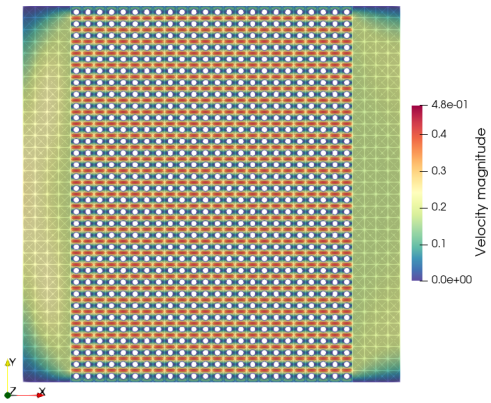
[4] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P3pnc_3d.cpp

of the coarse mesh size H which minimises the error is given by

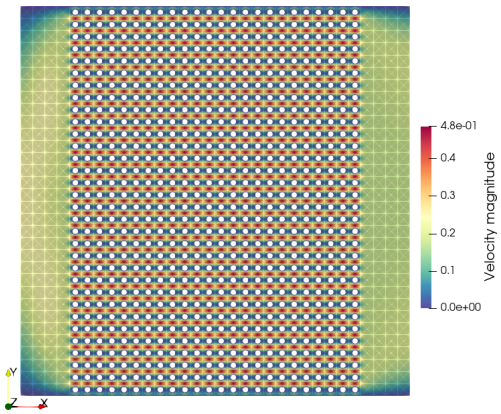
$$H_{opt} = \left(\frac{\sqrt{\varepsilon}}{2n}\right)^{\frac{1}{n+\frac{1}{2}}}.$$

Furthermore, using $H = H_{opt}$ in the error estimate, we note that the error becomes of order $\sqrt{\varepsilon}$ as $n \to \infty$ (we recover the order of convergence of homogenization).

Contribution 5 We extend the MsFEM, initially developed for the Stokes problem (see for example Figure 1.12), to the Oseen problem (1.2) in Part IV (see for example Figure 1.13). The main difference with MsFEM for the Stokes problem is that in this case local problems are Oseen type problems, with the same Oseen velocity $\boldsymbol{U_0}$ as in the global problem. Furthermore, through numerical experiments, we show that this MsFEM allows to deal with high Reynolds number, and that the MsFEM Petrov–Galerkin formulation (in which the test functions are solutions of local problems derived from the Oseen adjoint problem and the trial functions are solutions of local problems derived from the Oseen problem) does not lead to more accurate results.

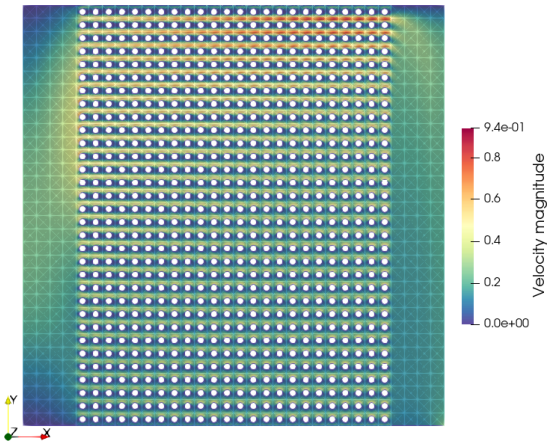


(a) MsFEM approximation ($n = 1$, 4096 coarse elements).

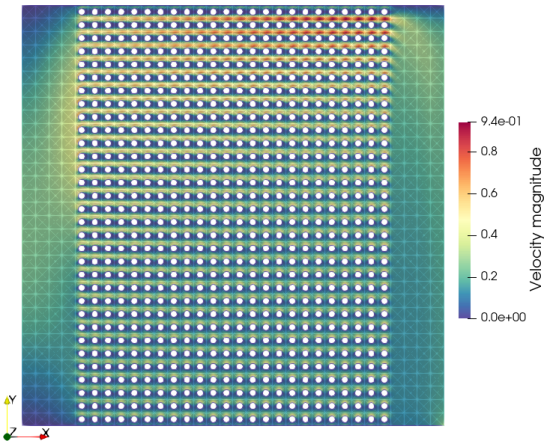


(b) Reference solution computed on a fine mesh.

Figure 1.12: Simulation of a Stokes flow in an open channel, $\nu = 1$, flow from left to right with a parabolic inlet, a free outlet and zero velocity at the top, the bottom and on osbtacles.



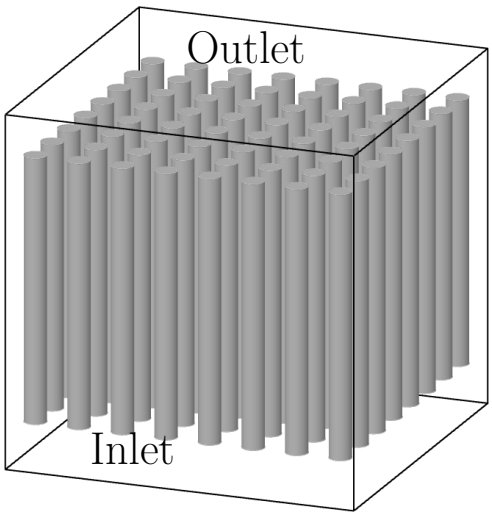
(a) MsFEM approximation ($n = 2$, 4096 coarse elements).

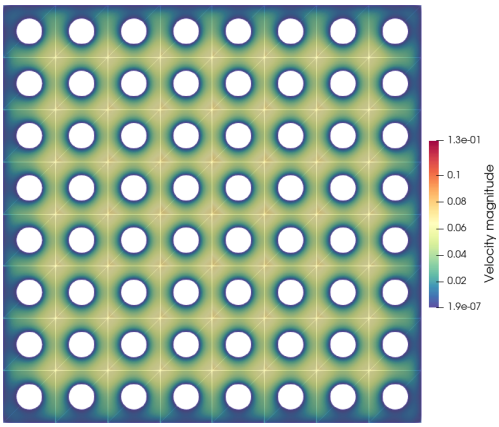


(b) Reference solution computed on a fine mesh.

Figure 1.13: Simulation of an Oseen problem in a two-dimensional open channel, Reynolds number $Re = 2000$ ($\nu = 5 \times 10^{-4}$, $\boldsymbol{U_0} = [y(1-x^2), -x(1-y^2)]^t$), flow from left to right with a parabolic inlet, a free outlet and zero velocity at the top, the bottom and on osbtacles.

Contribution 6 We implement the MsFEM for solving the Stokes and the Oseen problems in two and three dimensions (see for example Figure 1.14) in a high-performance framework in FreeFEM [93]. Indeed, for solving the local problems, we develop two levels of parallelism: a parallelism between each coarse element (allowing to consider each coarse element in parallel) and a parallelism inside each coarse element (allowing to compute local problems themselves in parallel). What is more, for the sake of comparison, we implement solvers to compute reference solutions on fine meshes ("reference meshes"). The details of these implementations are presented in Appendices D and E. In addition, we propose the first implementation at order $n = 2$ of the MsFEM under study (see for example Figure 1.13).

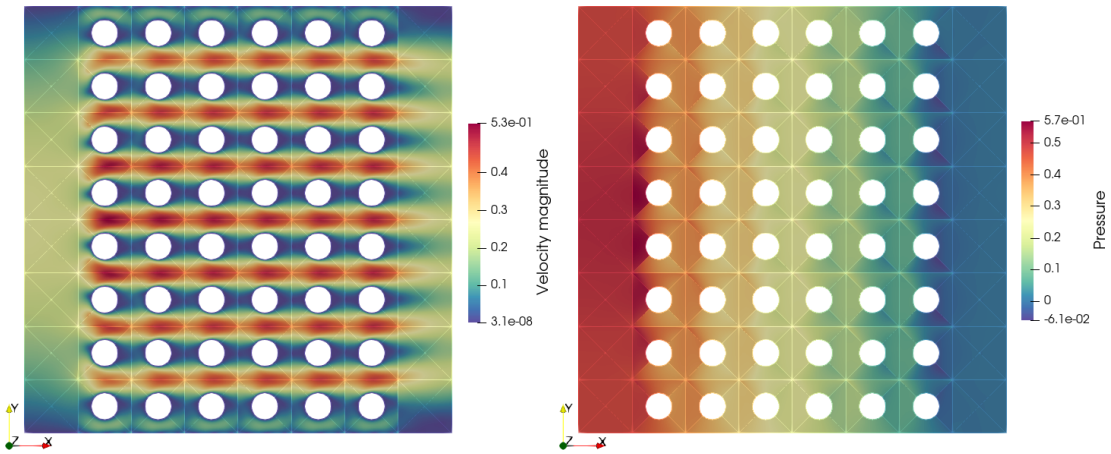




(a) Cut of the velocity magnitude field along the z axis (transverse to the cylinders).

Figure 1.14: MsFEM approximation of Stokes flow in three-dimensional open channel with a parabolic inlet, a free outlet and zero velocity on the other boundaries ($n = 2$, 3072 coarse elements).

Contribution 7 We propose a methodology to solve the stationnary Navier–Stokes problem (1.1) on a coarse grid using adaptative MsFEM basis functions in Chapter 16 (see for example Figure 1.15). As the problem involves non-linear terms, we propose a pseudo-transient approach to solve it. However, the main difficulty for solving iteratively the Navier–Stokes problem using a multi-scale approach is that this problem cannot be solved using only one fixed basis, derived either from the Oseen problem (with a given arbitrary Oseen velocity $\boldsymbol{U_0}$) or from the Stokes problem (equivalent to the Oseen problem with a null advection velocity). Indeed, at each iteration the global problem to be computed changes since the advection velocity changes. In particular, as soon as the advection velocity of the global problem differs too much from the one used for the basis functions, the basis functions become non-adapted to the problem. Consequently, to overcome this difficulty, at each iteration, new basis functions are computed so that the basis functions remain well-adapted to the problem. In addition, we propose a discussion on the possibility of solving such a problem using the Generalized Finite Element Method (GMsFEM) [69].



(a) Velocity magnitude.

(b) Pressure field.

Figure 1.15: MsFEM approximation ($n = 2$, 256 coarse elements) of the steady Navier–Stokes in a two-dimensional open-channel, Reynolds number $Re = 1000$ ($\nu = 1 \times 10^{-3}$), flow from the left to right with a parabolic inlet, a free outlet and zero velocity at the top, the bottom and on osbtacles.

Chapter 2

Introduction de la thèse en français

Contents

2.1 Motivation

2.1.1 Milieux multi-échelles

Cette thèse porte sur le développement de méthodes de calcul pour les écoulements incompressibles dans les milieux multi-échelles. Un milieu multi-échelle est un milieu dont les caractéristiques significatives, telles que le coefficient de diffusion ou la conductivité thermique,

ou la géométrie elle-même, varient à plusieurs échelles différentes (magnitudes) en espace ou présentent une nature oscillante. Un tel dernier milieu à géométrie multi-échelle peut être rencontré dans de nombreux domaines. On peut penser, par exemple, au sol, qui est constitué de matériaux de différentes tailles ou à une ville dense, où les bâtiments, les voitures ou d'autres installations créent des obstacles de différentes tailles à l'écoulement de l'air ou de l'eau. Dans cette thèse, nous nous intéressons à cette dernière catégorie de milieux multi-échelles. Dans la littérature, ils peuvent être désignés par différents noms : milieux poreux en géosciences, domaines congestionnés ou domaines hétérogènes dans le domaine nucléaire, ou encore domaines perforés.

En effet, au Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), nous nous intéressons au calcul de l'écoulement dans les cœurs de réacteurs nucléaires, comme dans une cuve sous pression d'un réacteur à eau pressurisée (REP), qui sont des milieux multi-échelles avec des obstacles dont les tailles varient sur plus de quatre ordres de grandeur et dont les échelles d'intérêt s'étendent sur plusieurs ordres de grandeur. Pour avoir une idée des échelles en jeu, nous donnons, suit, quelques valeurs numériques. La longueur totale de la cuve sous pression est d'environ 14 mètres. Le cœur du réacteur a un diamètre d'environ 4 à 5 mètres et comprend entre 157 et 200 assemblages de combustible (voir la Figure 2.1). L'assemblage combustible mesure environ 4 à 5 mètres de long et se compose de barres ou de crayons de combustible généralement regroupés en un réseau carré de 14×14 à 17×17. L'épaisseur des ailettes de mélange est d'environ 1 centimètre. En résumé, la cuve sous pression contient donc différentes échelles et l'échelle globale est environ 10^4 fois plus grande que l'échelle locale.

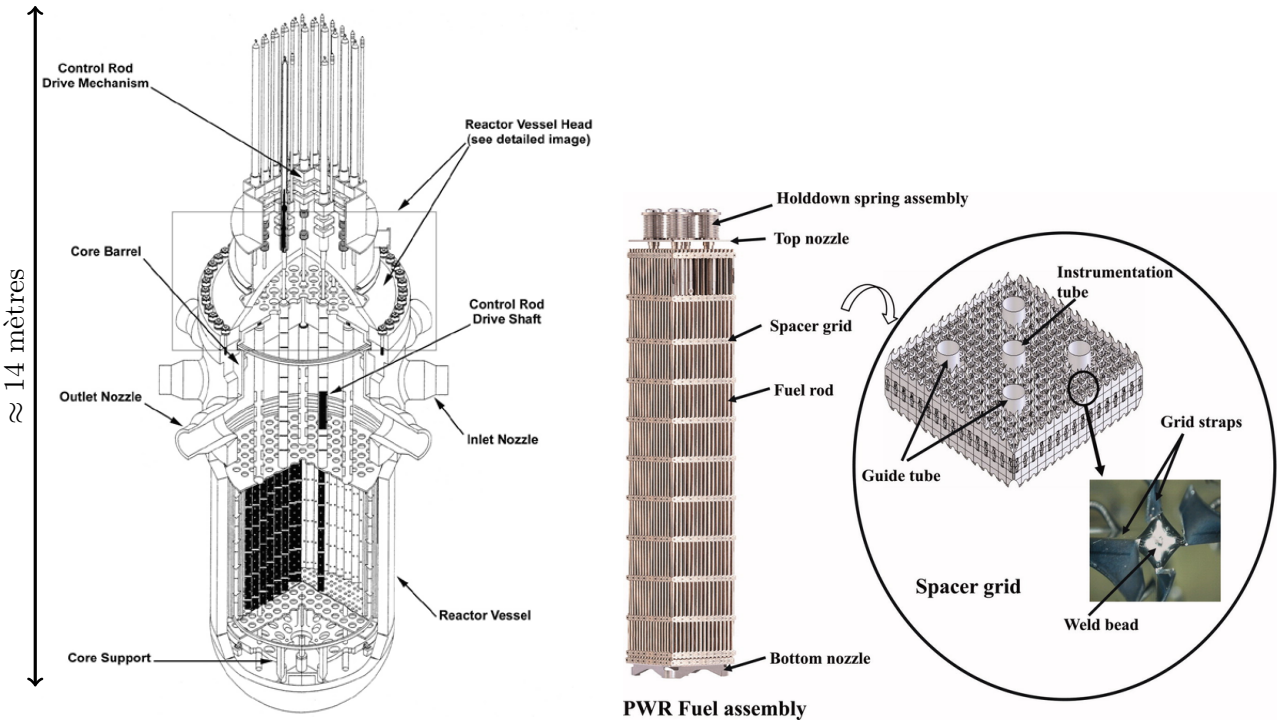


Figure 2.1: Schéma d'un REP (gauche) et d'un assemblage de combustible (droite) [1].

Une modélisation typique d'un tel milieu multi-échelle consiste à considérer un domaine $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ perforé par un ensemble d'obstacles périodiques B^ε, où $\varepsilon > 0$ est la période

(ou l'échelle de longueur des hétérogénéités). Il est important de noter que ce domaine peut également être non périodique. Cependant, pour d'autres résultats quantitatifs (voir par exemple le Chapitre 6 et le Chapitre 9), en particulier pour l'homogénéisation, la nature périodique est l'un des cadres possibles. Il est à noter que, des résultats quantitatifs d'homogénéisation peuvent également être obtenus dans le cadre stochastique, mais avec des hypothèses supplémentaires telles que l'ergodicité et la stationnarité [22].

Un exemple illustratif d'un domaine perforé est donné aux Figures 2.2 and 2.3 respectivement en deux et trois dimensions. Le domaine Ω est divisé en une partie solide fixe B^ε et sa partie fluide complémentaire Ω^ε, c'est-à-dire $\Omega^\varepsilon = \Omega \setminus B^\varepsilon$.

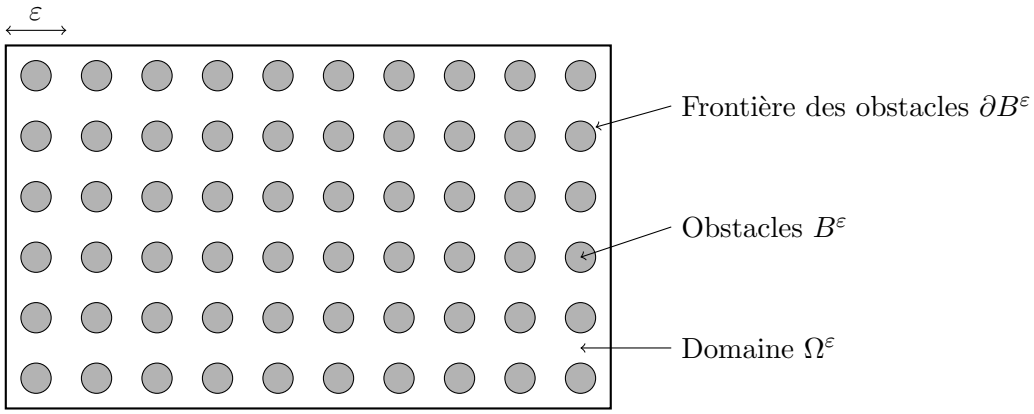


Figure 2.2: Domaine bidimensionnel poreux Ω^ε, obtenu à partir d'un domaine Ω perforé par un ensemble d'obstacles solides B^ε.

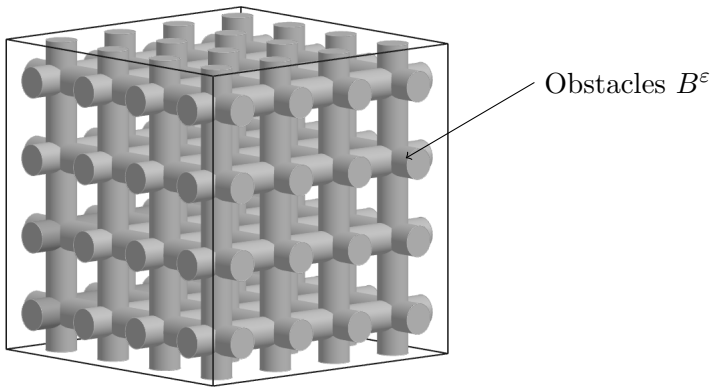


Figure 2.3: Domaine tridimensionnel poreux Ω^ε, obtenu à partir d'un domaine Ω perforé par un ensemble d'obstacles solides B^ε.

2.1.2 Modélisation de l'écoulement

Une fois le milieu multi-échelle donné, le problème de l'écoulement dans ce milieu doit être modélisé. Dans le cas des réacteurs à eau pressurisée (REP), l'eau dans le cœur atteint environ 325°C et est maintenue à une pression élevée d'environ 150 bars afin d'éviter qu'elle ne se transforme en vapeur. Dans les conditions nominales, le nombre de Reynolds dans le cœur est de l'ordre de 1×10^5, ce qui suggère que l'écoulement est entièrement turbulent [2, 59]. Dans ce travail, nous limitons notre attention aux problèmes incompressibles et monophasique comme

une étape préliminaire avant d'envisager des applications plus complexes. En outre, nous ne considérons pas les écoulements turbulents, car cela nécessiterait l'introduction d'un modèle de turbulence, tel que le modèle $k-\varepsilon$ proposé par Hanjalić et Launder [89] ou le modèle $k-\omega$ proposé par Wilcox [167]. Cela dépasse le cadre de cette thèse.

Nous notons la vitesse $\boldsymbol{u_\varepsilon}$ et la pression p_ε pour rappeler que nous résolvons l'écoulement dans un domaine multi-échelle avec des hétérogénéités d'échelle de longueur (ou de période) de taille ε. Un aperçu des différents problèmes d'écoulement est donné ci-dessous.

Les équations de Navier–Stokes Les équations de Navier–Stokes incompressible peuvent être utilisées pour modéliser les écoulements fluide dans le domaine Ω^ε. Le problème de Navier–Stokes incompressible et stationnaire avec des conditions de Dirichlet homogènes est de trouver la vitesse $\boldsymbol{u_\varepsilon} : \Omega^\varepsilon \to \mathbb{R}^d$ et la pression $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution de

$$\left\{\begin{array}{rcll} -\nu\Delta \boldsymbol{u_\varepsilon} + (\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon} + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{dans } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u_\varepsilon} &=& 0 & \text{dans } \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{sur } \partial\Omega^\varepsilon, \end{array}\right. \tag{2.1}$$

avec $\boldsymbol{f}$ la force appliquée et $\nu > 0$ la viscosité.

Les équation de Oseen L'approximation de Oseen est une linéarisation partielle des équations de Navier–Stokes. Le terme non-linéaire $(\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}$ est remplacé par le terme linéaire $(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}$ avec $\boldsymbol{U_0}$ une vitesse connue, appelée vitesse de Oseen. Le problème de Oseen en régime stationnaire avec des conditions de Dirichlet homogènes est de trouver la vitesse $\boldsymbol{u_\varepsilon} : \Omega^\varepsilon \to \mathbb{R}^d$ et la pression $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution de

$$\left\{\begin{array}{rcll} -\nu\Delta \boldsymbol{u_\varepsilon} + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon} + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{dans } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u_\varepsilon} &=& 0 & \text{dans } \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{sur } \partial\Omega^\varepsilon, \end{array}\right. \tag{2.2}$$

avec $\boldsymbol{U_0}$ la vitesse de Oseen, $\boldsymbol{f}$ la force appliquée et $\nu > 0$ la viscosité.

Dans les deux problèmes ci-dessus, le terme convectif, non-linéaire $(\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}$ pour le problème de Navier–Stokes, et linéaire $(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}$ pour le problème de Oseen, accroît la difficulté de l'analyse mathématique et des solutions numériques de ces problèmes. La prédominance de ce terme convectif peut être prédite par le nombre de Reynolds, noté Re, un nombre sans dimension, défini comme le rapport entre les forces d'inertie et les forces visqueuses dans l'écoulement,

$$Re = \frac{U_c L}{\nu},$$

où U_c est une vitesse caractéristique, ν la viscosité et L une longueur caractéristique.

Les équations de Stokes Le nombre de Reynolds des écoulements dans les milieux poreux est généralement faible et les forces d'inertie peuvent être négligées par rapport aux forces visqueuses. Cette simplification conduit au problème de Stokes. Le problème de Stokes en régime stationnaire avec des conditions aux limites homogènes de Dirichlet consiste à trouver la

vitesse $\boldsymbol{u_\varepsilon} : \Omega^\varepsilon \in \mathbb{R}^d$ et la pression $p_\varepsilon : \Omega^\varepsilon \in \mathbb{R}$ solution de

$$\left\{ \begin{array}{rcll} -\nu\Delta \boldsymbol{u_\varepsilon} + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{dans } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u_\varepsilon} &=& 0 & \text{dans } \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{sur } \partial\Omega^\varepsilon, \end{array} \right. \tag{2.3}$$

avec $\boldsymbol{f}$ la force appliquée et $\nu > 0$ la viscosité.

La première approche pour étudier l'écoulement dans un milieu multi-échelle est par le calcul direct de l'un des problèmes mentionnés ci-dessus (les problèmes de Stokes, d'Oseen ou de Navier–Stokes). Cependant, la simulation de l'écoulement dans un milieu multi-échelle avec de nombreux obstacles est très difficile. En effet, dans un tel milieu, l'écoulement est influencé par des phénomènes qui se produisent aux échelles les plus fines. Ainsi, pour capturer ces échelles les plus fines, il est nécessaire d'utiliser un maillage très fin, ce qui conduit souvent à des simulations irréalisables en raison du manque de ressources informatiques. Au CEA, les études de dynamique des fluides numérique (CFD) des coeurs de réacteurs nucléaires sont réalisées à l'aide du logiciel open-source TrioCFD [45]. Il s'agit d'un logiciel développé depuis une vingtaine d'années à la Division de l'Énergie Nucléaire du CEA. TrioCFD est un code orienté objet, massivement parallèle, implémenté en C++, qui a été développé pour être utilisé dans une variété d'études scientifiques et industrielles et d'applications de recherche. Le code a été conçu pour traiter efficacement différents problèmes physiques, tels que les écoulements turbulents, le couplage fluide/solide, les écoulements multiphasiques et les écoulements dans les milieux poreux [32, 153, 147]. Une vue d'ensemble du code TrioCFD est disponible dans [16, 17].

À titre d'illustration, définissons un maillage $\mathcal{T}_H$ d'un domaine polyédrique perforé périodique Ω^ε comme une décomposition de Ω^ε en polyèdres T, chacun ayant un diamètre d'au plus H. Nous définissons l'espace classique des vitesses $V = H^1_0(\Omega^\varepsilon)^d$ et l'espace des pressions $M = L^2_0(\Omega^\varepsilon) = \{q \in L^2(\Omega^\varepsilon),\ t.q.\ \int_{\Omega^\varepsilon} q = 0\}$ muni respectivement de la semi-norme H^1 $|\cdot|_1$ et de la norme L^2 $\|\cdot\|_0$. Pour les approches non-conformes, nous introduisons l'espace de Sobolev brisé

$$H^1(\mathcal{T}_H) = \left\{ \boldsymbol{u} \in L^2(\Omega^\varepsilon)^d \text{ telle que } \boldsymbol{u}|_T \in H^1(T)^d,\ \forall T \in \mathcal{T}_H \right\}.$$

Nous définissons aussi la semi-norme H^1-brisée $|\boldsymbol{u}|_{H,1}$ dans l'espace H^1 brisé par

$$|\boldsymbol{u}|_{H,1} = \left(\sum_{T \in \mathcal{T}_H} |\boldsymbol{u}|^2_{H^1(T)} \right)^{\frac{1}{2}}.$$

Nous pouvons alors envisager de résoudre les équations de Stokes (1.3) dans ce domaine Ω^ε. Supposons que le problème de Stokes soit discrétisé avec les éléments finis de Crouzeix–Raviart [61], il est alors bien connu que l'erreur entre la solution exacte $(\boldsymbol{u_\varepsilon}, p_\varepsilon)$ et son approximation par éléments finis $(\boldsymbol{u_H}, p_H)$ est donnée par

$$\frac{1}{|\boldsymbol{u_\varepsilon}|_{H^1(\Omega^\varepsilon)}} \left(|\boldsymbol{u_\varepsilon} - \boldsymbol{u_H}|_{H,1} + \|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \right) \leq CH \underbrace{\left(|\boldsymbol{u_\varepsilon}|_{H^2(\Omega^\varepsilon)} + |p_\varepsilon|_{H^1(\Omega^\varepsilon)} \right) \frac{1}{|\boldsymbol{u_\varepsilon}|_{H^1(\Omega^\varepsilon)}}}_{\beta_\varepsilon},$$

avec C une constante indépendante du maillage. Or, on sait que dans le cas périodique (voir le Chapitre 6 pour plus de détails), fonctionnellement, c'est-à-dire pour la fonction et ses dérivées d'ordre quelconque, on a $\boldsymbol{u_\varepsilon} \approx \varepsilon^2 \boldsymbol{u}(x, \frac{x}{\varepsilon})$ et donc le terme β_ε est d'ordre ε^{-1} (puisque à chaque dérivation de $\boldsymbol{u}(x, \frac{x}{\varepsilon})$, un facteur $1/\varepsilon$ apparaît). Par conséquent, l'obtention d'une bonne approximation de la solution nécessite une taille de maillage H beaucoup plus petite que ε. Comme ε peut être très petit en pratique, il est souvent impossible de satisfaire cette condition. Par exemple, pour capturer les échelles les plus fines de l'écoulement dans la cuve sous pression présentée ci-dessus, un maillage typique pour discrétiser ce domaine peut contenir des centaines de millions d'éléments. Il est clair que les simulations directes sur un tel maillage nécessitent une quantité massive de mémoire et de temps de calcul, qui peut dépasser la puissance de calcul actuelle. Même si un tel calcul est possible, il ne peut être effectué régulièrement, mais seulement dans des cas exceptionnels, tel que les calculs CFD des écoulements turbulents autour des grilles de mélange des REP avec plus d'un milliard de degrés de liberté effectués avec TrioCFD [32].

Afin de surmonter ces difficultés, une classe populaire de méthodes numériques, basée sur la stratégie de calcul de la moyenne ou "remontée d'échelle", a été développée. Elle correspond au processus d'approximation des modèles à l'échelle fine par des modèles à l'échelle grossière. Dans ce qui suit, nous présentons deux méthodes de remontée d'échelle, le volume élémentaire représentatif (VER) et l'homogénéisation.

2.1.3 Approches basées sur la stratégie de la moyenne ou de la remontée d'échelle

Volume élémentaire représentatif (VER) Le VER est une première méthode pour "mettre à l'échelle" ou moyenner des modèles à échelle fine [149]. L'idée de cette approche est la suivante [102]. Soit u une fonction à valeur réelle sur un domaine Ω qui décrit certaines quantités physiques avec des oscillations spatiales rapides. Pour lisser cette fonction, on considère des moyennes locales de la forme

$$\langle u \rangle(x) = \int_{V(x)} u(y) \mathrm{d}y,$$

où $V(x)$ est un petit voisinage du point x de la taille d'un VER. Si les oscillations de u reflètent le comportement de la grandeur physique en question à une "micro-échelle", la fonction moyennée $\langle u \rangle$ est censée décrire ses propriétés à une plus grande échelle, c'est-à-dire à une "macro-échelle". Plusieurs travaux [56, 120] ont appliqué cette stratégie pour obtenir un modèle à grande échelle de l'écoulement dans les milieux poreux.

L'homogénéisation L'homogénéisation est une autre méthode mathématique qui permet de "mettre à l'échelle" ou de moyenner des équations différentielles [102]. L'idée de l'homogénéisation est de considérer une famille de fonctions u_ε où $\varepsilon > 0$ est un paramètre d'échelle spatiale (longueur). Ensuite, il faut déterminer la limite

$$u = \lim_{\varepsilon \to 0} u_\varepsilon, \tag{2.4}$$

et en considérant cette limite comme le résultat de la procédure de mise à l'échelle. L'objectif de la procédure d'homogénéisation est de déterminer les équations différentielles auxquelles la limite u satisfait et de prouver que (2.4) est valable. Appliquée aux écoulements incompressibles, il est bien connu que l'homogénéisation des équations de Stokes (2.3), c'est-à-dire la recherche du système limite satisfait par la limite de $(\boldsymbol{u}_\varepsilon, p_\varepsilon)$ lorsque ε tends vers zéro [154, 160, 9], conduit à la loi de Darcy pour la vitesse $\boldsymbol{u}$ et la pression homogénéisée p^*,

$$\left\{\begin{array}{rcll} \boldsymbol{u} & = & \frac{1}{\nu}A^*(\boldsymbol{f}-\boldsymbol{\nabla}p^*) & \text{dans } \Omega, \\ \operatorname{div}\boldsymbol{u} & = & 0 & \text{dans } \Omega, \\ \boldsymbol{u}\cdot\boldsymbol{n} & = & 0 & \text{sur } \partial\Omega, \end{array}\right.$$

où A^* est un tenseur de perméabilité constant (voir le Chapitre 6 pour plus de détails). Cette méthode est souvent combinée avec une méthode de remontée d'échelle dans laquelle les propriétés à fine échelle du problème sont moyennées afin de trouver des paramètres macroscopiques effectifs, qui sont ensuite introduits dans le modèle à grande échelle défini par la théorie de l'homogénéisation. Par exemple, dans le cas du problème de Darcy, cette méthode consiste à trouver une perméabilité effective à partir des perméabilités locales. Cette méthode a été appliquée à plusieurs problèmes dans la littérature pour définir des problèmes d'écoulement à grande échelle dans des milieux fortement hétérogènes ou des matériaux composites [169, 51, 110, 53, 77, 109].

L'inconvénient de ces deux approches est que seul le comportement moyen de l'écoulement est pris en compte. En effet, elles ne peuvent pas saisir le comportement de l'écoulement aux échelles les plus fines, par exemple les caractéristiques oscillantes ou fluctuantes, à moins d'utiliser des correcteurs.

Pour surmonter cette limitation, une autre approche, inspirée de ces idées, a été développée pour tenter de résoudre les échelles inférieures à l'échelle du maillage grossier. Au lieu de calculer les propriétés effectives comme dans les approches de remontée d'échelle, les caractéristiques de l'échelle fine sont incorporées dans le calcul à l'échelle grossière par le biais d'une approche multi-échelle. La plupart des méthodes utilisant cette approche consistent à résoudre des problèmes à l'échelle fine, puis à utiliser ces solutions dans un problème à l'échelle grossière. Une revue de certaines méthodes multi-échelles est effectuée dans la Section 2.2. Dans cette thèse, nous nous concentrons sur l'une d'entre elles, à savoir la **Méthode des Éléments Finis Multi-échelle (MsFEM)**, puisqu'elle est la continuation de travaux antérieurs sur les MsFEMs pour les écoulements incompressibles [111, 78, 79]. Dans la Section 2.3, nous présentons les principales idées de la MsFEM, et nous illustrons cette méthode au travers d'un example simple dans la Section 2.4. Puis dans la Section 2.5, nous donnons un aperçu de l'évolution des MsFEMs dans la littérature.

2.2 Une vue d'ensemble des différentes méthodes multi-échelles

Dans cette section, nous donnons un aperçu des différentes méthodes multi-échelles développées dans la littérature.

2.2.1 La méthode variationnelle multi-échelle (VMS)

La méthode variationnelle multi-échelle (VMS) a été introduite pour la première fois par [106, 108] comme une procédure générale permettant d'élaborer des méthodes numériques capables de traiter des phénomènes multi-échelles. Cette méthode est basée sur la décomposition de la solution, $u = \overline{u} + \tilde{u}$, où $\overline{u}$ est la partie résolue ou échelle grossière et $\tilde{u}$ est la partie non résolue ou échelle fine. Il s'agit alors de résoudre numériquement $\overline{u}$ et de tenter de déterminer analytiquement $\tilde{u}$, en l'éliminant du problème pour $\overline{u}$.

2.2.2 Décomposition orthogonale localisée (LOD)

La décomposition orthogonale localisée (LOD) a été introduite pour la première fois dans [97, 98], dans lesquels les auteurs construisent une base d'éléments finis généralisée locale pour les problèmes elliptiques avec des coefficients hétérogènes et très variables. Les fonctions de base sont des solutions de problèmes locaux sur des patchs de noeuds. L'avantage de cette méthode est que les problèmes locaux sont localisés avec une diminution rapide permettant de restreindre le domaine de calcul des problèmes locaux sur un petit patch. Cette méthode a été appliquée avec succès dans la littérature, par exemple pour des problèmes de type Helmhotz [140], pour l'équation des ondes [6] ou pour effectuer l'homogénéisation numérique [95, 13].

2.2.3 Méthode hétérogène multi-échelle (HMM)

La méthode hétérogène multi-échelle (HMM) a été introduite pour la première fois dans [166, 67]. Cette méthode fournit une méthodologie générale pour le calcul numérique efficace des problèmes multi-échelles et multi-physiques, impliquant un modèle macroscopique U et un modèle microscopique u. Ces modèles sont définis sur des domaines différents et sont liés par des opérateurs de reconstruction et de compression. Le but de cette méthode est d'approximer l'échelle macroscopique du système sur un maillage macroscopique qui résout la grande échelle du problème, et d'estimer les données manquantes nécessaires pour le schéma macroscopique à partir du schéma microscopique. Cette méthode comporte donc deux composantes principales : un schéma macroscopique global pour U et une estimation des données manquantes du modèle microscopique pour u. Cette méthode est qualifiée d'hétérogène pour souligner le fait que différents modèles physiques et techniques numériques sont utilisés à différentes échelles et sur différents maillages. Cette méthode a été appliquée avec succès dans la littérature pour différents problèmes tels que les problèmes de diffusion [7], les problèmes d'homogénéisation [68, 66] ou pour la modélisation de fluides complexes [150].

2.2.4 Méthode des volumes finis multi-échelle (MsFVM)

La méthode des volumes finis multi-échelle (MsFVM) a été introduite pour la première fois par [112] pour les simulations d'écoulement souterrain et étendue dans [113] pour les écoulements multiphasiques et le transport dans les milieux poreux. Cette méthode est basée sur la construction de deux ensembles de fonctions de base : le premier est utilisé pour construire les transmissibilités pour le problème à grande échelle et le second est utilisé pour reconstruire le champ de vitesse à petite échelle à partir de la solution à grande échelle. Ces fonctions de

base, qui assurent l'équilibre de la masse aux deux échelles, sont calculées sur un maillage dual. Cette méthode a été principalement appliquée à l'écoulement souterrain et aux simulations de réservoirs [126, 168, 114, 137, 87, 127, 88].

2.2.5 Méthode hybride mixte multi-échelle (méthode MHM)

La méthode hybride mixte multi-échelle (MHM) a été introduite pour la première fois par [91] pour le problème de Darcy et par [19] pour les problèmes elliptiques. Cette méthode vise à incorporer des échelles multiples dans la construction des fonctions de base et est basée sur une formulation mixte : elle relâche la continuité de la composante normale primale par l'action des multiplicateurs de Lagrange, tout en assurant la continuité forte de la composante normale du flux. Les problèmes locaux consistent à trouver des fonctions de base qui approchent l'espace des multiplicateurs de Lagrange. Cette méthode a été appliquée pour des problèmes d'advection-réaction à coefficients hétérogènes [92], pour des problèmes elliptiques du second ordre à coefficients périodiques rugueux [143]. Elle a également été utilisée pour résoudre, dans des milieux hétérogènes, les problèmes de Stokes et de Brinkmann [20], les équations de Maxwell [119], l'équation de Helmholtz [47] ou le problème d'Oseen [18].

2.2.6 Autres approches multi-échelles

Il existe un certain nombre d'autres méthodes en plus de celles mentionnées ci-dessus. On peut citer par exemple la méthode des éléments finis multi-échelles de type Mortar [85] utilisée pour résoudre les écoulements de Stokes et de Darcy ou la méthode de Galerkin discontinu multi-échelle [73, 72] utilisée pour des problèmes elliptiques du second ordre.

2.3 Introduction à la Méthode des éléments finis multi-échelles (MsFEM)

L'idée principale de la méthode des éléments finis multi-échelles (MsFEM) est d'approximer un problème global sur un maillage grossier $\mathcal{T}_H$, qui ne prend pas en compte les détails géométriques (voir par exemple la Figure 2.4b), à l'aide d'une approximation de Galerkin. Cependant, les fonctions de base ne sont plus les fonctions de base polynomiales de Lagrange des éléments finis, mais résolvent des problèmes locaux sur les éléments T du maillage grossier. Ces fonctions sont elles-mêmes approximées numériquement sur un maillage fin $\mathcal{T}_h(T)$ prenant en compte tous les détails géométriques (voir par exemple la Figure 2.4c), ce qui donne l'aspect multi-échelle de cette méthode.

Jusqu'à présent, nous n'avons fait aucune hypothèse sur la forme des éléments du maillage grossier. Ces éléments peuvent être de différents types (triangles, carrés ou polygones en deux dimensions, et tétraèdres, parallélépipède ou polyèdres en trois dimensions). Toutefois, par commodité, les éléments du maillage grossier sont choisis comme étant des triangles et des tétraèdres (respectivement en deux et trois dimensions) dans les différentes illustrations et dans les simulations numériques.

Deux points essentiels doivent être pris en compte à ce stade. Le premier point est que ces fonctions de base locales ne dépendent pas des conditions aux limites et des chargements du problème global. Par conséquent, ces fonctions de base peuvent être réutilisées pour résoudre le problème global mais avec des conditions aux limites et des chargements différents, ce qui rend cette méthode très pertinente pour les études paramétriques. Le deuxième point est que les fonctions de base sont entièrement locales, c'est-à-dire que les fonctions de base associées à un élément grossier donné du maillage grossier sont indépendantes de celles des autres éléments grossiers. Ce point rend le calcul des fonctions de base locales très efficace, puisque leurs calculs peuvent être effectués entièrement en parallèle, c'est-à-dire sans aucune communication entre les différents éléments du maillage grossier.

Cette méthode se compose de deux étapes principales, décrites ci-dessous.

L'étape hors ligne La première étape, appelée étape hors ligne, correspond au calcul des fonctions de base locales dans chaque élément grossier du maillage grossier, en résolvant des problèmes locaux du type $\mathcal{A}_T(\phi, v) = \mathcal{F}_T(v)$. Au cours de cette étape, les matrices du problème global sont assemblées. Cette étape est considérée comme la plus coûteuse car elle nécessite de résoudre, éventuellement en parallèle, plusieurs problèmes locaux sur des mailles fines.

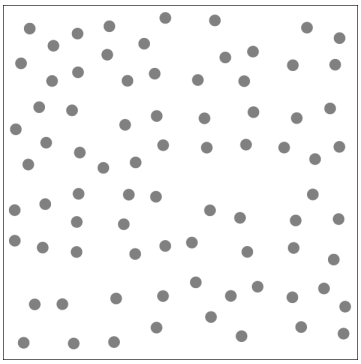
L'étape en ligne Une fois les fonctions de base calculées, la deuxième étape, appelée étape en ligne, consiste à assembler le second membre, à appliquer les conditions aux limites et à résoudre le problème global du type $\mathcal{A}_{\mathcal{T}_H}(u, v) = \mathcal{F}_{\mathcal{T}_H}(v)$ sur le maillage grossier. Cette étape est considérée comme l'étape peu coûteuse, puisqu'elle implique de résoudre un problème sur un maillage grossier qui contient peu d'éléments. Si le problème global doit être recalculé, mais avec des conditions aux limites et des charges différentes, seule l'étape en ligne, c'est-à-dire l'étape peu coûteuse, doit être recalculée.

Une troisième étape consiste ensuite à reconstruire les solutions sur les maillages fins et à post-traiter les résultats.

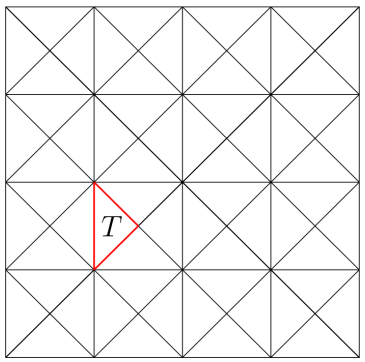
En résumé, la procédure de la MsFEM est la suivante.

Algorithm 2.1 (La procédure de la MsFEM)**.**

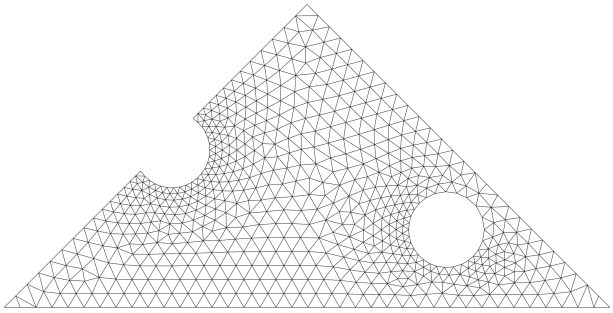
- Partitionner le domaine en un ensemble de mailles grossières qui forme le maillage grossier (voir la Figure 2.4b).
- Étape hors-ligne. Pour chaque élément du maillage grossier (**En parallèle**) :
  - Partitionner l'élément grossier en un maillage fin (voir la Figure 2.4c).
  - Construire les fonctions de base multi-échelles en résolvant les problèmes locaux sur ce maillage fin.
  - Calculer les matrices locales et assembler les matrices globale.
- Étape en ligne. Appliquer les conditions au limites, assembler le second membre et résoudre le problème global sur le maillage grossier.
- Pour chaque élément grossier (**En parallèle**) :
  - Reconstruire la solution fine sur les maillages fins.
  - Post-traiter les résultats.



(a) Domaine hétérogène Ω^ε.



(b) Maillage grossier $\mathcal{T}_H$ et un élement T.



(c) Maillage fin $\mathcal{T}_h(T)$.

Figure 2.4: Description de la précédure MsFEM.

A des fins de comparaison, les solutions peuvent être reconstruites sur un maillage fin, appelé "maillage de référence", donné dans la Figure 2.5.

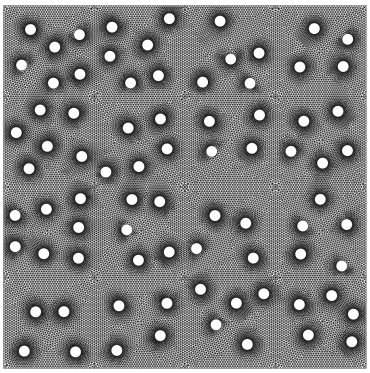


Figure 2.5: Maillage de référence $\mathcal{T}_h(\Omega^\varepsilon)$.

Dans la section suivante, pour une meilleure compréhension, nous illustrons la MsFEM par un exemple simple : un problème de diffusion avec un coefficient de diffusion oscillant en une dimension.

2.4 Un exemple illustratif simple en une dimension : un problème de diffusion oscillante

Soit I l'intervalle $[0,1] \subset \mathbb{R}$ and $f \in L^2(I)$ une fonction donnée. Nous considérons le problème de diffusion suivant avec conditions de Dirichlet homogènes.

$$\begin{aligned} -\frac{\mathrm{d}}{\mathrm{d}x}\left(A_\varepsilon(x)\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}x}(x)\right) &= f(x) \text{ dans } I, \\ u_\varepsilon(0) &= 0, \\ u_\varepsilon(1) &= 0. \end{aligned} \tag{2.5}$$

Nous considérons le coefficient de diffusion oscillant A_ε suivant

$$A_\varepsilon(x) = \frac{1}{2+\cos\left(2\pi(x/\varepsilon)\right)},$$

avec $\varepsilon > 0$ un petit paramètre. Les oscillations de A_ε sont présentées dans la Figure 2.6 pour $\varepsilon = 1 \times 10^{-1}$.

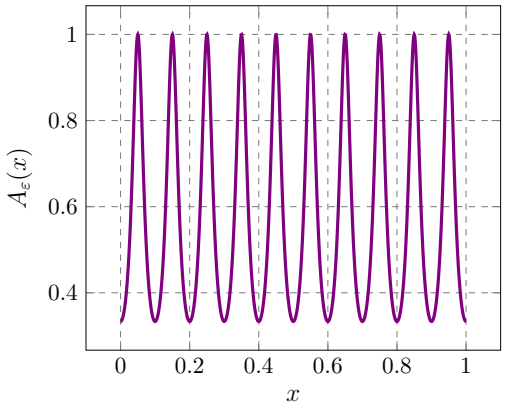


Figure 2.6: Coefficient de diffusion A_ε , $\varepsilon = 1 \times 10^{-1}$.

Pour le second membre particulier $f = 1$, la solution exacte du problème (2.5) est donnée par

$$u_\varepsilon(x) = x - x^2 + \frac{\varepsilon}{2\pi}\left(\frac{1}{2} - x\right)\sin\left(2\pi\frac{x}{\varepsilon}\right) + \frac{\varepsilon^2}{4\pi^2}\left(1 - \cos\left(2\pi\frac{x}{\varepsilon}\right)\right).$$

La formulation faible du problème (2.5) est de trouver $u_\varepsilon \in H_0^1(I)$ telle que, pour tout $v \in H_0^1(I)$,

$$a(u_\varepsilon, v) = \int_I f v \mathrm{d}x,$$

avec

$$a(u_\varepsilon, v) = \int_I A_\varepsilon(x)\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}x}\frac{\mathrm{d}v}{\mathrm{d}x}\mathrm{d}x.$$

Soit $N \in \mathbb{N}^*$ un entier donné. Soit le segment I divisé en $N+1$ sous-intervalles $I_i = [x_i, x_{i+1}]$ de taille $H = \frac{1}{N+1}$ avec $i \in [\![0, N]\!]$. Supposons que la solution exacte de (1.4) soit approchée par la solution du problème discret suivant. Trouver $u_H \in V_H$ tel que pour tout $v_H \in V_H$ nous avons

$$a(u_H, v_H) = \int_I f v_H \mathrm{d}x, \tag{2.6}$$

où V_H, l'espace d'approximation, est un sous-ensemble de $H_0^1(I)$ à définir.

Approximation avec des fonctions affines par morceaux Nous proposons tout d'abord de résoudre (2.6) avec des fonctions linéaires par morceaux. L'espace d'approximation est l'espace de Lagrange $\mathbb{P}_1$, défini comme suit

$$V_H = \left\{\varphi \in \mathcal{C}^0(I) \text{ telle que } \varphi = 0 \text{ sur } \partial I,\ \varphi|_{I_i} \in \mathbb{P}_1(I_i),\ \forall i \in [\![0, N]\!]\right\},$$

où $\mathbb{P}_1(I_i)$ désigne l'espace des polynômes de degré total au plus égal à 1 dans I_i. Une base de cet espace en une dimension est présentée dans la Figure 2.7.

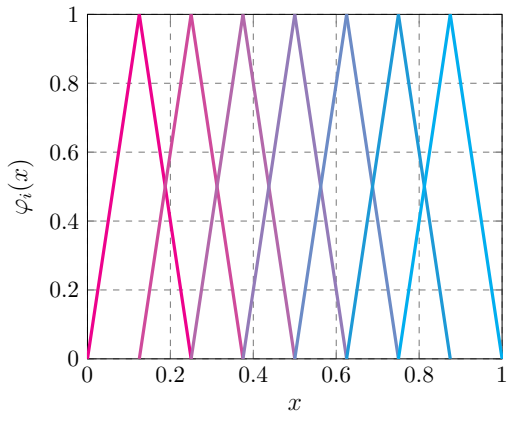


Figure 2.7: Fonctions de base $\mathbb{P}_1$, φ_i, $H = 1/8$.

L'approximation du problème oscillant (2.5) avec des fonctions affines par morceaux est donnée dans la Figure 2.8. On peut facilement constater que l'approximation $\mathbb{P}_1$ de ce problème ne parvient pas à capturer la solution oscillante, à moins d'utiliser un maillage très fin. Pour la résolution de ce problème, les intégrales ont été calculées en utilisant la formule de quadrature

par défaut de FreeFEM pour l'intégration unidimensionnelle, c'est-à-dire `qforder` $= 6$, qui est exacte pour des polynômes $\mathbb{P}_5$. Il convient de noter que les phénomènes observés dans la Figure 2.8 peuvent avoir deux causes : une intégration insuffisamment précise de $A_\varepsilon(x)$ lorsque H est grand et un aspect intrinsèque de l'approximation $\mathbb{P}_1$ qui ne parvient pas à capturer les oscillations [13].

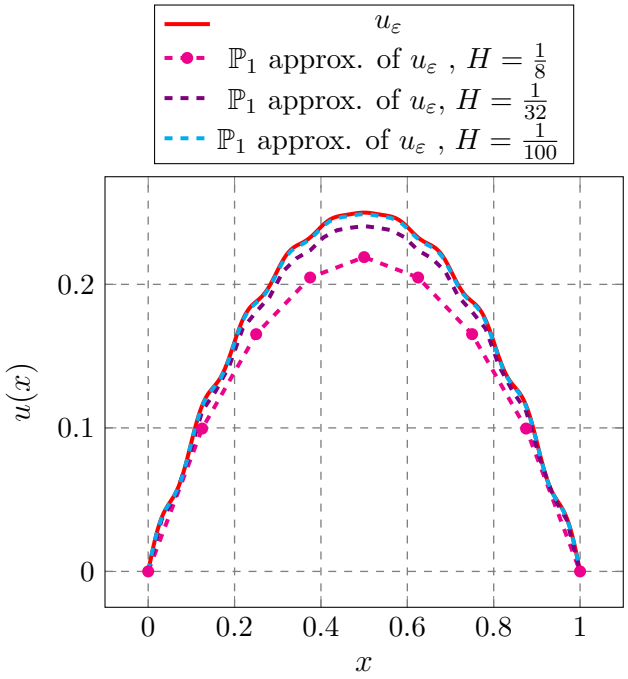


Figure 2.8: Approximation éléments finis du problème de diffusion oscillant.

Approximation avec une approche MsFEM Maintenant, au lieu de considérer un espace d'approximation composé de fonctions affines par morceaux, nous considérons un espace d'approximation composé de fonctions bien adaptées au problème oscillant. Nous définissons le nouvel espace d'approximation V_H comme suit

$$V_H = \left\{ \varphi \in \mathcal{C}^0(I),\ -\frac{\mathrm{d}}{\mathrm{d}x}\left(A_\varepsilon(x)\frac{\mathrm{d}}{\mathrm{d}x}\varphi(x)\right)|_{I_i} = 0,\ \forall i \in [\![0, N]\!] \right\},$$

La base de cet espace dans un domaine unidimensionnel est représentée dans la Figure 2.9. On peut voir que les fonctions de base encodent en elles-mêmes la nature oscillante du problème.

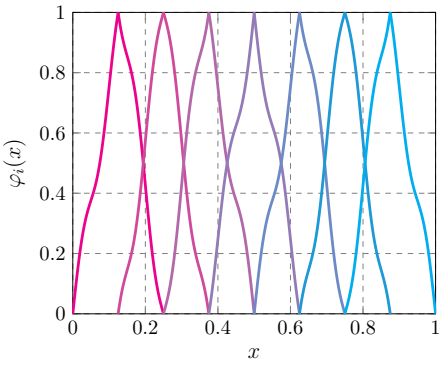


Figure 2.9: Fonctions de bases MsFEM, φ_i, $H = 1/8$.

L'approximation du problème oscillant (2.5) avec la méthode des éléments finis multi-échelles est donnée dans la Figure 2.10. On peut voir que la MsFEM permet maintenant d'approcher la solution du problème (2.5) beaucoup mieux que la méthode des éléments finis $\mathbb{P}_1$. En outre, le tracé du gradient de l'approximation multi-échelle dans la Figure 2.11 (calculé élément par élément puisque les gradients ne sont pas continus à l'interface des éléments grossiers) permet de voir que l'approche multi-échelle permet de capturer les oscillations de la solution contrairement à l'approximation $\mathbb{P}_1$ qui ne voit pas les oscillations.

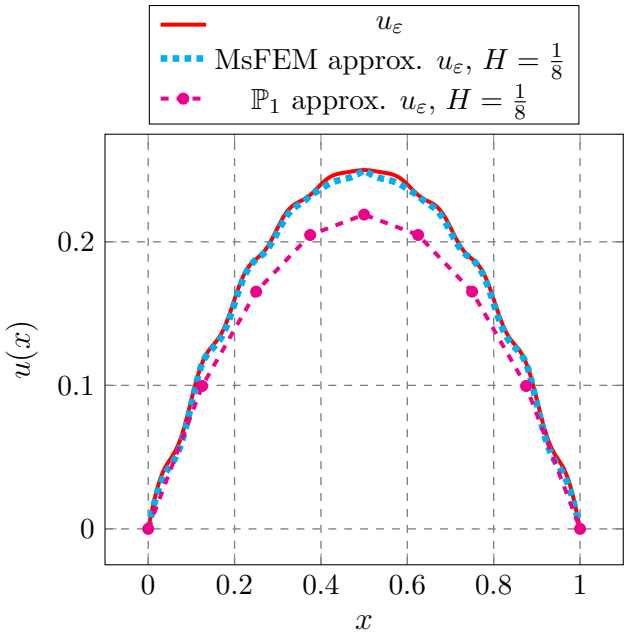


Figure 2.10: Approximation MsFEM du problème de diffusion oscillant.

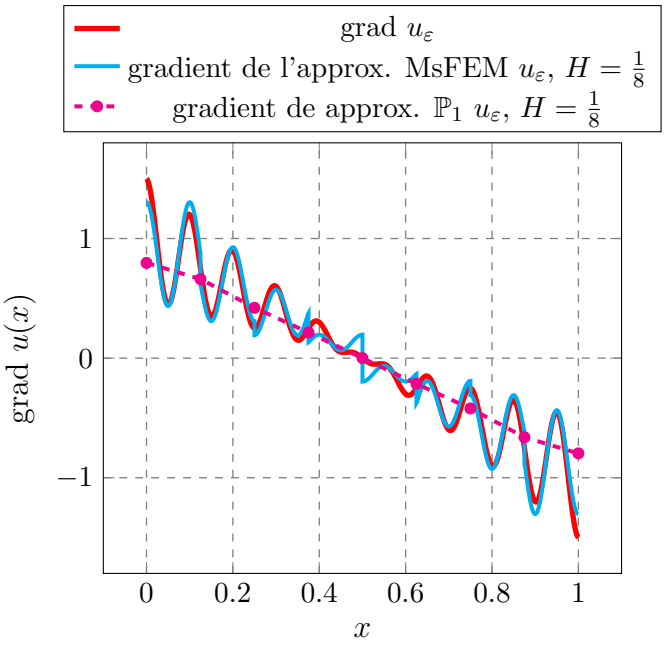


Figure 2.11: Gradient de l'approximation MsFEM du problème de diffusion oscillant.

La section suivante présente un aperçu du développement des méthodes d'éléments finis multi-échelles (MsFEM) dans la littérature.

2.5 Une vue d'ensemble de la méthode des éléments finis multi-échelles (MsFEM) dans la littérature

Étant donné que l'évolution de la MsFEM a été largement motivée par l'amélioration de l'estimation de l'erreur entre la solution exacte (dénotée par l'indice ε) et la solution approchée (dénotée par l'indice H), nous incluons en termes vagues les estimations d'erreur correspondantes dans cet aperçu. Dans ces estimations, H est la taille du maillage grossier, composé d'élément grossier T, et C une constante générique qui ne dépend pas de H. Il convient de noter que les estimations d'erreur suivantes ne sont valables que pour des coefficients périodiques et sont obtenues en supposant le calcul numérique des fonctions de base locales exact. Pour les approches non-conformes dans des milieux perforés, nous définissons l'espace de Sobolev brisé par

$$H^1(\mathcal{T}_H) = \left\{ \boldsymbol{u} \in L^2(\Omega^\varepsilon)^d \text{ telle que } \boldsymbol{u}|_T \in H^1(T \cap \Omega^\varepsilon)^d,\ \forall T \in \mathcal{T}_H \right\},$$

et la semi-norme H^1-brisée $|\boldsymbol{u}|_{H,1}$ dans l'espace H^1 brisé par

$$|\boldsymbol{u}|_{H,1} = \left(\sum_{T \in \mathcal{T}_H} |\boldsymbol{u}|^2_{H^1(T \cap \Omega^\varepsilon)} \right)^{\frac{1}{2}}.$$

2.5.1 MsFEM pour des problèmes elliptiques

Dans le cadre des travaux préliminaires effectués par [24, 133], la méthode originale des éléments finis multi-échelles a été introduite pour la première fois dans [104] pour les problèmes elliptiques

dans les matériaux composites ainsi que dans les milieux poreux et dans [103] pour les problèmes elliptiques avec des coefficients à oscillation rapide. Dans ces premiers travaux, le problème elliptique suivant a été considéré

$$-\operatorname{div}\left(A_\varepsilon \boldsymbol{\nabla} u_\varepsilon\right) = f \text{ dans } \Omega, \qquad u_\varepsilon = 0 \text{ sur } \partial\Omega, \tag{2.7}$$

avec A_ε un coefficient de diffusion oscillant. Les fonctions de base locales sont calculées à l'aide d'un problème elliptique similaire dans chaque élément grossier. Pour ces calculs, différentes conditions aux limites ont été choisies. En considérant d'abord les conditions aux limites linéaires sur chaque élément grossier, cette méthode conduit à l'estimation d'erreur suivante

$$|u_\varepsilon - u_H|_1 \leq C\left(\sqrt{\varepsilon} + H + \sqrt{\frac{\varepsilon}{H}}\right), \tag{2.8}$$

où u_H est la solution obtenue à l'aide de la MsFEM. Des conditions aux limites oscillantes (conditions aux limites données par la solution de l'équation différentielle ordinaire oscillante obtenue en prenant la trace de l'équation originale sur les bords des éléments grossiers) ont été aussi prises en compte. Il a été montré que cette approche mène à la même estimation de l'erreur (2.8). Néanmoins, à l'aide d'exemples numériques, il a été montré que l'utilisation de conditions aux limites oscillantes conduit à une amélioration significative de la précision des résultats numériques. Cependant, la prise en compte de l'un de ces deux types de conditions aux limites conduit à une résonance entre l'échelle du maillage grossier et la petite échelle dans la solution physique. Ceci est mis en évidence par la présence du terme $\sqrt{\frac{\varepsilon}{H}}$, appelé "erreur de résonance", dans l'estimation de l'erreur.

Ensuite, pour atténuer la résonance due aux couches limites et au choix des conditions limites, [104, 71] ont proposé la méthode dite de l'*oversampling*. Cette méthode consiste à résoudre les problèmes locaux sur un domaine plus grand que l'élément grossier et à tronquer la solution à l'intérieur de l'élément grossier. Cette méthode, qui rend la solution moins dépendante des conditions aux limites choisies, conduit à des problèmes non-conformes puisque les quantités ne sont pas continues sur les bords des éléments grossiers. L'utilisation de l'oversampling conduit alors à l'estimation d'erreur suivante

$$|u_\varepsilon - u_H|_{H,1} \leq C\left(\sqrt{\varepsilon} + H + \frac{\varepsilon}{H}\right). \tag{2.9}$$

A partir de (2.9), il est facile de voir que l'erreur de résonance a été réduit mais existe toujours. Par conséquent, pour éliminer la résonance, dans [105], l'approche a été modifiée, en introduisant une formulation de Petrov–Galerkin avec des fonctions d'essai non-conformes multi-échelles et des fonctions test linéaires. En utilisant cette nouvelle approche, l'estimation d'erreur suivante a été obtenue

$$|u_\varepsilon - u_H|_{H,1} \leq C\left(\sqrt{\varepsilon} + H + \varepsilon\right),$$

qui ne fait, cette fois ci, pas intervenir l'erreur de résonance. Une étude complète de ces premières méthodes peut être trouvée dans [70].

Dans [12], la MsFEM a été employée pour effectuer l'homogénéisation numérique des problèmes

de diffusion. Le même problème elliptique (2.7) a été considérée. L'une des nouveautés de ce travail est qu'une règle de composition (changement de variables) a été introduite pour la construction de bases d'éléments finis qui permet un traitement simple des méthodes d'éléments finis d'ordre élevé. Une taille de maillage fin h optimal qui minimise l'erreur a aussi été exhibée. Inspiré par [12], [101] développa une méthode d'éléments finis multi-échelle d'ordre élevé pour les problèmes elliptiques avec des coefficients fortement oscillants, mais en construisant un espace d'éléments finis multi-échelle plus explicite.

L'une des dernières avancées dans le domaine des méthodes d'éléments finis multi-échelles a été d'envisager une méthode dans la veine de la méthode classique des éléments finis de Crouzeix–Raviart, c'est-à-dire une méthode dans laquelle la continuité à travers les faces des éléments grossiers est uniquement considérée dans un sens faible : seule la valeur moyenne est continue à travers les faces. Puisque des conditions aux limites sont requises sur chaque élément grossier, la MsFEM à la Crouzeix-Raviart permet des conditions aux limites naturelles sur les faces de l'élément grossier. Cette méthode a été introduite pour la première fois dans [121] pour les problèmes elliptiques fortement oscillants (2.7). Pour résoudre ce problème, l'espace d'approximation est construite à partir d'idées similaires à celles de Crouzeix et Raviart dans leur construction d'un espace d'éléments finis classique [61]. L'estimation d'erreur suivante a été obtenue

$$|u_\varepsilon - u_H|_{H,1} \leq CH\|f\|_{L^2(\Omega)} + C\left(\sqrt{\varepsilon} + H + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{\nabla} u^*\|_{C^1(\overline{\Omega})},$$

où u^* est la solution du problème homogénéisé et avec la constante C indépendante de H, ε, f et u^*. Bien que l'estimation de l'erreur ne soit pas améliorée par rapport à la MsFEM avec des conditions aux limites linéaires sur chaque élément grossier, les expériences numériques montrent que cette approche est plus performante que l'approche classique.

Ce travail a été poursuivi dans [122], dans lequel l'approche a été adaptée pour traiter des problèmes multi-échelles dans un domaine perforé Ω^ε, en considérant le problème suivant. Trouver $u_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$, la solution de

$$-\Delta u_\varepsilon = f \text{ dans } \Omega^\varepsilon, \quad u_\varepsilon = 0 \text{ sur } \partial\Omega^\varepsilon.$$

La principale nouveauté de leur approche est l'enrichissement de l'espace des éléments finis multi-échelles à l'aide de fonctions bulle. Dans le cas de perforations périodiques en deux dimensions, avec une régularité suffisante du second membre f, l'approximation u_H satisfait

$$|u_\varepsilon - u_H|_{H,1} \leq C\varepsilon\left(\sqrt{\varepsilon} + H + \sqrt{\frac{\varepsilon}{H}}\right)\|f\|_{H^2(\Omega)},$$

avec une constante C indépendante de H, ε et f, mais dépendant de la géométrie du maillage et d'autres paramètres du problème.

Ensuite, dans [63], les auteurs l'application de la MsFEM à la Crouzeix-Raviart a été étendue aux problèmes d'advection-diffusion avec des conditions aux limites non homogènes. Le problème

suivant a été considéré. Trouver $u_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$, la solution de

$$\begin{cases} -\operatorname{div}(A_\varepsilon \boldsymbol{\nabla} u_\varepsilon) + \boldsymbol{w} \cdot \boldsymbol{\nabla} u_\varepsilon &= f \quad \text{dans } \Omega^\varepsilon, \\ u_\varepsilon &= 0 \quad \text{sur } \partial B^\varepsilon \cap \Omega^\varepsilon, \\ u_\varepsilon &= g \quad \text{sur } \partial\Omega \cap \Omega^\varepsilon, \end{cases}$$

avec $\boldsymbol{w}$ le terme d'advection. Au travers de plusieurs exemples numériques, les auteurs ont montré que cette méthode particulière est plus performante que les approches MsFEM plus conventionnelles. L'étude des MsFEMs pour les problèmes d'advection-diffusion a fait l'objet de nombreuses thèses [141, 128, 33].

Récemment, [123] a développé une autre approche MsFEM pour les problèmes de diffusion avec un coefficient de diffusion oscillant, dans laquelle les fonctions de base associées aux faces et aux éléments sont enrichies à l'aide de polynômes de Legendre. Cette approche permet de réduire l'erreur de résonance dans le régime où la taille du maillage grossier H est de l'ordre de la petite échelle ε des oscillations. De plus, des MsFEMs pour les problèmes de contrôle optimal elliptiques avec des coefficients rugueux ont été étudiés dans [38, 39].

En outre, d'autres travaux [99, 46] traitent des MsFEMs adaptatives en calculant une estimation de l'erreur a posteriori qui dépend de la taille du maillage grossier H et de la taille du maillage fin h.

2.5.2 Extension de la MsFEM pour les écoulements incompressibles

La première utilisation de la MsFEM pour les écoulements incompressibles a été faite dans [136], dans lequel a été proposé une MsFEM à la Crouzeix–Raviart (c'est-à-dire que seule la valeur moyenne de la vitesse est continue à travers les faces des éléments grossiers) pour résoudre le problème de Stokes (2.3). Cependant, l'estimation d'erreur obtenue dans ce travail n'était pas entièrement satisfaisante car dépendant du caractère oscillant de la solution exacte sur les faces.

Ensuite, dans [111], cette méthode a été généralisée à des ensembles arbitraires de fonctions de poids utilisées pour imposer la continuité entre les faces des éléments grossiers. Pour un ensemble particulier de fonctions de poids dans un cadre périodique en deux dimensions (la MsFEM dite CR3, dans laquelle les trois fonctions de poids sur les faces sont $\boldsymbol{e_1}$, $\boldsymbol{e_2}$, la base canonique de $\mathbb{R}^2$ et la composante normale d'un polynôme linéaire), la borne d'erreur suivante a été établie

$$\begin{aligned} &|\boldsymbol{u_\varepsilon} - \boldsymbol{u_H}|_{H,1} + \varepsilon \|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \\ &\leq C\varepsilon \left[\left(H + \sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}} \right) \left(\|\boldsymbol{f}\|_{H^2(\Omega)\cap C^1(\overline{\Omega})} + \|p^*\|_{H^2(\Omega)} \right) \right], \end{aligned}$$

où p^* est la pression homogénéisée et C est une constante dépendant seulement de la régularité du maillage et du motif des perforations.

Ensuite, [135] a étendu cette approche pour résoudre le problème d'Oseen (2.2) dans des milieux hétérogènes. Cependant, les applications étaient limitées aux petits nombres de Reynolds. Récemment [79] a développé une MsFEM non conforme enrichie pour résoudre le problème de

Stokes dans des milieux hétérogènes, qui généralise la méthode proposée dans [111], en considérant également des fonctions de poids d'ordre élevé sur les faces et dans les éléments, plutôt que de simplement considérer des fonctions de poids d'ordre élevé sur les faces.

2.5.3 Conclusion

Dans cette section, plusieurs MsFEM ont été décrits. Il est pertinent de souligner que dans tous les travaux présentés dans cette revue, les applications numériques ont été considérées uniquement dans des domaines bidimensionnels et les différentes estimations d'erreur ont été obtenues dans un cadre périodique et en supposant que le calcul des fonctions de base sont exactes. De plus, l'étude des différentes estimations d'erreurs présentées dans cette section n'est pas simple. En effet, ces erreurs convergent vers zéro lorsque H et ε tendent vers zéro ensemble (la convergence de H vers zéro dépendant toutefois de celle de ε). Cependant, pour un ε fixé, l'erreur entre la solution exacte et l'approximation multi-échelle ne devrait pas converger vers zéro à mesure que la taille du maillage grossier H diminue (dans le régime néanmoins dans lequel $H > \varepsilon$), mais sera bornée par une quantité dépendante de ε. Une valeur optimale de H peut être trouvée afin que l'erreur soit minime.

Ce travail de thèse s'inscrit dans la continuité de l'étude menée dans ce dernier travail [79]. En particulier, cette thèse se concentre sur le niveau discret en montrant le caractère bien posé des problèmes locaux discrets et en dérivant une estimation d'erreur pour les approximations multi-échelles. Nous proposons également une extension de cette méthode pour résoudre le problème d'Oseen. Les principales contributions de la thèse sont répertoriées dans la Section 2.6. Les méthodes d'éléments finis multi-échelles (MsFEM) décrites dans cette thèse sont destinées à être incorporées dans le logiciel TrioCFD. Cependant, par souci de commodité, les applications numériques préliminaires de cette méthode sont effectuées dans le logiciel FreeFEM [93], un logiciel plus académique, choisi pour sa flexibilité à mettre en œuvre de nouveaux éléments finis et sa facilité d'utilisation pour la mise en œuvre et le test de différentes méthodes. FreeFEM est un logiciel, open-source, parallèle d'éléments finis. Il permet de résoudre des problèmes impliquant des équations aux dérivées partielles issues de différentes branches de la physique, telles que la mécanique des fluides, les interactions fluide-structure et bien d'autres. Dans FreeFEM, le problème à résoudre est défini en termes de formulation variationnelle. FreeFEM est écrit en C++ et son langage est un idiome du C++. Pour les applications plus complexes, nous profitons également du fait que FreeFEM s'interface avec différentes bibliothèques telles que le partitionneur METIS [118] ou la bibliothèque PETSc [28].

2.6 Contributions majeures de la thèse

Les principales contributions de cette thèse peuvent être divisées en deux parties. En effet, les premières contributions (**Contributions 1 à 3**) concernent le développement d'outils préliminaires, utiles pour l'analyse de la méthode des éléments finis multi-échelles (MsFEM). Cependant, il est important de noter que ces outils préliminaires ne sont pas limités à l'étude de la MsFEM mais peuvent être réutilisés dans d'autres contextes. La **Contribution 1** con-

cerne la quantification de la convergence de l'homogénéisation du problème de Stokes dans un domaine perforé. Le théorème principal de cette contribution est l'un des principaux ingrédients de l'élaboration d'une borne d'erreur pour la MsFEM appliquée au problème de Stokes. La **Contribution 2** concerne l'introduction d'une nouvelle famille d'éléments finis non-conformes en trois dimensions sur des tétraèdres dans la veine des éléments finis non-conformes introduits dans [132] en deux dimensions sur des triangles. Ces familles d'éléments finis non-conformes, en deux et trois dimensions, permettent de prouver le caractère bien posé discrète des problèmes locaux issus de la MsFEM appliquée au problème de Stokes ou d'Oseen. En effet, pour ces problèmes locaux impliquant des multiplicateurs de Lagrange, le caractère bien posé ne peut pas être démontré avec des paires classiques d'éléments finis. En conséquence, à des fins numériques, la **Contribution 3** concerne l'implémentation des nouveaux éléments finis non-conformes dans le code source de FreeFEM [93]. Ensuite, les autres contributions (**Contributions 4 à 7**) concernent plus spécifiquement l'analyse de la MsFEM d'un point de vue théorique et numérique.

Contribution 1 La première contribution concerne la dérivation d'un taux de convergence quantitatif pour l'homogénéisation du problème de Stokes (2.3) dans un domaine perforé périodique. Cette contribution est présentée dans le Chapitre 6 et a donné lieu à une publication [30]. Il est bien connu que les équations homogénéisées ou effectives pour le système de Stokes, dans un domaine perforé périodique, est la loi de Darcy pour la vitesse $\boldsymbol{u}$ et la pression homogénéisée p^*,

$$\left\{\begin{array}{rcll} \boldsymbol{u} & = & \frac{1}{\nu}A^*(\boldsymbol{f}-\boldsymbol{\nabla} p^*) & \text{dans } \Omega, \\ \operatorname{div}\boldsymbol{u} & = & 0 & \text{dans } \Omega, \\ \boldsymbol{u}\cdot\boldsymbol{n} & = & 0 & \text{sur } \partial\Omega. \end{array}\right.$$

où A^* est un tenseur de perméabilité constant (voir la Proposition 6.5). Nous établissons un taux de convergence $O(\sqrt{\varepsilon})$ pour la norme d'énergie de la différence des vitesses et des pressions, où ε représente la taille des obstacles solides. Ceci est réalisé en utilisant une expansion asymptotique à deux échelles des équations de Stokes et une nouvelle construction d'une fonction de coupure qui évite l'introduction de couches limites. Cette estimation de l'erreur $\sqrt{\varepsilon}$ a d'abord été démontrée par [111] dans un espace bidimensionnel, et a été indépendamment améliorée dans [156] pour n'importe quelle dimension d'espace $d \geq 2$, dans les deux cas sous l'hypothèse d'obstacles solides isolés (comme dans la Figure 6.1). Cependant, cette hypothèse n'est pas physiquement réaliste en dimension $d \geq 3$. Par conséquent, nous étendons cette estimation d'erreur $\sqrt{\varepsilon}$ au cas d'obstacles solides connectés dans tout espace de dimension $d \geq 2$.

Avant d'énoncer notre résultat principal (en termes vagues, voir le Théorème 6.9 pour un énoncé précis), nous devons introduire quelques notations. En fait, la vitesse de Darcy $\boldsymbol{u}$ ne voit pas les obstacles solides et ne peut donc pas être une bonne approximation de la vitesse originale $\boldsymbol{u_\varepsilon}$. Elle doit être corrigée en introduisant une meilleure approximation qui est

$$\boldsymbol{u_2}\left(x,\frac{x}{\varepsilon}\right) = \frac{1}{\nu}\sum_{i=1}^{d}\boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\left(f_i(x)-\frac{\partial p^*}{\partial x_i}(x)\right),$$

où les vitesses locales $\boldsymbol{\omega_i}$ sont des solutions de problèmes de cellule (6.6). Le lien entre $\boldsymbol{u_2}$ et la

vitesse de Darcy $\boldsymbol{u}$ est que $\boldsymbol{u}(x)$ est la moyenne de $\boldsymbol{u_2}(x, y)$ par rapport à la variable périodique y. La principale nouveauté est que notre analyse s'applique au cas physiquement pertinent d'un milieu poreux où les parties fluides et solides sont connectées. En termes simples, cette contribution se lit comme suit.

Theorem 2.2. Soit $\boldsymbol{u_\varepsilon}$, p_ε la solution du problème de Stokes (2.3) and $\boldsymbol{u_2}$, p^*, leur approximations homogénéisées. En supposant que $\boldsymbol{f}$ est suffisamment lisse, il existe une constante C, indépendante de ε, telle que

$$\frac{\|p_\varepsilon - p^*\|_{L^2(\Omega^\varepsilon)}}{\|p^*\|_{L^2(\Omega^\varepsilon)}} \leq C\varepsilon^{\frac{1}{2}},$$

$$\frac{|\boldsymbol{u_\varepsilon} - \varepsilon^2\boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)}}{|\varepsilon^2\boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)}} \leq C\varepsilon^{\frac{1}{2}},$$

$$\frac{\|\boldsymbol{u_\varepsilon} - \varepsilon^2\boldsymbol{u_2}\|_{L^2(\Omega^\varepsilon)}}{\|\varepsilon^2\boldsymbol{u_2}\|_{L^2(\Omega^\varepsilon)}} \leq C\varepsilon^{\frac{1}{2}}.$$

Le Théorème 2.2 est énoncé en termes d'erreurs relatives puisque $\varepsilon^2\boldsymbol{u_2}$ (et donc $\boldsymbol{u_\varepsilon}$) est petit. Ce théorème est utilisé dans la dérivation d'une estimation d'erreur pour l'approximation MsFEM du problème de Stokes dans un domaine perforé périodique (voir la Section 9.2).

Contribution 2 La deuxième contribution est le développement d'une famille d'éléments finis scalaires non-conformes d'ordre de précision deux et trois par rapport à la norme H^1 sur des tétraèdres. Cette contribution est présentée dans le Chapitre 5 et a conduit à une publication [29]. En effet, pour la classe de MsFEM développée dans cette thèse, les problèmes locaux impliquent une divergence polynomiale et des multiplicateurs de Lagrange, et ne peuvent donc pas être résolus en utilisant des paires classiques d'éléments finis. En deux dimensions, les éléments finis non-conformes introduits par [132] (et présentés dans le Chapitre 4) associés à des pressions polynomiales discontinues d'ordre n permettent de résoudre ces problèmes locaux. Cependant, en trois dimensions, aucun couple d'éléments finis existant n'a conduit à un problème discret dont le caractère bien posé a pu être prouvé, à l'exception de ceux développés ici. Par conséquent, dans la veine de [132], nous introduisons une famille d'éléments finis scalaires non-conformes d'ordre deux et trois par rapport à la norme H^1 sur des tétraèdres. Leur version vectorielle génère, avec une approximation de pression discontinue d'ordre un et deux respectivement, une paire d'éléments finis vérifiant la condition inf-sup d'ordre de convergence deux et trois pour le problème de Stokes dans la norme d'énergie. Les degrés de liberté scalaires (locaux) sont définis par des moments sur les faces F_α et dans les tétraèdres K,

$$N_j^{F_\alpha}(v) = \int_{F_\alpha} vL_j^{F_\alpha} \qquad j \geq 0,\ \alpha = 1, 2, 3, 4,$$

$$N_j^K(v) = \int_K vM_j^K \qquad j \geq 0,$$

où $(L_j^{F_\alpha})_j$ définit une base arbitraire de $\mathbb{P}_n(F_\alpha)$ de dimension $(n+1)(n+2)/2$, et $(M_j^K)_j$ définit une base arbitraire de $\mathbb{P}_{n-1}(K)$ de dimension $n(n+1)(n+2)/6$. Pour chaque élément K du maillage $\mathcal{T}_h$ et tout entier $n \geq 1$, nous définissons l'ensemble des degrés de liberté $\mathcal{N}_{n+1}(K)$ par

$$\begin{aligned}\mathcal{N}_{n+1}(K) := &\left\{ N_j^{F_\alpha} \qquad 1 \leq j \leq \frac{(n+1)(n+2)}{2},\ \alpha = 1,2,3,4 \right\} \\ &\cup \left\{ N_j^K \qquad 1 \leq j \leq \frac{n(n+1)(n+2)}{6} \right\}.\end{aligned} \tag{2.10}$$

Suivant [132], nous enrichissons l'espace local $\mathbb{P}_{n+1}(K)$ avec un sous-espace propre de $\mathbb{P}_{n+2}(K)$, dénoté par $\Sigma_{n+2}(K) \subset \mathbb{P}_{n+2}(K)$, avec une intersection triviale avec $\mathbb{P}_{n+1}(K)$. Ainsi, l'espace enrichi $V_{n+1}(K)$ est donné par

$$V_{n+1}(K) = \mathbb{P}_{n+1}(K) \oplus \Sigma_{n+2}(K).$$

De tels espaces V_{n+1} n'ont été trouvés que pour les cas $n = 1$ et $n = 2$. Le lemme principal de cette contribution est donc le suivant.

Lemma 2.3. Soit K un tétraèdre avec des coordonnées barycentriques λ_1, λ_2, λ_3, λ_4. Alors, les espaces éléments finis

$$V_2 = \mathbb{P}_2 + \text{span}\{\lambda_1\lambda_2^2,\ \lambda_1\lambda_3^2,\ \lambda_2\lambda_3^2\},$$

et

$$V_3 = \mathbb{P}_3 + \text{span}\{\lambda_1^3\lambda_2,\ \lambda_2^3\lambda_3,\ \lambda_3^3\lambda_4,\ \lambda_4^3\lambda_1,\ \lambda_2^3\lambda_1,\ \lambda_1^3\lambda_4,\ \lambda_4^3\lambda_3,\ \lambda_3^3\lambda_2\},$$

sont unisolvants, respectivement, par rapport à l'ensemble des degrés de liberté $\mathcal{N}_2$ and $\mathcal{N}_3$ décrits dans (2.10).

Il faut noter que les espaces d'éléments finis V_2 et V_3 décrits dans le Lemme 2.3 ne sont pas uniques (voir par exemple la liste de toutes les bases convenables V_3 dans l'Annexe A.2). De plus, pour le cas $n = 1$, nous retrouvons l'élément fini proposé par [50].

Contribution 3 Comme nous l'avons expliqué plus haut, la caractère bien posé des problèmes locaux impliqués dans la MsFEM est démontrée à l'aide d'éléments finis non classiques. En effet, en deux dimensions, nous avons utilisé la famille d'éléments finis introduite dans [132] et présentée dans le Chapitre 4. En trois dimensions, nous avons utilisé la famille d'éléments finis introduite dans le Chapitre 5 et dans [29]. Dans ces deux cas, ces éléments finis sont associés à une approximation polynomiale discontinue d'ordre n. Par conséquent, à des fins numériques, c'est-à-dire pour le calcul des fonctions de base locales, ces nouveaux éléments finis ont été implémentés dans le code source de FreeFEM [93]. Tout d'abord, nous avons implémenté l'élément fini non-conforme d'ordre de précision trois, introduit dans [132], dans le code source de FreeFEM, et il est connu sous le nom *P3pnc*[1]. L'élément fini non-conforme d'ordre de précision

[1] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P3pnc.cpp

deux, connu sous le nom $P2pnc$[2], était déjà implémenté. Pour les applications numériques en trois dimensions, nous avons implémenté les deux éléments finis non-conformes, d'ordre deux et trois, présentés dans le Chapitre 5 et dans [29]. Ils sont connus respectivement sous le nom de $P2pnc3d$[3] et $P3pnc3d$[4]. **Il est pertinent de souligner qu'en plus d'être utiles pour cette thèse, ces éléments finis non-conformes peuvent maintenant être utilisés par tous les utilisateurs de FreeFEM pour résoudre des problèmes physiques.**

Contribution 4 Nous proposons la première estimation d'erreur pour l'approximation du problème de Stokes dans un domaine perforé périodique en utilisant la MsFEM développée dans cette thèse, ce qui permet de montrer la convergence de la méthode proposée. La démonstration de l'estimation d'erreur est présentée dans la Section 9.2 et est basée sur le lemme de Strang, qui est le lemme classique pour l'estimation d'erreur pour les méthodes d'éléments finis non-conformes, et la convergence quantitative de l'homogénéisation du problème de Stokes introduite dans le Chapitre 6 et dans [30]. Cette estimation de l'erreur est obtenue en supposant que le calcul des fonctions de base locales est exact. La méthodologie utilisée s'inspire de la démonstration de l'erreur effectuée dans [111]. Le résultat principal se lit en termes vagues (voir le Théorème 9.27 pour un énoncé précis) comme suit.

Theorem 2.4. Soit $\boldsymbol{f}$ suffisamment lisse. L'erreur entre la solution $(\boldsymbol{u_\varepsilon}, p_\varepsilon)$ du problème de Stokes (2.3) dans un domaine périodique perforé et à son approximation MsFEM $(\boldsymbol{u_H}, p_H)$ est donnée par

$$\begin{aligned}&\varepsilon^{-1}|\boldsymbol{u_\varepsilon} - \boldsymbol{u_H}|_{H,1} + \|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \\ &\leq C\Big(H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)} + H^n|p^*|_{H^{n+1}(\Omega)} + \left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\Big),\end{aligned}$$

où la constante C ne dépend pas de ε, H et de $\boldsymbol{f}$.

Plus précisément, à H fixé, le terme $p_\varepsilon - p_H$ est $O(1)$ lorsque ε tends vers zéro. De plus, en rappelant que dans un domaine périodique, on a $\boldsymbol{u_\varepsilon} \approx \varepsilon^2 \boldsymbol{u}(x, \frac{x}{\varepsilon})$ (à chaque dérivation de $\boldsymbol{u}(x, \frac{x}{\varepsilon})$, un facteur $1/\varepsilon$ apparaît), les estimations d'erreur ci-dessus pour la vitesse et la pression indiquent essentiellement que les erreurs relatives sont de l'ordre de $\left(H^n + \sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)$. Dans cette somme, le terme $\sqrt{\varepsilon}$, qui provient des couches limites dans le processus d'homogénéisation, est négligeable. En effet, le terme dominant est l'erreur de résonance classique $\sqrt{\frac{\varepsilon}{H}}$, qui indique que la taille de la maille grossière H doit être beaucoup plus grande que ε pour obtenir une bonne approximation. Cependant, la présence du terme H^n permet de choisir un H suffisamment grand sans réduire significativement la précision de l'approximation, ce qui montre l'intérêt de considérer des méthodes d'ordre élevé. La valeur optimale du maillage grossier H qui minimise

[2] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P2pnc.cpp

[3] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P2pnc_3d.cpp

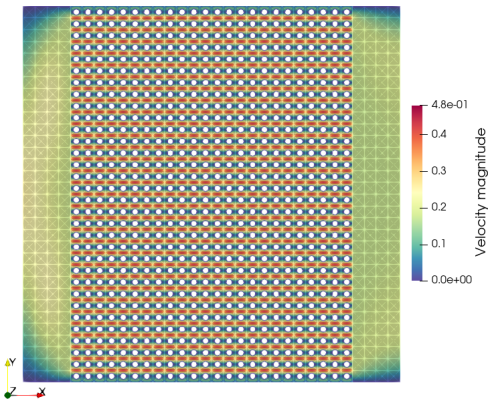
[4] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P3pnc_3d.cpp

l'erreur est donnée par

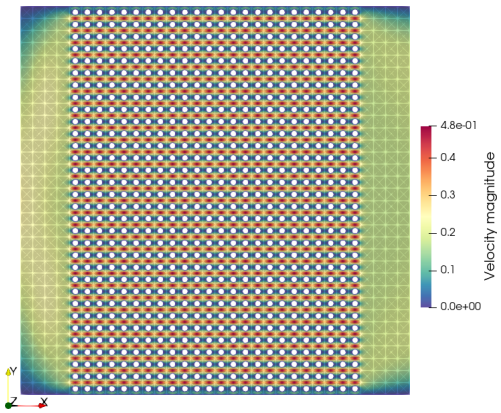
$$H_{opt} = \left(\frac{\sqrt{\varepsilon}}{2n}\right)^{\frac{1}{n+\frac{1}{2}}}.$$

De plus, en utilisant $H = H_{opt}$ dans l'estimation de l'erreur, nous remarquons que l'erreur devient de l'ordre $\sqrt{\varepsilon}$ lorsque $n \to \infty$ (nous retrouvons l'ordre de convergence de l'homogénéisation).

Contribution 5 Nous étendons la MsFEM, initialement développée pour le problème de Stokes (voir par exemple la Figure 2.12), au problème d'Oseen (2.2) dans la Partie IV (voir par exemple la Figure 2.13). La principale différence avec la MsFEM pour le problème de Stokes est que dans ce cas, les problèmes locaux sont des problèmes de type Oseen, avec la même vitesse d'Oseen $\boldsymbol{U_0}$ que dans le problème global. De plus, à travers des expériences numériques, nous montrons que cette MsFEM permet de traiter des nombres de Reynolds élevés, et que la formulation MsFEM Petrov–Galerkin (dans laquelle les fonctions test sont solutions de problèmes locaux dérivés du problème adjoint d'Oseen et les fonctions d'essai sont solutions de problèmes locaux dérivés du problème d'Oseen) n'aboutit pas à des résultats numériques plus précis.

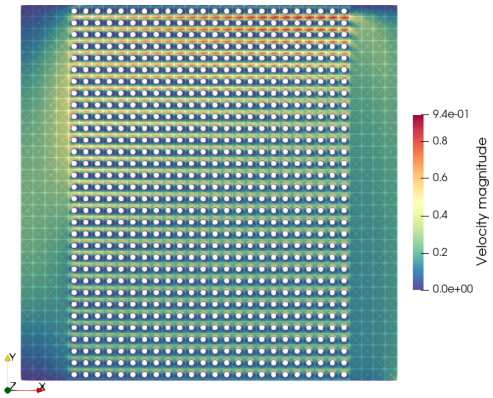


(a) Approximation MsFEM ($n = 1$, 4096 éléments grossiers).

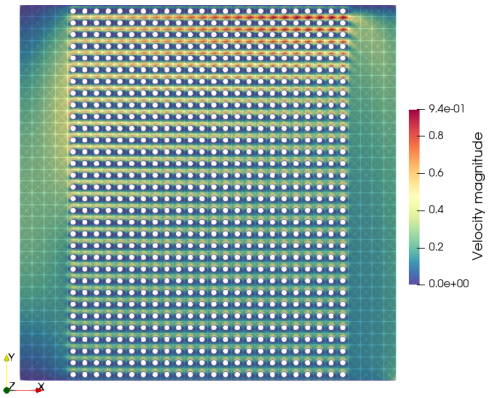


(b) Solution de référence calculée sur un maillage fin.

Figure 2.12: Simulation bi-dimensionnelle d'un écoulement de Stokes dans un canal ouvert, $\nu = 1$, écoulement de la gauche vers la droite avec une entrée parabolique, une sortie libre, et une vitesse nulle sur les bords haut, bas et sur les obstacles.



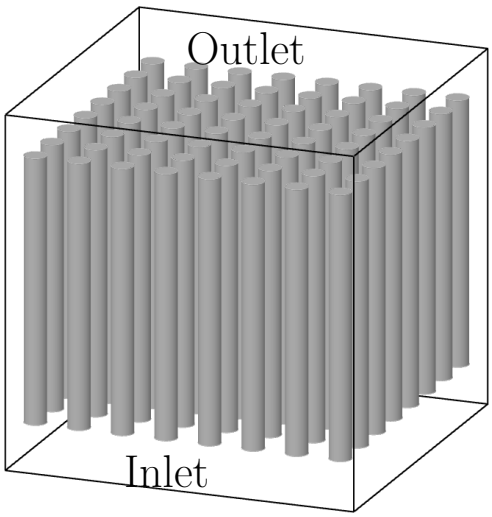
(a) Approximation MsFEM ($n = 2$, 4096 éléments grossiers).

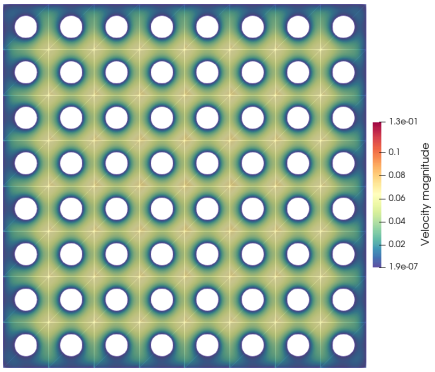


(b) Solution de référence calculée sur un maillage fin.

Figure 2.13: Simulation bi-dimensionnelle d'un problème de Oseen dans un canal ouvert, nombre de Reynolds $Re = 2000$ ($\nu = 5 \times 10^{-4}$, $\boldsymbol{U_0} = [y(1-x^2), -x(1-y^2)]^t$), écoulement de la gauche vers la droite, avec une entrée parabolique, une sortie libre, et une vitesse nulle sur les bords haut, bas et sur les obstacles.

Contribution 6 Nous implémentons la MsFEM pour résoudre les problèmes de Stokes et d'Oseen en deux et trois dimensions (voir par exemple la Figure 2.14) dans un cadre de haute performance dans FreeFEM [93]. En effet, pour résoudre les problèmes locaux, nous développons deux niveaux de parallélisme : un parallélisme entre chaque élément grossier (permettant de considérer chaque élément grossier en parallèle) et un parallélisme à l'intérieur de chaque élément grossier (permettant de calculer les problèmes locaux eux-mêmes en parallèle). De plus, à des fins de comparaison, nous implémentons des solveurs pour calculer des solutions de référence sur des maillages fins. Les détails de ces implémentations sont présentés dans les Annexes E et D. En outre, nous proposons la première implémentation à l'ordre $n = 2$ pour la MsFEM étudiée (voir par exemple la Figure 2.13).

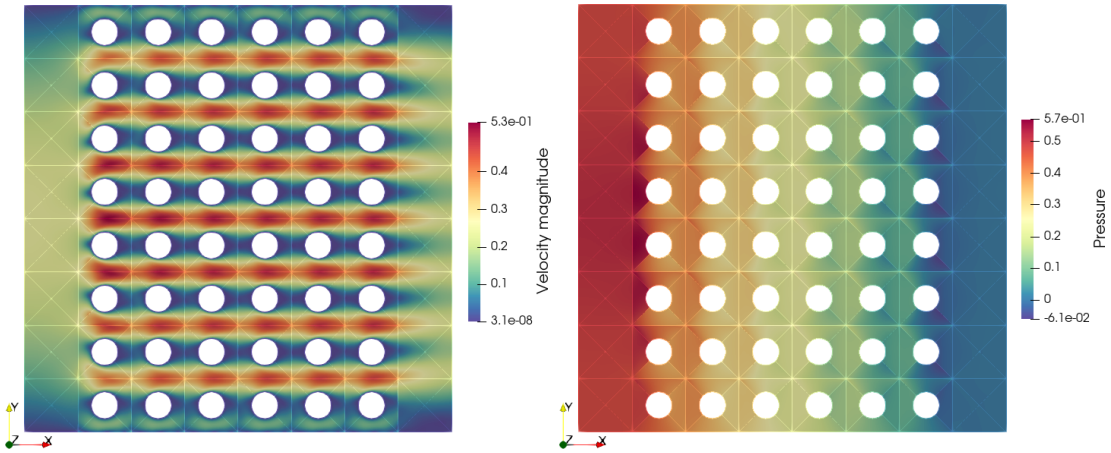




(a) Coupe dans le champ de magnitude de la vitesse suivant l'axe z (perpendiculaire aux cylindres).

Figure 2.14: Approximation MsFEM d'un écoulement de Stokes dans un canal ouvert tridimensionnel avec une entrée parabolique, une sortie libre et une vitesse nulle sur les autres bords ($n = 2$, 3072 éléments grossiers).

Contribution 7 Nous proposons une méthodologie pour résoudre le problème stationnaire de Navier–Stokes (2.1) sur une grille grossière en utilisant des fonctions de base MsFEM adaptatives dans le Chapitre 16 (voir par exemple la Figure 2.15). Comme le problème comporte des termes non linéaires, nous proposons une approche pseudo-transitoire pour le résoudre. Cependant, la principale difficulté pour résoudre itérativement le problème de Navier-Stokes en utilisant une approche multi-échelle est que ce problème ne peut pas être résolu en utilisant une seule base fixée, dérivée soit du problème d'Oseen (avec une vitesse d'Oseen arbitraire donnée $\boldsymbol{U_0}$) soit du problème de Stokes (équivalent au problème d'Oseen avec une vitesse d'advection nulle). En effet, à chaque itération, le problème global à calculer change puisque la vitesse d'advection change. En particulier, dès que la vitesse d'advection du problème global diffère trop de celle utilisée pour les fonctions de base, ces dernières deviennent inadaptées au problème. Par conséquent, pour surmonter cette difficulté, à chaque itération, de nouvelles fonctions de base sont calculées afin que les fonctions de base restent bien adaptées au problème. En outre, nous proposons une discussion sur la possibilité de résoudre un tel problème à l'aide de la méthode des éléments finis généralisée (GMsFEM) [69].



(a) Velocity magnitude. (b) Pressure field.

Figure 2.15: Approximation MsFEM ($n = 2$, 256 éléments grossiers) d'un problème de Navier–Stokes stationnaire sans un canal ouvert en deux dimensions, nombre de Reynolds $Re = 1000$ ($\nu = 1 \times 10^{-3}$), écoulement de la gauche vers la droite, avec une entrée parabolique, une sortie libre, et une vitesse nulle sur les bords haut, bas et sur les obstacles.

Part II

Various preliminary tools: finite element and homogenization

Chapter 3

Some useful reminders

Contents

3.1 The Crouzeix–Raviart finite element

Let Ω be a connected and bounded open set in $\mathbb{R}^d$ with $d = 2$ or 3. We denote by $\mathcal{T}_h$ a discretization of Ω by triangles in two dimensions or tetrahedra in three dimensions. Let K be an element of $\mathcal{T}_h$. Let $\mathcal{F}_h$ be the set of faces in the discretization and let F be an element of $\mathcal{F}_h$.

The Crouzeix–Raviart finite element was introduced first in [61]. The velocity unknown of this finite element is the value of the shape functions at the centers of the faces and the pressure degrees of freedom are located at the barycenter of each element (see Figure 3.1).

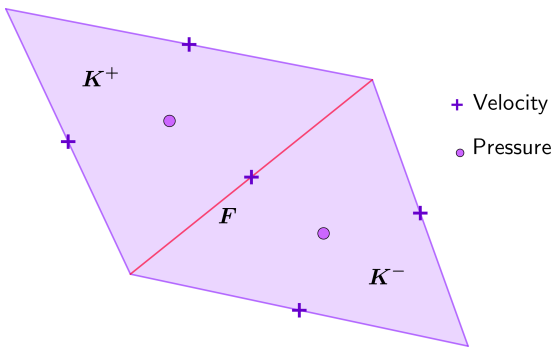


Figure 3.1: Crouzeix–Raviart element with locations of unknowns in two dimensions.

We note V_h and M_h the approximation space of the velocity $\boldsymbol{u_h}$ and the pressure p_h respectively. The spaces V_h and M_h are defined as

$$V_h = \left\{ \boldsymbol{v_h} \in L^2(\Omega)^d \ s.t. \ \boldsymbol{v_h}|_K \in \mathbb{P}_1(K)^d, \ \boldsymbol{v_h} \text{ is continuous at the middle of the faces, } \forall K \in \mathcal{T}_h \right\},$$

$$M_h = \{ q_h \in L^2_0(\Omega) \ s.t. \ q_h|_K \in \mathbb{P}_0(K), \ \forall K \in \mathcal{T}_h \}.$$

The approximation space V_h is not included in $V = H^1_0(\Omega)^d$ and thus it is non-conforming. Let $\{\boldsymbol{a_0}, \ldots, \boldsymbol{a_d}\}$ be the vertices of K, F_i be the face of K opposite to $\boldsymbol{a_i}$ and $\boldsymbol{n_i}$ be the outward normal to the face F_i. The associated barycentric coordinates $(\lambda_0, \ldots, \lambda_d)$ [75] are defined by

$$\lambda_i : \begin{array}{ccc} \mathbb{R}^d & \to & \mathbb{R} \\ \boldsymbol{x} & \mapsto & \lambda_i(\boldsymbol{x}) = 1 - \frac{(\boldsymbol{x}-\boldsymbol{a_i})\cdot\boldsymbol{n_i}}{(\boldsymbol{a_j}-\boldsymbol{a_i})\cdot\boldsymbol{n_i}} \quad \text{for } 0 \leq i \leq d, \end{array}$$

where $\boldsymbol{a_j}$ is an arbitrary vertex in F_i. The barycentric coordinate λ_i is an affine function, it is equal to 1 at $\boldsymbol{a_i}$ and vanishes at F_i, i.e. $\lambda_i(\boldsymbol{a_j}) = \delta_{ij}$ (with δ_{ij} the Kronecker symbol). In particular, the barycentric coordinates satisfy the following properties. For all $\boldsymbol{x} \in K$, $0 \leq \lambda_i(\boldsymbol{x}) \leq 1$, and for all $\boldsymbol{x} \in \mathbb{R}^d$,

$$\sum_{i=0}^{d} \lambda_i(\boldsymbol{x}) = 1 \quad \text{and} \quad \sum_{i=0}^{d} \lambda_i(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{a_i}) = 0.$$

Let $\{\phi_0, \ldots, \phi_d\}$ be the basis functions of each component of V_h (the scalar version of V_h is of dimension $d+1$). Using the barycentric coordinates $(\lambda_0, \ldots, \lambda_d)$, ϕ_i reads on each element K

$$\phi_i|_K = 1 - d\lambda_i|_K, \ \text{for } 0 \leq i \leq d.$$

By noting $\boldsymbol{x_i}$ the center of the face F_i, we have $\phi_i|_K(\boldsymbol{x_i}) = \delta_{ij}$. Consequently, any function of V_h can be written as

$$\boldsymbol{u_h} = \sum_{i=0}^{d} \boldsymbol{u_h}(\boldsymbol{x_i})\phi_i(\boldsymbol{x}).$$

By introducing $[[\boldsymbol{v_h}]]$ the jump of $\boldsymbol{v_h}$ across an internal face F defined by

$$[[\boldsymbol{v_h}]]_F = \boldsymbol{v_h}^- \boldsymbol{n}^- + \boldsymbol{v_h}^+ \boldsymbol{n}^+,$$

where F is the common face of $K^\pm \in \mathcal{T}_h$, $\boldsymbol{v_h}^\pm = \boldsymbol{v_h}|_{K^\pm}$ and $\boldsymbol{n}^\pm$ are the outward unit normal of $K^\pm$ along F, and with $[[\boldsymbol{v_h}]] = \boldsymbol{v_h}$ on $\partial\Omega$, then the space V_h can be written as

$$\tilde{V_h} = \{ \boldsymbol{v_h} \in L^2(\Omega)^d \ s.t. \ \boldsymbol{v_h}|_K \in \mathbb{P}_1(K)^d, \int_F [[\boldsymbol{v_h}]] = 0, \ \forall K \in \mathcal{T}_h, \ \forall F \in \mathcal{F}_h \}$$

Indeed, since $\phi_i|_K \in \mathbb{P}_1(K)$, the mean value over a face is equal to the value at the barycenter. The condition $\int_{F_i} [[\boldsymbol{v_h}]] = 0$ for $F_i \in \mathcal{F}_h$ implies that functions of $\tilde{V_h}$ are only continuous at the center of the face F_i. Therefore, V_h and $\tilde{V_h}$ are the same spaces.

3.2 Inf-sup related theorems

3.2.1 Continuous version

Let V and M be two Hilbert spaces. The scalar product defined on these spaces are denoted respectively by $(\cdot,\cdot)_V$ and $(\cdot,\cdot)_M$. The norms associated with these scalar products are denoted respectively by $\|\cdot\|_V$ and $\|\cdot\|_M$. Let V' and M' be the dual spaces of V and M and let $\|\cdot\|_{V'}$ and $\|\cdot\|_{M'}$ be the associated dual norms. The dual space V' (respectively M') is the space of continuous linear forms defined on V (respectively M). We denote by $< \cdot,\cdot >$ the product of an element in the Hilbert space and an element of its dual space. Let f and g be element of V' and M' respectively, i.e. f and g are two linear forms.

Let $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ be two continuous bi-linear forms, defined by

$$a(\cdot,\cdot) : V \times V \to \mathbb{R} \quad \text{and} \quad b(\cdot,\cdot) : M \times V \to \mathbb{R}.$$

Then we consider the following variational problem. Given $f \in V'$ and $g \in M'$, find $(u,p) \in V \times M$ such that

$$\begin{cases} a(u,v) + b(p,v) =< f, v > \ \forall v \in V, \\ b(q,u) =< g, q > \ \forall q \in M. \end{cases} \tag{3.1}$$

First, we suppose that $g = 0$, and introduce the following subspace W of V defined by

$$W = \{v \in V \ s.t. \ \forall q \in M, \ b(q,v) = 0\}.$$

The problem (3.1) can be rewritten as: find $u \in W$ such that

$$a(u,v) =< f, v > \ \forall v \in W. \tag{3.2}$$

The well-posedness of problem (3.2) is ensured by the Lax-Milgram lemma [75, Lemma 2.2].

Lemma 3.1 (Lax-Milgram)**.** We assume that

- $a(\cdot,\cdot)$ is continuous, i.e. there exists a constant β such that

$$|a(u,v)| \leq \beta \|u\|_W \|v\|_W, \ \forall u, v \in W.$$

- $a(\cdot,\cdot)$ is coercive on W, i.e. there exists a constant α such that

$$a(v,v) \geq \alpha \|v\|_W^2, \ \forall v \in W.$$

Then, the problem (3.2) is well-posed i.e. has one and only one solution $u \in W$, with the a priori estimate

$$\|u\|_W \leq \frac{1}{\alpha} \|f\|_{V'}.$$

Proof. This lemma is a consequence of the Banach-Nečas–Babuška (BNB) Theorem (Theorem 3.2). □

A generalization of Lax-Milgram's lemma and its proof is presented in [75, Theorem 2.6]. Indeed as noted in [75] the Lax-Milgram Lemma holds in Hilbert spaces only, since the coercivity (an Hilbertian property) is required, while the Banach-Nečas–Babuška (BNB) Theorem holds in a Banach space.

Theorem 3.2 (Banach-Nečas–Babuška (BNB))**.** Let X be a Banach space and let Y be a reflexive Banach space. Let $a(\cdot,\cdot)$ be a bi-linear form: $X \times Y \to \mathbb{R}$ and $f \in Y'$. Then, the problem (3.2) is a well-posed if and only if

$$\exists \alpha > 0,\ \inf_{w\in X} \sup_{v\in Y} \frac{a(w,v)}{\|w\|_X \|v\|_Y} \geq \alpha,$$
$$\forall v \in Y,\ (\forall w \in X, a(w,v) = 0) \implies (v = 0).$$

Moreover, the following a priori estimate holds

$$\forall f \in Y', \|u\|_X \leq \frac{1}{\alpha}\|f\|_{Y'}.$$

For more general problems, in the case $g \neq 0$, we cannot introduce a similar subspace W. Consequently, we need necessary and sufficient conditions so that the saddle point problem (3.1) is well-posed, i.e. has one and only one solution. This case has been considered in [84].

Theorem 3.3. Assume that

1. $a(\cdot,\cdot)$ is a bi-linear form continuous on $V \times V$,
2. $b(\cdot,\cdot)$ is a bi-linear form continuous on $M \times V$,
3. $a(\cdot,\cdot)$ is V-elliptic, i.e. there exists a constant $\alpha > 0$ such that
$$a(v,v) \geq \alpha \|v\|_V^2 \ \forall v \in V,$$
4. $b(\cdot,\cdot)$ verifies the inf-sup condition: there exists a constant $\beta > 0$ such that
$$\inf_{q\in M,\ q\neq 0} \sup_{v\in V} \frac{b(q,v)}{\|v\|_V \|q\|_M} \geq \beta. \tag{3.3}$$

Then, the problem (3.1) is well-posed and it has one unique solution $u \in V$, $p \in M$ for any $f \in V'$, $g \in M'$.

The inf-sup condition (3.3), also called the LBB condition for Ladyzhenskaya–Babuška–Brezzi, was introduced by Babuška [23] and Brezzi [40].

3.2.2 Discrete version

When the problem (3.1) is discretized using finite elements, it is important that the discrete inf-sup condition is satisfied. Let h denote the discretization parameter tending to zero. Let V_h and M_h two finite dimensional spaces, called approximation spaces. We introduce two bi-linear forms $a_h(\cdot,\cdot) : V_h \times V_h \to \mathbb{R}$ and $b_h(\cdot,\cdot) : M_h \times V_h \to \mathbb{R}$. We approximate the problem (3.1) by the following discrete problem. Find $(u_h, p_h) \in V_h \times M_h$ such that

$$\left\{ \begin{array}{rcll} a_h(u_h, v_h) + b(p_h, v_h) & = & < f, v_h > & \forall v_h \in V_h, \\ b_h(q_h, u_h) & = & < g, q_h > & \forall q_h \in M_h, \end{array} \right. \tag{3.4}$$

with $f \in V'$ and $g \in M'$. The existence and uniqueness of a solution to system (3.4) is guaranteed by the following discrete inf-sup condition, which is a discrete version of Theorem 3.3.

Theorem 3.4. Assume that

1. $a_h(\cdot,\cdot)$ is a bi-linear form continuous on $V_h \times V_h$,
2. $b_h(\cdot,\cdot)$ is a bi-linear form continuous on $M_h \times V_h$,
3. $a_h(\cdot,\cdot)$ is V_h-elliptic, i.e. there exists a constant $\alpha_h > 0$ such that

$$a_h(v_h, v_h) \geq \alpha_h \|v_h\|^2_{V_h} \ \forall v_h \in V_h,$$

4. $b_h(\cdot,\cdot)$ verifies the inf-sup condition: there exists a constant $\beta_h > 0$ such that

$$\inf_{q_h \in M_h,\ q_h \neq 0} \sup_{v_h \in V_h} \frac{b_h(q_h, v_h)}{\|v_h\|_{V_h} \|q_h\|_{M_h}} \geq \beta_h.$$

Then, the problem (3.4) is well-posed and it has one unique solution $u_h \in V_h$, $p_h \in M_h$ for any $f \in V'$, $g \in M'$.

3.3 Introduction to homogenization

In Chapter 6, we address the homogenization of the Stokes problem in a periodic perforated domain. Therefore, for the sake of understanding, in this section we make a brief introduction to homogenization, largely based on [10]. For a more advanced presentation of homogenization the reader may refer to the following books [31, 26, 164, 57, 161]. Homogenization is a rigorous version of what is known as averaging. To put it another way, homogenization enables the extraction of homogeneous, effective parameters from disordered or heterogeneous media. The technique of homogenization was initially developed for periodic structures. Indeed, in numerous fields of science and technology, one is required to solve boundary value problems in periodic media. However, homogenization is not limited to the periodic case and can be applied to any kind of disordered media. In general, the size of the period is relatively small compared to the size of a sample of the medium. By denoting by ε their ratio, an asymptotic analysis, is therefore necessary as ε goes to zero. Homogenization is the process whereby an asymptotic

analysis is conducted in order to derive an averaged formulation. It enables the transition from a microscopic description of a problem to a macroscopic, or effective, description, making it easier to analyse or to solve.

3.3.1 General procedure of homogenization

In what follows, we give the main steps of homogenization.

1. Identify the small parameter ε that characterises the ratio between the microscopic and macroscopic scales.

2. Expand the solution u_ε in powers of ε (asymptotic expansion) generally as

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

3. Replace the asymptotic expansion into the governing equations and gather terms of the same order in ε, in order to derive the effective or macroscopic equation. The leading-order terms usually yield the effective macroscopic equations.

4. Integrate the microscopic properties to derive the effective (homogenized) coefficients for the macroscopic equations.

As result, we obtain an effective macroscopic model that captures the averaged behaviour of the microscopic model. This model is often simpler and more tractable than the original microscopic one. For example, in the case of heat conduction, the homogenized equation could be a heat equation with effective thermal conductivity. For fluid flows in porous media, the homogenized equation could be the Darcy law with an effective permeability.

To introduce the mathematical approach, in what follows, we apply this procedure to a simple problem of conductivity.

3.3.2 Setting of the problem

Let Ω be a periodic domain (see for example Figure 3.2) which is a bounded open set in $\mathbb{R}^d$ (d is the space dimension) with periodicity $\varepsilon > 0$ (assumed to be very small in comparison with the size of the domain), and Y be the re-scaled unit periodic cell $Y = (0,1)^d$. The conductivity in Ω is not constant, but varies periodically with period ε in each direction. It is a matrix tensor $A(y)$, where $y = x/\varepsilon \in Y$ is the fast variable, while $x \in \Omega$ is the slow variable. Equivalently, x is called the macroscopic variable, and y the microscopic variable. The matrix A can be any second order tensor that is bounded and positive definite, i.e. there exist two positive constants $\beta \geq \alpha > 0$ such that, for any vector $\boldsymbol{\xi} \in \mathbb{R}^d$ and at any point $y \in Y$,

$$\alpha|\boldsymbol{\xi}|^2 \leq \sum_{i,j=1}^{d} A_{ij}(y)\xi_i\xi_j \leq \beta|\boldsymbol{\xi}|^2.$$

The matrix $A(y)$ is a periodic function of y, with period Y, and it may be discontinuous in y (to model the discontinuity of conductivities from one phase to the other).

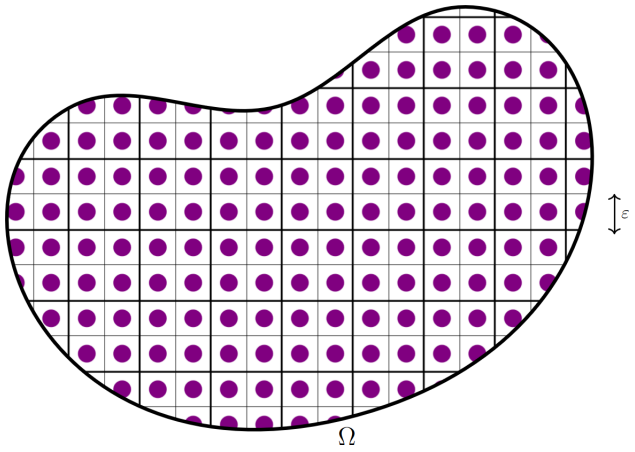


Figure 3.2: A periodic domain Ω.

Denoting by $f(x)$ the source term (a scalar function defined in Ω), and enforcing a Dirichlet boundary condition (for simplicity), the model problem of conductivity reads

$$\begin{cases} -\operatorname{div}\left(A\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla} u_\varepsilon\right) &= f \quad \text{in } \Omega, \\ u_\varepsilon &= 0 \quad \text{on } \partial\Omega, \end{cases} \tag{3.5}$$

where $u_\varepsilon(x)$ is the unknown function, modelling physical quantities such as temperature, electrical potential. Considering $f \in L^2(\Omega)$, then the problem (3.5) is well-posed due to the Lax–Milgram lemma which implies existence and uniqueness of the solution u_ε in the Sobolev space $H_0^1(\Omega)$.

The domain Ω, with is conductivity $A\left(\frac{x}{\varepsilon}\right)$, is highly heterogeneous with periodic heterogeneities of length-scale ε. In practice, it is not necessary to have the full details of the variations of u_ε, rather, it is sufficient to have some global or averaged behaviour of the domain, considered as a homogeneous domain. What is more, from a numerical point of view, solving problem (3.5), by any method will require too much computational resources, in particular if ε is small since the number of elements for a fixed level of accuracy grows like $1/\varepsilon^d$. It is thus preferable, to find an effective or equivalent macroscopic conductivity of the domain Ω and compute an approximation of u_ε on a coarse mesh.

The mathematical theory of homogenization works in a completely different way to traditional physical approaches such as the Representative Volume Element (RVE), which consists of averaging physical quantities in a sample of heterogeneous medium. In fact, instead of considering a single heterogeneous medium with a fixed length scale, the problem is first embedded in a sequence of similar problems for which the length scale ε becomes smaller and smaller, goes to zero (see Figure 3.3). Then, an asymptotic analysis is performed as ε tends to zero, and the conductivity tensor of the limit problem is said to be the effective or homogenized properties. In the case of a periodic medium Ω, this asymptotic analysis of system (3.5), as the period ε goes to zero, is simple. Formally, the solution u_ε is written as a power series of ε, i.e.

$$u_\varepsilon = \sum_{i=0}^{+\infty} \varepsilon^i u_i.$$

The first term u_0 of this series is identified with the solution of the so-called homogenized equation, for which effective conductivity A^* can be exactly computed. We can show that A^* is a constant tensor, describing a homogeneous medium, which is independent of f and of the boundary conditions. Consequently, numerical computations of the homogenized equation do not require a fine mesh since the heterogeneities of size ε have been averaged out. It is important to note that this homogenized tensor A^* is rarely an ordinary average.

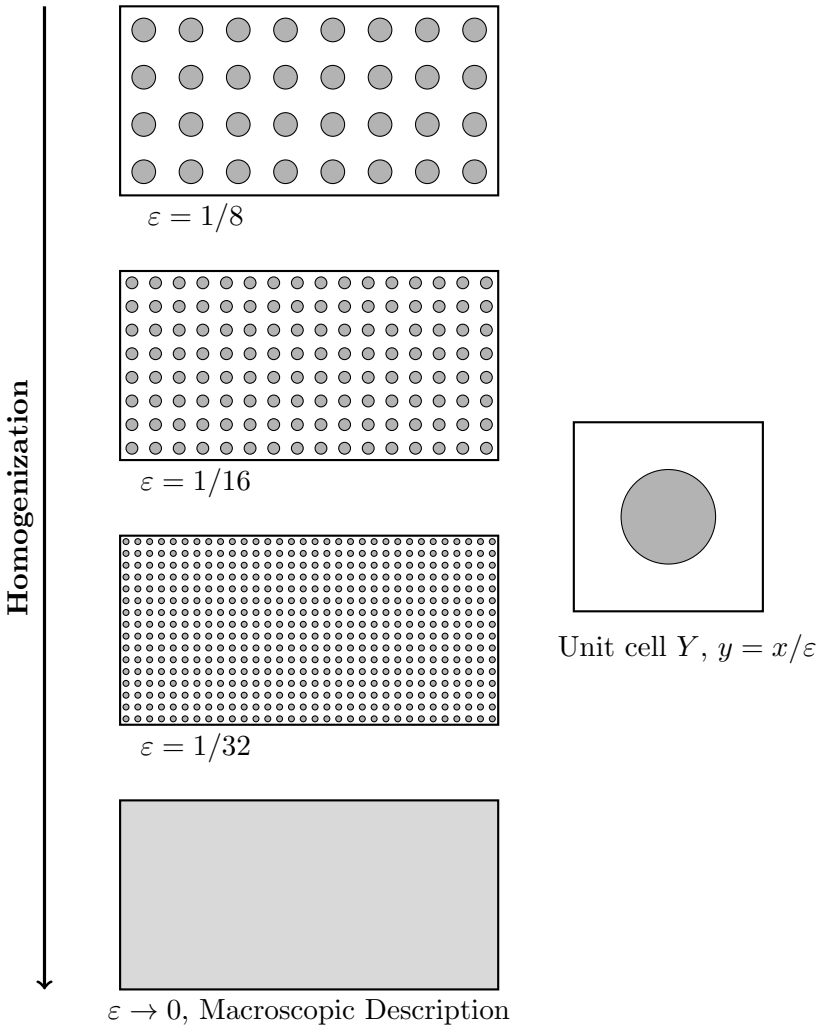


Figure 3.3: Homogenization procedure.

3.3.3 Two-scale asymptotic expansions

3.3.3.1 Ansatz

The method of two-scale asymptotic expansions is an heuristic method, which allows one to formally homogenize a great variety of models or equations posed in a periodic domain. The starting point is to consider the following two-scale asymptotic expansion (also called ansatz), for the solution u_ε of problem (3.5),

$$u_\varepsilon(x) = \sum_{i=0}^{+\infty} \varepsilon^i u_i(x, \frac{x}{\varepsilon}), \tag{3.6}$$

where each term $u_i(x,y)$ is a function of both variables x and y, periodic in y with period $Y=(0,1)^d$, u_i is thus called a periodic function with respect to y. This series is plugged into (3.5), and the following derivation rule is used

$$\boldsymbol{\nabla}\left(u_i\left(x,\frac{x}{\varepsilon}\right)\right)=(\varepsilon^{-1}\boldsymbol{\nabla}_{\boldsymbol{y}}u_i+\boldsymbol{\nabla}u_i)\left(x,\frac{x}{\varepsilon}\right), \tag{3.7}$$

where $\boldsymbol{\nabla}$ and $\boldsymbol{\nabla}_{\boldsymbol{y}}$ denote the partial derivative with respect to the first and second variable of $u_i\left(x,\frac{x}{\varepsilon}\right)$. For example, one has

$$\boldsymbol{\nabla}u_\varepsilon(x)=\varepsilon^{-1}\boldsymbol{\nabla}_{\boldsymbol{y}}u_0\left(x,\frac{x}{\varepsilon}\right)+\sum_{i=0}^{+\infty}\varepsilon^i(\boldsymbol{\nabla}_{\boldsymbol{y}}u_{i+1}+\boldsymbol{\nabla}u_i)\left(x,\frac{x}{\varepsilon}\right).$$

Problem (3.5) becomes a series in ε,

$$\begin{aligned}
&-\varepsilon^{-2}[\operatorname{div}_y A\boldsymbol{\nabla}_{\boldsymbol{y}}u_0]\left(x,\frac{x}{\varepsilon}\right)\\
&-\varepsilon^{-1}[\operatorname{div}_y A(\boldsymbol{\nabla}_{\boldsymbol{x}}u_0+\boldsymbol{\nabla}_{\boldsymbol{y}}u_1)+\operatorname{div}A\boldsymbol{\nabla}_{\boldsymbol{y}}u_0]\left(x,\frac{x}{\varepsilon}\right)\\
&-\varepsilon^{0}[\operatorname{div}A(\boldsymbol{\nabla}u_0+\boldsymbol{\nabla}_{\boldsymbol{y}}u_1)+\operatorname{div}_y A(\boldsymbol{\nabla}u_1+\boldsymbol{\nabla}_{\boldsymbol{y}}u_2)]\left(x,\frac{x}{\varepsilon}\right)\\
&-\sum_{i=1}^{+\infty}\varepsilon^i\left[\operatorname{div}A(\boldsymbol{\nabla}u_i+\boldsymbol{\nabla}_{\boldsymbol{y}}u_{i+1})+\operatorname{div}_y A(\boldsymbol{\nabla}u_{i+1}+\boldsymbol{\nabla}_{\boldsymbol{y}}u_{i+2})\right]\left(x,\frac{x}{\varepsilon}\right)\\
&=f(x).
\end{aligned} \tag{3.8}$$

Identifying each coefficient of (3.8) as an individual equation yields a cascade of equations (a series of the variable ε is zero for all values of ε if each coefficient is zero). The three first equations are enough for our purpose. The ε^{-2} equation is

$$-\operatorname{div}_y A(y)\boldsymbol{\nabla}_{\boldsymbol{y}}u_0(x,y)=0,$$

which is nothing else that an equation in the unit cell Y with periodic boundary condition. In this equation, y is the variable, and x plays the role of a parameter. It can be checked that there exists a unique solution of this equation up to a constant (i.e. a function of x independent of y since x is just a parameter). This implies that u_0 is a function that does not depend on y, i.e. there exists a function $u(x)$ such that

$$u_0(x,y)=u(x).$$

Since $\boldsymbol{\nabla}_{\boldsymbol{y}}u_0=0$, the ε^{-1} equation is

$$-\operatorname{div}_y A(y)\boldsymbol{\nabla}_{\boldsymbol{y}}u_1(x,y)=\operatorname{div}_y A(y)\boldsymbol{\nabla}u(x), \tag{3.9}$$

which is an equation for the unknown u_1 in the periodic cell Y. Again, it is a well-posed problem, which admits a unique solution up to a constant, as soon as the right-hand side is known. Equation (3.9) allows one to compute u_1 in terms of u, and it is easily seen that $u_1(x,y)$ depends linearly on the first derivative $\boldsymbol{\nabla}u(x)$.

Finally, the ε^0 equation is

$$-\operatorname{div}_y A(y)\nabla_{\boldsymbol{y}} u_2(x,y) = \operatorname{div}_y A(y)\nabla u_1 + \operatorname{div} A(y)(\nabla_{\boldsymbol{y}} u_1 + \nabla u) + f(x), \tag{3.10}$$

which is an equation for the unknown u_2 in the periodic unit cell Y. Equation (3.10) admits a solution if a compatibility condition is satisfied (the so-called Fredholm alternative, see Lemma 3.5). Indeed, integrating the left-hand side of (3.10) over Y, and using the periodic boundary condition for u_2, we obtain

$$\int_Y \operatorname{div}_y (A(y)\nabla_{\boldsymbol{y}} u_2(x,y))\mathrm{d}y = \int_{\partial Y} [A(y)\nabla_{\boldsymbol{y}} u_2(x,y)] \cdot \boldsymbol{n} ds = 0,$$

which implies that the right-hand side of (3.10) must have zero average over Y, i.e.

$$\int_Y [\operatorname{div}_y A(y)\nabla u_1 + \operatorname{div} A(y)(\nabla_{\boldsymbol{y}} u_1 + \nabla u) + f(x)]\mathrm{d}y = 0,$$

which simplifies to

$$-\operatorname{div}\left(\int_Y A(y)(\nabla_{\boldsymbol{y}} u_1 + \nabla u)\mathrm{d}y\right) = f(x) \text{ in } \Omega. \tag{3.11}$$

Since $u_1(x,y)$ depends linearly on $\nabla u(x)$, (3.11) is simply an equation for $u(x)$ involving only the second order derivatives of u.

3.3.4 The cell and homogenized problems

The method of two-scale asymptotic expansions leads to a couple of equations ((3.9) and (3.11)) that have a mathematical, as well as physical, interpretation. In order to compute u_1 and to simplify (3.11), we introduce the so-called cell problems. We denote by $(\boldsymbol{e_i})_{1\le i\le d}$ the canonical basis of $\mathbb{R}^d$. For each unit vector $\boldsymbol{e_i}$, consider the following conductivity problem in the periodic unit cell

$$\begin{cases} -\operatorname{div}_y A(y)(\boldsymbol{e_i} + \nabla_{\boldsymbol{y}}\omega_i(y)) = 0 & \text{in } Y, \\ y \to \omega_i(y) & Y-\text{periodic}, \end{cases} \tag{3.12}$$

where $\omega_i(y)$ is the local variation of potential created by an averaged (or macroscopic) gradient $\boldsymbol{e_i}$. The existence of a solution ω_i, to problem (3.12) is guaranteed by Lemma 3.5.

Lemma 3.5. Let $f(y) \in L^2_\#(Y)$ be a periodic function. There exists a solution in $H^1_\#(Y)$ (unique up to an additive constant) of

$$\begin{cases} -\operatorname{div}_y A(y)\nabla\omega(y) = f(y) & \text{in } Y, \\ y \to \omega(y) & Y-\text{periodic}, \end{cases}$$

if and only if $\displaystyle\int_Y f(y)\mathrm{d}y = 0$ (this is called the Fredholm alternative).

We recall that $L^2_\#(Y)$ denotes the L^2 space of Y-periodic functions i.e.

$$L^2_\#(Y) = \{\varphi \in L^2_{loc}(\mathbb{R}^d),\ \varphi \text{ is } Y-\text{periodic}\},$$

endowed with the usual scalar product and norm of $L^2(Y)$, and $H^1_\#(Y)$ denotes the H^1 space of Y-periodic functions, i.e.

$$H^1_\#(Y) = \{\varphi \in H^1_{loc}(\mathbb{R}^d),\ \varphi \text{ is } Y-\text{periodic}\},$$

endowed with the usual scalar product and norm of $H^1(Y)$.

By linearity, it is not difficult to compute $u_1(x,y)$, solution of (3.9) in terms of $u(x)$ and $\omega_i(y)$,

$$u_1(x,y) = \sum_{i=1}^{d} \frac{\partial u}{\partial x_i}(x)\omega_i(y).$$

In reality, $u_1(x,y)$ is merely defined up to the addition of a function $\tilde{u}_1(x)$ (depending only on x), but this not matter since only its gradient $\boldsymbol{\nabla_y} u_1(x,y)$ is used in the homogenized equation. Inserting this expression in (3.11), we obtain the homogenized equation for u that we supplement with a Dirichlet boundary condition on $\partial\Omega$,

$$\left\{ \begin{array}{rcll} -\operatorname{div}\left(A^*\boldsymbol{\nabla} u(x)\right) & = & f(x) & \text{in } \Omega, \\ u & = & 0 & \text{on } \partial\Omega. \end{array} \right.$$

The homogenized conductivity A^* is defined by its entries

$$A^*_{ij} = \int_Y \left[(A(y)\boldsymbol{\nabla_y}\omega_i)\cdot \boldsymbol{e_j} + A_{ij}(y)\right] \mathrm{d}y,$$

or equivalently, after a simple integration by parts in Y,

$$A^*_{ij} = \int_Y A(y)(\boldsymbol{e_i} + \boldsymbol{\nabla_y}\omega_i)\cdot(\boldsymbol{e_j} + \boldsymbol{\nabla_y}\omega_j)\mathrm{d}y.$$

The constant tensor A^* describes the effective or homogenized properties of the heterogeneous material $A\left(\frac{x}{\varepsilon}\right)$. It is important to note that A^* does not depend on the choice of domain Ω, source term f, or boundary condition on $\partial\Omega$. Therefore, the solution u_ε can be approximated (at the order 1) by

$$u_\varepsilon(x) \approx u(x) + \varepsilon u_1\left(x, \frac{x}{\varepsilon}\right) = u(x) + \varepsilon \sum_{i=1}^{d} \frac{\partial u_i}{\partial x_i}(x)\omega_i\left(\frac{x}{\varepsilon}\right),$$

where the term $\varepsilon u_1\left(x, \frac{x}{\varepsilon}\right)$ is the corrector term.

Chapter 4

Presentation of an inf-sup stable non-conforming family of finite elements on triangles of arbitrary accuracy order

Contents

Note: We would like to thank Prof. Frédéric Hecht for his help with the implementation of the finite elements in FreeFEM source code.

This chapter is dedicated to the presentation of the family of non-conforming finite elements of arbitrary accuracy order on triangles in two dimensions introduced in [132]. This family of finite element is useful for solving the discrete local problems in two dimensions arising from the Multi-scale Finite Element Method (MsFEM) for the Stokes or the Oseen problem (see

Chapter 8). For these local problems, the discrete well-posedness could not be shown with classical pairs of finite elements. For this purpose, we have implemented the non-conforming finite element of accuracy order 3 in FreeFEM [93] source code, and it is known under the name $P3pnc$[1]. The non-conforming finite element of accuracy order 2, known under the name $P2pnc$[2], was already implemented, and the one of accuracy order 1, corresponding to the well-known Crouzeix–Raviart finite element [61], was obviously already implemented in FreeFEM, and it is known under the name $P1nc$.

In this chapter, for the sake of completeness, and to help understand Chapter 8, as well as the next chapter, Chapter 5, which is about the extension of this non-conforming finite element in three dimensions on tetrahedra, we recall the definition of this family of non-conforming finite element and the main theoretical results of [132]. Then, we explain the general procedure to implement such a non-conforming finite element in FreeFEM and apply it for the finite element $P3pnc$. Finally, we carry out numerical experiments to assess the implementation of these finite elements.

4.1 Introduction

We consider the Stokes problem in two dimensions. Let $\Omega \subset \mathbb{R}^2$ be a bounded connected domain. The steady-state Stokes problem is to find the velocity $\boldsymbol{u} : \Omega \to \mathbb{R}^2$ and the pressure $p : \Omega \to \mathbb{R}$ solution to

$$\begin{aligned} -\nu\Delta\boldsymbol{u} + \boldsymbol{\nabla} p &= \boldsymbol{f} \quad \text{in } \Omega, \\ \operatorname{div} \boldsymbol{u} &= 0 \quad \text{in } \Omega, \\ \boldsymbol{u} &= \boldsymbol{0} \quad \text{on } \partial\Omega, \end{aligned} \tag{4.1}$$

with $\nu > 0$ the viscosity and $\boldsymbol{f}$ the applied force. We consider homogeneous boundary conditions but all the theoretical analyses hold for non-homogeneous boundary conditions. Vector valued functions are written in bold. We denote by

$$(u, v) = \int_\Omega uv \mathrm{d}x,$$

the scalar product in $L^2(\Omega)$. We use the same notation for vector-valued function, i.e. the $L^2(\Omega)$ scalar product for vector-valued functions is given by

$$(\boldsymbol{u}, \boldsymbol{v}) = \int_\Omega \boldsymbol{u} \cdot \boldsymbol{v} \, \mathrm{d}x,$$

where $\boldsymbol{u} \cdot \boldsymbol{v}$ denotes the Euclidian scalar product in $\mathbb{R}^2$. We define the classical velocity space $V = H_0^1(\Omega)^2$ and pressure space $M = L_0^2(\Omega) = \{q \in L^2(\Omega),\ s.t.\ \int_\Omega q = 0\}$ equipped respectively with the $|\cdot|_1$ H^1 semi-norm and the $\|\cdot\|_0$ L^2 norm. We introduce the bi-linear

[1] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P3pnc.cpp

[2] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P2pnc.cpp

forms $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ and $b(\cdot,\cdot) : M \times V \to \mathbb{R}$ defined by

$$a(\boldsymbol{u},\boldsymbol{v}) = (\boldsymbol{\nabla u}, \boldsymbol{\nabla v}), \qquad b(p,\boldsymbol{v}) = (p,\, \mathrm{div}\, \boldsymbol{v}).$$

Then, a weak formulation of the Stokes problem (4.1) reads as follows. Find $\boldsymbol{u} \in V$ and $p \in M$ such that

$$\left\{ \begin{array}{rcll} a(\boldsymbol{u},\boldsymbol{v}) + b(p,\boldsymbol{v}) & = & (\boldsymbol{f},\boldsymbol{v}) & \forall \boldsymbol{v} \in V, \\ b(q,\boldsymbol{u}) & = & 0 & \forall q \in M, \end{array} \right. \tag{4.2}$$

with $\boldsymbol{f} \in (L^2(\Omega))^2$. It is well-known [84] that there exists a unique weak solution to system (4.2), due to the Ladyzhenskaya–Babuška–Brezzi (LBB) condition, also called inf-sup condition (see Theorem 3.3), namely there exists $\beta > 0$ such that

$$\inf_{q \in M} \sup_{\boldsymbol{v} \in V} \frac{b(q,\boldsymbol{v})}{|\boldsymbol{v}|_1 \|q\|_0} \geq \beta.$$

Next, when the Stokes problem is discretized using finite elements, it is important that the discrete inf-sup condition (4.5) is satisfied uniformly in the mesh size h. This condition imposes constraints on the choice of the pairs of finite elements (for velocity and pressure), since this condition is not satisfied for every pair of finite elements.

4.2 Discrete setting

Let $\Omega \subset \mathbb{R}^2$ be a polyhedral domain. We note $\{\mathcal{T}_h\}$ a family of triangulations of Ω parametrized by a positive parameter h which tends to 0. Each triangulation $\mathcal{T}_h$ consists of a finite number of triangles K such that $\overline{\Omega} = \cup_{K \in \mathcal{T}_h} \overline{K}$. Let $h_K = \mathrm{diam}(K)$ and $h := \max h_K$. We assume that the triangulations are conformal in the sense that the intersection of the closures of two different cells K is either empty, a common vertex or a common edge. Besides, we assume that the triangulations are shape regular, i.e. there exists a positive constant C independent of the mesh parameter h such that

$$\frac{h_K}{\sigma_K} \leq C, \qquad \forall K \in \mathcal{T}_h,\ h_K > 0,$$

where σ_K denotes the maximum diameter of a ball which can be inscribed in $\overline{K}$. We denote by $\mathcal{E}_h$ the set of all edges E of $\mathcal{T}_h$. The length of E and the area of K are denoted respectively by $|E|$ and $|K|$.

For any integer n and any integer $1 \leq l \leq d$ (with d the space dimension, here $d = 2$), we denote by $\mathbb{P}_n^l$ the linear space spanned by l-variate polynomial functions of total degree at most n. The dimension of $\mathbb{P}_n^l$ is

$$N_n^l := \dim(\mathbb{P}_n^l) = \binom{n+l}{n}.$$

For any $K \in \mathcal{T}_h$, we denote by $\mathbb{P}_n(K)^2$ the restriction to K of polynomials in $(\mathbb{P}_n)^2$. For any $E \in \mathcal{E}_h$, we denote by $\mathbb{P}_n(E)$ the restriction to E of $\mathbb{P}_n$. For the sake of simplicity, we denote $\mathbb{P}_n(K)^2$ by $\mathbb{P}_n(K)$.

For a domain $D \subset \mathbb{R}^2$, we use the Sobolev spaces $H^m(D)$, $H_0^m(D)$, and $L^2(D) = H^0(D)$ for non-negative integer m. The norms and semi-norms in the scalar and vector-valued versions of $H^m(D)$ are denoted by $\|\cdot\|_{m,D}$ and $|\cdot|_{m,D}$ respectively. We define the H^1 broken space by

$$H^{1,h}(\Omega) := \{v \in L^2(\Omega),\ s.t.\ v|_K \in H^1(K),\ \forall K \in \mathcal{T}_h\}.$$

On the broken space $H^{1,h}(\Omega)$, an analogue of the semi-norm $|\cdot|_1$ is defined, for any $\boldsymbol{v} \in H^{1,h}(\Omega)^2$, by

$$|\boldsymbol{v}|_{1,h} := \left(\sum_{K \in \mathcal{T}_h} |\boldsymbol{v}|_{1,K}^2 \right)^{1/2},$$

where $|\boldsymbol{v}|_{1,K} = \|\boldsymbol{\nabla v}\|_{0,K}$. Let $\boldsymbol{\nabla_h v}$ and $\mathrm{div}_h\, \boldsymbol{v}$ denote the piecewise gradient and piecewise divergence of $\boldsymbol{v} \in H^{1,h}(\Omega)^2$, defined by

$$(\boldsymbol{\nabla_h v})|_K = \boldsymbol{\nabla}(\boldsymbol{v}|_K), \qquad (\mathrm{div}_h\, \boldsymbol{v})|_K = \mathrm{div}\,(\boldsymbol{v}|_K).$$

Let $V_h \subset \left(L^2(\Omega)\right)^2$ be a non-conforming finite element space and $M_h \subset M$. Let $\boldsymbol{f}$ be in $\left(L^2(\Omega)\right)^2$. Then, the discrete Stokes problem reads as follows. Find $\boldsymbol{u_h} \in V_h$ and $p_h \in M_h$ such that

$$\left\{ \begin{array}{rcll} a_h(\boldsymbol{u_h}, \boldsymbol{v_h}) + b_h(p_h, \boldsymbol{v_h}) & = & (\boldsymbol{f}, \boldsymbol{v_h}) & \forall \boldsymbol{v_h} \in V_h, \\ b_h(q_h, \boldsymbol{u_h}) & = & 0 & \forall q_h \in M_h, \end{array} \right. \tag{4.3}$$

where $a_h(\cdot,\cdot) : V_h \times V_h \to \mathbb{R}$ and $b_h(\cdot,\cdot) : M_h \times V_h \to \mathbb{R}$ are respectively the discrete versions of the bi-linear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$, defined by

$$a_h(\boldsymbol{u_h}, \boldsymbol{v_h}) = (\boldsymbol{\nabla_h u_h}, \boldsymbol{\nabla_h v_h}), \qquad b_h(p_h, \boldsymbol{v}) = (p_h,\, \mathrm{div}_h\, \boldsymbol{v_h}). \tag{4.4}$$

It is well-known that the discrete problem (4.4) is well-posed under the following conditions: $a_h(\cdot,\cdot)$ and $b_h(\cdot,\cdot)$ are continuous, $a_h(\cdot,\cdot)$ is symmetric and coercive and the spaces V_h and M_h satisfy the discrete Ladyzhenskaya–Babuška–Brezzi (LBB) condition (or the discrete inf-sup condition, or are said to be inf-sup stable), namely there exists $\beta^* > 0$, independent of h, such that

$$\inf_{q_h \in M_h} \sup_{\boldsymbol{v_h} \in V_h} \frac{b_h(q_h, \boldsymbol{v_h})}{|\boldsymbol{v_h}|_1 \|q_h\|_0} \geq \beta^*. \tag{4.5}$$

4.3 A family of scalar non-conforming finite element

4.3.1 Definition of the non-conforming finite element

The main goal of the work presented in [132] is to define a scalar non-conforming finite element space such that its vector-valued version associated with a discontinuous pressure approximation of polynomial order n form an inf-sup stable finite element pair of accuracy order $n+1$, $n \geq 0$ for the Stokes problem.

To construct the non-conforming finite element space, they propose to enrich the local polynomial space $\mathbb{P}_{n+1}(K)$ with a proper subspace of $\mathbb{P}_{n+2}(K)$, denoted by $\Sigma_{n+2} \subset \mathbb{P}_{n+2}(K)$, with

a trivial intersction with $\mathbb{P}_{n+1}(K)$. For $n \geq 1$, they define the space Σ_{n+2} as

$$\Sigma_{n+2} = \text{span}\left\{ b_K \lambda_1^{n-1-i} \lambda_2^i,\ i = 0, ..., n-1 \right\},$$

where

$$b_K = (\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_3 - \lambda_1),$$

and λ_1, λ_2, λ_3 are the three barycentric coordinates associated with the triangle K.

Remark 4.1. For $n = 0$, $\Sigma_{n+2} = \emptyset$.

The enriched space $V_{n+1}(K)$ obtained is therefore given by

$$V_{n+1}(K) = \mathbb{P}_{n+1}(K) \oplus \Sigma_{n+2}(K).$$

They introduce the following scalar (local) degrees of freedom, defined on $H^{1,h}(\Omega)$ by

$$N_j^{E_\alpha}(v) = \int_{E_\alpha} v L_j^{E_\alpha} \qquad j \geq 0,\ \alpha = 1, 2, 3, \tag{4.6}$$

$$N_j^{K}(v) = \int_{K} v M_j^{K} \qquad j \geq 0, \tag{4.7}$$

where $(L_j^{E_\alpha})_j$ define an arbitrary basis of $\mathbb{P}_n(E_\alpha)$ of dimension $(n+1)$, and $(M_j^K)_j$ define an arbitrary basis of $\mathbb{P}_{n-1}(K)$ of dimension $n(n+1)/2$. Thus, for each element $K \in \mathcal{T}_h$ and any integer $n \geq 1$, the set of degrees of freedom $\mathcal{N}_{n+1}(K)$ is defined by

$$\begin{aligned} \mathcal{N}_{n+1}(K) := &\left\{ N_j^{E_\alpha} \qquad 1 \leq j \leq (n+1),\ \alpha = 1, 2, 3 \right\} \\ &\cup \left\{ N_j^{K} \qquad 1 \leq j \leq \frac{n(n+1)}{2} \right\}. \end{aligned} \tag{4.8}$$

Remark 4.2. For the case $n = 0$, the set of degrees of freedom is reduced to

$$\mathcal{N}_1(v) = \left\{ \frac{1}{|E_\alpha|} \int_{E_\alpha} v\ ,\ \alpha = 1, 2, 3 \right\}$$

which corresponds clearly to the well-known Crouzeix–Raviart finite element [61].

4.3.2 Theoretical analysis

In this section, we recall the main theoretical results established in [132]. The proof of all the following results can be found in [132].

Theorem 4.3. For any integer $n \geq 0$, the finite element space $V_{n+1}(K)$ is unisolvent with respect to the set of degrees of freedom $\mathcal{N}_{n+1}(K)$.

Theorem 4.4. The finite element V_{n+1} associated with a discontinuous pressure approximation of polynomial order n (noted $\mathbb{P}_n^{\mathrm{dc}}$) satisfies the discrete inf-sup stability condition (4.5) for the discrete Stokes problem.

Lemma 4.5. There exists interpolation operators $\boldsymbol{I}_h : V \to V_{n+1}$ and $J_h : M \to M_h$ such that

$$|\boldsymbol{u} - \boldsymbol{I}_h\boldsymbol{u}|_{1,h} \leq Ch^{n+1}|\boldsymbol{u}|_{n+2} \qquad \forall \boldsymbol{u} \in V \cap H^{n+2}(\Omega),$$

$$\|p - J_h p\|_0 \leq Ch^{n+1}|p|_{n+1} \qquad \forall p \in M \cap H^{n+1}(\Omega).$$

Theorem 4.6. Assume that the approximation property (Lemma 4.5) and the discrete inf-sup condition (4.5) are satisfied. Let $(\boldsymbol{u}, p)$ be the solution to the Stokes problem (4.1) and assume that

$$\boldsymbol{u} \in V \cap (H^{n+2}(\Omega))^2, \; p \in H^{n+1}(\Omega).$$

Then, (4.3) has a unique solution $(\boldsymbol{u_h}, p_h)$ and the following error estimate holds

$$|\boldsymbol{u} - \boldsymbol{u_h}|_{1,h} + \|p - p_h\|_0 \leq Ch^{n+1}\left(|\boldsymbol{u}|_{n+2} + |p|_{n+1}\right),$$

with C a constant independent of the mesh size h.

4.4 Implementation in FreeFEM source code

For numerical applications of the Multi-scale Finite Element Method presented in Chapters 7 and 12, we need the non-conforming finite elements V_2 (finite element of accuracy order two corresponding to $n = 1$) and V_3 (finite element of accuracy order three, corresponding to $n = 2$). The non-conforming finite element V_2 (finite element $P2pnc$) being already implemented in FreeFEM, as explained in the introduction of this chapter, part of the thesis work was to implement the finite element V_3 (finite element $P3pnc$) in FreeFEM source code. In this section, we explain the implementation of such a non-conforming finite element in FreeFEM. We note $M := \mathrm{card}(\mathcal{N}_{n+1})$, the number of degrees of freedom of the finite element V_{n+1}. The implementation is done in the reference element $\widehat{K}$ and then an affine operator is applied to define the finite element in any element K.

4.4.1 Strategy to construct a finite element interpolant

When implementing a new finite element in FreeFEM, the main task is to build the local (scalar-valued) finite element interpolant $I_h^K : H^1(K) \to V_{n+1}(K)$. We recall that in the case of a non-conforming finite element, the structure of the local finite element interpolant is given by

$$I_h^K f = \sum_{l=0}^{M-1} N_l(f)\ \Phi_l^K, \tag{4.9}$$

with f the function to be interpolated, $N_l \in \mathcal{N}_{n+1}$, defined in (4.8), a local degree of freedom, and Φ_l^K the associated scalar local basis functions, satisfying

$$N_i(\Phi_j^K) = \delta_{ij}, \text{ for } 0 \leq i, j \leq M-1, \tag{4.10}$$

where δ_{ij} is the Kronecker symbol. Finally, the local finite element interpolant (4.9) satisfies

$$N_j(I_h^K f) = N_j(f), \text{ for } j = 0, \ldots, M-1.$$

Consequently, the construction of this local interpolant consists of two main steps. The first step is to be able to compute numerically the quantities $N_i(f)$ defined in (4.6) and (4.7). The second step consists in building a set of local scalar basis functions Φ_i^K, such that (4.10) holds.

In Section 4.4.2, we explain the computation of the quantities $N_i(f)$. In Section 4.4.3, we explain the computation of the local scalar basis functions Φ_i^K. Finally, in Section 4.4.5, we summarise the final structure of the local finite element interpolant.

4.4.2 Computation of the quantities $N_i(f)$

The quantities $N_i(f)$ in (4.9), i.e. integrals of type (4.6) and (4.7), are not determined analytically. They are approximated using quadratures rules.

Definition 4.7. A quadrature rule is a linear map of the form

$$I_D^{nQF} : \mathcal{F} \mapsto \sum_{l=0}^{nQF-1} \omega_l \mathcal{F}(\boldsymbol{\zeta_l}),$$

which associates to any function $\mathcal{F}$ a quantity $I_D^n(\mathcal{F})$ supposed to approximate $\displaystyle\int_D \mathcal{F}(\boldsymbol{x})\mathrm{d}x$ i.e.

$$\int_D \mathcal{F}(\boldsymbol{x}) \approx \sum_{l=0}^{nQF-1} \omega_l \mathcal{F}(\boldsymbol{\zeta_l}),$$

where nQF is the number of points of evaluation. The points $(\boldsymbol{\zeta_l})_l$ are called the integration points and the coefficients $(\omega_l)_l$ are called weights.

In this section, the domain D represents either the element K or the edge E_α.

Remark 4.8. Generally, the quadrature formula can be defined on the reference edge (the segment $[0, 1]$) thanks to a change of variable. For example, for one dimensional integration, between the points $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$, we have

$$\mathcal{F}(\boldsymbol{x}) = g(t),$$

with $0 \leq t \leq 1$ and $\boldsymbol{x}$ defined as $\boldsymbol{x} = (1-t)\boldsymbol{x_0} + t\boldsymbol{x_1}$. The two-dimensional case is treated in the same way.

4.4.2.1 Construction of the lists of integration points and weights for degrees of freedom defined over edges

Using quadrature formulas to compute the integrals (4.6) leads to the following approximation

$$\begin{aligned} N_j^{E_\alpha}(f) &= \sum_{l=0}^{nQFE-1} c_l L_j(P_l^{E_\alpha}) f(P_l^{E_\alpha}), \\ &= \sum_{l=0}^{nQFE-1} w_{j,l}^{E_\alpha} f(P_l^{E_\alpha}), \end{aligned}$$

for $j = 0 \ldots, n$ and $\alpha = 1, 2, 3$. The $P_l^{E_\alpha}$ designate the integration points, the $w_l^{E_\alpha}$ the associated weights, and $nQFE$ the number of integration points.

In the case of the non-conforming finite element V_3, for the integration over the edge E_α (one dimensional integration), we use the quadrature formula `qf4pE`, defined in FreeFEM. This quadrature corresponds to the quadrature rule of Gauss Legendre of order 4, which is exact for polynomial of order 7. This quadrature rule is constructed as (with the change of variable described in Remark 4.8)

$$\int_{E_\alpha} \mathcal{F}(\boldsymbol{x}) \approx \frac{|E_\alpha|}{72} \left(b_1 g(s_1) + b_1 g(s_2) + b_2 g(s_3) + b_2 g(s_4) \right),$$

where

$$b_1 = 18 - \sqrt{30}, \qquad b_2 = 18 + \sqrt{30},$$

$$s_1 = \frac{1 - \frac{\sqrt{525+70\sqrt{30}}}{35}}{2}, \qquad s_2 = \frac{1 + \frac{\sqrt{525+70\sqrt{30}}}{35}}{2},$$

$$s_3 = \frac{1 - \frac{\sqrt{525-70\sqrt{30}}}{35}}{2}, \qquad s_4 = \frac{1 + \frac{\sqrt{525-70\sqrt{30}}}{35}}{2}.$$

Noting $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$ the two vertices of the edge E_α, the list P^{E_α} of integration points are defined as

$$P^{E_\alpha} = \{ s_1\boldsymbol{x_0} + (1-s_1)\boldsymbol{x_1}, s_2\boldsymbol{x_0} + (1-s_2)\boldsymbol{x_1}, s_3\boldsymbol{x_0} + (1-s_3)\boldsymbol{x_1}, s_4\boldsymbol{x_0} + (1-s_4)\boldsymbol{x_1} \}$$

and the lists of weights are defined as

$$\begin{aligned} w_j^{E_\alpha} = \{ & b_1 L_j(s_1\boldsymbol{x_0} + (1-s_1)\boldsymbol{x_1}),\ b_1 L_j(s_2\boldsymbol{x_0} + (1-s_2)\boldsymbol{x_1}), \\ & b_2 L_j(s_3\boldsymbol{x_0} + (1-s_3)\boldsymbol{x_1}),\ b_2 L_j(s_4\boldsymbol{x_0} + (1-s_4)\boldsymbol{x_1}) \}. \end{aligned}$$

4.4.2.2 Construction of the lists of integration points and weights for degrees of freedom defined over element

As for the edges, using quadrature formulas to compute the integrals (4.7) leads to the following approximation

$$\begin{aligned} N_j^K(f) &= \sum_{l=0}^{nQFK-1} d_l M_j^K(P_l^K) f(P_l^K) \\ &= \sum_{l=0}^{nQFK-1} w_{j,l}^K f(P_l^K), \end{aligned}$$

for $j = 1, \ldots, \frac{n(n+1)}{2}$. The P_l^K designates the integration points, the w_l^K the associated weights, and $nQFK$ the number of integration points.

In the case of the non-conforming finite element V_3, for the integration over the elements (two dimensional integration), we use the quadrature formula `qf5pT`, defined in FreeFEM. This quadrature rule, exact for polynomials of order 5, is constructed (with the change of variable of Remark 4.8) as [159], i.e.

$$\begin{aligned} \int_K \mathcal{F}(\boldsymbol{x}) \approx \big(&c_1 g(t_1,t_1) + c_2 g(t_2,t_2) + c_2 g(t_2,t_4) + c_2 g(t_4,t_2) \\ &+ c_3 g(t_3,t_3) + c_3 g(t_3,t_5) + c_3 g(t_5,t_3)\big), \end{aligned}$$

with

$$\begin{aligned} c_1 &= 0.225, \qquad c_2 = \frac{155-\sqrt{15}}{1200}, \qquad c_3 = \frac{155+\sqrt{15}}{1200}, \\ t_1 &= \frac{1}{3}, \qquad t_2 = \frac{6-\sqrt{15}}{21}, \qquad t_3 = \frac{6+\sqrt{15}}{21}, \\ & \qquad t_4 = \frac{9+\sqrt{15}}{21}, \qquad t_5 = \frac{9-\sqrt{15}}{21}. \end{aligned}$$

Noting (x_i, x_j) the global coordinates of any point in the triangle K defined by the parametric coordinates (t_i, t_j), the list P^{E_α} of evaluation points are defined as

$$P^K = \{(x_1,x_1),(x_2,x_2),(x_2,x_4),(x_4,x_2),(x_3,x_3),(x_3,x_5),(x_5,x_3)\},$$

and the lists of weights are defined as

$$\begin{aligned} w_j^K = \big\{&c_1 M_j^K(x_1,x_1),\ c_2 M_j^K(x_2,x_2),\ c_2 M_j^K(x_2,x_4),\ c_2 M_j^K(x_4,x_2), \\ &c_3 M_j^K(x_3,x_3),\ c_3 M_j^K(x_3,x_5),\ c_3 M_j^K(x_5,x_3)\big\}. \end{aligned}$$

4.4.3 Computation of the local basis functions Φ_i^K

4.4.3.1 General idea

To build the local scalar-valued basis functions Φ_i^K, for $0 \leq i \leq M-1$, first we consider a basis of the space V_{n+1} composed of monomials p_k, for $0 \leq k \leq M-1$. Then, we construct Φ_j^K as a linear combination of the mononials p_k, i.e.

$$\Phi_i^K = \sum_{k=0}^{M-1} a_{i,k} p_k. \tag{4.11}$$

Now, the goal is to determine the coefficients $a_{i,k}$, such that (4.10) holds. To do this, we apply the degree of freedom N_i to both members of (4.11), leading to

$$N_j(\Phi_i^K) = \sum_{k=0}^{M-1} a_{i,k} N_i(p_k) = \delta_{ij}. \tag{4.12}$$

By defining the matrix $\mathcal{C}$ as

$$\mathcal{C} = \left(N_j\left(p_i\right)\right)_{0 \leq i,j \leq M-1},$$

and the unknown matrix $\mathcal{A} = (a_{ij})_{i,j}$, it follows that (4.12) is equivalent to

$$\mathcal{A}\mathcal{C} = I_M,$$

with I_M the identity matrix of size M. Therefore, the coefficients a_{ij} are given by

$$\mathcal{A} = \mathcal{C}^{-1},$$

where $\mathcal{C}^{-1}$ is the inverse matrix of $\mathcal{C}$, which is invertible since the $(p_k)_k$ is a basis of the space V_{n+1}. At the end, we compute the local basis functions Φ_i^K as

$$\begin{pmatrix} \Phi_0^K \\ \vdots \\ \Phi_{M-1}^K \end{pmatrix} = \mathcal{C}^{-1} \begin{pmatrix} p_0 \\ \vdots \\ p_{M-1} \end{pmatrix}.$$

For the sake of understanding, let us consider the following example.

Example 4.9. We consider a finite element W_0 with two degrees of freedom N_1 and N_2. We define also two monomials p_1 and p_2, basis of W_0. We note Φ_1 and Φ_2 the two basis functions that we are looking for. The matrix $\mathcal{C}$ is defined by

$$\mathcal{C} = \begin{pmatrix} N_1(p_1) & N_2(p_1) \\ N_1(p_2) & N_2(p_2) \end{pmatrix},$$

and it inverse $\mathcal{C}^{-1}$ is given by

$$\mathcal{C}^{-1} = \frac{1}{\det(\mathcal{C})} \begin{pmatrix} N_2(p_2) & -N_2(p_1) \\ -N_1(p_2) & N_1(p_1) \end{pmatrix},$$

where $\det(\mathcal{C}) = N_1(p_1)N_2(p_2) - N_1(p_2)N_2(p_1)$. The basis functions are therefore defined by

$$\begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} = \mathcal{C}^{-1} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \frac{1}{\det(\mathcal{C})} \begin{pmatrix} N_2(p_2)p_1 - N_2(p_1)p_2 \\ -N_1(p_2)p_1 - N_1(p_1)p_2 \end{pmatrix}.$$

Now, it is easy to check that

$$\begin{cases} N_1(\Phi_1) = 1, & N_2(\Phi_1) = 0, \\ N_1(\Phi_2) = 0, & N_2(\Phi_2) = 1. \end{cases}$$

Consequently, we have built a set of basis functions such as $N_i(\Phi_j) = \delta_{ij}$.

4.4.3.2 Numerical applications for V_3

Remark 4.10. Some of these basis functions depend on the edge orientation. Consequently, in FreeFEM, the implementation is not exactly as presented here. According to the edge orientation, one needs to swap some monomials.

To implement the finite element V_3, we choose the monomials p_k as

$$\begin{cases} p_0 = \lambda_0^3, & p_1 = \lambda_1^3, & p_2 = \lambda_2^3, \\ p_3 = \lambda_0^2\lambda_1, & p_4 = \lambda_0^2\lambda_2, & p_5 = \lambda_1^2\lambda_0, \\ p_6 = \lambda_1^2, & p_7 = \lambda_2^2\lambda_0, & p_8 = \lambda_2^2\lambda_0, \\ p_9 = \lambda_0\lambda_1\lambda_2, & p_{10} = b_K\lambda_0, & p_{11} = b_K\lambda_1. \end{cases}$$

where $\lambda_0, \lambda_1, \lambda_2$ are the barycentric coordinates of the triangle K, defined by

$$\lambda_0 = 1 - x - y, \qquad \lambda_1 = x, \qquad \lambda_2 = y,$$

and

$$b_K = (\lambda_0 - \lambda_1)(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_0).$$

To define the degrees of freedom $N_j^{E_\alpha}$, we choose for the set $\left\{L_j^{E_\alpha}\right\}_j$, which defines a basis of the space $\mathbb{P}_2(E_\alpha)$, the set $\{t, 1-t, t(1-t)\}$ with t the parametric coordinate on the edge. To define the degree of freedom N_j^K, we choose for the set $\left\{M_j^K\right\}_j$, which defines a basis of the

space $\mathbb{P}_1(K)$, the set $\{\lambda_0, \lambda_1, \lambda_2\}$. Using the following formula [163],

$$\int_D \prod_{i=0}^{d} \lambda_i^{\alpha_i} = \frac{1}{|D|} \frac{d! \prod_{i=0}^{d} \alpha_i!}{(d + \sum_{i=0}^{d} \alpha_i)!},$$

where $D = K$ and $d = 2$ for integrals over elements and $D = E_\alpha$ and $d = 1$ for integrals over edges, the matrix $\mathcal{C}^{-1}$ for the finite element V_3 is given by

$$\begin{pmatrix}
-1.2 & -1.2 & 1.2 & 1.2 & 6 & -15.6 & 6 & 1.2 & -15.6 & -9.6 & 4.8 & 4.8 \\
6 & 1.2 & -15.6 & -1.2 & -1.2 & 1.2 & 1.2 & 6 & -15.6 & 4.8 & -9.6 & 4.8 \\
1.2 & 6 & -15.6 & 6 & 1.2 & -15.6 & -1.2 & -1.2 & 1.2 & 4.8 & 4.8 & -9.6 \\
-8.4 & -73.2 & 298.8 & 73.2 & 10.8 & -306 & -46.8 & 20.4 & 97.2 & 189.6 & -192 & 2.4 \\
10.8 & 75.6 & -205.2 & -63.6 & -46.8 & 349.2 & 10.8 & -10.8 & -54 & 21.6 & 86.4 & -108 \\
10.8 & 73.2 & -306 & -73.2 & -8.4 & 298.8 & 20.4 & -46.8 & 97.2 & -192 & 189.6 & 2.4 \\
-46.8 & -63.6 & 349.2 & 75.6 & 10.8 & -205.2 & -10.8 & 10.8 & -54 & 86.4 & 21.6 & -108 \\
-10.8 & -73.2 & 198 & 37.2 & 20.4 & -154.8 & -8.4 & 10.8 & 46.8 & -24 & -81.6 & 105.6 \\
20.4 & 37.2 & -154.8 & -73.2 & -10.8 & 198 & 10.8 & -8.4 & 46.8 & -81.6 & -24 & 105.6 \\
12 & 12 & -180 & 12 & 12 & -180 & 12 & 12 & -180 & 60 & 60 & 60 \\
0 & -84 & 252 & 84 & 84 & -504 & -84 & 0 & 252 & 84 & -168 & 84 \\
-84 & -84 & 504 & 84 & 0 & -252 & 0 & 84 & -252 & 168 & -84 & -84
\end{pmatrix}.$$

4.4.4 Computation of the derivatives of the local basis functions Φ_i^K

The local basis function Φ_i^K being defined as a linear combination of the monomials p_k

$$\Phi_i^K = \sum_{k=0}^{M-1} a_{i,k} p_k,$$

their derivatives are defined as linear combinations of derivatives of the monomials p_k, i.e.

$$\frac{\partial \Phi_i^K}{\partial x} = \sum_{k=0}^{M-1} a_{i,k} \frac{\partial p_k}{\partial x},$$
$$\frac{\partial \Phi_i^K}{\partial y} = \sum_{k=0}^{M-1} a_{i,k} \frac{\partial p_k}{\partial y}.$$

4.4.5 Final structure of the finite element interpolant

At the end, we implement the finite element interpolant using a summation indexed by k, i.e.

$$I_h^K f = \sum_{k=0}^{kPi-1} \alpha_k f_{j_k}(P_{p_k})\ \Phi_{i_k}^K. \tag{4.13}$$

Below, we detail all the indices and lists involved in (4.13). First, we have

$$kPi = 57,$$

since this finite element has three degrees of freedom on each edge and for each of them we use four integration points to compute integrals of type (4.6), and three degrees of freedom on the

element and for each of them we use seven integration points to compute integrals of type (4.7), i.e.

$$\begin{aligned} 57 =& 3 \text{ edges} \times 3 \text{ degrees of freedom} \times 4 \text{ integration points} \\ &+ 1 \text{ element} \times 3 \text{ degrees of freedom} \times 7 \text{ integration points}. \end{aligned}$$

Then, since we define a scalar-valued finite element interpolant, f has only one component. We define, consequently a list consisting of kPi zeros (0 to select the first component for f, 1 to select the second component of f if it is vectorial, and so on) i.e.

$$\begin{aligned} j_k = \big\{&0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\\ &0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\\ &0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0\big\}. \end{aligned}$$

Afterwards, we define the list i_k allowing to select the considered basis functions, numbered from 0 to 8 for the basis functions associated with edges and from 9 to 11 for basis functions associated with the element. The basis functions associated with edges are repeated four times, since we use four integration points to compute integrals of type (4.6) and the basis functions associated with the element are repeated seven times, since we use seven integration points to compute integrals of type (4.7), which gives

$$\begin{aligned} i_k = \big\{&0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 1,\ 2,\ 2,\ 2,\ 2,\ 3,\ 3,\ 3,\ 3,\ 4,\ 4,\ 4,\ 4,\\ &5,\ 5,\ 5,\ 5,\ 6,\ 6,\ 6,\ 6,\ 7,\ 7,\ 7,\ 7,\ 8,\ 8,\ 8,\ 8,\ 9,\ 9,\ 9,\ 9,\ 9,\ 9,\ 9,\\ &10,\ 10,\ 10,\ 10,\ 10,\ 10,\ 10,\ 11,\ 11,\ 11,\ 11,\ 11,\ 11,\ 11,\ 11\big\}. \end{aligned}$$

Next, we define the integration points. We have in total $npPi = 19$ integration points numbered from 0 to 18 (four integrations points on each edge and seven in the element), which gives the list of indices p_k

$$\begin{aligned} p_k = \big\{&0,\ 1,\ 2,\ 3,\ 0,\ 1,\ 2,\ 3,\ 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 4,\ 5,\ 6,\\ &7,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 8,\ 9,\ 10,\ 11,\ 8,\ 9,\ 10,\ 11,\ 12,\ 13,\\ &14,\ 15,\ 16,\ 17,\ 18,\ 12,\ 13,\ 14,\ 15,\ 16,\ 17,\ 18,\ 12,\ 13,\ 14,\ 15,\ 16,\ 17,\ 18\big\}. \end{aligned}$$

If we consider a set of indices of integration points associated with an edge, for example the set $\{0,\ 1,\ 2,\ 3\}$, associated let us say to the edge E_1, this set of indices is repeated three times since we have three degrees of freedom on the edge E_1, and similarly for the other two edges. Similarly the set of indices of integration points on the element $\{12,\ 13,\ 14,\ 15,\ 16,\ 17,\ 18\}$ is repeated three times since we have three degrees of freedom on the element. The integration

points are chosen using the index p_k in the list P defined as

$$\begin{aligned}P = \big\{ & s_1A + (1-s_1)B,\ s_2A + b(1-s_2)B,\ s_3A + (1-s_3)B,\\ & s_4A + (1-s_4)B,\ s_1B + (1-s_1)C,\ s_2B + b(1-s_2)C,\\ & s_3B + (1-s_3)C,\ s_4B + (1-s_4)C,\ s_1C + (1-s_1)A,\\ & s_2C + b(1-s_2)A, s_3C + (1-s_3)A,\ s_4C + (1-s_4)A,\\ & (x_1,x_1),\ (x_2,x_2),\ (x_2,x_4),\ (x_4,x_2),\ (x_3,x_3),\ (x_3,x_5),\ (x_5,x_3)\big\}.\end{aligned}$$

The kPi associated weights are defined by

$$\begin{aligned}\alpha = \big\{ & b_1L_0\big(s_1A + (1-s_1)B\big),\ b_1L_0\big(s_2A + (1-s_2)B\big), b_2L_0\big(s_3A + (1-s_3)B\big),\\ & b_2L_0\big(s_4A + (1-s_4)B\big),\ b_1L_1\big(s_1A + (1-s_1)B\big),\ b_1L_1\big(s_2A + (1-s_2)B\big),\\ & b_2L_1\big(s_3A + (1-s_3)B\big),\ b_2L_1\big(s_4A + (1-s_4)B\big),\ b_1L_2\big(s_1A + (1-s_1)B\big),\\ & b_1L_2\big(s_2A + (1-s_2)B\big),\ b_2L_2\big(s_3A + (1-s_3)B\big),\ b_2L_2\big(s_4A + (1-s_4)B\big),\\ & b_1L_0\big(s_1B + (1-s_1)C\big),\ b_1L_0\big(s_2B + (1-s_2)C\big),\ b_2L_0\big(s_3B + (1-s_3)C\big),\\ & b_2L_0\big(s_4B + (1-s_4)C\big),\ b_1L_1\big(s_1B + (1-s_1)C\big),\ b_1L_1\big(s_2B + (1-s_2)C\big)\\ & b_2L_1\big(s_3B + (1-s_3)C\big),\ b_2L_1\big(s_4B + (1-s_4)C\big),\ b_1L_2\big(s_1B + (1-s_1)C\big),\\ & b_1L_2\big(s_2B + (1-s_2)C\big),\ b_2L_2\big(s_3B + (1-s_3)C\big),\ b_2L_2\big(s_4B + (1-s_4)C\big),\\ & b_1L_0\big(s_1C + (1-s_1)A\big),\ b_1L_0\big(s_2C + (1-s_2)A\big),\ b_2L_0\big(s_3C + (1-s_3)A\big),\\ & b_2L_0\big(s_4C + (1-s_4)A\big),\ b_1L_1\big(s_1C + (1-s_1)A\big),\ b_1L_1\big(s_2C + (1-s_2)A\big),\\ & b_2L_1\big(s_3C + (1-s_3)A\big),\ b_2L_1\big(s_4C + (1-s_4)A\big),\ b_1L_2\big(s_1C + (1-s_1)A\big),\\ & b_1L_2\big(s_2C + (1-s_2)A\big),\ b_2L_2\big(s_3C + (1-s_3)A\big),\ b_2L_2\big(s_4C + (1-s_4)A\big),\\ & c_1M_0^K(x_1,x_1),\ c_2M_0^K(x_2,x_2),\ c_2M_0^K(x_2,x_4),\ c_2M_0^K(x_4,x_2),\ c_3M_0^K(x_3,x_3),\\ & c_3M_0^K(x_3,x_5),\ c_3M_0^K(x_5,x_3), c_1M_1^K(x_1,x_1),\ c_2M_1^K(x_2,x_2),\ c_2xM_1^K(x_2,x_4),\\ & c_2M_1^K(x_4,x_2),\ c_3M_1^K(x_3,x_3),\ c_3M_1^K(x_3,x_5),\ c_3M_1^K(x_5,x_3),\ c_1M_2^K(x_1,x_1)\\ & c_2M_2^K(x_2,x_2),\ c_2xM_2^K(x_2,x_4),\ c_2M_2^K(x_4,x_2),\ c_3M_2^K(x_3,x_3), c_3M_2^K(x_3,x_5),\\ & c_3M_2^K(x_5,x_3)\big\}.\end{aligned}$$

4.4.6 Plotting of non conforming finite element

In FreeFEM, for the plot, each triangle is cut in n_{sub}^2 sub-triangles, where the parameter n_{sub} is defined in the finite element builder. The vectors are then plotted on the vertices of this cutting. For the finite element V_3 we choose $n_{sub} = 4$ since it involves polynomials $\mathbb{P}_4$.

4.4.7 Assessment of the implementation of the finite element V_3

Following the implementation of the finite element, this section is dedicated to an assessment of the results. In particular, the implementation of the basis functions and their derivatives is verified.

Assessment of the basis functions First, to assess the implementation of the local basis functions $\mathbf{\Phi}_i^K$, we compute in FreeFEM the terms

$$N_j^{E_\alpha}(\Phi_i^K) = \frac{1}{|E_\alpha|}\int_{E_\alpha} \Phi_i^K L_j^{E_\alpha},\ 0 \le i,j \le 2,\ \alpha = 1,2,3,$$
$$N_j^K(\Phi_i^K) = \frac{1}{|K|}\int_K \Phi_i^K M_j^K,\ j = 1,2,3.$$

In order to compute these integrals, the quadrature formulas `qforder` $= 7$ are employed for integrations over edges and `qforder` $= 6$ for integrations over elements. Consequently, the result is the identity matrix, $\mathbb{I}_{12}$ as expected, which serves to validate the computation of the basis functions.

Assessment of the derivatives Afterwards, we assess the implementation of the derivatives of the basis functions $\mathbf{\Phi}_i^K$. For this purpose, we use the space of polynomial of order 4 fully discontinuous (noted $\mathbb{P}_4^{dc}$), as reference. We call $\Pi_h^{P_4^{dc}}$ the interpolation operator in the space $\mathbb{P}_4^{dc}$. We apply the following procedure.

1. For each basis function $\Phi_i^K \in V_3$, we consider $\widehat{\Phi_i^K} = \Pi_h^{P_4^{dc}}(\Phi_i^K)$, its interpolate in the space $\mathbb{P}_4^{dc}$. Consequently, $\widehat{\Phi_i^K}$ is an element of the space $\mathbb{P}_4^{dc}$.

2. We compute

$$d_x\Phi_i^K = \frac{\partial \Phi_i^K}{\partial x}, \qquad d_y\Phi_i^K = \frac{\partial \Phi_i^K}{\partial y},$$
$$\widehat{d_x\Phi_i^K} = \frac{\partial \widehat{\Phi_i^K}}{\partial x}, \qquad \widehat{d_y\Phi_i^K} = \frac{\partial \widehat{\Phi_i^K}}{\partial y}.$$

   The quantities $\widehat{d_x\Phi_i^K}$ and $\widehat{d_y\Phi_i^K}$ are used as reference since the computation of the derivatives in the space $\mathbb{P}_4^{dc}$ is already validated.

3. Then we compute the following error

$$e_x = \|\widehat{d_x\Phi_i^K} - \Pi_h^{P_4^{dc}}(d_x\Phi_i^K)\|_{L^2(\Omega)},$$
$$e_y = \|\widehat{d_y\Phi_i^K} - \Pi_h^{P_4^{dc}}(d_y\Phi_i^K)\|_{L^2(\Omega)}.$$

4. We check that e_x and e_y are smaller than 1×10^{-10}.

As result, we get $e_x < 1 \times 10^{-10}$ and $e_y < 10^{-10}$ which allows to validate the implementation of the derivatives of the basis functions.

4.5 Numerical experiments

At this stage, the finite element spaces V_2 and V_3 (corresponding respectively to the cases $n = 1$ and $n = 2$), are available in FreeFEM. Therein, they are referenced respectively under the names

$P2pnc$[3] and $P3pnc$[4]. In this section, we present the convergence orders obtained with these new finite elements in order to verify the theoretical analysis. The obtained results are compared with those obtained with classical Taylor–Hood finite elements of accuracy order two and three.

For this purpose, we solve the Stokes equations (4.1) with the test case proposed in [132]. Let $\Omega = (0,1)^2$, $\nu = 1$ and $\boldsymbol{f}$ and $\boldsymbol{g}$ are chosen such that

$$\begin{aligned} \boldsymbol{u}(x,y) &= (\sin(x)\sin(y), \cos(x)\cos(y))^T, \\ p(x,y) &= 2\cos(x)\sin(y) - 2\sin(1)(1-\cos(1)), \end{aligned}$$

is solution to (4.1), which leads to

$$\begin{aligned} \boldsymbol{f}(x,y) &= (0, 4\cos(x)\cos(y))^T, \\ \boldsymbol{g} &= \boldsymbol{u}|_{\partial\Omega}. \end{aligned} .$$

All numerical tests are performed in FreeFEM, considering a regular grid composed of triangular elements. For the different pairs of finite elements tested, we compare the finite element approximations to the exact solutions for the velocity and the pressure.

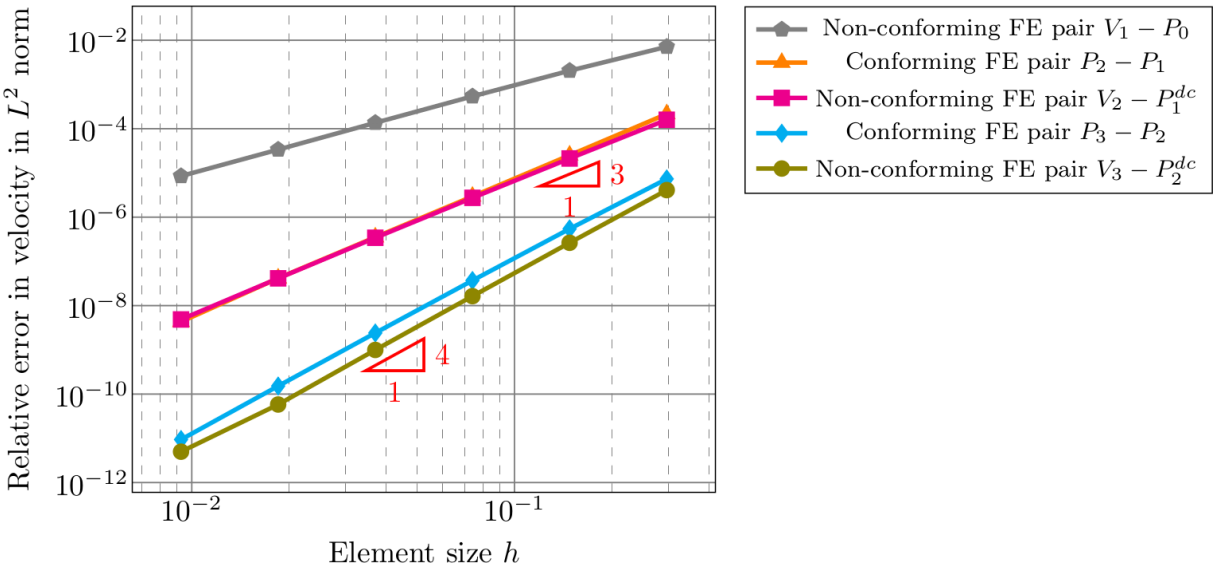


Figure 4.1: Error in norm L^2 between the computed velocity and the exact velocity for different pairs of finite elements.

[3] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P2pnc.cpp

[4] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P3pnc.cpp

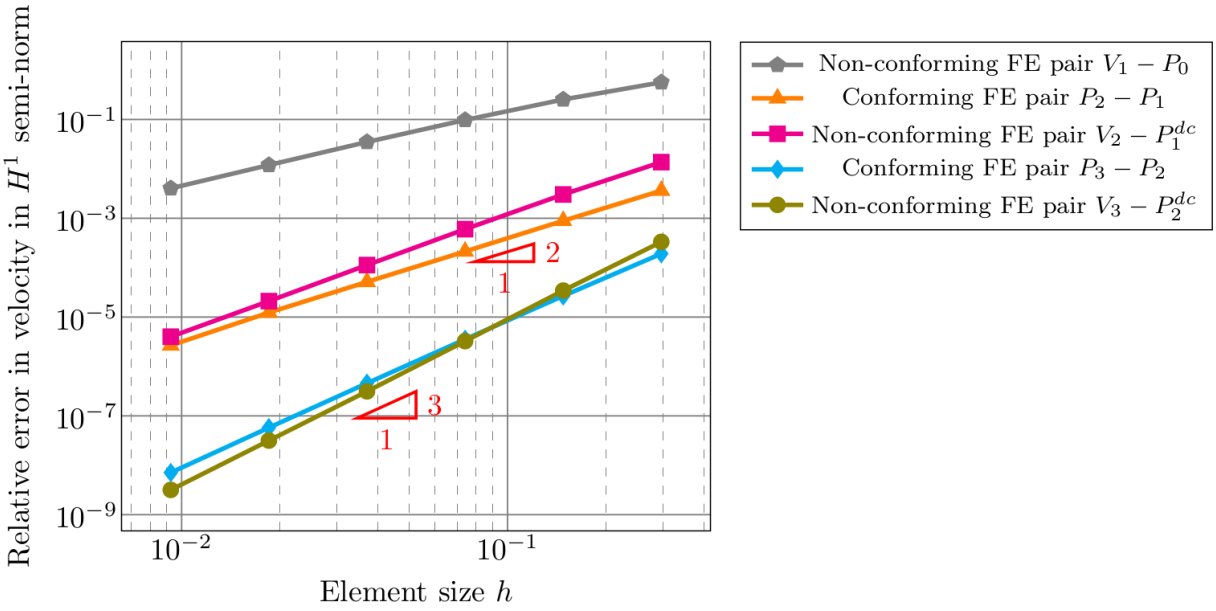


Figure 4.2: Error in norm H^1 between the computed velocity and the exact velocity for different pairs of finite elements.

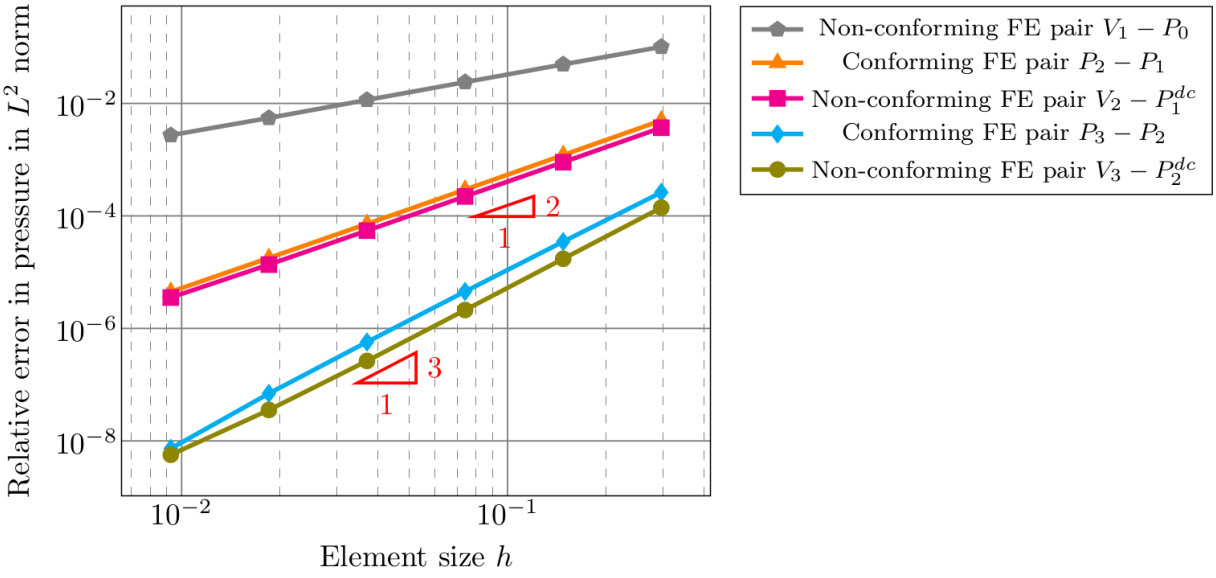


Figure 4.3: Error in norm L^2 between the computed pressure and the exact pressure for different pairs of finite elements.

It can be seen from Figures 4.1 to 4.3, that we obtain the expected orders of convergence for the non-conforming pairs V_2–$\mathbb{P}_1^{dc}$ and V_3–$\mathbb{P}_2^{dc}$. The results obtained are comparable with those obtained with conforming pairs of finite elements.

4.6 Conclusion

In this chapter, we have presented the family of non-conforming finite elements introduced in [132]. In particular, we have implemented in FreeFEM the finite element of order of accuracy three (corresponding to the case $n = 2$). We carried out numerical simulations to assess the

non-conforming finite element of accuracy order two and three. Finally, for the solution to the classical Stokes problem, the error obtained with the non-conforming pairs and with the conforming ones are comparable. However, we recall that we need these non-conforming finite elements to solve the local problems involved in MsFEM (see Chapter 8). In the next chapter, we extend this non-conforming finite element in three dimensions on tetrahedra.

Chapter 5

Inf-sup stable non-conforming finite elements on tetrahedra of accuracy order two and order three

Contents

Note: We would like to thank Prof. Frédéric Hecht for his help with the implementation of the finite elements in FreeFEM source code.

The content of this chapter has been published in [29].

[29] L. Balazi, G. Allaire, P. Jolivet, and P. Omnes. Inf-sup stable non-conforming finite elements on tetrahedra with second and third order accuracy. https://hal.science/ hal-04541809, April 2024.

This chapter is dedicated to the introduction of a family of scalar non-conforming finite elements of order two and three with respect to the H^1-norm on tetrahedra. This family of finite element is an extension in three dimensions of the non-conforming family of finite element introduced by [132] and presented in Chapter 4. As in the two-dimensional case, this family of finite element is useful for solving the discrete local problems in three dimensions arising from the Multi-scale Finite Element Method (see Chapter 8). More precisely, for the class of MsFEM developed (see Chapters 7 and 12), these local problems involve polynomial divergence and Lagrange multipliers, and therefore cannot be solved using classical pairs of finite elements. In two dimensions, the non-conforming finite elements introduced by [132] associated with discontinuous piecewise polynomial pressures of order n allow to solve these local problems. However, in three dimensions, no existing finite element pairs led to a discrete problem whose well-posedness could be proved, except the ones developed here. For numerical purposes, the two non-conforming finite element, of order two and three, have been implemented in FreeFEM (following the same procedure as described in Chapter 4). They are known respectively under the name $P2pnc3d$[1] and $P3pnc3d$[2].

In this chapter, for the sake of simplicity, we present the new family of scalar non-conforming finite elements in the context of the Stokes problem. Indeed, their vector-valued version generates, together with a discontinuous pressure approximation of order one and two respectively, an inf-sup stable finite element pair of convergence order two and three for the Stokes problem in energy norm.

5.1 Introduction

We consider the Stokes problem, which is a simple model for viscous incompressible flows and a first step towards more complex problems such as the Oseen problem or the Navier–Stokes problem. Let $\Omega \subset \mathbb{R}^3$ be a bounded connected domain. The steady-state Stokes problem is to find the velocity $\boldsymbol{u} : \Omega \to \mathbb{R}^3$ and the pressure $p : \Omega \to \mathbb{R}$ solution to

$$\begin{aligned} -\nu\Delta\boldsymbol{u} + \boldsymbol{\nabla} p &= \boldsymbol{f} \quad \text{in } \Omega, \\ \operatorname{div} \boldsymbol{u} &= 0 \quad \text{in } \Omega, \\ \boldsymbol{u} &= \boldsymbol{0} \quad \text{on } \partial\Omega, \end{aligned} \tag{5.1}$$

with $\nu > 0$ the kinematic viscosity and $\boldsymbol{f}$ the load. Although, throughout this paper, we consider homogeneous boundary conditions, the generalization of the study for non-homogeneous Dirichlet boundary conditions is straightforward.

[1] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P2pnc_3d.cpp

[2] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P3pnc_3d.cpp

We denote by

$$(u, v) = \int_\Omega uv \mathrm{d}x$$

the scalar product in $L^2(\Omega)$. We use the same notation for vector-valued functions (which are denoted using boldface letters), i.e. the $L^2(\Omega)$ scalar product for vector-valued functions is given by

$$(\boldsymbol{u}, \boldsymbol{v}) = \int_\Omega \boldsymbol{u} \cdot \boldsymbol{v} \, \mathrm{d}x,$$

where $\boldsymbol{u} \cdot \boldsymbol{v}$ denotes the Euclidean scalar product in $\mathbb{R}^3$. Then, we define the classical velocity space $V = \left(H_0^1(\Omega)\right)^3$ and pressure space $M = L_0^2(\Omega) = \{q \in L^2(\Omega), \ s.t. \ \int_\Omega q = 0\}$ equipped respectively with the $|\cdot|_1$ H^1 semi-norm and the $\|\cdot\|_0$ L^2 norm. We define the bi-linear forms $a(\cdot, \cdot) : V \times V \to \mathbb{R}$ and $b(\cdot, \cdot) : M \times V \to \mathbb{R}$ by

$$a(\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{\nabla u}, \boldsymbol{\nabla v}), \qquad b(p, \boldsymbol{v}) = (p, \operatorname{div} \boldsymbol{v}).$$

Then, a weak formulation of the Stokes problem (5.1) reads as follows. Find $\boldsymbol{u} \in V$ and $p \in M$ such that

$$\left\{ \begin{array}{rcll} a(\boldsymbol{u}, \boldsymbol{v}) + b(p, \boldsymbol{v}) & = & (\boldsymbol{f}, \boldsymbol{v}) & \forall \boldsymbol{v} \in V, \\ b(q, \boldsymbol{u}) & = & 0 & \forall q \in M, \end{array} \right. \tag{5.2}$$

with $\boldsymbol{f} \in L^2(\Omega)^3$. It is well-known [84] that system (5.2) has a unique solution due to the Ladyzhenskaya–Babuška–Brezzi (LBB) condition, also called inf-sup condition, namely there exists $\beta > 0$ such that

$$\inf_{q \in M} \sup_{\boldsymbol{v} \in V} \frac{b(q, \boldsymbol{v})}{|\boldsymbol{v}|_1 \|q\|_0} \geq \beta.$$

Then, when the Stokes problem is discretized using finite elements, it is important that the discrete inf-sup condition is satisfied uniformly in the mesh size h. This condition imposes constraints on the choice of finite element pairs for velocity and pressure, as not every pair satisfies this requirement.

Additionally, to our own interest, the design of inf-sup stable pairs for the Stokes equations, especially in three dimensions, is essential in many applications involving fluid mechanics (environmental flow, biological flow, in the energy sector, and others). A first class of such inf-sup stable methods is given by the conforming velocity/pressure pair $\mathbb{P}_{n+1} - \mathbb{P}_n$ (where $\mathbb{P}_{n+1}$ and $\mathbb{P}_n$ are respectively the spaces of polynomials of order less or equal to $n+1$ and n), called the Taylor–Hood method. Indeed, the classical Taylor–Hood method, obtained with $n = 1$ for tetrahedra in three dimensions, was first studied by R. Stenberg [158] who showed that it is sufficient that any element of the tetrahedral mesh has at least one vertex in the interior of Ω for this pair to be inf-sup stable for the Stokes problem. Then, D. Boffi [34] extended the Taylor–Hood method for higher-order polynomials for tetrahedra and proved that these pairs of finite elements satisfy the inf-sup condition for $n \geq 1$ under mesh restrictions less stringent than those in [158].

Another appealing approach to build inf-sup stable pairs for the Stokes problem is to consider non-conforming approximations. This class of approximations may have some advantages such as fewer geometric constraints on the mesh and in some cases fewer degrees of freedom and far fewer non-zero matrix entries than the corresponding conforming discretisations [131]. It also

allows to use the inf-sup stable lowest order approximation ($\mathbb{P}_1$-$\mathbb{P}_0$) which is not possible with conforming elements. Thus, in [61], families of non-conforming finite elements are introduced: one of accuracy order one (the Crouzeix–Raviart finite element) on triangles and tetrahedra, and one of accuracy order three on triangles in two dimensions, which are inf-sup stable for the Stokes problem. Then, non-conforming finite elements have been developed for quadrilaterals and hexahedra. Indeed, a constructive method for deriving finite elements of nodal type for rectangles and rectangle parallelepipeds was developed in [96]. More recently, [131] developed families of scalar non-conforming finite elements of arbitrary orders with optimal approximation properties on parallelograms and parallelepipeds. In two dimensions, [132] introduced a family of scalar non-conforming finite elements of accuracy order $n+1$ ($n \geq 0$) with respect to the H^1-norm on triangles, by enriching the local space of polynomials of order $n+1$ with a proper subspace of polynomials of order $n+2$. The authors of [132] showed that their vector-valued versions associated with discontinuous pressure approximations of polynomial order n form inf-sup stable finite element pairs of accuracy order $n+1$ for the Stokes problem. For $n = 0$, their non-conforming finite element recovers the well-known Crouzeix–Raviart element. However, the generalization of these non-conforming finite elements to three dimensions on tetrahedra has not been studied. Based on the general definition of non-conforming finite elements in [61], the authors of [155] introduce a specific example of such a family on tetrahedra with $\mathbb{P}_2$ velocities. This finite element associated with a discontinuous pressure of polynomial order one, form an inf-sup stable finite element pair with quadratic accuracy for the Stokes problem.

More recently, two families of scalar non-conforming finite elements of accuracy order two and three with respect to the H^1-norm on tetrahedra were introduced in [50]. The authors showed that their vector-valued versions, associated with a discontinuous pressure approximation of polynomial order one and two, respectively, form inf-sup stable finite element pairs with accuracy order two and three for the Stokes problem.

In this chapter, inspired by the work of [132], we develop a new non-conforming family of finite elements of order two (quadratic velocity approximation error) and three (cubic velocity approximation error). The main difference with [50] is that we consider moments of order $n-1$ in the element while in [50] only moments of order zero in the element are considered. For the case $n = 1$, the two methods are equivalent.

The content of this chapter is the following. Section 5.2 recalls the discretisation setting for the Stokes problem. Section 5.3 is concerned with the definition of a family of scalar non-conforming finite elements, that are used to approximate the velocity space. Explicit bases of order two and three are provided. Section 5.4 is devoted to the study of the discrete inf-sup condition and the approximation properties of the finite element pair for the Stokes problem defined by the previous family of non-conforming finite elements for velocity and a piecewise discontinuous polynomial pressure of order n. To complete this study, we present an efficient preconditioning strategy for the resulting linear system in Section 5.5. In Section 5.6, we carry out numerical experiments allowing to assess the new non-conforming pairs of finite elements and we compare them with conforming pairs. Finally, Section 5.7 gives a comparison of the non-conforming and conforming pairs of finite elements in terms of degrees of freedom and computational times.

5.2 Discrete setting

Let $\Omega \subset \mathbb{R}^3$ be a polyhedral domain. We note $\{\mathcal{T}_h\}$ a family of triangulations of Ω parameterized by a positive parameter h which tends to 0. Each triangulation $\mathcal{T}_h$ consists of a finite number of tetrahedra K such that $\overline{\Omega} = \cup_{K\in\mathcal{T}_h}\overline{K}$. Let $h_K = \mathrm{diam}(K)$ and $h := \max\ h_k$. We assume that the triangulations are conformal in the sense that the intersection of the closures of two different cells K is either empty, a common vertex, a common edge, or a common face. Besides, we assume that the triangulations are shape regular, i.e. there exists a positive constant C independent of the mesh parameter h such that

$$\frac{h_K}{\sigma_K} \leq C, \qquad \forall K \in \mathcal{T}_h,\ h_K > 0,$$

where σ_K denotes the maximum diameter of a ball which can be inscribed in $\overline{K}$.

We denote by $\mathcal{F}_h$ the set of all faces F of $\mathcal{T}_h$. The set of inner faces will be denoted by $\mathcal{F}_h^i$ and the set of boundary faces by $\mathcal{F}_h^b$. The area of F and the volume of K are denoted respectively by $|F|$ and $|K|$. We note $\boldsymbol{n}_K$ the outer unit normal vector on ∂K. For any face F, we note $\boldsymbol{n}_F$ a fixed unit normal vector to F. If $F \in \mathcal{F}_h^b$, then $\boldsymbol{n}_F$ coincides with the outer normal vector to $\partial\Omega$.

For any integer n and any integer $1 \leq l \leq d$ (with d the space dimension, here $d = 3$), we denote by $\mathbb{P}_n^l$ the linear space spanned by l-variate polynomial functions of total degree at most n. The dimension of $\mathbb{P}_n^l$ is

$$N_n^l := \dim(\mathbb{P}_n^l) = \binom{n+l}{n}.$$

For any $K \in \mathcal{T}_h$, we denote by $\mathbb{P}_n^3(K)$ the restriction to K of polynomials in $\mathbb{P}_n^3$. For any $F \in \mathcal{F}_h$, we denote by $\mathbb{P}_n^2(F)$ the restriction to F of $\mathbb{P}_n^2$. For the sake of simplicity, we denote $\mathbb{P}_n^3(K)$ and $\mathbb{P}_n^2(F)$ respectively by $\mathbb{P}_n(K)$ and $\mathbb{P}_n(F)$.

We define the H^1 broken space by

$$H^{1,h}(\Omega) := \{v \in L^2(\Omega),\ s.t.\ v|_K \in H^1(K),\ \forall K \in \mathcal{T}_h\}.$$

For $v \in H^{1,h}(\Omega)$, we define the jump across a face $F \in \mathcal{F}_h^i$ by

$$[v]_F = (v|_K)|_F - (v|_{\tilde{K}})|_F,$$

where K and $\tilde{K}$ are the two cells which are adjacent to the face F oriented such that the unit normal vector $\boldsymbol{n}_F$ points into $\tilde{K}$. For a domain $D \subset \mathbb{R}^3$, we use the Sobolev spaces $H^m(D)$, $H_0^m(D)$, and $L^2(D) = H^0(D)$ for non-negative integer m. The norms and semi-norms in the scalar and vector-valued versions of $H^m(D)$ are denoted by $\|\cdot\|_{m,D}$ and $|\cdot|_{m,D}$ respectively. For any $\boldsymbol{v}$ in the broken space $\left(H^{1,h}(\Omega)\right)^3$, an analogue of its semi-norm $|\cdot|_1$ is defined, by

$$|\boldsymbol{v}|_{1,h} := \left(\sum_{K\in\mathcal{T}_h} |\boldsymbol{v}|_{1,K}^2\right)^{1/2},$$

where $|\boldsymbol{v}|_{1,K} = \|\boldsymbol{\nabla v}\|_{0,K}$. Let $\boldsymbol{\nabla_h v}$ and $\mathrm{div}_h\, \boldsymbol{v}$ denote the piecewise gradient and piecewise divergence of $\boldsymbol{v} \in H^{1,h}(\Omega)^3$, defined by

$$(\boldsymbol{\nabla_h}\ \boldsymbol{v})|_K = \boldsymbol{\nabla}(\boldsymbol{v}|_K), \qquad (\,\mathrm{div}_h\, \boldsymbol{v})|_K = \,\mathrm{div}\,(\boldsymbol{v}|_K).$$

Let $V_h \subset L^2(\Omega)^3$ be a non-conforming finite element space and $M_h \subset M$. Let $\boldsymbol{f}$ be in $L^2(\Omega)^3$. Then, the discrete Stokes problem reads as follow. Find $\boldsymbol{u_h} \in V_h$ and $p_h \in M_h$ such that

$$\left\{ \begin{array}{rcll} a_h(\boldsymbol{u_h}, \boldsymbol{v_h}) + b_h(p_h, \boldsymbol{v_h}) & = & (\boldsymbol{f}, \boldsymbol{v_h}) & \forall \boldsymbol{v_h} \in V_h, \\ b_h(q_h, \boldsymbol{u_h}) & = & 0 & \forall q_h \in M_h, \end{array} \right. \tag{5.3}$$

where $a_h(\cdot,\cdot) : V_h \times V_h \to \mathbb{R}$ and $b_h(\cdot,\cdot) : M_h \times V_h \to \mathbb{R}$ are respectively the discrete versions of the bi-linear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$, defined by

$$a_h(\boldsymbol{u_h}, \boldsymbol{v_h}) = (\boldsymbol{\nabla_h u_h}, \boldsymbol{\nabla_h v_h}), \qquad b_h(p_h, \boldsymbol{v_h}) = (p_h,\, \mathrm{div}_h\, \boldsymbol{v_h}). \tag{5.4}$$

It is well-known that the discrete problem (5.4) is well-posed under the following conditions: $a_h(\cdot,\cdot)$ and $b_h(\cdot,\cdot)$ are continuous, $a_h(\cdot,\cdot)$ is symmetric and coercive and the spaces V_h and M_h satisfy the discrete Ladyzhenskaya–Babuška–Brezzi (LBB) condition (or the discrete inf-sup condition, or are said to be inf-sup stable), namely there exists $\beta^\star > 0$, independent of h, such that

$$\inf_{q_h \in M_h} \sup_{\boldsymbol{v_h} \in V_h} \frac{b_h(q_h, \boldsymbol{v_h})}{|\boldsymbol{v_h}|_1 \|q_h\|_0} \geq \beta^\star.$$

5.3 A family of scalar non-conforming finite elements

In this section, we introduce a family of scalar non-conforming finite elements and a set of degrees of freedom.

5.3.1 Definition of the degrees of freedom

In the vein of [132], our aim is to define a scalar non-conforming finite element space of accuracy order $n+1$, $n \geq 0$ such that its vector-valued version $V_{h,n+1}$ is rich enough to satisfy the inf-sup condition when associated with the pressure approximation space $M_{h,n}$ made up of globally discontinuous piecewise $\mathbb{P}_n(K)$ pressures, with vanishing mean-value on Ω. This global pressure approximation space is also denoted $\mathbb{P}_n^{\mathrm{dc}}$ in what follows. The scalar (local) degrees of freedom are defined for any $v \in H^1(K)$ as

$$N_j^{F_\alpha}(v) = \int_{F_\alpha} v L_j^{F_\alpha} \qquad j \geq 0,\ \alpha = 1,2,3,4, \tag{5.5}$$

$$N_j^{K}(v) = \int_{K} v M_j^{K} \qquad j \geq 0, \tag{5.6}$$

where $(L_j^{F_\alpha})_j$ define an arbitrary basis of $\mathbb{P}_n(F_\alpha)$ of dimension $(n+1)(n+2)/2$, and $(M_j^K)_j$ define an arbitrary basis of $\mathbb{P}_{n-1}(K)$ of dimension $n(n+1)(n+2)/6$. Thus, for each element $K \in \mathcal{T}_h$

and any integer $n \geq 0$, we define the set of degrees of freedom $\mathcal{N}_{n+1}(K)$ by

$$\begin{aligned}\mathcal{N}_{n+1}(K) := &\left\{ N_j^{F_\alpha} \qquad 1 \leq j \leq \frac{(n+1)(n+2)}{2},\ \alpha = 1,2,3,4 \right\} \\ &\cup \left\{ N_j^{K} \qquad 1 \leq j \leq \frac{n(n+1)(n+2)}{6} \right\}.\end{aligned} \tag{5.7}$$

Remark 5.1. In the $n = 0$ case, (5.5) reduces to the integral of v on each of the 4 faces of the tetrahedron while (5.6) completely disappears. Since for $\mathbb{P}_1$ functions, the integral over a face is simply the area of the face multiplied by the value at the barycenter of that face, the degrees of freedom in the $n = 0$ case can be assimilated to those of the standard $\mathbb{P}_1$ non-conforming (Crouzeix–Raviart) finite element space. Then, the vector version of this finite element space associated with the $\mathbb{P}_0$ pressure space is known to satisfy the discrete inf-sup property and to provide a first-order accurate approximation of the velocity in the H^1 norm and of the pressure.

5.3.2 Definition of the finite element basis

Following [132], for $n \geq 1$, the idea is to enrich the local space $\mathbb{P}_{n+1}(K)$ with a proper subspace of $\mathbb{P}_{n+2}(K)$, denoted by $\Sigma_{n+2}(K) \subset \mathbb{P}_{n+2}(K)$, such that $\Sigma_{n+2}(K)$ has a trivial intersection with $\mathbb{P}_{n+1}(K)$. Thus, the enriched space $V_{n+1}(K)$ will be given by

$$V_{n+1}(K) = \mathbb{P}_{n+1}(K) \oplus \Sigma_{n+2}(K).$$

The first requirement is to ensure that

$$\dim(V_{n+1}) = \operatorname{card}(\mathcal{N}_{n+1}).$$

Given that

$$\dim(\mathbb{P}_{n+1}) = \frac{(n+2)(n+3)(n+4)}{6},$$

we should have

$$\begin{aligned}\dim(\Sigma_{n+2}) &= \dim(V_{n+1}) - \dim(\mathbb{P}_{n+1}) \\ &= \frac{n(n+1)(n+2)}{6} + 4\frac{(n+1)(n+2)}{2} - \frac{(n+2)(n+3)(n+4)}{6} \\ &= n(n+2),\end{aligned}$$

and we have to ensure that

$$\mathbb{P}_{n+1}(K) \cap \Sigma_{n+2}(K) = \{0\}.$$

Consequently, to build the space V_{n+1}, i.e. to complete the space $\mathbb{P}_{n+1}(K)$, we are looking for a proper subspace $\Sigma_{n+2}(K)$ of $\mathbb{P}_{n+2}$ of dimension $n(n+2)$.

Remark 5.2. By noting that

$$\dim(\mathbb{P}_{n+2} \setminus \mathbb{P}_{n+1}) = \frac{(n+3)(n+4)}{2},$$

we remark that, for $n > 5$,

$$\frac{(n+3)(n+4)}{2} - n(n+2) < 0.$$

Therefore, the proposed approach would work in theory at most for $n \leq 5$. However, in standard applications and for our purposes, considering accuracy order up to three is enough.

Now, we recall the notion of unisolvence. A finite element is defined by a set $(K, V(K), \mathcal{N}(K))$ where $V(K)$ is a space of functions of finite dimension M, and $\mathcal{N}(K)$ is a space of M independent linear forms $(\psi_i)_{i=1,\dots,M}$. The *unisolvence* property reads as follows.

Definition 5.3. The set $(K,\ V(K),\ \mathcal{N}(K))$ is said to be unisolvent if and only if the application

$$\begin{aligned} V(K) &\to \mathbb{R}^M \\ v &\mapsto (\psi_1(v), \dots, \psi_M(v)) \end{aligned}$$

is an isomorphism.

We can state the following result.

Lemma 5.4. Let K be a tetrahedron with barycentric coordinates λ_1, λ_2, λ_3, λ_4. Then, the finite element spaces

$$V_2 = \mathbb{P}_2 + \text{span}\{\lambda_1\lambda_2^2,\ \lambda_1\lambda_3^2,\ \lambda_2\lambda_3^2\},$$

and

$$V_3 = \mathbb{P}_3 + \text{span}\{\lambda_1^3\lambda_2,\ \lambda_2^3\lambda_3,\ \lambda_3^3\lambda_4,\ \lambda_4^3\lambda_1,\ \lambda_2^3\lambda_1,\ \lambda_1^3\lambda_4,\ \lambda_4^3\lambda_3,\ \lambda_3^3\lambda_2\},$$

are unisolvent with respect to the sets of degrees of freedom $\mathcal{N}_2$ and $\mathcal{N}_3$, respectively, described in (5.7).

Proof. We propose a numerical proof of the unisolvence of the finite element spaces V_2 and V_3 using Definition 5.3. On the reference tetrahedron and with the basis of $\mathbb{P}_{n+1}$ given by Lemma A.1 and the additional basis functions given in Lemma 5.4, we build $(\Phi_j)_j$ a basis of V_{n+1}. Then, we assemble the square matrix $\mathcal{M}_{i,j} = N_i(\Phi_j)$ for $i = 1, \dots, \text{card}(\mathcal{N}_{n+1})$ and $j = 1, \dots, \dim(V_{n+1})$, with $N_i \in \mathcal{N}_{n+1}$. It should be noted that the two resulting matrices are of small size (13×13 for V_2 and 28×28 for V_3). Since the matrices contain small coefficients, the value of their determinant can be non relevant (near to zero) to check for inversibility. Instead, we compute the conditioning of these matrices (in this paper, we use the NumPy linear algebra functions [3]). In the end, we find that the conditioning of matrix $\mathcal{M}$ has a value around 188 for the case $n = 1$ and around 2307 for the case $n = 2$, which are clearly different from infinite and allows to conclude on the inversibility of $\mathcal{M}$. It should be noted that a rigorous proof could be done using interval arithmetic, with software such as Xcas [144]. For the unisolvence of the finite element space V_2, a proof can be found in [50]. □

In Lemma 5.4, we have proposed a basis for V_2 and one for V_3. In Appendix A, we explain how these bases have been found and how to find additional bases.

Remark 5.5. For the case $n = 1$, we recover the finite element proposed by [50]. However, for the case $n = 2$, the proposed finite element space differs from that proposed in [50], since in their work the authors consider only moments of order zero in the element, while in our work, we consider moments of order $n - 1$ in the element (i.e. moment of order one for the case $n = 2$).

Remark 5.6. The finite element spaces V_2 and V_3 described in Lemma 5.4 are not unique. We propose a numerical strategy in Appendix A.1 to find relevant bases of functions. It has to be noted that none of these bases are symmetric, that is to say, the finite element space depends on the face numbering. However, we can suppose that in non-structured meshes, the non-symmetry of the global finite element space will have little influence.

Summarizing, for any integer $n \geq 0$, the finite element we consider is given by the set $(K,\ V_{n+1}(K),\ \mathcal{N}_{n+1}(K))$ provided that $V_{n+1}(K)$ is unisolvent with respect to the set $\mathcal{N}_{n+1}(K)$ (which is known for $n = 0$ and which we have proved for $n = 1, 2$). Then, we define the global, vector-valued finite element space $V_{h,n+1}$ by

$$V_{h,n+1} = \left\{ \begin{array}{c} \boldsymbol{v} \in \left(H^{1,h}(\Omega)\right)^3 \ :\ \boldsymbol{v}|_K \in \left(V_{n+1}(K)\right)^3,\ \forall K \in \mathcal{T}_h \\ \displaystyle\int_F q[\boldsymbol{v}]_F = 0,\ \forall F \in \mathcal{F}_h^i,\ q \in \mathbb{P}_n(F) \\ \displaystyle\int_F q\boldsymbol{v} = 0,\ \forall E \in \mathcal{F}_h^b,\ q \in \mathbb{P}_n(F) \end{array} \right\}. \tag{5.8}$$

5.4 Approximation properties and the discrete inf-sup condition

In this section, the order n is fixed. We show that the non-conforming finite element space $V_{h,n+1}$ (under Assumption 5.7) together with the $\mathbb{P}_n^{\mathrm{dc}}$ space for the pressure fulfil the discrete inf-sup condition and have an accuracy of order $n + 1$.

Assumption 5.7. In this section, we assume that such a space $V_{n+1}(K)$, unisolvent with respect to the set of degrees of freedom $\mathcal{N}_{n+1}(K)$, exists for $n \leq 5$.

In what follows, we introduce the local (scalar-valued) interpolation operators. We define the finite element interpolant as

$$I_h^K : H^1(K) \to V_{n+1}(K) \qquad J_h^K : L^2(K) \to \mathbb{P}_n(K).$$

Using the set of degrees of freedom from $\mathcal{N}_{n+1}(K)$, we define $I_h^K v$ for $v \in H^1(K)$ by

$$N_i(I_h^K v) = N_i(v), \quad \forall N_i \in \mathcal{N}_{n+1}(K). \tag{5.9}$$

Then, $I_h^K v$ can be written as

$$I_h^K v = \sum_{j=1}^{\mathrm{card}(\mathcal{N}_{n+1})} N_j(v)\varphi_j,$$

where $\{\varphi_j\}$ is the dual basis for $\mathcal{N}_{n+1}(K)$ i.e. $N_i(\varphi_j) = \delta_{ij}\ \forall(i,j)$. Next, we define $J_h^K p \in \mathbb{P}_n(K)$ for $p \in L^2(K)$ by the $L^2(K)$-projection in $\mathbb{P}_n(K)$ as

$$(J_h^K p, q)_K = (p, q)_K, \quad \forall q \in \mathbb{P}_n(K).$$

It can be seen that the interpolation operator I_h^K preserves $\mathbb{P}_{n+1}(K)$. Indeed, for $p \in V_{n+1}$, p and $I_h^K p$ have the same degrees of freedom and then by unisolvence $I_h^K p = p$. In particular, this holds for $\mathbb{P}_{n+1}$ which is included in V_{n+1}. From its definition, J_h^K preserves $\mathbb{P}_n(K)$ evidently.

Moreover, it can be seen from the definitions of I_h^K and its vector-valued version $\boldsymbol{I}_h^K$ and the definition of $V_{h,n+1}$ in (5.8) on the one hand, and from the definition of J_h^K on the other hand, that these local interpolants generate global interpolation operators $\boldsymbol{I}_h : V \to V_{h,n+1}$ and $J_h : M \to M_{h,n} := \mathbb{P}_n^{\mathrm{dc}} \bigcap L_0^2(\Omega)$.

5.4.1 The discrete inf-sup condition

In this section, we show that the discrete version of the inf-sup (LBB) condition, that we recall below, is valid uniformly in h, for the finite element pair $V_{h,n+1} - M_{h,n}$, i.e.

$$\exists \beta_n^\star > 0 \text{ such that } \forall h \qquad \inf_{q_h \in M_{h,n}} \sup_{\boldsymbol{v_h} \in V_{h,n+1}} \frac{b_h(q_h, \boldsymbol{v_h})}{|\boldsymbol{v_h}|_{1,h} \|q_h\|_0} \geq \beta_n^\star, \tag{5.10}$$

where the bi-linear form $b_h(\cdot,\cdot)$ is defined in (5.4).

Theorem 5.8. The finite element pair $V_{h,n+1} - M_{h,n}$ satisfies the discrete inf-sup stability condition (5.10).

In order to prove Theorem 5.8, we recall Fortin's lemma [80].

Lemma 5.9 (Fortin's lemma)**.** If the continuous inf-sup condition holds with the constant β and if there exists a linear operator $\Pi_h : V \to V_{h,n+1}$ such that

$$b_h(q_h, \Pi_h \boldsymbol{u}) = b(q_h, \boldsymbol{u}),\ \forall q_h \in M_{h,n}, \forall \boldsymbol{u} \in V,$$
$$|\Pi_h \boldsymbol{u}|_{1,h} \leq C|\boldsymbol{u}|_1, \forall \boldsymbol{u} \in V,$$

then the discrete inf-sup condition holds with $\beta_n^\star \geq \beta\ |||\Pi_h|||^{-1}$, where $|||\cdot|||$ denotes the operator norm.

Below, we prove Theorem 5.8.

Proof. Let $\boldsymbol{I}_h$ be the interpolant defined above, then we remark that for any $q_h \in M_h$ and for

any $\boldsymbol{v} \in V$, we have

$$\begin{aligned} b_h(q_h, \boldsymbol{I}_h\boldsymbol{v}) &= (\operatorname{div}_h \boldsymbol{I}_h\boldsymbol{v}, q_h) \\ &= \sum_{K\in\mathcal{T}_h} (\operatorname{div}_h \boldsymbol{I}_h^K\boldsymbol{v}, q_h)_K \\ &= \sum_{K\in\mathcal{T}_h} \left(-(\boldsymbol{I}_h^K\boldsymbol{v}, \boldsymbol{\nabla} q_h)_K + (\boldsymbol{I}_h\boldsymbol{v}\cdot\boldsymbol{n}_K, q_h)_{\partial K}\right), \end{aligned}$$

by integration by parts. Now, since $\boldsymbol{\nabla} q_h|_K \in (\mathbb{P}_{n-1}(K))^3$ and from the definition of the degrees of freedom (5.6) and the interpolation properties (5.9) we get $(\boldsymbol{I}_h^K\boldsymbol{v}, \boldsymbol{\nabla} q_h)_K = (\boldsymbol{v}, \boldsymbol{\nabla} q_h)_K$. In the same way, since $q_h|_{F_\alpha} \in \mathbb{P}_n(K)$ and from the definition of the degrees of freedom (5.5) and the interpolation properties (5.9), we get that $(\boldsymbol{I}_h\boldsymbol{v}\cdot\boldsymbol{n}_K, q_h)_{\partial K} = (\boldsymbol{v}\cdot\boldsymbol{n}_K, q_h)_{\partial K}$. It follows that

$$\begin{aligned} b_h(q_h, \boldsymbol{I}_h\boldsymbol{v}) &= \sum_{K\in\mathcal{T}_h} (-(\boldsymbol{v}, \boldsymbol{\nabla} q_h)_K + (\boldsymbol{v}\cdot\boldsymbol{n}_K, q_h)_{\partial K}) \\ &= \sum_{K\in\mathcal{T}_h} (\operatorname{div}\boldsymbol{v}, q_h)_K \\ &= (\operatorname{div}\boldsymbol{v}, q_h) = b(q_h, \boldsymbol{v}). \end{aligned}$$

Besides, from Lemma 5.10 (see Section 5.4.2) with $k = 0$, it follows that

$$|\boldsymbol{I}_h\boldsymbol{u}|_{1,h} \le C|\boldsymbol{u}|_1, \quad \forall \boldsymbol{u} \in V,$$

with a constant C that does not depend on the mesh. Finally, we conclude thanks to Fortin's lemma, that the pair of finite elements $V_{h,n+1} - M_{h,n}$ satisfies the discrete inf-sup condition for the Stokes problem. □

5.4.2 Approximation property and error analysis

In this section, we derive an error estimate for the new non-conforming family of finite elements. We follow the same procedure as developed in [132]. First, we introduce Lemma 5.10.

Lemma 5.10. The local interpolation operators $I_h^K : H^1(K) \to V_{n+1}(K)$ and $J_h^K : L^2(K) \to \mathbb{P}_n(K)$ satisfy for all $K \in \mathcal{T}_h$:

$$|u - I_h^K u|_{1,K} \le Ch^k |u|_{k+1,K} \qquad \forall u \in H^{k+1}(K),\ k = 0, \dots, n+1, \tag{5.11}$$

$$\|p - J_h^K p\|_{0,K} \le Ch^{n+1} |p|_{n+1,K} \qquad \forall p \in H^{n+1}(K). \tag{5.12}$$

with C a constant that does not depend on the mesh.

Proof. First, let K be an arbitrary element and $\widehat{K}$ the reference element and let us denote by

$$\begin{array}{rccl} \pi_K : & \widehat{K} & \longrightarrow & K \\ & \widehat{x} & \longmapsto & x = \pi_K(\widehat{x}) = \mathbb{A}_K\widehat{x} + b_K, \end{array}$$

the affine invertible mapping such that $K = \pi_K(\widehat{K})$. Using the standard scaling properties of the reference transformation and its inverse [76, chapter 11.1 - 11.2], it is sufficient to show the following estimates on the reference tetrahedron $\widehat{K}$

$$|\widehat{u} - I_h^{\widehat{K}}\widehat{u}|_{1,\widehat{K}} \leq \widehat{C}|\hat{u}|_{k+1,\widehat{K}}, \qquad \forall \widehat{u} \in H^{k+1}(\widehat{K}), \qquad k = 0, \ldots, n+1,$$

$$\|\widehat{p} - J_h^{\widehat{K}}\widehat{p}\|_{0,\widehat{K}} \leq \widehat{C}|\widehat{p}|_{n+1,\widehat{K}}, \qquad \forall \widehat{p} \in H^{n+1}(\widehat{K}),$$

where $I_h^{\widehat{K}}$ and $J_h^{\widehat{K}}$ denote the corresponding interpolation operators on the reference tetrahedron. In order to apply the Bramble–Hilbert Lemma [76, Lemma 11.9], we first need to show the continuity of the interpolation operators. From the continuity of the trace operator $\hat{u} \mapsto \hat{u}|_{\widehat{F}}$, we get for the degrees of freedom associated with the faces

$$|N_j^{\widehat{F_\alpha}}(\widehat{u})| \leq \widehat{C}\|\widehat{u}\|_{0,\widehat{F_\alpha}} \leq \widehat{C}\|\widehat{u}\|_{1,\widehat{K}}.$$

The degrees of freedom associated with the element satisfy

$$|N_j^{\widehat{K}}(\widehat{u})| \leq \widehat{C}\|\widehat{u}\|_{0,\widehat{K}} \leq \widehat{C}\|\widehat{u}\|_{1,\widehat{K}},$$

which leads to

$$|\widehat{u} - I_h^{\widehat{K}}\widehat{u}|_{1,\widehat{K}} \leq C\|\widehat{u}\|_{1,\widehat{K}} \leq C\|\widehat{u}\|_{k+1,\widehat{K}}. \tag{5.13}$$

We conclude that the mapping $\widehat{u} \mapsto (\widehat{u} - I_h^{\widehat{K}}\widehat{u})$ is continuous on the spaces $H^{k+1}(\widehat{K})$ for $k = 0, \ldots, n+1$. Besides, this mapping vanishes on the subspace $\mathbb{P}_k$. Thus, from [76, Theorem 11.13] based on the Bramble–Hilbert Lemma, estimate (5.13) becomes

$$|\widehat{u} - I_h^{\widehat{K}}\widehat{u}|_{1,\widehat{K}} \leq \widehat{C}|\widehat{u}|_{k+1,\widehat{K}}.$$

Then, using the standard scaling properties, we obtain (5.11). We prove estimate (5.12) using the same procedure. □

It is clear that the properties of the scalar local interpolation operator I_h^K carry over to its vector-valued version $\boldsymbol{I}_h^K$. Furthermore, for the generated global interpolation operators $\boldsymbol{I}_h : V \to V_{h,n+1}$ and $J_h : M \to M_{h,n}$, we have the following result.

Lemma 5.11. The global interpolation operators $\boldsymbol{I}_h$ and J_h satisfy

$$|\boldsymbol{u} - \boldsymbol{I}_h\boldsymbol{u}|_{1,h} \leq Ch^{n+1}|\boldsymbol{u}|_{n+2} \qquad \forall \boldsymbol{u} \in V \cap H^{n+2}(\Omega)$$

$$\|p - J_h p\|_0 \leq Ch^{n+1}|p|_{n+1} \qquad \forall p \in M \cap H^{n+1}(\Omega).$$

This leads to the error estimate for the solution to (5.1).

Theorem 5.12. Let $(\boldsymbol{u}, p)$ be the solution to the Stokes system (5.1) and assume that

$$\boldsymbol{u} \in V \cap H^{n+2}(\Omega)^3,\ p \in H^{n+1}(\Omega).$$

Then, system (5.3) has a unique solution $(\boldsymbol{u_h}, p_h)$ and the following error estimate holds

$$|\boldsymbol{u} - \boldsymbol{u_h}|_{1,h} + \|p - p_h\|_0 \leq Ch^{n+1}\left(|\boldsymbol{u}|_{n+2} + |p|_{n+1}\right), \tag{5.14}$$

with C a constant independent of the mesh size h.

Proof. Given that the approximation property (Lemma 5.11), the discrete inf-sup condition (5.10) and the patch test, which follows from the definition of V_h by (5.8), are satisfied, then we have the classical error estimate (5.14) that holds for non-conforming finite elements (see for example [61, Theorem 3] or [84, Theorem 2.6]). □

Remark 5.13. To establish the result given by Theorem 5.12, we only use the $\mathbb{P}_{n+1}$ part of V_{n+1} and not the subspace Σ_{n+2}. One could imagine that taking into account Σ_{n+2} could improve the error bound of Theorem 5.12. However, as shown in the numerical results, the presence of the subspace Σ_{n+2} does not improve the order of convergence but only the accuracy level. This observation does not encourage us to investigate more about the consideration of the subspace Σ_{n+2} in the error estimate.

We have shown, that the pair of finite elements $V_{h,n+1} - M_{h,n}$ is inf-sup stable for the Stokes equations and fulfils the approximation properties under Assumption 5.7. Thus, these properties are satisfied in particular for the non-conforming finite elements $V_{h,2} - M_{h,1}$ and $V_{h,3} - M_{h,2}$.

5.5 Efficient preconditioning strategy for the Stokes equations

Below, we present an efficient preconditioning strategy to solve the discrete Stokes problem in three dimensions. The solver is developed in FreeFEM [93] interfaced with PETSc [28]. The aim of this section is to show that standard preconditioners that are known to perform well for conforming finite elements may handle just as well the proposed non-conforming finite elements. This is an appealing feature, since it means there is no need to derive a new preconditioning strategy, which can be challenging for high-dimensional and ill-posed linear systems.

5.5.1 Stokes problem in matrix form

It is well-known that the discretisation of the Stokes problem (5.1) leads to a linear system featuring a two-by-two block matrix. Let m and n be two integers with $m < n$. Let A be a sparse $n \times n$ matrix, B be a sparse $n \times m$ full-rank matrix of constraints, and C an $m \times m$ matrix (in particular $C = 0$ is allowed). From this, a saddle point matrix is built as

$$\mathcal{A} = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

We also define

$$b = \begin{pmatrix} \boldsymbol{f} \\ g \end{pmatrix},$$

where $\boldsymbol{f}$ is the source term and g is the value of the divergence (g is usually equal to 0 for the Stokes problem since we consider a divergence-free velocity). The objective is thus to efficiently solve the following linear system

$$\mathcal{A}x = b, \tag{5.15}$$

where x is the vector of unknowns. In the case of the Stokes system, $x = \begin{pmatrix} \boldsymbol{u} \\ p \end{pmatrix}$, with $\boldsymbol{u}$ the velocity unknowns and p the pressure unknowns. Below, we introduce a preconditioning strategy based on a decomposition of the problem by blocks with respect to the physics. This strategy is not new [74, 41], but our goal here is to make the chapter self-contained. It is based on the Schur complement, which is an appealing approach for two-by-two blocks, as in the problem considered.

5.5.2 Fieldsplit preconditioning based on the Schur Complement

The Schur complement is defined as

$$S = C - B^T A^{-1} B.$$

For the sake of simplicity, from now on we consider $C = 0$. In our implementation, we consider a lower Schur factorization, in which the action of the lower Schur factorization is defined by

$$M_p = \begin{pmatrix} A & 0 \\ B^T & S \end{pmatrix}. \tag{5.16}$$

Remark 5.14. Calculating explicitly M_p^{-1} leads to

$$M_p^{-1} = \begin{pmatrix} A^{-1} & 0 \\ -S^{-1}B^T A^{-1} & S^{-1} \end{pmatrix}.$$

It is then easy to show that M_p is a good left preconditioner for (5.15). However, applying this strategy requires to solve exactly systems of the form $Sz = w$, which may be very expensive. That is why we use M_p as a right preconditioner with not fully converged solutions of systems of the form $Sz = w$, taking advantage of the Flexible Generalized Minimal Residual method (FGMRES [151]) as the outer Krylov method which allows for inexact inner solvers at each outer iteration.

For applying this preconditioning strategy, one needs to solve systems of the form $M_p y = z$. These systems will themselves be preconditioned with approximate inverses of A and $\widehat{S}$ defined by $\widehat{S} = -B^T \mathrm{diag}(A)^{-1} B$ (where $\mathrm{diag}(A)$ stands for the square diagonal matrix made of the diagonal entries of A), since $\widehat{S}$ can be computed explicitly while S cannot because it is dense and not stored explicitly in practice.

In Section 5.5.3.1, we present a preconditioner to solve linear systems involving matrix A. In Section 5.5.3.2, we present a preconditioner to solve linear systems involving matrix S. These preconditioners are called respectively A_p and S_p.

5.5.3 Approximations of the physical blocks

5.5.3.1 Solution of linear systems involving matrix A

The preconditioner A_p is defined in a domain-decomposition approach using the restricted additive Schwarz method (RAS, see [43]) with an algebraic overlap

$$A_p^{-1} = \sum_{j=1}^{N} \tilde{R}_j^T (R_j A R_j^T)^{-1} R_j,$$

where N is the number of blocks (or number of subdomains), which is by default the same as the number of MPI processes, R_j is the velocity restriction operator that returns a local vector from a global vector, and $\tilde{R}_j$ is the same operator, but scaled by 0 on the overlap, of the jth block. By default, this preconditioner has one layer of overlap.

Remark 5.15. In general, larger overlaps lead to faster convergence up to a certain point where increasing the overlap does not further improve the convergence rate. Unfortunately, larger overlap implies greater communication and computation requirements.

In each block, MUMPS [15] is used to compute the action of local inverses using an LU factorization.

Remark 5.16. Boundary conditions are applied using a symmetric method, based on row and column eliminations, which allows to keep matrix A symmetric. Since it is also symmetric definite positive, MUMPS could be used to compute a Cholesky factorization instead. In theory, this choice would lead to better performance, however, our numerical experiments show that LU factorizations give better performance (less undesirable numerical pivoting).

5.5.3.2 Solution of linear systems involving matrix S

The effectiveness of Schur complement-based factorization depends on the availability of a good preconditioner S_p for the Schur complement matrix. Determining a good preconditioner for the Schur complement is often a difficult task since we hardly know the structure of the Schur complement (we recall that it is not assembled explicitly). This is why we replace in the preconditioner formula S by its approximation $\widehat{S}$. The preconditioner S_p is defined in a domain-decomposition approach by using the block Jacobi method

$$S_p^{-1} = \sum_{j=1}^{N} R_j^T (R_j \widehat{S} R_j^T)^{-1} R_j,$$

where N is still the number of blocks (or number of subdomains), which is by default the same as the number of MPI processes, and R_j (resp. R_j^T) is the pressure restriction (resp. pressure

extension by 0) operator of the jth block. It has to be noted that the operator R_j is not the same as in Section 5.5.3.1 since here we are dealing with pressure unknowns and not velocity unknowns, and we do not consider overlap compared to the one layer of overlap in Section 5.5.3.1. This method does not require any communication as each block of the preconditioner acts only on unknowns internal to each MPI process. In each of the blocks, the solution is computed using the sparse direct solver MUMPS, which is internally computing an LU factorization.

5.5.4 Final structure of the solver

Now, we summarise the proposed method for solving iteratively (5.15) and detail associated PETSc instructions for the sake of reproducibility. We use the Flexible Generalized Minimal Residual algorithm (FGMRES, `-ksp_type fgmres`), which iteratively solves (5.15) using (5.16) as an outer right-preconditioner. The linear solves involving M_p require inner velocity and pressure solves, with matrices A and S respectively, which are themselves solved iteratively using the preconditioners A_p and S_p defined in Section 5.5.3.1 and Section 5.5.3.2, respectively. The inner Krylov solver for both physical blocks is GMRES, performing at most 5 iterations (`-fieldsplit_ksp_max_it 5`), which justifies the use of the flexible GMRES algorithm as an outer Krylov solver. Concerning some other solver settings:

- the relative tolerance of the outer Krylov method is set to 10^{-10} (`-ksp_rtol 1.0E-10`, ratio according to the norm of the initial unpreconditioned residual);
- for the setup of the velocity preconditioner, MUMPS parameter ICNTL(35) is set to 2 (`-fieldsplit_0_sub_mat_mumps_icntl_35 2`), which allows to activate its block low-rank (BLR) feature [14] during both the factorization and solution phases. This allows for memory gains by storing low-rank factors. Finally, the accuracy of the low-rank approximation is itself controlled by MUMPS parameter CNTL(7), which is set to 10^{-6} (`-fieldsplit_0_sub_mat_mumps_cntl_7 1.0E-6`).

5.6 Numerical experiments

The finite element spaces V_2 and V_3, presented in Lemma 5.4, have been implemented in the open source finite element software FreeFEM [93] following the same procedure as described in Section 4.4 for non-conforming finite element in two dimensions. Therein, they are referenced respectively under the name $P2pnc3d$[3] and $P3pnc3d$[4]. In this section, we present the convergence orders obtained with these new finite elements for two different test cases, in order to verify the theoretical analysis. The obtained results are compared with those obtained with classical Taylor–Hood finite elements of order two and three. The accuracy of these methods both in term of mesh size h and number of unknowns is also compared. For all the following cases, the computational domain is the cube $\Omega = [0, 1]^3$ and the viscosity ν is set to 1.

[3] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P2pnc_3d.cpp

[4] https://github.com/FreeFem/FreeFem-sources/blob/4307d439ca8313cd8fda1c6ce34384e096efea4a/plugin/seq/Element_P3pnc_3d.cpp

5.6.1 Case #1: Stokes with non-homogeneous Dirichlet boundary conditions

First, we consider the Stokes test case proposed in [145, Guide to Stokes]. We consider the Stokes problem (5.1) with non-homogeneous Dirichlet boundary conditions $\boldsymbol{g}$ on all the faces of the cube. The load function $\boldsymbol{f}$ and the velocity Dirichlet boundary conditions $\boldsymbol{g}$ are chosen such that the exact solution is

$$\begin{cases} u_x &= 2\sin(\pi x) + \sin(\pi y) + \sin(\pi z), \\ u_y &= -\pi\cos(\pi x)y, \\ u_z &= -\pi\cos(\pi x)z, \\ p &= \sin(2\pi x) + \sin(2\pi y) + \sin(2\pi z). \end{cases}$$

5.6.1.1 Relative errors according to the element size h

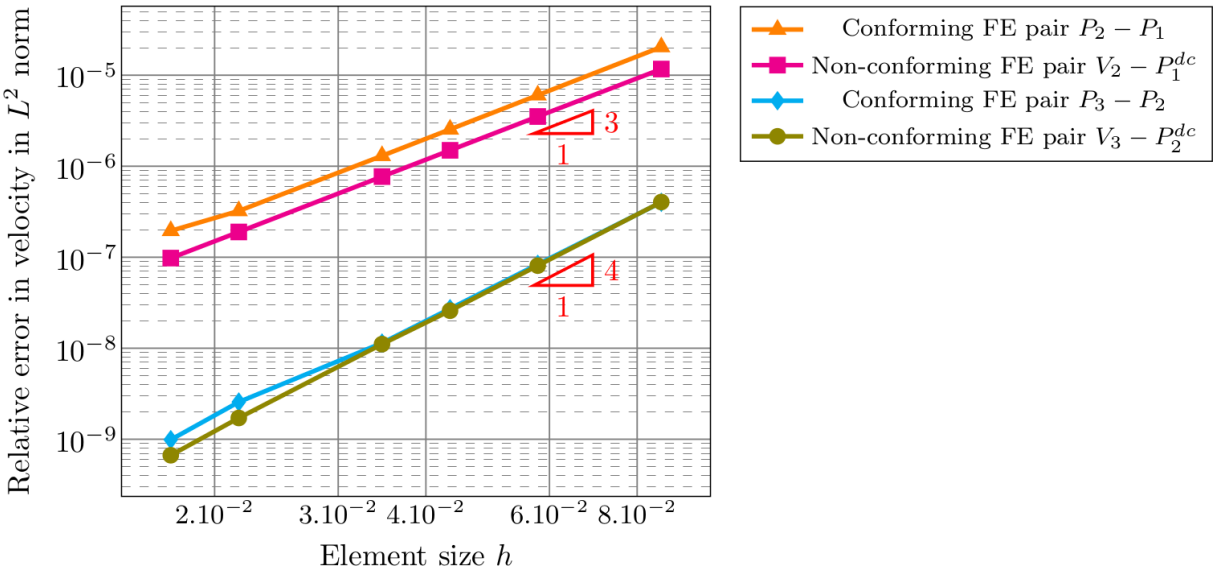


Figure 5.1: Relative error in L^2 norm between computed and exact velocities for case #1 according to the mesh size h.

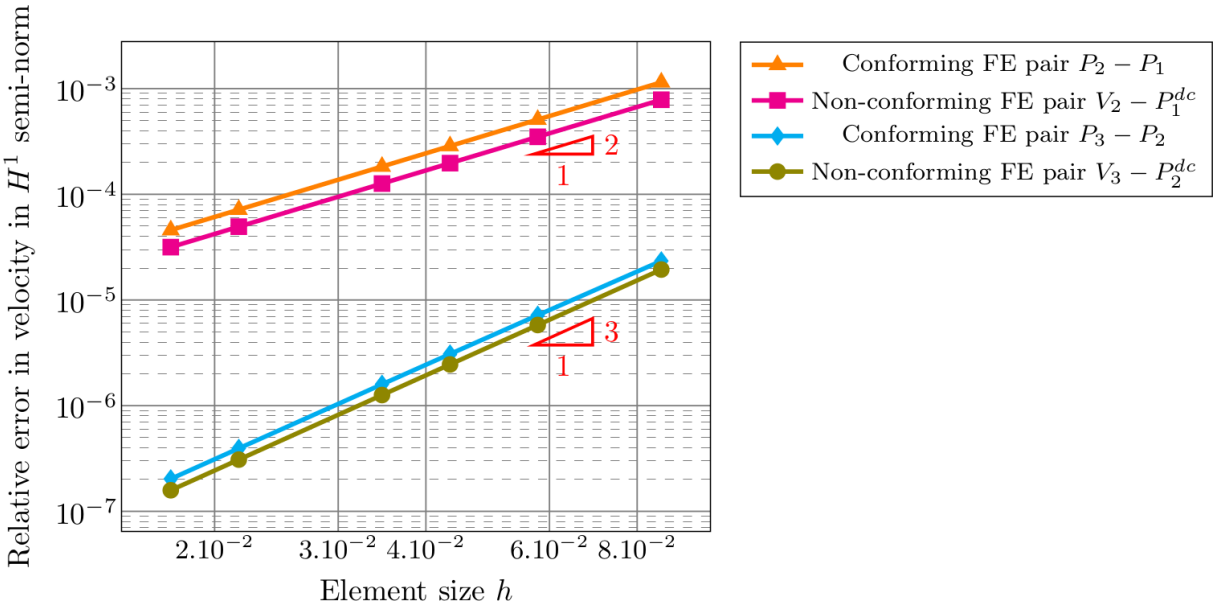


Figure 5.2: Relative error in H^1 semi-norm between computed and exact velocities for case #1 according to the mesh size h.

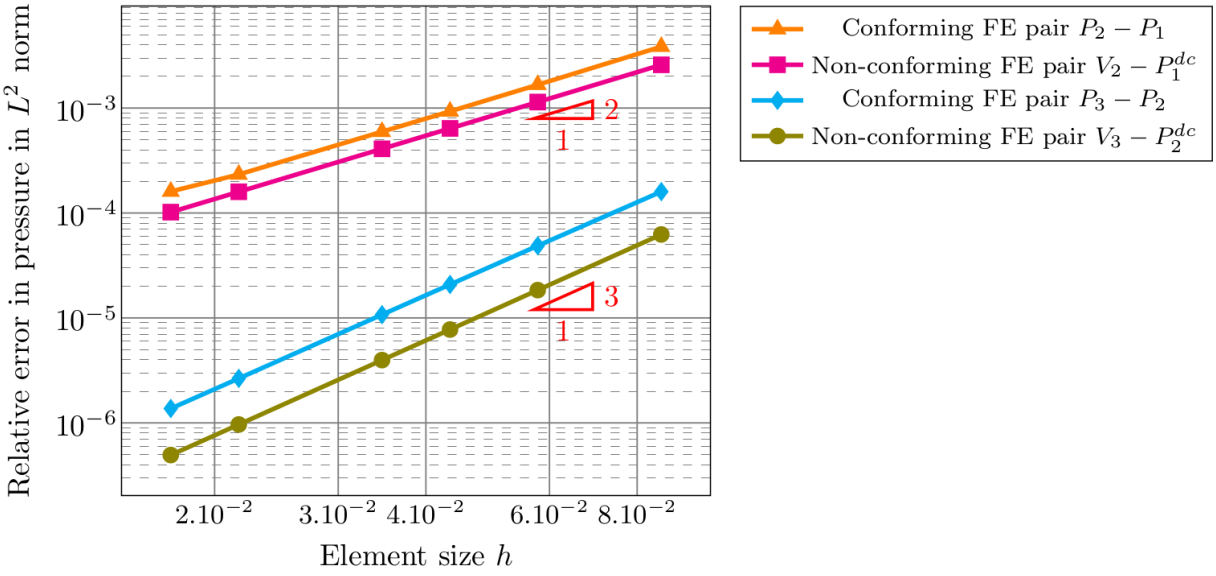


Figure 5.3: Relative error in L^2 norm between computed and exact pressures for case #1 according to the mesh size h.

Figures 5.1 to 5.3 display orders of convergence for different pairs of finite elements with respect to the mesh size h. It can be seen from these figures that the theoretical orders of convergence (see Theorem 5.12) are obtained. It should be noted that, for a given order n, the same orders of convergence are obtained for both the non-conforming and conforming pairs of finite elements. It seems that the addition of some basis functions to the space $\mathbb{P}_{n+1}$ does not improve the order of convergence. Similar behaviors can be seen for non-conforming finite elements of order n in two dimensions (see for example [132]). However, the non-conforming finite element pairs have a better level of accuracy compared to the conforming ones (i.e. the constant C in the error estimate is improved). This observation is more obvious for the relative pressure error in the L^2

norm.

5.6.1.2 Relative errors according to the number of unknowns

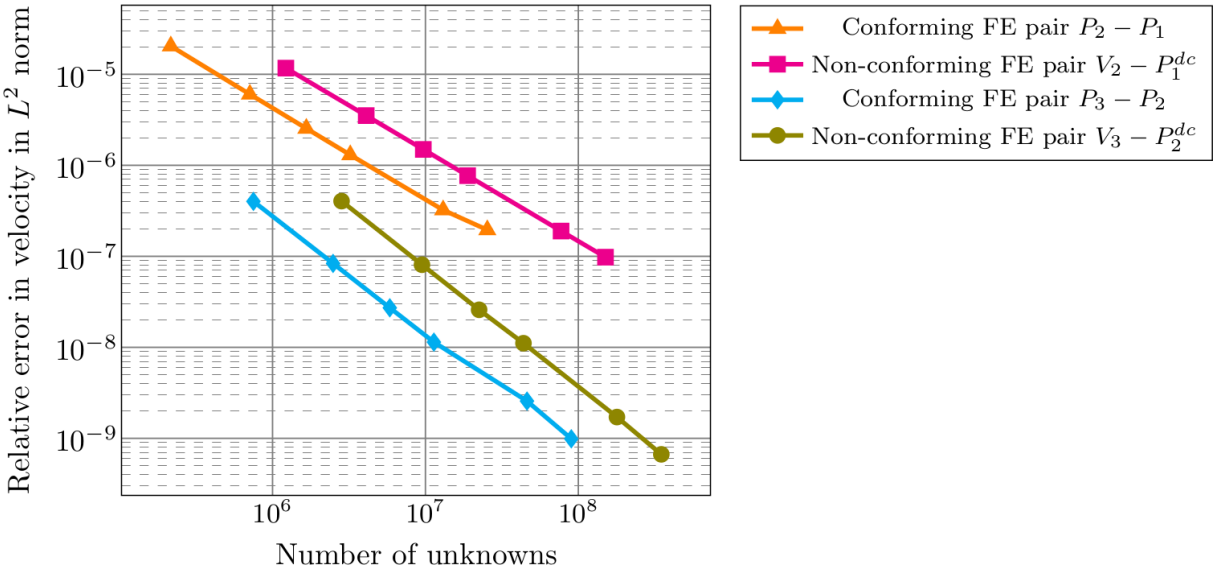


Figure 5.4: Relative error in L^2 norm between computed and exact velocities for case #1 according to the number of unknowns.

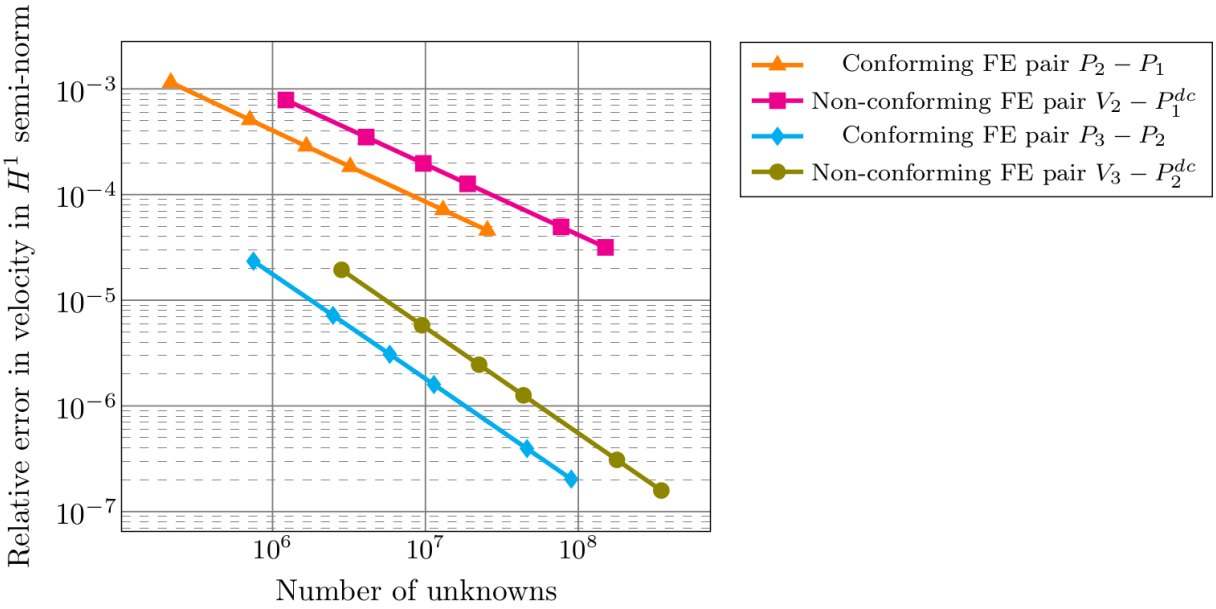


Figure 5.5: Relative error in H^1 semi-norm between computed and exact velocities for case #1 according to the number of unknowns.

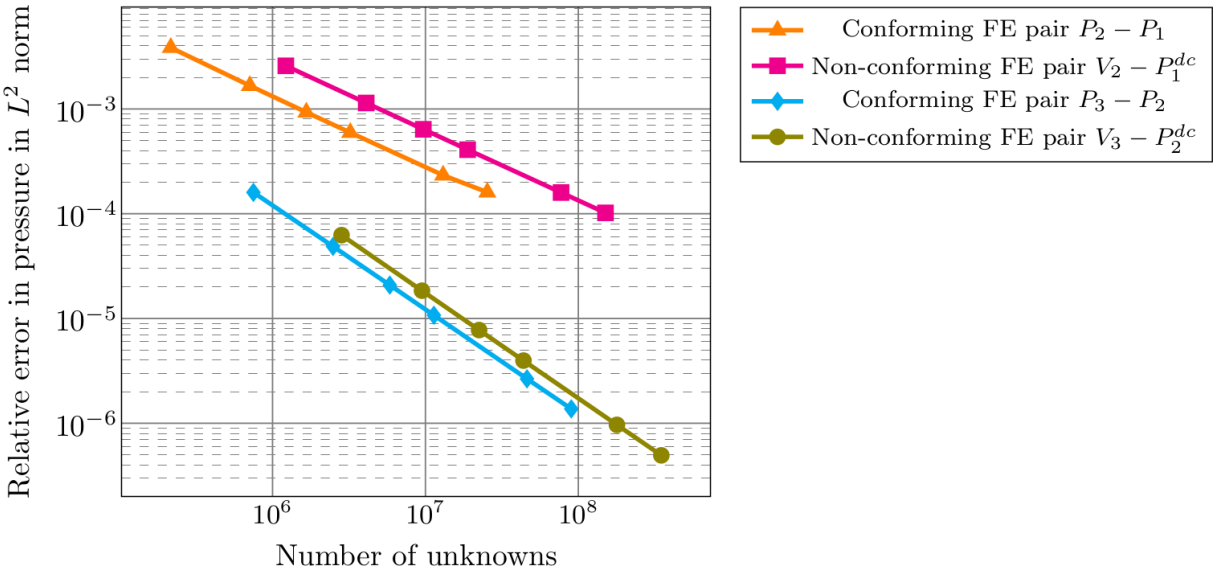


Figure 5.6: Relative error in L^2 norm between computed and exact pressures for case #1 according to the number of unknowns.

Figures 5.4 to 5.6 display the relative errors with respect to the number of degrees of freedom. These figures illustrate the well-known fact that, for regular solutions, it is better to choose higher order polynomials than to refine the mesh to decrease the errors since a given accuracy is achieved with fewer degrees of freedom.

5.6.2 Case #2: Stokes with partly Neumann boundary conditions

Second, we perform the test case proposed in [50]. The Stokes problem (5.1) is solved with a different boundary condition. More precisely, a homogeneous Neumann boundary condition is imposed on $\Gamma_N = \{z = 1\} \cap \partial\Omega$ and a homogeneous Dirichlet boundary condition is imposed on $\Gamma_D = \partial\Omega \setminus \Gamma_N$. The load function $\boldsymbol{f}$ is derived from the following exact solution

$$\boldsymbol{u} = \operatorname{curl} \begin{pmatrix} y^2(1-y)^2x(1-x)z^2(1-z)^3 \\ x^2(1-x)^2y(1-y)z^2(1-z)^3 \\ 0 \end{pmatrix} \qquad p = (x - \frac{1}{2})(y - \frac{1}{2})(1-z).$$

One can indeed check that the respective boundary conditions are homogeneous on Γ_N and Γ_D.

5.6.2.1 Relative errors according to the element size h

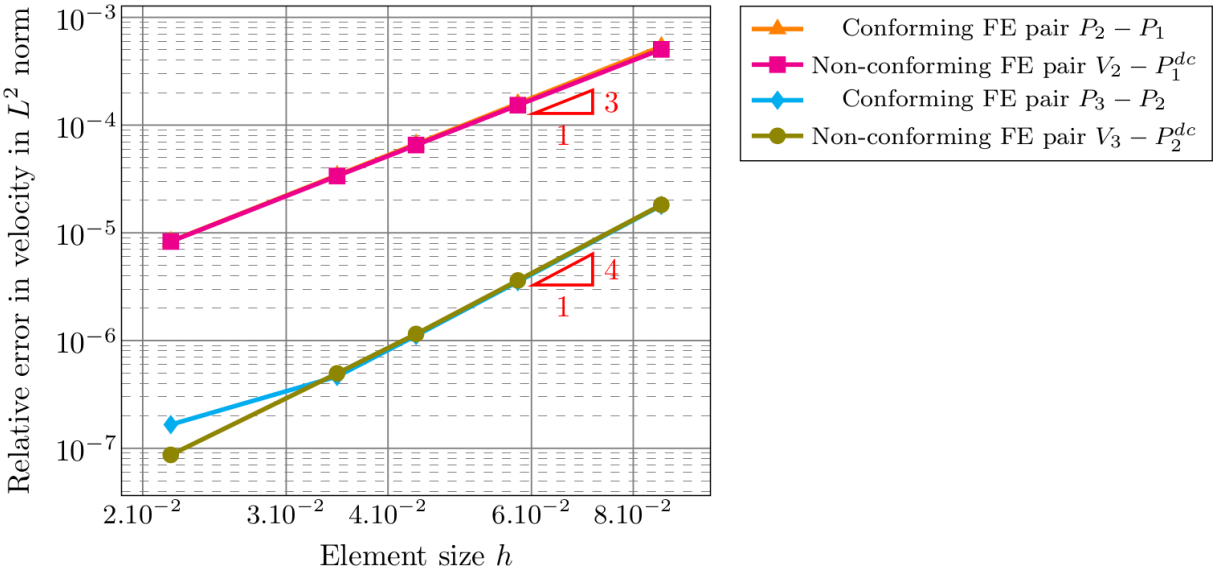


Figure 5.7: Relative error in L^2 norm between computed and exact velocities for case #2 according to the mesh size h.

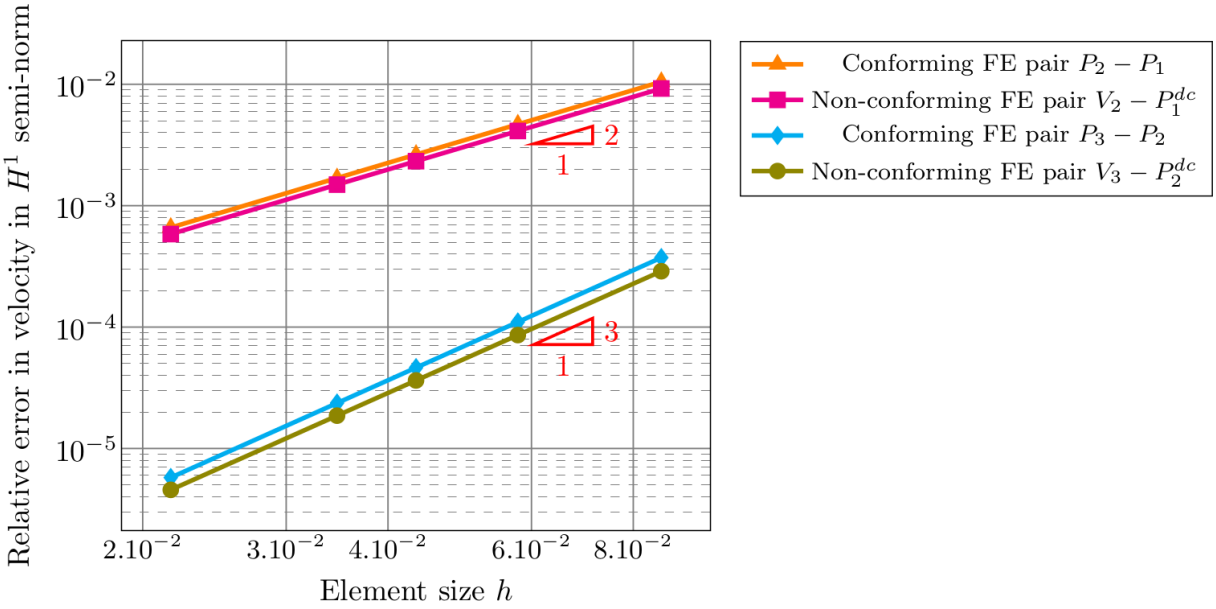


Figure 5.8: Relative error in H^1 semi-norm between computed and exact velocities for case #2 according to the mesh size h.

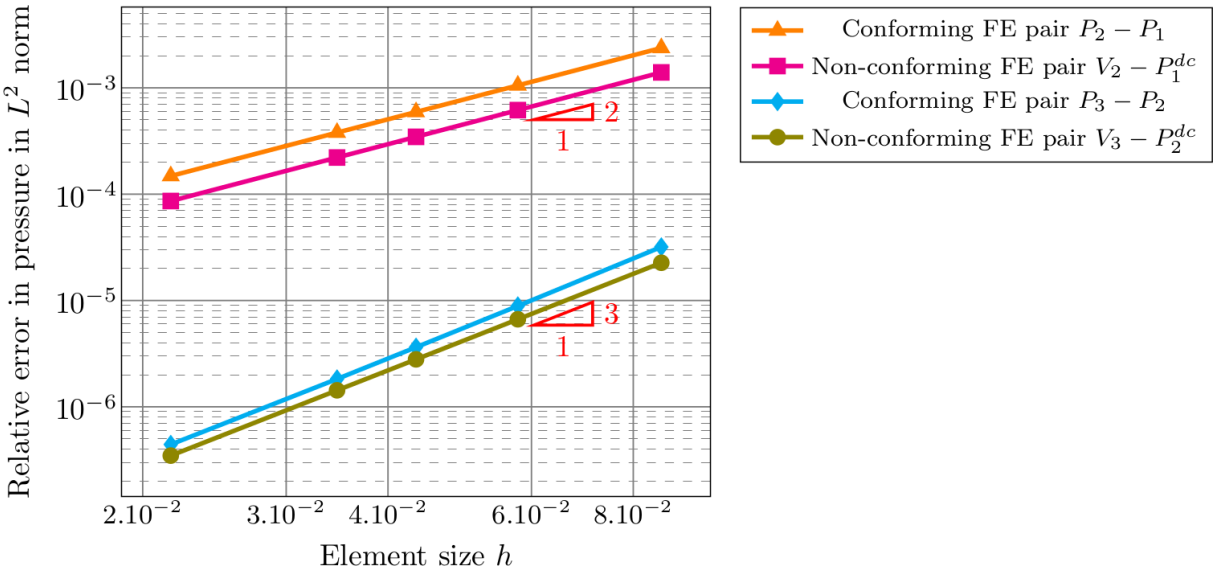


Figure 5.9: Relative error in L^2 norm between computed and exact pressures for case #2 according to the mesh size h.

5.6.2.2 Relative errors according to the number of unknowns

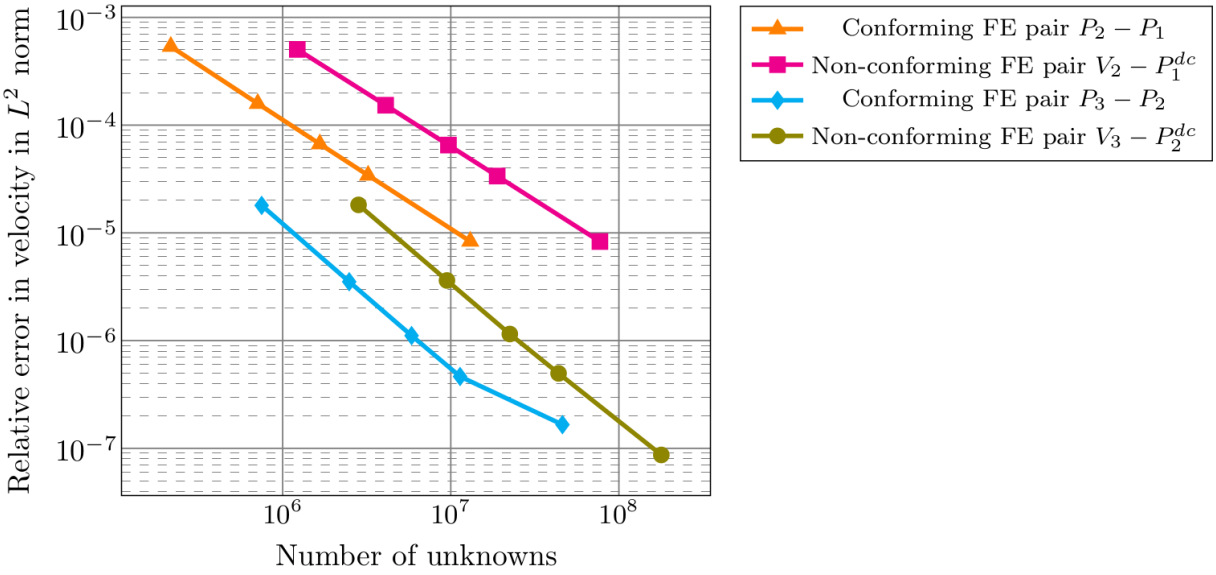


Figure 5.10: Relative error in L^2 norm between computed and exact velocities for case #2 according to the number of unknowns.

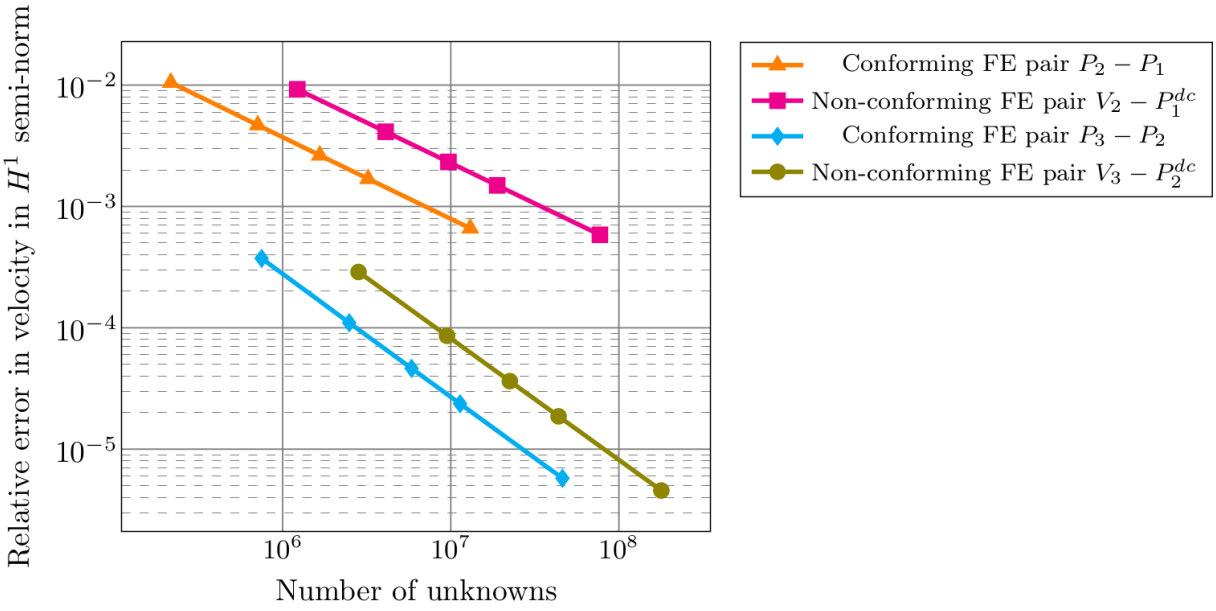


Figure 5.11: Relative error in H^1 semi-norm between computed and exact velocities for case #2 according to the number of unknowns.

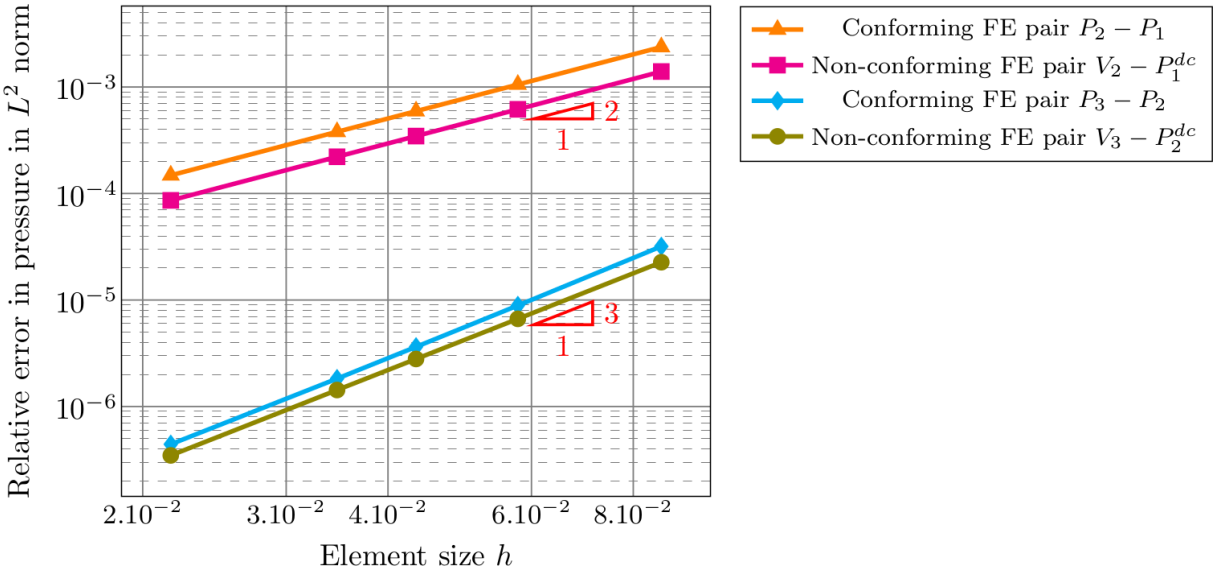


Figure 5.12: Relative error in L^2 norm between computed and exact pressures for case #2 according to the number of unknowns.

The cases presented above allow to assess the family of finite elements with Neumann boundary conditions. From Figures 5.7 to 5.9 and from Figures 5.10 to 5.12, similar observations can be done as in the case with Dirichlet boundary conditions (see Section 5.6.1).

5.7 Discussion about the number of degrees of freedom and the computational times

In this section, we compare the non-conforming pairs and the conforming ones both in terms of asymptotic number of degrees of freedom and in terms of computational times. We show

that the non-conforming discretisations have asymptotically more degrees of freedom than the conforming ones, leading to longer computational times. However, in return, the computational times per degree of freedom are lower for the non-conforming discretisations.

5.7.1 Discussion about the number of degrees of freedom

We consider a family of three-dimensional triangulations $(\mathcal{T}_{h_n})_{n\in\mathbb{N}}$, obtained from an initial mesh $\mathcal{T}_{h_0}$ in which a regular partition is recursively applied. The triangulation $\mathcal{T}_{h_n}$ is made of N_n nodes, E_n edges, F_n faces, and T_n tetrahedra. As noted in [146], the asymptotic behavior (as n goes to infinity) of the adjacency relations between the topological elements (nodes, edges, faces, and tetrahedra) in the mesh depends on the particular partition considered. First, we recall the distribution of the degrees of freedom of different finite elements according to the topological elements. Then, we make explicit the asymptotic number of degrees of freedom for different pairs of finite elements on four different partitions.

5.7.1.1 Distribution of the degrees of freedom

In Tables 5.1 to 5.3 the distribution of the degrees of freedom for different finite elements according to topological elements is recalled. This will be needed to calculate the number of the degrees of freedom for each conforming pair $\mathbb{P}_{n+1} - \mathbb{P}_n$ and of the proposed non-conforming pairs $V_2 - \mathbb{P}_1^{\mathrm{dc}}$ and $V_3 - \mathbb{P}_2^{\mathrm{dc}}$. In Table 5.2, for discontinuous finite elements, we consider that all the degrees of freedom are located in the tetrahedra since these degrees of freedom are not shared with neighboring elements.

	N_n	E_n	F_n	T_n
$\mathbb{P}_1$	1	0	0	0
$\mathbb{P}_2$	1	1	0	0
$\mathbb{P}_3$	1	2	0	4

Table 5.1: Number of unknowns of scalar conforming finite elements.

	N_n	E_n	F_n	T_n
$\mathbb{P}_1^{\mathrm{dc}}$	0	0	0	4
$\mathbb{P}_2^{\mathrm{dc}}$	0	0	0	10

Table 5.2: Number of unknowns of scalar fully discontinuous finite elements.

	N_n	E_n	F_n	T_n
V_2	0	0	3	1
V_3	0	0	6	4

Table 5.3: Number of unknowns of scalar non-conforming finite elements.

5.7.1.2 Computation of the number of degrees of freedom

Case of non-conforming pairs As noted in [146], the ratio T_n/F_n does not depend on the choice of a given partition. Indeed, for any partition $T_n/F_n = \frac{1}{2}$ asymptotically. Given that

for the non-conforming pairs, the degrees of freedom are located on faces and in tetrahedra, the number of unknowns of these pairs does not depend on the partition considered. Thus, an easy computation, using data of Tables 5.2 and 5.3 (considering three components for the velocity field and one for the pressure field) leads to

Finite element pairs	Asymptotic number of unknowns
$V_2 - \mathbb{P}_1^{\mathrm{dc}}$	$25T_n$
$V_3 - \mathbb{P}_2^{\mathrm{dc}}$	$58T_n$

Table 5.4: Number of unknowns of non-conforming pairs.

Case of conforming pairs Contrary to non-conforming pairs, for conforming pairs some unknowns are located on edges and vertices. As noted in [146], the relations between the number of edges (as well as the number of vertices) and the number of tetrahedra depend on the partition considered. Below, for the sake of completeness we give the definitions of common $3D$ partitions as presented in [146] (for more details the reader can refer to [146]).

Definition 5.17 ($3D$-Bey partition)**.** For any tetrahedron K, the $3D$-Bey partition is defined by dividing K into eight sub-tetrahedra by cutting off the four corners by the midpoints of the edges, and the remaining octahedron is subdivided further into four tetrahedra by one of the three possible interior diagonals.

Definition 5.18 ($8T-LE$ partition)**.** For any tetrahedron K, the $8T-LE$ partition is defined by dividing K into eight sub-tetrahedra by performing the $4T-LE$ partition of the faces, and then subdividing the interior of the tetrahedron in a manner consistent with the performed division in the 2-skeleton.

Definition 5.19 ($3D$ barycentric partition)**.** For any tetrahedron K, the barycentric partition is defined by inserting a new node P at the barycenter of K, putting new nodes at the barycenters of the faces of K, and putting new nodes at the midpoints of the edges of K. Then, on each face of K a barycentric triangular partition is performed. Finally, the node P is joined with all the vertices of K, and with all the new nodes introduced.

Definition 5.20 ($4T$ partition)**.** For any tetrahedron K, the $4T$ barycentric partition is defined by inserting a new node P at some interior point of K (for example at the barycenter of K) and joining P with the all the vertices of K.

Then, for each of these partitions, the following relations hold asymptotically.

	$3D$-Bey	$8T-LE$	$3D$ barycentric	$4T$
N_n/T_n	1/6	1/6	2/11	1/3
E_n/T_n	7/6	7/6	13/11	4/3
F_n/T_n	2	2	2	2

Table 5.5: Relations between topological elements for different partitions.

Using the relations in Table 5.5 with the data of Table 5.1 and considering three components for the velocity field and one for the pressure field, we get the asymptotic number of degrees of freedom for different pairs of finite elements given in Table 5.6.

Finite element pairs	$3D$-Bey	$8T-LE$	$3D$ barycentric	$4T$
$\mathbb{P}_2-\mathbb{P}_1$	$4.16T_n$	$4.16T_n$	$4.27T_n$	$5.33T_n$
$\mathbb{P}_3-\mathbb{P}_2$	$20.83T_n$	$20.83T_n$	$21T_n$	$22.66T_n$

Table 5.6: Number of unknowns of different conforming pairs according to the partition considered.

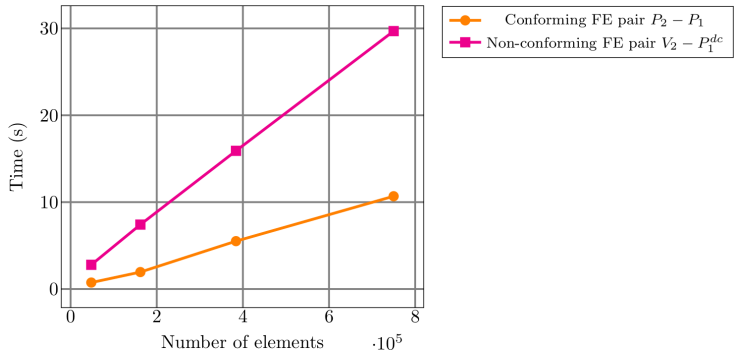
Thus, for the non-conforming $V_2-\mathbb{P}_1^{\mathrm{dc}}$ pair, we asymptotically have $25T_n$ degrees of freedom. In comparison, for the conforming $\mathbb{P}_2-\mathbb{P}_1$ pair, we have asymptotically around $5T_n$ degrees of freedom (approximately four times less than the non-conforming case). For the non-conforming $V_3-\mathbb{P}_2^{\mathrm{dc}}$ pair, we asymptotically have $58T_n$ degrees of freedom. In comparison, for the conforming $\mathbb{P}_3-\mathbb{P}_2$ pair, we have asymptotically around $22T_n$ degrees of freedom (approximately 2.2 times less than the non-conforming case). In view of these results, it is clear that using conforming methods for high-order is more advantageous.

5.7.2 Discussion about the computational times

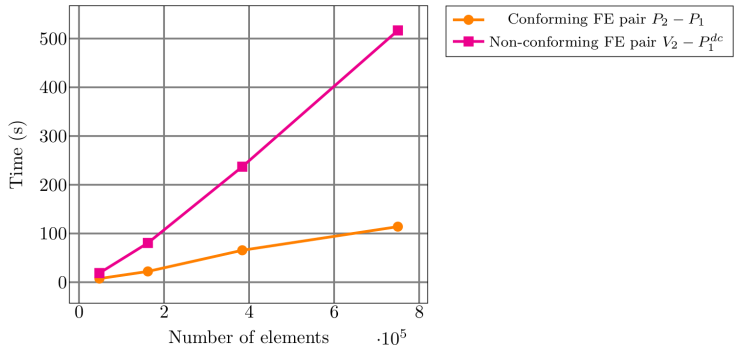
In this section, we consider case #1 (Stokes with non-homogeneous Dirichlet boundary conditions) seen in Section 5.6. We compare the computational times needed by the non-conforming $V_2-\mathbb{P}_1^{\mathrm{dc}}$ finite element pair with that needed by the conforming $\mathbb{P}_2-\mathbb{P}_1$ finite element pair. To do this, we consider the following computational steps:

- the assembly phase, during which the stiffness matrix and the right-hand side are assembled in FreeFEM;
- the solution phase, during which the preconditioner is setup and the linear system is solved algebraically.

The computations are run on ORCUS a cluster from CEA Saclay hosted at TGCC. It is composed of AMD EPYC 7281 CPUs, clocked at 2.3 GHz. For each test, the average number of outer iterations is about 40. Deriving a more efficient preconditioner to decrease the number of outer iterations is out of the scope of this paper.



(a) Assemply phase.



(b) Solution phase.

Figure 5.13: Comparison of the computational times of the non-conforming FE pair $V_2 - \mathbb{P}_1^{\text{dc}}$ and the conforming FE pair $\mathbb{P}_2 - \mathbb{P}_1$ for different mesh sizes computed on 64 processes.

First, in Figure 5.13, we compare different computational times needed by the two pairs of finite elements with different mesh sizes computed on 64 processes. As expected, the computational times needed by the non-conforming pair of finite elements are higher than those needed by the conforming one, since it has more degrees of freedom. Besides, we clearly see that the different computational times increase linearly with the mesh size (and thus with the number of degrees of freedom), a clear manifestation of the good scaling of the implemented solver. In Table 5.8, we compute the ratio of the computational times needed by the two finite element pairs for different mesh sizes presented in Table 5.7.

Mesh ID	1	2	3	4	5	6
# of elements (6 ×)	20^3	30^3	40^3	50^3	80^3	100^3

Table 5.7: Different considered meshes.

Mesh ID	# of processes	÷ # of unknowns	÷ assembly phase	÷ solution phase	÷ total computational time
1	64	5.6	3.7	2.5	2.6
2		5.8	3.8	3.6	3.6
3		5.8	2.9	3.6	3.5
4		5.8	2.8	4.5	4.3
5	512	5.9	2.9	6.0	5.0
6		5.9	2.8	7.6	5.6

Table 5.8: Ratio (represented by the symbol ÷) between different computational times obtained with the conforming $\mathbb{P}_2 - \mathbb{P}_1$ finite element pair and the non-conforming $V_2 - \mathbb{P}_1^{\text{dc}}$ finite element pair.

From Table 5.8, we see that the ratio of the computational times needed with the two finite element pairs is lower than the ratio of the number of degrees of freedom of the two finite element pairs. Indeed, although the non-conforming finite element pair has about six times more degrees of freedom than the conforming one, the computational time ranges between 2.6 and 5.6 times higher.

Remark 5.21. Similar comparisons can be done between the conforming $\mathbb{P}_3-\mathbb{P}_2$ finite element pair and the non-conforming $V_3-\mathbb{P}_2^{\mathrm{dc}}$ finite element pair. The same conclusions apply.

5.8 Conclusion

In this chapter, two inf-sup stable non-conforming discretisations with accuracy order two and three on tetrahedra were developed for the Stokes problem in three dimensions. From a given set of degrees of freedom, we propose a strategy to find unisolvent finite element families with respect to those degrees of freedom. The family of finite elements fulfils the consistency, approximability and stability conditions, guaranteeing optimal orders of convergence which are indeed observed in the numerical results. We recall that this family of finite elements has been designed for solving the local problems involved in MsFEM (further details can be found in Part III). Given the theoretical and numerical results presented in this paper, the two new non-conforming finite elements proposed meet the requirements of these complex problems. However, for the solution to the classical Stokes problem, the non-conforming discretisations are more expensive than the conforming ones both in terms of number of degrees of freedom and computational times, for ultimately comparable errors.

Chapter 6

Sharp convergence rates for the homogenization of the Stokes equations in a perforated domain

Contents

Note: The content of this chapter has been published in [30].

[30] L. Balazi, G. Allaire, and P. Omnes. Sharp convergence rates for the homogenization of the Stokes equations in a perforated domain, *Discrete and Continuous Dynamical Systems - B*, 2024.

This chapter is concerned with the homogenization of the Stokes equations in a periodic perforated domain. We establish a sharp convergence rate $O(\sqrt{\varepsilon})$ for the energy norm of the difference of the velocities and pressures, where ε represents the size of the solid obstacles. This is achieved by using a two-scale asymptotic expansion of the Stokes equations and a new construction of a cut-off function which avoids the introduction of boundary layers. The main novelty is that our analysis applies for the physically relevant case of a porous medium where both the fluid and solid parts are connected. The main result of this chapter (Theorem 6.9) is one

of the main ingredient for the derivation of an error estimate for the Multi-scale Finite Element Method (MsFEM) applied to the Stokes equations in perforated domains (see Chapter 9).

6.1 Introduction

Let $\Omega \subset \mathbb{R}^d$ be a regular bounded open set. As shown in Figure 6.1, we divide the domain Ω into a fixed solid part B^ε and its complementary fluid part Ω^ε. Here ε denotes a small parameter equal to the ratio between the characteristic length of the periodic heterogeneities and the characteristic length of the domain. A first typical example of such a porous domain Ω^ε is given by Figure 6.1 where the solid obstacles B^ε are a collection of isolated and periodically repeated obstacles. A second typical case, in dimension $d \geq 3$, is a porous domain Ω^ε perforated by a regular lattice of interconnected solid cylinders as presented in Figure 6.2. A precise definition of Ω^ε is given in Section 6.2.

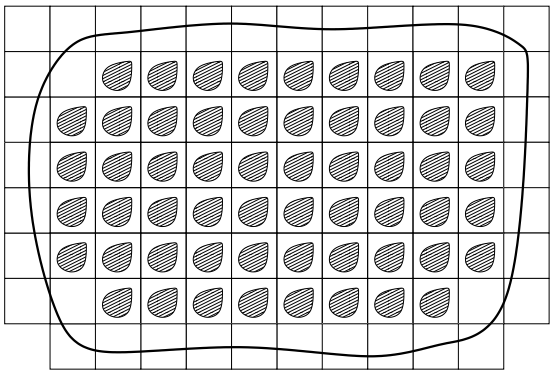


Figure 6.1: Porous medium Ω^ε, obtained from a domain Ω perforated by a set of solid isolated obstacles B^ε.

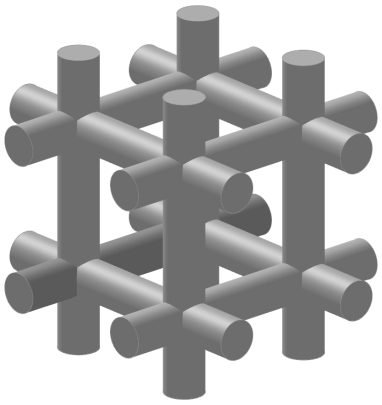


Figure 6.2: Regular lattice of interconnected cylinders.

We recall that the steady-state Stokes problem with homogeneous Dirichlet boundary condition in a perforated domain is to find the velocity $\boldsymbol{u_\varepsilon} : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$

solution to

$$\left\{\begin{array}{rcll} -\nu\Delta \boldsymbol{u_\varepsilon} + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{in } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u_\varepsilon} &=& 0 & \text{in } \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{on } \partial B^\varepsilon \cap \partial\Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{on } \partial\Omega \cap \partial\Omega^\varepsilon, \end{array}\right. \tag{6.1}$$

where $\nu > 0$ is the viscosity and $\boldsymbol{f}$ the applied force. Vector valued functions are written in bold. In the literature, some authors consider a different scaling of (6.1), replacing the velocity $\boldsymbol{u_\varepsilon}$ by $\boldsymbol{u_\varepsilon} = \varepsilon^2 \tilde{\boldsymbol{u}}_\varepsilon$. This does not change anything to our methodology and our results, up to this ε^2 factor.

We denote by $(\cdot,\cdot)$ the usual scalar product in $L^2(\Omega^\varepsilon)$, we use the same notation for vector-valued functions. We introduce the classical velocity space $V = H_0^1(\Omega^\varepsilon)^d = \{\boldsymbol{v} \in H^1(\Omega^\varepsilon)^d \ s.t. \ \boldsymbol{v}|_{\partial\Omega^\varepsilon} = \boldsymbol{0}\}$ and pressure space $M = L_0^2(\Omega^\varepsilon) = \{q \in L^2(\Omega^\varepsilon),\ s.t.\ \int_{\Omega^\varepsilon} q = 0\}$. In V, we shall often use the semi-norm $|\cdot|_{H^1}$, equivalent to the usual norm and defined by

$$|\boldsymbol{v}|_{H^1(\Omega^\varepsilon)} = \|\boldsymbol{\nabla v}\|_{L^2(\Omega^\varepsilon)}, \quad \boldsymbol{v} \in H^1(\Omega^\varepsilon)^d.$$

We introduce the bi-linear forms $a : V \times V \to \mathbb{R}$ and $b : M \times V \to \mathbb{R}$ by

$$a(\boldsymbol{u},\boldsymbol{v}) = \nu(\boldsymbol{\nabla u}, \boldsymbol{\nabla v}), \qquad b(p,\boldsymbol{v}) = (p,\ \operatorname{div} \boldsymbol{v}).$$

Then, assuming that $\boldsymbol{f} \in L^2(\Omega)^d$, a weak formulation of the Stokes problem (6.1) reads as follows: find $\boldsymbol{u_\varepsilon} \in V$ and $p_\varepsilon \in M$ such that

$$\left\{\begin{array}{rcll} a(\boldsymbol{u_\varepsilon},\boldsymbol{v}) + b(p_\varepsilon,\boldsymbol{v}) &=& (\boldsymbol{f},\boldsymbol{v}) & \forall \boldsymbol{v} \in V, \\ b(q,\boldsymbol{u_\varepsilon}) &=& 0 & \forall q \in M. \end{array}\right. \tag{6.2}$$

It is well-known that, if Ω^ε is connected (to ensure that the zero-average condition of the pressure is enough to remove the undetermined constant of the pressure), there exists a unique weak solution to (6.2) [84].

The homogenization of the Stokes equations, i.e. finding the limit system satisfied by the limit of $(\boldsymbol{u_\varepsilon}, p_\varepsilon)$ as ε goes to zero, was first investigated by [154, 160, 9]. A review of these results can be found in [102], [48]. The homogenized or effective equations for the Stokes system, in a periodic perforated domain, is Darcy's law for the velocity $\boldsymbol{u}$ and the homogenized pressure p^*,

$$\left\{\begin{array}{rcll} \boldsymbol{u} &=& \frac{1}{\nu} A^*(\boldsymbol{f} - \boldsymbol{\nabla} p^*) & \text{in } \Omega, \\ \operatorname{div} \boldsymbol{u} &=& 0 & \text{in } \Omega, \\ \boldsymbol{u}\cdot\boldsymbol{n} &=& 0 & \text{on } \partial\Omega, \end{array}\right.$$

where A^* is a constant permeability tensor (see Proposition 6.5).

The first convergence result for the homogenization of the Stokes equations was established in [160], where the author proved the weak L^2 convergence of the velocity and the strong L^2 convergence of the pressure. The strong L^2 convergence of the velocity with a corrector was proven later in [9]. The first quantitative result of convergence in the H^1-norm was obtained in [130] with a relative error estimate of order $\varepsilon^{1/6}$ in a two-dimensional domain ($d = 2$). This

result has been improved to $\sqrt{\varepsilon}$ in [111] for the case of isolated solid obstacles (still in two space dimension $d = 2$). This $\sqrt{\varepsilon}$ error estimate was independently and further improved in [156] for any space dimension $d \geq 2$, still under the assumption of isolated solid obstacles (as in Figure 6.1). The main technical idea in [156] is to construct boundary correctors which control the boundary layers appearing in the proof of the error estimate. Note that the assumption of isolated solid obstacles is not physically realistic in dimension $d \geq 3$.

The goal of the work present in this chapter is to extend this $\sqrt{\varepsilon}$ error estimate to the case of connected solid obstacles (as in Figure 6.2). Before stating our main result (in loose terms, see Theorem 6.9 for a precise statement) we need to introduce some notations. Actually the Darcy velocity $\boldsymbol{u}$ does not see the solid obstacles and thus cannot be a good approximation of the original velocity $\boldsymbol{u_\varepsilon}$. It must be corrected by introducing a better approximation which is

$$\boldsymbol{u_2}\left(x, \frac{x}{\varepsilon}\right) = \frac{1}{\nu} \sum_{i=1}^{d} \boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right) \left(f_i(x) - \frac{\partial p^*}{\partial x_i}(x)\right),$$

where the local velocities $\boldsymbol{\omega_i}$ are solutions of cell problems (6.6). Actually the link between $\boldsymbol{u_2}$ and the Darcy velocity $\boldsymbol{u}$ is that $\boldsymbol{u}(x)$ is the average of $\boldsymbol{u_2}(x, y)$ with respect to the periodic variable y.

Theorem 6.1. Let $\boldsymbol{u_\varepsilon}$, p_ε be the solution to the Stokes problem (6.1) and $\boldsymbol{u_2}$, p^*, their homogenized approximations. Assuming that $\boldsymbol{f}$ is smooth enough, there exists a constant C, independent of ε , such that

$$\frac{\|p_\varepsilon - p^*\|_{L^2(\Omega^\varepsilon)}}{\|p^*\|_{L^2(\Omega^\varepsilon)}} \leq C\varepsilon^{\frac{1}{2}},$$

$$\frac{|\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)}}{|\varepsilon^2 \boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)}} \leq C\varepsilon^{\frac{1}{2}},$$

$$\frac{\|\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}\|_{L^2(\Omega^\varepsilon)}}{\|\varepsilon^2 \boldsymbol{u_2}\|_{L^2(\Omega^\varepsilon)}} \leq C\varepsilon^{\frac{1}{2}}.$$

Theorem 6.1 is stated in terms of relative errors since $\varepsilon^2 \boldsymbol{u_2}$ (and thus $\boldsymbol{u_\varepsilon}$) is small. Its proof follows the same strategy as that in [111] (but extends it to dimensions larger than 2) and differs significantly from the proof in [156]. Actually, our proof is simpler in the sense that no boundary layers (and thus correctors of their effect) appear. It turns out that we are able to use a cut-off argument as in the standard elliptic case [31] but without compromising the divergence-free condition for the velocities. Of course, the main novelty is that Theorem 6.1 applies to any case of solide obstacles, isolated or connected. In dimension $d \geq 3$, the physically relevant case is that of connected obstacles and not isolated obstacles.

The content of this chapter is the following. Section 6.2 gives a precise definition of the geometrical assumptions of a periodic porous medium. Section 6.3 is devoted to the formal two-scale asymptotic expansion method which delivers the homogenized problem, as well as some cell correctors, including the velocities $\boldsymbol{\omega_i}$ appearing in the definition of $\boldsymbol{u_2}$. In Section 6.4 is

stated our main result (Theorem 6.9 which is just a rigorous statement of Theorem 6.1) and its proof is decomposed in several technical lemmas.

6.2 Geometric modeling of the porous medium

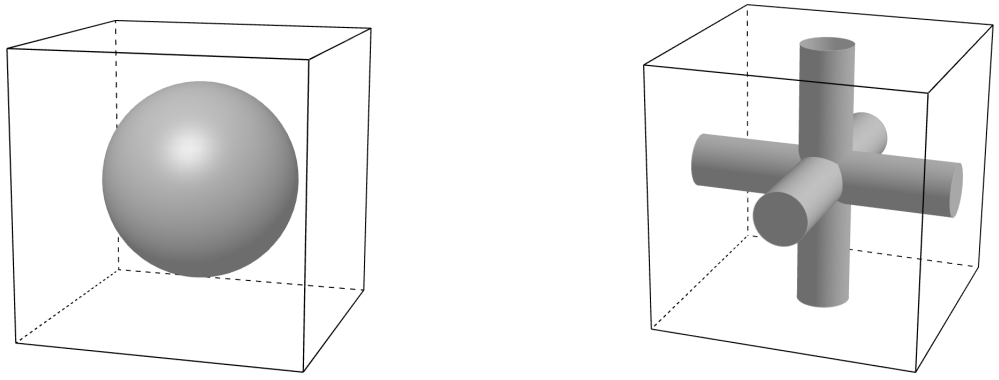
As usual in periodic homogenization theory [31, 102, 154], we consider a porous medium obtained by the periodic repetition of an elementary cell of size ε, in a bounded domain of $\mathbb{R}^d$. We first define the corresponding dimensionless elementary cell Y.

6.2.1 Definition of the elementary cell Y

Let $Y =]0,1[^d$ be the open unit cube of $\mathbb{R}^d$, $d \geq 2$. Let Y_S be a closed subset of $\overline{Y}$. We define Y_F, an open set of $\mathbb{R}^d$, by $Y_F = Y \setminus Y_S$, where Y_S represents the part of Y occupied by the solid and Y_F represents the part of Y occupied by the fluid. The fluid part Y_F and the solid part Y_S satisfy $Y_F \cup Y_S = Y$ and $Y_F \cap Y_S = \emptyset$ (see Figure 6.3). The closed set Y_S is repeated by Y-periodicity and fills the entire space $\mathbb{R}^d$, in order to obtain a closed set of $\mathbb{R}^d$, noted E_S. Let the open set E_F be the complementary of E_S in $\mathbb{R}^d$, i.e. $E_F = \mathbb{R}^d \setminus E_S$. We assume the following hypotheses on Y_F and E_F:

1. Y_F and Y_S have strictly positive measure on $\overline{Y}$ (the elementary cell Y contains fluid and solid together).
2. Y_F is an open connected set with a locally Lipschitz boundary.
3. E_F and the interior of E_S are open sets with smooth boundaries of class C^{m+2}, with $m > \frac{d}{2}$, and are locally located on one side of their boundary. Moreover E_F is connected.

Remark that the solid part E_S can be connected or not, corresponding to the two different geometric cases of Figures 6.1 and 6.2 which are later called isolated obstacles or connected obstacles. The assumption that E_F is smooth and connected implies that, on each face of Y, there is a fluid sub-domain of non-zero (surface) measure.



(a) (b)

Figure 6.3: Isolated (left) or connected (right) solid obstacles in the unit cell $Y = Y_F \cup Y_S$.

6.2.2 Definition of the open set Ω^ε

Let Ω be a bounded and connected open set of $\mathbb{R}^d$ ($d \geq 2$) which is assumed to have a smooth boundary of class $C^{3,\alpha}$, for some $0 < \alpha < 1$. The set Ω is covered with a regular square mesh of size ε, each cell being a cube Y_i^ε. We define for any $i \in \mathbb{Z}^d$:

$$Y_i = Y + i, \quad Y_{i,F} = Y_F + i, \quad Y_{i,S} = Y_S + i,$$

$$Y_i^\varepsilon = \varepsilon(Y + i), \quad Y_{i,F}^\varepsilon = \varepsilon(Y_F + i), \quad Y_{i,S}^\varepsilon = \varepsilon(Y_F + i).$$

The periodic set of solid obstacles (or perforations) inside Ω is defined as

$$B^\varepsilon = \bigcup_{i \in \mathcal{I}} Y_{i,S}^\varepsilon,$$

where $\mathcal{I} \subset \mathbb{Z}^d$ is a set of indices which is precisely defined in Assumption 6.2 below. The role of $\mathcal{I}$ is to remove obstacles from Ω only if they are sufficiently away from the boundary $\partial\Omega$. Finally, the fluid part Ω^ε of the perforated medium is defined by $\Omega^\varepsilon = \Omega \setminus B^\varepsilon$, i.e.

$$\Omega^\varepsilon = \Omega \setminus \bigcup_{i \in \mathcal{I}} Y_{i,S}^\varepsilon.$$

The definition of $\mathcal{I}$ depends on the two cases of isolated or connected solid obstacles (see Figure 6.3).

Assumption 6.2. We consider two different sets of assumptions depending on the two cases of isolated or connected solid obstacles.

1. If the solid part Y_S is isolated, namely strictly included in Y (not touching its boundary), then $\mathcal{I} = \{i \in \mathbb{Z}^d \ s.t. \ Y_i^\varepsilon \subset \Omega\}$. In other words, only entire obstacles $Y_{i,S}^\varepsilon$ are removed from Ω and thus no obstacles are cut by the boundary $\partial\Omega$.

2. If the solid part Y_S touches the boundary of Y, meaning that obstacles are connected, then we define a first set of indices $\mathcal{I}_1 = \{i \in \mathbb{Z}^d \ s.t. \ Y_i^\varepsilon \subset \Omega\}$ and an open subset of Ω defined by its closure
$$\overline{C_1^\varepsilon} = \bigcup_{i \in \mathcal{I}_1} \overline{Y_i^\varepsilon}.$$
Then, a second smaller set of indices is
$$\mathcal{I}_2 = \{i \in \mathcal{I}_1 \ \text{ such that } \ \overline{Y_i^\varepsilon} \cap \partial C_1^\varepsilon = \emptyset\},$$
meaning that no cell Y_i^ε in $\mathcal{I}_2$ touches the boundary ∂C_1^ε. Similarly, we define an open subset of Ω defined by its closure
$$\overline{C_2^\varepsilon} = \bigcup_{i \in \mathcal{I}_2} \overline{Y_i^\varepsilon}.$$
Finally, we choose $\mathcal{I} = \mathcal{I}_2$, meaning that the obstacles are not cut by the boundary $\partial\Omega$ and stay away from $\partial\Omega$ and ∂C_1^ε at a distance of the order of ε (see Figure 6.4, which is 2-d sketch of a 3-d situation).

Remark 6.3. The second item of Assumption 6.2 allows us to treat the realistic case of a connected solid part, which is possible only for space dimensions $d \geq 3$ because in 2-d E_F and E_S cannot be connected simultaneously. The second item of Assumption 6.2 is similar to an assumption in [48] and is slightly different from the assumption made in [9]. It is a necessary assumption for several technical reasons, including the fact that, when E_S is connected, there may be some pathological cut obstacles near the boundary $\partial\Omega$. In particular, the fluid domain may not be connected.

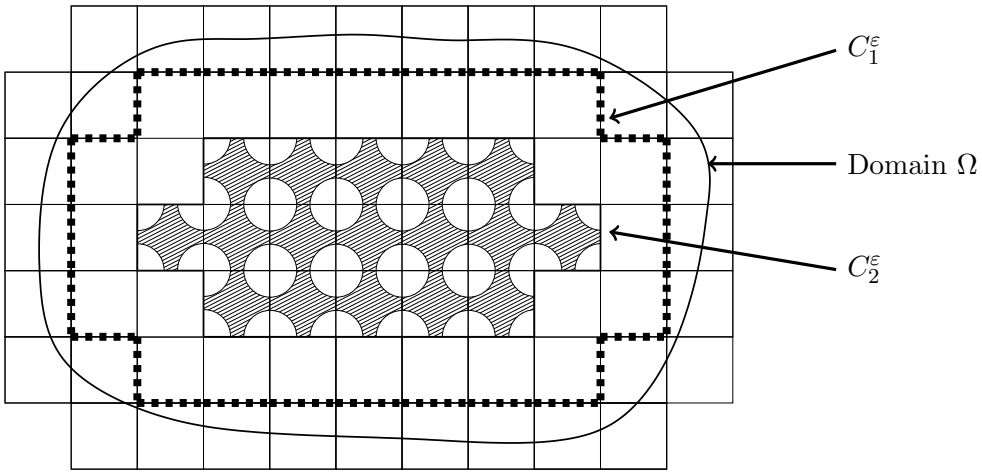


Figure 6.4: Porous medium Ω^ε when the solid part Y_S touches the boundary of Y.

6.3 Two-scale asymptotic expansion

This section recalls how to formally obtain the homogenized problem for the Stokes equations (6.1), as well as the definitions of several corrector terms in the unit cell which are required for the sequel. This is achieved by applying the classical method of two-scale asymptotic expansion [31, 154]. The following results are already well-known but are recalled since the notations are required for the next sections. The formal method of two-scale asymptotic expansion starts with the following ansatz for the velocity $\boldsymbol{u_\varepsilon}$ and pressure p_ε,

$$\boldsymbol{u_\varepsilon}(x) = \sum_{k=2}^{+\infty} \varepsilon^k \boldsymbol{u_k}\left(x, \frac{x}{\varepsilon}\right), \qquad p_\varepsilon(x) = \sum_{k=0}^{+\infty} \varepsilon^k p_k\left(x, \frac{x}{\varepsilon}\right). \tag{6.3}$$

All functions $\boldsymbol{u_k}(x,y)$, $p_k(x,y)$ are assumed $\mathbb{Z}^d$-periodic in y i.e. 1-periodic with respect to each component $y_1, \ldots, y_d$.

Remark 6.4. We begin the asymptotic expansion of the velocity $\boldsymbol{u_\varepsilon}$ at $k = 2$ since it is well-known in the literature that the first terms $\boldsymbol{u_0}$ and $\boldsymbol{u_1}$ vanish.

We recall the following derivation rule

$$\boldsymbol{\nabla}\left(\boldsymbol{u_k}(x, \frac{x}{\varepsilon})\right) = \left(\varepsilon^{-1}\boldsymbol{\nabla_y}\boldsymbol{u_k} + \boldsymbol{\nabla}\boldsymbol{u_k}\right)\left(x, \frac{x}{\varepsilon}\right), \tag{6.4}$$

where $\boldsymbol{\nabla}$ and $\boldsymbol{\nabla_y}$ denote the partial derivatives with respect to the macroscopic variable x and the microscopic variable y. Then, introducing the series (6.3) in the Stokes equations (6.1), using the derivation rule (6.4), we have for the velocity $\boldsymbol{u_\varepsilon}$

$$\begin{aligned}\boldsymbol{\nabla u_\varepsilon}(x) &= \sum_{k=2}^{+\infty} \varepsilon^k \left(\boldsymbol{\nabla u_k} + \varepsilon^{-1}\boldsymbol{\nabla_y u_k}\right)\\ &= \sum_{k=2}^{+\infty} \varepsilon^k \left(\boldsymbol{\nabla u_k} + \boldsymbol{\nabla_y u_{k+1}}\right) + \varepsilon \boldsymbol{\nabla_y u_2},\end{aligned}$$

$$\begin{aligned}\operatorname{div} \boldsymbol{u_\varepsilon}(x) &= \sum_{k=2}^{+\infty} \varepsilon^k \left(\operatorname{div} \boldsymbol{u_k} + \varepsilon^{-1} \operatorname{div}_y \boldsymbol{u_k}\right)\\ &= \sum_{k=2}^{+\infty} \varepsilon^k \left(\operatorname{div} \boldsymbol{u_k} + \operatorname{div}_y \boldsymbol{u_{k+1}}\right) + \varepsilon \operatorname{div}_y \boldsymbol{u_2},\end{aligned}$$

$$\begin{aligned}\Delta \boldsymbol{u_\varepsilon}(x) &= \sum_{k=2}^{+\infty} \varepsilon^k \left(\operatorname{div} \nabla \boldsymbol{u_k} + \varepsilon^{-1} \operatorname{div} \boldsymbol{\nabla_y u_k} + \varepsilon^{-1} \operatorname{div}_y \nabla \boldsymbol{u_k} + \varepsilon^{-2} \operatorname{div}_y \boldsymbol{\nabla_y u_k}\right)\\ &= \sum_{k=0}^{\infty} \varepsilon^k \left(\operatorname{div} \boldsymbol{\nabla u_k} + \operatorname{div} \boldsymbol{\nabla_y u_{k+1}} + \operatorname{div}_y \boldsymbol{\nabla u_{k+1}} + \operatorname{div}_y \boldsymbol{\nabla_y u_{k+2}}\right)\\ &\quad + \varepsilon \left(\operatorname{div} \boldsymbol{\nabla_y u_2} + \operatorname{div}_y \boldsymbol{\nabla u_2} + \operatorname{div}_y \boldsymbol{\nabla_y u_3}\right) + \operatorname{div}_y \boldsymbol{\nabla_y u_2},\end{aligned}$$

and for the pressure p_ε,

$$\begin{aligned}\boldsymbol{\nabla} p_\varepsilon(x) &= \sum_{k=0}^{\infty} \varepsilon^k \left(\boldsymbol{\nabla} p_k + \varepsilon^{-1}\boldsymbol{\nabla_y} p_k\right)\\ &= \sum_{k=0}^{\infty} \varepsilon^k \left(\boldsymbol{\nabla} p_k + \boldsymbol{\nabla_y} p_{k+1}\right) + \varepsilon^{-1}\boldsymbol{\nabla_y} p_0.\end{aligned}$$

Finally, we obtain for the momentum equation

$$\begin{aligned}-\nu\Delta \boldsymbol{u_\varepsilon}(x) + \boldsymbol{\nabla} p_\varepsilon(x) &= \varepsilon^{-1}\boldsymbol{\nabla_y} p_0 + \varepsilon^0 \left\{-\nu \operatorname{div}_y \boldsymbol{\nabla_y u_2} + \boldsymbol{\nabla} p_0 + \boldsymbol{\nabla_y} p_1\right\}\\ &\quad + \varepsilon \left\{-\nu\left(\operatorname{div} \boldsymbol{\nabla_y u_2} + \operatorname{div}_y \boldsymbol{\nabla u_2} + \operatorname{div}_y \boldsymbol{\nabla_y u_3}\right) + \boldsymbol{\nabla} p_1 + \boldsymbol{\nabla_y} p_2\right\}\\ &\quad + \sum_{k=2}^{+\infty} \varepsilon^k \left\{-\nu\left(\operatorname{div} \boldsymbol{\nabla u_k} + \operatorname{div} \boldsymbol{\nabla_y u_{k+1}} + \operatorname{div}_y \boldsymbol{\nabla u_{k+1}} + \operatorname{div}_y \boldsymbol{\nabla_y u_{k+2}}\right)\right.\\ &\quad \left. + \boldsymbol{\nabla} p_k + \boldsymbol{\nabla_y} p_{k+1}\right\} = \boldsymbol{f}\end{aligned}$$

and for the incompressiblity condition

$$\operatorname{div} \boldsymbol{u_\varepsilon}(x) = \varepsilon \operatorname{div}_y \boldsymbol{u_2} + \sum_{k=2}^{+\infty} \varepsilon^k \left\{\operatorname{div} \boldsymbol{u_k} + \operatorname{div}_y \boldsymbol{u_{k+1}}\right\} = 0.$$

Now, we identify the quantities associated with the different orders of ε. The leading order ε^{-1}

of the momentum equation is

$$\begin{cases} \boldsymbol{\nabla_y} p_0 = 0 & \text{in } Y_F, \\ y \to p_0(x,y) & Y-\text{periodic.} \end{cases} \tag{6.5}$$

From (6.5) one deduces that p_0 does not depend on y and there exists a function $p^*(x)$ such that $p_0(x,y) = p^*(x)$. At the next order ε^0 for the momentum equation and ε for the incompressibility condition, we get

$$\begin{cases} -\nu \Delta_y \boldsymbol{u_2} + \boldsymbol{\nabla_y} p_1 = \boldsymbol{f} - \boldsymbol{\nabla} p^* & \text{in } Y_F, \\ \operatorname{div}_y \boldsymbol{u_2} = 0 & \text{in } Y_F, \\ \boldsymbol{u_2} = \boldsymbol{0} & \text{on } \partial Y_S, \\ y \to \boldsymbol{u_2}(x,y),\ p_1(x,y) & Y-\text{periodic.} \end{cases}$$

It follows that $\boldsymbol{u_2}$ and p_1 are linear combinations of the solutions to the following cell Stokes problems: for $i = 1, \dots, d$ find $\boldsymbol{\omega_i} : Y \to \mathbb{R}^d$ and $\pi_i : Y \to \mathbb{R}$, $\mathbb{Z}^d$-periodic and solutions to

$$\begin{cases} -\Delta_y \boldsymbol{\omega_i} + \boldsymbol{\nabla_y} \pi_i = \boldsymbol{e_i} & \text{in } Y_F, \\ \operatorname{div}_y \boldsymbol{\omega_i} = 0 & \text{in } Y_F, \\ \boldsymbol{\omega_i} = \boldsymbol{0} & \text{on } \partial Y_S, \\ \displaystyle\int_Y \pi_i = 0, & \\ y \to \boldsymbol{\omega_i},\ \pi_i & Y-\text{periodic.} \end{cases} \tag{6.6}$$

Velocity and pressure are extended by 0 in Y_S. Using the Einstein summation convention on repeated indices, it follows that

$$\begin{cases} \boldsymbol{u_2}(x,y) &= \dfrac{1}{\nu} \boldsymbol{\omega_i}(y) \left(f_i(x) - \dfrac{\partial p^*}{\partial x_i}(x) \right), \\ p_1(x,y) &= \pi_i(y) \left(f_i(x) - \dfrac{\partial p^*}{\partial x_i}(x) \right). \end{cases} \tag{6.7}$$

Furthermore, averaging on Y_F the ε^2 order term of the incompressibility condition (taking into account the boundary conditions for $\boldsymbol{u_3}$ in (6.11)) yields

$$\operatorname{div} \langle \boldsymbol{u_2} \rangle = 0, \tag{6.8}$$

where $\langle \cdot \rangle$ stands for the average over Y_F, i.e.

$$\langle \boldsymbol{v} \rangle = \frac{1}{|Y_F|} \int_{Y_F} \boldsymbol{v}(y) \mathrm{d}y.$$

Inserting formula (6.7) for $\boldsymbol{u_2}$ in (6.8) leads to the Darcy equation for the pressure p^*. Furthermore, a similar asymptotic analysis can be performed on the boundary $\partial\Omega$ but the homogenized boundary condition turns out to be weaker, namely only the normal component of $\boldsymbol{u}$ vanishes on $\partial\Omega$ [9, 102, 111].

Proposition 6.5. The Darcy velocity is defined as

$$\boldsymbol{u} = \langle \boldsymbol{u_2} \rangle = \frac{1}{\nu} \langle \boldsymbol{\omega_i} \rangle \left(f_i(x) - \frac{\partial p^*}{\partial x_i}(x) \right) \tag{6.9}$$

and the homogenized pressure p^* is the solution to the following Darcy problem

$$\begin{cases} \operatorname{div} A^*(\boldsymbol{f} - \boldsymbol{\nabla} p^*) &= 0 \quad \text{in } \Omega, \\ \boldsymbol{n} \cdot (A^*(\boldsymbol{f} - \boldsymbol{\nabla} p^*)) &= 0 \quad \text{on } \partial\Omega, \end{cases} \tag{6.10}$$

where the constant tensor A^* is defined by its columns

$$A^* \boldsymbol{e_i} = \langle \boldsymbol{\omega_i} \rangle = \frac{1}{|Y_F|} \int_{Y_F} \boldsymbol{\omega_i}(y) \mathrm{d}y.$$

There exists a solution $p^* \in H^1(\Omega)$ of (6.10), which is unique up to an additive constant. Furthermore, under Assumption 6.2 and assuming $\boldsymbol{f} \in C^{1,\alpha}(\overline{\Omega})^d$ with $0 < \alpha < 1$, then $\boldsymbol{\nabla} p^*$ is of class $C^{1,\alpha}(\overline{\Omega})$ and p^* is of class $C^{2,\alpha}(\overline{\Omega})$. Similarly, if $\boldsymbol{f} \in H^2(\Omega)^d$, then $\boldsymbol{\nabla} p^* \in H^2(\Omega)^d$ too.

Proof. Problem (6.10) is well-posed in $H^1(\Omega)/\mathbb{R}$ because it is just a second-order elliptic equation for the pressure p^*, complemented by a Neumann boundary condition (one can check that A^* is positive definite). Recall that $C^{1,\alpha}(\overline{\Omega})$ is the space of functions in $C^1(\overline{\Omega})$ with derivatives which are α-Hölderian. Since $\boldsymbol{f} \in C^{1,\alpha}(\overline{\Omega})$ then it is known that $\boldsymbol{\nabla} p^*$ is of class $C^{1,\alpha}(\overline{\Omega})$ (see for example [83, Theorems 8.33 and 8.34]). Finally, $\boldsymbol{f} \in H^2(\Omega)^d$ implies $\boldsymbol{\nabla} p^* \in H^2(\Omega)^d$ by standard Sobolev regularity theory for elliptic equations. □

At the next order ε for the momentum equation and ε^2 for the incompressibility condition, we get

$$\begin{cases} -\nu \Delta_y \boldsymbol{u_3} + \boldsymbol{\nabla_y} p_2 = \nu \operatorname{div} \boldsymbol{\nabla_y} \boldsymbol{u_2} + \nu \operatorname{div}_y \boldsymbol{\nabla} \boldsymbol{u_2} - \boldsymbol{\nabla} p_1 & \text{in } Y_F, \\ \operatorname{div}_y \boldsymbol{u_3} = -\operatorname{div} \boldsymbol{u_2} & \text{in } Y_F, \\ \boldsymbol{u_3} = \boldsymbol{0} & \text{on } \partial Y_S, \\ y \to \boldsymbol{u_3}(x,y),\ p_2(x,y) & Y - \text{periodic}. \end{cases} \tag{6.11}$$

Substituting the expressions (6.7) of $\boldsymbol{u_2}$ and p_1 in (6.11) leads to

$$-\nu \Delta_y \boldsymbol{u_3} + \boldsymbol{\nabla_y} p_2 = (2 \boldsymbol{\nabla_y} \boldsymbol{\omega_i}(y) - \pi_i(y) I_d) \boldsymbol{\nabla} \left(f_i(x) - \frac{\partial p^*}{\partial x_i}(x) \right).$$

By linearity, we deduce that

$$\begin{cases} \boldsymbol{u_3}(x,y) &= \dfrac{1}{\nu} \boldsymbol{\gamma_{ij}}(y) \dfrac{\partial}{\partial x_j} \left(f_i(x) - \dfrac{\partial p^*}{\partial x_i}(x) \right), \\ p_2(x,y) &= \theta_{ij}(y) \dfrac{\partial}{\partial x_j} \left(f_i(x) - \dfrac{\partial p^*}{\partial x_i}(x) \right), \end{cases}$$

where, for $i, j = 1, \ldots, d$, $\boldsymbol{\gamma_{ij}} : Y \to \mathbb{R}^d$ and $\theta_{ij} : Y \to \mathbb{R}$ are $\mathbb{Z}^d$-periodic solutions to another

cell Stokes problem

$$\begin{cases} -\Delta_y \boldsymbol{\gamma_{ij}} + \boldsymbol{\nabla_y} \theta_{ij} = 2\dfrac{\partial \boldsymbol{\omega_i}}{\partial y_j} - \pi_i \boldsymbol{e_j} & \text{in } Y_F, \\ \operatorname{div}_y \boldsymbol{\gamma_{ij}} = -\boldsymbol{\omega_i} \cdot \boldsymbol{e_j} + \langle \boldsymbol{\omega_i} \cdot \boldsymbol{e_j} \rangle & \text{in } Y_F, \\ \boldsymbol{\gamma_{ij}} = \mathbf{0} & \text{on } \partial Y_S, \\ \displaystyle\int_Y \theta_{ij} = 0, & \\ y \to \boldsymbol{\gamma_{ij}},\ \theta_{ij} & Y - \text{periodic.} \end{cases} \tag{6.12}$$

We recall some notations on Sobolev spaces of periodic functions (see [102, Appendix B.2.2] for details). The L^2 space of Y-periodic functions is

$$L^2_{\#}(Y) = \{\varphi \in L^2_{loc}(\mathbb{R}^d),\ \varphi \text{ is } Y - \text{periodic}\},$$

endowed with the usual scalar product and norm of $L^2(Y)$. It turns out that $L^2_{\#}(Y)$ coincides with $L^2(Y)$. Higher-order Sobolev spaces are defined, for $m \geq 1$, by

$$H^m_{\#}(Y) = \{\varphi \in H^m_{loc}(\mathbb{R}^d),\ \varphi \text{ is } Y - \text{periodic}\},$$

endowed with the usual scalar product and norm of $H^m(Y)$. Let us recall a classical regularity result [44].

Lemma 6.6. Let $m \geq 0$ and Y_F be of class C^{m+2}, consider $\boldsymbol{f}(y) \in H^m_{\#}(Y)^d$ and $g(y) \in H^{m+1}_{\#}(Y)$ with zero mean on Y. The problem

$$\begin{cases} -\Delta_y \boldsymbol{u} + \boldsymbol{\nabla_y} p = \boldsymbol{f}(y) & \text{in } Y_F, \\ \operatorname{div}_y \boldsymbol{u} = g(y) & \text{in } Y_F, \\ \boldsymbol{u} = \mathbf{0} & \text{on } \partial Y_S, \\ y \to \boldsymbol{u}(y),\ p(y) & Y - \text{periodic}, \end{cases} \tag{6.13}$$

admits a unique solution $\boldsymbol{u} \in H^{m+2}_{0,\#}(Y_F)^d$ and $p \in H^{m+1}_{\#}(Y_F)/\mathbb{R}$, where the space $H^k_{0,\#}(Y_F)$ is defined by

$$H^k_{0,\#}(Y_F) = \left\{\varphi \in H^k_{\#}(Y_F) \text{ such that } \varphi = 0 \text{ on } \partial Y_S\right\}.$$

Remark 6.7. In Lemma 6.6, the regularity has to be understood in the unit torus (i.e. Y_F repeated by Y-periodicity in $\mathbb{R}^d$).

Proof. In this proof we propose to show that the problem (6.13) admits a unique solution $(\boldsymbol{u}, p) \in H^1_{0,\#}(Y_F)^d \times L^2(Y_F)$. First, we note that the function g is constrained by the relation $\displaystyle\int_Y g = 0$, since the divergence formula implies

$$\int_Y g = \int_Y \operatorname{div}_y \boldsymbol{v} = \int_{\partial Y} \boldsymbol{v} \cdot \boldsymbol{n} = \int_{\partial Y_S} \boldsymbol{v} \cdot \boldsymbol{n} + \int_{\partial Y \setminus \partial Y_S} \boldsymbol{v} \cdot \boldsymbol{n} = 0.$$

We multiply (6.13) by $\boldsymbol{v} \in H^1_{0,\#}(Y_F)^d$ and by integration by parts, it follows

$$\begin{aligned}
-\int_Y \Delta_y \cdot \boldsymbol{w}\ \boldsymbol{v} + \int_Y \boldsymbol{\nabla_y} p \cdot \boldsymbol{v} &= \int_Y \boldsymbol{\nabla w} \cdot \boldsymbol{\nabla v} - \int_{\partial Y} \boldsymbol{\nabla w} \cdot \boldsymbol{n}\ \boldsymbol{v} - \int_Y p \operatorname{div}_y \boldsymbol{v} + \int_{\partial Y} p \boldsymbol{n} \cdot \boldsymbol{v} \\
&= \int_Y \boldsymbol{\nabla w} \cdot \boldsymbol{\nabla v} - \int_{\partial Y_S} \boldsymbol{\nabla w} \cdot \boldsymbol{n}\ \boldsymbol{v} - \int_{\partial Y \setminus \partial Y_S} \boldsymbol{\nabla w} \cdot \boldsymbol{n}\ \boldsymbol{v} \\
&\quad - \int_Y p \operatorname{div}_y \boldsymbol{v} + \int_{\partial Y_S} p\ \boldsymbol{n} \cdot \boldsymbol{v} + \int_{\partial Y \setminus \partial Y_S} p\ \boldsymbol{n} \cdot \boldsymbol{v} \\
&= \int_Y \boldsymbol{\nabla w} \cdot \boldsymbol{\nabla v} - \int_Y p \operatorname{div}_y \boldsymbol{v} = \int_Y \boldsymbol{f} \cdot \boldsymbol{v},
\end{aligned}$$

where the second terms of the second and third lines is are equal to zero since $\boldsymbol{v}$ vanishes on ∂Y_S, and the third terms of the second and third lines are equal to zero by Y-periodicity of $\boldsymbol{v}$. Consequently, the variational form of (6.13) reads as follows. Find $(\boldsymbol{u}, p) \in H^1_{0,\#}(Y_F) \times L^2(Y_F)$ such that

$$\begin{aligned}
a(\boldsymbol{u}, \boldsymbol{v}) + b(p, \boldsymbol{v}) &= L_1(\boldsymbol{v}), \quad \forall \boldsymbol{v} \in H^1_{0,\#}(Y_F), \\
b(q, \boldsymbol{u}) &= L_2(q), \quad \forall q \in L^2(Y_F),
\end{aligned} \tag{6.14}$$

where

$$\begin{aligned}
a(\boldsymbol{u}, \boldsymbol{v}) &= \int_{Y_F} \boldsymbol{\nabla u} \cdot \boldsymbol{\nabla v}, \qquad b(q, \boldsymbol{u}) = -\int_{Y_F} \operatorname{div}_y \boldsymbol{u}\ q, \\
L_1(\boldsymbol{v}) &= \int_{Y_F} \boldsymbol{f} \cdot \boldsymbol{v}, \qquad L_2(q) = \int_{Y_F} gq.
\end{aligned}$$

It remains to verify that problem (6.14) is well-posed. First the bi-linear form $a(\cdot, \cdot)$ is continuous on $H^1_{0,\#}(Y_F) \times H^1_{0,\#}(Y_F)$, since for all $(\boldsymbol{u}, \boldsymbol{v}) \in H^1_{0,\#}(Y_F) \times H^1_{0,\#}(Y_F)$,

$$\begin{aligned}
|a(\boldsymbol{u}, \boldsymbol{v})| &= \left| \int_{Y_F} \boldsymbol{\nabla u} \cdot \boldsymbol{\nabla v} \right| \leq \int_{Y_F} |\boldsymbol{\nabla u} \cdot \boldsymbol{\nabla v}| \\
&\leq \|\boldsymbol{\nabla u}\|_{L^2(Y_F)} \|\boldsymbol{\nabla v}\|_{L^2(Y_F)} \leq \|\boldsymbol{v}\|_{H^1(Y_F)} \|\boldsymbol{v}\|_{H^1(Y_F)},
\end{aligned}$$

by the Cauchy–Schwarz inequality. The bi-linear form $a(\cdot, \cdot)$ is coercive, since for all $\boldsymbol{v} \in H^1_{0,\#}(Y_F)$,

$$a(\boldsymbol{v}, \boldsymbol{v}) = \int_{Y_F} (\boldsymbol{\nabla v})^2 = \|\boldsymbol{\nabla v}\|^2_{L^2(Y_F)} \leq C \left(\|\boldsymbol{v}\|^2_{L^2(Y_F)} + \|\boldsymbol{\nabla v}\|^2_{L^2(Y_F)} \right),$$

with the Poincaré inequality. The bi-linear form $b(\cdot, \cdot)$ is continuous on $H^1_{0,\#}(Y_F) \times L^2(Y_F)$, since

$$|b(q, \boldsymbol{v})| = \left| -\int_{Y_F} q \operatorname{div}_y \boldsymbol{v} \right| \leq \|\boldsymbol{v}\|_{H^1(Y_F)} \|q\|_{L^2(Y_F)}.$$

The bi-linear form $b(\cdot, \cdot)$ verifies the "inf-sup" condition. Indeed, considering $q \in L^2(Y_F)$ and according to [84], there exists $\boldsymbol{w} \in H^1_0(Y_F)$ such that $q = -\operatorname{div} \boldsymbol{w}$ and $\|\boldsymbol{w}\|_{H^1(Y_F)} \leq C \|q\|_{L^2(Y_F)}$, with C a constant independent of $\boldsymbol{w}$ and g. Consequently, we have

$$b(q, \boldsymbol{w}) = -\int_{Y_F} q \operatorname{div} \boldsymbol{w} = \|q\|_{L^2(Y_F)} \tag{6.15}$$

leading to,

$$\frac{b(q,\boldsymbol{w})}{\|\boldsymbol{w}\|_{H^1(Y_F)}} \geq \frac{1}{C}\frac{b(q,\boldsymbol{w})}{\|q\|_{L^2(Y_F)}} = \frac{1}{C}\|q\|_{L^2(Y_F)}.$$

We conclude that the problem (6.13) admits a unique solution $(\boldsymbol{u},p) \in H^1_{0,\#}(Y_F)^d \times L^2(Y_F)$. □

Remark 6.8. If $\boldsymbol{f}=\boldsymbol{0}$, then $\boldsymbol{u}=\boldsymbol{0}$ and $p=0$ in Y are obvious solution, hence the nullity of the first two terms in the two-scale asymptotic expansion of $\boldsymbol{u_\varepsilon}$.

Using Lemma 6.6, it is clear that (6.6) and (6.12) admit each a unique solution since the right-hand side of the divergence condition has zero-average in Y_F. In particular $\boldsymbol{\omega_i}$ and $\boldsymbol{\gamma_{i,j}}$ are in $H^{m+2}_{0,\#}(Y_F)^d$, and, π_i and $\theta_{i,j}$ are in $H^{m+1}_{\#}(Y_F)/\mathbb{R}$.

The conclusion of this section is that we obtained an approximation of order 3 of the velocity

$$\boldsymbol{u_\varepsilon}(x) \approx \varepsilon^2 \boldsymbol{u_2}\left(x,\frac{x}{\varepsilon}\right) + \varepsilon^3 \boldsymbol{u_3}\left(x,\frac{x}{\varepsilon}\right), \tag{6.16}$$

and of order 1 of the pressure

$$p_\varepsilon(x) \approx p^*(x) + \varepsilon p_1\left(x,\frac{x}{\varepsilon}\right). \tag{6.17}$$

The derivation of (6.16) and (6.17) is merely formal but these approximations are key ingredients in the rigorous proof of our main result of quantitative homogenization.

6.4 A quantitative homogenization result

6.4.1 Main result

Our main result is a quantitative error estimate for the homogenization of Stokes equations in a periodic porous medium, which was loosely announced as Theorem 6.1 in the introduction. A more precise statement of this result is given below.

Theorem 6.9. Let $\boldsymbol{u_\varepsilon}$, p_ε be the solution to the Stokes equations (6.1), p^* be the solution to the Darcy equation (6.10) and $\boldsymbol{u_2}$ be defined by (6.7). Under Assumption 6.2 and assuming that $\boldsymbol{f}$ belongs to $H^2(\Omega)^d \cap C^{1,\alpha}(\overline{\Omega})^d$ with $0<\alpha<1$, there holds:

$$\|p_\varepsilon(x) - p^*(x)\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon^{\frac{1}{2}}\|\boldsymbol{f}(x) - \boldsymbol{\nabla} p^*(x)\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}, \tag{6.18}$$

$$|\boldsymbol{u_\varepsilon}(x) - \varepsilon^2 \boldsymbol{u_2}\left(x,\frac{x}{\varepsilon}\right)|_{H^1(\Omega^\varepsilon)} \leq C\varepsilon^{\frac{3}{2}}\|\boldsymbol{f}(x) - \boldsymbol{\nabla} p^*(x)\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}, \tag{6.19}$$

$$\|\boldsymbol{u_\varepsilon}(x) - \varepsilon^2 \boldsymbol{u_2}\left(x,\frac{x}{\varepsilon}\right)\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon^{\frac{5}{2}}\|\boldsymbol{f}(x) - \boldsymbol{\nabla} p^*(x)\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}, \tag{6.20}$$

where C is independent of ε and the data.

Theorem 6.9 was already proved in the easier case of isolated obstacles in [111] when $d=2$ and in [156] for any dimension. The main novelty of the work present in this chapter is that Theorem 6.9 applies to the physically relevant case of connected obstacles (in dimension $d \geq 3$)

Remark 6.10. In Theorem 6.9, the regularity of $\boldsymbol{f}$ ensures that $\boldsymbol{\nabla} p^*$ also belongs to $H^2(\Omega)^d \cap C^{1,\alpha}(\overline{\Omega})^d$ (see Proposition 6.5). We assume that $\boldsymbol{f} \in H^2(\Omega)^d \cap C^{1,\alpha}(\overline{\Omega})^d$, but all what follows only requires the assumption that the difference $(\boldsymbol{f} - \boldsymbol{\nabla} p^*)$ belongs to $H^2(\Omega)^d \cap C^{1,\alpha}(\overline{\Omega})^d$. The space $H^2(\Omega)^d \cap C^{1,\alpha}(\overline{\Omega})^d$ is equipped with the norm $\|\cdot\|_{H^2 \cap C^{1,\alpha}}$ defined by $\|\cdot\|_{H^2 \cap C^{1,\alpha}} = \|\cdot\|_{H^2} + \|\cdot\|_{C^{1,\alpha}}$.

Remark 6.11. The two estimates for the velocity essentially mean that the relative error for the velocity in the H^1-norm is of the order $\sqrt{\varepsilon}$. Indeed, the velocity is of order ε^2 (be it $\boldsymbol{u_\varepsilon}$ or $\varepsilon^2 \boldsymbol{u_2}$), but its gradient is of order ε (see the derivative rule (6.4)) since they oscillate on the length scale ε. Finally the relative error for the pressure is also of order $\sqrt{\varepsilon}$. Eventually, upon rescaling the velocity, we can write

$$\varepsilon^{-2}\|\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}\|_{L^2(\Omega^\varepsilon)} + \varepsilon^{-1}|\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)} + \|p_\varepsilon - p^*\|_{L^2(\Omega^\varepsilon)} \leq C\sqrt{\varepsilon}\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}.$$

The relative error of order $\sqrt{\varepsilon}$ is worse than the expected ε. It is a classical phenomenon in homogenization [31, 11, 111] due to boundary layer effects caused by the Dirichlet boundary condition on $\partial\Omega$. Technically, this is taken into account by introducing a suitably chosen cut-off function η^ε in the proof. If the boundary layers were absent, which would be the case, for example, with periodic boundary conditions over a rectangular box $\Omega = \prod_{i=1}^d (0, \varepsilon n_i)$ with $n_i \in \mathbb{N}$, the error estimate would improve and give a relative error of order ε

$$\varepsilon^{-2}\|\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}\|_{L^2(\Omega^\varepsilon)} + \varepsilon^{-1}|\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)} + \|p_\varepsilon - p^*\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}.$$

The proof of Theorem 6.9 is given in Section 6.4.5. Beforehand it requires three technical lemmas concerning (i) solving a non-homogeneous divergence equation in a porous domain (Lemma 6.17), (ii) correcting non-homogeneous boundary conditions (Lemma 6.21) and (iii) representing periodic divergence-free fields (Lemma 6.23).

6.4.2 Correction of the divergence condition

Notation 6.12. In what follows, for any $\phi \in H_0^1(\Omega^\varepsilon)$, we also denote by ϕ its extension by zero in $\Omega \setminus \Omega^\varepsilon$, which belongs to $H_0^1(\Omega)$.

Definition 6.13. We note π_i^ε the linear continuous invertible application, composed of a translation and an homothety of ratio $\frac{1}{\varepsilon}$, which maps any cell Y_i^ε onto the reference cell Y

$$\begin{array}{rccl} \pi_i^\varepsilon : & Y_i^\varepsilon & \longrightarrow & Y \\ & x & \longmapsto & y = \frac{x}{\varepsilon} + \text{translation}. \end{array} \tag{6.21}$$

First, we recall the Poincaré inequality in perforated domains (for a proof, see [102]).

Lemma 6.14. Assume that the hypotheses on the unit cell Y (Section 6.2.1) hold true, then there exists a constant $C > 0$, independent of ε, such that, for any $\phi \in H_0^1(\Omega^\varepsilon)$,

$$\|\phi\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon|\phi|_{H^1(\Omega^\varepsilon)}.$$

We now recall a result [160, 9] on a restriction operator for velocities defined on the full domain Ω which are restricted to the porous domain Ω^ε.

Lemma 6.15. Assume that the hypotheses on the unit cell Y (Section 6.2.1) hold true. Then there exists a linear continuous operator R_ε such that

1. $R_\varepsilon \in \mathcal{L}\left(H_0^1(\Omega)^d; H_0^1(\Omega^\varepsilon)^d\right)$;
2. $\boldsymbol{u}|_{\Omega^\varepsilon} \in H_0^1(\Omega^\varepsilon)^d$ implies $R_\varepsilon \boldsymbol{u} = \boldsymbol{u}|_{\Omega^\varepsilon}$ in Ω^ε;
3. $\operatorname{div} \boldsymbol{u} = 0$ in $\Omega \setminus \Omega^\varepsilon$ implies $\operatorname{div}(R_\varepsilon \boldsymbol{u}) = \operatorname{div} \boldsymbol{u}$ in Ω^ε;
4. there exists a constant C, which does not depend on ε, such that, for any $\boldsymbol{u} \in H_0^1(\Omega)^d$, we have

$$\|R_\varepsilon \boldsymbol{u}\|_{L^2(\Omega^\varepsilon)} + \varepsilon\|\boldsymbol{\nabla}(R_\varepsilon \boldsymbol{u})\|_{L^2(\Omega^\varepsilon)} \leq C\left(\|\boldsymbol{u}\|_{L^2(\Omega)} + \varepsilon\|\boldsymbol{\nabla}\boldsymbol{u}\|_{L^2(\Omega)}\right).$$

Remark 6.16. Lemma 6.15 was first proved in [160] for the case of a solid part Y_S strictly included in the cell Y (see Figure 6.3a) and was then extended in [9] for the more general case of a solid part Y_S which touches the boundary of Y (see Figure 6.3b).

Proof. Since we shall need specific properties of the operator R_ε in the sequel, we briefly recall how to construct this operator. We start with the simpler case of isolated obstacles, namely when the solid part Y_S is strictly included in the cell Y (the general case is presented afterwards). Consider a domain Y_M, strictly contained in Y, such that ∂Y_M is smooth and encloses Y_S as presented in Figure 6.5.

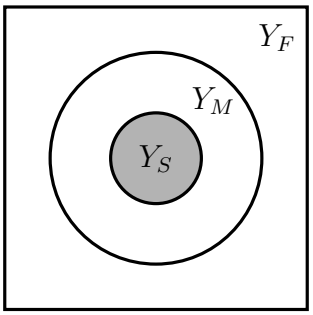


Figure 6.5: Schematic description of the domain Y_M around Y_S.

Then for a given $\boldsymbol{u} \in H^1(Y)^d$, there exists $\boldsymbol{v} \in H^1(Y_M \setminus Y_S)^d$, $q \in L^2(Y_M \setminus Y_S)/\mathbb{R}$ satisfying

$$\left\{\begin{array}{rcll} -\Delta \boldsymbol{v} + \boldsymbol{\nabla} q &=& -\Delta \boldsymbol{u} & \text{in } Y_M \setminus Y_S, \\ \operatorname{div} \boldsymbol{v} &=& \operatorname{div} \boldsymbol{u} + \dfrac{1}{|Y_M \setminus Y_S|} \displaystyle\int_{Y_S} \operatorname{div} \boldsymbol{u} & \text{in } Y_M \setminus Y_S, \\ \boldsymbol{v} &=& \boldsymbol{u} & \text{on } \partial Y_M, \\ \boldsymbol{v} &=& \boldsymbol{0} & \text{on } \partial Y_S. \end{array}\right.$$

This problem is well-posed, since the compatibility condition can easily be checked on the data, namely

$$\int_{Y_M \setminus Y_S} \operatorname{div} \boldsymbol{v} = \int_{\partial (Y_M \setminus Y_S)} \boldsymbol{v} \cdot \boldsymbol{n}.$$

Now, a restriction operator R is constructed as follow: for any $\boldsymbol{u} \in H^1(Y)^d$

$$R\boldsymbol{u} = \left\{\begin{array}{l} \boldsymbol{u} \text{ in } Y \setminus Y_M, \\ \boldsymbol{v} \text{ in } Y_M \setminus Y_S, \\ \boldsymbol{0} \text{ in } Y_S. \end{array}\right.$$

Clearly, R is linear continuous from $H^1(Y)^d$ to its subspace of functions vanishing on Y_S, satisfying

$$\|R\boldsymbol{u}\|_{H^1(Y)} \leq C \|\boldsymbol{u}\|_{H^1(Y)}, \tag{6.22}$$

and such that, if $\boldsymbol{u} = \boldsymbol{0}$ in Y_S, then $R\boldsymbol{u} = \boldsymbol{u}$, and, if $\operatorname{div} \boldsymbol{u} = 0$ in Y_S, then $\operatorname{div} R\boldsymbol{u} = \operatorname{div} \boldsymbol{u}$ in Y. Then the operator R_ε is defined by applying the mapping (6.21) to the operator R and, with standard scaling properties, it comes from (6.22)

$$\|R_\varepsilon \boldsymbol{u}\|^2_{L^2(\Omega^\varepsilon)} + \varepsilon^2 \|\boldsymbol{\nabla}(R_\varepsilon \boldsymbol{u})\|^2_{L^2(\Omega^\varepsilon)} \leq C \left(\|\boldsymbol{u}\|^2_{L^2(\Omega)} + \varepsilon^2 \|\boldsymbol{\nabla} \boldsymbol{u}\|^2_{L^2(\Omega)} \right),$$

which implies that R_ε would satisfy all four properties of Lemma 6.15.

We now turn to the case of connected obstacles, namely when Y_S touches the cell boundary ∂Y as presented in Figure 6.3b. The construction of R_ε is slightly more involved. First, we define a linear continuous operator $Q_\varepsilon \in \mathcal{L}\left(H^1_0(\Omega)^d; H^1_0(\Omega^\varepsilon)^d\right)$ by

$$\left\{\begin{array}{rcll} -\Delta(Q_\varepsilon \boldsymbol{u}) &=& -\Delta \boldsymbol{u} & \text{in } \Omega^\varepsilon \cap C_1^\varepsilon, \\ Q_\varepsilon \boldsymbol{u} &=& \boldsymbol{0} & \text{on } \partial\Omega^\varepsilon \cap C_1^\varepsilon, \\ Q_\varepsilon \boldsymbol{u} &=& \boldsymbol{u} & \text{on } \partial C_1^\varepsilon, \\ Q_\varepsilon \boldsymbol{u} &=& \boldsymbol{u} & \text{in } \Omega \setminus C_1^\varepsilon, \end{array}\right.$$

where C_1^ε is defined by Assumption 6.2 as the union of entire cells inside Ω. Clearly, $Q_\varepsilon \boldsymbol{u} = \boldsymbol{u}$ if and only if $\boldsymbol{u}|_{\Omega^\varepsilon} \in H^1_0(\Omega^\varepsilon)^d$. Thanks to the definition of C_1^ε and C_2^ε, there exists a smooth cut-off function ζ^ε such that $\zeta^\varepsilon(x) = 1$ on ∂C_1^ε, $\zeta^\varepsilon(x) = 0$ in C_2^ε and $\|\boldsymbol{\nabla} \zeta^\varepsilon\|_{L^\infty(C_1^\varepsilon)} \leq \frac{C}{\varepsilon}$. One can rewrite (6.4.2) for the new unknown $\boldsymbol{w^\varepsilon} = Q_\varepsilon \boldsymbol{u} - \zeta^\varepsilon \boldsymbol{u}$ as

$$\left\{\begin{array}{rcll} -\Delta \boldsymbol{w^\varepsilon} &=& -\Delta((1-\zeta^\varepsilon)\boldsymbol{u}) & \text{in } \Omega^\varepsilon \cap C_1^\varepsilon, \\ \boldsymbol{w^\varepsilon} &=& \boldsymbol{0} & \text{on } \partial\Omega^\varepsilon \cap C_1^\varepsilon, \\ \boldsymbol{w^\varepsilon} &=& \boldsymbol{0} & \text{on } \partial C_1^\varepsilon. \end{array}\right. \tag{6.23}$$

Multiplying (6.23) by $\boldsymbol{w^\varepsilon}$ leads to an a priori estimate, which, combined with the Poincaré inequality (Lemma 6.14), implies that there exists a constant C, which does not depend on ε, such that, for any $\boldsymbol{u} \in H^1_0(\Omega)^d$, we have

$$\|Q_\varepsilon \boldsymbol{u}\|_{L^2(\Omega^\varepsilon)} + \varepsilon \|\boldsymbol{\nabla}(Q_\varepsilon \boldsymbol{u})\|_{L^2(\Omega^\varepsilon)} \leq C \left(\|\boldsymbol{u}\|_{L^2(\Omega)} + \varepsilon \|\boldsymbol{\nabla u}\|_{L^2(\Omega)} \right).$$

Then, for each cell Y_i^ε, we define a linear operator Q_i, from $H^1(Y)^d$ to its subspace of functions vanishing on Y_S, as follows: for any $\boldsymbol{u} \in H^1_0(\Omega)^d$,

$$\left(Q_i \left(\boldsymbol{u} \circ (\pi_i^\varepsilon)^{-1} \right) \right) \circ \pi_i^\varepsilon = Q_\varepsilon \boldsymbol{u} \text{ in } Y_i^\varepsilon. \tag{6.24}$$

Recall from Assumption 6.2 that we defined two sets of indices $\mathcal{I}_1$ and $\mathcal{I}_2$, corresponding to the cells Y_i^ε which cover C_1^ε and C_2^ε, respectively. Define a third set $\mathcal{I}_0 = \{i \in \mathbb{Z}^d \ s.t. \ Y_i^\varepsilon \cap \Omega \neq \emptyset\}$, so that $\mathcal{I}_2 \subset \mathcal{I}_1 \subset \mathcal{I}_0$. We now define a family of restriction operators R_i, depending on which set $\mathcal{I}_2, \mathcal{I}_1, \mathcal{I}_0$ the index i belongs to.

For $i \in \mathcal{I}_2$ (meaning that the corresponding cell contains a solid obstacle), R_i is defined, for any $\boldsymbol{u} \in H^1(Y)^d$, by

$$\left\{ \begin{array}{rcll} -\Delta R_i \boldsymbol{u} + \boldsymbol{\nabla} p &=& -\Delta \boldsymbol{u} & \text{in } Y_F, \\ \operatorname{div} R_i \boldsymbol{u} &=& \operatorname{div} \boldsymbol{u} + \dfrac{1}{|Y_F|} \displaystyle\int_{Y_S} \operatorname{div} \boldsymbol{u} & \text{in } Y_F, \\ R_i \boldsymbol{u} &=& Q_i \boldsymbol{u} + \phi_k \left(\int_{\Sigma_k} (\boldsymbol{u} - Q_i \boldsymbol{u}) \cdot \boldsymbol{e_k} \right) \boldsymbol{e_k} & \text{on } \Sigma_k \cap \overline{Y_F}, \\ R_i \boldsymbol{u} &=& \boldsymbol{0} & \text{on } \partial Y_S, \end{array} \right. \tag{6.25}$$

where $(\Sigma_k)_{1 \leq |k| \leq d}$ are the $2d$ faces of the unit cube (Σ_k and Σ_{-k} are opposite faces), with unit normal vector $\boldsymbol{e_k}$, and $(\phi_k)_{1 \leq |k| \leq d}$ is a family of smooth, non-negative functions, compactly supported in the periodic repetition of $\overline{Y_F}$, satisfying

$$\int_{\Sigma_k} \phi_k = 1, \quad \phi_{k|_{\Sigma_k}} \equiv \phi_{-k|_{\Sigma_{-k}}} \text{ and } \phi_{k|_{\Sigma_{k'}}} \equiv 0 \text{ if } k \neq k'.$$

The existence of such a family of functions ϕ_k is guaranteed by the assumptions on the unit cell in Section 6.2.1, which imply that the surface measure of $\Sigma_k \cap \overline{Y_F}$ is non-zero (see [9] for more details). Problem (6.25) is well-posed, since the data of the trace of $R_i \boldsymbol{u}$ belongs to $H^{1/2}(Y_F)^d$ and the compatibility condition on the divergence data is satisfied, namely

$$\int_{Y_F} \operatorname{div} R_i \boldsymbol{u} = \int_{\partial Y_F} R_i \boldsymbol{u} \cdot \boldsymbol{n}.$$

For $i \in \mathcal{I}_1 \setminus \mathcal{I}_2$ (meaning that the corresponding cell does not contain a solid obstacle but one of its face may touch an obstacle), R_i is defined, for any $\boldsymbol{u} \in H^1(Y)^d$, by

$$\left\{ \begin{array}{rcll} -\Delta R_i \boldsymbol{u} + \boldsymbol{\nabla} p &=& -\Delta \boldsymbol{u} & \text{in } Y, \\ \operatorname{div} R_i \boldsymbol{u} &=& \operatorname{div} \boldsymbol{u} & \text{in } Y, \\ R_i \boldsymbol{u} &=& Q_i \boldsymbol{u} + \phi_k \left(\displaystyle\int_{\Sigma_k} (\boldsymbol{u} - Q_i \boldsymbol{u}) \cdot \boldsymbol{e_k} \right) \boldsymbol{e_k} & \text{on } \Sigma_k, \\ R_i \boldsymbol{u} &=& \boldsymbol{u} & \text{on } \tilde{\Sigma}_k, \end{array} \right. \tag{6.26}$$

where Σ_k are the faces of the cell which touch an obstacle and $\tilde{\Sigma}_k$ are the faces of the cell which do not touch any obstacle. Problem (6.26) is well-posed, since the data of the trace of $R_i\boldsymbol{u}$ belongs to $H^{1/2}(Y)^d$ and the compatibility condition on the divergence data is satisfied, namely

$$\int_Y \operatorname{div} R_i\boldsymbol{u} = \int_{\partial Y} R_i\boldsymbol{u}\cdot\boldsymbol{n}.$$

For $i \in \mathcal{I}_0 \setminus \mathcal{I}_1$ (implying that the corresponding cell does not contain a solid obstacle and none of its face touches an obstacle), R_i is simply defined as $R_i\boldsymbol{u} = \boldsymbol{u}$. Note that this case corresponds to cells Y_i^ε which are cut by the boundary $\partial\Omega$. Inside such a cell, with a possibly wild geometry, it is not clear that a Stokes problem of the type of (6.26) satisfies a priori estimates, independent of the geometry (this is precisely why the two sets of indices $\mathcal{I}_1$ and $\mathcal{I}_2$ have been introduced). A similar estimate to (6.22) holds true for all operators R_i with a constant C independent of i and ε.

Finally, the restriction operator $R_\varepsilon \in \mathcal{L}\left(H_0^1(\Omega)^d; H_0^1(\Omega^\varepsilon)^d\right)$ is defined, for any $\boldsymbol{u} \in H_0^1(\Omega)^d$, by

$$R_\varepsilon\boldsymbol{u} = \left(R_i\left(\boldsymbol{u}\circ(\pi_i^\varepsilon)^{-1}\right)\right)\circ\pi_i^\varepsilon \text{ in } Y_i^\varepsilon.$$

It is easy again to check that R_ε satisfies all four properties of Lemma 6.15. Note however that, here, the definition of R_ε is not local in each cell since they are all coupled through the operator Q_ε. □

We now state a key lemma for our analysis. It was already presented in [111] in two dimensions for the case when Y_S is strictly included in Y. It was also stated in [48, Theorem 14.2] for the general case of any Y_S and any space dimension but its proof was rather technical, decomposed in several exercises. Here, we present a new and detailed proof, using different arguments.

Lemma 6.17. For any $q \in L_0^2(\Omega^\varepsilon)$ there exists $\boldsymbol{v_{d\varepsilon}} \in H_0^1(\Omega^\varepsilon)^d$ such that

$$\operatorname{div}\boldsymbol{v_{d\varepsilon}} = q \text{ in } \Omega^\varepsilon \text{ and } |\boldsymbol{v_{d\varepsilon}}|_{H^1(\Omega^\varepsilon)} \leq \frac{C}{\varepsilon}\|q\|_{L^2(\Omega^\varepsilon)} \tag{6.27}$$

where $C > 0$ is a constant independent of ε.

Proof. For any $q \in L_0^2(\Omega^\varepsilon)$ we still denote by q its extension by zero inside the obstacles B^ε. One can easily check that its average on Ω still vanishes and thus $q \in L_0^2(\Omega)$ with $q = 0$ in B^ε. Using [84, corollary 2.4], there exists a (non unique) $\boldsymbol{w} \in H_0^1(\Omega)^d$ such that

$$\operatorname{div}\boldsymbol{w} = q \text{ in } \Omega \quad \text{and} \quad \|\boldsymbol{w}\|_{H^1(\Omega)} \leq C\|q\|_{L^2(\Omega)}.$$

Since $q = 0$ in B^ε, we have

$$\operatorname{div}\boldsymbol{w} = 0 \text{ in } \Omega\setminus\Omega^\varepsilon.$$

Using the restriction operator R_ε defined in Lemma 6.15, we have

$$\operatorname{div}(R_\varepsilon\boldsymbol{w}) = q \text{ in } \Omega^\varepsilon.$$

Considering $\varepsilon < 1$ in the fourth property of Lemma 6.15, it follows

$$\|R_\varepsilon \boldsymbol{w}\|_{L^2(\Omega^\varepsilon)} + \varepsilon \|\boldsymbol{\nabla} R_\varepsilon \boldsymbol{w}\|_{L^2(\Omega^\varepsilon)} \leq C \|\boldsymbol{w}\|_{H^1(\Omega)}$$

and consequently

$$\varepsilon \|\boldsymbol{\nabla}(R_\varepsilon \boldsymbol{w})\|_{L^2(\Omega^\varepsilon)} \leq C \|\boldsymbol{w}\|_{H^1(\Omega)} \leq C \|q\|_{L^2(\Omega^\varepsilon)}.$$

Therefore, $\boldsymbol{v_{d\varepsilon}} = R_\varepsilon \boldsymbol{w}$ is a solution to (6.27). □

6.4.3 Correction of the boundary condition

In this section, we build a divergence-free field which satisfies a given boundary condition. Recall the definition of the curl* operator, proposed in [62].

Definition 6.18. The operator curl* is defined as

$$\begin{array}{rccl} \text{curl}^* : & C^1(\mathbb{R}^d)^{d(d-1)/2} & \longrightarrow & C^0(\mathbb{R}^d)^d \\ & \boldsymbol{\mathcal{A}} & \longmapsto & \left(\displaystyle\sum_{j=1}^{i-1} \frac{\partial \mathcal{A}_{ji}}{\partial x_j} - \sum_{j=i+1}^{d} \frac{\partial \mathcal{A}_{ij}}{\partial x_j} \right)_{1 \leq i \leq d} \end{array}$$

where $\boldsymbol{\mathcal{A}} = (\mathcal{A}_{ij})_{1 \leq i < j \leq d} \in \mathbb{R}^{d(d-1)/2}$.

A simple computation shows that the image of this curl* operator is divergence-free, namely $\operatorname{div} \text{curl}^* \boldsymbol{\mathcal{A}} = 0$ for any $\boldsymbol{\mathcal{A}} \in C^1(\mathbb{R}^d)^{d(d-1)/2}$. This is an expected property since curl* is an extension, to any dimension d, of the standard curl operator in 2-d and 3-d. We now state a preliminary lemma, also adapted from [62] .

Lemma 6.19. Let $\Omega \subset \mathbb{R}^d$ be a bounded connected open set with a smooth boundary of class $C^{3,\alpha}$, for some $0 < \alpha < 1$. Let $\boldsymbol{g}$ in $C^{1,\alpha}\left(\overline{\Omega}\right)^d$ with $\boldsymbol{g} \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. Then there exists $\boldsymbol{\mathcal{A}}$ in $C^{2,\alpha}(\overline{\Omega})^{d(d-1)/2}$ such that

$$\text{curl}^*(\boldsymbol{\mathcal{A}}) = \boldsymbol{g} \text{ on } \partial\Omega \text{ and } \boldsymbol{\mathcal{A}} = \boldsymbol{0} \text{ on } \partial\Omega.$$

In other words, the vector field curl$^*(\boldsymbol{\mathcal{A}})$ is divergence-free and its trace on $\partial\Omega$ coincides with $\boldsymbol{g}$.

Proof. [62, Lemma 1] ensures that there exists $\boldsymbol{\mathcal{A}} \in C^{2,\alpha}(\overline{\Omega})^{d(d-1)/2}$, with $\boldsymbol{\mathcal{A}} = \boldsymbol{0}$ on $\partial\Omega$, such that

$$\boldsymbol{\nabla} \mathcal{A}_{ij} = (g_j n_i - g_i n_j) \boldsymbol{n} \text{ on } \partial\Omega,$$

where $\boldsymbol{n}$ is the outward unit normal vector, g_i and n_i are respectively the ith coordinates of $\boldsymbol{g}$ and $\boldsymbol{n}$ in the canonical basis of $\mathbb{R}^d$. It remains to check that curl$^*(\boldsymbol{\mathcal{A}}) = \boldsymbol{g}$ on $\partial\Omega$. Indeed, we

have, on $\partial\Omega$,

$$(\mathrm{curl}^*\boldsymbol{\mathcal{A}})_i = \sum_{j=1}^{i-1}(g_i n_j - g_j n_i)n_j - \sum_{j=i+1}^{d}(g_j n_i - g_i n_j)n_j$$
$$= \sum_{j=1,\ j\neq i}^{d} g_i n_j^2 - \sum_{j=1,\ j\neq i}^{d} g_j n_j n_i = \sum_{j=1}^{d} g_i n_j^2 = g_i,$$

where we used the assumption that $\boldsymbol{g}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$. □

The next step is to localize in the vicinity of the boundary $\partial\Omega$ the divergence-free field which lifts a given boundary condition. To do this, we introduce the following classical cut-off function as in [31, 111].

Definition 6.20 (Cut-off function)**.** For a small enough constant $c > 0$, define a neighborhood of the boundary $\partial\Omega$, denoted by $O^\varepsilon := \{x \in \Omega : \mathrm{dist}(x, \partial\Omega) \leq c\varepsilon\}$ (band of width $c\varepsilon$ near $\partial\Omega$). Then we choose a cut-off function $\theta^\varepsilon \in C^\infty(\Omega)$ such that $\theta^\varepsilon(x) = 1$ in $O^{\varepsilon/2}$, $\theta^\varepsilon(x) = 0$ in $\Omega \setminus O^\varepsilon$, and

$$\|\boldsymbol{\nabla}\theta^\varepsilon\|_{L^\infty(\Omega)} \leq \frac{C}{\varepsilon}, \quad \|\boldsymbol{\nabla}^2\theta^\varepsilon\|_{L^\infty(\Omega)} \leq \frac{C}{\varepsilon^2}.$$

The main result of this section is the following lemma.

Lemma 6.21. For any $\boldsymbol{g} \in C^{1,\alpha}\left(\overline{\Omega}\right)^d$ with $\mathrm{div}\,\boldsymbol{g} = 0$ in Ω, $\boldsymbol{g}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$, there exists $\boldsymbol{v_{b\varepsilon}} \in H^1(\Omega)^d$ such that *supp* $\boldsymbol{v_{b\varepsilon}} \subset O^\varepsilon := \{x \in \Omega : \mathrm{dist}(x, \partial\Omega) < c\varepsilon\}$ and

$$\boldsymbol{v_{b\varepsilon}} = \boldsymbol{g} \text{ on } \partial\Omega, \ \ \mathrm{div}\,\boldsymbol{v_{b\varepsilon}} = 0 \text{ in } \Omega \text{ and } |\boldsymbol{v_{b\varepsilon}}|_{H^1(\Omega)} \leq \frac{C}{\sqrt{\varepsilon}}\|\boldsymbol{g}\|_{C^{1,\alpha}(\Omega)}$$

where $C > 0$ is a constant independent of ε.

Remark 6.22. Lemma 6.21 was already proved in [111] in two dimensions. Here, we extend this result to any dimension.

Proof. Let $\boldsymbol{g} \in C^{1,\alpha}\left(\overline{\Omega}\right)^d$ with $\mathrm{div}\,\boldsymbol{g} = 0$ in Ω, $\boldsymbol{g}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$. By virtue of Lemma 6.19, there exists a function $\boldsymbol{\mathcal{A}} \in C^{2,\alpha}(\overline{\Omega})^{d(d-1)/2}$, such that $\mathrm{curl}^*(\boldsymbol{\mathcal{A}}) = \boldsymbol{g}$ on $\partial\Omega$. Define

$$\boldsymbol{v_{b\varepsilon}} = \mathrm{curl}^*(\theta^\varepsilon\boldsymbol{\mathcal{A}}),$$

where θ^ε is the cut-off function introduced in Definition 6.20. Obviously, $\mathrm{div}\,\boldsymbol{v_{b\varepsilon}} = 0$ in Ω. Then, $\boldsymbol{v_{b\varepsilon}} = \boldsymbol{g}$ on $\partial\Omega$ since, in the vicinity $O^{\varepsilon/2}$ of $\partial\Omega$, we recall that $\theta^\varepsilon \equiv 1$. Finally, it remains to bound $|\boldsymbol{v_{b\varepsilon}}|_{H^1(\Omega)}$. We write

$$\mathrm{curl}^*(\theta^\varepsilon\boldsymbol{\mathcal{A}}) = \theta^\varepsilon\mathrm{curl}^*(\boldsymbol{\mathcal{A}}) + \Gamma(\boldsymbol{\mathcal{A}}, \boldsymbol{\nabla}\theta^\varepsilon)$$

where $\Gamma(\boldsymbol{\mathcal{A}}, \boldsymbol{\nabla}\theta^\varepsilon)$ is defined as

$$\Gamma(\boldsymbol{\mathcal{A}}, \boldsymbol{\nabla}\theta^\varepsilon)_i = \sum_{j=1}^{i-1} \mathcal{A}_{ji} \frac{\partial \theta^\varepsilon}{\partial x_j} - \sum_{j=i+1}^{d} \mathcal{A}_{ij} \frac{\partial \theta^\varepsilon}{\partial x_j}.$$

It follows therefore

$$\begin{aligned} |\boldsymbol{v_{b\varepsilon}}|_{H^1(\Omega)} &\leq \|\theta^\varepsilon \boldsymbol{\nabla}\mathrm{curl}^*(\boldsymbol{\mathcal{A}})\|_{L^2(O^\varepsilon)} + \|\boldsymbol{\nabla}\theta^\varepsilon \mathrm{curl}^*(\boldsymbol{\mathcal{A}})\|_{L^2(O^\varepsilon)} + \|\boldsymbol{\nabla}\Gamma(\boldsymbol{\mathcal{A}}, \boldsymbol{\nabla}\theta^\varepsilon)\|_{L^2(O^\varepsilon)} \\ &\leq C\left(\sqrt{\varepsilon}\|\boldsymbol{\nabla}\boldsymbol{\nabla}\boldsymbol{\mathcal{A}}\|_{L^\infty(O^\varepsilon)} + \frac{1}{\sqrt{\varepsilon}}\|\boldsymbol{\nabla}\boldsymbol{\mathcal{A}}\|_{L^\infty(O^\varepsilon)}\right) + \|\boldsymbol{\nabla}\Gamma(\boldsymbol{\mathcal{A}}, \boldsymbol{\nabla}\theta^\varepsilon)\|_{L^2(O^\varepsilon)}, \end{aligned}$$

since $meas(O^\varepsilon) \leq c\varepsilon$, and

$$\begin{aligned} \|\boldsymbol{\nabla}\Gamma(\boldsymbol{\mathcal{A}}, \boldsymbol{\nabla}\theta^\varepsilon)\|_{L^2(O^\varepsilon)} &\leq C\left(\|\boldsymbol{\nabla}\theta^\varepsilon\|_{L^\infty(O^\varepsilon)}\|\boldsymbol{\nabla}\boldsymbol{\mathcal{A}}\|_{L^2(O^\varepsilon)} + \|\boldsymbol{\nabla}\boldsymbol{\nabla}\theta^\varepsilon\|_{L^\infty(O^\varepsilon)}\|\boldsymbol{\mathcal{A}}\|_{L^2(O^\varepsilon)}\right) \\ &\leq C\left(\frac{1}{\varepsilon}\|\boldsymbol{\nabla}\boldsymbol{\mathcal{A}}\|_{L^2(O^\varepsilon)} + \frac{1}{\varepsilon^2}\|\boldsymbol{\mathcal{A}}\|_{L^2(O^\varepsilon)}\right) \\ &\leq C\left(\frac{1}{\sqrt{\varepsilon}}\|\boldsymbol{\nabla}\boldsymbol{\mathcal{A}}\|_{L^\infty(\Omega)} + \frac{1}{\varepsilon\sqrt{\varepsilon}}\|\boldsymbol{\mathcal{A}}\|_{L^\infty(O^\varepsilon)}\right). \end{aligned}$$

We observe now that any point $x \in O^\varepsilon$ can be connected to a point $x' \in \partial\Omega$ by a segment of length not greater than $c\varepsilon$ lying in O^ε. Recalling that $\boldsymbol{\mathcal{A}}(x') = 0$ and using a Taylor expansion at order 1 with integral remainder gives $|\boldsymbol{\mathcal{A}}|(x) \leq c\varepsilon\|\boldsymbol{\nabla}\boldsymbol{\mathcal{A}}(z)\|_{L^\infty(O^\varepsilon)}$. Consequently, $\|\boldsymbol{\mathcal{A}}\|_{L^\infty(O^\varepsilon)} \leq c\varepsilon\|\boldsymbol{\nabla}\boldsymbol{\mathcal{A}}\|_{L^\infty(O^\varepsilon)}$ and

$$\|\boldsymbol{\nabla}\Gamma(\boldsymbol{\mathcal{A}}, \boldsymbol{\nabla}\theta^\varepsilon)\|_{L^2(O^\varepsilon)} \leq \frac{C}{\sqrt{\varepsilon}}\|\boldsymbol{\nabla}\boldsymbol{\mathcal{A}}\|_{L^\infty(O^\varepsilon)},$$

which concludes the proof. □

6.4.4 Representation of divergence-free fields

A key ingredient in the proof of Theorem 6.9 is a representation of periodic divergence-free and zero-mean fields as the image of a differential operator, the range of which is always divergence-free. This representation is given by the following.

Lemma 6.23. Let $\boldsymbol{g} \in L^2_{\#}(Y)^d$ be periodic, satisfying

$$\int_Y \boldsymbol{g} \mathrm{d}y = 0, \quad \operatorname{div}_y \boldsymbol{g} = 0 \text{ in } Y.$$

There exists a (non-unique) periodic solution $\boldsymbol{\varphi} \in H^1_{\#}(Y)^d \cap H^2_{\#}(Y)^d$ to the problem

$$\begin{cases} -\Delta_y \boldsymbol{\varphi} + \boldsymbol{\nabla}_y \operatorname{div}_y \boldsymbol{\varphi} = \boldsymbol{g} & \text{in } Y, \\ y \to \boldsymbol{\varphi}(y) & Y - \text{periodic}, \end{cases} \tag{6.28}$$

which satisfies, for some constant $C > 0$,

$$\|\boldsymbol{\varphi}\|_{H^2_{\#}(Y)} \leq C \|\boldsymbol{g}\|_{L^2_{\#}(Y)}.$$

Furthermore, for any $\boldsymbol{\psi} \in H^2_{\#}(Y)^d$, it holds that

$$\operatorname{div}_y \left(-\Delta_y \boldsymbol{\psi} + \boldsymbol{\nabla}_y \operatorname{div}_y \boldsymbol{\psi}\right) = 0 \quad \text{in } Y.$$

Remark 6.24. More generally, the identity

$$\operatorname{div} \left(-\Delta \boldsymbol{\psi} + \boldsymbol{\nabla} \operatorname{div} \boldsymbol{\psi}\right) = 0 \quad \text{in } \mathbb{R}^d$$

holds true for any vector field $\boldsymbol{\psi} \in (H^2(\mathbb{R}^d))^d$.

Proof. Using Fourier series the periodic solution $\boldsymbol{\varphi}$ and the periodic right-hand side $\boldsymbol{g}$ are written as

$$\boldsymbol{\varphi}(y) = \sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} \widehat{\boldsymbol{\varphi}}(\boldsymbol{\xi}) e^{2i\pi \boldsymbol{\xi} \cdot y}, \quad \boldsymbol{g}(y) = \sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} \widehat{\boldsymbol{g}}(\boldsymbol{\xi}) e^{2i\pi \boldsymbol{\xi} \cdot y},$$

where $\widehat{\boldsymbol{g}}(\mathbf{0}) = \mathbf{0}$ and $\widehat{\boldsymbol{g}}(\boldsymbol{\xi}) \cdot \boldsymbol{\xi} = 0$ for any $\boldsymbol{\xi} \in \mathbb{Z}^d$, because of the assumptions on $\boldsymbol{g}$. Then, (6.28) becomes

$$4\pi^2 |\boldsymbol{\xi}|^2 \widehat{\boldsymbol{\varphi}} - 4\pi^2 (\boldsymbol{\xi} \cdot \widehat{\boldsymbol{\varphi}}) \boldsymbol{\xi} = \widehat{\boldsymbol{g}}. \tag{6.29}$$

By writing (6.29) under matrix form, we have

$$A \widehat{\boldsymbol{\varphi}} = \widehat{\boldsymbol{g}}, \quad A(\boldsymbol{\xi}) = 4\pi^2 \left(|\boldsymbol{\xi}|^2 I_d - \boldsymbol{\xi} \otimes \boldsymbol{\xi}\right).$$

We easily see that $\ker(A) = \operatorname{span}(\boldsymbol{\xi})$. Since A is symmetric, we have $\operatorname{Im}(A) = \ker(A)^\perp = \{\boldsymbol{u} \in \mathbb{R}^d \ s.t. \ \boldsymbol{u} \cdot \boldsymbol{\xi} = 0\}$ and $\widehat{\boldsymbol{g}} \in \ker(A)^\perp$. Now, if we restrict the matrix A to its image, it is reduced to $|\boldsymbol{\xi}|^2 I_d$. Consequently, its inverse is simply given by $|\boldsymbol{\xi}|^{-2} I_d$ which gives the following solution

$$\widehat{\boldsymbol{\varphi}} = \frac{\widehat{\boldsymbol{g}}}{4\pi^2 |\boldsymbol{\xi}|^2} \text{ for any } \boldsymbol{\xi} \neq 0,$$

which is unique, up to the addition of a multiple of $\boldsymbol{\xi}$. In other words, we obtained a solution $\boldsymbol{\varphi}(y)$

of (6.28) which is unique up to the addition of a constant and of a periodic gradient. Now, recalling that $\boldsymbol{\xi}$ is an integer vector, we note that

$$|\widehat{\boldsymbol{\varphi}}| = \frac{|\widehat{\boldsymbol{g}}|}{4\pi^2|\boldsymbol{\xi}|^2} \leq C|\widehat{\boldsymbol{g}}|, \quad |\boldsymbol{\xi}\widehat{\boldsymbol{\varphi}}| \leq C|\widehat{\boldsymbol{g}}|, \quad |\boldsymbol{\xi}\boldsymbol{\xi}\widehat{\boldsymbol{\varphi}}| \leq C|\widehat{\boldsymbol{g}}|,$$

which allows to conclude that $\boldsymbol{\varphi}$ belongs to $H^2_{\#}(Y)^d$. □

In view of the proof of Theorem 6.9, Lemma 6.23 is applied to the solution $\boldsymbol{\omega_i}$ of the cell problem (6.6). More precisely, introduce

$$\boldsymbol{\omega_i}' = \boldsymbol{\omega_i} - |Y_F|\langle\boldsymbol{\omega_i}\rangle = \boldsymbol{\omega_i} - \int_Y \boldsymbol{\omega_i},$$

which satisfies

$$\begin{cases} \operatorname{div}_y \ \boldsymbol{\omega_i}' = 0 \text{ on } Y, \\ \boldsymbol{\omega_i}' \text{ is } \mathbb{Z}^d - \text{periodic}, \\ \boldsymbol{\omega_i}' \text{ has zero mean over } Y. \end{cases}$$

According to Lemma 6.23, there exists a periodic function $\boldsymbol{\psi_i} \in H^2_{\#}(Y)^d$ such that

$$-\Delta_y \boldsymbol{\psi_i} + \boldsymbol{\nabla}_y \operatorname{div}_y \boldsymbol{\psi_i} = \boldsymbol{\omega_i}' \text{ in } Y. \tag{6.30}$$

In particular, since $\boldsymbol{\omega_i} \in H^1_{\#}(Y)^d$, then $\boldsymbol{\psi_i} \in H^3_{\#}(Y)^d$. This regularity cannot be improved since, although $\boldsymbol{\omega_i}$ is smooth inside Y_F, it is not in the full cell Y. In particular, it implies that, at least for large dimension d, $\boldsymbol{\psi_i}$ and $\boldsymbol{\nabla}_y\boldsymbol{\psi_i}$ may not belong to $L^\infty(Y)$. This lack of regularity requires special care in the next subsection.

6.4.5 Proof of Theorem 6.9

Inspired by the approximation Equation (6.16) of the two-scale asymptotic expansion, namely $\boldsymbol{u_\varepsilon} \approx \varepsilon^2\boldsymbol{u_2} + \varepsilon^3\boldsymbol{u_3}$, the main idea is to compare the velocity $\boldsymbol{u_\varepsilon}$ with $\varepsilon^2\boldsymbol{u_2} + \varepsilon^3\boldsymbol{u_3}$. The difficulty is that the velocity correctors $\boldsymbol{u_2}, \boldsymbol{u_3}$, built in Section 6.3, do not satisfy the boundary condition, i.e. $\boldsymbol{u_2}, \boldsymbol{u_3}$ do not vanish on $\partial\Omega$, and are not divergence-free. To satisfy this Dirichlet boundary condition for an approximation of $\boldsymbol{u_\varepsilon}$, we introduce a cut-off function η^ε as in [111].

Definition 6.25. For a small enough constant $c > 0$, define a neighborhood of the boundary $\partial\Omega$, denoted by $O^\varepsilon := \{x \in \Omega : \operatorname{dist}(x, \partial\Omega) \leq c\varepsilon\}$ (band of width $c\varepsilon$ near $\partial\Omega$). Then we choose a cut-off function $\eta^\varepsilon \in C^\infty(\overline{\Omega})$ such that η^ε and all its derivatives vanish on $\partial\Omega$, while $\eta^\varepsilon(x) = 1$ on $\Omega \setminus O^\varepsilon$ and

$$\|1 - \eta^\varepsilon\|_{L^2(\Omega)} \leq C\sqrt{\varepsilon}, \quad \|\eta^\varepsilon\|_{L^\infty(\Omega)} = 1, \quad \|\boldsymbol{\nabla}\eta^\varepsilon\|_{L^\infty(\Omega)} \leq \frac{C}{\varepsilon}, \quad \|\boldsymbol{\nabla}\boldsymbol{\nabla}\eta^\varepsilon\|_{L^\infty(\Omega)} \leq \frac{C}{\varepsilon^2}. \tag{6.31}$$

If we simply multiply the approximation $\varepsilon^2\boldsymbol{u_2} + \varepsilon^3\boldsymbol{u_3}$ by the cut-off function η^ε, of course the Dirichlet boundary condition will be satisfied on $\partial\Omega$ but the divergence of the product is not under control. Therefore, a more clever use of η^ε is required. It turns out that only $\boldsymbol{u_2}$ needs a

special treatment. Recall from (6.7) that

$$\boldsymbol{u_2}(x,y) = \frac{1}{\nu}\boldsymbol{\omega_i}(y)\left(f_i(x) - \frac{\partial p^*}{\partial x_i}(x)\right).$$

The main idea is to decompose $\boldsymbol{\omega_i}$ as

$$\boldsymbol{\omega_i} = \boldsymbol{\omega_i}' + |Y_F|\langle\boldsymbol{\omega_i}\rangle,$$

and to use (6.30) to replace $\boldsymbol{\omega_i}'$ by its representation as derivatives of the periodic function $\boldsymbol{\psi_i}$. Then, the cut-off function is inserted inside this differential representation as

$$(-\Delta + \boldsymbol{\nabla}\operatorname{div})(\eta^\varepsilon \boldsymbol{\psi_i}),$$

which has the advantage of being divergence-free and vanishing on the boundary $\partial\Omega$. We shall not apply the cut-off function to the other (non-oscillating) term

$$\frac{1}{\nu}|Y_F|\langle\boldsymbol{\omega_i}\rangle\left(f_i(x) - \frac{\partial p^*}{\partial x_i}(x)\right) \equiv |Y_F|\boldsymbol{u}(x),$$

where $\boldsymbol{u}$ is divergence-free (because it is the Darcy velocity) and satisfies $\boldsymbol{u}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$. To correct this boundary condition we apply Lemma 6.21 to $\boldsymbol{g} = |Y_F|\boldsymbol{u}$. The assumptions of Lemma 6.21 are satisfied, by its definition (6.9), the Darcy velocity $\boldsymbol{u}$ belongs to $C^{1,\alpha}(\overline{\Omega})^d$, satisfies $\operatorname{div}\boldsymbol{u} = 0$ in Ω and $\boldsymbol{u}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$. Thus Lemma 6.21 ensures that there exists $\boldsymbol{v_{b\varepsilon}} \in H^1(\Omega)^d$ supported in O^ε such that

$$\begin{gathered}\operatorname{div}\boldsymbol{v_{b\varepsilon}} = 0 \text{ on } \Omega, \qquad \boldsymbol{v_{b\varepsilon}} = |Y_F|\boldsymbol{u} \text{ on } \partial\Omega,\\ |\boldsymbol{v_{b\varepsilon}}|_{H^1(\Omega^\varepsilon)} \le \tfrac{C}{\sqrt{\varepsilon}}\|\boldsymbol{u}\|_{C^{1,\alpha}(\overline{\Omega})}.\end{gathered} \tag{6.32}$$

The key observation is that, because of our construction of Ω^ε, where obstacles are removed from Ω only if the entire cell is included in Ω, all solid obstacles are uniformly away of the boundary $\partial\Omega$ by a distance of the order of ε. Therefore, in the choice of the cut-off function η^ε (see Definition 6.25), one can choose the constant c small enough so that no obstacles are contained or intersect the neighborhood O^ε of the boundary $\partial\Omega$, where the support of the derivatives of η^ε is restricted.

Based on the asymptotic expansion of order 3 for the velocity, $\varepsilon^2\boldsymbol{u_2} + \varepsilon^3\boldsymbol{u_3}$, we define the following velocity estimator,

$$\begin{aligned}\boldsymbol{u_{\varepsilon,3}}(x) =& \frac{\varepsilon^2}{\nu}|Y_F|\langle\boldsymbol{\omega_i}\rangle\left(f_i(x) - \frac{\partial p^*}{\partial x_i}(x)\right) - \varepsilon^2\boldsymbol{v_{b\varepsilon}}\\ &+ \frac{\varepsilon^4}{\nu}(-\Delta + \boldsymbol{\nabla}\operatorname{div})\left(\eta^\varepsilon(x)\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\right)\left(f_i(x) - \frac{\partial p^*}{\partial x_i}(x)\right)\\ &+ \frac{\varepsilon^3}{\nu}\boldsymbol{\gamma_{ij}}\left(\frac{x}{\varepsilon}\right)\eta^\varepsilon(x)\frac{\partial}{\partial x_j}\left(f_i(x) - \frac{\partial p^*}{\partial x_i}(x)\right) - \boldsymbol{v_{d\varepsilon}},\end{aligned} \tag{6.33}$$

where $\boldsymbol{v_{b\varepsilon}}$, defined in (6.32), corrects the boundary condition of the first term above and $\boldsymbol{v_{d\varepsilon}} \in H_0^1(\Omega^\varepsilon)^d$ is going to be defined in the proof of Lemma 6.27 to correct the divergence-free condition

of $\boldsymbol{u_{\varepsilon,3}}$. Recall that $\boldsymbol{\psi_i}$ is defined by (6.30), $\boldsymbol{\omega_i}$ by (6.6), $\boldsymbol{\gamma_{ij}}$ by (6.12) and all functions are extended by 0 inside the obstacles so that $\boldsymbol{u_{\varepsilon,3}}$ is well defined on the whole Ω. Similarly, we define an estimator for the pressure, which is exactly the asymptotic expansion at first order of the pressure (no need of using a cut-off function), as

$$p_{\varepsilon,1} = p^* + \varepsilon p_1, \tag{6.34}$$

where p^* is the homogenized pressure, solution to the Darcy problem (6.10) and p_1 is defined by (6.7).

Remark 6.26. A first key idea in the velocity estimator (6.33) is to apply a cut-off function η^ε only to the mean-zero oscillating velocity $\boldsymbol{\omega_i}'$ (see the second line of (6.33)). For the mean, non-oscillating velocity (see the first line of (6.33)), since its normal component vanishes on the boundary $\partial\Omega$, one can build a boundary corrector $\boldsymbol{v_{b\varepsilon}}$, thanks to Lemma 6.21. The same idea is used in [165] (see e.g. (1.19) in [165]) but instead of Lemma 6.21 the authors rely on a so-called radial cut-off function which leverages the Dirichlet boundary condition for the normal component of the Darcy velocity. A second key idea is to introduce the cut-off function inside the differential representation (6.30) of $\boldsymbol{\omega_i}'$, which thus does not destroy its divergence-free character. This second idea is not found in [165], where (6.30) and Lemma 6.23 are replaced by a weaker result, Propositions 2.1 and 2.2, expressing $\boldsymbol{\omega_i}'$ as a divergence (called there a flux corrector). Therefore [165] needs to build another boundary layer corrector, which requires a delicate decomposition of the boundary to optimally estimate its norm.

Lemma 6.27. There exists $\boldsymbol{v_{d\varepsilon}} \in H^1_0(\Omega^\varepsilon)^d$ such that the velocity estimator $\boldsymbol{u_{\varepsilon,3}}$, defined by (6.33), satisfies

$$\boldsymbol{u_{\varepsilon,3}} = \varepsilon^2 \boldsymbol{u_2} + \varepsilon^3 \boldsymbol{u_3} - \boldsymbol{v_{d\varepsilon}} \text{ in } \Omega \setminus O^\varepsilon, \tag{6.35}$$

$$\boldsymbol{u_{\varepsilon,3}} = \boldsymbol{0} \text{ on } \partial\Omega, \tag{6.36}$$

$$\boldsymbol{u_{\varepsilon,3}} = \boldsymbol{0} \text{ on } \partial B^\varepsilon, \tag{6.37}$$

$$\operatorname{div} \boldsymbol{u_{\varepsilon,3}} = 0 \text{ in } \Omega, \tag{6.38}$$

and

$$|\boldsymbol{v_{d\varepsilon}}|_{H^1(\Omega^\varepsilon)} \leq C\varepsilon^{\frac{3}{2}} \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}, \tag{6.39}$$

with C independent of ε.

Proof. To check (6.35), simply recall that $\eta^\varepsilon \equiv 1$ in $\Omega \setminus O^\varepsilon$. Thus by virtue of (6.30) the second line of (6.33) simplifies and, combined with the first line, yields $\varepsilon^2 \boldsymbol{u_2}$, since $\boldsymbol{v_{b\varepsilon}}$ is supported in O^ε. Similarly, the third line of (6.33) coincides with $\varepsilon^3 \boldsymbol{u_3}$ in $\Omega \setminus O^\varepsilon$.

To check (6.36), we recognize that the first line of (6.33) vanishes on $\partial\Omega$ by construction of $\boldsymbol{v_{b\varepsilon}}$, while the second and third line vanish too because η^ε and all its derivatives vanish on $\partial\Omega$, as well as $\boldsymbol{v_{d\varepsilon}}$ by definition.

By construction of the cut-off function and Assumption 6.2 on the solid obstacles which

stay away from the boundary $\partial\Omega$, we have $\eta^\varepsilon \equiv 1$ in the vicinity of B^ε. Therefore $\boldsymbol{u_{\varepsilon,3}} = \varepsilon^2\boldsymbol{u_2} + \varepsilon^3\boldsymbol{u_3} - \boldsymbol{v_{d\varepsilon}}$ on ∂B^ε. Furthermore, since $\boldsymbol{\omega_i}$ and $\boldsymbol{\gamma_{ij}}$ vanish on Y_S, as well as $\boldsymbol{v_{d\varepsilon}}$ on ∂B^ε by definition, we deduce that (6.37) holds true.

It remains to compute the divergence of the estimator $\boldsymbol{u_{\varepsilon,3}}$ and to prove (6.38) and (6.39). The first line of (6.33) is divergence-free by construction and recall that $\operatorname{div}(-\Delta + \boldsymbol{\nabla}\operatorname{div}) = 0$. Therefore,

$$\begin{aligned}\operatorname{div}\boldsymbol{u_{\varepsilon,3}}(x) =& \frac{\varepsilon^4}{\nu}(-\Delta + \boldsymbol{\nabla}\operatorname{div})\left(\eta^\varepsilon(x)\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\right)\cdot\boldsymbol{\nabla}(f_i - \partial_i p^*)(x)\\ &+ \frac{\varepsilon^3}{\nu}\operatorname{div}\left(\boldsymbol{\gamma_{ij}}\left(\frac{x}{\varepsilon}\right)\eta^\varepsilon(x)\frac{\partial}{\partial x_j}(f_i - \partial_i p^*)(x)\right) - \operatorname{div}\boldsymbol{v_{d\varepsilon}},\end{aligned}$$

where, for the sake of clarity, the partial derivative with respect to x_i is simply denoted by ∂_i, An easy computation shows that

$$\begin{aligned}(-\Delta + \boldsymbol{\nabla}\operatorname{div})(\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\eta^\varepsilon(x)) =& \frac{1}{\varepsilon^2}\eta^\varepsilon(x)(-\Delta_y + \boldsymbol{\nabla_y}\operatorname{div}_y)\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right) + \frac{2}{\varepsilon}(\boldsymbol{\nabla}\eta^\varepsilon(x)\cdot\boldsymbol{\nabla_y}\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\\ &- \operatorname{div}_y\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}\eta^\varepsilon(x)) + (-\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\Delta\eta^\varepsilon(x) + \boldsymbol{\nabla}\boldsymbol{\nabla}\eta^\varepsilon(x)\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)),\end{aligned}$$

which leads to

$$\operatorname{div}\boldsymbol{u_{\varepsilon,3}} = L_\varepsilon + G_\varepsilon - \operatorname{div}\boldsymbol{v_{d\varepsilon}},$$

where L_ε is made of terms of order 2,

$$\begin{aligned}L_\varepsilon =& \frac{\varepsilon^2}{\nu}\eta^\varepsilon(x)(-\Delta_y + \boldsymbol{\nabla_y}\operatorname{div}_y)\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\cdot\boldsymbol{\nabla}(f_i - \partial_i p^*)(x)\\ &+ \frac{\varepsilon^2}{\nu}\eta^\varepsilon(x)\operatorname{div}_y\boldsymbol{\gamma_{ij}}\left(\frac{x}{\varepsilon}\right)\frac{\partial}{\partial x_j}(f_i - \partial_i p^*)(x),\end{aligned}$$

and G_ε consists of the remaining terms of order 3 and 4,

$$\begin{aligned}G_\varepsilon =& \frac{\varepsilon^3}{\nu}\Big(\boldsymbol{\nabla}\eta^\varepsilon(x)\operatorname{div}_y\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right) - \boldsymbol{\nabla_y}\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}\eta^\varepsilon(x)\Big)\cdot\boldsymbol{\nabla}(f_i - \partial_i p^*)(x)\\ &+ \frac{\varepsilon^3}{\nu}\boldsymbol{\gamma_{ij}}\left(\frac{x}{\varepsilon}\right)\cdot\boldsymbol{\nabla}\eta^\varepsilon(x)\frac{\partial}{\partial x_j}(f_i - \partial_i p^*)(x) + \frac{\varepsilon^3}{\nu}\boldsymbol{\gamma_{ij}}\left(\frac{x}{\varepsilon}\right)\eta^\varepsilon(x)\cdot\boldsymbol{\nabla}\frac{\partial}{\partial x_j}(f_i - \partial_i p^*)(x)\\ &+ \frac{\varepsilon^4}{\nu}\Big(-\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\Delta\eta^\varepsilon(x) + \boldsymbol{\nabla}\boldsymbol{\nabla}\eta^\varepsilon(x)\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\Big)\cdot\boldsymbol{\nabla}(f_i - \partial_i p^*)(x).\end{aligned} \tag{6.40}$$

Recalling that $\operatorname{div}_y\boldsymbol{\gamma_{ij}} = -\boldsymbol{\omega_i}\cdot\boldsymbol{e_j} + \langle\boldsymbol{\omega_i}\cdot\boldsymbol{e_j}\rangle$ and $(-\Delta_y + \boldsymbol{\nabla_y}\operatorname{div}_y)\boldsymbol{\psi_i} = \boldsymbol{\omega_i} - |Y_F|\langle\boldsymbol{\omega_i}\rangle$, one can check that L_ε vanishes since

$$\begin{aligned}L_\varepsilon =& \frac{\varepsilon^2}{\nu}\eta^\varepsilon(x)(\boldsymbol{\omega_i} - |Y_F|\langle\boldsymbol{\omega_i}\rangle)\cdot\boldsymbol{\nabla}(f_i - \partial_i p^*)(x) - \frac{\varepsilon^2}{\nu}\eta^\varepsilon(x)(\boldsymbol{\omega_i} - \langle\boldsymbol{\omega_i}\rangle)\cdot\boldsymbol{\nabla}(f_i - \partial_i p^*)(x)\\ =& \frac{\varepsilon^2}{\nu}\eta^\varepsilon(x)(1 - |Y_F|)\langle\boldsymbol{\omega_i}\rangle\cdot\boldsymbol{\nabla}(f_i - \partial_i p^*)(x) = 0\end{aligned}$$

in view of the homogenized equation (6.10). Therefore, the estimator divergence simplifies as

$$\operatorname{div}\boldsymbol{u_{\varepsilon,3}} = G_\varepsilon - \operatorname{div}\boldsymbol{v_{d\varepsilon}},$$

and we now prove the following bound

$$\|G_\varepsilon\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon^{\frac{5}{2}}\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}. \tag{6.41}$$

Indeed, the third term of G_ε in (6.40) is bounded by $C\varepsilon^3\|(f_i - \partial_i p^*)\|_{H^2(\Omega)}$ since $\|\eta^\varepsilon\|_{L^\infty(\Omega)} = 1$ and $\boldsymbol{\gamma_{ij}} \in H^{m+2}(Y_F)^d \subset L^\infty(Y_F)^d$ by Sobolev embeddings (note that this bound is slightly better than (6.41) and it is the only term in (6.40) to involve the H^2-norm). All other terms in (6.40) feature derivatives of η^ε and have thus compact supports in O^ε. In all those terms we bound $(f_i - \partial_i p^*)$ by its $C^{1,\alpha}$-norm. Although $\boldsymbol{\gamma_{ij}}$ is bounded in $L^\infty(Y_F)^d$ for the second term of (6.40), it is not clear that it is the case for $\boldsymbol{\psi_i}$ and $\boldsymbol{\nabla_y \psi_i}$, at least for any space dimension. Therefore, we rely on another argument that we explain on the first term of (6.40) (the other ones being treated similarly). Since $\|\boldsymbol{\nabla}\eta^\varepsilon\|_{L^\infty(\Omega)} \leq C\varepsilon^{-1}$, this first term is bounded in $L^2(\Omega^\varepsilon)$ by

$$\begin{aligned} &C\varepsilon^2\|(f_i - \partial_i p^*)\|_{C^{1,\alpha}(\overline{\Omega})}\left(\int_{O^\varepsilon} |\boldsymbol{\nabla_y \psi_i}\left(\frac{x}{\varepsilon}\right)|^2 \mathrm{d}x\right)^{1/2} \\ &\leq C\varepsilon^2\|(f_i - \partial_i p^*)\|_{C^{1,\alpha}(\overline{\Omega})}\left(n_{O^\varepsilon}\varepsilon^d\|\boldsymbol{\nabla_y \psi_i}(y)\|^2_{L^2(Y_F)}\right)^{1/2}, \end{aligned} \tag{6.42}$$

where we used the periodicity of $\boldsymbol{\nabla_y \psi_i}$ and $n_{O^\varepsilon} = O(\varepsilon^{-d+1})$ is the number of cells Y_i^ε which cover O^ε. This implies exactly the bound (6.41). Finally, remark that G_ε has zero mean on Ω^ε since

$$\int_{\Omega^\varepsilon} \operatorname{div}(\boldsymbol{u_{\varepsilon,3}} + \boldsymbol{v_{d\varepsilon}}) = \int_{\partial\Omega^\varepsilon} (\boldsymbol{u_{\varepsilon,3}} + \boldsymbol{v_{d\varepsilon}}) \cdot \boldsymbol{n} = 0,$$

because $\boldsymbol{u_{\varepsilon,3}}$ and $\boldsymbol{v_{d\varepsilon}}$ vanish on ∂B^ε and $\partial\Omega$, Consequently, to prove (6.38) we can rely on Lemma 6.17 which ensures that there exists $\boldsymbol{v_{d\varepsilon}} \in H_0^1(\Omega^\varepsilon)^d$ such that

$$\operatorname{div} \boldsymbol{v_{d\varepsilon}} = G_\varepsilon \text{ and } |\boldsymbol{v_{d\varepsilon}}|_{H^1(\Omega^\varepsilon)} \leq \frac{C}{\varepsilon}\|G_\varepsilon\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon^{\frac{3}{2}}\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})},$$

which proves (6.38) and (6.39). □

Now that all terms in the velocity estimator (6.33) are well defined, we turn to the estimates for the residuals of the homogenization process. Define the velocity residual $\boldsymbol{R_{u_\varepsilon}}$ and the pressure residual R_{p_ε} by

$$\boldsymbol{R_{u_\varepsilon}} = \boldsymbol{u_\varepsilon} - \boldsymbol{u_{\varepsilon,3}}, \quad R_{p_\varepsilon} = p_\varepsilon - p_{\varepsilon,1},$$

where the estimators $\boldsymbol{u_{\varepsilon,3}}$ and $p_{\varepsilon,1}$ are given respectively in (6.33) and (6.34).

Lemma 6.28. The velocity and pressure homogenization residuals verify

$$\left\{\begin{array}{rcll} -\nu\Delta \boldsymbol{R_{u_\varepsilon}} + \boldsymbol{\nabla} R_{p_\varepsilon} &=& \boldsymbol{F_\varepsilon} + \operatorname{div} \boldsymbol{H_\varepsilon} & \text{in } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{R_{u_\varepsilon}} &=& 0 & \text{in } \Omega^\varepsilon, \\ \boldsymbol{R_{u_\varepsilon}} &=& \boldsymbol{0} & \text{on } \partial B^\varepsilon, \\ \boldsymbol{R_{u_\varepsilon}} &=& \boldsymbol{0} & \text{on } \partial\Omega, \end{array}\right. \tag{6.43}$$

where the source terms are bounded by

$$\|\boldsymbol{F_\varepsilon}\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}, \tag{6.44}$$

$$\|\boldsymbol{H_\varepsilon}\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon^{\frac{3}{2}} \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}, \tag{6.45}$$

with C independent of ε.

Proof. Obviously, $\operatorname{div} \boldsymbol{R_{u_\varepsilon}} = 0$ in Ω^ε and $\boldsymbol{R_{u_\varepsilon}} = \boldsymbol{0}$ on $\partial B^\varepsilon \cup \partial\Omega$. We decompose the velocity residual as $\boldsymbol{R_{u_\varepsilon}} = (\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}) + (\varepsilon^2 \boldsymbol{u_2} - \boldsymbol{u_{\varepsilon,3}})$ which leads to the first line of (6.43), where

$$\boldsymbol{F_\varepsilon} = -\nu\Delta(\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}) + \boldsymbol{\nabla}(p_\varepsilon - p_{\varepsilon,1}) \quad \text{and} \quad \boldsymbol{H_\varepsilon} = -\nu\boldsymbol{\nabla}(\varepsilon^2 \boldsymbol{u_2} - \boldsymbol{u_{\varepsilon,3}}).$$

To prove estimate (6.44) for $\boldsymbol{F_\varepsilon}$, we compute

$$\begin{aligned} \boldsymbol{F_\varepsilon} = \boldsymbol{f} - \boldsymbol{\nabla} p^* &+ (\Delta_y \boldsymbol{\omega_i} - \boldsymbol{\nabla_y} \pi_i)\left(\frac{x}{\varepsilon}\right)(f_i - \partial_i p^*)(x) + 2\varepsilon \boldsymbol{\nabla_y} \boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}(f_i - \partial_i p^*)(x) \\ &+ \varepsilon^2 \boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\Delta(f_i - \partial_i p^*)(x) - \varepsilon \pi_i\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}(f_i - \partial_i p^*)(x), \end{aligned}$$

and, rearranging the terms,

$$\begin{aligned} \boldsymbol{F_\varepsilon} &= (\Delta_y \boldsymbol{\omega_i} - \boldsymbol{\nabla_y} \pi_i + \boldsymbol{e_i})\left(\frac{x}{\varepsilon}\right)(f_i - \partial_i p^*)(x) \\ &\quad + 2\varepsilon \boldsymbol{\nabla_y} \boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}(f_i - \partial_i p^*)(x) - \varepsilon \pi_i\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}(f_i - \partial_i p^*)(x) \\ &\quad + \varepsilon^2 \boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\Delta(f_i - \partial_i p^*)(x). \end{aligned} \tag{6.46}$$

The first term in (6.46) actually vanishes because of the cell equation (6.6). The next three terms above are of order ε or higher. To bound the fourth term requires the H^2 regularity of $(f_i - \partial_i p^*)$, while its $C^{1,\alpha}$ regularity is sufficient for all other terms. In the end, we deduce the bound (6.44) from (6.46).

Next, to prove estimate (6.45) for $\boldsymbol{H_\varepsilon}$, we rewrite

$$\begin{aligned} \boldsymbol{u_{\varepsilon,3}} - \varepsilon^2 \boldsymbol{u_2} =& \frac{\varepsilon^2}{\nu}(1 - \eta^\varepsilon(x))(|Y_F|\langle \boldsymbol{\omega_i}\rangle - \boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right))(f_i - \partial_i p^*)(x) - \varepsilon^2 \boldsymbol{v_{b\varepsilon}} - \boldsymbol{v_{d_\varepsilon}} \\ &+ \frac{\varepsilon^3}{\nu}\eta^\varepsilon(x)\boldsymbol{\gamma_{ij}}\left(\frac{x}{\varepsilon}\right)\partial_j(f_i - \partial_i p^*)(x) \\ &+ \frac{\varepsilon^3}{\nu}\left(\boldsymbol{\nabla}\eta^\varepsilon(x)\operatorname{div}_y \boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right) - \boldsymbol{\nabla_y}\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}\eta^\varepsilon(x)\right)(f_i - \partial_i p^*)(x) \\ &+ \frac{\varepsilon^4}{\nu}\left(-\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\Delta\eta^\varepsilon(x) + \boldsymbol{\nabla}\boldsymbol{\nabla}\eta^\varepsilon(x)\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\right)(f_i - \partial_i p^*)(x). \end{aligned} \tag{6.47}$$

The gradient of (6.47), equal to $\boldsymbol{H_\varepsilon}$ up to a ν-factor, is

$$\begin{aligned}
&\boldsymbol{\nabla}(\boldsymbol{u_{\varepsilon,3}} - \varepsilon^2 \boldsymbol{u_2}) = -\varepsilon^2 \boldsymbol{\nabla} \boldsymbol{v_{b\varepsilon}} - \boldsymbol{\nabla} \boldsymbol{v_{d\varepsilon}} - \frac{\varepsilon^2}{\nu} \boldsymbol{\nabla} \eta^\varepsilon (|Y_F| \langle \boldsymbol{\omega_i} \rangle - \boldsymbol{\omega_i}) (f_i - \partial_i p^*) \\
&- \frac{\varepsilon}{\nu} (1 - \eta^\varepsilon) \boldsymbol{\nabla_y} \boldsymbol{\omega_i} (f_i - \partial_i p^*) + \frac{\varepsilon^2}{\nu} (\boldsymbol{\nabla} \eta^\varepsilon \boldsymbol{\nabla_y} \operatorname{div}_y \boldsymbol{\psi_i} - \boldsymbol{\nabla_y} \boldsymbol{\nabla_y} \boldsymbol{\psi_i} \boldsymbol{\nabla} \eta^\varepsilon) (f_i - \partial_i p^*) \\
&+ \frac{\varepsilon^3}{\nu} (\boldsymbol{\nabla} \boldsymbol{\nabla} \eta^\varepsilon \operatorname{div}_y \boldsymbol{\psi_i} - \boldsymbol{\nabla_y} \boldsymbol{\psi_i} \boldsymbol{\nabla} \boldsymbol{\nabla} \eta^\varepsilon) (f_i - \partial_i p^*) \\
&+ \frac{\varepsilon^3}{\nu} (-\boldsymbol{\nabla_y} \boldsymbol{\psi_i} \Delta \eta^\varepsilon + (\boldsymbol{\nabla} \boldsymbol{\nabla} \eta^\varepsilon) \boldsymbol{\nabla_y} \boldsymbol{\psi_i}) (f_i - \partial_i p^*) \\
&+ \frac{\varepsilon^4}{\nu} (-\boldsymbol{\psi_i} \boldsymbol{\nabla} \Delta \eta^\varepsilon + \boldsymbol{\nabla} (\boldsymbol{\nabla} \boldsymbol{\nabla} \eta^\varepsilon) \boldsymbol{\psi_i}) (f_i - \partial_i p^*) \\
&+ \frac{\varepsilon^2}{\nu} (1 - \eta^\varepsilon) (|Y_F| \langle \boldsymbol{\omega_i} \rangle - \boldsymbol{\omega_i}) \boldsymbol{\nabla} (f_i - \partial_i p^*) + \frac{\varepsilon^3}{\nu} \boldsymbol{\nabla} \eta^\varepsilon \boldsymbol{\gamma_{ij}} \partial_j (f_i - \partial_i p^*) \\
&+ \frac{\varepsilon^3}{\nu} (\boldsymbol{\nabla} \eta^\varepsilon \operatorname{div}_y \boldsymbol{\psi_i} - \boldsymbol{\nabla_y} \boldsymbol{\psi_i} \boldsymbol{\nabla} \eta^\varepsilon) \boldsymbol{\nabla} (f_i - \partial_i p^*) \\
&+ \frac{\varepsilon^4}{\nu} (-\boldsymbol{\psi_i} \Delta \eta^\varepsilon + (\boldsymbol{\nabla} \boldsymbol{\nabla} \eta^\varepsilon) \boldsymbol{\psi_i}) \boldsymbol{\nabla} (f_i - \partial_i p^*) \\
&+ \frac{\varepsilon^2}{\nu} \eta^\varepsilon \boldsymbol{\nabla_y} \boldsymbol{\gamma_{ij}} \partial_j (f_i - \partial_i p^*) + \frac{\varepsilon^3}{\nu} \eta^\varepsilon \boldsymbol{\gamma_{ij}} \boldsymbol{\nabla} \partial_j (f_i - \partial_i p^*) .
\end{aligned} \tag{6.48}$$

We bound (6.48) in the same way as for the estimate of G_ε. The first two terms of (6.48) are bounded by (6.32) and (6.39). The last term of (6.48) is the only one requiring the H^2-regularity of $(f_i - \partial_i p^*)$: for all other terms the $C^{1,\alpha}$-regularity of $(f_i - \partial_i p^*)$ is sufficient. By Lemma 6.6 the cell solutions $\boldsymbol{\omega_i}$, $\boldsymbol{\gamma_{ij}}$ and the gradient $\boldsymbol{\nabla_y} \boldsymbol{\omega_i}$ belong to $L^\infty(Y)$. This is not the case for $\boldsymbol{\psi_i}$ and its derivatives (at least for any space dimension d): therefore we rely on a similar trick as in (6.42) for all terms involving $\boldsymbol{\psi_i}$. Using the bounds (6.31) on the cut-off function η^ε, it follows that the terms in the first five lines of (6.48) are of order $\varepsilon^{3/2}$. The terms of the three following lines are of order $\varepsilon^{5/2}$ and eventually in the last line the terms are respectively of order ε^2 and ε^3. We do not detail the bounds except for the seemingly higher order term

$$\begin{aligned}
\| \frac{\varepsilon}{\nu} (1 - \eta^\varepsilon) \boldsymbol{\nabla_y} \boldsymbol{\omega_i} (f_i - \partial_i p^*) \|_{L^2(\Omega^\varepsilon)} &\leq C\varepsilon \| 1 - \eta^\varepsilon \|_{L^2(\Omega^\varepsilon)} \| \boldsymbol{\nabla_y} \boldsymbol{\omega_i} \|_{L^\infty(Y)} \| (f_i - \partial_i p^*) \|_{C^{1,\alpha}(\overline{\Omega})} \\
&\leq C\varepsilon^{3/2} \| (f_i - \partial_i p^*) \|_{C^{1,\alpha}(\overline{\Omega})}
\end{aligned}$$

because of (6.31). In the end, we deduce estimate (6.45). □

Remark 6.29. We cannot have a better estimate due the presence of boundary layers. Indeed, the lower order $\sqrt{\varepsilon}$ is due to the boundary condition $\boldsymbol{u_\varepsilon} = \boldsymbol{0}$ on $\partial\Omega$, which worsens the approximation near the boundary. Technically, this is taken into account by the introduction of the cut-off function η^ε in the proof. This term satisfies the bound $|\eta^\varepsilon|_{H^1(\Omega)} \leq \frac{C}{\sqrt{\varepsilon}}$, which is the origin of the $\sqrt{\varepsilon}$-order in the estimate.

Since the residual $\boldsymbol{R_{u_\varepsilon}}$ is divergence-free and belongs to $H_0^1(\Omega^\varepsilon)^d$, it can be used as test

function, leading to

$$\begin{aligned}\nu\int_{\Omega^\varepsilon}\boldsymbol{\nabla R_{u_\varepsilon}}\cdot\boldsymbol{\nabla R_{u_\varepsilon}} &= \nu\int_{\Omega^\varepsilon}\boldsymbol{\nabla R_{u_\varepsilon}}\cdot\boldsymbol{\nabla R_{u_\varepsilon}} - \int_{\Omega^\varepsilon} R_{p_\varepsilon}\operatorname{div}\boldsymbol{R_{u_\varepsilon}}\\ &= \int_{\Omega^\varepsilon}\boldsymbol{F_\varepsilon}\cdot\boldsymbol{R_{u_\varepsilon}} - \int_{\Omega^\varepsilon}\boldsymbol{H_\varepsilon}\cdot\boldsymbol{\nabla R_{u_\varepsilon}}\\ &\le \|\boldsymbol{F_\varepsilon}\|_{L^2(\Omega^\varepsilon)}\|\boldsymbol{R_{u_\varepsilon}}\|_{L^2(\Omega^\varepsilon)} + \|\boldsymbol{H_\varepsilon}\|_{L^2(\Omega^\varepsilon)}|\boldsymbol{R_{u_\varepsilon}}|_{H^1(\Omega^\varepsilon)}.\end{aligned}$$

Using estimate (6.44) and the Poincaré inequality in perforated domains (Lemma 6.14), it follows

$$\|\boldsymbol{F_\varepsilon}\|_{L^2(\Omega^\varepsilon)}\|\boldsymbol{R_{u_\varepsilon}}\|_{L^2(\Omega^\varepsilon)} \le C\varepsilon^2\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}|\boldsymbol{R_{u_\varepsilon}}|_{H^1(\Omega^\varepsilon)}.$$

Then, using (6.45), we conclude that

$$|\boldsymbol{R_{u_\varepsilon}}|_{H^1(\Omega^\varepsilon)} \le C\varepsilon^{\frac{3}{2}}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}. \tag{6.49}$$

Now, writing $\boldsymbol{u_\varepsilon}-\varepsilon^2\boldsymbol{u_2} = \boldsymbol{R_{u_\varepsilon}}+\boldsymbol{u_{\varepsilon,3}}-\varepsilon^2\boldsymbol{u_2}$, using (6.49) and (6.45) yields

$$|\boldsymbol{u_\varepsilon}-\varepsilon^2\boldsymbol{u_2}|_{H^1(\Omega^\varepsilon)} \le |\boldsymbol{R_{u_\varepsilon}}|_{H^1(\Omega^\varepsilon)}+\|\boldsymbol{H_\varepsilon}\|_{L^2(\Omega^\varepsilon)} \le C\varepsilon^{\frac{3}{2}}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})},$$

which is precisely the H^1-estimate (6.19). The L^2-estimate (6.20) is then a consequence of the Poincaré inequality in perforated domains (Lemma 6.14). To prove the remaining pressure estimate (6.18), recall that pressure is always defined up to a constant and choose the original pressure p_ε, the homogenized pressure p^* and the pressure estimator $p_{\varepsilon,1}$ with zero mean in Ω^ε (if it is not the case, it suffices to consider $p_{\varepsilon,1} = p^* + \varepsilon(p_1 - \int_{\Omega^\varepsilon} p_1)$). Lemma 6.17 ensures the existence of $\boldsymbol{v}\in H_0^1(\Omega^\varepsilon)^d$ such that $\operatorname{div}\boldsymbol{v} = p_\varepsilon - p_{\varepsilon,1}$. Then by integration by parts,

$$\begin{aligned}&\int_{\Omega^\varepsilon}(p_\varepsilon-p_{\varepsilon,1})^2 = \int_{\Omega^\varepsilon}(p_\varepsilon-p_{\varepsilon,1})\operatorname{div}\boldsymbol{v} = -\int_{\Omega^\varepsilon}\boldsymbol{F_\varepsilon}\cdot\boldsymbol{v} + \int_{\Omega^\varepsilon}\nu\boldsymbol{\nabla}(\boldsymbol{u_\varepsilon}-\varepsilon^2\boldsymbol{u_2})\cdot\boldsymbol{\nabla v}\\ &\le C\varepsilon^{\frac{3}{2}}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}|\boldsymbol{v}|_{H^1(\Omega^\varepsilon)} \le C\varepsilon^{\frac{1}{2}}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\|p_\varepsilon-p_{\varepsilon,1}\|_{L^2(\Omega^\varepsilon)},\end{aligned}$$

using the estimate in Lemma 6.17. Thus,

$$\|p_\varepsilon-p^*\|_{L^2(\Omega^\varepsilon)} \le \|p_\varepsilon-p_{\varepsilon,1}\|_{L^2(\Omega^\varepsilon)} + \|p_{\varepsilon,1}-p^*\|_{L^2(\Omega^\varepsilon)} \le C\varepsilon^{\frac{1}{2}}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})},$$

since $(p_{\varepsilon,1}-p^*)$ is of order ε as seen from (6.34), which concludes the proof of Theorem 6.9.

Remark 6.30. In this chapter, we have considered that the obstacles do not touch the border of Ω for simplicity of presentation. However, the proof of Theorem 6.9 can be extended in the case in which the obstacles touch the border of Ω. This extension is presented in Appendix B.

Part III

MsFEM for the Stokes problem

Chapter 7

An MsFEM à la Crouzeix–Raviart for the Stokes problem

Contents

In this chapter, for the sake of completeness, we present the non-conforming high-order Multi-scale Finite Element Method (MsFEM) introduced in [79] for solving the Stokes problem in heterogeneous media. This method generalized the method proposed in [111], by considering high order sets of weighting functions on the faces as well as in the elements. In what follows, we recall the notations and the main theoretical results introduced in [79], which will be useful for the additional theoretical analysis developed in Chapters 8 and 9, which are dedicated respectively to the well-posedness of the discrete local problems involved in the MsFEM and to the derivation of an error estimate for the MsFEM under study.

7.1 Problem definition

The steady state Stokes problem with homogeneous Dirichlet boundary conditions in a perforated domain is to find the velocity $\boldsymbol{u_\varepsilon} : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\left\{\begin{array}{rcll} -\nu\Delta \boldsymbol{u_\varepsilon} + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{in } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u_\varepsilon} &=& 0 & \text{in } \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{on } \partial B^\varepsilon \cap \partial\Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{on } \partial\Omega \cap \partial\Omega^\varepsilon, \end{array}\right. \tag{7.1}$$

with $\boldsymbol{f}$ the applied force and $\nu > 0$ the viscosity. We introduce the velocity space

$$V = H_0^1(\Omega^\varepsilon)^d = \{\boldsymbol{u} \in H^1(\Omega^\varepsilon)^d \text{ such that } \boldsymbol{u}|_{\partial\Omega^\varepsilon} = \boldsymbol{0}\},$$

and the pressure space

$$M = L^2_0(\Omega^\varepsilon) = \{p \in L^2(\Omega^\varepsilon) \text{ such that } \int_{\Omega^\varepsilon} p = 0\}. \tag{7.2}$$

We note $X = V \times M$. For simplicity the fluid domain is assumed to be connected in order for the pressure to be uniquely defined in M. We introduce the following bi-linear forms

$$a(\boldsymbol{u}, \boldsymbol{v}) = \nu \int_{\Omega^\varepsilon} \boldsymbol{\nabla u} \cdot \boldsymbol{\nabla v}, \quad b(q, \boldsymbol{v}) = -\int_{\Omega^\varepsilon} q \operatorname{div} \boldsymbol{v}.$$

Then a weak formulation of the Stokes problem (7.1) reads: find $(\boldsymbol{u_\varepsilon}, p_\varepsilon) \in X$ such that

$$\left\{ \begin{array}{rcll} a(\boldsymbol{u_\varepsilon}, \boldsymbol{v}) + b(p_\varepsilon, \boldsymbol{v}) & = & \displaystyle\int_{\Omega^\varepsilon} \boldsymbol{f} \cdot \boldsymbol{v} & \forall \boldsymbol{v} \in V, \\ b(q, \boldsymbol{u_\varepsilon}) & = & 0 & \forall q \in M. \end{array} \right. \tag{7.3}$$

To obtain the existence and uniqueness result for (7.1), the load $\boldsymbol{f}$ is assumed to have the usual regularity $\boldsymbol{f} \in L^2(\Omega)^d$. It should be noted that $\boldsymbol{f} \in (H^{-1}(\Omega))^d$ is sufficient for the well-posedness of (7.1) (by replacing then $(\boldsymbol{f}, \boldsymbol{v})_{L^2}$ by $< \boldsymbol{f}, \boldsymbol{v} >_{H^{-1}}$), which is indeed the case if $\boldsymbol{f} \in L^2(\Omega)^d$. Then, existence and uniqueness of a solution to problem (7.3) is guaranteed by the Ladyzhenskaya–Babuška–Brezzi (LBB) condition (see Theorem 3.3). We introduced another bi-linear form $c(\cdot, \cdot)$ continuous on $V \times M$

$$c\left((\boldsymbol{u_\varepsilon}, p_\varepsilon), (\boldsymbol{v}, q)\right) = \int_{\Omega^\epsilon} \left(\nu \boldsymbol{\nabla u_\varepsilon} \cdot \boldsymbol{\nabla v} - p_\varepsilon \operatorname{div} \boldsymbol{v} - q \operatorname{div} \boldsymbol{u_\varepsilon}\right).$$

The variational formulation of (7.1) is equivalent to: find $(\boldsymbol{u_\varepsilon}, p_\varepsilon) \in X$ such that

$$c\left((\boldsymbol{u_\varepsilon}, p_\varepsilon), (\boldsymbol{v}, q)\right) = \int_{\Omega^\epsilon} \boldsymbol{f} \cdot \boldsymbol{v} \qquad \forall (\boldsymbol{v}, q) \in X, \tag{7.4}$$

Then, existence and uniqueness of a solution to problem (7.4) is guaranteed by the Banach-Nečas–Babuška (BNB) Theorem (see Theorem 3.2). Problem (7.4) is well-posed (admits a unique solution) if and only if the bi-linear form $c(\cdot, \cdot)$ is continuous and satisfies the so-called inf-sup condition, i.e. there exits $\alpha > 0$ such that

$$\inf_{(\boldsymbol{u_\varepsilon}, p_\varepsilon) \in X} \sup_{(\boldsymbol{v}, q) \in X} \frac{c((\boldsymbol{u_\varepsilon}, p_\varepsilon), (\boldsymbol{v}, q))}{\|(\boldsymbol{u_\varepsilon}, p_\varepsilon)\|_X \|(\boldsymbol{v}, q)\|_X} \geq \alpha.$$

It is well-known (see for example [162]), that the Stokes problem (7.4) verifies the inf-sup condition and thus admits a unique solution

$$\boldsymbol{u_\varepsilon} \in H^1_0(\Omega^\varepsilon)^d, \quad p_\varepsilon \in L^2(\Omega^\varepsilon)/\mathbb{R}.$$

7.2 Crouzeix–Raviart Multiscale Finite Element Method

Assumption 7.1. We assume that Ω is a polygonal domain. We define a mesh $\mathcal{T}_H$ on Ω as a decomposition of Ω into non-overlapping polyhedra T, each of diameter at most H, and

denote $\mathcal{F}_H$ the set of all faces of $\mathcal{T}_H$, $\mathcal{F}_H^{int} \subset \mathcal{F}_H$ the internal faces and $\mathcal{F}(T) \subset \mathcal{F}_H$ the set of faces of $T \in \mathcal{T}_H$. We assume also that the mesh does not have any hanging nodes, i.e. each internal face is shared by exactly two mesh elements.

7.2.1 Crouzeix–Raviart functional spaces

For any integer n and any integer $1 \leq l \leq d$, we denote by $\mathbb{P}_n^l$ the linear space spanned by l-variate polynomial of total degree at most n. The dimension of $\mathbb{P}_n^l$ is

$$N_n^l := \dim(\mathbb{P}_n^l) = \binom{n+l}{n}.$$

For any $T \in \mathcal{T}_H$, we denote by $\mathbb{P}_n^d(T)$ the restriction to T of polynomials in $\mathbb{P}_n^d$. For any $F \in \mathcal{F}_H$, we denote $\mathbb{P}_n^{d-1}(F)$ the restriction to F of polynomials in $\mathbb{P}_n^{d-1}$.

We now define weighting functions for velocity and pressure. Let s be a positive integer and d the dimension of Ω. We associate vector-valued functions $\boldsymbol{\omega_{F,i}} : F \to \mathbb{R}^d$ to each face $F \in \mathcal{F}_H$ for $i = 1, \ldots, s$. Let r be a positive integer. We associate vector-valued functions $\boldsymbol{\varphi_{T,k}} : T \to \mathbb{R}^d$ to each coarse element $T \in \mathcal{T}_H$ for $k = 1, \ldots, r$. Let t be a positive integer. We associate scalar functions $\varpi_{T,j} : T \to \mathbb{R}$ to each coarse element $T \in \mathcal{T}_H$ for $j = 1, \ldots, t$.

Assumption 7.2. For $n = 0$, for any $T \in \mathcal{T}_H$ and for any $F \in \mathcal{F}_H$, we choose

$$\begin{cases} s = d, \ \boldsymbol{\omega_{F,1}} = \boldsymbol{e_1}, \ldots, \boldsymbol{\omega_{F,s}} = \boldsymbol{e_d}, \\ r = 0, \\ t = 1, \ \varpi_{T,1} = 1, \end{cases} \tag{7.5}$$

where $(\boldsymbol{e_1}, \ldots, \boldsymbol{e_d})$ is the canonical basis of $\mathbb{R}^d$.

For $n \geq 1$, for any $F \in \mathcal{F}_H$ and for $T \in \mathcal{T}_H$, we assume that:

1. $s = dN_n^{d-1}$ and $(\boldsymbol{\omega_{F,i}})_{1 \leq i \leq s}$ is a set of basis functions of $(\mathbb{P}_n(F))^d$;
2. $r = dN_{n-1}^d$ and $(\boldsymbol{\varphi_{T,k}})_{1 \leq k \leq r}$ is a set of basis functions of $(\mathbb{P}_{n-1}(T))^d$;
3. $t = N_n^d$ and $(\varpi_{T,j})_{1 \leq j \leq t}$ is a set of basis functions of $\mathbb{P}_n(T)$.

Before constructing the approximation spaces of enriched Crouzeix–Raviart MsFEM, we define the functional space V_H^{ext} as presented in [79],

$$V_H^{ext} = \left\{ \begin{array}{l} \boldsymbol{u} \in (L^2(\Omega^\varepsilon))^d \ s.t. \ \boldsymbol{u}|_T \in (H^1(T \cap \Omega^\varepsilon))^d \text{ for any } T \in \mathcal{T}_H, \\ \boldsymbol{u} = \boldsymbol{0} \text{ on } \partial B^\varepsilon, \int_{F \cap \partial\Omega^\varepsilon} [[\boldsymbol{u}]] \cdot \boldsymbol{\omega_{F,j}} = 0 \text{ for all } F \in \mathcal{F}_H, j = 1, \ldots, s \end{array} \right\}$$

where $[[\boldsymbol{u}]]$ denotes the jump of $\boldsymbol{u}$ across an internal face and $[[\boldsymbol{u}]] = \boldsymbol{u}$ on $\partial\Omega$. The space V_H^{ext} plays the role of the H^1 space for the classical Finite Element Method, but it is bigger that the latter. The space V_H^{ext} enhances the natural velocity space $(H_0^1(\Omega^\varepsilon))^d$ so that we have at our disposal discontinuous vector fields across faces of the coarse mesh $\mathcal{T}_H$. Therefore V_H^{ext} is not

included in $(H_0^1(\Omega^\varepsilon))^d$, the MsFEM is non-conforming and the boundary condition is weakly enforced. The extended velocity-pressure space is defined as

$$X_H^{ext} = V_H^{ext} \times M, \tag{7.6}$$

with M defined by (7.2). The objective is to decompose X_H^{ext} in an unresolved space of infinite dimension X_H^0 and a resolved space of finite dimension X_H, which is the approximation space, such that

$$X_H^{ext} = X_H^0 \oplus X_H.$$

We define the infinite dimensional space X_H^0 as follows.

Definition 7.3. The velocity-pressure space X_H^0 is defined as a subspace of X_H^{ext} by

$$X_H^0 = V_H^0 \times M_H^0, \quad \text{with}$$

$$V_H^0 = \left\{ \begin{array}{l} \boldsymbol{u} \in V_H^{ext} \ s.t. \ \displaystyle\int_{F\cap\Omega^\varepsilon} \boldsymbol{u}\cdot\boldsymbol{\omega_{F,j}} = 0, \displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{u}\cdot\boldsymbol{\varphi_{T,k}} = 0, \\ \forall T \in \mathcal{T}_H, \ \forall F \in \mathcal{F}_H, \ j = 1,\dots,s, \ k = 1,\dots,r. \end{array} \right\} \tag{7.7}$$

and

$$M_H^0 = \left\{ p \in M \ s.t. \ \int_{T\cap\Omega^\varepsilon} p\varpi_{T,j} = 0, \forall T \in \mathcal{T}_H, \ j = 1,\dots,t \right\}. \tag{7.8}$$

As noted in [79], enriching only the set of edge weights $\boldsymbol{\omega_{F,j}}$ seems insufficient: indeed, in that case, a given function $\boldsymbol{u}$ vanishing on the faces of any T would belong to the unresolved fine scales whatever the number of face weights, but imposing the condition $\displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{u}\cdot\boldsymbol{\varphi_{T,k}} = 0$ for an increasing number of element weights will reduce the norm of its component in the unresolved space.

Afterwards, we define the subspace X_H containing coarse scales.

Definition 7.4. The velocity-pressure space X_H is defined as a subspace of X_H^{ext}, being the "orthogonal" complement of X_H^0 with respect to the bi-linear form c_H as follows

$$(\boldsymbol{u_H}, p_H) \in X_H \iff \begin{array}{l} (\boldsymbol{u_H}, p_H) \in X_H^{ext} \text{ such that} \\ c_H((\boldsymbol{u_H}, p_H), (\boldsymbol{v}, q)) = 0 \ \forall (\boldsymbol{v}, q) \in X_H^0 \end{array}$$

where c_H is defined by

$$c_H((\boldsymbol{u_H}, p_H), (\boldsymbol{v}, q)) = \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} (\nu \boldsymbol{\nabla u_H} \cdot \boldsymbol{\nabla v} - p_H \operatorname{div} \boldsymbol{v} - q \operatorname{div} \boldsymbol{u_H}).$$

Remark 7.5. The word "orthogonal" is written between quotes since the bi-linear form c_H is not a scalar product (not positive definite).

In what follows, we explicit the definition of the subspace X_H by introducing the functional spaces V_H and M_H.

Definition 7.6. We introduce the functional spaces M_H and V_H by

$$M_H = \left\{ q \in M \text{ such that } q|_T \in \mathbb{P}_n(T),\ \forall T \in \mathcal{T}_H \right\}, \tag{7.9}$$

and

$$V_H = \begin{cases} \boldsymbol{v} \in V_H^{ext} :\ \forall T\ \in\ \mathcal{T}_H,\ \exists\ \zeta_T \in M_H^0(T) \text{ such that} \\ -\nu\Delta\boldsymbol{v} + \boldsymbol{\nabla}\zeta_T\ \in \text{span}(\boldsymbol{\varphi_{T,1}}, \dots \boldsymbol{\varphi_{T,r}}) \text{ in } T \cap \Omega^\varepsilon, \\ \text{div}\,\boldsymbol{v}\ \in \text{span}(\varpi_{T,1}, \dots, \varpi_{T,t}) \text{ in } T \cap \Omega^\varepsilon, \\ \nu\boldsymbol{\nabla v n} - \zeta_T \boldsymbol{n}\ \in\ \text{span}(\boldsymbol{\omega_{F,1}}, \dots, \boldsymbol{\omega_{F,s}}) \text{ on } F \cap \Omega^\varepsilon,\ \forall F \in \mathcal{F}(T), \end{cases} \tag{7.10}$$

where $\mathcal{F}(T)$ is the set of faces composing ∂T.

Remark 7.7. It is easy to check that M_H, defined by (7.9), is orthogonal to M_H^0, defined by (7.8).

Lemma 7.8. In definition (7.10), the pressure ζ_T is uniquely defined for a given velocity $\boldsymbol{v} \in V_H$. Therefore gluing together the pressure ζ_T on all elements $T \in \mathcal{T}_H$ yields a single function $\pi_H(\boldsymbol{v}) \in M_H^0$ such that $\pi_H(\boldsymbol{v}) = \zeta_T$ on any element $T \in \mathcal{T}_H$. The operator $\pi_H : V_H \to M_H^0$ is linear and continuous.

Proof. Lemma 7.8 is proved in [79, Lemma 3.11]. □

Theorem 7.9. Using the notations of (7.9) and (7.10), we introduce the space

$$\underline{X}_H = \text{span}\{(\boldsymbol{u_H}, \pi_H(\boldsymbol{u_H}) + \overline{p}_H),\ \boldsymbol{u_H} \in V_H,\ \overline{p}_H \in M_H\}.$$

Then it coincides with the subspace X_H, i.e.

$$X_H = \underline{X}_H.$$

Proof. Theorem 7.9 is proved in [79, Theorem 3.12]. □

Theorem 7.10. The space X_H^{ext} can be decomposed as

$$X_H^{ext} = X_H \oplus X_H^0 \tag{7.11}$$

with the space X_H^0 from Definition 7.3 and the space X_H from Definition 7.4.

Proof. Theorem 7.10 is proved in [79, Theorem 3.15]. □

7.2.2 Local problem defined by the Stokes equations

Now, we construct some functions associated with faces and coarse elements of the coarse mesh. From their definitions below, they evidently belong to V_H^{ext}.

Basis functions associated with faces We first construct basis functions associated with faces of the coarse mesh. For any $F \in \mathcal{F}_H$, for $i = 1, \dots, s$, find the function $\boldsymbol{\Phi_{F,i}} : \Omega^\varepsilon \to \mathbb{R}^d$, the pressure $\pi_{F,i} : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\begin{cases} -\nu\Delta\boldsymbol{\Phi_{F,i}} + \boldsymbol{\nabla}\pi_{F,i} \in \text{span}(\boldsymbol{\varphi_{T,1}}, \dots, \boldsymbol{\varphi_{T,r}}) \text{ in } T \cap \Omega^\varepsilon, \\ \text{div}\, \boldsymbol{\Phi_{F,i}} \in \text{span}(\varpi_{T,1}, \dots, \varpi_{T,t}) \text{ in } T \cap \Omega^\varepsilon, \\ \nu\boldsymbol{\nabla}\boldsymbol{\Phi_{F,i}}\boldsymbol{n} - \pi_{F,i}\boldsymbol{n} \in \text{span}(\boldsymbol{\omega_{E,1}}, \dots, \boldsymbol{\omega_{E,s}}) \text{ on } E \cap \Omega^\varepsilon, \ \forall E \in \mathcal{F}(T), \\ \boldsymbol{\Phi_{F,i}} = \mathbf{0} \text{ on } \partial B^\varepsilon \cap T, \\ \displaystyle\int_{E\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}} \cdot \boldsymbol{\omega_{E,j}} = \begin{cases} \delta_{ij}, & E = F \\ 0, & E \neq F \end{cases} \quad \forall E \in \mathcal{F}(T), \ \forall j = 1, \dots, s, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}} \cdot \boldsymbol{\varphi_{T,l}} = 0, \ \forall l = 1, \dots, r, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \pi_{F,i}\varpi_{T,m} = 0, \ \forall m = 1, \dots, t. \end{cases} \tag{7.12}$$

Basis functions associated with elements Now, we construct basis functions associated with elements of the coarse mesh. For each $T \in \mathcal{T}_H$, for $k = 1, \dots, r$, we find $\boldsymbol{\Psi_{T,k}} : \Omega^\varepsilon \to \mathbb{R}^d$ and $\pi_{T,k} : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\begin{cases} -\nu\Delta\boldsymbol{\Psi_{T,k}} + \boldsymbol{\nabla}\pi_{T,k} \in \text{span}(\boldsymbol{\varphi_{T,1}}, \dots, \boldsymbol{\varphi_{T,r}}) \text{ in } T \cap \Omega^\varepsilon, \\ \text{div}\, \boldsymbol{\Psi_{T,k}} \in \text{span}(\varpi_{T,1}, \dots, \varpi_{T,t}) \text{ in } T \cap \Omega^\varepsilon, \\ \nu\boldsymbol{\nabla}\boldsymbol{\Psi_{T,k}}\boldsymbol{n} - \pi_{T,k}\boldsymbol{n} \in \text{span}(\boldsymbol{\omega_{E,1}}, \dots, \boldsymbol{\omega_{E,s}}) \text{ on } E \cap \Omega^\varepsilon, \ \forall E \in \mathcal{F}(T), \\ \boldsymbol{\Psi_{T,k}} = \mathbf{0} \text{ on } \partial B^\varepsilon \cap T, \\ \displaystyle\int_{E\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}} \cdot \boldsymbol{\omega_{E,j}} = 0, \ \forall E \in \mathcal{F}(T), \ \forall j = 1, \dots, s, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}} \cdot \boldsymbol{\varphi_{T,l}} = \delta_{kl}, \ \forall l = 1, \dots, r, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \pi_{T,k}\varpi_{T,m} = 0, \ \forall m = 1, \dots, t. \end{cases} \tag{7.13}$$

Theorem 7.11. The functions $\boldsymbol{\Phi_{F,i}}$ for $F \in \mathcal{F}_H$ and $i = 1, \dots, s$ defined by (7.12) and the functions $\boldsymbol{\Psi_{T,k}}$ for $T \in \mathcal{T}_H$ and $k = 1, \dots, r$ defined by (7.13) form a basis of V_H defined by (7.10). In other words,

$$V_H = \text{span}\{\boldsymbol{\Phi_{F,i}}, \ \boldsymbol{\Psi_{T,k}}, \ F \in \mathcal{F}_H, \ T \in \mathcal{T}_H, \ i = 1 \dots, s, \text{ and } k = 1, \dots, r\}$$

and $\{\boldsymbol{\Phi_{F,i}}, \ F \in \mathcal{F}_H, \ i = 1 \dots, s\} \cup \{\boldsymbol{\Psi_{T,k}}, \ T \in \mathcal{T}_H, \ k = 1, \dots, r\}$ forms a linearly independent family.

Proof. The reader can refer to [79, section 3.5]. □

Remark 7.12. We do not have explicit formulations of the functions in V_H. However in a domain without obstacles discretized with coarse triangles in two dimensions or coarse tetrahedra in three dimensions, it is easy to show that the space of polynomials of order $n+1$ is included in V_H. Consequently, in such a domain, the space V_H can be written as $V_H = \mathbb{P}_{n+1} \oplus \Sigma$, Σ being a subspace of V_H^{ext}. In particular, during this thesis, we looked for Σ as a subspace of $\mathbb{P}_{n+2}$ with trivial intersection with $\mathbb{P}_{n+1}$. However, we have shown that Σ is not a subspace of $\mathbb{P}_{n+2}$. To date, the characterization of Σ is still an open question.

Remark 7.13. For $n = 0$ the choice (7.5) of weighting functions implies that the enriched Crouzeix–Raviart MsFEM becomes exactly the Crouzeix–Raviart MsFEM proposed in [136]. This also corresponds to the so-called $CR2$ method investigated in [111]. Furthermore, in a domain without obstacles, i.e. $B^\varepsilon = \emptyset$, discretized with triangles in two dimensions or tetrahedra in three dimensions, the Crouzeix–Raviart MsFEM space with $n = 0$ becomes the classical Crouzeix–Raviart finite element space (see Section 3.1 for the description of the Crouzeix–Raviart space).

According to Theorem 7.11, any function of V_H can be represented as

$$\forall \boldsymbol{u_H} \in V_H,\ \boldsymbol{u_H}|_T = \sum_{F\in\mathcal{F}(T)} \sum_{i=1}^{s} u_{F,i} \boldsymbol{\Phi_{F,i}} + \sum_{k=1}^{r} u_{T,k} \boldsymbol{\Psi_{T,k}},$$

Similarly, an explicit formulation of the linear operator π_H (see Lemma 7.8) is

$$\forall \boldsymbol{u_H} \in V_H, \pi_H(\boldsymbol{u_H})|_T = \sum_{F\in\mathcal{F}(T)} \sum_{i=1}^{s} u_{F,i} \pi_{F,i} + \sum_{k=1}^{r} u_{T,k} \pi_{T,k}.$$

7.2.3 MsFEM approximation

The coarse-scale formulation of the Stokes problem (7.1) reads: find $(\boldsymbol{u_H}, p_H) \in X_H$ such that

$$c_H((\boldsymbol{u_H}, p_H), (\boldsymbol{v}, q)) = (\boldsymbol{f}, \boldsymbol{v}) \quad \forall (\boldsymbol{v}, q) \in X_H.$$

Theorem 7.9 implies that p_H can be decomposed as $p_H = \pi_H(\boldsymbol{u_H}) + \overline{p}_H$ with $\pi_H(\boldsymbol{u_H}) \in M_H^0$ and $\overline{p}_H \in M_H$. It is easy to verify that $\int_{T\cap\Omega^\varepsilon} \pi_H(\boldsymbol{u_H}) \operatorname{div} \boldsymbol{v} = 0$ for all $\boldsymbol{u_H}, \boldsymbol{v} \in V_H$. The problem above can consequently be reformulated as: find $\boldsymbol{u_H} \in V_H$ and $\overline{p}_H \in M_H$ such that

$$\begin{aligned} a_H(\boldsymbol{u_H}, \boldsymbol{v}) + b_H(\overline{p}_H, \boldsymbol{v}) &= F_H(\boldsymbol{v}) && \forall \boldsymbol{v} \in V_H, \\ b_H(q, \boldsymbol{u_H}) &= 0 && \forall q \in M_H, \end{aligned} \tag{7.14}$$

where

$$\begin{aligned} a_H(\boldsymbol{u_H}, \boldsymbol{v}) &= \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \nu \boldsymbol{\nabla u_H} \cdot \boldsymbol{\nabla v}, \\ b_H(\overline{p}_H, \boldsymbol{v}) &= -\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \overline{p}_H \operatorname{div} \boldsymbol{v}, \\ F_H(\boldsymbol{v}) &= \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{f} \cdot \boldsymbol{v}. \end{aligned} \tag{7.15}$$

In order to prove the well-posedness of (7.14) we introduce the following theorem.

Theorem 7.14. The space V_H and the space M_H have the following relation

$$\operatorname{div}_H V_H = M_H$$

where div_H is the broken divergence operator defined element by element

$$(\operatorname{div}_H \boldsymbol{v})_{|T} = \operatorname{div} \boldsymbol{v}_{|T}.$$

In other words, for any $p_H \in M_H$, there exist $\boldsymbol{v_H} \in V_H$ and a constant $\beta > 0$ such that

$$\operatorname{div} \boldsymbol{v_H} = p_H \text{ on } T\cap\Omega^\varepsilon,\ \forall T \in \mathcal{T}_H \text{ and } |\boldsymbol{v_H}|_{H,1} \leq \frac{1}{\beta}\|p_H\|_{L^2(\Omega^\varepsilon)}. \tag{7.16}$$

Proof. The first part of Theorem 7.14 is proved in [79, Theorem 3.16]. We prove the second part of the (7.16). For any $q \in M_H$, definition of M_H shows that $q \in L^2_0(\Omega^\varepsilon)$. Thus, there exists $\boldsymbol{v} \in (H^1_0(\Omega^\varepsilon))^d$ and a constant $\beta > 0$ (the same as in the inf-sup condition) such that $\operatorname{div} \boldsymbol{v} = q$ and $|\boldsymbol{v}|_1 \leq \frac{1}{\beta}\|q\|_{L^2(\Omega^\varepsilon)}$ [84]. As in [79, Theorem 3.16], we construct $\boldsymbol{v_H}$ as

$$\boldsymbol{v} = \boldsymbol{v_H} + \boldsymbol{v_H^0} \text{ with } \boldsymbol{v_H} \in V_H,\ \boldsymbol{v_H^0} \in V_H^0.$$

We have by orthogonality between X_H and X_H^0,

$$\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla v_H^0}\cdot\boldsymbol{\nabla w_H} - \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} p_H^0 \operatorname{div} \boldsymbol{w_H} - \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} q_H \operatorname{div} \boldsymbol{v_H^0} = 0 \quad \forall (\boldsymbol{w_H}, q_H) \in X_H.$$

Now, we remark that the second term vanishes by definition of M_H^0 and the fact that $\operatorname{div} \boldsymbol{w_H} \in \mathbb{P}_n(T)$. Now, by taking $q_H = 0$, and by noting $\boldsymbol{v_H^0} = \boldsymbol{v} - \boldsymbol{v_H}$, it comes

$$\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{v} - \boldsymbol{v_H}) \cdot \boldsymbol{\nabla w_H} = 0 \quad \forall \boldsymbol{w_H} \in V_H.$$

By taking $\boldsymbol{w_H} = \boldsymbol{v_H}$ and using the Cauchy-Schwarz inequality, it follows

$$|\boldsymbol{v_H}|^2_{H,1} = \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla v} \cdot \boldsymbol{\nabla v_H} \leq |\boldsymbol{v}|_1 |\boldsymbol{v_H}|_{H,1}.$$

Finally,

$$|\boldsymbol{v_H}|_{H,1} \leq |\boldsymbol{v}|_1 \leq \frac{1}{\beta}\|q\|_{L^2(\Omega^\varepsilon)}.$$

□

Making use of Theorem 7.14, we deduce that $\operatorname{div} \boldsymbol{u_H} = 0$ in $T \cap \Omega^\varepsilon$ for $T \in \mathcal{T}_H$. We can therefore eliminate the pressure unknown from (7.14) by introducing a subspace of V_H

$$Z_H = \{\boldsymbol{v} \in V_H \text{ such that } \operatorname{div}_H \boldsymbol{v} = 0,\ \forall T \in \mathcal{T}_H\}.$$

Therefore (7.14) is equivalent for the velocity part $\boldsymbol{u_H}$ to: find $\boldsymbol{u_H} \in Z_H$ such that

$$\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla u_H} \cdot \boldsymbol{\nabla v} = \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{f}\cdot\boldsymbol{v} \qquad \forall \boldsymbol{v} \in Z_H.$$

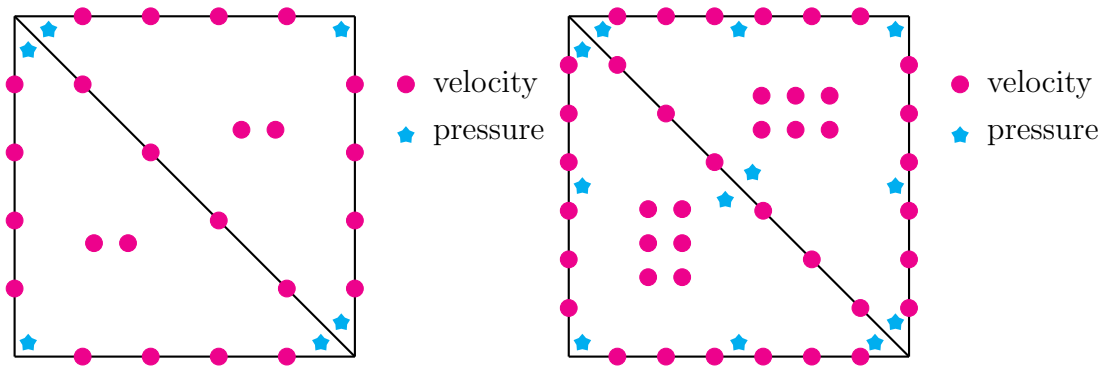
The existence and uniqueness of a solution $\boldsymbol{u_H}$ to this problem is guaranteed by the coercivity of the bi-linear form over V_H. Then, the existence and uniqueness of pressure $\overline{p}_H$ follows from the fact that $\operatorname{div}_H V_H = M_H$. As a result we have proved that (7.14) has one and only one solution $(\boldsymbol{u_H}, p_H) \in V_H \times M_H$.

After obtaining coarse solutions $\boldsymbol{u_H} = (u_{T,1},\dots,u_{T,r})_{T\in\mathcal{T}_H} \cup (u_{F,1},\dots,u_{F,s})_{F\in\mathcal{F}_H}$ and $\overline{p}_H = (\overline{p}_H|_T)_{T\in\mathcal{T}_H}$, we reconstruct on any coarse element $T \in \mathcal{T}_H$ fine-scale solutions by

$$\boldsymbol{u_H}|_T = \sum_{F\in\mathcal{F}(T)} \sum_{i=1}^{s} u_{F,i}\boldsymbol{\phi_{F,i}} + \sum_{i=1}^{r} u_{T,k}\boldsymbol{\psi_{T,k}},$$

$$p_H|_T = \sum_{F\in\mathcal{F}(T)} \sum_{i=1}^{s} u_{F,i}\pi_{F,i} + \sum_{i=1}^{r} u_{T,k}\pi_{T,k} + \overline{p}_H.$$

For the sake of understanding, in Figures 7.1a and 7.1b we represent schematically the degrees of freedom of the high-order MsFEM developed for the case $n = 1$ and $n = 2$ in two dimensions. It should be noted that the figures presented here are only schematic representations, since we recall that the degrees of freedom for the velocity are not nodal, but defined by moments.



(a) MsFEM for $n = 1$. (b) MsFEM for $n = 2$.

Figure 7.1: Scheme of the degrees of freedom of the MsFEM in two dimensions.

7.2.4 Definition of an interpolant

An important ingredient, useful in Chapters 8 and 9, is the definition of an interpolant. We note $\mathcal{N}_{MsFEM}$ the set of degrees of freedom of the MsFEM space. Using the degrees of freedom of the MsFEM space, we introduce the local interpolation operators $\boldsymbol{I_h^T} : H^1(T) \to V_H$, for $v \in H^1(T)$ by

$$N_i(\boldsymbol{I_h^T v}) = N_i(\boldsymbol{v}) \quad \forall N_i \in \mathcal{N}_{MsFEM}(T),$$

which can be expressed explicitly by

$$\boldsymbol{I_h^T u_\varepsilon} = \boldsymbol{w_h} = \sum_{F\in\mathcal{F}_H} \sum_{i=1}^{s} \left(\int_{F\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\omega_{F,i}} \right) \boldsymbol{\Phi_{F,i}} + \sum_{T\in\mathcal{T}_H} \sum_{l=1}^{r} \left(\int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{T,l}} \right) \boldsymbol{\Psi_{T,l}}. \tag{7.17}$$

Similarly, we define the following interpolant for the pressure

$$\zeta_h^K \boldsymbol{u_\varepsilon} = q_h = \sum_{F\in\mathcal{F}_H} \sum_{i=1}^{s} \left(\int_{F\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\omega_{F,i}} \right) \pi_{F,i} + \sum_{T\in\mathcal{T}_H} \sum_{l=1}^{r} \left(\int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{T,l}} \right) \pi_{T,l}. \tag{7.18}$$

In this chapter, we have recapitulated the Multi-scale Finite Element Method (MsFEM) introduced in [79]. In the subsequent two chapters, we perform a discrete analysis of this MsFEM. Indeed, in Chapter 8, we prove the well-posedness of the discrete local problems involved in the MsFEM and in Chapter 9, we derive the first error estimate for the MsFEM under study applied to the Stokes problem in perforated domains.

Chapter 8

Well-posedness of the discrete local problems

Contents

As analytical solutions of the local problems are not known, at this stage, the implementation of the MsFEM requires a numerical approximation of the local basis functions on local fine meshes. In order to obtain this approximation, it is possible to select any appropriate numerical method. In this thesis, we choose to solve the local problems using Finite Element Method. Consequently, in this chapter, we propose a discretization of the local problems using Finite Element Method. Afterwards, the main part of this chapter is dedicated to the proof of the well-posedness of the discrete local problems involved in the Multi-scale Finite Element Method (MsFEM). First, we investigate the well-posedness of the discrete local problems with usual pair of finite elements. We show their limitations for solving the local problems. Then, in the last section, we show the discrete well-posedness of the discrete local problems with the new pair of finite elements introduced in Chapters 4 and 5.

8.1 Notation

Choosing a coarse element $T \in \mathcal{T}_H$, let n_F be the number of faces composing ∂T and $n_s = n_F \times s$ be the dimension of the vector containing the Lagrange multipliers $\lambda_{F,j}$ for all $F \in \mathcal{F}(T)$ and $j = 1, \dots, s$, denoted by $\boldsymbol{\lambda_F} := (\lambda_{F,j})_{F\in\mathcal{F}(T),\ 1\le j\le s}$. Moreover, r is the dimension of the vector containing the Lagrange multipliers $\lambda^v_{T,l}$ for $l = 1, \dots, r$, denoted by $\boldsymbol{\lambda^v_T} := (\lambda^v_{T,l})_{1\le l\le r}$. In addition t is the dimension of the vector containing the Lagrange multipliers $\lambda^q_{T,m}$ for $m = 1, \dots, t$, denoted by $\boldsymbol{\lambda^q_T} := (\lambda^q_{T,m})_{1\le m\le t}$.

We introduce the velocity space

$$V_T = \left\{ \boldsymbol{v} \in (H^1(T \cap \Omega^\varepsilon))^d,\ \boldsymbol{v} = \boldsymbol{0} \text{ on } \partial B^\varepsilon \cap T \right\},$$

as well as $M^0_H(T)$, the pressure space defined by

$$M^0_H(T) = \left\{ p \in M \text{ such that } \int_{T\cap\Omega^\varepsilon} p\ \varpi_{T,i} = 0,\ i = 1, \dots, t \right\},$$

where $(\varpi_{T,i})_{1\le i\le t}$ is a set of basis function of $\mathbb{P}_n(T)$.

8.2 Weak form of the local problems

Below, we present the weak form of the local problems.

Basis functions associated with faces The weak form of (7.12) is: for any $F \in \mathcal{F}_H$, for $i = 1, \dots, s$, on the coarse element T, find $\boldsymbol{\Phi_{F,i}} \in V_T$, $\pi_{F,i} \in L^2(T \cap \Omega^\varepsilon)$, $\boldsymbol{\lambda_F} \in \mathbb{R}^{n_s}$, $\boldsymbol{\lambda^v_T} \in \mathbb{R}^r$, $\boldsymbol{\lambda^q_T} \in \mathbb{R}^t$ by solving, for all $(\boldsymbol{v}, q, \boldsymbol{\mu_F}, \boldsymbol{\mu^v_T}, \boldsymbol{\mu^q_T}) \in V_T \times L^2(T \cap \Omega^\varepsilon) \times \mathbb{R}^{n_s} \times \mathbb{R}^r \times \mathbb{R}^t$,

$$\begin{cases} \displaystyle\int_{T\cap\Omega^\varepsilon} \nu \boldsymbol{\nabla\Phi_{F,i}} \cdot \boldsymbol{\nabla v} - \int_{T\cap\Omega^\varepsilon} \pi_{F,i} \operatorname{div} \boldsymbol{v} + \sum_{E\in\mathcal{F}(T)} \sum_{j=1}^{s} \lambda_{E,j} \int_{E\cap\Omega^\varepsilon} \boldsymbol{v} \cdot \boldsymbol{\omega_{E,j}} - \\ \displaystyle\sum_{l=1}^{r} \lambda^v_{T,l} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\varphi_{T,l}} \cdot \boldsymbol{v} = 0, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} q \operatorname{div} \boldsymbol{\Phi_{F,i}} + \sum_{m=1}^{t} \lambda^q_{T,m} \int_{T\cap\Omega^\varepsilon} \varpi_{T,m} q = 0, \\ \displaystyle\sum_{E\in\mathcal{F}(T)} \sum_{j=1}^{s} \mu_{E,j} \int_{E\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}} \cdot \boldsymbol{\omega_{E,j}} = \mu_{F,i}, \\ \displaystyle\sum_{l=1}^{r} \mu^v_{T,l} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}} \cdot \boldsymbol{\varphi_{T,l}} = 0, \\ \displaystyle\sum_{m=1}^{t} \mu^q_{T,m} \int_{T\cap\Omega^\varepsilon} \pi_{F,i} \varpi_{T,m} = 0. \end{cases} \tag{8.1}$$

Basis functions associated with elements The weak form of (7.13) is: for any $T \in \mathcal{T}_H$, for $k = 1, \dots, r$, find $\boldsymbol{\Psi_{T,k}} \in V_T$, $\pi_{T,k} \in L^2(T \cap \Omega^\varepsilon)$, $\boldsymbol{\lambda_F} \in \mathbb{R}^{n_s}$, $\boldsymbol{\lambda^v_T} \in \mathbb{R}^r$, $\boldsymbol{\lambda^q_T} \in \mathbb{R}^t$ by solving, for all $(\boldsymbol{v}, q, \boldsymbol{\mu_F}, \boldsymbol{\mu^v_T}, \boldsymbol{\mu^q_T}) \in V_T \times L^2(T \cap \Omega^\varepsilon) \times \mathbb{R}^{n_s} \times \mathbb{R}^r \times \mathbb{R}^t$,

$$
\begin{cases}
\displaystyle\int_{T\cap\Omega^\varepsilon} \nu\boldsymbol{\nabla}\boldsymbol{\Psi}_{\boldsymbol{T},\boldsymbol{k}}\cdot\boldsymbol{\nabla}\boldsymbol{v} - \int_{T\cap\Omega^\varepsilon}\pi_{T,k}\operatorname{div}\boldsymbol{v} + \sum_{E\in\mathcal{F}(T)}\sum_{j=1}^{s}\lambda_{E,j}\int_{E\cap\Omega^\varepsilon}\boldsymbol{v}\cdot\boldsymbol{\omega}_{E,j} - \\
\displaystyle\sum_{l=1}^{r}\lambda^v_{T,l}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\varphi}_{T,l}\cdot\boldsymbol{v} = 0, \\
\displaystyle\int_{T\cap\Omega^\varepsilon} q\operatorname{div}\boldsymbol{\Psi}_{\boldsymbol{T},\boldsymbol{k}} + \sum_{m=1}^{t}\lambda^q_{T,m}\int_{T\cap\Omega^\varepsilon}\varpi_{T,m}q = 0, \\
\displaystyle\sum_{E\in\mathcal{F}(T)}\sum_{j=1}^{s}\mu_{E,j}\int_{E\cap\Omega^\varepsilon}\boldsymbol{\Psi}_{\boldsymbol{T},\boldsymbol{k}}\cdot\boldsymbol{\omega}_{\boldsymbol{E},\boldsymbol{j}} = 0, \\
\displaystyle\sum_{l=1}^{r}\mu^v_{T,l}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\Psi}_{\boldsymbol{T},\boldsymbol{k}}\cdot\boldsymbol{\varphi}_{\boldsymbol{T},\boldsymbol{l}} = \mu^v_{T,k}, \\
\displaystyle\sum_{m=1}^{t}\mu^q_{T,m}\int_{T\cap\Omega^\varepsilon}\pi_{T,k}\varpi_{T,m} = 0.
\end{cases}
\tag{8.2}
$$

Remark 8.1. Ideally, we would like to choose the pressure $\pi_{T,k}\in M^0_H(T)$, reducing the second equations of (8.1) and (8.2) respectively to $\displaystyle\int_{T\cap\Omega^\varepsilon} q\operatorname{div}\boldsymbol{\Phi}_{\boldsymbol{F},\boldsymbol{i}} = 0$ and $\displaystyle\int_{T\cap\Omega^\varepsilon} q\operatorname{div}\boldsymbol{\Psi}_{\boldsymbol{T},\boldsymbol{k}} = 0$. However, numerically, it is difficult to choose the test function $q\in M^0_H(T)$. Consequently, we impose this condition by using Lagrangian multipliers and we solve the local problems as written in (8.1) and (8.2).

The variational formulation of local problems can be re-written as follows. Find $(\boldsymbol{u}, p, \boldsymbol{\lambda_F}, \boldsymbol{\lambda^v_T}, \boldsymbol{\lambda^q_T}) \in V_T\times L^2(T\cap\Omega^\varepsilon)\times\mathbb{R}^{n_s}\times\mathbb{R}^r\times\mathbb{R}^t$ such that

$$
a_T(\boldsymbol{u},\boldsymbol{v}) + b_T(\boldsymbol{v},p) + c_T(\boldsymbol{v},\boldsymbol{\lambda_F}) + d_T(\boldsymbol{v},\boldsymbol{\lambda^v_T}) = 0 \qquad \forall\boldsymbol{v}\in V_T, \tag{8.3}
$$

$$
b_T(\boldsymbol{u},q) + e_T(\boldsymbol{\lambda^q_T},q) = 0 \qquad \forall q\in L^2(T\cap\Omega^\varepsilon), \tag{8.4}
$$

$$
c_T(\boldsymbol{u},\boldsymbol{\mu_F}) = l_1(\boldsymbol{\mu_F}) \qquad \forall\boldsymbol{\mu_F}\in\mathbb{R}^{n_s}, \tag{8.5}
$$

$$
d_T(\boldsymbol{u},\boldsymbol{\mu^v_T}) = l_2(\boldsymbol{\mu^v_T}) \qquad \forall\boldsymbol{\mu^v_T}\in\mathbb{R}^r, \tag{8.6}
$$

$$
e_T(\boldsymbol{\mu^q_T},p) = 0 \qquad \forall\boldsymbol{\mu^q_T}\in\mathbb{R}^t. \tag{8.7}
$$

where the bi-linear forms are defined by

$$
\begin{aligned}
a_T(\boldsymbol{u},\boldsymbol{v}) &= \int_{T\cap\Omega^\varepsilon}\nu\boldsymbol{\nabla}\boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{v}, \\
b_T(\boldsymbol{v},p) &= -\int_{T\cap\Omega^\varepsilon} p\operatorname{div}\boldsymbol{v}, \\
c_T(\boldsymbol{v},\boldsymbol{\lambda_F}) &= \sum_{E\in\mathcal{F}(T)}\sum_{j=1}^{s}\lambda_{E,j}\int_{E\cap\Omega^\varepsilon}\boldsymbol{v}\cdot\boldsymbol{\omega}_{\boldsymbol{E},\boldsymbol{j}}, \\
d_T(\boldsymbol{v},\boldsymbol{\lambda^v_T}) &= \sum_{l=1}^{r}\lambda^v_{T,l}\int_{T\cap\Omega^\varepsilon}\boldsymbol{v}\cdot\boldsymbol{\varphi}_{\boldsymbol{T},\boldsymbol{l}}, \\
e_T(\boldsymbol{\lambda^q_T},q) &= \sum_{m=1}^{t}\lambda^q_{T,l}\int_{T\cap\Omega^\varepsilon}\pi_{T,m}q,
\end{aligned}
$$

and where the linear forms l_1 and l_2 are defined by

$$\begin{aligned} l_1(\boldsymbol{\mu_F}) &= \mu_{F,i}, \qquad l_2(\boldsymbol{\mu_T^v}) = 0 \quad \text{for (8.1)} \\ l_1(\boldsymbol{\mu_F}) &= 0, \qquad l_2(\boldsymbol{\mu_T^v}) = \mu^v_{T,k} \quad \text{for (8.2)} \end{aligned}$$

System (8.3)-(8.7) is called a twofold saddle point problem in the literature and can be viewed as a single saddle point problem defined on $V_T \times (M_H^0(T) \times \mathbb{R}^{n_s} \times \mathbb{R}^r)$. By introducing the following bi-linear form $\forall (\boldsymbol{u}, (p, \boldsymbol{\lambda_F}, \boldsymbol{\lambda_T^v})) \in V_T \times (M_H^0(T) \times \mathbb{R}^{n_s} \times \mathbb{R}^r)$,

$$\tilde{b}_T(\boldsymbol{v}, (p, \boldsymbol{\lambda_F}, \boldsymbol{\lambda_T^v})) = b_T(\boldsymbol{v}, p) + c_T(\boldsymbol{v}, \boldsymbol{\lambda_F}) + d_T(\boldsymbol{v}, \boldsymbol{\lambda_T^v}), \tag{8.8}$$

the system (8.3)-(8.7) can be reformulated as

$$\begin{aligned} &a_T(\boldsymbol{u}, \boldsymbol{v}) + \tilde{b}_T(\boldsymbol{v}, (p, \boldsymbol{\lambda_F}, \boldsymbol{\lambda_T^v})) = 0 \quad \forall \boldsymbol{v} \in V_T, \\ &\tilde{b}_T(\boldsymbol{u}, (q, \boldsymbol{\mu_F}, \boldsymbol{\mu_T^v})) = l_1(\boldsymbol{\mu_F}) + l_2(\boldsymbol{\mu_T^v}) \quad \forall (q, \boldsymbol{\mu_F}, \boldsymbol{\mu_T^v}) \in M_H^0(T) \times \mathbb{R}^{n_s} \times \mathbb{R}^r. \end{aligned} \tag{8.9}$$

Lemma 8.2. The local problems defined by (8.1) and (8.2) are well-posed.

Proof. For details of the proof, the reader can refer to [79, section 3.4]. The proof is based on the inf-sup theory. It is easy to check that the bi-linear form a_T and $\tilde{b}_T$ are both continuous over their spaces of definition. Additionnaly, the linear forms l_1 and l_2 are also continuous. Moreover, the bi-linear form a_T is coercive over the velocity kernel of the bi-linear form $\tilde{b}_T$, since this kernel, for all $n \geq 0$, contains vector functions which, at least, have mean values that vanish on the faces of T, and this implies a Poincaré-like inequality. Finally, [79] proves that the bi-linear form $\tilde{b}_T$ satisfies the following inf-sup condition: there exists $\tilde{\alpha} > 0$ such that $\forall (q, \boldsymbol{\mu_F}, \boldsymbol{\mu_T^v}) \in M_H^0(T) \times \mathbb{R}^{n_s} \times \mathbb{R}^r$,

$$\sup_{\boldsymbol{v} \in V_T} \frac{\tilde{b}_T(\boldsymbol{v}, (p, \boldsymbol{\lambda_F}, \boldsymbol{\lambda_T^v}))}{(\|p\|_{L^2(T \cap \Omega^\epsilon)} + \|\boldsymbol{\lambda_F}\| + \|\boldsymbol{\lambda_T^v}\|)} \geq \tilde{\alpha}.$$

□

8.3 Discretization of the local problems

To solve the local problems using Finite Element Method, we define a local fine mesh $\mathcal{T}_h(T)$ in each coarse element $T \in \mathcal{T}_H$. We note $\{\mathcal{T}_h(T)\}$ a family of triangulations of the coarse element T parametrised by a positive parameter h which tends to 0. Each triangulation $\mathcal{T}_h(T)$ consists of a finite number of triangles in two dimensions or tetrahedra in three dimensions, τ, such that $\overline{T} = \cup_{\tau \in \mathcal{T}_h(T)} \overline{\tau}$. Let $h_\tau = \text{diam}(\tau)$ and $h := \max h_\tau$. We assume that the triangulations are conformal in the sense that the intersection of the closures of two different cells τ is either empty, a common vertex, a common edge, or a common face. As usual, we assume that the triangulations are shape regular. We denote by f a face of the element τ.

We write the local problems (7.12) and (7.13) in matrix form. First, we recall and introduce some useful notations. Let $\mathcal{F}_H$ be the set of all faces of $\mathcal{T}_H$ including those on the domain

boundary $\partial\Omega$. Let $\mathcal{F}(T)$ be the set of faces of a coarse element T. We have $\mathcal{F}(T) = \{F_1, \ldots, F_{n_F}\}$ with n_F the number of faces of an element T.

Let $\boldsymbol{\Phi_{F,i}}$ be the local basis function number i associated with the face F and $\pi_{F,i}$ the associated pressure. Let $\boldsymbol{\Psi_{T,k}}$ be the MsFEM local basis function number k associated with the element T and $\pi_{T,k}$ is the associated pressure. Let V_h be the finite element space used to compute the basis functions $\boldsymbol{\Phi_{F,i}}$ and $\boldsymbol{\Psi_{T,k}}$. Let M_h be the finite element space used to compute the associated pressure $\pi_{F,i}$ and $\pi_{T,k}$. The subscript h allows to recall the dependency of these finite element spaces to the fine mesh $\mathcal{T}_h(T)$. We also note

$$M = \dim(V_h) \qquad \text{and} \qquad N = \dim(M_h)$$

and let $\{\boldsymbol{\Lambda_1^h}, \ldots, \boldsymbol{\Lambda_M^h}\}$ be a basis of the space V_h and $\{\Theta_1^h, \ldots, \Theta_N^h\}$ be a basis of the space M_h.

The discretization of the local problems (8.1)-(8.2) using the finite element spaces V_h and M_h can be written under matrix form

$$\begin{bmatrix} A_T & B_T & C_T & D_T & 0 \\ (B_T)^t & 0 & 0 & 0 & E_T \\ (C_T)^t & 0 & 0 & 0 & 0 \\ (D_T)^t & 0 & 0 & 0 & 0 \\ 0 & (E_T)^t & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{X_T} \\ \boldsymbol{\Pi_T} \\ \boldsymbol{\lambda_F} \\ \boldsymbol{\lambda_T^v} \\ \boldsymbol{\lambda_T^q} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{B_{F,i}} \\ \boldsymbol{B_{T,k}} \\ 0 \end{bmatrix} \tag{8.10}$$

where the sub-matrices are defined by

$$\begin{cases} (A_T)_{a,b} = \displaystyle\int_{T\cap\Omega^\varepsilon} \nu \boldsymbol{\nabla\Lambda_a^h} \cdot \boldsymbol{\nabla\Lambda_b^h}, & 1 \le a,b \le M \\ (B_T)_{a,b} = -\displaystyle\int_{T\cap\Omega^\varepsilon} \Theta_b^h \operatorname{div} \boldsymbol{\Lambda_a^h}, & 1 \le a \le M \qquad 1 \le b \le N \\ (C_T)_{a,b} = \left[\displaystyle\int_{F_1\cap\Omega^\varepsilon} \boldsymbol{\Lambda_a^h} \cdot \boldsymbol{\omega_{F_1,b}}, \ldots, \int_{F_{n_F}\cap\Omega^\varepsilon} \boldsymbol{\Lambda_a^h} \cdot \boldsymbol{\omega_{F_{n_F},b}}\right], & 1 \le a \le M, \quad 1 \le b \le s \\ (D_T)_{a,b} = -\displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Lambda_a^h} \cdot \boldsymbol{\varphi_{T,b}}, & 1 \le a \le M, \qquad 1 \le b \le r \\ (E_T)_{a,b} = \displaystyle\int_{T\cap\Omega^\varepsilon} \Theta_a^h \varpi_{T,b}, & 1 \le a \le N, \qquad 1 \le b \le t \end{cases}$$

where $\boldsymbol{X_T} \in \mathbb{R}^M$ is the vector of the basis coefficients in the finite element space used to compute the solution $\boldsymbol{\Phi_{F,i}}$, $\boldsymbol{\Pi_T} \in \mathbb{R}^N$ is the vector of the basis coefficients in the finite element space used to compute the solution $\pi_{F,i}$. The right-hand side, for basis functions associated with the faces, i.e. defined by local problems (8.1), are defined by

$$\begin{aligned} (\boldsymbol{B_{F,i}})_a &= \left[\delta_{F,F_1} \ldots, \delta_{F,F_{n_F}}\right]^T \delta_{i,a}, \quad 1 \le a \le s \\ (\boldsymbol{B_{T,k}})_a &= 0, \end{aligned}$$

and those for basis functions associated with the elements, i.e. defined by local problems (8.2), are defined by

$$\begin{aligned} (\boldsymbol{B_{F,i}})_a &= 0, \\ (\boldsymbol{B_{T,k}})_a &= \delta_{k,a} \quad 1 \le a \le r. \end{aligned}$$

Once the local problems discretized, it is necessary to choose the appropriate Finite Element Method (a method that allows to solve the local problems effectively). It is possible to choose any ad hoc methods, let say, discontinuous Galerkin Method, HHO, ... However, for the sake of simplicity, in what follows, we explore the resolution of the local problems using usual pairs of finite elements (Crouzeix–Raviart, Taylor–Hood, ...), which are implemented in most of finite elements software. Indeed, if such pairs of finite elements are suitable, this renders the method straightforward to implement. However, in what follows, we notice that usual pairs of finite elements cannot solve the local problems involved in the two-dimensional MsFEM for $n \geq 2$, and in the three-dimensional MsFEM. Hence the Section 8.5, in which we show the discrete well-posedness of the discrete local problems with new pairs of finite elements (introduced in Chapters 4 and 5).

8.4 Investigations of the discrete well-posedness of the local problems with usual pairs of finite elements

8.4.1 Preliminaries

To show the well-posedness of the discrete local problems, we have to prove that the kernel of the matrix involved in (8.10) is reduced to $\mathbf{0}$. Therefore, we want to show that if $(\boldsymbol{X_T}, \boldsymbol{\Pi_T}, \boldsymbol{\lambda_F}, \boldsymbol{\lambda_T^v}, \boldsymbol{\lambda_T^q})$ satisfy

$$A_T \boldsymbol{X_T} + B_T \boldsymbol{\Pi_T} + C_T \boldsymbol{\lambda_F} + D_T \boldsymbol{\lambda_T^v} = \mathbf{0}, \tag{8.11}$$

$$(B_T)^t \boldsymbol{X_T} + E_T \boldsymbol{\lambda_T^q} = \mathbf{0}, \tag{8.12}$$

$$(C_T)^t \boldsymbol{X_T} = \mathbf{0}, \tag{8.13}$$

$$(D_T)^t \boldsymbol{X_T} = \mathbf{0}, \tag{8.14}$$

$$(E_T)^t \boldsymbol{\Pi_T} = \mathbf{0}, \tag{8.15}$$

then $\boldsymbol{X_T} = \mathbf{0}$, $\boldsymbol{\Pi_T} = \mathbf{0}$, $\boldsymbol{\lambda_F} = \mathbf{0}$, $\boldsymbol{\lambda_T^v} = \mathbf{0}$ and $\boldsymbol{\lambda_T^q} = \mathbf{0}$.

First, let us first prove that $\boldsymbol{X_T} = \mathbf{0}$. Multiplying (8.11) by $\boldsymbol{X_T}$ yields

$$(A_T \boldsymbol{X_T}, \boldsymbol{X_T}) + (\boldsymbol{\Pi_T}, (B_T)^t \boldsymbol{X_T}) + (\boldsymbol{\lambda_F}, (C_T)^t \boldsymbol{X_T}) + (\boldsymbol{\lambda_T^v}, (D_T)^t \boldsymbol{X_T}) = \mathbf{0}. \tag{8.16}$$

With (8.12), we have

$$(\boldsymbol{\Pi_T}, (B_T)^t \boldsymbol{X_T}) = -(\boldsymbol{\Pi_T}, E_T \boldsymbol{\lambda_T^q}) = -((E_T)^t \boldsymbol{\Pi_T}, \boldsymbol{\lambda_T^q}).$$

Then, (8.15) implies that

$$((E_T)^t \boldsymbol{\Pi_T}, \boldsymbol{\lambda_T^q}) = 0.$$

Equation (8.13) implies that

$$(\boldsymbol{\lambda_F}, (C_T)^t \boldsymbol{X_T}) = 0.$$

Equation (8.14) implies that

$$(\boldsymbol{\lambda_T^v}, (D_T)^t \boldsymbol{X_T}) = 0.$$

Combining these results, (8.16) becomes

$$(A_T \boldsymbol{X_T}, \boldsymbol{X_T}) = 0. \tag{8.17}$$

Equation (8.17) implies that $\boldsymbol{\nabla X_T} = \mathbf{0}$ and therefore $\boldsymbol{X_T}$ is constant. Taking advantage of (8.14), we conclude that $\boldsymbol{X_T} = \mathbf{0}$. Consequently, Equations (8.11) to (8.15) are reduced to

$$B_T \boldsymbol{\Pi_T} + C_T \boldsymbol{\lambda_F} + D_T \boldsymbol{\lambda_T^v} = \mathbf{0} \tag{8.18}$$

$$E_T \boldsymbol{\lambda_T^q} = \mathbf{0} \tag{8.19}$$

$$(E_T)^t \boldsymbol{\Pi_T} = \mathbf{0} \tag{8.20}$$

It is straightforward to deduce from (8.19) that $\boldsymbol{\lambda_T^q} = \mathbf{0}$. It remains to prove that $\boldsymbol{\Pi_T} = \mathbf{0}$, $\boldsymbol{\lambda_F} = \mathbf{0}$ and $\boldsymbol{\lambda_T^v} = \mathbf{0}$. Let f be a given face and $\omega_f = \{\tau_1, \tau_2\}$ be the set of elements that share the face f. Let $\boldsymbol{\Lambda_i^h}$ be the velocity basis function that has a degree of freedom located on the face f (this is legitimate since we are considering in the subsections that follow the non-conforming $\mathbb{P}_1$ or the Lagrange $\mathbb{P}_2$ finite element spaces for the velocity). Its support is included in ω_f and then the line i of the matrix equation (8.18) can be written as

$$\int_{\omega_f} \operatorname{div} \boldsymbol{\Lambda_i^h}\ q_h = \sum_{l=1}^{r} \int_{\omega_f} \lambda_{T,l}^v \boldsymbol{\varphi_{T,l}} \cdot \boldsymbol{\Lambda_i^h}. \tag{8.21}$$

In (8.21), q_h is the unknown pressure in ω_f, obtained by a linear combination of the pressure basis functions (i.e. the basis functions of the finite element space M_h used to discretize the functions $\pi_{F,i}$ and $\pi_{T,k}$) associated with the pressure unknowns (the components of $\boldsymbol{\Pi_T}$) i.e.

$$q_h = \sum_{j \in \mathcal{J}} \theta_j^h q_j,$$

where q_j is the j^{th} component of $\boldsymbol{\Pi_T}$ and $\mathcal{J} = \{1 \leq j \leq N \text{ such that } \operatorname{supp}(\theta_j^h) \cap \omega_f \neq \emptyset\}$.

It remains to prove that the pressure q_h and the Lagrange multipliers $\lambda_{T,l}^v$ are equal to zero. However, the proof of this latter assertion is contingent upon several parameters. These include the MsFEM order, the dimensions of the problem (two or three dimensions) and the pair of finite elements chosen to solve the local problems. Consequently, in what follows, we examine whether the aforementioned equality holds for different pairs of finite elements.

8.4.2 Well-posedness of the discrete local problems for the MsFEM in the case $n = 1$ in two dimensions (2D)

Definition 8.3. The weighting functions for the MsFEM in the case $n = 1$ are defined in two dimensions by

$$\begin{cases} s = 4,\ \boldsymbol{\omega_{F,1}} = \boldsymbol{e_1},\ \boldsymbol{\omega_{F,2}} = \boldsymbol{e_2},\ \boldsymbol{\omega_{F,3}} = \boldsymbol{n_F}\Psi_F,\ \boldsymbol{\omega_{F,4}} = \boldsymbol{\tau_F}\Phi_F, \\ r = 2,\ \boldsymbol{\varphi_{T,1}} = \boldsymbol{e_1},\ \boldsymbol{\varphi_{T,2}} = \boldsymbol{e_2}, \\ t = 3, \varpi_{T,1} = 1,\ \varpi_{T,2} = x,\ \varpi_{T,3} = y, \end{cases}$$

where Ψ_F and Φ_F are non-vanishing functions in $\mathbb{P}_1(F)$ with vanishing mean values on $F \cap \Omega^\varepsilon$.

In this section, we note $\omega_f = \{\tau_1, \tau_2\}$ the set of elements that share a face f (see Figure 8.1). We note $\boldsymbol{n_f} = (n_{x,f}, n_{y,f})^t$ a normal vector to the face f oriented from τ_1 to τ_2. We note s_1 and s_2 the two vertices of the face f, and $\boldsymbol{t_f}$ a tangential vector to the face f oriented from s_1 to s_2.

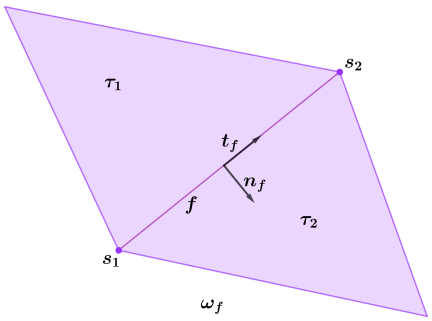


Figure 8.1: Scheme of the domain ω_f.

Besides, in particular, for this MsFEM, (8.21) becomes in the domain ω_f,

$$\int_{\omega_f} \operatorname{div} \boldsymbol{\Lambda_i^h}\ q_h = \int_{\omega_f} \lambda^v_{T,1} \boldsymbol{\varphi_{T,1}} \cdot \boldsymbol{\Lambda_i^h} + \int_{\omega_f} \lambda^v_{T,2} \boldsymbol{\varphi_{T,2}} \cdot \boldsymbol{\Lambda_i^h}. \tag{8.22}$$

8.4.2.1 Discrete well-posedness with the Crouzeix–Raviart Finite Element

Definition 8.4. The functional spaces of the Crouzeix–Raviart Finite Element are defined by

$$V_h = \{\boldsymbol{v_h} \in (\mathcal{C}^0(\overline{T}))^2 \ s.t.\ \boldsymbol{v_h}|_\tau \in \mathbb{P}_1(\tau)^2 \text{ and continuous at the middle of the faces } \forall \tau \in \mathcal{T}_h(T)\},$$

$$M_h = \left\{q_h \in \mathcal{C}^0(\overline{T}) \text{ such that } q_h|_\tau \in \mathbb{P}_0(\tau),\ \forall \tau \in \mathcal{T}_h(T)\right\}.$$

A scheme of this discretization is proposed in Figure 8.2.

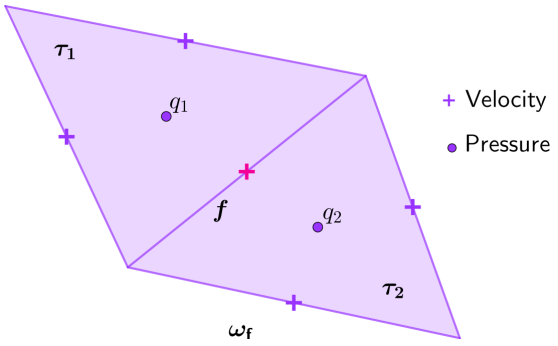


Figure 8.2: Scheme of the domain ω_f discretized with Crouzeix–Raviart Finite Element.

Lemma 8.5. The MsFEM discrete local problems in the case $n = 1$ in two dimensions are well-posed with the Crouzeix–Raviart Finite Element.

Proof. If velocity and pressure unknowns of the local problems are discretized using the Crouzeix–Raviart Finite Element, then the pressure is constant in each element $\tau \in \mathcal{T}_h(T)$. First, we choose the test function $\boldsymbol{\Lambda_i^h}$ which is equal to $(1,0)$ on the face f and $(0,0)$ on the other faces. Then, (8.22) becomes

$$(q_1 - q_2)|f|n_{x,f} = \frac{1}{3}\text{meas}(\omega_f)\lambda^v_{T,1}. \tag{8.23}$$

where we have used the fact that q_h is constant and the Green formula on the left of (8.23) and the fact that $\boldsymbol{\varphi_{T,1}} = \boldsymbol{e_1}$ and $\boldsymbol{\varphi_{T,2}} = \boldsymbol{e_2}$ on the right of (8.23). Repeat the same procedure with the test function $\boldsymbol{\Lambda_i^h}$ which is equal to $(0,1)$ on the face f and $(0,0)$ on the other faces, (8.22) becomes

$$(q_1 - q_2)|f|n_{y,f} = \frac{1}{3}\text{meas}(\omega_f)\lambda^v_{T,2}. \tag{8.24}$$

Then, by gathering (8.23) and (8.24), it follows

$$(q_1 - q_2)|f|\boldsymbol{n_f} = \frac{1}{3}\text{meas}(\omega_f)\boldsymbol{\lambda^v_T}. \tag{8.25}$$

Using (8.25), we deduce that $\boldsymbol{\lambda^v_T}$ is co-linear with $\boldsymbol{n_f}$. Besides, the same result is obtained for the other faces f_i of the element τ_1. This result implies that $\boldsymbol{\lambda^v_T}$ is also co-linear with all vectors $\boldsymbol{n_{f_i}}$. In conclusion, $\boldsymbol{\lambda^v_T} = \boldsymbol{0}$ which implies that $q_1 = q_2$. Using the same arguments for any internal face $f \in \mathcal{F}_h$ of the mesh $\mathcal{T}_h(T)$, we verify that the pressure q_h has the same value on all elements $\tau \in \mathcal{T}_h(T)$. In other words, q_h is in the space $\mathbb{P}_0(T)$. Moreover, making use of (8.20), it is straightforward to deduce that $q_h = 0$. From (8.18), we deduce that $\boldsymbol{\lambda_F} = \boldsymbol{0}$. □

8.4.2.2 Discrete well-posedness with the $\mathbb{P}_1$-nonconforming/$\mathbb{P}_1$ finite element

Definition 8.6. The $\mathbb{P}_1$-nonconforming / $\mathbb{P}_1$ functional spaces are defined by

$$V_h = \{\boldsymbol{v_h} \in (\mathcal{C}^0(\overline{T}))^2 \; s.t. \; \boldsymbol{v_h}|_\tau \in \mathbb{P}_1(\tau)^2 \text{ and continuous at the middle of the faces, } \forall \tau \in \mathcal{T}_h(T)\}$$

$$M_h = \{q_h \in \mathcal{C}^0(\overline{T}) \text{ such that } q_h|_\tau \in \mathbb{P}_1(\tau), \; \forall \tau \in \mathcal{T}_h(T)\}$$

A scheme of this discretization is proposed in Figure 8.3.

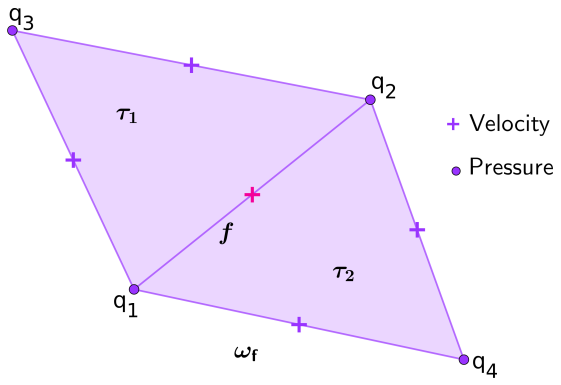


Figure 8.3: Scheme of the domain ω_f discretized with $\mathbb{P}_1$-nonconforming / $\mathbb{P}_1$ Finite Element.

Lemma 8.7. The MsFEM discrete local problems in the case $n = 1$ in two dimensions are not well-posed with the $\mathbb{P}_1$-nonconforming/$\mathbb{P}_1$ finite element.

Proof. We propose to calculate directly the term $\int_{\omega_f} \operatorname{div} \boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}}\, q_h$. Since $\boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}} \in \mathbb{P}_1-$nonconforming, we have $\operatorname{div} \boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}} = cste$. First, we choose the test function $\boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}}$ which is equal to $(1,0)$ on the face f and $(0,0)$ on the other faces. Observe that with this choice of $\boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}}$, the following formula holds

$$\operatorname{div} \boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}} = \frac{1}{\operatorname{meas}(\tau_1)} \int_{\tau_1} \operatorname{div} \boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}} = \frac{1}{\operatorname{meas}(\tau_1)} \int_{\partial\tau_1} \boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}} \cdot \boldsymbol{n_f} = \frac{|f|}{\operatorname{meas}(\tau_1)} n_{x,f}.$$

Consequently,

$$\begin{aligned}
\int_{\omega_f} \operatorname{div} \boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}}\, q_h &= \frac{|f|}{\operatorname{meas}(\tau_1)} n_{x,f} \int_{\tau_1} q_h - \frac{|f|}{\operatorname{meas}(\tau_2)} n_{x,f} \int_{\tau_2} q_h \\
&= \frac{|f|}{\operatorname{meas}(\tau_1)} \operatorname{meas}(\tau_1)(q_1+q_2+q_3) n_{x,f} - \frac{|f|}{\operatorname{meas}(\tau_2)} \operatorname{meas}(\tau_2)(q_1+q_2+q_4) n_{x,f} \\
&= |f|(q_3 - q_4) n_{x,f}.
\end{aligned}$$

Besides, we also have,

$$\int_{\omega_f} \lambda_{T,1}^v \boldsymbol{\varphi_{T,1}} \cdot \boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}} + \int_{\omega_f} \lambda_{T,2}^v \boldsymbol{\varphi_{T,2}} \cdot \boldsymbol{\Lambda}_{\boldsymbol{i}}^{\boldsymbol{h}} = \frac{1}{3} \operatorname{meas}(\omega_f) \lambda_{T,1}^v,$$

given that $\boldsymbol{\lambda_T^v} \in \mathbb{R}^2$ $(r = 2)$ and $\boldsymbol{\varphi_{T,1}} = \boldsymbol{e_1}$ and $\boldsymbol{\varphi_{T,2}} = \boldsymbol{e_2}$. Therefore, it follows

$$|f|(q_3 - q_4) n_{x,f} = \frac{1}{3} \operatorname{meas}(\omega_f) \lambda_{T,1}^v. \tag{8.26}$$

Repeat the same procedure with the test function $\boldsymbol{v_h}$ which is equal to $(0,1)$ on the face f and $(0,0)$ on the other faces, we have

$$|f|(q_3 - q_4) n_{y,f} = \frac{1}{3} \operatorname{meas}(\omega_f) \lambda_{T,2}^v. \tag{8.27}$$

Now, gathering (8.26) and (8.27), we have

$$|f|(q_3 - q_4)\boldsymbol{n_f} = \frac{1}{3}\text{meas}(\omega_f)\boldsymbol{\lambda_T^v}.$$

We conclude that $\boldsymbol{\lambda_T^v}$ is co-linear to all $\boldsymbol{n}$, which implies that $\boldsymbol{\lambda_T^v} = \boldsymbol{0}$. And then, we conclude that $q_3 = q_4$. We have shown that if we choose a face f, the values of the pressure defined on the two opposed vertices to f are equal. It remains to show that all q_i are equal. If all the pressure nodes are connected through together trougth opposed faces, then all q_i are equal. However, this condition is not statifies for a given arbitrary mesh. Indeed, as shown in Figure 8.4, all q_i opposed by a face are not linked together. In Figure 8.4, all the pressure of the same color are equal but the three values taken by the pressure are a priori not equal.

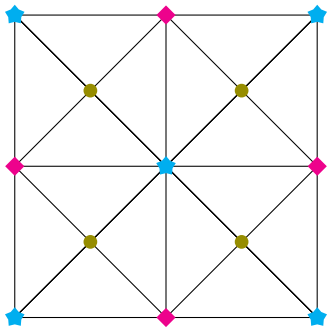


Figure 8.4: Pressure unknowns (pressures with same symbol are equal).

□

8.4.2.3 Discrete well-posedness with the $\mathbb{P}_2/\mathbb{P}_1$ finite element (Taylor–Hood finite element)

Definition 8.8. The Taylor–Hood functional spaces are defined by

$$V_h = \{\boldsymbol{v_h} \in (\mathcal{C}^0(\overline{T}))^2 \text{ such that } \boldsymbol{v_h}|_\tau \in \mathbb{P}_2(\tau)^2,\ \forall \tau \in \mathcal{T}_h(T)\}$$

with the function values at the principal lattice of order 2 as degrees of freedom.

$$M_h = \{q_h \in \mathcal{C}^0(\overline{T}) \text{ such that } q_h|_\tau \in \mathbb{P}_1(\tau),\ \forall \tau \in \mathcal{T}_h(T)\}$$

A scheme of this discretization is proposed in Figure 8.5.

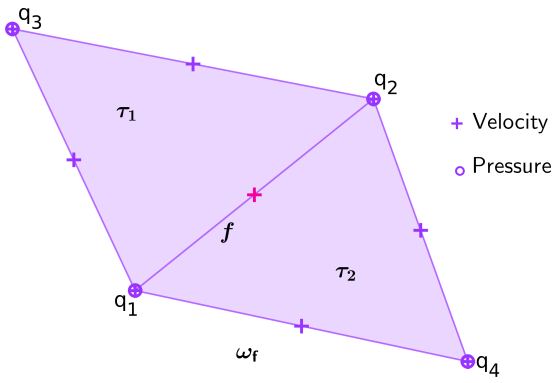


Figure 8.5: Scheme of the domain ω_f discretized with $\mathbb{P}_2$ / $\mathbb{P}_1$ Finite Element.

Lemma 8.9. The MsFEM discrete local problems in the case $n = 1$ in two dimensions are well-posed with the Taylor–Hood finite element.

Proof. First, we choose the test function $\boldsymbol{\Lambda_i^h}$ which is equal to $(1, 0)$ at the middle of the edge f and $(0, 0)$ on the other degrees of freedom. Since $\boldsymbol{\Lambda_i^h}$ vanishes on $\partial\omega_f$ and q_h belongs to $H^1(\omega_f)$, we have

$$\begin{aligned}\int_{\omega_f} \operatorname{div} \boldsymbol{\Lambda_i^h}\, q_h \mathrm{d}x &= -\int_{\omega_f} \boldsymbol{\Lambda_i^h} \cdot \boldsymbol{\nabla} q_h \mathrm{d}x \\ &= -\left(\int_{\tau_1} \boldsymbol{\Lambda_i^h} \cdot \boldsymbol{\nabla} q_h \mathrm{d}x + \int_{\tau_2} \boldsymbol{\Lambda_i^h} \cdot \boldsymbol{\nabla} q_h \mathrm{d}x\right).\end{aligned}$$

Using the exact quadrature formula for quadratic function (see Appendix C.1) and the fact that $\boldsymbol{\nabla} q_h$ is constant on each τ_i, it follows

$$\int_{\tau_i} \boldsymbol{\Lambda_i^h} \cdot \boldsymbol{\nabla} q_h \mathrm{d}x = \boldsymbol{\nabla} q_h|_{\tau_i} \cdot \int_{\tau_i} \boldsymbol{\Lambda_i^h} \mathrm{d}x = \frac{\operatorname{meas}(\tau_i)}{3} \boldsymbol{\nabla} q_h|_{\tau_i} \cdot \boldsymbol{e_1}.$$

Consequently,

$$\int_{\omega_f} \operatorname{div} \boldsymbol{\Lambda_i^h}\, q_h \mathrm{d}x = \frac{\operatorname{meas}(\tau_1)}{3} \boldsymbol{\nabla} q_h|_{\tau_1} \cdot \boldsymbol{e_1} + \frac{\operatorname{meas}(\tau_2)}{3} \boldsymbol{\nabla} q_h|_{\tau_2} \cdot \boldsymbol{e_1}.$$

Besides, we have also

$$\int_{\omega_f} \lambda^v_{T,1} \boldsymbol{\varphi_{T,1}} \cdot \boldsymbol{\Lambda_i^h} + \int_{\omega_f} \lambda^v_{T,2} \boldsymbol{\varphi_{T,2}} \cdot \boldsymbol{\Lambda_i^h} = \operatorname{meas}(\omega_f) \lambda^v_{T,1},$$

given that $\boldsymbol{\lambda_T^v} \in \mathbb{R}^2$ $(r = 2)$ and $\boldsymbol{\varphi_{T,1}} = \boldsymbol{e_1}$ and $\boldsymbol{\varphi_{T,2}} = \boldsymbol{e_2}$. Therefore, it follows

$$\frac{\operatorname{meas}(\tau_1)}{3} \boldsymbol{\nabla} q_h|_{\tau_1} \cdot \boldsymbol{e_1} + \frac{\operatorname{meas}(\tau_2)}{3} \boldsymbol{\nabla} q_h|_{\tau_2} \cdot \boldsymbol{e_1} = \operatorname{meas}(\omega_f) \lambda^v_{T,1}, \tag{8.28}$$

Repeat the same procedure with the test function $\boldsymbol{\Lambda_i^h}$ which is equal to $(0, 1)$ on the face f and $(0, 0)$ on the other degrees of freedom, we have

$$\frac{\operatorname{meas}(\tau_1)}{3} \boldsymbol{\nabla} q_h|_{\tau_1} \cdot \boldsymbol{e_2} + \frac{\operatorname{meas}(\tau_2)}{3} \boldsymbol{\nabla} q_h|_{\tau_2} \cdot \boldsymbol{e_2} = \operatorname{meas}(\omega_f) \lambda^v_{T,2}, \tag{8.29}$$

Finally, combining (8.28) and (8.29), it follows

$$\frac{\text{meas}(\tau_1)}{3}\boldsymbol{\nabla} q_h|_{\tau_1} + \frac{\text{meas}(\tau_2)}{3}\boldsymbol{\nabla} q_h|_{\tau_2} = \text{meas}(\omega_f)\boldsymbol{\lambda_T^v}. \tag{8.30}$$

Now, the definition of pressure gradient implies that

$$\begin{aligned}\boldsymbol{\nabla} q_h|_{\tau_1} \cdot \boldsymbol{t_f} &= q_2 - q_1,\\ \boldsymbol{\nabla} q_h|_{\tau_2} \cdot \boldsymbol{t_f} &= q_2 - q_1.\end{aligned}$$

Consequently, by multiplying both sides of (8.30) by $\boldsymbol{t_f}$, it follows

$$\left(\frac{\text{meas}(\tau_1)}{3} + \frac{\text{meas}(\tau_2)}{3}\right)\boldsymbol{\nabla} q_h|_{\tau_1} \cdot \boldsymbol{t_f} = \text{meas}(\omega_f)\boldsymbol{\lambda_T^v} \cdot \boldsymbol{t_f},$$

which can be reduced to

$$\boldsymbol{\nabla} q_h|_{\tau_1} \cdot \boldsymbol{t_f} = \boldsymbol{\lambda_T^v} \cdot \boldsymbol{t_f}.$$

Using the same arguments, it is straightforward to verify that this equality holds true for all $\boldsymbol{t_{f_i}}$ in the element τ_1. In this case, we deduce that

$$\boldsymbol{\nabla} q_h|_{\tau_1} = \boldsymbol{\lambda_T^v}.$$

Proceeding in this method for any internal face f of the mesh $\mathcal{T}_h(T)$, we deduce that q_h is in $\mathbb{P}_1(T)$ space. Combining (8.20) we conclude that $q_h = 0$. Consequently, we have $\boldsymbol{\lambda_T^v} = \mathbf{0}$ and then (8.18) reveals that $\boldsymbol{\lambda_F} = \mathbf{0}$. □

8.4.3 Discussions about the well-posedness of the discrete local problems in the case $n \geq 2$ in two dimensions (2D)

In this section, we once again refer to V_h as finite element space of velocity approximation, and M_h the space of pressure approximation. We note $\omega_f = \{\tau_1, \tau_2\}$ the set of elements that share a common face f. In particular for this MsFEM, making use of definition of D_T, we have in the domain ω_f,

$$\int_{\omega_f} \text{div}\, \boldsymbol{\Lambda_i^h}\ q_h = \sum_{l=1}^{r} \int_{\omega_f} \lambda_{T,l}^v \boldsymbol{\varphi_{T,l}} \cdot \boldsymbol{\Lambda_i^h}, \tag{8.31}$$

with $\boldsymbol{\lambda_T^v} \in \mathbb{R}^r$ and $(\boldsymbol{\varphi_{T,l}})_{1\leq j\leq r}$ a basis of the space $(\mathbb{P}_{n-1}(T))^d$. In order to proceed, we introduce the Taylor–Hood method for higher-order polynomials, referred to as the $\mathbb{P}_{n+1}/\mathbb{P}_n$ finite element.

Definition 8.10. The Taylor–Hood functional spaces for higher-order polynomials are defined by

$$V_h = \{\boldsymbol{v_h} \in (\mathcal{C}^0(\overline{T}))^2 \text{ such that } \boldsymbol{v_h}|_\tau \in \mathbb{P}_{n+1}(\tau)^2,\ \forall \tau \in \mathcal{T}_h(T)\},$$

with the function values at the principal lattice of order 2 as degrees of freedom, and

$$M_h = \{q_h \in \mathcal{C}^0(\overline{T}) \text{ such that } q_h|_\tau \in \mathbb{P}_n(\tau),\ \forall \tau \in \mathcal{T}_h(T)\}.$$

Given the fact that some quantities $\lambda_{T,l}^{v}\boldsymbol{\varphi_{T,l}}$ are of order $n-1$, we would like to choose q_h of polynomial order n so that $\boldsymbol{\nabla} q_h$ is also of order $n-1$. We associate this pressure with a velocity $\boldsymbol{\Lambda_i^h}$ of order $n+1$. Then, the right-hand side of (8.31) is of order $2n$. We would like therefore integrate a polynomial of order $2n$ with quadrature points corresponding to the degrees of freedom of $\mathbb{P}_{n+1}$. However, no general quadrature formula can been found for high-order polynonials [170, 94] (see Appendix C.2). For example, with the pair $\mathbb{P}_2-\mathbb{P}_1$, this leads to wanting to integrate a polynomial of order three with the quadrature points corresponding to the principal lattices of order two in the triangle. Such a quadrature rule has not been found.

Remark 8.11. Numerical experiments have also shown that solving the local problems in two dimensions for $n=2$ with Crouzeix–Raviart finite element or with Taylor–Hood finite element lead to instabilities in the numerical solutions.

8.4.4 Discussions about the well-posedness of the discrete local problems for any order n in three dimensions (3D)

For the three-dimensional MsFEM with $n=1$, we may think that the pair of finite elements $\mathbb{P}_2-\mathbb{P}_1$ allows to satisfy the discrete inf-sup condition for the discrete local problems, as for the case $n=1$ in two dimensions. We recall that in two dimensions, in the case $n=1$, the discrete inf-sup condition for the pair of finite elements $\mathbb{P}_2-\mathbb{P}_1$ holds due to the existence of an exact quadrature formula for polynomials of degree 2 on triangle using the 6 points of Lagrange interpolation (see Appendix C.1). However to my knowledge, and after a literature review, such a formula does not exist in 3D. We cannot integrate exactly a polynomial of degree 2 on a tetrahedron using the 10 points of Lagrange interpolation.

In the two previous discussions, we highlighted the limitations of usual pair of finite elements to solve the discrete local problems in two-dimensions for $n\geq 2$ and in three dimensions for any order n. Consequently, to prove the well-posedness of the discrete local problems in two and three dimensions, we proceed in another way: we use the so-called Fortin operator as in [80, 36, 35, 49] associated with the new pairs of non-conforming finite elements introduced in Chapters 4 and 5.

8.5 Unified theory: well-posedness of the discrete local problems at any order in two and three dimensions

In this section we prove the well-posedness of the discrete local problems, with the Fortin Lemma [80], in two dimensions with the family of finite elements introduced by [132] (and presented in Chapter 4) and in three dimensions with the family of finite elements introduced in Chapter 5. We take advantage of the fact that these families of finite element have the same degrees of freedom as the MsFEM space. In particular in this section, we keep the notations of Chapter 5 for defining the different Sobolev spaces and associated norms. Besides, let T be an element of $\mathcal{T}_H$ and $\widehat{T}$ be a reference element, such that there exists a geometric transformation π_T from $\widehat{T}$ to $T=\pi_T(\widehat{T})$.

First, for the sake of completeness, we recall the degrees of the new non-conforming families of finite elements.

Definition 8.12 (Degrees of freedom of the family of finite elements V_{n+1} in 2D and 3D)**.** The finite element space V_{n+1} is characterized by the following (local) degrees of freedom defined on $H^{1,h}(T)$

$$N_j^{f_\alpha}(v) = \int_{f_\alpha} v L_j^{f_\alpha}, \quad j \geq 0,\ \alpha = 1, \dots, d+1,$$

$$N_j^{\tau}(v) = \int_{\tau} v M_j^{\tau}, \quad j \geq 0,$$

where the $L_j^{f_\alpha}$ define an arbitrary basis of $\mathbb{P}_n(f_\alpha)$ and $(M_j^\tau)_j$ define an arbitrary basis of $\mathbb{P}_{n-1}(\tau)$.

Lemma 8.13. If the continuous inf-sup conditions holds with the constant β for the operator $\tilde{b}_T$ defined in (8.8) and if there exists a linear operator: $\boldsymbol{\Pi_h} : (H_0^1(\Omega))^d \to V_h$ such that $\forall (q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T}) \in M_h \times \mathbb{R}^{n_s} \times \mathbb{R}^r$ and $\boldsymbol{v} \in (H_0^1(\Omega))^d$, it holds

$$\tilde{b}_T\left(\boldsymbol{\Pi_h v}, (q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T^v})\right) = \tilde{b}_T\left(\boldsymbol{v}, (q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T^v})\right) \tag{8.32}$$

$$|\boldsymbol{\Pi_h v}|_{1,h} \leq C |\boldsymbol{v}|_1 \tag{8.33}$$

where the constant C is independent of h, then the discrete inf-sup condition holds with $\beta_h \leq \beta\ |||\boldsymbol{\Pi_h}|||^{-1}$ where $|||\cdot|||$ denotes the operator norm.

Theorem 8.14. The local problems (8.9) are well-posed when discretizing the local basis functions $\boldsymbol{\Phi_{F,i}}$ or $\boldsymbol{\Psi_{T,k}}$ with the finite element V_{n+1}, i.e. $V_h = V_{n+1}$ and the pressures $\pi_{F,i}$ or $\pi_{T,k}$ with fully discontinuous polynomial of order n, i.e. $M_h = \mathbb{P}_n^{dc} \bigcap L^2(T \cap \Omega^\varepsilon)$.

Proof. We prove below, that the interpolant $\boldsymbol{I_h}$ defined in (7.17) satisfies the properties defined in the Fortin lemma. First, we prove that the interpolation operator $\boldsymbol{I_h}$ satisfies (8.32). Indeed, we have, for any $(q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T}) \in M_h \times \mathbb{R}^{n_s} \times \mathbb{R}^r$ and $\boldsymbol{v} \in (H_0^1(T \cap \Omega^\varepsilon))^d$,

$$\begin{aligned}\tilde{b}_T\left(\boldsymbol{I_h v}, (q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T})\right) = &- \sum_{\tau \in \mathcal{T}_h(T)} \left(\operatorname{div} \boldsymbol{I_h v}, q_h\right)_\tau + \sum_{\tau \in \mathcal{T}_h(T)} \sum_{l=1}^{r} \mu_{T,l} \left(\boldsymbol{I_h v}, \boldsymbol{\varphi_{T,l}}\right)_\tau \\ &+ \sum_{F \in \mathcal{F}(T)} \sum_{j=1}^{s} \mu_{F,j} \left(\boldsymbol{I_h v}, \boldsymbol{\omega_{F,j}}\right)_F.\end{aligned}$$

By integration by parts, it comes

$$\tilde{b}_T\left(\boldsymbol{I_h v}, (q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T})\right) = \sum_{\tau\in\mathcal{T}_h(T)} \Big(\big(\boldsymbol{I_h v}, \boldsymbol{\nabla} q_h\big)_\tau - \sum_{f\in\partial\tau} \big(\boldsymbol{I_h v}\cdot\boldsymbol{n_f}, q_h|_f\big)_f \Big)$$
$$+ \sum_{\tau\in\mathcal{T}_h(T)} \sum_{l=1}^{r} \mu_{T,l}\big(\boldsymbol{I_h v}, \boldsymbol{\varphi_{T,l}}\big)_\tau + \sum_{F\in\mathcal{F}(T)} \sum_{f\in F} \sum_{j=1}^{s} \mu_{F,j}\big(\boldsymbol{I_h v}, \boldsymbol{\omega_{F,j}}\big)_f.$$

Now, using the fact that $\boldsymbol{\nabla} q_h|_\tau \in \mathbb{P}_{n-1}(\tau)$, $q_h|_f \in \mathbb{P}_n(f)$, $(\boldsymbol{\varphi_{T,l}}|_\tau)$ is a basis of $\mathbb{P}_{n-1}(\tau)$, $(\boldsymbol{\omega_{F,j}}|_f)$ is a basis of $\mathbb{P}_n(f)$ and the definition of the degrees of freedom, it follows

$$\tilde{b}_T\left(\boldsymbol{I_h v}, (q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T})\right) = \sum_{\tau\in\mathcal{T}_h(T)} \Big(\big(\boldsymbol{v}, \boldsymbol{\nabla} q_h\big)_\tau - \sum_{f\in\partial\tau} \big(\boldsymbol{v}\cdot\boldsymbol{n_f}, q_h|_f\big)_f \Big)$$
$$+ \sum_{\tau\in\mathcal{T}_h(T)} \sum_{l=1}^{r} \mu_{T,l}\big(\boldsymbol{v}, \boldsymbol{\varphi_{T,l}}\big)_\tau + \sum_{F\in\mathcal{F}(T)} \sum_{f\in F} \sum_{j=1}^{s} \mu_{F,j}\big(\boldsymbol{v}, \boldsymbol{\omega_{F,j}}\big)_f.$$

Finally, by integration by parts, we have

$$\begin{aligned}
\tilde{b}_T\left(\boldsymbol{I_h v}, (q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T})\right) &= -\sum_{\tau\in\mathcal{T}_h(T)} \big(\operatorname{div}\boldsymbol{v}, q_h\big)_\tau + \sum_{\tau\in\mathcal{T}_h(T)} \sum_{l=1}^{r} \mu_{T,l}\big(\boldsymbol{v}, \boldsymbol{\varphi_{T,l}}\big)_\tau \\
&\quad + \sum_{F\in\mathcal{F}(T)} \sum_{j=1}^{s} \mu_{F,j}\big(\boldsymbol{v}, \boldsymbol{\omega_{F,j}}\big)_F \\
&= -\big(\operatorname{div}\boldsymbol{v}, q_h\big)_T + \sum_{l=1}^{r} \mu_{T,l}\big(\boldsymbol{v}, \boldsymbol{\varphi_{T,l}}\big)_T + \sum_{F\in\mathcal{F}(T)} \sum_{j=1}^{s} \mu_{F,j}\big(\boldsymbol{v}, \boldsymbol{\omega_{F,j}}\big)_F \\
&= \tilde{b}_T\left(\boldsymbol{v}, (q_h, \boldsymbol{\mu_F}, \boldsymbol{\mu_T})\right).
\end{aligned}$$

Then, we prove that the interpolation operator $\boldsymbol{I_h}$ satisfies (8.33). Using standard scaling properties of the reference transformation and its inverse it is sufficient to show the following estimate on the reference element

$$|\boldsymbol{I_h^{\widehat{T}}}\widehat{\boldsymbol{u_\varepsilon}}|_{1,\widehat{T}} \leq \widehat{C}|\widehat{\boldsymbol{u_\varepsilon}}|_{1,\widehat{T}}.$$

On the one hand, from the continuity of the trace operator $\widehat{\boldsymbol{u_\varepsilon}} \mapsto \widehat{\boldsymbol{u_\varepsilon}}|_{\widehat{F}}$ we get

$$\left|\int_{\widehat{F}} \widehat{\boldsymbol{u_\varepsilon}}\cdot\boldsymbol{\omega_{F,i}}\right| \leq \widehat{C}\|\widehat{\boldsymbol{u_\varepsilon}}\|_{0,\widehat{F}} \leq \widehat{C}|\widehat{\boldsymbol{u_\varepsilon}}|_1,$$

and on the other hand, we have

$$\left|\int_{\widehat{T}} \widehat{\boldsymbol{u_\varepsilon}}\cdot\boldsymbol{\varphi_{T,l}}\right| \leq \widehat{C}\|\widehat{\boldsymbol{u_\varepsilon}}\|_{0,\widehat{T}} \leq \widehat{C}|\widehat{\boldsymbol{u_\varepsilon}}|_1,$$

which allows to conclude the proof. □

In this chapter, we have demonstrated the well-posedness of the discrete local problems in two and three dimensions. However, this necessitates the utilisation of novel finite elements, introduced in Chapters 4 and 5. This may present a limitation to the utilisation of this MsFEM,

particularly for high-order methods, given that the resolution of the local problems necessitates the implementation of new finite elements. In particular, in three dimensions, due to the absence of a generalisation of the family of finite elements introduced in Chapter 5 for any order, it is currently possible to use the MsFEM only up to the order $n = 2$.

Remark 8.15. In this section, we have chosen to prove the well-posedness of the discrete local problems by constructing a Fortin operator. However, it is important to note that it is not the only way to prove the well-posedness of complex and non-standard problems. Instead of constructing a Fortin operator, stabilisation technique can be used to address difficulties in ensuring that the inf-sup condition holds in finite element approximations. The stabilization techniques are particularly relevant for complex problems where standard discretization methods may fail. The choice of stabilization method depends on the problem at hand and requires balancing stability and accuracy. For example in the context of Hybrid High-Order (HHO) methods, stabilization technique is used to prove the inf-sup condition. The stabilization in HHO methods is designed to ensure both the stability and accuracy of the discrete solution, while allowing for a flexible choice of local approximation spaces. For HHO methods, the challenge is to ensure that this condition holds when high-order polynomials are used for approximation. In HHO, the stabilization is typically face-based, meaning that it involves terms that penalize the jump of the solution or its gradient across the interfaces between elements (faces) [64, 60].

Chapter 9

An error estimate

Contents

This chapter is dedicated to the derivation of the first error estimate for the Multi-scale Finite Element Method (MsFEM), introduced in Chapter 7, applied to solve the Stokes problem in perforated domains. The main result of this chapter reads in loose term

Theorem 9.1. Let $\boldsymbol{f}$ be sufficiently smooth. The following error bound holds between the solution $(\boldsymbol{u_\varepsilon}, p_\varepsilon)$ to the Stokes problem (7.1) in perforated periodic domains and its MsFEM approximation $(\boldsymbol{u_H}, p_H)$.

$$\begin{aligned}&\varepsilon^{-1}|\boldsymbol{u_\varepsilon} - \boldsymbol{u_H}|_{H,1} + \|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)}\\ &\leq C\Big(H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)} + H^n|p^*|_{H^{n+1}(\Omega)} + \left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\Big)\end{aligned}$$

where the constant C is independent of ε, H and $\boldsymbol{f}$.

The derivation of Theorem 9.1 is based on the usual finite element theory and homogenization of the Stokes problem, in particular on the homogenization bounds introduced in Chapter 6. The methodology used is inspired by the error derivation done in [111]. **In this chapter, we adopt definition of the perforated domain Ω^ε of Section 6.2.** In addition, in this chapter we eliminate the viscosity ν of the Stokes problem for the sake of clarity.

Through this chapter, for a domain $D \subset \mathbb{R}^d$, we use the Sobolev spaces $H^m(D)$, $H_0^m(D)$ for integer $m > 1$. The norms and semi-norms in the scalar and vector-valued versions of $H^m(D)$ are denoted by $\|\cdot\|_{H^m(D)}$ and $|\cdot|_{H^m(D)}$ respectively. For the sobolev space $L^2(D) = H^0(D)$, the associated norm is denoted $\|\cdot\|_{L^2(D)}$. Since the approach is non-conforming, we introduce the broken Sobolev space

$$H^1(\mathcal{T}_H) = \left\{ \boldsymbol{u} \in L^2(\Omega^\varepsilon) \; s.t. \; \boldsymbol{u}|_T \in H^1(T \cap \Omega^\varepsilon)^d, \; \forall \; T \in \mathcal{T}_H \right\}.$$

We also define the H^1-broken norm $|\boldsymbol{u}|_{H,1}$ in the H^1 broken space as

$$|\boldsymbol{u}|_{H,1} = \left(\sum_{T \in \mathcal{T}_H} |\boldsymbol{u}|^2_{H^1(T \cap \Omega^\varepsilon)} \right)^{\frac{1}{2}}.$$

This chapter is organized as follows. First, we introduce some useful technical lemmas for the derivation of the error estimate. Then, the last part of this chapter is dedicated to the proof of the main theorem.

9.1 Technical lemmas

Before proving the main theorem, in this section we introduce technical lemmas useful for the error estimate. First, we present some lemmas borrowed from the usual finite element theory and then we introduce lemmas related to perforated domains and oscillating functions.

9.1.1 Lemmas related to the reference element

As usual in finite element theory, most of the lemmas introduced in this part are proved first in the reference element and then in any element using scaling properties. Let $\widehat{T}$ be the reference polyhedron in $\mathbb{R}^d$ and let $T \in \mathcal{T}_H$ be a mesh element. Let π_T be the geometric mapping from $\widehat{T}$ to T. Classical examples of reference elements are the unit interval $[0,1]$ in dimension one, the unit simplex with vertices $\{(0,0),(1,0),(0,1)\}$ or the unit square $[0,1]^2$ in dimension two, the unit simplex with vertices $\{(0,0,0),(1,0,0),(0,1,0),(0,0,1)\}$ or the unit cube $[0,1]^3$ in dimension three. Then, for a given mesh element $T = \pi_T(\widehat{T})$, the vertices, edges, and faces are defined to be the image by the geometric transformation π_T of the vertices, edges, and faces of the reference element $\widehat{T}$. To avoid some technical complications, we assume that the mapping π_T is affine, i.e.

$$\begin{array}{rccl} \pi_T : & \widehat{T} & \longrightarrow & T \\ & \widehat{x} & \longmapsto & x = \pi_T(\widehat{x}) = \mathbb{A}_T \widehat{x} + b_T \end{array} \tag{9.1}$$

where $\mathbb{A}_T$ is an invertible matrix defined on $\mathbb{R}^{d \times d}$ and $b_T \in \mathbb{R}^d$.

Remark 9.2. The results presented in this chapter remain valid for non-affine transformations, i.e. on general mesh shape (see for example [76, Chapter 13] for results on non-affine meshes). Another option is to consider a finite family (possibly large) of reference elements, allowing to recover all the elements $T \in \mathcal{T}_H$ with affine mappings.

We note $\widehat{v} = v \circ \pi_T$, the function v on T expressed on local coordinates on $\widehat{T}$. The chain rule for weak derivatives is given by

$$\frac{\partial \widehat{v}}{\partial \widehat{x_i}} = \sum_{j=1}^{n} \frac{\partial v}{\partial x_j} \frac{\partial x_j}{\partial \widehat{x_i}} = \sum_{j=1}^{n} (\mathbb{A}_T)_{ij} \frac{\partial v}{\partial x_j},$$

and the transformation rule for integrals by

$$\int_{T = \pi_T(\widehat{T})} v \mathrm{d}x = \int_{\widehat{T}} (v \circ \pi_T) |\det(\mathbb{A}_T)| \mathrm{d}\widehat{x} = \int_{\widehat{T}} \widehat{v} |\det(\mathbb{A}_T)| \mathrm{d}\widehat{x}.$$

Then the following transformations of Sobolev semi-norms holds [76, Chapter 11].

Lemma 9.3. Let $k \geq 0$ and $1 \leq p \leq \infty$. there exists $C > 0$ such that for all T and $\boldsymbol{v} \in W^{k,p}(T)$, the function $\widehat{\boldsymbol{v}} = \boldsymbol{v} \circ \pi_T$ satisfies

$$|\widehat{\boldsymbol{v}}|_{W^{k,p}(\widehat{T})} \leq C \, \|\mathbb{A}_T\|^k \, |\det(\mathbb{A}_T)|^{-\frac{1}{p}} \, |\boldsymbol{v}|_{W^{k,p}(T)},$$

$$|\boldsymbol{v}|_{W^{k,p}(T)} \leq C \, \|\mathbb{A}_T^{-1}\|^k \, |\det(\mathbb{A}_T)|^{\frac{1}{p}} \, |\widehat{\boldsymbol{v}}|_{W^{k,p}(\widehat{T})}.$$

For a given element T, we define the *diameter* $H_T := \max_{x_1, x_2 \in T} \|x_1 - x_2\|$, *the insphere diameter* $\rho_T := 2 \max\{\rho > 0 : B_\rho(x) \subset T \text{ for some } x \in T\}$ (i.e. the diameter of the largest ball contained in T), *the condition number* $\sigma_T := \dfrac{H_T}{\rho_T}$.

Lemma 9.4. Let π_T be an affine mapping defined as in (9.1), such that $T = \pi_T(\widehat{T})$. Then,

$$|\det(\mathbb{A}_T)| = \frac{|T|}{|\widehat{T}|}, \qquad \|\mathbb{A}_T\| \leq \frac{H_T}{\rho_{\widehat{T}}}, \qquad \|\mathbb{A}_T^{-1}\| \leq \frac{H_{\widehat{T}}}{\rho_T}.$$

9.1.2 Lemmas borrowed from the usual finite element theory

The following two lemmas, Lemmas 9.5 and 9.6, are Poincaré type inequalities.

Lemma 9.5. There exists a constant C (depending only on the regularity of the mesh) independent of H such that for any $T \in \mathcal{T}_H$, $\forall \boldsymbol{v} \in H^1(T)$, verifying $\displaystyle\int_F \boldsymbol{v} = 0$ for some $F \subset \partial T$,

$$\|\boldsymbol{v}\|_{L^2(T)} \leq CH \|\boldsymbol{\nabla v}\|_{L^2(T)}. \tag{9.2}$$

Proof. First, we consider the reference element $\widehat{T}$, with face $\widehat{F} \subset \partial \widehat{T}$. If there does not exist

$C > 0$ such that, for all functions $\boldsymbol{v} \in H^1(\widehat{T})$ verifying $\displaystyle\int_{\widehat{F}} \boldsymbol{v} = 0$,

$$\|\boldsymbol{v}\|^2_{L^2(\widehat{T})} \leq C\|\boldsymbol{\nabla v}\|^2_{L^2(\widehat{T})}.$$

then, there exists a sequence $\boldsymbol{v_n} \in H^1(\widehat{T})$ such that

$$1 = \int_{\widehat{T}} |\boldsymbol{v_n}(x)|^2 \mathrm{d}x > n \int_{\widehat{T}} |\boldsymbol{\nabla v_n}(x)|^2 \mathrm{d}x. \tag{9.3}$$

In particular (9.3), implies that $\boldsymbol{v_n}$ is bounded in $H^1(\widehat{T})$. By applying the Rellich theorem, there exists a sub-sequence $\boldsymbol{v'_n}$ which converges in $L^2(\widehat{T})$. In addition, (9.3) shows that $\boldsymbol{\nabla v'_n}$ converges to zero in $L^2(\widehat{T})$. Therefore $\boldsymbol{v'_n}$ is a Cauchy sequence in $H^1(\widehat{T})$, which is a Hilbert space, thus it converges in $H^1(\widehat{T})$ to a limit $\boldsymbol{v}$. As we have

$$\int_{\widehat{T}} |\boldsymbol{\nabla v}(x)|^2 \mathrm{d}x = \lim_{n\to\infty} \int_{\widehat{T}} |\boldsymbol{\nabla v_n}(x)|^2 \mathrm{d}x < \lim_{n\to\infty} \frac{1}{n} = 0,$$

we deduce that $\boldsymbol{v}$ is constant in $\widehat{T}$. Besides, since $\boldsymbol{v}$ is of zero integral on the border $\partial\widehat{T}$, $\boldsymbol{v}$ is null on the border $\partial\widehat{T}$. We conclude that $\boldsymbol{v}$ is null in $\widehat{T}$. Besides, we have

$$\int_{\widehat{T}} |\boldsymbol{v}(x)|^2 \mathrm{d}x = \lim_{n\to\infty} \int_{\widehat{T}} |\boldsymbol{v_n}(x)|^2 \mathrm{d}x = 1,$$

which is contradictory with the fact that $\boldsymbol{v} = \boldsymbol{0}$. Therefore, we have proved that in the reference element $\widehat{T}$ there exists $C > 0$ such that, for all functions $\boldsymbol{v} \in H^1(\widehat{T})$ verifying $\displaystyle\int_{\widehat{F}} \boldsymbol{v} = 0$,

$$\|\boldsymbol{v}\|^2_{L^2(\widehat{T})} \leq C\|\boldsymbol{\nabla v}\|^2_{L^2(\widehat{T})}.$$

Now, with usual scaling arguments (see Lemmas 9.3 and 9.4), it comes, that in any element T of size H,

$$\|\boldsymbol{v}\|^2_{L^2(T)} \leq CH^d \|\widehat{\boldsymbol{v}}\|^2_{L^2(\widehat{T})} \leq CH^d \|\boldsymbol{\nabla}\widehat{\boldsymbol{v}}\|^2_{L^2(\widehat{T})} \leq CH^d \frac{H^2}{H^d} \|\boldsymbol{\nabla v}\|^2_{L^2(T)} \leq CH^2 \|\boldsymbol{\nabla v}\|^2_{L^2(T)},$$

which proves (9.2). □

Lemma 9.6. There exists a constant C independent of H (depending on the regularity of the mesh) such that for any $T \in \mathcal{T}_H$, for all $\boldsymbol{v} \in H^1(T)$ and any face $F \subset \partial T$,

$$\|\boldsymbol{v}\|^2_{L^2(F)} \leq C \left(H^{-1} \|\boldsymbol{v}\|^2_{L^2(T)} + H \|\boldsymbol{\nabla} \boldsymbol{v}\|^2_{L^2(T)} \right). \tag{9.4}$$

Under the additional assumption that $\displaystyle\int_F \boldsymbol{v} = 0$, we have

$$\|\boldsymbol{v}\|^2_{L^2(F)} \leq CH \|\boldsymbol{\nabla} \boldsymbol{v}\|^2_{L^2(T)}, \tag{9.5}$$

and

$$\|\boldsymbol{v}\|^2_{H^{1/2}(F)} \leq C \|\boldsymbol{\nabla} \boldsymbol{v}\|^2_{L^2(T)}. \tag{9.6}$$

Proof. The inequality (9.4) is obtained using the continuity of the trace operator from $H^1(\widehat{T})$ to $L^2(\widehat{F})$ in the reference element which reads

$$\|\boldsymbol{v}\|^2_{L^2(\widehat{F})} \leq C \left(\|\boldsymbol{v}\|^2_{L^2(\widehat{T})} + \|\boldsymbol{\nabla} \boldsymbol{v}\|^2_{L^2(\widehat{T})} \right). \tag{9.7}$$

and, then usual scaling arguments. Next, combining (9.7) and Lemma 9.5, we obtain the inequality (9.5). For the inequality (9.6), we use the continuity of the trace operator from $H^1(\widehat{T})$ to $H^{1/2}(\widehat{T})$, which reads

$$\|\boldsymbol{v}\|^2_{H^{1/2}(\widehat{F})} \leq C \left(\|\boldsymbol{v}\|^2_{L^2(\widehat{T})} + \|\boldsymbol{\nabla} \boldsymbol{v}\|^2_{L^2(\widehat{T})} \right). \tag{9.8}$$

Then, usual scaling arguments lead to

$$\begin{aligned}
\|\boldsymbol{v}\|^2_{H^{1/2}(F)} &= \|\boldsymbol{v}\|^2_{L^2(F)} + |\boldsymbol{v}|^2_{H^{1/2}(F)} \\
&\leq CH^{d-1} \|\widehat{\boldsymbol{v}}\|^2_{L^2(\widehat{F})} + CH^{d-2} |\widehat{\boldsymbol{v}}|^2_{H^{1/2}(\widehat{F})} \\
&\leq CH^{d-1} |\widehat{\boldsymbol{v}}|^2_{H^1(\widehat{T})} + CH^{d-2} |\widehat{\boldsymbol{v}}|^2_{H^1(\widehat{T})} \\
&\leq CH^{d-1} \frac{H^2}{H^d} |\boldsymbol{v}|^2_{H^1(T)} + CH^{d-2} \frac{H^2}{H^d} |\boldsymbol{v}|^2_{H^1(T)} \\
&\leq C(1+H) |\boldsymbol{v}|^2_{H^1(T)} \\
&\leq C |\boldsymbol{v}|^2_{H^1(T)},
\end{aligned}$$

which proves (9.6). □

Lemma 9.7. Let Π_H^{n-1} be the $L^2(\Omega)$-orthogonal projection on the space of piecewise polynomials of degree smaller or equal to $n-1$ on $\mathcal{T}_H$. Then, for any $\boldsymbol{f} \in H^1(\Omega)^d$,

$$\|\boldsymbol{f} - \Pi_H^{n-1} \boldsymbol{f}\|_{L^2(\Omega)} \leq CH^n |\boldsymbol{f}|_{H^n(\Omega)},$$

with a constant $C > 0$ depending only on the regularity of $\mathcal{T}_H$.

Proof. This is a standard finite element interpolation result [75, Proposition 1.12]). □

Lemma 9.8. There exists a bounded linear operator $I_H^n : H^2(\Omega)^d \to H^1(\Omega)^d$ such that $I_H^n \boldsymbol{v}$ is a polynomial of degree smaller or equal to n on any face $F \in \mathcal{F}_H$ for any $\boldsymbol{v} \in H^2(\Omega)^d$ and

$$|I_H^n \boldsymbol{v} - \boldsymbol{v}|_{H^1(\Omega)} \leq C H^n |\boldsymbol{v}|_{H^{n+1}(\Omega)},$$

and

$$\|I_H^n \boldsymbol{v} - \boldsymbol{v}\|_{L^2(\Omega)} \leq C H^{n+1} |\boldsymbol{v}|_{H^{n+1}(\Omega)},$$

with a constant $C > 0$ depending only on the regularity of $\mathcal{T}_H$.

Proof. Given that we consider $\boldsymbol{v} \in H^2(\Omega)^d$, one can simply take I_H^n as the usual Lagrange interpolation operator on $\mathbb{P}_n$ finite element if $\mathcal{T}_H$ is a triangular mesh. Otherwise, we consider $\widehat{\mathcal{T}_H}$ a submesh of $\mathcal{T}_H$ which consists of triangles only. To construct $\widehat{\mathcal{T}_H}$, one only needs to remesh all the elements T in triangles, without adding nodes on ∂T. We can now define I_H^n as the usual Lagrange interpolation operator on $\mathbb{P}_n$ finite elements on $\widehat{\mathcal{T}_H}$. Then, the error bounds are given by classical Finite Element results (see for example [75, Proposition 1.12]). □

9.1.3 Lemmas related to perforated domains and oscillating functions

In this section, we introduce lemmas related to perforated domains and oscillating functions. The main lemmas of this section are Lemmas 9.21, 9.22 and 9.24 which are Poincaré type inequalities in perforated domains. In order to prove these lemmas, first, we introduce preliminary lemmas, Lemmas 9.11 to 9.13, which are also Poincaré type inequalities but in periodicity cells Y_i^ε, and Lemma 9.14 which introduces a classical extension operator.

As previously, most of the following lemmas are proved first in the reference periodicity cell Y (see for example Figure 6.3) and then in any cell Y_i^ε using scaling properties. The precise definition of the unit cell can be found in Section 6.2.1. We recall that the reference cell is the unit square in two dimensions and the unit cube in three dimensions. Then, for a given cell $Y_i^\varepsilon = \pi_i^\varepsilon(Y)$, the vertices, edges, and faces are defined to be the image by the geometric transformation π_i^ε of the vertices, edges, and faces of the reference cell Y. We recall that the affine transformation from Y is defined as follows.

Definition 9.9. We note π_i^ε the linear continuous invertible application, composed of a translation and an homothety of ratio $\frac{1}{\varepsilon}$, which maps any cell Y_i^ε onto the reference cell Y

$$\begin{array}{rccl} \pi_i^\varepsilon : & Y_i^\varepsilon & \longrightarrow & Y \\ & x & \longmapsto & y = \frac{x}{\varepsilon} + \text{translation} \end{array} \tag{9.9}$$

For the proof of Lemma 9.13, we need also to introduce the notion of "adjacent cells".

Definition 9.10. Two cells Y_1^ε and Y_2^ε are said to be adjacent if they share a common face Γ^ε.

9.1.3.1 Preliminary results

Lemma 9.11. There exists a constant $C > 0$ independent of ε such that, for any Y_i^ε and for all $\boldsymbol{v} \in H^1(Y^\varepsilon)$ vanishing on $\partial Y_{i,S}^\varepsilon$,

$$\|\boldsymbol{v}\|_{L^2(Y_i^\varepsilon)} \leq C\varepsilon \|\boldsymbol{\nabla v}\|_{L^2(Y_i^\varepsilon)}.$$

Proof. Lemma 9.11 is a classical Poincaré inequality. It can be proved with a proof similar to Lemma 9.5 □

Lemma 9.12. There exists a constant $C > 0$ independent of ε such that, for any cell Y_i^ε and for any face $\Gamma_i^\varepsilon \subset \partial Y_i^\varepsilon$ and for all $\boldsymbol{v} \in H^1(Y_i^\varepsilon)$ vanishing on $\partial Y_{i,S}^\varepsilon$,

$$\|\boldsymbol{v}\|_{L^2(\Gamma_i^\varepsilon)} \leq C\sqrt{\varepsilon} \|\boldsymbol{\nabla v}\|_{L^2(Y_i^\varepsilon)}.$$

Proof. Lemma 9.12 is proved using the continuity of the trace operator from $H^1(Y)$ to $L^2(\Gamma)$, combined with Lemma 9.11 and usual scaling arguments applying the mapping (9.9) (as in Lemma 9.6). □

Lemma 9.13. There exists a constant $C > 0$ independent of ε such that, for any adjacent cells Y_1^ε and Y_2^ε (in the meaning of Definition 9.10) and for all $\boldsymbol{v} \in H^1(Y_1^\varepsilon \cup Y_2^\varepsilon)$,

$$\|\boldsymbol{v}\|_{L^2(Y_2^\varepsilon)} \leq C \left(\|\boldsymbol{v}\|_{L^2(Y_1^\varepsilon)} + \varepsilon \|\boldsymbol{\nabla v}\|_{L^2(Y_1^\varepsilon \cup Y_2^\varepsilon)} \right).$$

Proof. Lemma 9.13 is a classical Poincaré inequality. It can be proved with a proof similar to Lemma 9.5. □

Lemma 9.14. Let $T \subset \mathcal{T}_H$. Let $Y_{i_{in}}^\varepsilon$ and $Y_{i_{out}}^\varepsilon$ be two adjacent cells (in the meaning of Definition 9.10), such that $Y_{i_{in}}^\varepsilon \subset T$, $Y_{i_{out}}^\varepsilon \not\subset T$ and $Y_{i_{out}}^\varepsilon \cap T \neq \emptyset$. Assume that the cells do not contain perforations. Then, for all $\boldsymbol{v} \in H^1(Y_{i_{in}}^\varepsilon \cup Y_{i_{out}}^\varepsilon \cap T)$ there exists an extension operator

$$\tilde{P} \in \mathcal{L}\left(H^1(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon \cap T); H^1(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon)\right),$$

such that

$$\|\boldsymbol{\nabla} \tilde{P}\boldsymbol{v}\|_{L^2(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon)} \leq C \|\boldsymbol{\nabla v}\|_{L^2(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon \cap T)},$$

and

$$\|\tilde{P}\boldsymbol{v}\|_{L^2(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon)} \leq C \|\boldsymbol{v}\|_{L^2(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon \cap T)}.$$

Proof. Let us construct an inclined cell $\tilde{Y}_i^\varepsilon$ obtained by the translation of $\partial T \cap Y_{i_{out}}^\varepsilon$ of a distance ε towards $Y_{i_{in}}^\varepsilon$. Then, the extension of $\boldsymbol{v}$ in $Y_{i_{out}}^\varepsilon \setminus (Y_{i_{out}}^\varepsilon \cap T)$ is given by the horizontal-symmetry of $\tilde{Y}_i^\varepsilon$ with respect to the axis passing by $\partial T \cap Y_{i_{out}}^\varepsilon$. Afterward, we use the trick of [58, Lemma 3] (which consists in applying the extension to $\boldsymbol{v} - \mathcal{M}(\boldsymbol{v})$, where $\mathcal{M}(\boldsymbol{v})$ stands for the average of $\boldsymbol{v}$ over $Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon \cap T$), which gives the estimates. It should be noted that the constant C in the estimates depends on the face considered, but we recall that ∂T consists only in a finite number of faces (we consider the maximun of these constants). We illustrate this construction with a simple example in two dimensions in Figure 9.1. For each point (y_1, y_2) in $Y_{i_{out}}^\varepsilon \setminus (Y_{i_{out}}^\varepsilon \cap T)$, we define the unique value $y_1(y_2)$ of the abcisse such that the point $(y_1(y_2), y_2) \in \partial T \cap Y_{i_{out}}^\varepsilon$. Then, $\tilde{P}\boldsymbol{v}(y_1, y_2) = \boldsymbol{v}(2y_1(y_2) - y_1, y_2)$. The generalization in any dimension is straightforward.

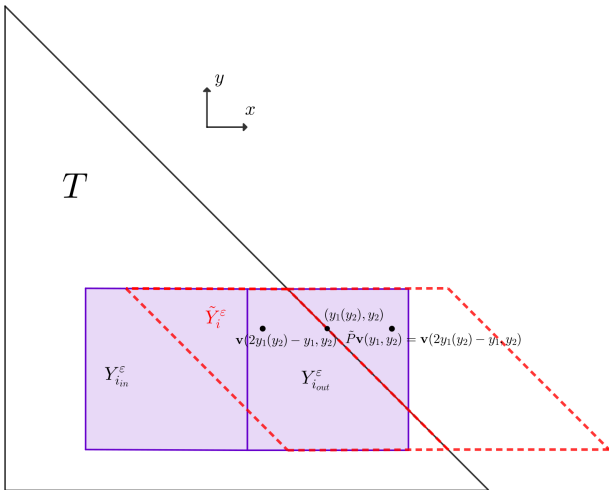


Figure 9.1: Example, in two dimensions, of the geometric configuration of Lemma 9.14.

□

9.1.3.2 Main lemmas

Assumption 9.15. We assume (and this is a purely technical assumption that does not matter for the numerical practice) that the boundary ∂T of any element $T \in \mathcal{T}_H$ consists of a set of faces F, included in a hyperplane whose unit normal vector $\boldsymbol{n_F}$ belongs to $\mathbb{R}\mathbb{Z}^d$ (proportional to a vector with integer coordinates), i.e.

$$T = \bigcap_{k=1}^{\mathcal{N}_F} \{\boldsymbol{x} \text{ such that } \boldsymbol{x} \cdot \boldsymbol{n_{F_k}} \leq c_{F_k}\}$$

with $\mathcal{N}_F$ the number of faces of the element T, $\boldsymbol{n_F} \in S^d$, $c_F \in \mathbb{R}$ [81, 82, 139]. The unit normal vector is of the form $\frac{1}{q}(p_1, p_2, \ldots, p_d)^t$, where p_1, p_2, …, p_d are integers, and $q = \sqrt{\sum_{i=1}^{d} p_i^2}$ a real number. We supposed that the p_i are not too large, i.e. $q^2 \leq C^2$ for a constant C, independent of the face considered in the mesh and of the mesh size H. Figure 9.2 shows an example of such a configuration in two dimensions.

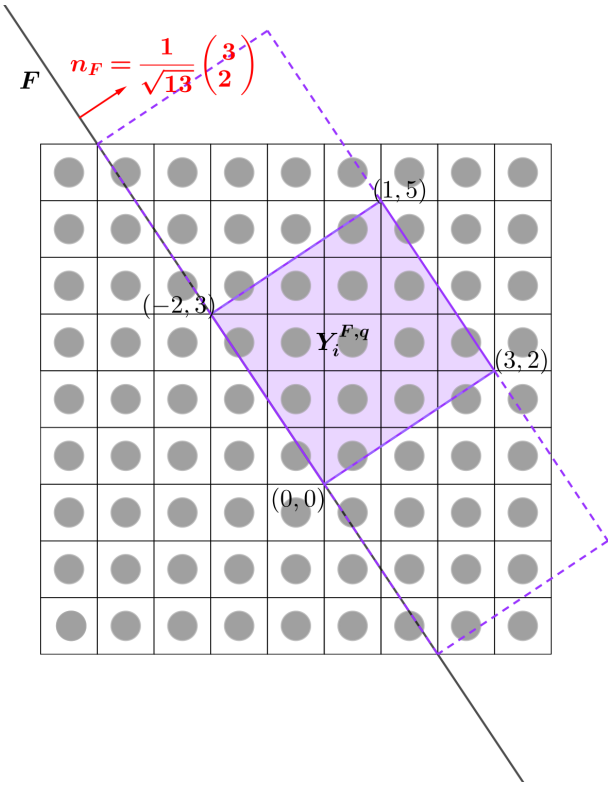


Figure 9.2: Example, in two dimensions, of the geometric configuration of Assumption 9.15.

Remark 9.16. Assumption 9.15 is a classical assumption when dealing with Multi-scale Finite Element Methods (see for example [122, Theorem 2.2] or the works of [134, 81, 82]).

Assumption 9.15 yields periodicity tangentially to the hyperplane F. Indeed, we can consider a new rotated cell, depending on the orientation of the face F and of size q, denoted $Y_i^{F,q}$, whose one of its face is parallel to F. Figure 9.2 shows an example, in two dimensions, of such a cell, which corresponds to the square of size $\sqrt{p_1^2+p_2^2}$ with the corners $(0,0)$, (p_1,p_2), (p_1-p_2,p_1+p_2), $(-p_2,p_1)$, with $p_1=3$ and $p_2=2$. This cell can be treated as the new periodic cell. Indeed, one can choose the new set of coordinates x' (obtained by the transformation which maps the canonical vector $\boldsymbol{e_d}=(0,\dots,0,1)$ to the normal vector $\boldsymbol{n_F}$), such that the quantities are periodic along F with period q possibly greater than 1. Now, any function which is periodic in the coordinates x, with period ε, is also periodic in the coordinates x' with period $q\varepsilon$. A function that is a function of the fast variable $y=\dfrac{x}{\varepsilon}$ remains a function of the corresponding fast variable $y'=\dfrac{x'}{\varepsilon}$ and if it is periodic in y, with period 1, it is periodic in y', with period q.

Remark 9.17. If one has to choose a period q, one can always choose an integer multiple of q such that each cell $Y_i^{F,q}$ contains one perforation at a distance of order at least 1 of the boundary $\partial Y_i^{F,q}$, which ensures that one perforation is at a distance of at least 1 of the border ∂T. This precaution guarantees that the Poincaré constant C of Lemma 9.12 does not depend on the way the boundary ∂T cuts the cells $Y_i^{F,q}$.

In what follows, we consider a large enough period q such that each cell $Y_i^{F,q}$ contains a

perforation at a distance of at least 1 of the border $\partial Y_i^{F,q}$, and we consider cells of size $q\varepsilon$.

Definition 9.18. For any $T \in \mathcal{T}_H$, we define the domain S_T (a domain which is slightly larger than T such that $T \subset S_T$) as the dilation centered at T with a scale factor δ, for some $\delta > 1$ (see Figure 9.3).

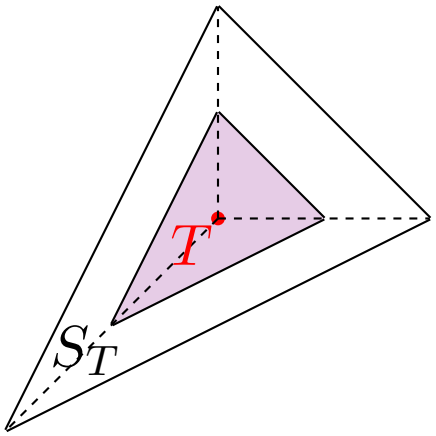


Figure 9.3: Definition of the domain S_T.

Remark 9.19. The use of such a domain S_T is very common when dealing with Multi-scale Finite Element Methods, for example in the case of oversamplig method (see [71, 105, 46]).

Assumption 9.20. Let $\Gamma^{F,q\varepsilon}$ be any face separating two cells $Y_+^{F,q\varepsilon}$ and $Y_-^{q\varepsilon}$ ($Y_+^{F,q\varepsilon}$ and $Y_-^{F,q\varepsilon}$ being samples of $Y_i^{F,q\varepsilon}$, the ε-scaled version of $Y_i^{F,q}$). For any $T \in \mathcal{T}_H$, the domain S_T is chosen such that there exists $\varepsilon_0 > 0$, such that for all $\varepsilon < \varepsilon_0$, if $\Gamma^{F,q\varepsilon} \cap \partial T \neq \emptyset$ then either $Y_+^{F,q\varepsilon}$ or $Y_-^{F,q\varepsilon}$ is included in S_T.

Lemma 9.21. Suppose $H \geq \gamma\varepsilon$ with some big enough γ. Let $T \subset \mathcal{T}_H$ and consider a domain S_T as built in Definition 9.18. Then, under Assumptions 9.15 and 9.20, for any $\boldsymbol{v} \in H^1(S_T)$ vanishing on $B^\varepsilon \cap S_T$,

$$\|\boldsymbol{v}\|_{L^2(\partial T)} \leq C\sqrt{\varepsilon}|\boldsymbol{v}|_{H^1(S_T)}.$$

The constants $\gamma > 0$ and $C > 0$ here depend only on the regularity of the mesh $\mathcal{T}_H$ and on the perforation pattern B^ε.

Proof. Let F be a face of T (there are only a finite number of such faces F). Assumption 9.15 and the assumption $H > \gamma\varepsilon$ (with some large enough γ) ensure that each face F consists of a finite number of small edges $\Gamma_i^{F,q\varepsilon} \subset \partial Y_i^{F,q\varepsilon}$ (see for example Figure 9.4), i.e.

$$F = \sum_{i \in I_{in}} \Gamma_i^{F,q\varepsilon} + \sum_{i \in I_{out}} \Gamma_i^{F,q\varepsilon} \cap F,$$

with

$$\begin{cases} I_{in} = \{i \in \mathbb{N} \text{ such that } \Gamma_i^{F,q\varepsilon} \subset F\}, \\ I_{out} = \{i \in \mathbb{N} \text{ such that } \Gamma_i^{F,q\varepsilon} \not\subset F \text{ and } \Gamma_i^{F,q\varepsilon} \cap F \neq \emptyset\}. \end{cases}$$

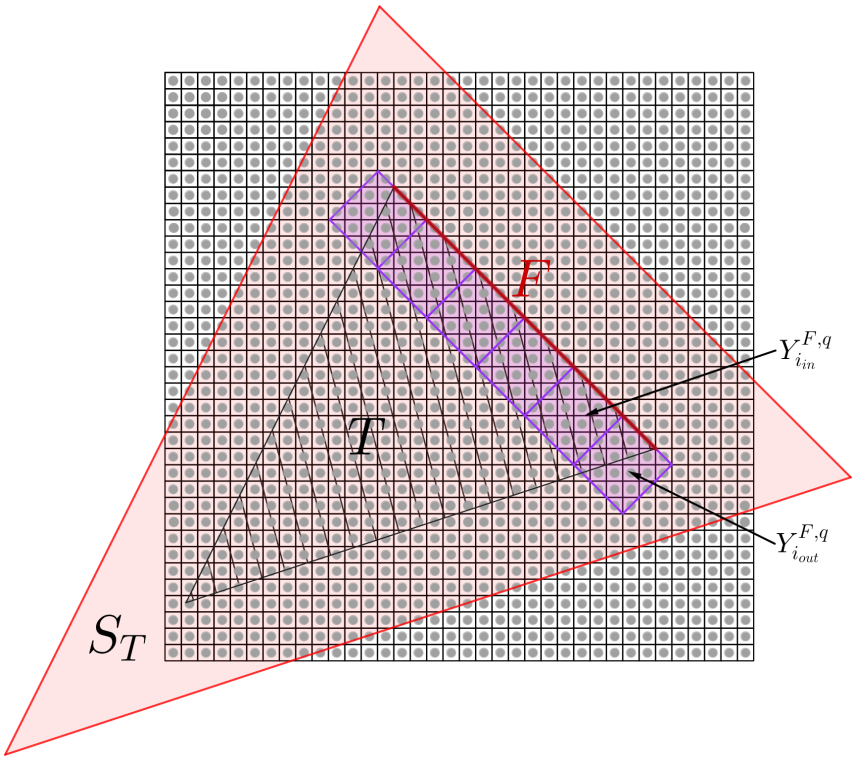


Figure 9.4: Example, in two dimensions, of the geometric configuration considered in the proof of Lemma 9.21.

Considering the domain S_T instead of T with Assumption 9.20 ensures that for all $\Gamma_i^{F,q\varepsilon}$, with $i \in I_{in} \cup I_{out}$, one of the cells $Y_i^{F,q\varepsilon}$ sharing the face $\Gamma_i^{F,q\varepsilon}$ is included in S_T. Besides, according to Lemma 9.12, there exits $C > 0$ such that

$$\|\boldsymbol{v}\|^2_{L^2(\Gamma_i^{F,q\varepsilon})} \leq C\varepsilon \|\boldsymbol{\nabla} \boldsymbol{v}\|^2_{L^2(Y_i^{F,q\varepsilon})}.$$

We have,

$$\begin{aligned}
\|\boldsymbol{v}\|^2_{L^2(F)} &= \sum_{i \in I_{in}} \|\boldsymbol{v}_{\boldsymbol{\Gamma_i^{F,q\varepsilon}}}\|^2_{L^2(\Gamma_i^{F,q\varepsilon})} + \sum_{i \in I_{out}} \|\boldsymbol{v}_{\boldsymbol{\Gamma_i^{F,q\varepsilon}}}\|^2_{L^2(\Gamma_i^{F,q\varepsilon} \cap T)} \\
&\leq \sum_{i \in I_{in} \cup I_{out}} C\varepsilon \|\boldsymbol{\nabla} \boldsymbol{v}_{\boldsymbol{Y_i^{F,q\varepsilon}}}\|^2_{L^2(Y_i^{F,q\varepsilon})} \\
&\leq C\varepsilon \|\boldsymbol{\nabla} \boldsymbol{v}\|^2_{L^2(S_T)}.
\end{aligned}$$

Finally, by summing over the faces F,

$$\|\boldsymbol{v}\|^2_{L^2(\partial T)} \leq \max_{F \in \partial T} \{C_F\} \varepsilon \|\boldsymbol{\nabla} \boldsymbol{v}\|^2_{L^2(S_T)},$$

which concludes the proof. □

Lemma 9.22. Suppose $H \geq \gamma\varepsilon$. Then, for any $T \in \mathcal{T}_H$ and any $\boldsymbol{v} \in H^1(T)$ vanishing on $B^\varepsilon \cap T$,

$$\|\boldsymbol{v}\|_{L^2(T)} \leq C\varepsilon|\boldsymbol{v}|_{H^1(T)}, \tag{9.10}$$

with $C > 0$ independent of ε and of H.

Proof. First, we note

$$\begin{cases} I_{in} = \{i \in \mathbb{N} \text{ such that } Y_i^{q\varepsilon} \subset T\}, \\ I_{out} = \{i \in \mathbb{N} \text{ such that } Y_i^{q\varepsilon} \not\subset \partial T \text{ and } Y_i^{q\varepsilon} \cap T \neq \emptyset\}. \end{cases}$$

Step 1 For $i \in I_{in}$, the cell Y_i^ε is fully included in the element T, which ensures that $Y_{i,S}^\varepsilon \cap T \neq \emptyset$, allowing to use the Poincaré inequality. From Lemma 9.11, it comes

$$\|\boldsymbol{v}\|_{L^2(Y_i^\varepsilon)} \leq C\varepsilon\|\boldsymbol{\nabla v}\|_{L^2(Y_i^\varepsilon)}.$$

Step 2 For $i \in I_{out}$, Y_i^ε is not fully included in T. Then it is possible that $Y_{i,S}^\varepsilon \cap T = \emptyset$ which prevents us to use the Poincaré inequality in $Y_i^\varepsilon \cap T$. We overcome this difficulty, by considering for each cell partially included in T an adjacent cell which is fully included in T. Indeed the assumption $H \geq \gamma\varepsilon$ (with some big enough γ) ensures that for any $i_{out} \in I_{out}$, there exists at least a $i_{in} \in I_{in}$ such that $Y_{i_{out}}^\varepsilon$ and $Y_{i_{in}}^\varepsilon$ are adjacent (in the meaning of Definition 9.10). Now, we consider $\boldsymbol{v}(x) \in H^1(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon \cap T)$, and we rescale it by ε which leads to $\hat{\boldsymbol{v}}(y) = \boldsymbol{v}(\frac{x}{\varepsilon}) \in H^1(Y_{i_{out}} \cup Y_{i_{in}} \cap \frac{1}{\varepsilon}T)$. We extend $\boldsymbol{v}(x)$ by $\mathbf{0}$ inside $B^\varepsilon \cap T$. Then, we are able to consider an extension $\boldsymbol{w}(x)$ of $\boldsymbol{v}(x)$, as built in Lemma 9.14, such that $\boldsymbol{w}(x) \in H^1(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon)$ and similarly we rescale it by ε which leads to $\hat{\boldsymbol{w}} \in H^1(Y_{i_{out}} \cup Y_{i_{in}})$. First, according to Lemma 9.13, there exists a constant $C(Y_{i_{out}} \cup Y_{i_{in}}) > 0$ such that

$$\|\hat{\boldsymbol{w}}\|_{L^2(Y_{i_{out}})} \leq C(Y_{i_{out}} \cup Y_{i_{in}}) \left(\|\hat{\boldsymbol{w}}\|_{L^2(Y_{i_{in}})} + \|\boldsymbol{\nabla}\hat{\boldsymbol{w}}\|_{L^2(Y_{i_{out}} \cup Y_{i_{in}})}\right).$$

With Lemma 9.14, we have

$$\|\boldsymbol{\nabla}\hat{\boldsymbol{w}}\|_{L^2(Y_{i_{out}} \cup Y_{i_{in}})} \leq C\|\boldsymbol{\nabla}\hat{\boldsymbol{v}}\|_{L^2(Y_{i_{out}} \cup Y_{i_{in}} \cap \frac{1}{\varepsilon}T)}.$$

which leads to

$$\|\hat{\boldsymbol{v}}\|_{L^2(Y_{i_{out}} \cap \frac{1}{\varepsilon}T)} \leq C\left(\|\hat{\boldsymbol{v}}\|_{L^2(Y_{i_{in}})} + \|\boldsymbol{\nabla}\hat{\boldsymbol{v}}\|_{L^2(Y_{i_{out}} \cup Y_{i_{in}} \cap \frac{1}{\varepsilon}T)}\right),$$

and then by usual scaling arguments, it comes

$$\|\boldsymbol{v}\|_{L^2(Y_{i_{out}}^\varepsilon \cap T)} \leq C\left(\|\boldsymbol{v}\|_{L^2(Y_{i_{in}}^\varepsilon)} + C\varepsilon\|\boldsymbol{\nabla v}\|_{L^2(Y_{i_{out}}^\varepsilon \cup Y_{i_{in}}^\varepsilon \cap T)}\right).$$

By using the Poincaré inequality in $Y_{i_{in}}^\varepsilon$, it follows

$$\|\boldsymbol{v}\|_{L^2(Y_{i_{in}}^\varepsilon)} \leq C\varepsilon\|\boldsymbol{\nabla v}\|_{L^2(Y_{i_{in}}^\varepsilon)}.$$

At the end, we conclude that

$$\|\boldsymbol{v}\|_{L^2(Y^\varepsilon_{i_{out}}\cap T)} \leq C\varepsilon\|\boldsymbol{\nabla v}\|_{L^2(Y^\varepsilon_{i_{out}}\cup Y^\varepsilon_{i_{in}}\cap T)}.$$

Step 3 We are now able to prove the estimation given in the lemma. Summing over the cells in the element T, it comes

$$\begin{aligned}\|\boldsymbol{v}\|_{L^2(T)} &= \sum_{i\in I_{in}} \|\boldsymbol{v}\|_{L^2(Y_i^\varepsilon)} + \sum_{i\in I_{out}} \|\boldsymbol{v}\|_{L^2(Y_i^\varepsilon\cap T)} \\ &\leq \sum_{i\in I_{in}} C\varepsilon\|\boldsymbol{\nabla v}\|_{L^2(Y_i^\varepsilon)} + \sum_{i\in I_{out}} C\varepsilon\|\boldsymbol{\nabla v}\|_{L^2(Y_i^\varepsilon\cup Y^\varepsilon_{i_{in}}\cap T)} \\ &\leq C\varepsilon\|\boldsymbol{\nabla v}\|_{L^2(T)} = C\varepsilon|\boldsymbol{v}|_{H^1(T)},\end{aligned}$$

where $Y^\varepsilon_{i_{in}}$ ($i_{in}\in I_{in}$) is an adjacent cell to Y_i^ε for $i\in I_{out}$. □

Remark 9.23. In this proof, for each $i\in I_{out}$, we add to the cell $Y^\varepsilon_{i_{out}}$ a cell $Y^\varepsilon_{i_{in}}$ with $i_{in}\in I_{in}$. Consequently, in this estimate we consider an extra $\mathrm{card}(I_{out})$ contributions of the cell $Y^\varepsilon_{i_{in}}$. However, this is not a problem since the set I_{out} is finite in virtue of Assumption 7.1. This fact only worsens the constant C in this estimate which is therefore not optimal.

Lemma 9.24. Suppose $H\geq\gamma\varepsilon$. Then, for any $\boldsymbol{v}\in H^1(\Omega^\varepsilon)$,

$$\|\boldsymbol{v}\|_{L^2(\Omega^\varepsilon)} \leq C\varepsilon|\boldsymbol{v}|_{H^1(\Omega^\varepsilon)},$$

with $C>0$ independent of ε and of H.

Proof. This lemma is a simple corollary of Lemma 9.22 obtained by summing (9.10) over all the mesh elements. □

9.1.4 Other Lemmas

Lemma 9.25. Assume that the continuous velocity-pressure inf-sup property holds on Ω^ε with a constant $\beta > 0$, i.e.

$$\inf_{p\in L^2_0(\Omega^\varepsilon)} \sup_{\boldsymbol{v}\in H^1_0(\Omega^\varepsilon)^d} \frac{\displaystyle\int_{\Omega^\varepsilon} p \operatorname{div} \boldsymbol{v}}{\|p\|_{L^2(\Omega^\varepsilon)} |\boldsymbol{v}|_{H^1(\Omega^\varepsilon)}} \geq \beta,$$

Then, the discrete *inf-sup* property holds on $V_H \times M_H$ with the same constant $\beta > 0$

$$\inf_{p\in M_H} \sup_{\boldsymbol{v}\in V_H} \frac{\displaystyle\int_{\Omega^\varepsilon} p_H \operatorname{div} \boldsymbol{v_H}}{\|p_H\|_{L^2(\Omega^\varepsilon)} |\boldsymbol{v_H}|_{H,1}} \geq \beta.$$

In particular, for any $p_H \in M_H$, there exists $\boldsymbol{v_H} \in V_H$ such that

$$\operatorname{div} \boldsymbol{v_H} = p_H \text{ on } T \cap \Omega^\varepsilon, \ \forall T \in \mathcal{T}_H \text{ and } |\boldsymbol{v_H}|_{H,1} \leq \frac{1}{\beta}\|p_H\|_{L^2(\Omega^\varepsilon)}. \tag{9.11}$$

Proof. This theorem is a direct consequence of Theorem 7.14. □

Lemma 9.26. Let p^*_H be the L^2-orthogonal projection of $I^n_H p^*$ on M_H, i.e.

$$p^*_H = \Pi^n_H I^n_H p^* \in M_H,$$

where I^n_H is the operator defined in Lemma 9.8.

Then, for any $p_H \in M_H$, there exists $\boldsymbol{v_H} \in V_H$ and a constant C independent of the mesh, such that for any $T \in \mathcal{T}_H$,

$$\operatorname{div} \boldsymbol{v_H} = p_H - p^*_H \text{ on } T \cap \Omega^\varepsilon \text{ and } |\boldsymbol{v_H}|_{H,1} \leq \frac{C}{\varepsilon}\|p_H - p^*_H\|_{L^2(\Omega^\varepsilon)}.$$

Proof. We note that $p_H - p^*_H \in M_H$. Then, Lemma 9.26 is a direct consequence of Lemma 9.25 and Theorem 6.9, with the inf-sup constant $\beta = \dfrac{\varepsilon}{C}$. □

9.2 Proof of Theorem 9.27 (Error estimate)

9.2.1 Main Result

Theorem 9.27. Adopt definition of the perforated domain Ω^ε of Section 6.2, Assumption 7.1 about the mesh and Assumption 7.2 about the choice of the weighting functions. Suppose also that $\boldsymbol{f}$ is sufficiently smooth. The following error bound holds between the solution to the Stokes equations in perforated domains and its MsFEM approximation.

$$\begin{aligned}&\varepsilon^{-1}\left|\boldsymbol{u_\varepsilon}-\boldsymbol{u_H}\right|_{H,1}+\left\|p_\varepsilon-p_H\right\|_{L^2(\Omega^\varepsilon)}\\&\leq C\left(H^n\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^n(\Omega)}+H^n|p^*|_{H^{n+1}(\Omega)}+\left(\sqrt{\varepsilon}+\sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\right)\end{aligned}\tag{9.12}$$

where the constant C is independent of ε, H and $\boldsymbol{f}$.

Remark 9.28. In Theorem 9.27, p^* is the homogenized pressure. As soon as $\boldsymbol{f}$ and the domain Ω are sufficiently smooth (which is effectively the case for the domain Ω given Assumption 7.1), then as p^* is the solution to an elliptic problem, it follows that p^* is also smooth (see for example [86] for results about regularity of the solution to an elliptic problem).

Remark 9.29. At H fixed, the term $p_\varepsilon - p_H$ is $O(1)$ as ε goes to zero. Besides, recalling that in a periodic domain we have $\boldsymbol{u_\varepsilon} \approx \varepsilon^2 \boldsymbol{u}(x, \frac{x}{\varepsilon})$, then the above error estimate for the velocity and pressure essentially says that the relative errors are of the order $\left(H^n + \sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)$. In this sum, the term $\sqrt{\varepsilon}$, which comes from the boundary layers in the homogenization process, is negligible. Indeed, the dominant term is the classical "resonnance error" $\sqrt{\frac{\varepsilon}{H}}$, which says that the coarse mesh size H should be much larger than ε to have a good approximation. However, the presence of the term H^n allows to choose an H big enough without significantly reducing the accuracy of the approximation, which shows the interest of considering high-order methods. We can determine the optimal value of the coarse mesh size H so that the error is minimal, which is given by

$$H_{opt} = \left(\frac{\sqrt{\varepsilon}}{2n}\right)^{\frac{1}{n+\frac{1}{2}}}.$$

Furthermore, using $H = H_{opt}$ in the error estimate, we note that the error becomes of order $\sqrt{\varepsilon}$ as $n \to \infty$ (we recover the order of convergence of homogenization).

Corollary 9.30. Suppose that $\boldsymbol{f}$ is sufficiently smooth. The following error bound holds between the solution to the Stokes equations in periodic perforated domains and its MsFEM approximation.

$$\varepsilon^{-2}\|\boldsymbol{u_\varepsilon}-\boldsymbol{u_H}\|_{L^2(\Omega^\varepsilon)} \leq C\left(H^n\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^n(\Omega)}+H^n|p^*|_{H^{n+1}(\Omega)}+\left(\sqrt{\varepsilon}+\sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\right)$$

where the constant C is independent of ε, H and $\boldsymbol{f}$.

Corollary 9.30 is proved by applying the Poincaré inequality (Lemma 9.22) in each element $T \in \mathcal{T}_H$, which allows to improve the error bound for the velocity by a factor ε.

9.2.2 Preliminaries

9.2.2.1 Strang Lemma

The error estimation is essentially based on a Strang Lemma which is the classical lemma for error estimate in case of non-conforming finite element method.

Lemma 9.31. Let $\boldsymbol{u_\varepsilon}$ be the exact solution and $\boldsymbol{u_H}$ the approximated solution. Let $a_H(\cdot,\cdot)$ and $F_H(\cdot)$ be respectively the bi-linear and linear forms defined in (7.15). Then,

$$\|\boldsymbol{u_\varepsilon}-\boldsymbol{u_H}\|_{V_H} \leq \inf_{\boldsymbol{w_h}\in Z_H}\|\boldsymbol{u_\varepsilon}-\boldsymbol{w_h}\|_{V_H}+\sup_{\boldsymbol{w_h}\in Z_H\setminus\{0\}}\frac{|F_H(\boldsymbol{w_h})-a_H(\boldsymbol{u_\varepsilon},\boldsymbol{w_h})|}{\|\boldsymbol{w_h}\|_{V_H}}. \tag{9.13}$$

The first term in (9.13) is the usual best approximation error present in the classical Céa Lemma. The second term of (9.13) is the non-conformity error, that is roughly speaking, how far Z_H is from the divergence free subspace of $H_0^1(\Omega)$.

9.2.2.2 Interpolant properties

We note $\boldsymbol{\delta_\varepsilon}$ the difference between the velocity $\boldsymbol{u_\varepsilon}$ and its interpolant $\boldsymbol{w_h}$, defined in (7.17), i.e. $\boldsymbol{\delta_\varepsilon}=\boldsymbol{u_\varepsilon}-\boldsymbol{w_h}$. In what follows, we emphasize two important properties of $\boldsymbol{\delta_\varepsilon}$, summarized in Theorem 9.32.

Theorem 9.32. For any element $T \subset \mathcal{T}_H$ and for any polynomial $\boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} \in \big(\mathbb{P}_{n-1}(T)\big)^d$,

$$\int_{T\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} = 0. \tag{9.14}$$

For any element $T \subset \mathcal{T}_H$ and face $F \subset \partial T$, for any polynomial $\boldsymbol{\Pi}_{\boldsymbol{n}}^{\boldsymbol{F}} \in (\mathbb{P}_n(F))^d$,

$$\int_{F\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon} \cdot \boldsymbol{\Pi}_{\boldsymbol{n}}^{\boldsymbol{F}} = 0. \tag{9.15}$$

Proof. First, let T be an element of $\mathcal{T}_H$ and let $\boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}}$ be a polynomial in $\big(\mathbb{P}_{n-1}(T)\big)^d$. It follows

$$\begin{aligned}\int_{T\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} = \int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} - \Big[\sum_{F\in\mathcal{F}_H} \sum_{i=1}^{s} \Big(\int_{F\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\omega_{F,i}} \Big) \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} \Big) \\ + \sum_{K\in\mathcal{T}_H} \sum_{l=1}^{r} \Big(\int_{K\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{K,l}} \Big) \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Psi_{K,l}} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} \Big) \Big].\end{aligned}$$

Now since $\displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} = 0$ for all $F \in \mathcal{F}_H$ and $1 \le i \le s$, it comes

$$\int_{T\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} = \int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} - \sum_{K\in\mathcal{T}_H} \sum_{l=1}^{r} \Big(\int_{K\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{K,l}} \Big) \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Psi_{K,l}} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} \Big)$$

Now, we consider $\boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} = \boldsymbol{\varphi_{T,j}} \in (\mathbb{P}_{n-1}(T))^d$, the $(\boldsymbol{\varphi_{T,j}})_j$ being the weighting functions associated with elements in the high-order MsFEM, which leads to

$$\begin{aligned}\int_{T\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon} \cdot \boldsymbol{\varphi_{T,j}} &= \int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{T,j}} - \sum_{K\in\mathcal{T}_H} \sum_{l=1}^{r} \Big(\int_{K\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{K,l}} \Big) \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Psi_{K,l}} \cdot \boldsymbol{\varphi_{T,j}} \Big) \\ &= \int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{T,j}} - \sum_{K\in\mathcal{T}_H} \sum_{l=1}^{r} \Big(\int_{K\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{K,l}} \Big) \delta_{l,j}\delta_{T,K} \\ &= \int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{T,j}} - \int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon} \cdot \boldsymbol{\varphi_{T,j}} \\ &= 0.\end{aligned}$$

Since $\big(\boldsymbol{\varphi_{T,j}}\big)_j$ is a basis of $\big(\mathbb{P}_{n-1}(T)\big)^d$, we have, for all polynomials $\boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} \in \big(\mathbb{P}_{n-1}(T)\big)^d$,

$$\int_{T\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon} \cdot \boldsymbol{\Pi}_{\boldsymbol{n-1}}^{\boldsymbol{T}} = 0$$

which proves (9.14).

Then, let F be an element of $\mathcal{F}_H$ and let $\boldsymbol{\Pi_n^F}$ be a polynomial in $(\mathbb{P}_n(F))^d$. It follows

$$\int_{F\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon}\cdot\boldsymbol{\Pi_n^F} = \int_{F\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon}\cdot\boldsymbol{\Pi_n^F} - \Big[\sum_{E\in\mathcal{F}_H}\sum_{i=1}^{s}\Big(\int_{E\cap\Omega^\varepsilon}\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\omega_{E,i}}\Big)\Big(\int_{F\cap\Omega^\varepsilon}\boldsymbol{\Phi_{E,i}}\cdot\boldsymbol{\Pi_n^F}\Big) + \sum_{T\in\mathcal{T}_H}\sum_{l=1}^{r}\Big(\int_{T\cap\Omega^\varepsilon}\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\varphi_{T,l}}\Big)\Big(\int_{F\cap\Omega^\varepsilon}\boldsymbol{\Psi_{T,l}}\cdot\boldsymbol{\Pi_n^F}\Big)\Big].$$

Since $\displaystyle\int_F \boldsymbol{\Psi_{T,l}}\cdot\boldsymbol{\Pi_n^F} = 0$ for all $T\in\mathcal{T}_H$ and $1\le l\le r$, it comes

$$\int_{F\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon}\cdot\boldsymbol{\Pi_n^F} = \int_{F\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon}\cdot\boldsymbol{\Pi_n^F} - \sum_{E\in\mathcal{F}_H}\sum_{i=1}^{s}\Big(\int_{E\cap\Omega^\varepsilon}\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\omega_{E,i}}\Big)\Big(\int_{F\cap\Omega^\varepsilon}\boldsymbol{\Phi_{E,i}}\cdot\boldsymbol{\Pi_n^F}\Big).$$

We consider $\boldsymbol{\Pi_n^F} = \boldsymbol{\omega_{F,j}} \in (\mathbb{P}_n(F))^d$, the $(\boldsymbol{\omega_{F,j}})_j$ being the weighting functions associated with faces in the local problems. It follows

$$\begin{aligned}
\int_{F\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon}\cdot\boldsymbol{\omega_{F,j}} &= \int_{F\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon}\cdot\boldsymbol{\omega_{F,j}} - \sum_{E\in\mathcal{F}_H}\sum_{i=1}^{s}\Big(\int_{E\cap\Omega^\varepsilon}\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\omega_{E,i}}\Big)\Big(\int_{F\cap\Omega^\varepsilon}\boldsymbol{\Phi_{E,i}}\cdot\boldsymbol{\omega_{F,j}}\Big)\\
&= \int_{F\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon}\cdot\boldsymbol{\omega_{F,j}} - \sum_{E\in\mathcal{F}_H}\sum_{i=1}^{s}\Big(\int_{E\cap\Omega^\varepsilon}\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\omega_{E,i}}\Big)\delta_{ij}\delta_{EF}\\
&= \int_{F\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon}\cdot\boldsymbol{\omega_{F,j}} - \int_{F\cap\Omega^\varepsilon}\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\omega_{F,j}}\\
&= 0.
\end{aligned}$$

Since $\big(\boldsymbol{\omega_{F,i}}\big)_i$ is a basis of $\big(\mathbb{P}_n(F)\big)^d$, we have, for all $\boldsymbol{\Pi_n^F}\in(\mathbb{P}_n(F))^d$,

$$\int_{F\cap\Omega^\varepsilon}\boldsymbol{\delta}\cdot\boldsymbol{\Pi_n^F} = 0,$$

which proves (9.15). □

Remark 9.33. These properties can be re-written as follows. For any element $T\subset\mathcal{T}_H$ and face $F\subset\partial T$, we have, for all $\boldsymbol{\Pi_n^F}\in\big(\mathbb{P}_n(F)\big)^d$ and for all $\boldsymbol{\Pi_{n-1}^T}\in\big(\mathbb{P}_{n-1}(T)\big)^d$,

$$\int_{F\cap\Omega^\varepsilon}\boldsymbol{w_h}\cdot\boldsymbol{\Pi_n^F} = \int_{F\cap\Omega^\varepsilon}\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\Pi_n^F},$$

and

$$\int_{T\cap\Omega^\varepsilon}\boldsymbol{w_h}\cdot\boldsymbol{\Pi_{n-1}^T} = \int_{T\cap\Omega^\varepsilon}\boldsymbol{u_\varepsilon}\cdot\boldsymbol{\Pi_{n-1}^T}.$$

Lemma 9.34. The interpolant $\boldsymbol{w_h}$, defined in (7.17), belongs to the space Z_H, where the space Z_H is defined as

$$Z_H = \{\boldsymbol{v} \in V_H \text{ such that } \operatorname{div}_H \boldsymbol{v} = 0,\ \forall T \in \mathcal{T}_H\}.$$

Proof. By construction, we have clearly $\boldsymbol{w_h} \in V_H$. Besides, it is easy to see that $\boldsymbol{w_h} \in Z_H$. Indeed, we have, for all $P_n^T \in \mathbb{P}_n(T)$,

$$\int_{T\cap\Omega^\varepsilon} \operatorname{div} \boldsymbol{w_h}\ P_n^T = \int_{\partial T\cap\Omega^\varepsilon} P_n^T \boldsymbol{n}\cdot\boldsymbol{w_h} - \int_{T\cap\Omega^\varepsilon} \boldsymbol{w_h}\cdot\boldsymbol{\nabla} P_n^T,$$

by integration by parts. Since $P_n^T\boldsymbol{n}|_{\partial T} \in \big(\mathbb{P}_n(F)\big)^d$ and $\boldsymbol{\nabla} P_n^T \in \big(\mathbb{P}_{n-1}(T)\big)^d$, it follows

$$\begin{aligned}\int_{T\cap\Omega^\varepsilon} \operatorname{div} \boldsymbol{w_h}\ P_n^T &= \int_{\partial T\cap\Omega^\varepsilon} P_n^T \boldsymbol{n}\cdot\boldsymbol{u_\varepsilon} - \int_{T\cap\Omega^\varepsilon} \boldsymbol{u_\varepsilon}\cdot\boldsymbol{\nabla} P_n^T\\ &= \int_{T\cap\Omega^\varepsilon} \operatorname{div} \boldsymbol{u_\varepsilon}\ P_n^T\\ &= 0,\end{aligned}$$

by integration by parts in the second line and using the fact $\operatorname{div} \boldsymbol{u_\varepsilon} = 0$ in the last line. Now, noting that $\operatorname{div} \boldsymbol{w_h} \in \mathbb{P}_n(T)$, it comes that $\operatorname{div} \boldsymbol{w_h} = 0$ in $T\cap\Omega^\varepsilon$ for $T \in \mathcal{T}_H$, which concludes the proof. □

Using Lemma 9.34 allows to eliminate the pressure unknown from (7.14) by introducing the space Z_H. Therefore, (7.14) is equivalent to find $\boldsymbol{u_H} \in Z_H$ such that

$$\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla u_H}\cdot\boldsymbol{\nabla v} = \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{f}\cdot\boldsymbol{v} \quad \forall \boldsymbol{v} \in Z_H.$$

9.2.3 Error estimate for velocity in perforated domains

9.2.3.1 Approximation error in perforated domains

We recall that $\boldsymbol{\delta_\varepsilon}$ is defined as $\boldsymbol{\delta_\varepsilon} = \boldsymbol{u_\varepsilon} - \boldsymbol{w_h}$, where $\boldsymbol{u_\varepsilon}$ is the velocity and $\boldsymbol{w_h}$ is its interpolant. First, we have

$$\begin{aligned}c_H(\boldsymbol{\delta_\varepsilon},\boldsymbol{\delta_\varepsilon}) &= \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon}-\boldsymbol{w_h})\cdot\boldsymbol{\nabla}\boldsymbol{\delta_\varepsilon}\\ &= \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon}-\boldsymbol{w_h})\cdot\boldsymbol{\nabla}\boldsymbol{\delta_\varepsilon} + \int_{T\cap\Omega^\varepsilon} q_h \operatorname{div}\boldsymbol{\delta_\varepsilon}\\ &= \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla}\boldsymbol{\delta_\varepsilon} + \int_{T\cap\Omega^\varepsilon} (\Delta\boldsymbol{w_h} - \boldsymbol{\nabla} q_h)\cdot\boldsymbol{\delta_\varepsilon} - \int_{\partial T\cap\Omega^\varepsilon} (\boldsymbol{\nabla w_h} - q_h I)\boldsymbol{n}\cdot\boldsymbol{\delta_\varepsilon},\end{aligned}$$

where we have considered $\operatorname{div}\boldsymbol{\delta_\varepsilon} = 0$ in the element T in the second line to introduce the term q_h and made an integration by parts in the third line.

Now, since $(-\Delta \boldsymbol{w_h} + \boldsymbol{\nabla} q_h) \in \left(\mathbb{P}_{n-1}(T)\right)^d$, $\forall T \in \mathcal{T}_H$, and $(\boldsymbol{\nabla w_h} - q_h I)\boldsymbol{n} \in \left(\mathbb{P}_n(F)\right)^d$, $\forall F \in \mathcal{F}_H$, using Theorem 9.32, we observe that the quantity $c_H(\boldsymbol{\delta_\varepsilon}, \boldsymbol{\delta_\varepsilon})$ is reduced to

$$c_H(\boldsymbol{\delta_\varepsilon}, \boldsymbol{\delta_\varepsilon}) = \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} \boldsymbol{\nabla u_\varepsilon} \cdot \boldsymbol{\nabla \delta_\varepsilon}. \tag{9.16}$$

To handle expression (9.16) we introduce quantities which come from the homogeneisation of the Stokes problem. First, we introduce $\varepsilon^2 \boldsymbol{u_2}$, where $\boldsymbol{u_2}(x,y) = \boldsymbol{\omega_i}(y)\left(f_i(x) - \dfrac{\partial p^*}{\partial x_i}(x)\right)$ from (6.7) (we eliminate the viscosity ν in this chapter), the first term of the asymptotic expansion of the velocity $\boldsymbol{u_\varepsilon}$, which leads to

$$\begin{aligned} c_H(\boldsymbol{\delta_\varepsilon}, \boldsymbol{\delta_\varepsilon}) &= \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} \boldsymbol{\nabla u_\varepsilon} \cdot \boldsymbol{\nabla \delta_\varepsilon} \\ &= \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}) \cdot \boldsymbol{\nabla \delta_\varepsilon} + \int_{T \cap \Omega^\varepsilon} \varepsilon^2 \boldsymbol{\nabla u_2} \cdot \boldsymbol{\nabla \delta_\varepsilon} \\ &= \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}) \cdot \boldsymbol{\nabla \delta_\varepsilon} - \int_{T \cap \Omega^\varepsilon} \varepsilon^2 \Delta \boldsymbol{u_2} \cdot \boldsymbol{\delta_\varepsilon} + \int_{\partial T \cap \Omega^\varepsilon} \varepsilon^2 \boldsymbol{\nabla u_2}\, \boldsymbol{n} \cdot \boldsymbol{\delta_\varepsilon}, \end{aligned} \tag{9.17}$$

by integration by parts in the third line.

Now, to handle the second term of the third line of (9.17), we make appear the residual $\boldsymbol{F_\varepsilon}$ due to homogenization introduced in Lemma 6.28. To do this, we introduce the source term $\boldsymbol{f}$ and the quantity $p_{\varepsilon,1}$, defined as $p_{\varepsilon,1} = p^* + \varepsilon p_1$ (see (6.34)), the expansion of the pressure p_ε at the order 1 in ε, with $p_1(x,y) = \pi_i(y)\left(f_i(x) - \dfrac{\partial p^*}{\partial x_i}(x)\right)$ from (6.7). We remark first, since $\operatorname{div} \boldsymbol{\delta_\varepsilon} = 0$ in the element T, that

$$\int_{T \cap \Omega^\varepsilon} \boldsymbol{\nabla} p_{\varepsilon,1} \cdot \boldsymbol{\delta_\varepsilon} = \int_{T \cap \Omega^\varepsilon} \boldsymbol{\nabla} p^* \cdot \boldsymbol{\delta_\varepsilon} + \int_{\partial T \cap \Omega^\varepsilon} \varepsilon p_1 \boldsymbol{n} \cdot \boldsymbol{\delta_\varepsilon}.$$

Introducing these terms leads to

$$\begin{aligned} c_H(\boldsymbol{\delta_\varepsilon}, \boldsymbol{\delta_\varepsilon}) = \sum_{T \in \mathcal{T}_H} \Bigg(& \int_{T \cap \Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}) \cdot \boldsymbol{\nabla \delta_\varepsilon} - \int_{T \cap \Omega^\varepsilon} (\boldsymbol{f} + \varepsilon^2 \Delta \boldsymbol{u_2} - \boldsymbol{\nabla} p_{\varepsilon,1}) \cdot \boldsymbol{\delta_\varepsilon} \\ & + \int_{T \cap \Omega^\varepsilon} (\boldsymbol{f} - \boldsymbol{\nabla} p^*) \cdot \boldsymbol{\delta_\varepsilon} + \int_{\partial T \cap \Omega^\varepsilon} (\varepsilon^2 \boldsymbol{\nabla u_2} - \varepsilon p_1 I)\boldsymbol{n} \cdot \boldsymbol{\delta_\varepsilon} \Bigg) \\ = \sum_{T \in \mathcal{T}_H} \Bigg(& \int_{T \cap \Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}) \cdot \boldsymbol{\nabla \delta_\varepsilon} - \int_{T \cap \Omega^\varepsilon} \boldsymbol{F_\varepsilon} \cdot \boldsymbol{\delta_\varepsilon} \\ & + \int_{T \cap \Omega^\varepsilon} (\boldsymbol{f} - \boldsymbol{\nabla} p^*) \cdot \boldsymbol{\delta_\varepsilon} + \int_{\partial T \cap \Omega^\varepsilon} (\varepsilon^2 \boldsymbol{\nabla u_2} - \varepsilon p_1 I)\boldsymbol{n} \cdot \boldsymbol{\delta_\varepsilon} \Bigg). \end{aligned} \tag{9.18}$$

Now, we bound separately the four terms, appearing in the right-hand side of (9.18), using estimates derived from homogenization of the Stokes equations (see Chapter 6).

First term of approximation error To bound the first term on the right-hand side of (9.18) we apply the Cauchy–Schwarz inequality, and then use homogenization bounds introduced in

Theorem 6.9, which leads to

$$\left| \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \varepsilon^2 \boldsymbol{u_2}) \cdot \boldsymbol{\nabla}\boldsymbol{\delta_\varepsilon} \right| \leq C\varepsilon\sqrt{\varepsilon} \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)} \, |\boldsymbol{\delta_\varepsilon}|_{H,1} \,. \tag{9.19}$$

Second term of approximation error For the second term on the right-hand side of (9.18), we use the convergence result for the residual $\boldsymbol{F_\varepsilon}$ due to the homogenization established in (6.44) and the Poincaré inequality in perforated domains (Lemma 9.24), which leads to

$$\left| \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{F_\varepsilon} \cdot \boldsymbol{\delta_\varepsilon} \right| \leq C\varepsilon \, |\boldsymbol{\delta_\varepsilon}|_{H,1} \, \sqrt{\varepsilon} \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}. \tag{9.20}$$

Remark 9.35. One may think that we can remove from the term $\boldsymbol{F_\varepsilon}$ its projection on the polynomial space $(\mathbb{P}_{n-1}(T))^d$, i.e. considering the term $\Pi_H^{n-1} \boldsymbol{F_\varepsilon}$ which vanished against the term $\boldsymbol{\delta_\varepsilon}$. However this method will not improve the error estimate since terms of order ε^{-n} will appear due to the derivation of the oscillating term $\boldsymbol{F_\varepsilon}$.

Third term of approximation error We note that the third term on the right-hand side of (9.18) involves the non-oscillating term $(\boldsymbol{f} - \boldsymbol{\nabla} p^*)$. We can consider thus its projection in the polynomial space $(\mathbb{P}_{n-1}(T))^d$, i.e. the term $\Pi_H^{n-1}(\boldsymbol{f} - \boldsymbol{\nabla} p^*)$, by taking advantage that $\boldsymbol{\delta_\varepsilon}$ is orthogonal to any polynomial in $(\mathbb{P}_{n-1}(T))^d$. We have consequently,

$$\begin{aligned}
\left| \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} (\boldsymbol{f} - \boldsymbol{\nabla} p^*) \cdot \boldsymbol{\delta_\varepsilon} \right| &= \left| \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \Big((\boldsymbol{f} - \boldsymbol{\nabla} p^*) - \Pi_H^{n-1}(\boldsymbol{f} - \boldsymbol{\nabla} p^*) \Big) \cdot \boldsymbol{\delta_\varepsilon} \right| \\
&\leq \sum_{T\in\mathcal{T}_H} \left(\int_{T\cap\Omega^\varepsilon} \Big((\boldsymbol{f} - \boldsymbol{\nabla} p^*) - \Pi_H^{n-1}(\boldsymbol{f} - \boldsymbol{\nabla} p^*) \Big)^2 \right)^{\frac{1}{2}} \left(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\delta_\varepsilon}^2 \right)^{\frac{1}{2}} \\
&\leq \sum_{T\in\mathcal{T}_H} \|(\boldsymbol{f} - \boldsymbol{\nabla} p^*) - \Pi_H^{n-1}(\boldsymbol{f} - \boldsymbol{\nabla} p^*)\|_{L^2(T)} \|\boldsymbol{\delta_\varepsilon}\|_{L^2(T\cap\Omega^\varepsilon)},
\end{aligned}$$

using the Cauchy–Schwarz inequality in the second line. Now using Lemma 9.7 and the Poincaré inequality in perforated domains (Lemma 9.22), it follows

$$\begin{aligned}
\left| \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} (\boldsymbol{f} - \boldsymbol{\nabla} p^*) \cdot \boldsymbol{\delta_\varepsilon} \right| &\leq \sum_{T\in\mathcal{T}_H} H^n \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(T)} C\varepsilon \|\boldsymbol{\nabla}\boldsymbol{\delta_\varepsilon}\|_{L^2(T\cap\Omega^\varepsilon)} \\
&\leq C\varepsilon H^n \left(\sum_{T\in\mathcal{T}_H} \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|^2_{H^n(T)} \right)^{1/2} \left(\sum_{T\in\mathcal{T}_H} \|\boldsymbol{\nabla}\boldsymbol{\delta_\varepsilon}\|^2_{L^2(T\cap\Omega^\varepsilon)} \right)^{1/2} \\
&\leq C\varepsilon H^n \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)} \, |\boldsymbol{\delta_\varepsilon}|_{H,1} \,,
\end{aligned} \tag{9.21}$$

where we have used the discrete Cauchy–Schwarz inequality in the second term and then summed over cells.

Fourth term of approximation error To handle the fourth term on the right-hand side of (9.18), we use the expressions of $\boldsymbol{u_2}$ and p_1. Consequently, $\forall T \in \mathcal{T}_H$, we have

$$
\begin{aligned}
&\left|\int_{\partial T\cap\Omega^\varepsilon}\left(\varepsilon^2\boldsymbol{\nabla u_2}-\varepsilon p_1 I\right)\boldsymbol{n}\cdot\boldsymbol{\delta_\varepsilon}\right| \\
&=\left|\int_{\partial T\cap\Omega^\varepsilon}\left(\varepsilon^2\boldsymbol{\nabla}\left(\boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\right)(f_i-\partial_i p^*)+\varepsilon^2\boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}(f_i-\partial_i p^*)-\varepsilon\pi_i(f_i-\partial_i p^*)I\right)\boldsymbol{n}\cdot\boldsymbol{\delta_\varepsilon}\right| \\
&=\left|\int_{\partial T\cap\Omega^\varepsilon}\left(\varepsilon^2\frac{1}{\varepsilon}\left(\boldsymbol{\nabla_y\omega_i}\right)\left(\frac{x}{\varepsilon}\right)(f_i-\partial_i p^*)+\varepsilon^2\boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}(f_i-\partial_i p^*)-\varepsilon\pi_i(f_i-\partial_i p^*)I\right)\boldsymbol{n}\cdot\boldsymbol{\delta_\varepsilon}\right| \\
&\leq\int_{\partial T\cap\Omega^\varepsilon}\left(\left|\varepsilon\left(\boldsymbol{\nabla_y\omega_i}\right)\left(\frac{x}{\varepsilon}\right)(f_i-\partial_i p^*)\right|+\left|\varepsilon^2\boldsymbol{\omega_i}\left(\frac{x}{\varepsilon}\right)\boldsymbol{\nabla}(f_i-\partial_i p^*)\right|+\left|\varepsilon\pi_i(f_i-\partial_i p^*)\right|\right)|\boldsymbol{\delta_\varepsilon}|.
\end{aligned}
$$

where for the sake of clarity, the partial derivative with respect to x_i is simply denoted by ∂_i. Now, we note that the terms $\boldsymbol{\nabla_y\omega_i}$, $\boldsymbol{\omega_i}$ and π_i are bounded, which allows to write, $\forall T \in \mathcal{T}_H$,

$$
\begin{aligned}
&\left|\int_{\partial T\cap\Omega^\varepsilon}\left(\boldsymbol{\nabla}\varepsilon^2\boldsymbol{u_2}-\varepsilon p_1 I\right)\boldsymbol{n}\cdot\boldsymbol{\delta_\varepsilon}\right|\leq C\int_{\partial T\cap\Omega^\varepsilon}\left(|\varepsilon(f_i-\partial_i p^*)|+|\varepsilon^2\boldsymbol{\nabla}(f_i-\partial_i p^*)|\right)|\boldsymbol{\delta_\varepsilon}| \\
&\leq C\left(\left(\int_{\partial T\cap\Omega^\varepsilon}|\varepsilon(f_i-\partial_i p^*)|^2\right)^{\frac{1}{2}}+\left(\int_{\partial T\cap\Omega^\varepsilon}|\varepsilon^2\boldsymbol{\nabla}(f_i-\partial_i p^*)|^2\right)^{\frac{1}{2}}\right)\left(\int_{\partial T\cap\Omega^\varepsilon}|\boldsymbol{\delta_\varepsilon}|^2\right)^{\frac{1}{2}} \\
&\leq C\|\boldsymbol{\delta_\varepsilon}\|_{L^2(\partial T\cap\Omega^\varepsilon)}\left(\varepsilon\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{L^2(\partial T)}+\varepsilon^2\|\boldsymbol{\nabla}(\boldsymbol{f}-\boldsymbol{\nabla}p^*)\|_{L^2(\partial T)}\right),
\end{aligned}
$$

using the Cauchy–Schwarz inequality in the second line. Now using the Trace inequality in perforated domains (Lemma 9.21) and the classical Trace inequality (Lemma 9.6), it follows, $\forall T \in \mathcal{T}_H$,

$$
\begin{aligned}
&\left|\int_{\partial T\cap\Omega^\varepsilon}\left(\varepsilon^2\boldsymbol{\nabla u_2}-\varepsilon p_1 I\right)\boldsymbol{n}\cdot\boldsymbol{\delta_\varepsilon}\right|\leq C\sqrt{\varepsilon}|\boldsymbol{\delta_\varepsilon}|_{H^1(T\cap\Omega^\varepsilon)}\varepsilon\left(\frac{1}{H}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|^2_{L^2(T)}\right. \\
&\left.+H\|\boldsymbol{\nabla}(\boldsymbol{f}-\boldsymbol{\nabla}p^*)\|^2_{L^2(T)}+\frac{\varepsilon}{H}\|\boldsymbol{\nabla}(\boldsymbol{f}-\boldsymbol{\nabla}p^*)\|^2_{L^2(T)}+\varepsilon H\|\boldsymbol{\nabla}\boldsymbol{\nabla}(\boldsymbol{f}-\boldsymbol{\nabla}p^*)\|^2_{L^2(T)}\right)^{\frac{1}{2}} \\
&\leq C\varepsilon\sqrt{\frac{\varepsilon}{H}}|\boldsymbol{\delta_\varepsilon}|_{H^1(T\cap\Omega^\varepsilon)}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(T)},
\end{aligned}
$$

where we have considered $H \leq H^{-1}$ to reach the last inequality. Now, by summing over elements, we have

$$
\begin{aligned}
&\left|\sum_{T\in\mathcal{T}_H}\int_{\partial T\cap\Omega^\varepsilon}\left(\varepsilon^2\boldsymbol{\nabla u_2}-\varepsilon p_1 I)\right)\boldsymbol{n}\cdot\boldsymbol{\delta_\varepsilon}\right|\leq C\varepsilon\sqrt{\frac{\varepsilon}{H}}\sum_{T\in\mathcal{T}_H}|\boldsymbol{\delta_\varepsilon}|_{H^1(T\cap\Omega^\varepsilon)}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(T)} \\
&\leq C\varepsilon\sqrt{\frac{\varepsilon}{H}}\left(\sum_{T\in\mathcal{T}_H}|\boldsymbol{\delta_\varepsilon}|^2_{H^1(T\cap\Omega^\varepsilon)}\right)^{\frac{1}{2}}\left(\sum_{T\in\mathcal{T}_H}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|^2_{H^2(T)}\right)^{\frac{1}{2}} \\
&\leq C\varepsilon\sqrt{\frac{\varepsilon}{H}}\,|\boldsymbol{\delta_\varepsilon}|_{H,1}\,\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)},
\end{aligned}
\tag{9.22}
$$

using the discrete Cauchy–Schwarz inequality.

Conclusion of approximation error Finally, by regrouping estimates (9.19) to (9.22), we conclude that

$$\begin{aligned}|\boldsymbol{\delta_\varepsilon}|_{H,1} &\leq C\varepsilon H^n \|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)} + C\varepsilon\left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})} \\ &\leq C\varepsilon\left[\left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})} + H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)}\right].\end{aligned} \tag{9.23}$$

9.2.3.2 Non conformity error in perforated domains

We turn now to the non conformity error defined in the Strang Lemma (Lemma 9.31). We consider first the following quantity,

$$a_H(\boldsymbol{u_\varepsilon}, \boldsymbol{w_h}) - F_H(\boldsymbol{w_h}) = \sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla w_h} - \sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{f}\cdot\boldsymbol{w_h}.$$

Now, by noting that $\operatorname{div}\boldsymbol{w_h} = 0$ in any element T, we apply the same procedure as for deriving the approximation error, which leads to

$$\begin{aligned}a_H(\boldsymbol{u_\varepsilon}, \boldsymbol{w_h}) - F_H(\boldsymbol{w_h}) = \sum_{T\in\mathcal{T}_H}\Bigg(&\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \varepsilon^2\boldsymbol{u_2})\cdot\boldsymbol{\nabla w_h} - \int_{T\cap\Omega^\varepsilon}\boldsymbol{F_\varepsilon}\cdot\boldsymbol{w_h} \\ &+ \int_{\partial T\cap\Omega^\varepsilon}(\boldsymbol{\nabla}\varepsilon^2\boldsymbol{u_2} + \varepsilon p_1 I)\boldsymbol{n}\cdot\boldsymbol{w_h} - \int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla} p^*\cdot\boldsymbol{w_h}\Bigg).\end{aligned} \tag{9.24}$$

We bound the first three terms appearing on the right-hand side of (9.24) using estimates already shown for the approximation error, i.e. (9.19), (9.20) and (9.22).

To bound the last term on the right-hand side of (9.24), we use the operator I_H^n as constructed in Lemma 9.8 and observe that, $\forall T\in\mathcal{T}_H$,

$$\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla} I_H^n p^*\cdot\boldsymbol{w_h} = -\int_{T\cap\Omega^\varepsilon} I_H^n p^*\operatorname{div}\boldsymbol{w_h} + \int_{\partial T\cap\Omega^\varepsilon} I_H^n p^*\boldsymbol{n}\cdot[[\boldsymbol{w_h}]] = 0.$$

It follows consequently,

$$\begin{aligned}\left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla} p^*\cdot\boldsymbol{w_h}\right| &= \left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}(\boldsymbol{\nabla} p^* - \boldsymbol{\nabla} I_H^n p^*)\cdot\boldsymbol{w_h}\right| \\ &\leq \sum_{T\in\mathcal{T}_H}\left(\int_{T\cap\Omega^\varepsilon}\left(\boldsymbol{\nabla}(p^* - I_H^n p^*)\right)^2\right)^{\frac{1}{2}}\left(\int_{T\cap\Omega^\varepsilon}\boldsymbol{w_h}^2\right)^{\frac{1}{2}} \\ &\leq \sum_{T\in\mathcal{T}_H}|p^* - I_H^n p^*|_{H^1(T)}\|\boldsymbol{w_h}\|_{L^2(T\cap\Omega^\varepsilon)},\end{aligned}$$

with the Cauchy–Schwarz inequality. Now, applying the discrete Cauchy–Schwarz inequality

and summing over elements, it follows

$$\left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla}p^*\cdot\boldsymbol{w_h}\right| \leq \left(\sum_{T\in\mathcal{T}_H}|p^*-I_H^n p^*|^2_{H^1(T)}\right)^{\frac{1}{2}}\left(\sum_{T\in\mathcal{T}_H}\|\boldsymbol{w_h}\|^2_{L^2(T\cap\Omega^\varepsilon)}\right)^{\frac{1}{2}}$$
$$\leq C|p^*-I_H^n p^*|_{H^1(\Omega)}\|\boldsymbol{w_h}\|_{L^2(\Omega^\varepsilon)}.$$

Now, using Lemma 9.8 and the Poincaré inequality in perforated domains (Lemma 9.24), it follows

$$\left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla}p^*\cdot\boldsymbol{w_h}\right| \leq CH^n|p^*|_{H^{n+1}(\Omega)}\varepsilon\,|\boldsymbol{w_h}|_{H,1}\,.$$

Gathering the estimates for the four terms appearing on the right-hand side of (9.24), we conclude that

$$\begin{aligned}&\sup_{\boldsymbol{w_h}\in Z_H}\frac{|F_H(\boldsymbol{w_h})-a_H(\boldsymbol{u_\varepsilon},\boldsymbol{w_h})|}{\|\boldsymbol{w_h}\|_{V_H}}\\ &\leq C\varepsilon\left[\left(\sqrt{\varepsilon}+\sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}+H^n|p^*|_{H^{n+1}(\Omega)}\right].\end{aligned}\tag{9.25}$$

9.2.3.3 Conclusion of error estimate for velocity

Finally, combining the approximation error (9.23) and the non-conformity error (9.25), we conclude that

$$\begin{aligned}&|\boldsymbol{u_\varepsilon}-\boldsymbol{u_H}|_{H,1}\leq\inf_{\boldsymbol{w_h}\in Z_H}|\boldsymbol{u_\varepsilon}-\boldsymbol{w_h}|_{H,1}+\sup_{\boldsymbol{w_h}\in Z_H}\frac{|F_H(\boldsymbol{w_h})-a_H(\boldsymbol{u_\varepsilon},\boldsymbol{w_h})|}{|\boldsymbol{w_h}|_{H,1}}\\ &\leq C\varepsilon\left[\left(\sqrt{\varepsilon}+\sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}+H^n\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^n(\Omega)}+H^n|p^*|_{H^{n+1}(\Omega)}\right].\end{aligned}\tag{9.26}$$

9.2.4 Error estimate for pressure in perforated domains

We now turn to the error estimation for pressure. In order to obtain an error estimate for pressure, we consider the pressure p_H^* defined as

$$p_H^*=\Pi_H^n I_H^n p^*\in M_H$$

i.e. the L^2-orthogonal projection of $I_H^n p^*$ on M_H. Using the triangle inequality, it follows

$$\begin{aligned}\|p_\varepsilon-p_H\|_{L^2(\Omega^\varepsilon)}&=\|p^*-p_\varepsilon+p_H-p^*\|_{L^2(\Omega^\varepsilon)}\\ &\leq\|p^*-p_\varepsilon\|_{L^2(\Omega^\varepsilon)}+\|p_H-p^*\|_{L^2(\Omega^\varepsilon)}.\end{aligned}\tag{9.27}$$

First, we bound the first term and then the second term appearing in the right-hand side of (9.27).

First term of (9.27) To bound the first term on the right-hand side of (9.27), we use the interpolation estimates given in Lemmas 9.7 and 9.8 and homogenization bounds given in Theorem 6.9, leading to

$$\begin{aligned}\|p_H^* - p_\varepsilon\|_{L^2(\Omega^\varepsilon)} &= \|\Pi_H^n I_H^n p^* - p_\varepsilon\|_{L^2(\Omega^\varepsilon)} \\ &= \|\Pi_H^n I_H^n p^* - \Pi_H^n p^* + \Pi_H^n p^* - p^* + p^* - p_\varepsilon\|_{L^2(\Omega^\varepsilon)} \\ &\leq \|\Pi_H^n \left(I_H^n p^* - p^*\right)\|_{L^2(\Omega)} + \|\Pi_H^n p^* - p^*\|_{L^2(\Omega)} + \|p^* - p_\varepsilon\|_{L^2(\Omega^\varepsilon)} \\ &\leq \|I_H^n p^* - p^*\|_{L^2(\Omega)} + \|\Pi_H^n p^* - p^*\|_{L^2(\Omega)} + \|p^* - p_\varepsilon\|_{L^2(\Omega^\varepsilon)} \\ &\leq CH^{n+1}|p^*|_{H^{n+1}(\Omega)} + CH^{n+1}|p^*|_{H^{n+1}(\Omega)} + C\sqrt{\varepsilon}\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})},\end{aligned}$$

which allows to conclude that

$$\|p_H^* - p_\varepsilon\|_{L^2(\Omega^\varepsilon)} \leq C\left(H^{n+1}|p^*|_{H^{n+1}(\Omega)} + \sqrt{\varepsilon}\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\right). \tag{9.28}$$

Second term of (9.27) To bound the second term of the pressure estimation (9.27), we use the velocity $\boldsymbol{v_H}$ as defined in Lemma 9.26 which leads to

$$\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \int_{\Omega^\varepsilon} (p_H - p_H^*)(p_H - p_H^*) = \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} (p_H - p_H^*)\ \operatorname{div} \boldsymbol{v_H}. \tag{9.29}$$

To bound (9.29), we consider first the weak form of the problem and we introduce successively terms from homogenization to make appear already known quantities.

First, considering the weak form of the problem leads to

$$\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla u_H} \cdot \boldsymbol{\nabla v_H} - \int_{T\cap\Omega^\varepsilon} \boldsymbol{f}\cdot\boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} p_H^*\ \operatorname{div}\boldsymbol{v_H}.$$

Then, we introduce the velocity $\boldsymbol{u_\varepsilon}$, which leads to

$$\begin{aligned}\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \sum_{T\in\mathcal{T}_H} \Bigg(&\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_H} - \boldsymbol{u_\varepsilon})\cdot\boldsymbol{\nabla v_H} + \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla v_H} \\ &- \int_{T\cap\Omega^\varepsilon} \boldsymbol{f}\cdot\boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} p_H^*\ \operatorname{div}\boldsymbol{v_H}\Bigg).\end{aligned}$$

Next, we introduce the term $\varepsilon^2\boldsymbol{u_2}$, the first term of the asymptotic expansion of the velocity $\boldsymbol{u_\varepsilon}$, which leads to

$$\begin{aligned}\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \sum_{T\in\mathcal{T}_H} \Bigg(&\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_H} - \boldsymbol{u_\varepsilon})\cdot\boldsymbol{\nabla v_H} + \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \boldsymbol{\varepsilon^2 u_2})\cdot\boldsymbol{\nabla v_H} \\ &+ \int_{T\cap\Omega^\varepsilon} \varepsilon^2\boldsymbol{\nabla u_2}\cdot\boldsymbol{\nabla v_H} - \int_{T\cap\Omega^\varepsilon} \boldsymbol{f}\cdot\boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} p_H^*\ \operatorname{div}\boldsymbol{v_H}\Bigg).\end{aligned}$$

By integration by parts, it follows

$$\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \sum_{T\in\mathcal{T}_H} \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_H} - \boldsymbol{u_\varepsilon}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} + \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \boldsymbol{\varepsilon^2 u_2}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} \\ - \int_{T\cap\Omega^\varepsilon} (\boldsymbol{f} + \varepsilon^2 \Delta \boldsymbol{u_2}) \cdot \boldsymbol{v_H} + \int_{\partial T\cap\Omega^\varepsilon} \varepsilon^2 \boldsymbol{\nabla} \boldsymbol{u_2}\, \boldsymbol{n} \cdot \boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} p_H^* \ \operatorname{div} \boldsymbol{v_H} \Big).$$

To make appear the residual due to the homogenization, we introduce the quantity $p_{\varepsilon,1}$, the expansion of the pressure p_ε at the order 1 in ε as defined in (6.34).

$$\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \sum_{T\in\mathcal{T}_H} \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_H} - \boldsymbol{u_\varepsilon}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} + \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \boldsymbol{\varepsilon^2 u_2}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} \\ - \int_{T\cap\Omega^\varepsilon} (\boldsymbol{f} + \varepsilon^2 \Delta \boldsymbol{u_2} - \boldsymbol{\nabla} p_{\varepsilon,1}) \cdot \boldsymbol{v_H} + \int_{\partial T\cap\Omega^\varepsilon} \varepsilon^2 \boldsymbol{\nabla} \boldsymbol{u_2}\, \boldsymbol{n} \cdot \boldsymbol{v_H} \\ - \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla} p_{\varepsilon,1} \cdot \boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} p_H^* \ \operatorname{div} \boldsymbol{v_H} \Big).$$

Finally we introduce the term p^*, which leads to

$$\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \sum_{T\in\mathcal{T}_H} \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_H} - \boldsymbol{u_\varepsilon}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} + \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \boldsymbol{\varepsilon^2 u_2}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} \\ - \int_{T\cap\Omega^\varepsilon} \boldsymbol{F_\varepsilon} \cdot \boldsymbol{v_H} + \int_{\partial T\cap\Omega^\varepsilon} \varepsilon^2 \boldsymbol{\nabla} \boldsymbol{u_2}\, \boldsymbol{n} \cdot \boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(p_{\varepsilon,1} - p^*) \cdot \boldsymbol{v_H} \\ - \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla} p^* \cdot \boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} p_H^* \ \operatorname{div} \boldsymbol{v_H} \Big).$$

By integration by parts, it follows

$$\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \sum_{T\in\mathcal{T}_H} \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_H} - \boldsymbol{u_\varepsilon}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} + \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \boldsymbol{\varepsilon^2 u_2}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} \\ - \int_{T\cap\Omega^\varepsilon} \boldsymbol{F_\varepsilon} \cdot \boldsymbol{v_H} + \int_{\partial T\cap\Omega^\varepsilon} (\varepsilon^2 \boldsymbol{\nabla} \boldsymbol{u_2} - (p_{\varepsilon,1} - p^*) I)\, \boldsymbol{n} \cdot \boldsymbol{v_H} \\ + \int_{T\cap\Omega^\varepsilon} (p_{\varepsilon,1} - p^*) \operatorname{div} \boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla} p^* \cdot \boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} p_H^* \ \operatorname{div} \boldsymbol{v_H} \Big).$$

Now, we introduce the term $I_H^n p^*$, where I_H^n is the operator defined in Lemma 9.8, which leads, at the end, to

$$\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} = \sum_{T\in\mathcal{T}_H} \Big(\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_H} - \boldsymbol{u_\varepsilon}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} + \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(\boldsymbol{u_\varepsilon} - \boldsymbol{\varepsilon^2 u_2}) \cdot \boldsymbol{\nabla} \boldsymbol{v_H} \\ - \int_{T\cap\Omega^\varepsilon} \boldsymbol{F_\varepsilon} \cdot \boldsymbol{v_H} + \int_{\partial T\cap\Omega^\varepsilon} (\varepsilon^2 \boldsymbol{\nabla} \boldsymbol{u_2} - (p_{\varepsilon,1} - p^*) I)\, \boldsymbol{n} \cdot \boldsymbol{v_H} \\ + \int_{T\cap\Omega^\varepsilon} (p_{\varepsilon,1} - p^*) \operatorname{div} \boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla}(p^* - I_H^n p^*) \cdot \boldsymbol{v_H} \\ - \int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla} I_H^n p^* \cdot \boldsymbol{v_H} - \int_{T\cap\Omega^\varepsilon} p_H^* \ \operatorname{div} \boldsymbol{v_H} \Big). \tag{9.30}$$

Now, we remark that the last two terms on the right-hand side of (9.30) cancel each other.

Indeed, by integration by parts and using the definition of p_H^*, we have

$$\begin{aligned}
&-\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla}(I_H^n p^*)\cdot\boldsymbol{v_H}-\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}p_H^*\ \operatorname{div}\boldsymbol{v_H}\\
&=\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}I_H^n p^*\ \operatorname{div}\boldsymbol{v_H}-\sum_{F\in\mathcal{F}_H}\int_{F\cap\Omega^\varepsilon}I_H^n p^*\ [[\boldsymbol{n}\cdot\boldsymbol{v_H}]]-\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\Pi_H^n I_H^n p^*\ \operatorname{div}\boldsymbol{v_H}\\
&=-\sum_{F\in\mathcal{F}_H}\int_{F\cap\Omega^\varepsilon}I_H^n p^*\ [[\boldsymbol{n}\cdot\boldsymbol{v_H}]]+\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}(I-\Pi_H^n)(I_H^n p^*)\ \operatorname{div}\boldsymbol{v_H}.
\end{aligned}$$

Now, given $[[\boldsymbol{n}\cdot\boldsymbol{v_H}]]$ is orthogonal to any polynomials of degree $\leq n$ on the faces, we have

$$\sum_{F\in\mathcal{F}_H}\int_{F\cap\Omega^\varepsilon}I_H^n p^*\ [[\boldsymbol{n}\cdot\boldsymbol{v_H}]]=0,$$

and given Π_H^n is defined as the L^2-orthogonal projection on M_H, $(I-\Pi_H^n)$ is orthogonal to any polynomial in M_H and by recalling that $\operatorname{div}\boldsymbol{v_H}\in M_H$, we have

$$\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}(I-\Pi_H^n)(I_H^n p^*)\ \operatorname{div}\boldsymbol{v_H}=0,$$

which allows to conclude that

$$\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla}(I_H^n p^*)\cdot\boldsymbol{v_H}-\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}p_H^*\ \operatorname{div}\boldsymbol{v_H}=0.$$

Then, we bound the other terms appearing in the right-hand side of (9.30), using estimates already shown in the estimation of velocity error.

For the first term on the right-hand side of (9.30), we use the Cauchy–Schwarz inequality, then the discrete Cauchy–Schwarz inequality and sum over the elements. We conclude by using the error bound for velocity (see (9.26)), which leads to

$$\left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla}(\boldsymbol{u_H}-\boldsymbol{u_\varepsilon})\cdot\boldsymbol{\nabla}\boldsymbol{v_H}\right|\leq C\,|\boldsymbol{u_H}-\boldsymbol{u_\varepsilon}|_{H,1}\,|\boldsymbol{v_H}|_{H,1}\,. \tag{9.31}$$

For the second term on the right-hand side of (9.30), we use first the Cauchy–Schwarz inequality, then the discrete Cauchy–Schwarz inequality and sum over the elements. We conclude by using the homogenization bound for velocity (see Theorem 6.9), which leads to

$$\begin{aligned}
\left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla}(\boldsymbol{u_\varepsilon}-\boldsymbol{\varepsilon^2 u_2})\cdot\boldsymbol{\nabla}\boldsymbol{v_H}\right|&\leq C\,\left|\boldsymbol{u_\varepsilon}-\varepsilon^2\boldsymbol{u_2}\right|_{H,1}|\boldsymbol{v_H}|_{H,1}\\
&\leq C\varepsilon\sqrt{\varepsilon}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega^\varepsilon)\cap C^{1,\alpha}(\overline{\Omega^\varepsilon})}\ |\boldsymbol{v_H}|_{H,1}\,.
\end{aligned} \tag{9.32}$$

For the third term on the right-hand side of (9.30), similar arguments as for (9.20) leads to

$$\left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{F_\varepsilon}\cdot\boldsymbol{v_H}\right| \leq C\varepsilon\sqrt{\varepsilon}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\,|\boldsymbol{v_H}|_{H,1}\,. \tag{9.33}$$

For the fourth term on the right-hand side of (9.30), we have as for (9.22),

$$\left|\sum_{T\in\mathcal{T}_H}\int_{\partial T\cap\Omega^\varepsilon}(\boldsymbol{\nabla}\varepsilon^2\boldsymbol{u_2}-(p_{\varepsilon,1}-p^*)I)\;\boldsymbol{n}\cdot\boldsymbol{v_H}\right| \leq C\varepsilon\sqrt{\frac{\varepsilon}{H}}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)}\,|\boldsymbol{v_H}|_{H,1}\,. \tag{9.34}$$

For the fifth term of the right-hand side of (9.30), considering

$$p_{\varepsilon,1}-p^*=\varepsilon p_1(x,y)=\varepsilon\pi_i(y)\left(f_i(x)-\frac{\partial p^*}{\partial x_i}(x)\right),$$

and that π_i is bounded, we have by using first the Cauchy–Schwarz inequality, then the discrete Cauchy–Schwarz inequality and sum over the elements,

$$\begin{aligned}\left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}(p_{\varepsilon,1}-p^*)\,\mathrm{div}\,\boldsymbol{v_H}\right| &\leq C\varepsilon\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)}\|\,\mathrm{div}_H\,\boldsymbol{v_H}\|_{L^2(\Omega^\varepsilon)}\\ &\leq C\varepsilon\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)}\|p_H-p_H^*\|_{L^2(\Omega^\varepsilon)}\\ &\leq C\sqrt{\varepsilon}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)}\|p_H-p_H^*\|_{L^2(\Omega^\varepsilon)},\end{aligned} \tag{9.35}$$

where the second line is given by Lemma 9.26. For the sixth term on the right-hand side of (9.30), we use first the Cauchy–Schwarz inequality, then the discrete Cauchy–Schwarz inequality and sum over the elements. We conclude by using Lemma 9.8 and the Poincaré inequality in perforated domains (Lemma 9.24), which leads to

$$\begin{aligned}\left|\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{\nabla}(p^*-I_H^np^*)\cdot\boldsymbol{v_H}\right| &\leq \|\boldsymbol{\nabla}(p^*-I_H^np^*)\|_{L^2(\Omega^\varepsilon)}\;\|\boldsymbol{v_H}\|_{L^2(\Omega^\varepsilon)}\\ &\leq |p^*-I_H^np^*|_{H^1(\Omega^\varepsilon)}\;\|\boldsymbol{v_H}\|_{L^2(\Omega^\varepsilon)}\\ &\leq H^n|p^*|_{H^{n+1}(\Omega^\varepsilon)}\varepsilon\,|\boldsymbol{v_H}|_{H,1}\,.\end{aligned}$$

Grouping all the estimates (9.31) to (9.35), it comes

$$\begin{aligned}\|p_H-p_H^*\|^2_{L^2(\Omega^\varepsilon)} \leq C\varepsilon\Bigg(&\left(\sqrt{\varepsilon}+\sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\\ &+H^n|p^*|_{H^{n+1}(\Omega^\varepsilon)}+H^n\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^n(\Omega)}\Bigg)\,|\boldsymbol{v_H}|_{H,1}\\ &+C\sqrt{\varepsilon}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)}\|p_H-p_H^*\|_{L^2(\Omega^\varepsilon)}.\end{aligned}$$

Now, using $|\boldsymbol{v_H}|_{H,1} \leq \dfrac{C}{\varepsilon}\|p_H - p_H^*\|_{L^2(\Omega^\varepsilon)}$ from Lemma 9.26, it comes

$$\begin{aligned}\|p_H - p_H^*\|^2_{L^2(\Omega^\varepsilon)} \leq C\Bigg(&\left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})} \\ &+ H^n|p^*|_{H^{n+1}(\Omega^\varepsilon)} + H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)}\Bigg)\|p_H - p_H^*\|_{L^2(\Omega^\varepsilon)},\end{aligned}$$

which allows to conclude that

$$\begin{aligned}\|p_H - p_H^*\|_{L^2(\Omega^\varepsilon)} \leq C\Bigg(&\left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})} \\ &+ H^n|p^*|_{H^{n+1}(\Omega)} + H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)}\Bigg).\end{aligned} \tag{9.36}$$

Now, from (9.27), (9.28) and (9.36), we get

$$\begin{aligned}\|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \leq\ & C\Big(H^{n+1}|p^*|_{H^{n+1}(\Omega^\varepsilon)} + \sqrt{\varepsilon}\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\Big) \\ &+ C\Bigg(\left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})} \\ &+ H^n|p^*|_{H^{n+1}(\Omega^\varepsilon)} + H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)}\Bigg).\end{aligned}$$

Noting that $H^{n+1} < H^n$, we conclude that

$$\begin{aligned}&\|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \\ &\leq C\left(\left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})} + H^n|p^*|_{H^{n+1}(\Omega)} + H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)}\right).\end{aligned} \tag{9.37}$$

Now, gathering the estimates for velocity error (9.26) and pressure error (9.37), we reach the estimate (9.12) given in Theorem 9.27, i.e,

$$\begin{aligned}&|\boldsymbol{u_\varepsilon} - \boldsymbol{u_H}|_{H,1} + \varepsilon\|p_\varepsilon - p_H\|_{L^2(\Omega^\varepsilon)} \\ &\leq C\varepsilon\left(H^n\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^n(\Omega)} + H^n|p^*|_{H^{n+1}(\Omega)} + \left(\sqrt{\varepsilon} + \sqrt{\frac{\varepsilon}{H}}\right)\|\boldsymbol{f} - \boldsymbol{\nabla} p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}\right).\end{aligned}$$

Chapter 10

Perspective for MsFEM applied to the Stokes problem: ideas for deriving an a posteriori error estimate

10.1 Motivations

In a multi-scale framework, one important tool is the a posteriori error estimate. In particular, it allows to identify an appropriate trade-off between accuracy and computational cost of the multi-scale simulations. For example for the MsFEM, it allows to choose between decreasing the fine mesh size h or the coarse mesh size H to reach a given accuracy. Several works have studied a posteriori error estimates for incompressible flows in the framework of multi-scale methods. We can cite, for example, [21, 18] which derived an a posteriori error estimate respectively for the Stokes and Brinkman equations and the Oseen equations for the Multi-scale Hybrid-Mixed (MHM) method, or [5] which established an a posteriori error estimate for the Stokes flow in porous media for the Heterogeneous Multi-scale Method (HMM). However, concerning the Multi-scale Finite Element Method (MsFEM) a posteriori error estimates have been developed only for elliptic problems [99]. One of the most advanced work in this field is [46]. However, we can hardly follow the methodology proposed in [46] to develop an a posteriori error estimation for MsFEM for the Stokes problem since we cannot recast the MsFEM developed under a unified formulation for coarse-scale and fine-scale computations (in particular due to the presence of Lagrange multipliers). The main idea of a posteriori error estimates for MsFEM is to decompose the error total, in a microscopic error e_{micro}, which comes from the computation of the basis functions and a macroscopic error e_{macro}, which comes from the computation of the global problem on the coarse mesh, i.e. $e_{total} = e_{micro} + e_{macro}$. The aim is not to get a certified estimator for the error stemming from using a finite value h for the fine mesh size, but to develop a methodology to assess the fine scale and coarse scale computations and to get an approximation respectively of e_{micro} and e_{macro}.

Due to this absence of work for deriving a posteriori error estimates for MsFEMs for incom-

pressible flows, as starting point, the purpose of this chapter is to take ideas from the derivation of a posteriori error estimates for classical non-conforming finite element method. Indeed, we recall that, the presented MsFEM of lowest order ($n = 0$), in absence of obstacles, allows to recover the classical non-conforming Crouzeix–Raviart finite element. In this chapter, we present the method for deriving a posteriori error estimates for the Stokes problem presented in [90]. This paper presents an unified framework to derive an a posteriori error for the Stokes problem which works for a large class of method. In particular, this framework works for non-conforming finite element such as the Crouzeix–Raviart element, and it is based on $(H_0^1(\Omega))^d$-conforming velocity reconstruction and $\underline{\boldsymbol{H}}(\operatorname{div}, \Omega)$-conforming locally conservative flux.

10.2 Problem presentation

We consider the classical Stokes problem (7.1) with homogeneous boundary conditions. The main idea of this methodology, is to introduce a "stress" tensor $\underline{\boldsymbol{\sigma}} \in \underline{\boldsymbol{H}}(\operatorname{div}, \Omega)$, which allows to write the Stokes problems as a system consisting of the "constitutive law"

$$\underline{\boldsymbol{\sigma}} = \boldsymbol{\nabla} \boldsymbol{u} - p \underline{\boldsymbol{I}}, \tag{10.1}$$

the equilibrium equation

$$\operatorname{div} \underline{\boldsymbol{\sigma}} + \boldsymbol{f} = \boldsymbol{0}, \tag{10.2}$$

and the divergence free constraint

$$\operatorname{div} \boldsymbol{u} = 0,$$

where $\underline{\boldsymbol{I}}$ is the identity matrix $d \times d$.

10.3 Estimator

In this part, we present the estimators that can be used for the a posteriori error estimates. We consider an approximation $(\boldsymbol{u_H}, p_H) \in V_H \times M_H$ of the Stokes problem. The velocity u_H can be non-conforming ($\boldsymbol{u_H} \notin V$) and non divergence free. Let $\underline{\boldsymbol{\sigma}}_{\boldsymbol{H}} \in \underline{\boldsymbol{H}}(\operatorname{div}, \Omega)$. We introduce the following estimators:

- Residual estimator
$$\eta_{R,T} = c_{P,T} H_T \| \operatorname{div} \underline{\boldsymbol{\sigma}}_{\boldsymbol{H}} + \boldsymbol{f} \|_T$$
where $C_{P,T}$ is the constant from the Poincaré inequality, related to the possible violation of the equilibrium equation (10.2) in the approximate solution.

- Diffusive flux estimator
$$\eta_{DF,T} = \| \boldsymbol{\nabla} \boldsymbol{u_H} - p_H \underline{\boldsymbol{I}} - \underline{\boldsymbol{\sigma}}_{\boldsymbol{H}} \|_T$$
related to the fact that the constitutive law (10.1) is not satisfied exactly by the approximate solution.

- The non-conformity estimator

$$\eta_{NC,T} = \|\boldsymbol{\nabla}(\boldsymbol{u_H} - \boldsymbol{s_H})\|_T$$

  where $\boldsymbol{s_H}$ is arbitrary in V, that estimates the fact that $\boldsymbol{u_H}$ is not necessarily in V.

- The divergence free estimator

$$\eta_{D,T} = \frac{\|\operatorname{div} \boldsymbol{s_H}\|_T}{\beta}$$

  related to the divergence-free constraint.

The key for the a posteriori error estimates is to construct a flux (stress field) $\underline{\boldsymbol{\sigma}}_{\boldsymbol{H}} \in \underline{\boldsymbol{H}}(\operatorname{div}, \Omega)$ that approximates the local equilibrium i.e. satisfying (10.3), which reads

$$(\operatorname{div} \underline{\boldsymbol{\sigma}}_{\boldsymbol{H}} + \boldsymbol{f}, \boldsymbol{e_i})_T = 0, \ i = 1, \dots, d, \quad \forall T \in \mathcal{T}_H \tag{10.3}$$

Then according to [90] the following error holds.

Theorem 10.1 (Estimate for general approximation)**.** Let $(\boldsymbol{u}, p) \in V \times M$ be the weak solution to the Stokes problem. Let $(\boldsymbol{u_H}, p_H) \in V_H \times M_H$ be arbitrary. Choose an arbitrary $\boldsymbol{s_H} \in V$ and $\underline{\boldsymbol{\sigma}}_{\boldsymbol{H}} \in \underline{\boldsymbol{H}}(\operatorname{div}, \Omega)$ which satisfies (10.3). Then, it holds

$$|||(\boldsymbol{u} - \boldsymbol{u_H}, p - p_H)||| \leq \left(\sum_{T \in \mathcal{T}_H} \eta_{NC,T}^2 \right)^{1/2} + \frac{1}{C_s} \left(\sum_{T \in \mathcal{T}_H} \left[(\eta_{R,T} + \eta_{DF,T})^2 + \eta_{D,T}^2 \right] \right)^{1/2}$$

where

$$|||(\boldsymbol{u}, p)||| := \|\boldsymbol{\nabla u}\|^2 + \beta^2 \|q\|^2,$$

with β the constant from the inf-sup condition.

In [90], the methodology to construct the flux $\underline{\boldsymbol{\sigma}}_{\boldsymbol{H}}$ is presented for the Crouzeix–Raviart element. This allows us to have an a posteriori error estimate for the MsFEM applied to the Stokes problem at the order $n = 0$ in a domain without obstacles. For future work along these lines, it must be studied how the presence of obstacles in the domain affects this methodology. What is more, during this thesis, we have specified the construction of the flux $\underline{\boldsymbol{\sigma}}_{\boldsymbol{H}}$ for high-order Crouzeix–Raviart element. We do not present the calculation here, since it goes beyond the scope of this thesis, and this is still preliminary work. This last point is motivated by the fact that in a domain without obstacles, the space of polynomial of order $n + 1$ is included in the multi-scale approximation space.

Part IV

Extension of the MsFEM to the Oseen problem

Chapter 11

Introduction to convection dominated problem

Contents

In this introductory chapter, we study the Multi-scale Finite Element Method (MsFEM) in the context of advection-diffusion problems in one dimension. The results presented in this chapter are not new, and have been the subject of many works [128, 33]. The goal of this chapter is first to emphasize the difficulty to solve Oseen type problems and second, to see how to use this theoretical analysis to solve Oseen problems with the MsFEM.

11.1 Numerical difficulties for solving advection-diffusion problems

11.1.1 Problem definition

Through this chapter, we consider a simple one-dimensional example, inspired from [148, section 8 - page 258-260]. Let I be the interval $[0,1] \subset \mathbb{R}$ and $f \in L^2(I)$ be a given function. The problem

reads as follows. Find $u : I \to \mathbb{R}$ solution to

$$\left\{ \begin{array}{rcll} -\nu u'' + bu' & = & f & \text{in } I, \\ u(0) & = & 0, & \\ u(1) & = & 1, & \end{array} \right. \tag{11.1}$$

with constants $\nu > 0$ and $b > 0$. For the particular right-hand side $f = 0$, problem (11.1) admits for exact solution

$$u(x) = (e^{b/\nu} - 1)^{-1}(e^{bx/\nu} - 1),$$

and exhibits a boundary layer of width $O(\nu/b)$ near to $x = 1$ if ν/b is small enough.

The weak formulation of problem (11.1) is to find $u \in H_0^1(I)$ such that, for all $w \in H_0^1(I)$, we have

$$a^{adv}(u, w) = \int_I f w \mathrm{d}x, \tag{11.2}$$

with

$$a^{adv}(u, w) = \int_I \left(\nu u'(x)w'(x) + bu'(x)w(x)\right) \mathrm{d}x.$$

We note that since $f \in L^2(I)$, the solution u of (11.2) belongs to $H^2(I)$. Let $N \in \mathbb{N}^*$ be a given integer. Let the segment I be divided into $N + 1$ sub-intervals $I_i = [x_i, x_{i+1}]$ of size $H = \frac{1}{N+1}$ with $i \in [\![0, N]\!]$. Let us suppose that the exact solution to (11.2) is approached by the solution to the following discrete problem. Find $u_H \in V_H$ such that, for all $w_H \in W_H$, we have

$$a^{adv}(u_H, w_H) = \int_I f w_H \mathrm{d}x, \tag{11.3}$$

where V_H (the space of trial functions) and W_H (the space of test functions) are two conforming subsets of $H_0^1(I)$ with equal dimension to be defined.

11.1.2 Identification of instabilities through galerkin approximation and central difference scheme

First, we consider the Galerkin approximation of (11.2), i.e. considering $V_H = W_H$. We approximate (11.1) using usual Galerkin method with piecewise linear finite element (see Figure 11.1) over a uniform grid, i.e. $V_H = W_H = \left\{\varphi \in \mathcal{C}^0(I) \; s.t. \; \varphi|_{I_i} \in \mathbb{P}_1, \; \forall i \in [\![0, N]\!]\right\}$. Taking $f = 0$, the finite element approximation leads to the following linear system

$$A\xi = F, \tag{11.4}$$

with ξ the unknowns vector

$$\xi = \{u_H(x_i)\}_{i=1,\dots,N}.$$

The matrix A is defined by

$$A = \text{tridiag}\left(-\frac{\nu}{H} - \frac{b}{2}, \frac{2\nu}{H}, -\frac{\nu}{H} + \frac{b}{2}\right),$$

and the right-hand side is defined by

$$F = \left(0, \ldots, 0, \frac{\nu}{H} - \frac{b}{2}\right).$$

Assuming that $2\nu \neq bH$, we show that the solution to (11.4) is given by

$$\xi_i = u_H(x_i) = \frac{(\frac{1+Pe}{1-Pe})^i - 1}{(\frac{1+Pe}{1-Pe})^{N+1} - 1}, \quad i = 1, \ldots, N, \tag{11.5}$$

where Pe is the Péclet number defined by

$$Pe = \frac{bH}{2\nu}.$$

If $Pe > 1$, then $(1 + Pe)/(1 - Pe) < 0$, and consequently the approximation u_H exhibits an oscillatory behaviour (see Figure 11.2). For fixed ν and b it is always possible to choose the mesh size H small enough so that $Pe < 1$, thus avoiding oscillations. However, this is often impractical if ν is very small compared to b, since one would obtain a linear system with too many unknowns.

Another approach to highlight these instabilities, is to remark that the linear system (11.4) is equivalent to the one obtained by approximating (11.1) by central finite difference scheme, which reads

$$\begin{cases} -\nu \dfrac{u_{i+1} - 2u_i + u_{i-1}}{H^2} + b\dfrac{u_{i+1} - u_{i-1}}{2H} &= 0, \quad \text{for } i = 1, \ldots, N, \\ u_0 &= 0, \\ u_{N+1} &= 1, \end{cases} \tag{11.6}$$

where u_i is the approximation of $u(x_i)$.

Now it is easy to see that the solution will be inaccurate due to the lack of stability of the discrete problem. Indeed, the first line of (11.6), can be rewritten as

$$(Pe - 1)u_{i+1} + 2u_i - (1 + Pe)u_{i-1} = 0,$$

an equation whose product of roots of its characteristic polynomial is equal to $\frac{1+Pe}{1-Pe}$. Consequently, if $Pe > 1$, the characteristic polynomial has one positive root and one negative root, resulting in instabilities in the numerical solution.

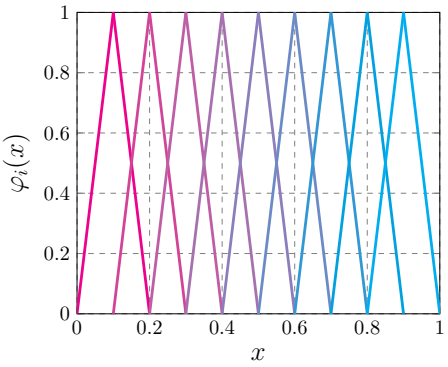


Figure 11.1: $\mathbb{P}_1$ basis functions φ_i.

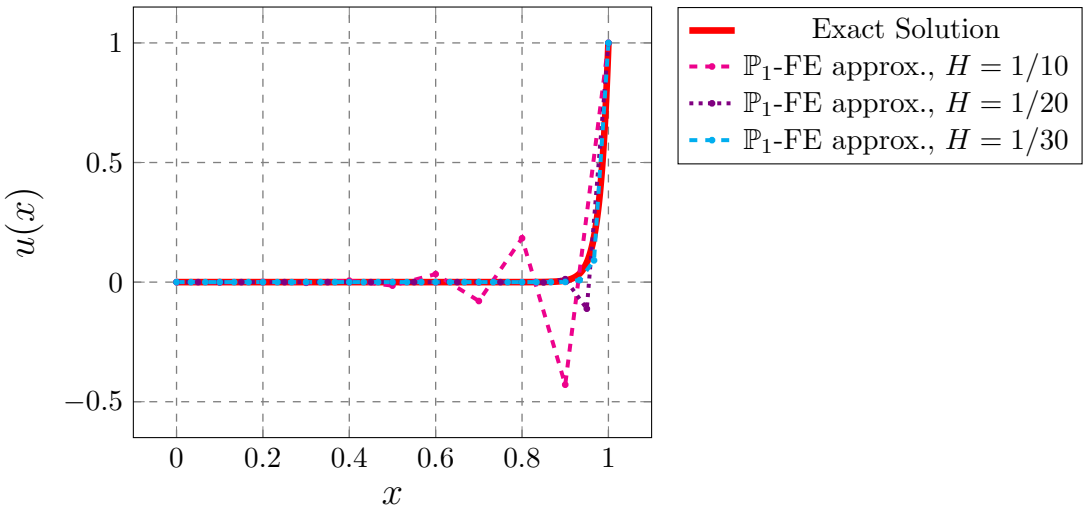


Figure 11.2: $\mathbb{P}_1$ Finite Element approximations of (11.1) for different mesh size H.

To mitigate or remove these numerical instabilities, different stabilisation methods have been developed in the literature for such problems. In the next section, we study a first way to mitigate the instabilities using an upwind finite difference method, which can be shown to be equivalent to a Petrov–Galerkin formulation, i.e. considering well-chosen different spaces for the trial and test spaces.

11.1.3 First strategy of stabilisation: Petrov-Galerking formulation and Upwind Finite Difference

As noted in [148], the simplest stable discretization for the problem (11.1) is the upwind scheme, defined as

$$\begin{cases} -\nu \dfrac{u_{i+1} - 2u_i + u_{i-1}}{H^2} - b \dfrac{u_i - u_{i-1}}{H} &= 0, \quad \text{for } i = 1, \dots, N, \\ u_0 &= 0, \\ u_{N+1} &= 1, \end{cases} \tag{11.7}$$

which admits for solution

$$u_i = \frac{(1+2Pe)^i - 1}{(1+2Pe)^{N+1} - 1}, \quad i = 1, \dots, N.$$

Now, it is easy to remark that u_i no longer oscillates, whatever the value of the Péclet number Pe. What is more, we remark that (11.7) is equivalent to

$$-\nu\frac{u_{i+1}-2u_i+u_{i-1}}{H^2}-b\frac{u_{i+1}-u_{i-1}}{H}-b\frac{H}{2}\frac{u_{i+1}-2u_i+u_{i-1}}{H^2}=0.$$

Consequently, the upwind scheme can be regarded as a stabilisation technique for the central difference scheme (11.6) by introducing a numerical dissipation with the discretization of the artificial viscous term $-b\frac{H}{2}u''$. As noted in [148], it is also possible to interpret the upwind difference scheme as a Petrov–Galerkin approximation of (11.1), with a space of trial functions V_H and a space of test functions W_H that are different. The space of trial functions is chosen as $V_H = X_H^1 \cap H_0^1(\Omega)$ where

$$X_H^1 = \left\{\varphi \in \mathcal{C}^0(I)\ s.t.\ \varphi|_{I_i} \in \mathbb{P}_1,\ \forall i \in [\![0, N]\!]\right\},$$

To define the space of test functions W_H, we consider the piecewise quadratic functions

$$\sigma_2(x) = \begin{cases} -3x(1+x), & -1 \le x \le 0, \\ -3x(1-x), & 0 \le x \le 1, \\ 0, & |x| \ge 1. \end{cases}$$

We define then the finite dimensional space W_H as

$$W_H = \text{span}\{\psi_1, \dots, \psi_N\},$$

where

$$\psi_i(x) = \varphi_i(x) + \sigma_2(H^{-1}x - i), \quad i = 1, \dots, N,$$

φ_i being the basis functions of V_H corresponding to the node x_i.

As conclusion, two methods for stabilising convection dominated problems have been presented. The first one is adding new terms (called stabilisation terms) in the discretization of the problem. The second one, is using a Petrov–Galerkin formulation, i.e. using different spaces of test and trial functions. In what follows, we focus on stabilisation techniques.

11.1.4 Stabilisation strategy

In view of (11.5), to eliminate the oscillations, one can refine the mesh so that the convection no longer dominates at the element level. However, this is in contradiction with the paradigm of MsFEM, whose goal is to capture macroscopic features of the problem on a rather coarse mesh. Another approach is consequently to use stabilisation techniques. In the literature, it has been proposed several stabilisation techniques to overcome advection-dominated regime. They consist in adding a stabilisation term in the Galerkin formulation, by considering a new test function $\tilde{w}$ as

$$\tilde{w} = w + \tau(\mathcal{L}_{ss} + \rho\mathcal{L}_s)w,$$

where $\mathcal{L}_s : u \mapsto -\nu u''$ and $\mathcal{L}_{ss} : u \mapsto bu'$ are the symmetric part and the skew symmetric part (provided that $\operatorname{div} b = 0$) of the advection–diffusion operator $\mathcal{L} : u \mapsto -\nu u'' + bu'$ and $\tau > 0$ the

stabilisation parameter. According to the value of the parameter ρ, different stabilised methods can be recovered. If $\rho = -1$, we recover the DWG (Gouglas-Wang/Galerkin) method, if $\rho = 1$, we recover the Galerkin/Least-Squares method (GLS), and if $\rho = 0$, we recover the Streamline Upwind / Petrov-Galerkin method (SUPG).

In what follows, we focus only on the SUPG (streamline upwind / Petrov-Galerkin) formulation that has been successfully applied for convection dominated problems [125, 37]. With the SUPG formulation, the standard Galerkin test functions are modified by adding a streamline upwind perturbation, which acts only in the flow direction, a priori eliminating the possibility of any crosswind diffusion. The modified test functions can be applied to all terms in the equations, resulting in a strongly consistent weighted residual formulation. The SUPG test function $\tilde{w}$ is defined by

$$\tilde{w} = w + \tau b w',$$

where w is the Galerkin type test function and τ is the stabilisation parameter. One of the most common τ is defined as

$$\tau = \frac{H}{2|b|}\left(\coth(\beta H) - \frac{1}{\beta H}\right), \tag{11.8}$$

with βH the Péclet number Pe.

We introduce the following bi-linear forms associated with the problem

$$a^{adv,stab}(u, v) = \int_I \tau(-\nu u'' + bu')bv' \mathrm{d}x, \tag{11.9}$$

with τ the stabilisation parameter. We introduce also the linear form

$$F(v) = \int_I fv \mathrm{d}x.$$

The stabilised version of the advection–diffusion problem (11.1) is to find $u_H \in V_H$ solution to

$$a_H^{adv}(u_H, w_H) + a_H^{adv,stab}(u_H, w_H) = F(w_H) \quad \forall w_H \in W_H. \tag{11.10}$$

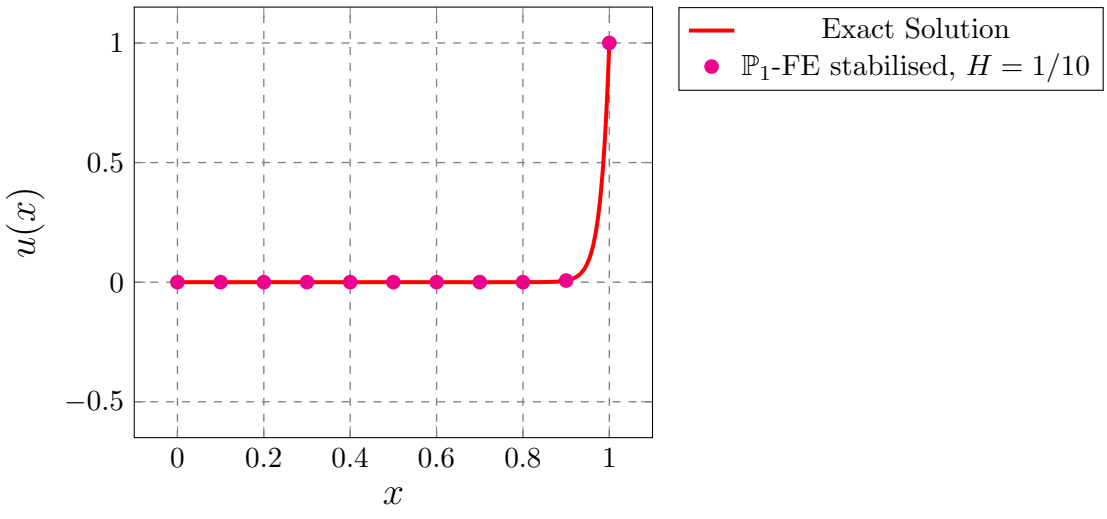


Figure 11.3: Stabilised $\mathbb{P}_1$ Finite Element approximation of (11.1).

It can be seen from Figure 11.3, the stabilised finite element approximation with SUPG method using the stabilisation parameter τ defined by (11.8) does not exhibit oscillatory behaviour compared to Figure 11.2.

Remark 11.1. The choice of an optimal parameter τ is a difficult and sensitive question, since it affects the quality of the numerical approximation. The value of τ given in (11.8), is one of the most used in the literature. It is derived in the framework of Variational Multi-scale Methods (VMS) using Green's function [107].

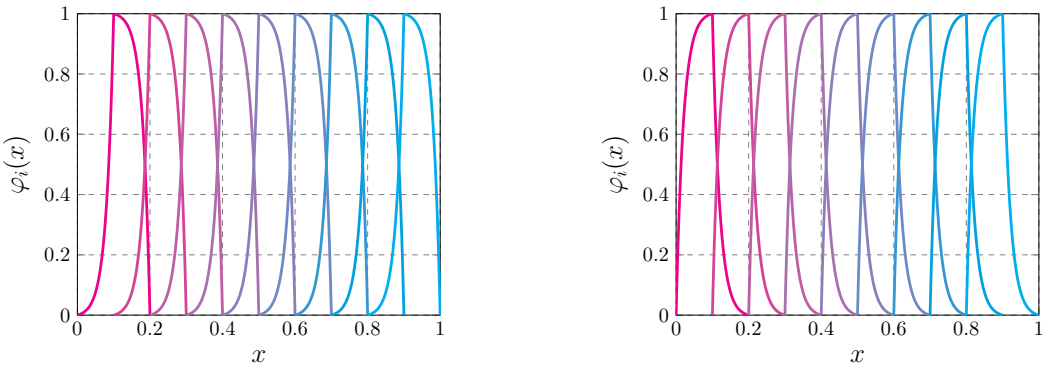
In the next section, we give an interpretation of the stabilisation parameter τ given in (11.8) in the context of Multi-scale Finite Element Method.

11.2 Advection-Diffusion in the MsFEM context

The goal of this section is to give a justification to the choice of the parameter τ, defined in (11.8), in the context of MsFEM. Now, instead of considering $\mathbb{P}_1$ finite element, we consider well-adapted multi-scale basis functions to the problem (11.1). We suppose that V_H and W_H are conforming in $H_0^1(I)$ and defined respectively by functions satisfying the direct and the adjoint problems on each segment I_i in the following way

$$V_H = \left\{\varphi \in \mathcal{C}^0(I), -\nu(\varphi|_{I_i})'' + b(\varphi|_{I_i})' = 0,\ \forall i \in [\![0, N]\!]\right\},$$
$$W_H = \left\{\varphi \in \mathcal{C}^0(I), -\nu(\varphi|_{I_i})'' - b(\varphi|_{I_i})' = 0,\ \forall i \in [\![0, N]\!]\right\}.$$

Figures 11.4a and 11.4b represent respectively examples of basis functions in V_H and W_H.



(a) Multi-scale basis functions of V_H.

(b) Multi-scale basis functions of W_H.

Figure 11.4: Multi-scale basis functions.

The Galerkin formulation of problem (11.1) leads to the following linear system

$$A^{adv}(V_H, V_H)\xi = F(V_H), \tag{11.11}$$

where $A^{adv}(V_H, V_H)$ is the matrix form of (11.3) whose coefficients are given by

$$\begin{cases} A_{i,i-1} = \dfrac{-b}{1 - e^{-\frac{b}{\nu}H}}, \\ A_{i,i+1} = \dfrac{-b\, e^{-\frac{b}{\nu}H}}{1 - e^{-\frac{b}{\nu}H}}, \\ A_{i,i} = |A_{i,i-1}| + |A_{i,i+1}| = b \coth\left(\dfrac{bH}{2\nu}\right). \end{cases}$$

Now, it is easy to conclude that this formulation (11.11) is always stable and no oscillations appear whatever the size of $\frac{bH}{\nu}$ due to the dominance of the diagonal (see for example Figure 11.5). What is more, the following equality holds

$$A^{adv}(V_H, V_H) = A^{adv}(V_H, W_H). \tag{11.12}$$

where $A^{adv}(V_H, W_H)$ is the stiffness matrix obtained with a Petrov–Galerkin formulation of problem (11.3). As we know that the discrete problem is well-posed with the Galerkin formulation, then consequently due to (11.12), the discrete problem arising from the Petrov-Galerkin formulation is also well-posed. The two formulations become equivalent if the right-hand side is equal to 0. In particular, the Petrov–Galerkin formulation leads to an approximation that is exact at the nodes (see Lemma 11.2).

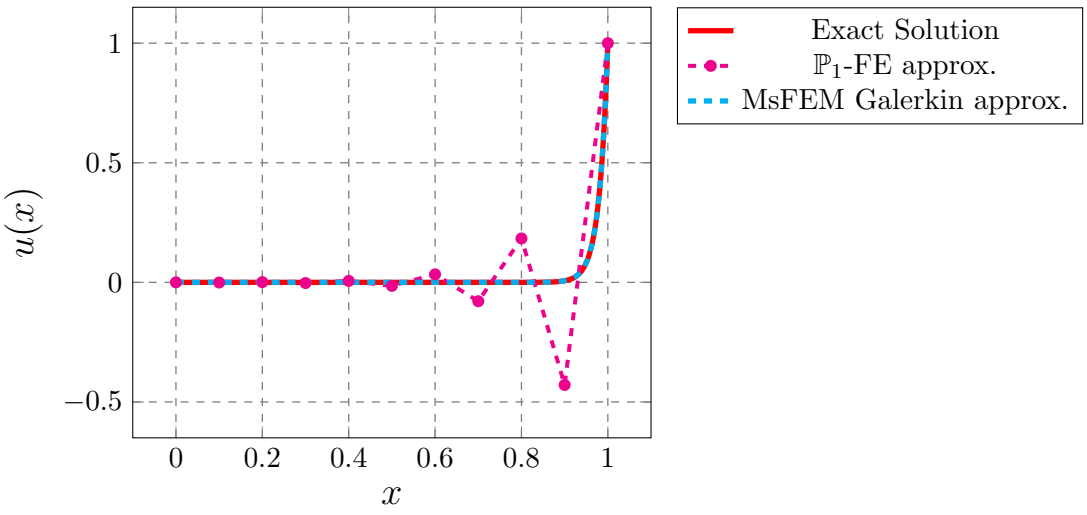


Figure 11.5: MsFEM approximation compared to $\mathbb{P}_1$ approximation.

In the two following paragraphs, we give an interpretation of the choice of the stabilisation parameter in the context of MsFEM and we prove that with a Petrov-Galerkin approach the approximated solution of (11.1) is exact at the nodes.

11.2.1 Interpretation of the stabilisation parameter

Let us choose the stabilisation parameter τ as

$$\tau = \frac{H}{2b}\left(\coth(\beta H) - \frac{1}{\beta H}\right), \tag{11.13}$$

with βH the Péclet number i.e.

$$Pe = \beta H = \frac{bH}{2\nu},$$

then the following equalities hold

$$A^{adv}(V_H, V_H) = A^{adv}(V_H, W_H) = A^{adv}(\mathbb{P}_1, \mathbb{P}_1) + A^{adv,stab}(\mathbb{P}_1, \mathbb{P}_1), \tag{11.14}$$

where $A^{adv}(\mathbb{P}_1, \mathbb{P}_1)$ and $A^{adv,stab}(\mathbb{P}_1, \mathbb{P}_1)$ are respectively the matrix forms of (11.3) and (11.9) obtained with linear test and trial functions, which are defined by

$$A^{adv}(\mathbb{P}_1, \mathbb{P}_1) = \text{tridiag}\left(-\frac{\nu}{H} - \frac{b}{2}\ ,\ \frac{2\nu}{H}\ ,\ -\frac{\nu}{H} + \frac{b}{2}\right),$$

and

$$A^{adv,stab}(\mathbb{P}_1, \mathbb{P}_1) = \text{tridiag}\left(-\tau\frac{b^2}{H}\ ,\ 2\tau\frac{b^2}{H}\ ,\ -\tau\frac{b^2}{H}\right).$$

The equalities (11.14) allow to give an explanation to the choice of the stabilisation parameter τ. Indeed, the choice of τ (11.13) leads to a stable formulation when discretising (11.10) with affine piecewise polynomials. Besides, if the right-hand side is equal to 0, this choice of τ leads to a numerical approximation that is exact at the nodes. One another important point is that $A^{adv,stab}(V_H, V_H) = 0$, resulting that no stabilisation is required to solve problem (11.1) with MsFEM.

11.2.2 High-order MsFEM and exactness of the solution

In what follows, we prove that the MsFEM Petrov–Galerkin formulation of (11.1) leads to an approximation that is exact at the nodes. In particular, we consider a more general case by considering an high-order MsFEM. Let $k \in \mathbb{N}$ be a given integer. We add the subscript k to denote the approximations and the spaces, in order to emphasize their dependency regarding to k. Let us suppose that the exact solution to (11.2) is approached by the solution to the following discrete problem. Find $u_{H,k} \in V_{H,k}$ such that, for all $w_{H,k} \in W_{H,k}$, we have

$$a^{adv}(u_{H,k}, w_{H,k}) = \int_I f w_{H,k} \mathrm{d}x,$$

where $V_{H,k}$ and $W_{H,k}$ are two conforming subsets of $H_0^1(I)$ with equal dimension to be defined. Let us now turn to the actual construction of $W_{H,k}$. Since exactness properties essentially rely on the definition of $W_{H,k}$, we start by its definition. We suppose that $W_{H,k}$ is conforming in $H_0^1(I)$ and defined by functions satisfying adjoint problems on each segment in the following way

$$W_{H,k} = \left\{w_{H,k} \in \mathcal{C}^0(I), -\nu(w_{H,k}|_{I_i})'' - b(w_{H,k}|_{I_i})' \in \mathbb{P}_{k-1}(I_i)\ \forall i \in [\![0, N]\!]\right\},$$

with the convention that, for the case $k = 0$, $\mathbb{P}_{-1}(I_i) = \{0\}$. Note that it is easily shown that

$$-\nu(w_{H,k}|_{I_i})'' - a(w_{H,k}|_{I_i})' \in \mathbb{P}_{k-1}(I_i) \iff w_{H,k}|_{I_i} \in \mathbb{P}_k(I_i) \oplus \text{span}\left\{x \mapsto \exp\left(-\frac{bx}{\nu}\right)\right\}.$$

We have then the following result.

Lemma 11.2. Provided that (11.3) yields a well-posed problem and that functions in $V_{H,k}$ are uniquely defined by their values at the vertices $(x_i)_{i\in[\![0,N+1]\!]}$ and by their moments against all monomials $x \mapsto x^q$ with $q \in [\![0, k-1]\!]$ on each I_i with $i \in [\![0, N]\!]$, the solution to (11.3) has the following exactness properties,

$$\begin{aligned} u_{H,k}(x_i) &= u(x_i)\ ,\ \forall i \in [\![0, N+1]\!], \\ \int_{I_i} x^q u_{H,k}|_{I_i}(x)\mathrm{d}x &= \int_{I_i} x^q (u|_{I_i})(x)\mathrm{d}x\ ,\ \forall i \in [\![0, N]\!]\ ,\ \forall q \in [\![0, k-1]\!]. \end{aligned}$$

Proof. Let us detail the calculation of $a^{adv}(u, w_{H,k})$.

$$a^{adv}(u, w_{H,k}) = \sum_{i=0}^{N} \int_{I_i} \left(\nu (u|_{I_i})'(x)(w_{H,k}|_{I_i})'(x) + b(u|_{I_i})'(x)(w_{H,k}|_{I_i})(x)\right) \mathrm{d}x. \tag{11.15}$$

On each interval I_i we can integrate by parts in (11.15) and, using that $u \in H^2(I) \subset \mathcal{C}^0(\overline{I})$, we get

$$\begin{aligned} &\int_{I_i} \left(\nu (u|_{I_i})'(x)(w_{H,k}|_{I_i})'(x) + b(u|_{I_i})'(x)(w_{H,k}|_{I_i})(x)\right)\mathrm{d}x \\ &= \nu u(x_{i+1})(w_{H,k}|_{I_i})'(x_{i+1}) + bu(x_{i+1})(w_{H,k}|_{I_i})(x_{i+1}) \\ &- \left[\nu u(x_i)(w_{H,k}|_{I_i})'(x_i) + bu(x_i)(w_{H,k}|_{I_i})(x_i)\right] \\ &+ \int_{I_i} \left(-\nu (w_{H,k}|_{I_i})''(x) - b(w_{H,k}|_{I_i})'(x)\right)(u|_{I_i})(x)\mathrm{d}x. \end{aligned} \tag{11.16}$$

Now, using the fact that functions in $V_{H,k}$ are uniquely defined by their values at the vertices $(x_i)_{i\in[\![0,N+1]\!]}$ and by their moments against all monomials $x \mapsto x^q$ with $q \in [\![0, k-1]\!]$ on each I_i with $i \in [\![0, N]\!]$, we can define a function $\Pi_{h,k}u \in V_{H,k}$ such that

$$\begin{aligned} \Pi_{h,k}u(x_i) &= u(x_i) \quad \forall i \in [\![0, N+1]\!], \\ \int_{I_i} x^q (\Pi_{h,k}u|_{I_i})(x)\mathrm{d}x &= \int_{I_i} x^q (u|_{I_i})(x)\mathrm{d}x \quad \forall i \in [\![0, N]\!]\ ,\ \forall q \in [\![0, k-1]\!]. \end{aligned}$$

Using this definition and the fact that $(-\nu(w_{H,k}|_{I_i})'' - b(w_{H,k}|_{I_i})') \in \mathbb{P}_{k-1}(I_i)$ for all $i \in [\![0, N]\!]$, function u can be replaced by function $\Pi_{h,k}u$ in the right-hand side of (11.16) and integration by parts can be performed the other way round to get

$$a^{adv}(u, w_{H,k}) = a^{adv}(\Pi_{h,k}u, w_{H,k}) = (f, w_{H,k})\ ,\ \forall w_{H,k} \in W_{H,k}, \tag{11.17}$$

the last equality being obtained using $w_{H,k}$ as test functions in (11.2). Now, since we have supposed that (11.3) has a unique solution, then (11.17) proves that $\Pi_{h,k}u$ is this unique solution. Therefore, the solution to (11.3) is such that it is exact at the vertices of the mesh and that it has exact moments against all monomials $x \mapsto x^q$ with $q \in [\![0, k-1]\!]$ on each I_i with $i \in [\![0, N]\!]$. □

Remark 11.3. As long as we have not specified what the space $V_{H,k}$ is, these exactness properties are not enough to ensure that $\Pi_{h,k}u$ is an accurate approximation of u.

11.3 Conclusion

In this chapter, we have studied the MsFEM in the context of advection dominated problems in one dimension. Before considering the genuine Oseen problem in two and three dimensions in the next chapter, two main points should be kept from this chapter. The first one is that no stabilisation is required for MsFEM approximations of advection dominated problems since advection is already encoded in the basis functions. The second one, is that the MsFEM Petrov–Galerkin formulation of the advection-diffusion problem in one dimension leads to an approximation that is exact at the nodes. The question now is to check whether these observations hold or not for the Oseen problem. In particular, does the MsFEM Petrov–Galerkin formulation of the Oseen problem lead to more accurate results?

Chapter 12

MsFEM for the Oseen problem

Contents

This chapter is dedicated to the extension of the Multi-scale Finite Element Method (MsFEM) presented in Chapter 7 to the Oseen problem. In particular, we also propose an MsFEM Petrov–Galerkin formulation of the Oseen problem, the trial and test functions solving respectively local problems deriving from the direct and the adjoint Oseen problems. In Section 12.1, we derive the local basis functions defined by the Oseen operator. In Section 12.2, we derive the local basis functions defined by the Oseen adjoint operator. In Section 12.3, we explicit the coarse Galerkin and Petrov–Galerkin formulation of the Oseen problem.

12.1 Multi-scale basis functions defined by the Oseen operator

12.1.1 Problem formulation

The steady state Oseen problem with homogeneous boundary conditions is to find the velocity $\boldsymbol{u}_\varepsilon : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\begin{aligned} -\nu\Delta\boldsymbol{u}_\varepsilon + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u}_\varepsilon + \boldsymbol{\nabla}p &= \boldsymbol{f} \quad \text{in} \quad \Omega^\varepsilon, \\ \operatorname{div}\boldsymbol{u}_\varepsilon &= 0 \quad \text{in} \quad \Omega^\varepsilon, \\ \boldsymbol{u}_\varepsilon &= \boldsymbol{0} \quad \text{on} \quad \partial B^\varepsilon \cap \partial\Omega^\varepsilon, \\ \boldsymbol{u}_\varepsilon &= \boldsymbol{0} \quad \text{on} \quad \partial\Omega \cap \partial\Omega^\varepsilon, \end{aligned} \tag{12.1}$$

with $\boldsymbol{U_0}$ a known velocity, $\nu > 0$ the viscosity and $\boldsymbol{f}$ a given force. We introduce the velocity space

$$V = H_0^1(\Omega^\varepsilon)^d = \{\boldsymbol{u} \in H^1(\Omega^\varepsilon)^d \ s.t. \ \boldsymbol{u}|_{\partial\Omega^\varepsilon} = \boldsymbol{0}\},$$

and the pressure space

$$M = L_0^2(\Omega^\varepsilon) = \{p \in L^2(\Omega^\varepsilon) \ s.t. \ \int_{\Omega^\varepsilon} p = 0\}.$$

We note $X = V \times M$. For simplicity the fluid domain is assumed to be connected in order for the pressure to be uniquely defined in M. We introduce the bi-linear form $a^{Os}(\cdot,\cdot)$ for the Oseen problem

$$a^{Os}(\boldsymbol{u},\boldsymbol{v}) = \int_{\Omega^\varepsilon} \nu \boldsymbol{\nabla u} \cdot \boldsymbol{\nabla v} + (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u} \cdot \boldsymbol{v}$$

and the bi-linear form

$$b(q,\boldsymbol{v}) = -\int_{\Omega^\varepsilon} q \operatorname{div} \boldsymbol{v}.$$

The variational formulation of the Oseen problem is to find $\boldsymbol{u_\varepsilon} \in V$ and $p_\varepsilon \in M$ such that

$$\left\{ \begin{array}{rcll} a^{Os}(\boldsymbol{u_\varepsilon},\boldsymbol{v}) + b(p_\varepsilon,\boldsymbol{v}) & = & \displaystyle\int_{\Omega^\varepsilon} \boldsymbol{f} \cdot \boldsymbol{v} & \forall \boldsymbol{v} \in V, \\ b(q,\boldsymbol{u_\varepsilon}) & = & 0 & \forall q \in M. \end{array} \right. \tag{12.2}$$

Theorem 3.3 guarantees the existence and uniqueness of a solution to problem (12.2). Indeed, we introduce another continuous bi-linear form equivalent to $a^{Os}(\cdot,\cdot)$, defined as

$$\begin{aligned} \widehat{a}^{Os}(\boldsymbol{u_\varepsilon},\boldsymbol{v}) &= \int_{\Omega^\varepsilon} \nu \boldsymbol{\nabla u_\varepsilon} \cdot \boldsymbol{\nabla v} + \int_{\Omega^\varepsilon} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u_\varepsilon} \cdot \boldsymbol{v} \\ &= \int_{\Omega^\varepsilon} \nu \boldsymbol{\nabla u_\varepsilon} \cdot \boldsymbol{\nabla v} + \frac{1}{2}\int_{\Omega^\varepsilon} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u_\varepsilon} \cdot \boldsymbol{v} + \frac{1}{2}\int_{\Omega^\varepsilon} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u_\varepsilon} \cdot \boldsymbol{v} \\ &= \int_{\Omega^\varepsilon} \nu \boldsymbol{\nabla u_\varepsilon} \cdot \boldsymbol{\nabla v} + \frac{1}{2}\int_{\Omega^\varepsilon} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u_\varepsilon} \cdot \boldsymbol{v} - \frac{1}{2}\int_{\Omega^\varepsilon} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{v} \cdot \boldsymbol{u_\varepsilon} - \frac{1}{2}\int_{\Omega^\varepsilon} \operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon} \cdot \boldsymbol{v} \\ &= \int_{\Omega^\varepsilon} \nu \boldsymbol{\nabla u_\varepsilon} \cdot \boldsymbol{\nabla v} + \frac{1}{2}\int_{\Omega^\varepsilon} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u_\varepsilon} \cdot \boldsymbol{v} - \frac{1}{2}\int_{\Omega^\varepsilon} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{v} \cdot \boldsymbol{u_\varepsilon} - \frac{1}{2}\int_{\Omega^\varepsilon} \operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon} \cdot \boldsymbol{v} \end{aligned}$$

with $\boldsymbol{u_\varepsilon} = \boldsymbol{0}$ on $\partial\Omega$. Since the two bi-linear forms $a^{Os}(\cdot,\cdot)$ and $\widehat{a}^{Os}(\cdot,\cdot)$ are equivalent, the coercivity of $\widehat{a}^{Os}(\cdot,\cdot)$ implies the coercivity of $a^{Os}(\cdot,\cdot)$. It is easy to observe that

$$\widehat{a}^{Os}(\boldsymbol{u_\varepsilon},\boldsymbol{u_\varepsilon}) = \int_{\Omega^\varepsilon} \nu \boldsymbol{\nabla u_\varepsilon} \cdot \boldsymbol{\nabla u_\varepsilon} - \frac{1}{2}\int_{\Omega^\varepsilon} \operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon} \cdot \boldsymbol{u_\varepsilon} \geq \int_{\Omega^\varepsilon} \nu |\boldsymbol{\nabla u_\varepsilon}|^2$$

as soon as $\operatorname{div}(\boldsymbol{U_0}) \leq 0$.

Remark 12.1. It is easy to show that $\widehat{a}^{Os}(\cdot,\cdot)$ remains coercive if there exists $\alpha > 0$ such that

$$\nu - \frac{1}{2}C\varepsilon \|\operatorname{div} \boldsymbol{U_0}\|_\infty \geq \alpha$$

where C is the Poincaré constant.

Since the semi-norm $\|\boldsymbol{\nabla v}\|_{L^2(\Omega^\varepsilon)}$ is equivalent to the full H^1 norm by Poincaré inequality, we

obtain that $\widehat{a}^{Os}(\cdot,\cdot)$, and consequently $a^{Os}(\cdot,\cdot)$ is elliptic on V. Besides $b(\cdot,\cdot)$ verifies the inf-sup condition (similar as for the Stokes problem). Consequently, using Theorem 3.3, it comes that the following variational formulation of the Oseen problem is well-posed.

$$\left\{\begin{array}{rcll} \widehat{a}^{Os}(\boldsymbol{u_\varepsilon},\boldsymbol{v}) + b(p_\varepsilon,\boldsymbol{v}) & = & \displaystyle\int_{\Omega^\varepsilon} \boldsymbol{f}\cdot\boldsymbol{v} & \forall \boldsymbol{v}\in V, \\ b(q,\boldsymbol{u_\varepsilon}) & = & 0 & \forall q\in M. \end{array}\right.$$

We introduce the bi-linear form $\widehat{c}^{Os}(\cdot,\cdot)$,

$$\begin{aligned}\widehat{c}^{Os}\left((\boldsymbol{u_\varepsilon},p_\varepsilon),(\boldsymbol{v},q)\right) = \int_{\Omega^\varepsilon}\Big(\nu\boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla v} + \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v} - \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{u_\varepsilon} - \frac{1}{2}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v}\Big) \\ - \int_{\Omega^\varepsilon} p_\varepsilon \operatorname{div}\boldsymbol{v} - \int_{\Omega^\varepsilon} q\operatorname{div}\boldsymbol{u_\varepsilon}.\end{aligned}$$

The variational formulation of the Oseen problem (12.1) can be written as: find $(\boldsymbol{u_\varepsilon},p_\varepsilon)\in V\times M$ such that

$$\widehat{c}^{Os}\left((\boldsymbol{u_\varepsilon},p_\varepsilon),(\boldsymbol{v},q)\right) = \int_{\Omega^\varepsilon}\boldsymbol{f}\cdot\boldsymbol{v} \quad \forall(\boldsymbol{v},q)\in V\times M. \tag{12.3}$$

Theorem 3.2 guarantees the existence and uniqueness of a solution to problem (12.3).

12.1.2 The high-order Crouzeix–Raviart Multi-Scale Finite Element Method defined by Oseen equations

To define the high-order Crouzeix–Raviart Multi-scale Finite Element for the Oseen problem, we follow the same procedure as for the Stokes problem in Chapter 7.

Definition 12.2. The velocity-pressure space X_H^{Os} is defined as a subspace of X_H^{ext}, defined in (7.6), being the "orthogonal" complement of X_H^0 (see Definition 7.3), with respect to the bi-linear form $\widehat{c}_H^{Os}(\cdot,\cdot)$ as follows

$$(\boldsymbol{u_H},p_H)\in X_H^{Os} \iff \begin{array}{l}(\boldsymbol{u_H},p_H)\in X_H^{ext} \text{ such that} \\ \widehat{c}_H^{Os}((\boldsymbol{u_H},p_H),(\boldsymbol{v},q)) = 0\ \forall(\boldsymbol{v},q)\in X_H^0\end{array},$$

where $\widehat{c}_H^{Os}$ is defined by

$$\begin{aligned}\widehat{c}_H^{Os}\left((\boldsymbol{u_H},p_H),(\boldsymbol{v},q)\right) = \sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\Big(\nu\boldsymbol{\nabla u_H}\cdot\boldsymbol{\nabla v} + \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_H}\cdot\boldsymbol{v} \\ -\frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{u_H} - \frac{1}{2}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_H}\cdot\boldsymbol{v}\Big) \\ + \sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\left(-p_H\operatorname{div}\boldsymbol{v} - q\operatorname{div}\boldsymbol{u_H}\right).\end{aligned}$$

Remark 12.3. The word "orthogonal" is written between quotes since the bi-linear form $\widehat{c}_H^{Os}$ is not a scalar product (not positive definite).

In what follows, we explicit the definition of the subspace X_H^{Os} by introducing the functional spaces V_H^{Os} and M_H.

Definition 12.4. We introduce the functional spaces $M_H \subset M$ and $V_H^{Os} \subset V_H^{ext}$ by

$$M_H = \{q \in M \text{ such that } q|_T \in \mathbb{P}_n(T),\ \forall T \in \mathcal{T}_H\}, \tag{12.4}$$

$$V_H^{Os} = \begin{cases} \boldsymbol{v} \in V_H^{ext} \ : \ \forall T \ \in \ \mathcal{T}_H,\ \exists\ \zeta_T \in M_H^0(T) \text{ such that} \\ -\nu\boldsymbol{\Delta v} + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v} + \boldsymbol{\nabla}\zeta_T \ \in \ \text{span}(\boldsymbol{\varphi_{T,1}}, \dots \boldsymbol{\varphi_{T,r}}) \text{ in } T \cap \Omega^\varepsilon, \\ \text{div}\,\boldsymbol{v} \ \in \ \text{span}(\varpi_{T,1}, \dots, \varpi_{T,t}) \text{ in } T \cap \Omega^\varepsilon, \\ \nu\boldsymbol{\nabla v n} - \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{n})\boldsymbol{v} - \zeta_T\boldsymbol{n} \ \in \ \text{span}(\boldsymbol{\omega_{F,1}}, \dots, \boldsymbol{\omega_{F,s}}) \text{ on } F \cap \Omega^\varepsilon,\ \forall F \in \mathcal{F}(T), \end{cases} \tag{12.5}$$

where $\mathcal{F}(T)$ is the set of faces composing ∂T, and M_H^0 defined (7.8).

Lemma 12.5. In definition (12.5), the pressure ζ_T is uniquely defined for a given velocity $\boldsymbol{v} \in V_H^{Os}$. Therefore gluing together the pressure ζ_T on all triangles $T \in \mathcal{T}_H$ yields a single function $\pi_H(\boldsymbol{v}) \in M_H^0$ such that $\pi_H(\boldsymbol{v}) = \zeta_T$ on any triangle $T \in \mathcal{T}_H$. The operator $\pi_H : V_H^{Os} \to M_H^0$ is linear and continuous.

Proof. The proof of Lemma 12.5 is similar to the proof of [79, Lemma 3.11]. □

Theorem 12.6. Using M_H (12.4) and V_H^{Os} (12.5), we introduce the space:

$$\underline{X}_H^{Os} = \text{span}\{(\boldsymbol{u_H}, \pi_H(\boldsymbol{u_H}) + \overline{p}_H),\ \boldsymbol{u_H} \in V_H^{Os},\ \overline{p}_H \in M_H\}$$

Then it coincides with the subspace X_H^{Os}, i.e.

$$X_H^{Os} = \underline{X}_H^{Os}.$$

Proof. The proof of Theorem 12.6 is similar to the proof of [79, Theorem 3.12]. □

12.1.2.1 The local problems defined by Oseen equations

In what follows, we construct a basis of the space V_H^{Os} which consists of functions associated with faces (face-based basis functions) and functions associated with coarse elements (element-based basis functions) of the coarse mesh.

12.1.2.2 Strong form of local problems

Basis functions associated with faces We first construct basis functions associated with faces of the coarse mesh. For any $F \in \mathcal{F}_H$, for $i = 1, \dots, s$, find the function $\boldsymbol{\Phi_{F,i}} : \Omega^\varepsilon \to \mathbb{R}^d$,

the pressure $\pi_{F,i} : \Omega^\varepsilon \to \mathbb{R}$ by solving on T

$$\begin{cases} -\nu\Delta\boldsymbol{\Phi_{F,i}} + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{\Phi_{F,i}} + \boldsymbol{\nabla}\pi_{F,i} \in \text{span}(\boldsymbol{\varphi_{T,1}},\dots,\boldsymbol{\varphi_{T,r}}) \text{ in } T\cap\Omega^\varepsilon, \\ \text{div}\,\boldsymbol{\Phi_{F,i}} \in \text{span}(\varpi_{T,1},\dots,\varpi_{T,t}) \text{ in } T\cap\Omega^\varepsilon, \\ \nu\boldsymbol{\nabla}\boldsymbol{\Phi_{F,i}}\boldsymbol{n} - \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{n})\boldsymbol{\Phi_{F,i}} - \pi_{F,i}\boldsymbol{n} \in \text{span}(\boldsymbol{\omega_{E,1}},\dots,\boldsymbol{\omega_{E,s}}) \text{ on } E\cap\Omega^\varepsilon,\ \forall E\in\mathcal{F}(T), \\ \boldsymbol{\Phi_{F,i}} = \boldsymbol{0} \text{ on } \partial B^\varepsilon\cap T, \\ \displaystyle\int_{E\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\omega_{E,j}} = \begin{cases}\delta_{ij}, & E=F\\ 0, & E\neq F\end{cases} \quad \forall E\in\mathcal{F}(T),\ \forall j=1,\dots,s, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\varphi_{T,l}} = 0\ \forall l=1,\dots,r, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \pi_{F,i}\varpi_{T,m} = 0\ \forall m=1,\dots,t. \end{cases} \tag{12.6}$$

Basis functions associated with elements Next, we construct basis functions associated with elements of the coarse mesh. For each $T\in\mathcal{T}_H$, for $k=1,\dots,r$, the support of the function $\boldsymbol{\Psi_{T,k}}$ is reduced to $T\cap\Omega^\varepsilon$. We find $\boldsymbol{\Psi_{T,k}} : \Omega^\varepsilon\to\mathbb{R}^d$ and $\pi_{T,k} : \Omega^\varepsilon\to\mathbb{R}$ by solving on T

$$\begin{cases} -\nu\Delta\boldsymbol{\Psi_{T,k}} + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{\Psi_{T,k}} + \boldsymbol{\nabla}\pi_{T,k} \in \text{span}(\boldsymbol{\varphi_{T,1}},\dots,\boldsymbol{\varphi_{T,r}}) \text{ in } T\cap\Omega^\varepsilon, \\ \text{div}\,\boldsymbol{\Psi_{T,k}} \in \text{span}(\varpi_{\mathrm{T},1},\dots,\varpi_{\mathrm{T,t}}) \text{ in } \mathrm{T}\cap\Omega^\varepsilon, \\ \nu\boldsymbol{\nabla}\boldsymbol{\Psi_{T,k}}\boldsymbol{n} - \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{n})\boldsymbol{\Psi_{T,k}} - \pi_{T,k}\boldsymbol{n} \in \text{span}(\boldsymbol{\omega_{E,1}},\dots,\boldsymbol{\omega_{E,s}}) \text{ on } E\cap\Omega^\varepsilon,\ \forall E\in\mathcal{F}(T), \\ \boldsymbol{\Psi_{T,k}} = \boldsymbol{0} \text{ on } \partial B^\varepsilon\cap T, \\ \displaystyle\int_{E\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{\omega_{E,j}} = 0,\ \forall E\in\mathcal{F}(T),\ \forall j=1,\dots,s, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{\varphi_{T,l}} = \delta_{kl}\ \forall l=1,\dots,r, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \pi_{T,k}\varpi_{T,m} = 0\ \forall m=1,\dots,t. \end{cases} \tag{12.7}$$

Theorem 12.7. The functions $\boldsymbol{\Phi_{F,i}}$ for $F\in\mathcal{F}_H$ and $i=1,\dots,s$ defined by their restrictions (12.6) over each element T whose boundary includes F and $\boldsymbol{\Psi_{T,k}}$ for $T\in\mathcal{T}_H$ and $k=1,\dots,r$ defined by (12.7) form a basis of V_H^{Os} defined by (12.5). In other words,

$$V_H^{Os} = \text{span}\{\boldsymbol{\Phi_{F,i}},\ \boldsymbol{\Psi_{T,k}},\ F\in\mathcal{F}_H,\ T\in\mathcal{T}_H,\ i=1\dots,s,\ k=1,\dots,r\}$$

and $\{\boldsymbol{\Phi_{F,i}},\ F\in\mathcal{F}_H,\ i=1\dots,s\}\cup\{\boldsymbol{\Psi_{T,k}},\ T\in\mathcal{T}_H,\ k=1,\dots,r\}$ forms a linearly independent family.

Proof. The proof of Theorem 12.7 is similar to the proof in [79, section 3.5]. □

12.1.2.3 Weak form of local problems

Using the same notations as in Chapter 8, the weak form of the local problems reads as follows.

Basis functions associated with faces The weak form of (12.6) is: for any $F\in\mathcal{F}_H$, for $i=1,\dots,s$, on the coarse element T find $\boldsymbol{\Phi_{F,i}}\in V_T$, $\pi_{F,i}\in L^2(T\cap\Omega^\varepsilon)$, $\boldsymbol{\lambda_F}\in\mathbb{R}^{n_s}$, $\boldsymbol{\lambda_T^v}\in\mathbb{R}^r$,

$\boldsymbol{\lambda_T^q} \in \mathbb{R}^t$ by solving, for all $(\boldsymbol{v}, q, \boldsymbol{\mu_F}, \boldsymbol{\mu_T^v}, \boldsymbol{\mu_T^q}) \in V_T \times L^2(T \cap \Omega^\varepsilon) \times \mathbb{R}^{n_s} \times \mathbb{R}^r \times \mathbb{R}^t$,

$$\begin{cases} \displaystyle\int_{T\cap\Omega^\varepsilon} \nu \boldsymbol{\nabla \Phi_{F,i}} \cdot \boldsymbol{\nabla v} + \int_{T\cap\Omega^\varepsilon} \left(\frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{v} - \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{\Phi_{F,i}} - \frac{1}{2}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{v} \right) \\ \displaystyle - \int_{T\cap\Omega^\varepsilon} \pi_{F,i} \operatorname{div}\boldsymbol{v} + \sum_{E\in\mathcal{F}(T)} \sum_{j=1}^{s} \lambda_{E,j} \int_{F\cap\Omega^\varepsilon} \boldsymbol{v}\cdot\boldsymbol{\omega_{E,j}} - \sum_{l=1}^{r} \lambda_{T,l}^v \int_{T\cap\Omega^\varepsilon} \boldsymbol{\varphi_{T,l}}\cdot\boldsymbol{v} = 0, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} q \operatorname{div}\boldsymbol{\Phi_{F,i}} + \sum_{m=1}^{t} \lambda_{T,m}^q \int_{T\cap\Omega^\varepsilon} \varpi_{T,m} q = 0, \\ \displaystyle\sum_{E\in\mathcal{F}(T)} \sum_{j=1}^{s} \mu_{E,j} \int_{E\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\omega_{E,j}} = \mu_{F,i}, \\ \displaystyle\sum_{l=1}^{r} \mu_{T,l}^v \int_{T\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\varphi_{T,l}} = 0, \\ \displaystyle\sum_{m=1}^{t} \mu_{T,m}^q \int_{T\cap\Omega^\varepsilon} \pi_{F,i}\varpi_{T,m} = 0. \end{cases} \tag{12.8}$$

Basis functions associated with elements The weak form of (12.7) is: for any $T \in \mathcal{T}_H$, for $k = 1, \dots, r$, find: $\boldsymbol{\Psi_{T,k}} \in V_T$, $\pi_{T,k} \in L^2(T \cap \Omega^\varepsilon)$, $\boldsymbol{\lambda_F} \in \mathbb{R}^{n_s}$, $\boldsymbol{\lambda_T^v} \in \mathbb{R}^r$, $\boldsymbol{\lambda_T^q} \in \mathbb{R}^t$ by solving, for all $(\boldsymbol{v}, q, \boldsymbol{\mu_F}, \boldsymbol{\mu_T^v}, \boldsymbol{\mu_T^q}) \in V_T \times L^2(T \cap \Omega^\varepsilon) \times \mathbb{R}^{n_s} \times \mathbb{R}^r \times \mathbb{R}^t$,

$$\begin{cases} \displaystyle\int_{T\cap\Omega^\varepsilon} \nu \boldsymbol{\nabla \Psi_{T,k}} \cdot \boldsymbol{\nabla v} + \int_{T\cap\Omega^\varepsilon} \left(\frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{v} - \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{\Psi_{T,k}} - \frac{1}{2}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{v} \right) \\ \displaystyle - \int_{T\cap\Omega^\varepsilon} \pi_{T,k} \operatorname{div}\boldsymbol{v} + \sum_{E\in\mathcal{F}(T)} \sum_{j=1}^{s} \lambda_{E,j} \int_{F\cap\Omega^\varepsilon} \boldsymbol{v}\cdot\boldsymbol{\omega}_{E,j} - \sum_{l=1}^{r} \lambda_{T,l}^v \int_{T\cap\Omega^\varepsilon} \boldsymbol{\varphi}_{T,l}\cdot\boldsymbol{v} = 0, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} q \operatorname{div}\boldsymbol{\Psi_{T,k}} + \sum_{m=1}^{t} \lambda_{T,m}^q \int_{T\cap\Omega^\varepsilon} \varpi_{T,m} q = 0, \\ \displaystyle\sum_{E\in\mathcal{F}(T)} \sum_{j=1}^{s} \mu_{E,j} \int_{E\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{\omega_{E,j}} = 0, \\ \displaystyle\sum_{l=1}^{r} \mu_{T,l}^v \int_{T\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{\varphi_{T,l}} = \mu_{T,k}^v, \\ \displaystyle\sum_{m=1}^{t} \mu_{T,m}^q \int_{T\cap\Omega^\varepsilon} \pi_{T,k}\varpi_{T,m} = 0. \end{cases} \tag{12.9}$$

Remark 12.8. In this work, we used a skew-symmetric version of the advective term, i.e. the bi-linear form $\widehat{a}^{Os}(\cdot,\cdot)$, so that the problem is coercive as long as $\operatorname{div}(U_0) \leq 0$. However, [33] stated that using the skew-symmetric formulation leads to less accurate results, than using the direct formulation $a^{Os}(\cdot,\cdot)$. In this work, we have not made the numerical comparisons. However, at the theoretical point of view, it is not clear how a change in the basis functions leads to more of less accurate results at the coarse scale. It could be interesting, in the future, to perform this comparison.

12.2 Multi-scale basis functions defined by the Oseen adjoint operator

12.2.1 Problem formulation

By integration by parts, it is easy to show that the adjoint of the Oseen operator is given by

$$\boldsymbol{u} \mapsto -\Delta \boldsymbol{u} - (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u} - \operatorname{div}(\boldsymbol{U_0})\boldsymbol{u}.$$

The steady state Oseen adjoint problem with homogeneous boundary conditions is to find the velocity $\boldsymbol{u} : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\begin{aligned} -\nu\Delta \boldsymbol{u_\varepsilon} - (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u_\varepsilon} - \operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon} + \boldsymbol{\nabla} p &= \boldsymbol{f} \quad \text{in} \quad \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u_\varepsilon} &= 0 \quad \text{in} \quad \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &= \boldsymbol{0} \quad \text{on} \quad \partial B^\varepsilon \cap \partial\Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &= \boldsymbol{0} \quad \text{on} \quad \partial\Omega \cap \partial\Omega^\varepsilon, \end{aligned} \tag{12.10}$$

with $\boldsymbol{U_0}$ a known velocity, ν the kinematic viscosity and $\boldsymbol{f}$ a given force. As previously, we introduce for the velocity the space

$$V = H_0^1(\Omega^\varepsilon)^d = \{\boldsymbol{u} \in H^1(\Omega^\varepsilon)^d \ s.t. \ \boldsymbol{u}|_{\partial\Omega^\varepsilon} = \boldsymbol{0}\},$$

and for the pressure

$$M = L_0^2(\Omega^\varepsilon) = \{p \in L^2(\Omega^\varepsilon) \ s.t. \ \int_{\Omega^\varepsilon} p = 0\}.$$

We note $X = V \times M$. For simplicity the fluid domain is assumed to be connected in order for the pressure to be uniquely defined in M.

We introduce the bi-linear form $a^{Os,ad}(\cdot,\cdot)$ for the Oseen adjoint problem,

$$a^{Os,ad}(\boldsymbol{u},\boldsymbol{v}) = \int_{\Omega^\varepsilon} \nu\boldsymbol{\nabla u} \cdot \boldsymbol{\nabla v} - (\boldsymbol{U_0} \cdot \boldsymbol{\nabla})\boldsymbol{u} \cdot \boldsymbol{v} - \operatorname{div}(\boldsymbol{U_0})\boldsymbol{u} \cdot \boldsymbol{v}.$$

The variational formulation of the Oseen adjoint problem is to find $\boldsymbol{u_\varepsilon} \in V$ and $p \in M$ such that

$$\begin{aligned} a^{Os,ad}(\boldsymbol{u_\varepsilon},\boldsymbol{v}) + b(p_\varepsilon,\boldsymbol{v}) &= \int_{\Omega^\varepsilon} \boldsymbol{f} \cdot \boldsymbol{v} \quad \forall \boldsymbol{v} \in V, \\ b(q,\boldsymbol{u_\varepsilon}) &= 0 \qquad\quad \forall q \in M. \end{aligned} \tag{12.11}$$

Theorem 3.3 guarantees the problem (12.11) is well-posed. Indeed, we introduce another con-

tinuous bi-linear form which is equivalent to $a^{Os,ad}(\cdot,\cdot)$,

$$\begin{aligned}
\widehat{a}^{Os,ad}(\boldsymbol{u_\varepsilon},\boldsymbol{v}) &= \int_{\Omega^\varepsilon} \nu\boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla v} - \int_{\Omega^\varepsilon}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v} - \int_{\Omega^\varepsilon}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v}\\
&= \int_{\Omega^\varepsilon} \nu\boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla v} - \frac{1}{2}\int_{\Omega^\varepsilon}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v} - \frac{1}{2}\int_{\Omega^\varepsilon}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v} - \int_{\Omega^\varepsilon}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v}\\
&= \int_{\Omega^\varepsilon} \nu\boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla v} - \frac{1}{2}\int_{\Omega^\varepsilon}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v} + \frac{1}{2}\int_{\Omega^\varepsilon}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{u_\varepsilon} + \frac{1}{2}\int_{\Omega^\varepsilon}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v}\\
&\quad - \int_{\Omega^\varepsilon}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v}\\
&= \int_{\Omega^\varepsilon} \nu\boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla v} - \frac{1}{2}\int_{\Omega^\varepsilon}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v} + \frac{1}{2}\int_{\Omega^\varepsilon}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{u_\varepsilon} - \frac{1}{2}\int_{\Omega^\varepsilon}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v}.
\end{aligned}$$

Since the two bi-linear forms $a^{Os,ad}(\cdot,\cdot)$ and $\widehat{a}^{Os,ad}(\cdot,\cdot)$ are equivalent, the coercivity of $\widehat{a}^{Os,ad}(\cdot,\cdot)$ implies the coercivity of $a^{Os,ad}(\cdot,\cdot)$. It is easy to observe that

$$\widehat{a}^{Os,ad}(\boldsymbol{u_\varepsilon},\boldsymbol{u_\varepsilon}) = \int_{\Omega^\varepsilon}\nu\boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla u_\varepsilon} - \frac{1}{2}\int_{\Omega^\varepsilon}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{u_\varepsilon} \geq \int_{\Omega^\varepsilon}\nu|\boldsymbol{\nabla u_\varepsilon}|^2,$$

as soon as $\operatorname{div}(\boldsymbol{U_0}) \leq 0$.

Since the semi-norm $\|\boldsymbol{\nabla v}\|_{L^2(\Omega)}$ is equivalent to the full H^1 norm by Poincaré inequality, we obtain that $\widehat{a}^{Os,ad}(\cdot,\cdot)$, and consequently $a^{Os,ad}(\cdot,\cdot)$ is elliptic on V. Besides $b(\cdot,\cdot)$ verifies the inf-sup condition (similar as for the Stokes problem). Consequently, using Theorem 3.3, it comes that the following variational formulation of the Oseen adjoint problem is well-posed.

$$\left\{\begin{array}{rcll}
\widehat{a}^{Os,ad}(\boldsymbol{u_\varepsilon},\boldsymbol{v}) + b(p_\varepsilon,\boldsymbol{v}) &=& \displaystyle\int_{\Omega^\varepsilon}\boldsymbol{f}\cdot\boldsymbol{v} & \forall\boldsymbol{v}\in V,\\
b(q,\boldsymbol{u_\varepsilon}) &=& 0 & \forall q\in M.
\end{array}\right.$$

We introduce also the bi-linear form $\widehat{c}^{Os,ad}(\cdot,\cdot)$,

$$\begin{aligned}
\widehat{c}^{Os,ad}\left((\boldsymbol{u_\varepsilon},p_\varepsilon),(\boldsymbol{v},q)\right) &= \int_{\Omega^\varepsilon}\left(\nu\boldsymbol{\nabla u_\varepsilon}\cdot\boldsymbol{\nabla v} - \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v} + \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{u_\varepsilon} - \frac{1}{2}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{u_\varepsilon}\cdot\boldsymbol{v}\right)\\
&\quad - \int_{\Omega^\varepsilon}p_\varepsilon\operatorname{div}\boldsymbol{v} - \int_{\Omega^\varepsilon}q\operatorname{div}\boldsymbol{u_\varepsilon}.
\end{aligned}$$

The variational formulation of the Oseen problem can be written as: find $(\boldsymbol{u_\varepsilon},p_\varepsilon)\in V\times M$ such that

$$\widehat{c}^{Os,ad}\left((\boldsymbol{u_\varepsilon},p_\varepsilon),(\boldsymbol{v},q)\right) = \int_{\Omega^\varepsilon}\boldsymbol{f}\cdot\boldsymbol{v} \quad \forall(\boldsymbol{v},q)\in V\times M. \tag{12.12}$$

Theorem 3.2 guarantees the existence and uniqueness of a solution to problem (12.12).

12.2.2 The high-order Crouzeix–Raviart multi-scale finite element method defined by Oseen adjoint equations

Definition 12.9. The velocity-pressure space $X_H^{Os,ad}$ is defined as a subspace of X_H^{ext}, defined in (7.6), being the "orthogonal" complement of X_H^0 (see Definition 7.3), with respect to the bi-linear form $\widehat{c}_H^{Os,ad}(\cdot,\cdot)$ as follows

$$(\boldsymbol{u_H}, p_H) \in X_H^{Os,ad} \iff \begin{array}{l} (\boldsymbol{u_H}, p_H) \in X_H^{ext} \text{ such that} \\ \widehat{c}_H^{Os,ad}((\boldsymbol{u_H}, p_H), (\boldsymbol{v}, q)) = 0 \ \forall (\boldsymbol{v}, q) \in X_H^0, \end{array}$$

where $\widehat{c}_H^{Os,ad}$ is defined by

$$\begin{aligned} \widehat{c}_H^{Os,ad}((\boldsymbol{u_H}, p_H), (\boldsymbol{v}, q)) = \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} \Big(& \nu \boldsymbol{\nabla u_H} \cdot \boldsymbol{\nabla v} - \frac{1}{2} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla}) \boldsymbol{u_H} \cdot \boldsymbol{v} \\ & + \frac{1}{2} (\boldsymbol{U_0} \cdot \boldsymbol{\nabla}) \boldsymbol{v} \cdot \boldsymbol{u_H} - \frac{1}{2} \operatorname{div}(\boldsymbol{U_0}) \boldsymbol{u_H} \cdot \boldsymbol{v} \Big) \\ & + \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} (-p_H \operatorname{div} \boldsymbol{v} - q \operatorname{div} \boldsymbol{u_H}) \,. \end{aligned}$$

Remark 12.10. The word "orthogonal" is written between quotes since the bi-linear form $\widehat{c}_H^{Os,ad}$ is not a scalar product (not positive definite).

In what follows, we make explicit the definition of the subspace $X_H^{Os,ad}$ by introducing the functional spaces $V_H^{Os,ad}$ and M_H.

Definition 12.11. We introduce the functional spaces M_H and $V_H^{Os,ad}$ by

$$M_H = \{ q \in M \ \text{ such that } q|_T \in \mathbb{P}_n(T), \ \forall T \in \mathcal{T}_H \}, \tag{12.13}$$

$$V_H^{Os,ad} = \left\{ \begin{array}{l} \boldsymbol{v} \in V_H^{ext} \ : \ \forall T \ \in \ \mathcal{T}_H, \ \exists \ \zeta_T \in M_H^0(T) \ \text{ such that} \\ -\nu \Delta \boldsymbol{v} - (\boldsymbol{U_0} \cdot \boldsymbol{\nabla}) \boldsymbol{v} - \operatorname{div}(\boldsymbol{U_0}) \boldsymbol{v} + \boldsymbol{\nabla} \zeta_T \ \in \ \operatorname{span}(\boldsymbol{\varphi_{T,1}}, \ldots \boldsymbol{\varphi_{T,r}}) \text{ in } T \cap \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{v} \ \in \ \operatorname{span}(\varpi_{T,1}, \ldots, \varpi_{T,t}) \text{ in } T \cap \Omega^\varepsilon, \\ \nu \boldsymbol{\nabla v n} + \frac{1}{2} (\boldsymbol{U_0} \cdot \boldsymbol{n}) \boldsymbol{v} - \zeta_T \boldsymbol{n} \ \in \ \operatorname{span}(\boldsymbol{\omega_{F,1}}, \ldots, \boldsymbol{\omega_{F,s}}) \text{ on } F \cap \Omega^\varepsilon, \ \forall F \in \mathcal{F}(T). \end{array} \right. \tag{12.14}$$

where $\mathcal{F}(T)$ is the set of faces composing ∂T and M_H^0 defined (7.8).

Lemma 12.12. In definition (12.14), the pressure ζ_T is uniquely defined for a given velocity $\boldsymbol{v} \in V_H^{Os,ad}$. Therefore gluing together the pressure ζ_T on all triangles $T \in \mathcal{T}_H$ yields a single function $\pi_H(\boldsymbol{v}) \in M_H^0$ such that $\pi_H(\boldsymbol{v}) = \zeta_T$ on any triangle $T \in \mathcal{T}_H$. The operator $\pi_H : V_H^{Os,ad} \to M_H^0$ is linear and continuous.

Proof. The proof of Lemma 12.12 is similar to [79, Lemma 3.11]. □

Theorem 12.13. Using M_H (12.13) and $V_H^{Os,ad}$ (12.14), we introduce the space

$$\underline{X}_H^{Os,ad} = \text{span}\{(\boldsymbol{u_H}, \pi_H(\boldsymbol{u_H}) + \overline{p}_H),\ \boldsymbol{u_H} \in V_H^{Os,ad},\ \overline{p}_H \in M_H\}$$

Then it coincides with the subspace $X_H^{Os,ad}$, i.e.

$$X_H^{Os,ad} = \underline{X}_H^{Os,ad}.$$

Proof. The proof of Theorem 12.13 is similar to [79, Theorem 3.12]. □

Now, we construct a basis of the space $V_H^{Os,ad}$ which consists of functions associated with faces (face-based basis functions) and to coarse element (element-based basis functions) of the coarse mesh.

12.2.2.1 Strong form of local problems

Basis functions associated with faces We first construct basis functions associated with faces of the coarse mesh. For any $F \in \mathcal{F}_H$, for $i = 1, \dots, s$, find the function $\boldsymbol{\Phi_{F,i}} : \Omega^\varepsilon \to \mathbb{R}^d$, the pressure $\pi_{F,i} : \Omega^\varepsilon \to \mathbb{R}$ and solve on T

$$\begin{cases} -\nu\Delta\boldsymbol{\Phi_{F,i}} - (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{\Phi_{F,i}} - \text{div}\,(\boldsymbol{U_0})\boldsymbol{\Phi_{F,i}} + \boldsymbol{\nabla}\pi_{F,i} \in \text{span}(\boldsymbol{\varphi_{T,1}}, \dots, \boldsymbol{\varphi_{T,r}}) \text{ in } T\cap\Omega^\varepsilon, \\ \text{div}\,\boldsymbol{\Phi_{F,i}} \in \text{span}(\varpi_{T,1}, \dots, \varpi_{T,t}) \text{ in } T\cap\Omega^\varepsilon, \\ \nu\boldsymbol{\nabla}\boldsymbol{\Phi_{F,i}}\boldsymbol{n} + \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{n})\boldsymbol{\Phi_{F,i}} - \pi_{F,i}\boldsymbol{n} \in \text{span}(\boldsymbol{\omega_{E,1}}, \dots, \boldsymbol{\omega_{E,s}}) \text{ on } E\cap\Omega^\varepsilon,\ \forall E \in \mathcal{F}(T), \\ \boldsymbol{\Phi_{F,i}} = \boldsymbol{0} \text{ on } \partial B^\varepsilon \cap T, \\ \displaystyle\int_{E\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\omega_{E,j}} = \begin{cases} \delta_{ij}, & E = F \\ 0, & E \neq F \end{cases} \quad \forall E \in \mathcal{F}(T),\ \forall j = 1, \dots, s, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\varphi_{T,l}} = 0\ \forall l = 1, \dots, r, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \pi_{F,i}\varpi_{T,m} = 0\ \forall m = 1, \dots, t. \end{cases} \tag{12.15}$$

Basis functions associated with elements Now we construct basis functions associated with elements of the coarse mesh. For each $T \in \mathcal{T}_H$, for $k = 1, \dots, r$, the support of the function

$\boldsymbol{\Psi_{T,k}}$ is reduced to $T \cap \Omega^\varepsilon$. We find $\boldsymbol{\Psi_{T,k}} : \Omega^\varepsilon \to \mathbb{R}^d$ and $\pi_{T,k} : \Omega^\varepsilon \to \mathbb{R}$ by solving on T,

$$\begin{cases} -\nu\Delta\boldsymbol{\Psi_{T,k}} - (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{\Psi_{T,k}} - \operatorname{div}(\boldsymbol{U_0})\boldsymbol{\Psi_{T,k}} + \boldsymbol{\nabla}\pi_{T,k} \in \operatorname{span}(\boldsymbol{\varphi_{T,1}},\dots,\boldsymbol{\varphi_{T,r}}) \text{ in } T\cap\Omega^\varepsilon, \\ \operatorname{div}\boldsymbol{\Psi_{T,k}} \in \operatorname{span}(\varpi_{T,1},\dots,\varpi_{T,t}) \text{ in } T\cap\Omega^\varepsilon, \\ \nu\boldsymbol{\nabla}\boldsymbol{\Psi_{T,k}}\boldsymbol{n} + \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{n})\boldsymbol{\Psi_{T,k}} - \pi_{T,k}\boldsymbol{n} \in \operatorname{span}(\boldsymbol{\omega_{E,1}},\dots,\boldsymbol{\omega_{E,s}}) \text{ on } F\cap\Omega^\varepsilon,\ \forall E\in\mathcal{F}(T), \\ \boldsymbol{\Psi_{T,k}} = \boldsymbol{0} \text{ on } \partial B^\varepsilon \cap T, \\ \displaystyle\int_{E\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{\omega_{E,j}} = 0,\ \forall E\in\mathcal{F}(T),\ \forall j=1,\dots,s, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{\varphi_{T,l}} = \delta_{kl}\ \forall l=1,\dots,r, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} \pi_{T,k}\varpi_{T,m} = 0\ \forall m=1,\dots,t. \end{cases} \tag{12.16}$$

Theorem 12.14. The functions $\boldsymbol{\Phi_{F,i}}$ for $F\in\mathcal{F}_H$ and $i=1,\dots,s$ defined by their restrictions (12.15) over each element T whose boundary includes F and $\boldsymbol{\Psi_{T,k}}$ for $T\in\mathcal{T}_H$ and $k=1,\dots,r$ defined by (12.16) form a basis of $V_H^{Os,ad}$ defined by (12.14). In other words,

$$V_H^{Os,ad} = \operatorname{span}\{\boldsymbol{\Phi_{F,i}},\ \boldsymbol{\Psi_{T,k}},\ F\in\mathcal{F}_H,\ T\in\mathcal{T}_H,\ i=1\dots,s,\ k=1,\dots,r\}$$

and $\{\boldsymbol{\Phi_{F,i}},\ F\in\mathcal{F}_H,\ i=1\dots,s\}\cup\{\boldsymbol{\Psi_{T,k}},\ T\in\mathcal{T}_H,\ k=1,\dots,r\}$ forms a linearly independent family.

Proof. The proof of Theorem 12.14 is similar to the proof in [79, section 3.5]. □

12.2.2.2 Weak form of local problems

Using the same notations as in Chapter 8, the weak form of the local problems reads as follows.

Basis functions associated with faces The weak form of (12.15) is: for any $F\in\mathcal{F}_H$, for $i=1,\dots,s$, on the coarse element T, find $\boldsymbol{\Phi_{F,i}}\in V_T$, $\pi_{F,i}\in L^2(T\cap\Omega^\varepsilon)$, $\boldsymbol{\lambda_F}\in\mathbb{R}^{n_s}$, $\boldsymbol{\lambda_T^v}\in\mathbb{R}^r$, $\boldsymbol{\lambda_T^q}\in\mathbb{R}^t$ by solving, for all $(\boldsymbol{v},q,\boldsymbol{\mu_F},\boldsymbol{\mu_T^v},\boldsymbol{\mu_T^q})\in V_T\times L^2(T\cap\Omega^\varepsilon)\times\mathbb{R}^{n_s}\times\mathbb{R}^r\times\mathbb{R}^t$,

$$\begin{cases} \displaystyle\int_{T\cap\Omega^\varepsilon} \nu\boldsymbol{\nabla}\boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\nabla}\boldsymbol{v} + \int_{T\cap\Omega^\varepsilon}\left(-\frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{v} + \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{\Phi_{F,i}} - \frac{1}{2}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{v}\right) \\ \displaystyle -\int_{T\cap\Omega^\varepsilon}\pi_{F,i}\operatorname{div}\boldsymbol{v} + \sum_{E\in\mathcal{F}(T)}\sum_{j=1}^{s}\lambda_{E,j}\int_{F\cap\Omega^\varepsilon}\boldsymbol{v}\cdot\boldsymbol{\omega_{E,j}} - \sum_{l=1}^{r}\lambda_{T,l}^v\int_{T\cap\Omega^\varepsilon}\boldsymbol{\varphi_{T,l}}\cdot\boldsymbol{v} = 0, \\ \displaystyle\int_{T\cap\Omega^\varepsilon} q\operatorname{div}\boldsymbol{\Phi_{F,i}} + \sum_{m=1}^{t}\lambda_{T,m}^q\int_{T\cap\Omega^\varepsilon}\varpi_{T,m}q = 0, \\ \displaystyle\sum_{E\in\mathcal{F}(T)}\sum_{j=1}^{s}\mu_{E,j}\int_{E\cap\Omega^\varepsilon}\boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\omega_{E,j}} = \mu_{F,i}, \\ \displaystyle\sum_{l=1}^{r}\mu_{T,l}^v\int_{T\cap\Omega^\varepsilon}\boldsymbol{\Phi_{F,i}}\cdot\boldsymbol{\varphi_{T,l}} = 0, \\ \displaystyle\sum_{m=1}^{t}\mu_{T,l}^q\int_{T\cap\Omega^\varepsilon}\pi_{F,i}\varpi_{T,m} = 0. \end{cases} \tag{12.17}$$

Basis functions associated with elements The weak form of (12.16) is: for any $T \in \mathcal{T}_H$, for $k = 1, \dots, r$, find $\boldsymbol{\Psi_{T,k}} \in V_T$, $\pi_{T,k} \in L^2(T \cap \Omega^\varepsilon)$, $\boldsymbol{\lambda_F} \in \mathbb{R}^{n_s}$, $\boldsymbol{\lambda_T^v} \in \mathbb{R}^r$, $\boldsymbol{\lambda_T^q} \in \mathbb{R}^t$ by solving, for all $(\boldsymbol{v}, q, \boldsymbol{\mu_F}, \boldsymbol{\mu_T^v}, \boldsymbol{\mu_T^q}) \in V_T \times L^2(T \cap \Omega^\varepsilon) \times \mathbb{R}^{n_s} \times \mathbb{R}^r \times \mathbb{R}^t$,

$$
\begin{cases}
\displaystyle\int_{T\cap\Omega^\varepsilon} \nu \boldsymbol{\nabla\Psi_{T,k}} \cdot \boldsymbol{\nabla v} + \int_{T\cap\Omega^\varepsilon} \left(-\frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{v} + \frac{1}{2}(\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{v}\cdot\boldsymbol{\Psi_{T,k}} - \frac{1}{2}\operatorname{div}(\boldsymbol{U_0})\boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{v} \right) \\
\displaystyle - \int_{T\cap\Omega^\varepsilon} \pi_{T,k} \operatorname{div}\boldsymbol{v} + \sum_{E\in\mathcal{F}(T)} \sum_{j=1}^{s} \lambda_{E,j} \int_{E\cap\Omega^\varepsilon} \boldsymbol{v}\cdot\boldsymbol{\omega_{E,j}} - \sum_{l=1}^{r} \lambda_{T,l}^v \int_{T\cap\Omega^\varepsilon} \boldsymbol{\varphi}_{T,l}\cdot\boldsymbol{v} = 0, \\
\displaystyle\int_{T\cap\Omega^\varepsilon} q \operatorname{div}\boldsymbol{\Psi_{T,k}} + \sum_{m=1}^{t} \lambda_{T,m}^q \int_{T\cap\Omega^\varepsilon} \varpi_{T,m} q = 0, \\
\displaystyle\sum_{E\in\mathcal{F}(T)} \sum_{j=1}^{s} \mu_{E,j} \int_{E\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{\omega_{E,j}} = 0, \\
\displaystyle\sum_{l=1}^{r} \mu_{T,l}^v \int_{T\cap\Omega^\varepsilon} \boldsymbol{\Psi_{T,k}}\cdot\boldsymbol{\varphi_{T,l}} = \mu_{T,k}^v, \\
\displaystyle\sum_{m=1}^{t} \mu_{T,m}^q \int_{T\cap\Omega^\varepsilon} \pi_{T,k}\varpi_{T,m} = 0.
\end{cases}
\tag{12.18}
$$

12.3 The Oseen coarse scale problem

12.3.1 The coarse scale problem

The discrete variational formulation of the Oseen problem (12.1) is: find $(\boldsymbol{u}_H, p_H) \in X_H^{Os}$ such that

$$
\begin{aligned}
c_H^{Os}((\boldsymbol{u_H}, p_H), (\boldsymbol{v}, q)) = &\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} (\nu\boldsymbol{\nabla u_H}\cdot\boldsymbol{\nabla v} + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_H}\cdot\boldsymbol{v} \\
&+ \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} (-p_H \operatorname{div}\boldsymbol{v} - q\operatorname{div}\boldsymbol{u_H}),
\end{aligned}
$$

$\forall (\boldsymbol{v}, q) \in X_H^{Os}$ or $X_H^{Os,ad}$ repectively if we consider a Galerkin or a Petrov–Galerkin formulation.

Theorem 12.6 implies that p_H can be decomposed as $p_H = \pi_H(\boldsymbol{u_H}) + \overline{p}_H$ with $\pi_H(\boldsymbol{u_H}) \in M_H^0$ and $\overline{p}_H \in M_H$. It is easy to verify that $\displaystyle\int_{T\cap\Omega^\varepsilon} \pi_H(\boldsymbol{u_H}) \operatorname{div}\boldsymbol{v} = 0$ for all $\boldsymbol{u_H}, \boldsymbol{v} \in V_H$. The problem above can be reformulated as: find $\boldsymbol{u_H} \in V_H$ and $\overline{p}_H \in M_H$ such that

$$
\begin{aligned}
a_H^{Os}(\boldsymbol{u_H}, \boldsymbol{v}) + b_H(\overline{p}_H, \boldsymbol{v}) &= F_H(\boldsymbol{v}) \qquad &\forall \boldsymbol{v} \in W_H, \\
b_H(q, \boldsymbol{u_H}) &= 0 \qquad &\forall q \in M_H,
\end{aligned}
\tag{12.19}
$$

where $V_H = V_H^{Os}$, and $W_H = V_H$ if we consider a Galerkin formulation or $W_H = V_H^{Os,ad}$ if we

consider a Petrov–Galerkin formulation, and

$$a_H(\boldsymbol{u_H}, \boldsymbol{v}) = \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \nu\boldsymbol{\nabla u_H}\cdot\boldsymbol{\nabla v} + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_H}\cdot\boldsymbol{v},$$
$$b_H(\overline{p}_H, \boldsymbol{v}) = -\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \overline{p}_H \operatorname{div}\boldsymbol{v},$$
$$F_H(\boldsymbol{v}) = \sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} \boldsymbol{f}\cdot\boldsymbol{v}.$$

Theorem 12.15. The space V_H^{Os} and the space M_H have the following relation

$$\operatorname{div}_H V_H^{Os} = M_H.$$

Proof. The proof of Theorem 12.15 is similar to [79, Theorem 3.16]. □

Making use of Theorem 12.15, we deduce from (12.19) that $\operatorname{div}\boldsymbol{u_H} = 0$ in $T\cap\Omega^\varepsilon$ for $T\in\mathcal{T}_H$. We can therefore eliminate the pressure unknown from (12.19) by introducing a subspace of V_H^{Os}, defined by

$$Z_H^{Os} = \{\boldsymbol{v}\in V_H^{Os} \text{ such that } \operatorname{div}_H\boldsymbol{v} = 0,\ \forall T\in\mathcal{T}_H\}.$$

Therefore (12.19) is equivalent to find $\boldsymbol{u_H}\in Z_H^{Os}$ such that

$$\sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon} \boldsymbol{\nabla u_H}\cdot\boldsymbol{\nabla v} + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_H}\cdot\boldsymbol{v} = \sum_{T\in\mathcal{T}_H}\int_{T\cap\Omega^\varepsilon}\boldsymbol{f}\cdot\boldsymbol{v} \qquad \forall\boldsymbol{v}\in Z_H^{Os}.$$

The existence and uniqueness of a solution $\boldsymbol{u_H}$ are guaranteed by Theorem 3.2. Then, the existence and uniqueness of pressure $\overline{p}_H$ follows from the fact that $\operatorname{div}_H V_H^{Os} = M_H$. As a result we have proved that (12.19) has one and only one solution $(\boldsymbol{u_H}, p_H)\in V_H^{Os}\times M_H$.

12.3.2 Reconstruction of the fine scale

After obtaining coarse solutions $\boldsymbol{u_H} = (u_{T,1},\dots,u_{T,r})_{T\in\mathcal{T}_H}\cup(u_{F,1},\dots,u_{F,s})_{F\in\mathcal{F}_H}$ and $\overline{p}_H = (\overline{p}_H|_T)_{F\in\mathcal{T}_H}$, we reconstruct on any coarse element $T\in\mathcal{T}_H$ fine scale solutions

$$\boldsymbol{u_H}|_T = \sum_{F\in\mathcal{F}(T)}\sum_{i=1}^{s} u_{F,i}\boldsymbol{\phi_{F,i}} + \sum_{i=1}^{r} u_{T,k}\boldsymbol{\psi_{T,k}},$$
$$p_H|_T = \sum_{F\in\mathcal{F}(T)}\sum_{i=1}^{s} u_{F,i}\pi_{F,i} + \sum_{i=1}^{r} u_{T,k}\pi_{T,k} + \overline{p}_H.$$

Part V

Numerical Results

Chapter 13

Settings of the Multi-scale Finite Element Methods

Contents

13.1 MsFEMs implemented

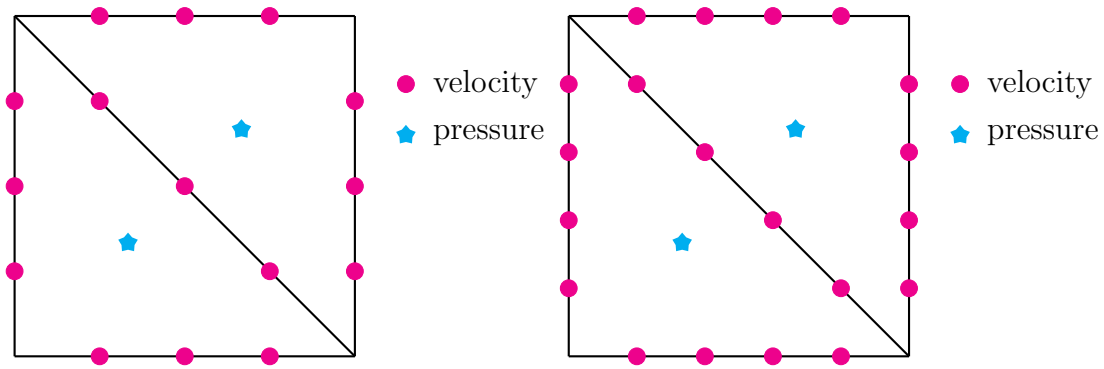
We implement the high-order MsFEM for the Stokes problem (see Chapter 7) and for the Oseen problem (see Chapter 12) in FreeFEM [93] for the cases $n = 1$ and $n = 2$ in two and three dimensions. For the sake of comparison, we also implement previous MsFEMs developed in the literature: the MsFEM called "CR3" (introduced in [111]) and the MsFEM called "CR4" (introduced in [78]), where "CR" stands for "Crouzeix–Raviart", for the Stokes and the Oseen problems in two dimensions. In two dimensions, the CR3 method is obtained by considering the following weighting functions on faces of the coarse elements,

$$\boldsymbol{\omega_{F,1}} = \boldsymbol{e_1},\ \boldsymbol{\omega_{F,2}} = \boldsymbol{e_2},\ \boldsymbol{\omega_{F,3}} = \boldsymbol{n_F}\psi_F$$

where ψ_F is a non-vanishing function in $\mathbb{P}_1(F)$ with vanishing mean-value on $F \cap \Omega^\varepsilon$, $(\boldsymbol{e_1}, \boldsymbol{e_2})$ is the canonical basis of $\mathbb{R}^2$ and $\boldsymbol{n_F}$ is a normal vector to F. In two dimensions, the CR4 method is obtained by enriching the CR3 method to reach the full $\mathbb{P}_1(F)^2$ weighting space on the faces, i.e.

$$\boldsymbol{\omega_{F,1}} = \boldsymbol{e_1},\ \boldsymbol{\omega_{F,2}} = \boldsymbol{e_2},\ \boldsymbol{\omega_{F,3}} = \boldsymbol{n_F}\psi_F,\ \boldsymbol{\omega_{F,4}} = \boldsymbol{t_F}\phi_F,$$

where ϕ_F is a non-vanishing function in $\mathbb{P}_1(F)$ with vanishing mean-value on $F \cap \Omega^\varepsilon$ and $\boldsymbol{t_F}$ a tangential vector to F. In these two cited methods, the pressure weighting functions are limited to constant polynomial, i.e. $\varpi_{T,1} = 1$. No weighting functions are considered in the elements contrary to the high-order MsFEM presented in this thesis. For the sake of understanding, Figures 13.1a and 13.1b illustrate the degrees of freedom for the CR3 and CR4 MsFEMs. The velocity degrees of freedom are represented in pink and pressure degrees of freedom are represented in blue. The main difference with high-order MsFEMs developed in this thesis is the absence of velocity degrees of freedom in the elements. It should be noted that the figures presented here are only schematic representations, since we recall that the degrees of freedom for the velocity are not nodal, but defined by moments.



(a) MsFEM CR3. (b) MsFEM CR4.

Figure 13.1: Scheme of the degrees of freedom in two-dimensions for the CR3 and CR4 MsFEMs.

Increasing the MsFEM order imply growing costs which have to be compared to guide a choice between the different approaches. Let us consider a coarse mesh made up of n_T coarse elements (triangles in two dimensions and tetrahedra in three dimensions) with n_F faces. Then, Tables 13.1 and 13.2 give the number of unknowns for the different methods. For the sake of comparison, we use the following asymptotic relations: $n_F \approx \frac{3}{2} n_T$ in two dimensions and $n_F \approx 2n_T$ in three dimensions.

Two-dimensional MsFEM				
Field	MsFEM in Lit.		High-order MsFEM	
	CR3	CR4	$n = 1$	$n = 2$
Pressure	n_T	n_T	$3n_T$	$6n_T$
Velocity	$3n_F$	$4n_F$	$2n_T$+$4n_F$	$6n_T$+$6n_F$
Total	n_T+$3n_F$	n_T+$4n_F$	$5n_T$+$4n_F$	$12n_T$+$6n_F$
Asymp. Total	$5.5n_T$	$7n_T$	$11n_T$	$21n_T$

Table 13.1: Number of unknowns in the different MsFEMs in two dimensions.

Three-dimensional MsFEM				
Field	MsFEM in Lit.		High-order MsFEM	
	CR3	CR4	$n=1$	$n=2$
Pressure	n_T	n_T	$4n_T$	$10n_T$
Velocity	$3n_F$	$9n_F$	$3n_T$+$9n_F$	$12n_T$+$18n_F$
Total	n_T+$3n_F$	n_T+$9n_F$	$7n_T$+$9n_F$	$22n_T$+$18n_F$
Asymp. Total	$7n_T$	$19n_T$	$25n_T$	$58n_T$

Table 13.2: Number of unknowns in the different MsFEMs in three dimensions.

13.2 Details of computation

Unknows in local problems of CR3 and CR4 MsFEMs are discretized with the pair of finite elements $\mathbb{P}_2-\mathbb{P}_1$. Unknowns in the local problems of the two-dimensional high-order MsFEM, presented in this thesis, are discretized respectively for $n=1$ and $n=2$ with the pairs of finite elements $P2pnc-\mathbb{P}_1^{dc}$ and $P3pnc-\mathbb{P}_2^{dc}$ (presented in Chapter 4). Unknowns in the local problems of the three-dimensional high-order MsFEM are discretized respectively for $n=1$ and $n=2$ with the pairs of finite elements $P2pnc3d-\mathbb{P}_1^{dc}$ and $P3pnc3d-\mathbb{P}_2^{dc}$ (presented in Chapter 5). To compute the local basis functions, no-slip boundary conditions are imposed on the obstacles.

To solve the coarse-scale problem on the coarse mesh, we approximate the non-homogeneous Dirichlet boundary conditions $\boldsymbol{u_H}=\boldsymbol{g}$ on $\partial\Omega$ in a weak form, i.e.

$$\int_{F\cap\partial\Omega}\boldsymbol{u_H}\cdot\boldsymbol{\omega_{F,j}}=\int_{F\cap\partial\Omega}\boldsymbol{g}\cdot\boldsymbol{\omega_{F,j}},\ \forall F\in\mathcal{F}_H\cap\partial\Omega,\ j=1,\dots,s. \tag{13.1}$$

Numerically, the conditions (13.1) are imposed using the penalization method. The readers may refer to Appendix E for details about the computation of the local and coarse problems.

For the sake of comparison, we compute reference solutions, for the Stokes and the Oseen problems, in two dimensions on fine meshes, called "reference meshes", as shown in Figure 1.5. The reference solutions are solved with the pair of finite element $\mathbb{P}_2-\mathbb{P}_1$ (see Appendix D for details about the computation of the reference solutions).

The different meshes used in this thesis are computed with the platform SALOME (an open source platform for numerical simulation) [4] in the format `.med`. These meshes are read in FreeFEM thanks to the module medio [25] (a library that enables input and output of mesh files for FreeFEM in `.med` format).

All the numerical results (MsFEM approximations and reference solutions) presented in this chapter are computed on Topaze a cluster from the CCRT. It is composed of AMD EPYC 7763 CPUs, clocked at 2.45 GHz.

We note $\boldsymbol{u_H}$ and p_H, respectively the MsFEM approximations of the velocity and pressure. We note $\boldsymbol{u}_{\text{ref}}$ and p_{ref}, respectively the reference velocity and pressure. In order to compare the MsFEM approximations and the reference solutions, numerically, we define the relative error in velocity in L^2-norm

$$L^2\ Rel.\ U=\frac{\|\boldsymbol{u_H}-\boldsymbol{u}_{\text{ref}}\|_{L^2}}{\|\boldsymbol{u}_{\text{ref}}\|_{L^2}},$$

the relative error in velocity in H^1-norm

$$H^1 \ Rel.\ U = \frac{|\boldsymbol{u_H} - \boldsymbol{u}_{\text{ref}}|_{H^1}}{|\boldsymbol{u}_{\text{ref}}|_{H^1}},$$

and the relative error in pressure in L^2-norm

$$L^2 \ Rel.\ P = \frac{\|\boldsymbol{p}_H - \boldsymbol{p}_{\text{ref}}\|_{L^2}}{\|\boldsymbol{p}_{\text{ref}}\|_{L^2}},$$

where we recall that the L^2-norm and the H^1-norm are defined respectively by

$$\|\boldsymbol{u}\|_{L^2} = \sqrt{\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} (\boldsymbol{u}|_T)^2},$$

and

$$|\boldsymbol{u}|_{H^1} = \sqrt{\sum_{T\in\mathcal{T}_H} \int_{T\cap\Omega^\varepsilon} (\boldsymbol{\nabla u}|_T)^2}.$$

13.3 Simulations in two-dimensional domains

13.3.1 Settings

In all following two-dimensional simulations, we fix the element size h of local fine meshes and vary the element size H of coarse meshes. We choose $h = 3.10^{-4}$ (the fine mesh contains more than 21 millions of fine elements, see for example Table 13.3), which ensures that it is small enough to well capture small obstacles in the domain Ω^ε, i.e. that $h \ll \varepsilon < H$ (this choice of h corresponds to $\varepsilon_{min}/25$ with $\varepsilon_{min} = 1/128$, the smallest ε considered). Besides, this choice ensures that the Péclet number is smaller than 1 for the computation of the local problems involved in the MsFEM for the Oseen problem, and the computation of the reference solutions for the Oseen problems, preventing thus oscillations to appear (see Chapter 11).

13.3.2 Test case: the channel flow

We consider a two-dimensional open channel $\Omega = [0 \leq x \leq 1, 0 \leq y \leq 1]$, where the heterogeneities are represented by solid obstacles. We assign the source term $\boldsymbol{f} = \boldsymbol{0}$. The parabolic inflow boundary condition $\boldsymbol{u} = y(1-y)\boldsymbol{e_1}$ is imposed at the inlet. The Neumann boundary condition $\nu\boldsymbol{\nabla u n} - p\boldsymbol{n} = \boldsymbol{0}$ is imposed at the outlet and the no-slip condition is applied on other boundaries. An example of computational domain is given in Figure 13.2 for the periodic cases or in Figure 13.3a for the random cases.

13.3.3 Convergence analysis in a two-dimensional periodic case

We perform a convergence analysis in the periodic case. We choose a periodic arrangement, of period ε, of circular obstacles (whose diameter is set to $\varepsilon/2$) in the domain Ω. We keep a region free of obstacles of size 1/8 both upstream and downstream of the arrangement of obstacles so that the flow develops. We consider different periods ε, from $\varepsilon = 1/8$ to $\varepsilon = 1/128$. Figure 13.2 illustrates the computational domain obtained for $\varepsilon = 1/16$. We fix the fine mesh size h and

consider different coarse mesh sizes H, from $H = 1/8$ to $H = 1/64$, which leads from 256 to 16384 coarse elements. We are interested only in the case in which $H \geq \varepsilon$, since the opposite case $H < \varepsilon$ is covered by classical finite element methods and MsFEMs are not needed. We compare the MsFEM approximations in the different configurations to reference solutions. We compute the relative errors for velocity in L^2-norm and H^1-norm and for pressure in L^2-norm between the MsFEM approximations and the reference solutions.

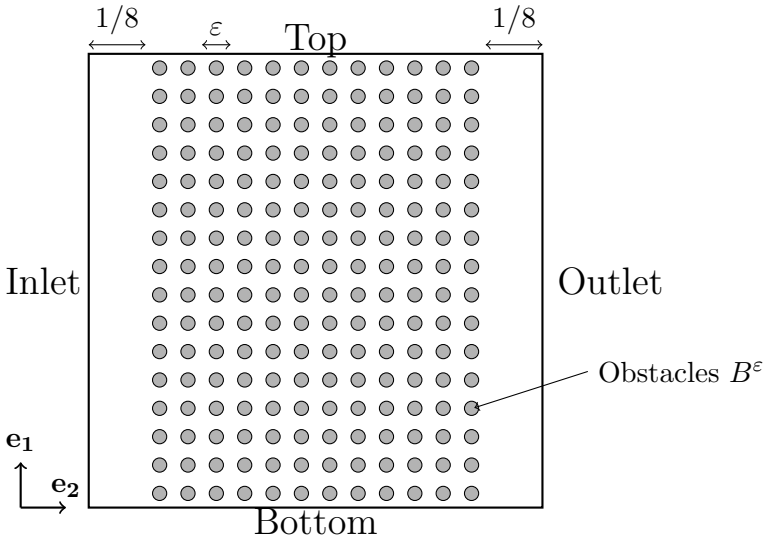


Figure 13.2: Scheme of the computational domain Ω^ε for $\varepsilon = 1/16$.

It should be noted that in most industrial cases, such as in nuclear reactor cores, the geometries considered are periodic.

13.3.4 Study of the number of unknowns

Mesh ID	ε	number of elements in ref. mesh	H	number of elements in coarse mesh	average number of fine elements in a coarse element
#1	1/64	21 millions	1/8	256	83 590
#2			1/16	1 024	20 950
#3			1/32	4 096	5 260
#4			1/64	16 384	1 300

Table 13.3: Characteristics of perforated meshes for $\varepsilon = 1/64$ in two dimensions.

		Mesh ID			
		#1	#2	#3	#4
number of unknowns in coarse problems	$n = 1$	2 880	11 392	45 312	180 736
	$n = 2$	5 472	21 696	86 400	344 832
number of unknowns in reference problems		$\approx$ 100 millions			

Table 13.4: Number of unknowns in the coarse problems and reference problem in two dimensions.

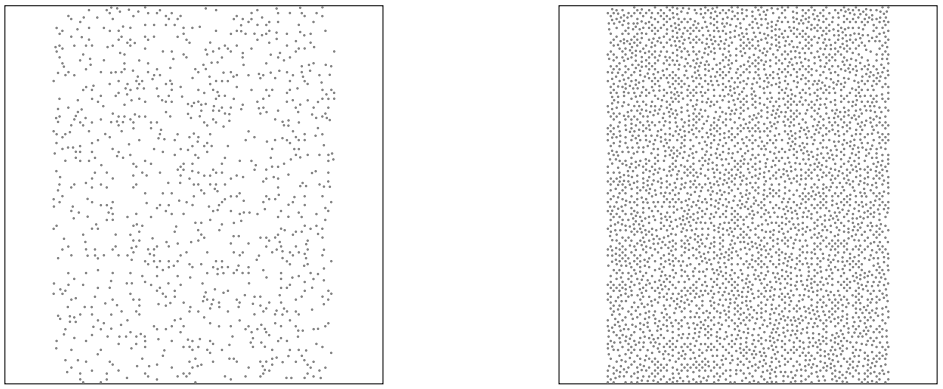
		Mesh ID			
		#1	#2	#3	#4
number of unknowns in local problems high-order MsFEM	$n = 1$	920 000	230 000	60 000	15 000
	$n = 2$	1 800 000	440 000	110 000	28 000

Table 13.5: Number of unknowns in the local problems in two dimensions.

Tables 13.3 and 13.4 illustrate the interest of MsFEM. Indeed, it can be seen that the number of unknowns in the coarse problems may be up to 30 000 times smaller than the number of unknowns in the reference problems. Table 13.5 shows the number of unknowns in the local problems. These tables illustrate the trade-off to find for MsFEMs. The greater the simplicity of the coarse problem (fewer unknowns), the more challenging the local problems become, due to the increased number of unknowns. For example, the local problems defined on the mesh #1 are computing by allocating 16 CPUs to each local process due to the high-number of unknowns. For the other meshes (#2, #3, #4), 1 CPU is allocated to each local process.

13.3.5 Two-dimensional domains with randomly placed obstacles

To show the robustness of the MsFEM, in addition to periodic domains, we consider random perforated domains which consists in random distributions of obstacles. Figures 13.3a and 13.3b show random perforated domains with respectively 1024 and 4096 obstacles of diameter 1/256.



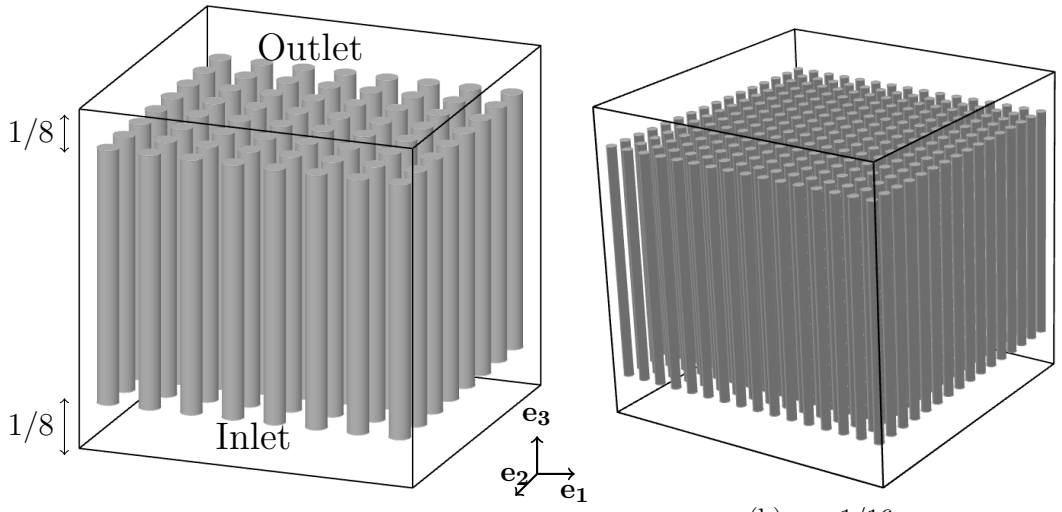
(a) 1024 obstacles. (b) 4096 obstacles.

Figure 13.3: Domain with randomly located obstacles.

Remark 13.1. When considering randomly located obstacles, especially when obstacles are densely packed, several meshing problems may appear. Indeed, the first one is the fact that if obstacles are too close of corners of coarse triangles, then, they can cut coarse triangles into two non-connected elements, which makes the method failed. The second one is the fact that obstacles may be tangent to interfaces between coarse elements. In this case, meshing tools may fail to generate meshes due to isolated tangent points. In addition, even if the meshing tools do not fail, there will not be enough fine elements between obstacles and interfaces for computation of the local problems. A solution to overcome these difficulties may be to allow general coarse element shapes or to compute first the fine mesh on the whole perforated domain and afterwards, to cut this fine mesh (possibly with some modifications) into polygonal coarse elements avoiding symptomatic cases.

13.4 Simulations in three-dimensional domains

We consider a three-dimensional open channel $\Omega = [0,1]^3$, perforated by a periodic arrangement, of period ε, of cylinders (whose diameter is set to $\varepsilon/2$). We keep a free region, of size 1/8, downstream of the inlet and upstream of the outlet so that the flow develops. Figures 13.4a and 13.4b present schemes of this domain respectively for $\varepsilon = 1/8$ and $\varepsilon = 1/16$. It should be noted that the domain for $\varepsilon = 1/16$ is closer to reality since we recall that in a reactor pressure vessel, the fuel rods are generally bundled in a square array of 14×14 to 17×17. The parabolic inflow boundary condition $\boldsymbol{u} = x(1-x)y(1-y)\boldsymbol{e_3}$ is imposed at the inlet. The flow is oriented in a longitudinal direction with respect to the cylinders. The Neumann boundary condition $\nu \boldsymbol{\nabla u n} - p\boldsymbol{n} = \boldsymbol{0}$ is imposed at the outlet and the no-slip condition is applied on other boundaries. The coarse mesh size is set to $H = 1/8$, which leads to 3072 coarse elements. Each coarse element is discretized with a fine mesh of size $h = 10^{-2}$ for the domain Figure 13.4a and $h = 10^{-3}$ for the domain Figure 13.4b.



(a) $\varepsilon = 1/8$.

(b) $\varepsilon = 1/16$.

Figure 13.4: Scheme of the computational domain Ω^ε in three dimensions.

Below, we give the characteristics of the meshes used.

Mesh ID	ε	number of elements in ref. mesh	number of elements in coarse mesh	average number of fine elements in a coarse element
#5	1/8	11 millions	3 072	3 500
#6	1/16	73 millions	3 072	23 000

Table 13.6: Characteristics of perforated meshes in three dimensions ($H = 1/8$).

		Mesh ID	
		#5	#6
number of unknowns in coarse problems	$n = 1$	80 256	80 256
	$n = 2$	185 088	185 088
number of unknowns in reference problems		$\approx$ 47 millions	$\approx$ 325 millions

Table 13.7: Number of unknowns in the coarse problems and reference problem in three dimensions.

		Mesh ID	
		#5	#6
number of unknowns in local problems high-order MsFEM	$n = 1$	91 000	650 000
	$n = 2$	210 000	1 500 000

Table 13.8: Number of unknowns in the local problems in three dimensions.

Tables 13.6 and 13.7 illustrate the interest of MsFEM. Indeed, it can be seen that the number of unknowns in the coarse problems may be up to 2 300 times smaller than the number of unknowns in the reference problems (if computed with the pair of finite elements $\mathbb{P}_2 - \mathbb{P}_1$. Table 13.8 shows the number of unknowns in the local problems. Due to the small number of degrees of freedom, each local problem can be computed using 1 CPU. However, for finer mesh, such as the mesh #6, which allows to consider more realistic geometries, more than one CPU is needed to compute in parallel each local problems (hence the development of parallelism for the local problems, see Appendix E.1.1).

Chapter 14

Numerical results for MsFEM applied to the Stokes problem

In this chapter, we solve the Stokes problem (7.1) on a coarse grid using basis functions defined by the Stokes problem, i.e. the local problems (8.1) and (8.2). Through this chapter, we set $\nu = 1$. For the two-dimensional simulations, we consider the channel flow test case presented in Section 13.3.2 and for the three-dimensional simulations, we consider the channel flow test case presented in Section 13.4.

14.1 Example of MsFEM local basis functions

To get an idea of the local basis functions, Figure 14.1 shows the 14 basis functions of the high-order MsFEM for $n = 1$ for the Stokes problem, obtained by solving (8.1) and (8.2). The first two correspond to basis functions associated with the element while the remaining correspond to basis functions associated with faces.

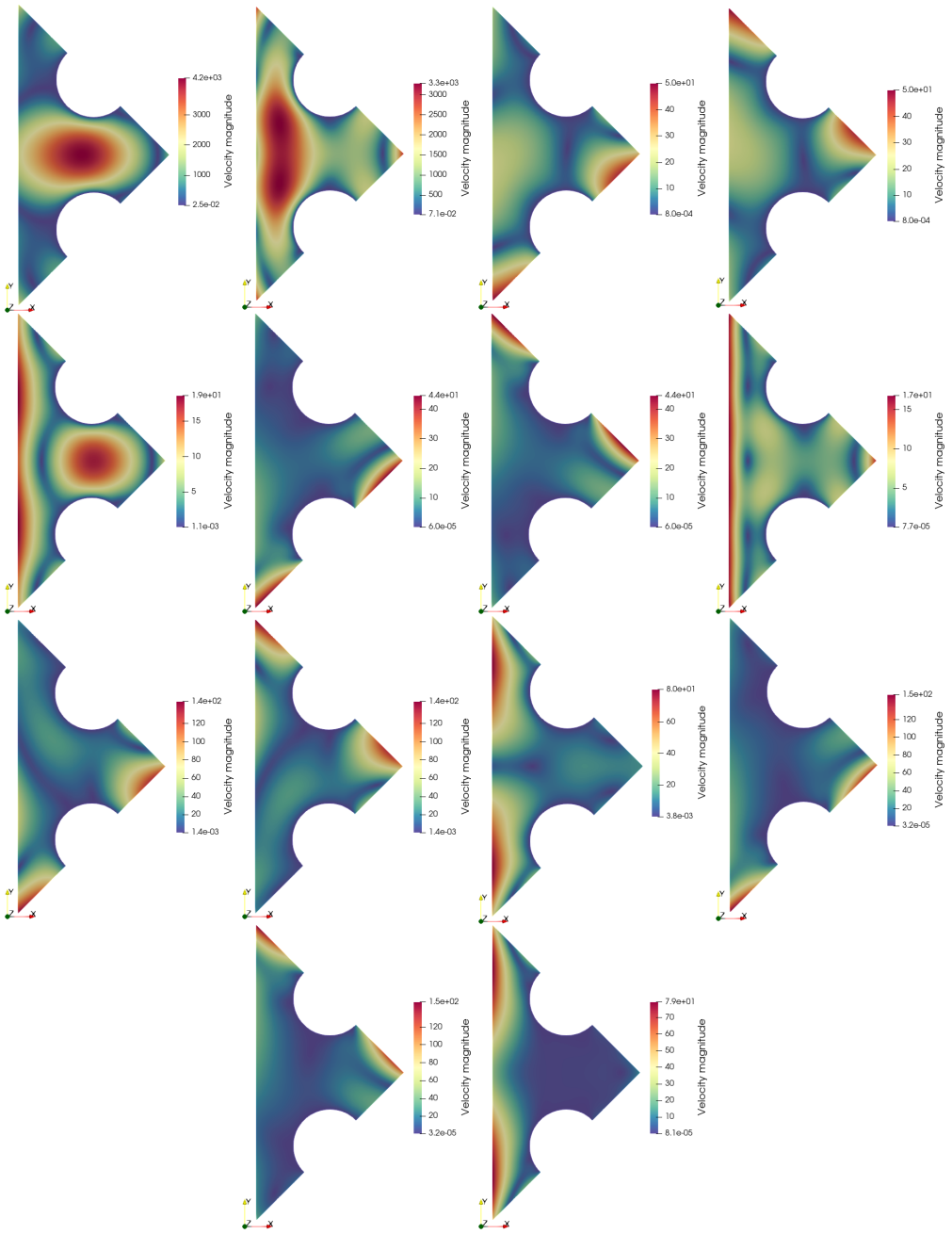


Figure 14.1: The 14 basis functions of the high-order MsFEM for the Stokes problem, $n = 1$.

14.2 Convergence analysis in two-dimensional periodic domains

In this section, we perform the convergence analysis proposed in Section 13.3.3. Figures 14.2 to 14.4, present the computed relative errors between the reference solutions computed on reference meshes and the MsFEM approximations computed on coarse mesh for the cases $n = 1$ and $n = 2$. In these figures, the different periods ε considered are represented by different colors. The high-order MsFEM for $n = 1$ is represented in solid line with square marks, while MsFEM for $n = 2$ is represented by dashed line with circle marks.

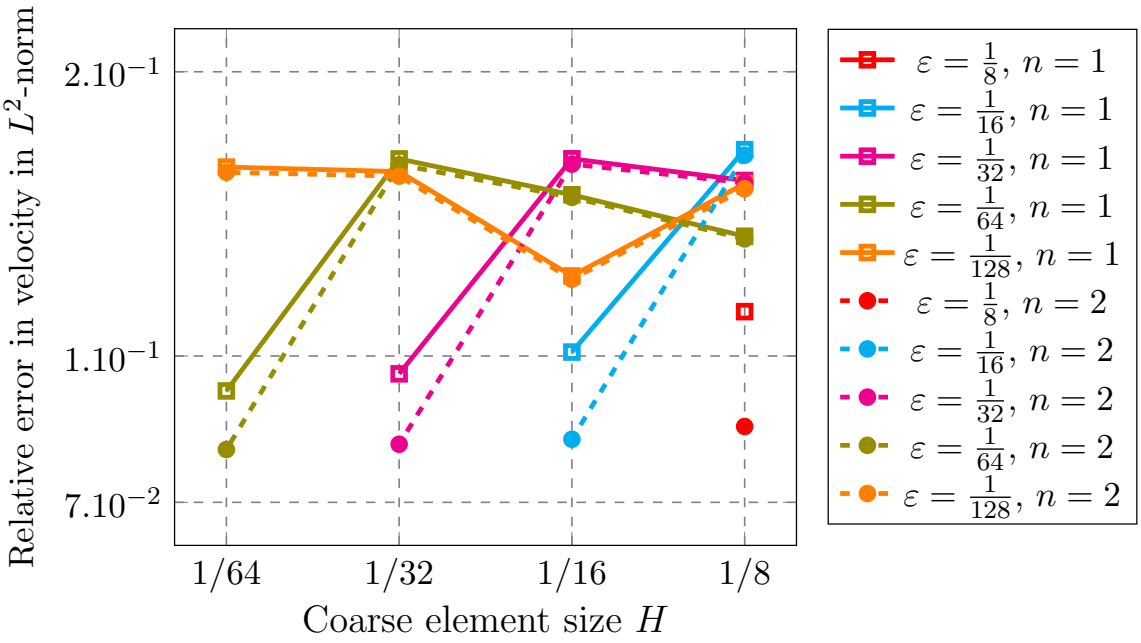


Figure 14.2: Relative errors between MsFEM approximations and reference solutions for velocity in L^2-norm (Stokes flow).

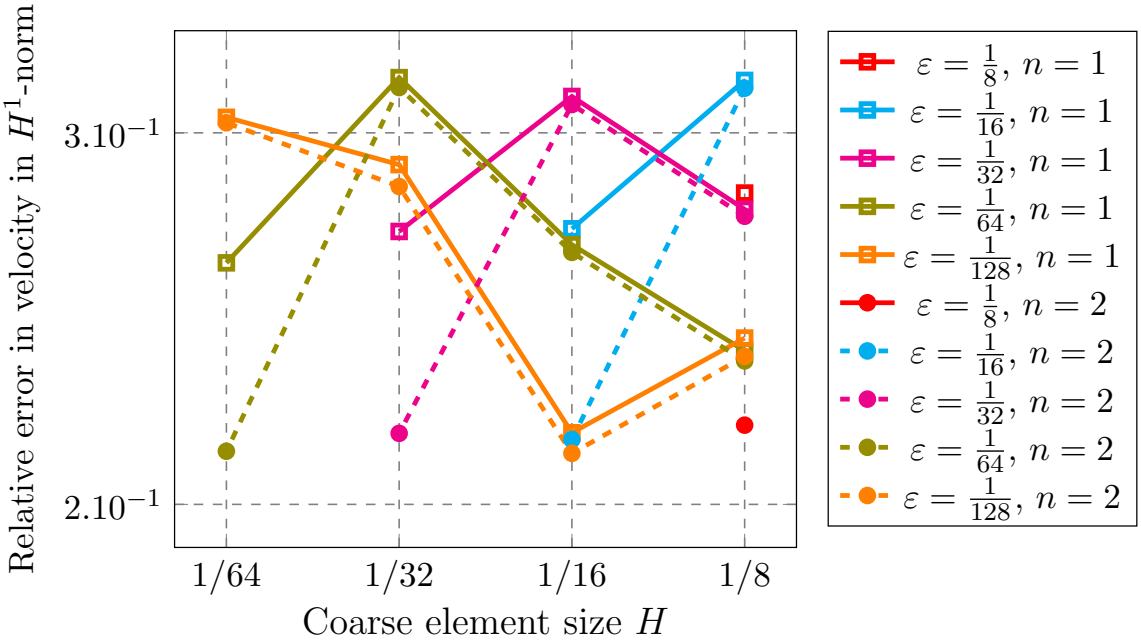


Figure 14.3: Relative errors between MsFEM approximations and reference solutions for velocity in H^1-norm (Stokes flow).

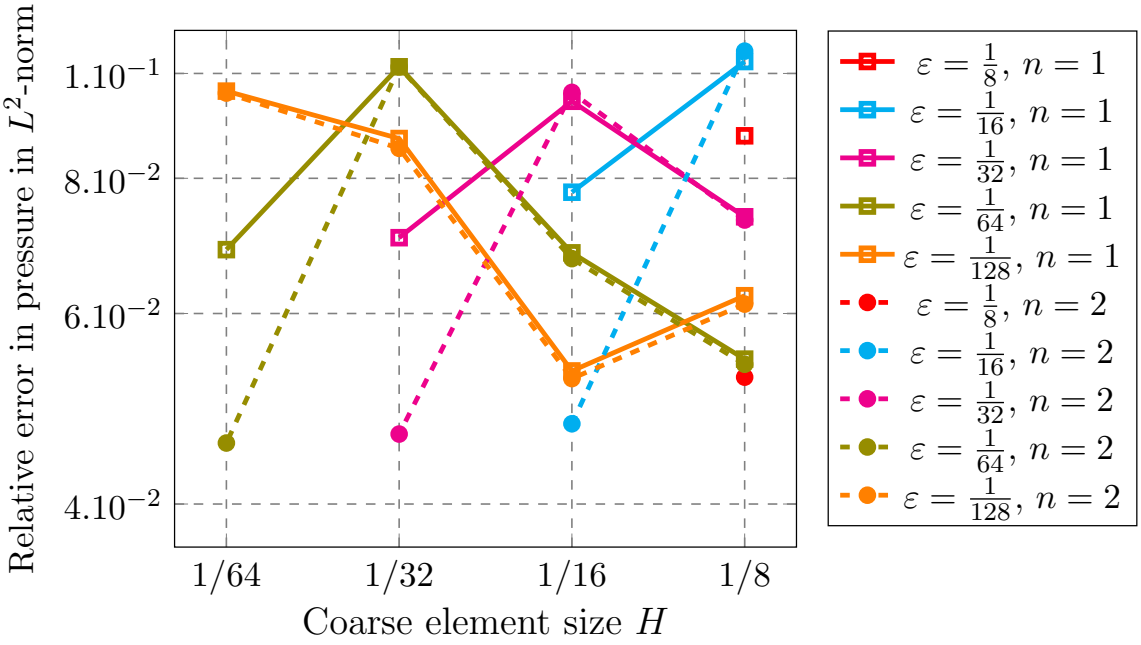


Figure 14.4: Relative errors between MsFEM approximations and reference solutions for pressure in L^2-norm (Stokes flow).

As expected, Figures 14.2 to 14.4 show that the relative error varies non-monotonically with the coarse element size H. Some cases show a decrease in error as H decreases (finer coarse mesh), while others exhibit an increase or fluctuations. However, in all cases, the relative errors are bounded. The relative errors in velocity in L^2-norm vary in a range approximately from 7×10^{-2} to 1.5×10^{-1}. The relative errors in velocity in H^1-norm vary in a range approximately from 2×10^{-1} to 3×10^{-1}. The relative errors in pressure in L^2-norm vary in a range approximately from 4×10^{-2} to 1×10^{-1}. These errors are typical errors for MsFEM approximations. It is difficult to identify the overall behaviour of the relative error. Indeed, for example if considering Figure 14.2, the case $\varepsilon = 1/128$, $n = 1$ (solid orange square) shows a significant fluctuation with the error dropping at $H = 1/16$ and peaking at $H = 1/32$, while the case $\varepsilon = 1/64$, $n = 1$ (solid green square) shows different error behaviors, with the error increasing until $H = 1/32$. In the cases $H \not\approx \varepsilon$, it seems that smaller ε values generally lead to lower relative errors for the same coarse element size H (see for example $H = 1/16$ or $H = 1/32$). To approximate the Stokes problem in the periodic cases, it seems that increasing the MsFEM order do not allow to significantly improve the accuracy of results, except for a particular configuration in which $H \approx \varepsilon$. In this particular case, obstacles cut triangles only at their vertices. However, we should not generalize the results obtained in the periodic settings for the Stokes problem to any more general cases. Indeed, the complex behaviour of the errors suggests the presence of multiple influencing factors, such as the ε, H, n but also the perforation patterns and the way the perforations cut each coarse elements.

Comparisons with previous MsFEMs Figure 14.5 presents similar convergence curves as previously but by including the relative errors obtained with previous MsFEMs, i.e. the CR3 and the CR4 MsFEMs, for $\varepsilon = 1/64$, in order to emphasize the interest of considering high-order

MsFEMs. Additional comparisons can be found in Figures F.1 to F.3 in Appendix for other periods ε.

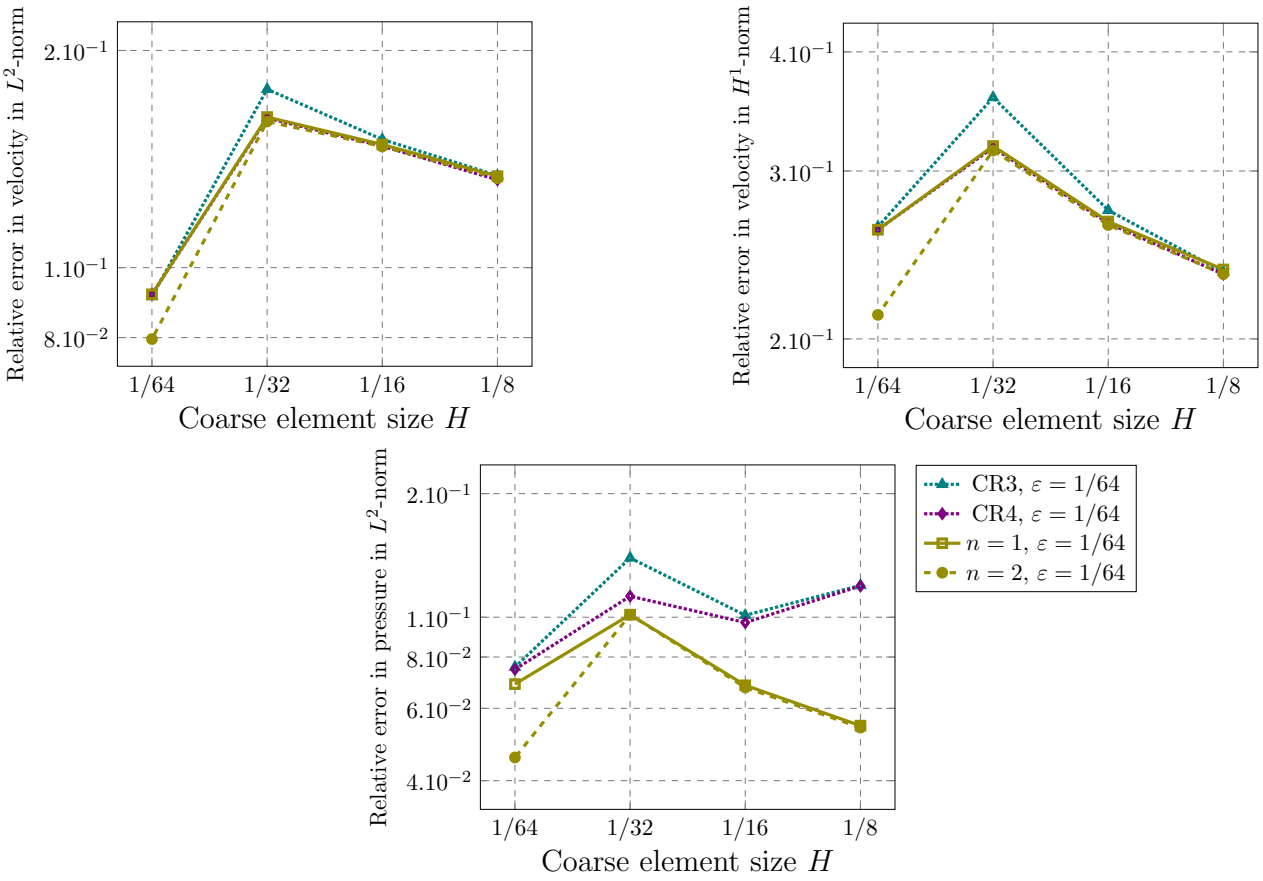
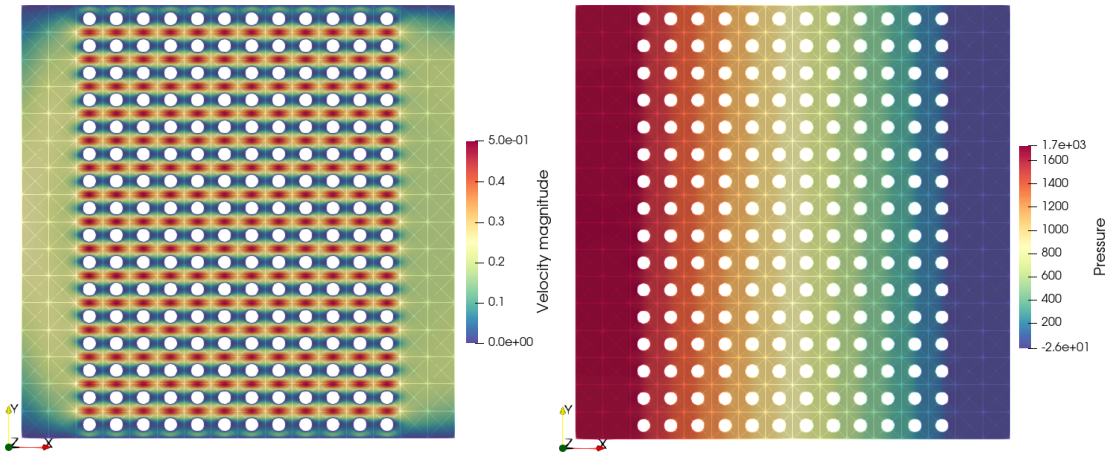


Figure 14.5: Relative errors between MsFEM approximations and the reference solution, $\varepsilon = 1/64$ (Stokes flow).

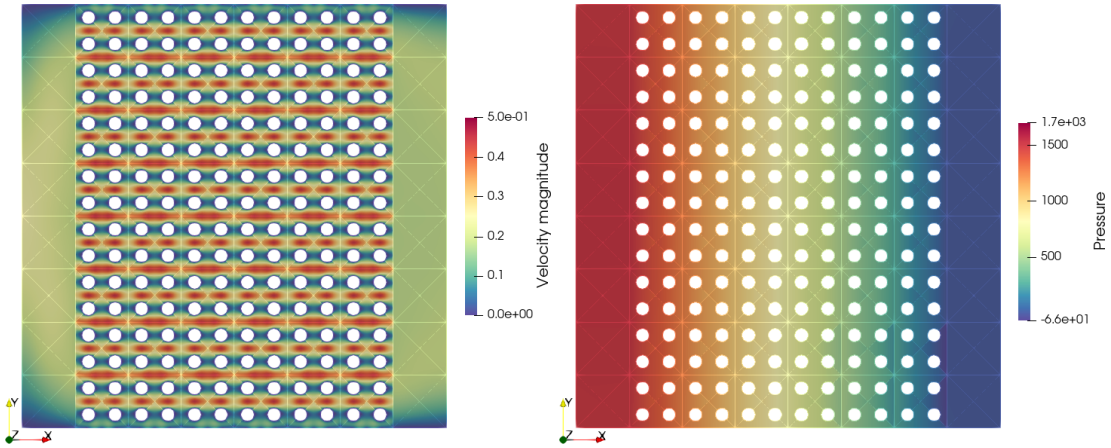
Figure 14.5 illustrates the fact that adding weighting functions in the volume allows to significantly improve mainly pressure approximations. This can be explained by the fact that for high-order MsFEM, the pressure approximation spaces, $\mathbb{P}_1^{dc}$ and $\mathbb{P}_2^{dc}$ respectively for the cases $n = 1$ and $n = 2$, are richer than the $\mathbb{P}_0$ space used to approximate the pressure in previous MsFEMs (the CR3 and CR4 MsFEMs). In contrast, adding weighting functions on the faces seems to improve mainly velocity approximations. Indeed, this can be explained by the fact that the more weighting functions there are on the faces, the more continuity there is between the coarse elements.

Some illustrative figures In what follows, we illustrate these errors with some numerical results. We focus on a periodic domain with $\varepsilon = 1/16$ for the ease of viewing.



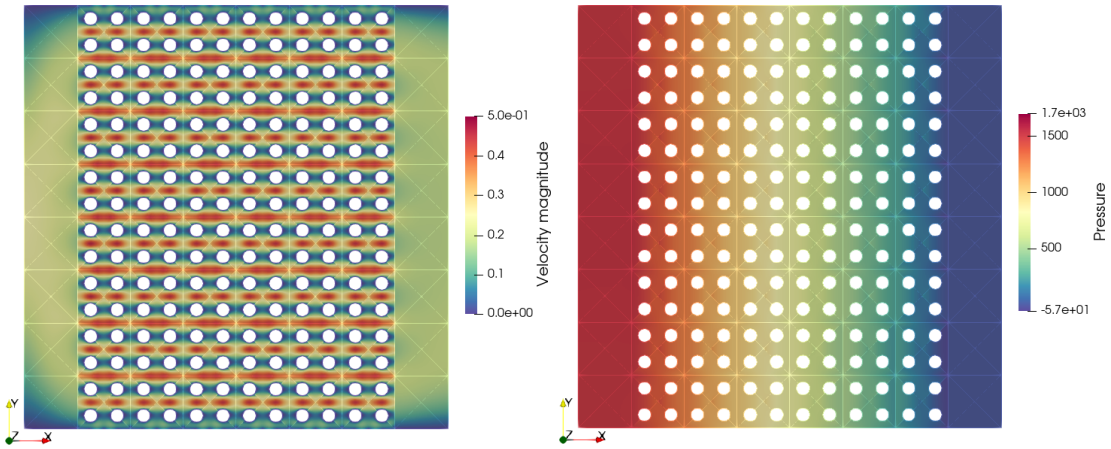
(a) Velocity magnitude. (b) Pressure field.

Figure 14.6: Reference solution (Stokes flow), $\varepsilon = 1/16$.



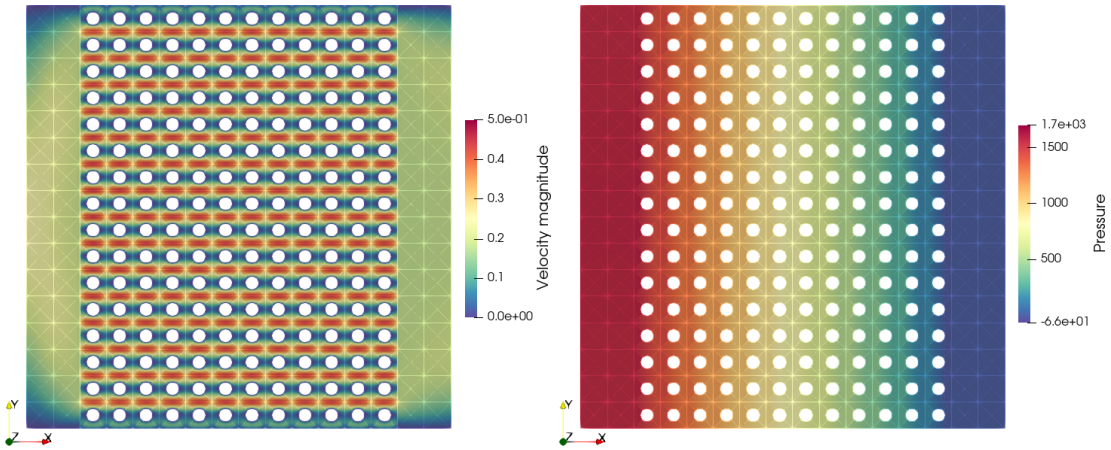
(a) Velocity magnitude. (b) Pressure field.

Figure 14.7: MsFEM approximation (Stokes flow), $n = 1$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/16$.



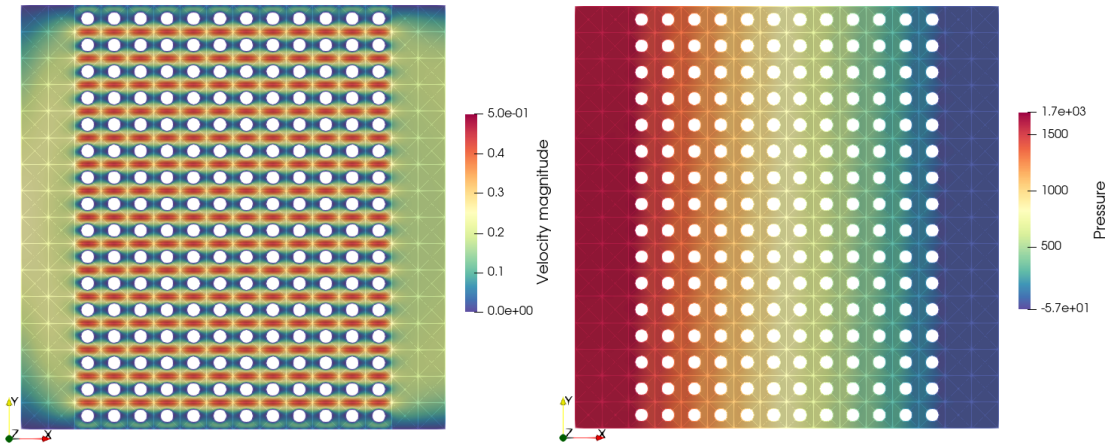
(a) Velocity magnitude. (b) Pressure field.

Figure 14.8: MsFEM approximation (Stokes flow), $n = 2$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/16$.



(a) Velocity magnitude. (b) Pressure field.

Figure 14.9: MsFEM approximation (Stokes flow), $n = 1$, $H = 1/16$ (1024 coarse elements), $\varepsilon = 1/16$.



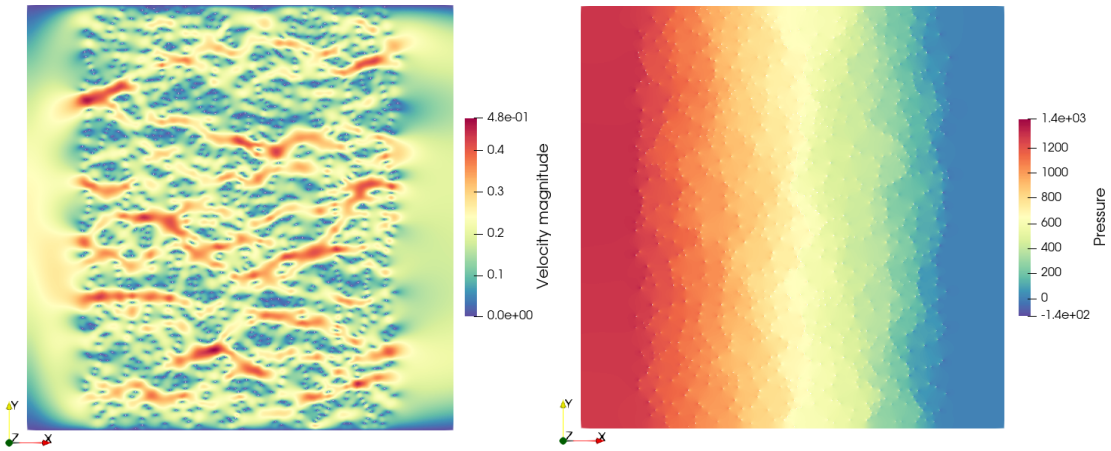
(a) Velocity magnitude. (b) Pressure field.

Figure 14.10: MsFEM approximation (Stokes flow), $n = 2$, $H = 1/16$ (1024 coarse elements), $\varepsilon = 1/16$.

A visual inspection of the solutions reveals that the MsFEM approximations are very similar for $n = 1$ (Figures 14.7 and 14.9) and for $n = 2$ (Figures 14.8 and 14.10). As expected, the utilisation of $H = 1/16$ (Figures 14.9 and 14.10) enables more precise approximations in comparison to the reference solution (Figure 14.6). For the sake of comparison, additional numerical results, for other periods ε, are presented in Appendix F.1.2.

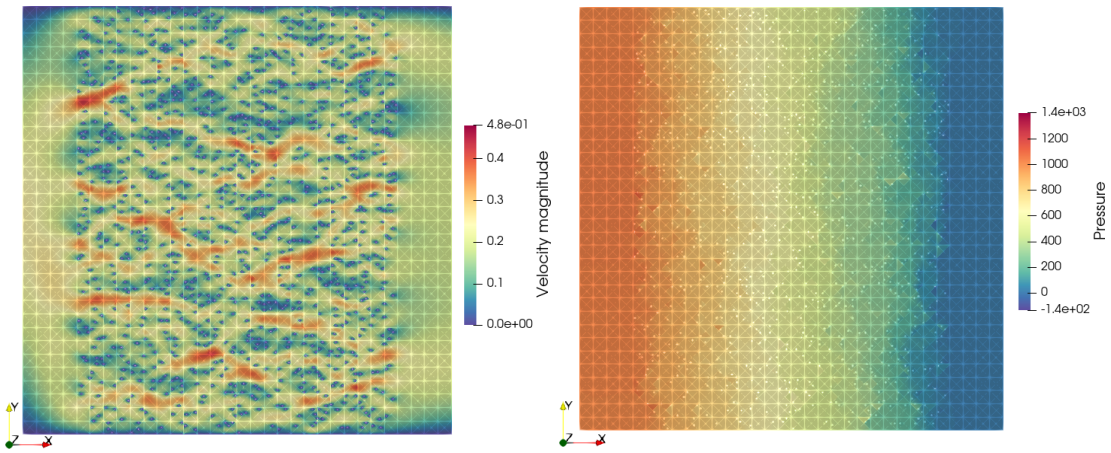
14.3 Simulations in a two-dimensional domain with ramdomly placed obstacles

In this section, we present simulations of the Stokes flows in the domain Figure 13.3a.



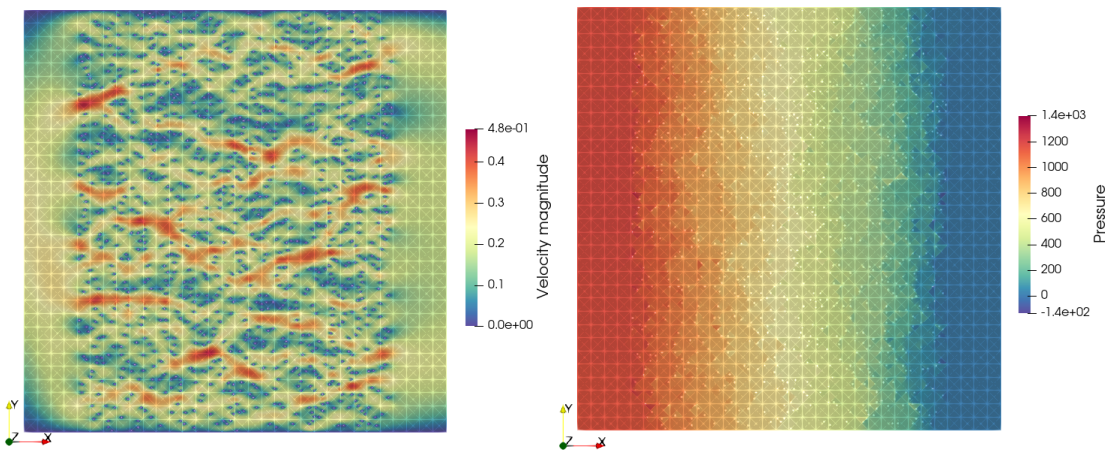
(a) Velocity magnitude. (b) Pressure field.

Figure 14.11: Reference solution (Stokes flow), domain with 1024 randomly placed obstacles.



(a) Velocity magnitude. (b) Pressure field.

Figure 14.12: MsFEM approximation (Stokes flow), $n = 1$, $H = 1/32$ (4096 coarse elements), domain with 1024 randomly placed obstacles.



(a) Velocity magnitude. (b) Pressure field.

Figure 14.13: MsFEM approximation (Stokes flow), $n = 2$, $H = 1/32$ (4096 coarse elements), domain with 1024 randomly placed obstacles.

A visual inspection of the solutions reveals that the MsFEM for $n = 1$ (Figure 14.12) appears to underestimate the velocity magnitude in comparison to the reference solution (Figure 14.11). The MsFEM for $n = 2$ (Figure 14.13) yields a more accurate approximation of pressure. For the sake of comparison, simulations of the Stokes flows in the domain Figure 13.3b are presented in Appendix F.1.2.

14.4 Simulations in a three-dimensional domain

In this section, we show the results of the channel test case presented in Section 13.4 in the domains Figures 13.4a and 13.4b.

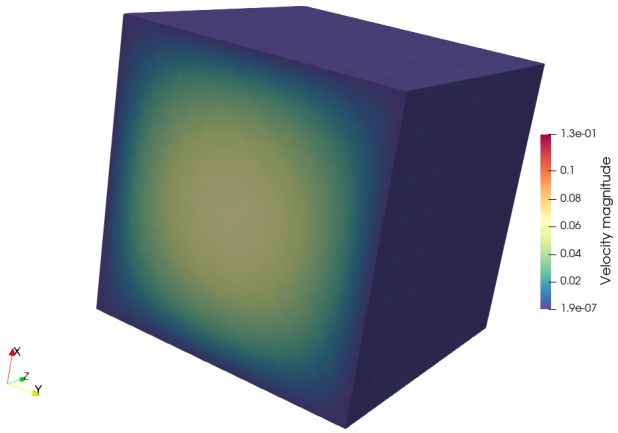
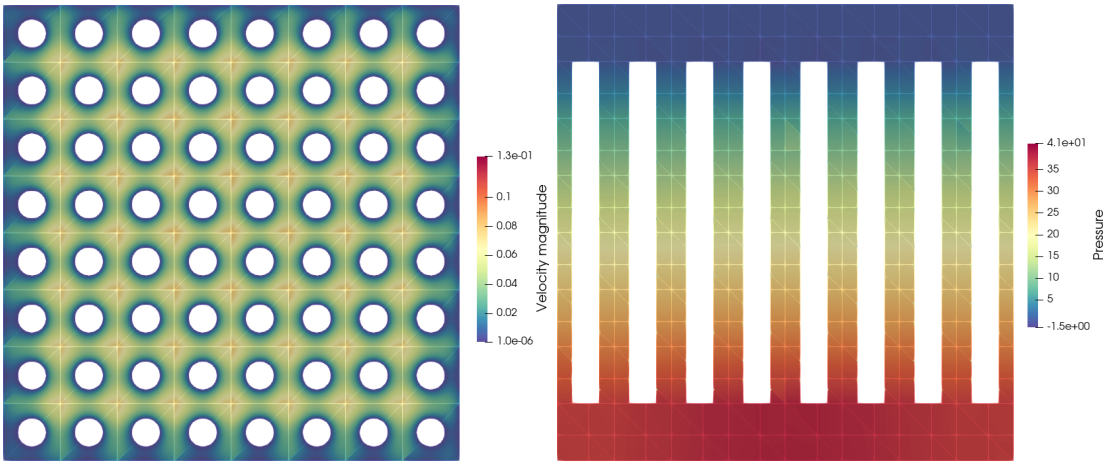
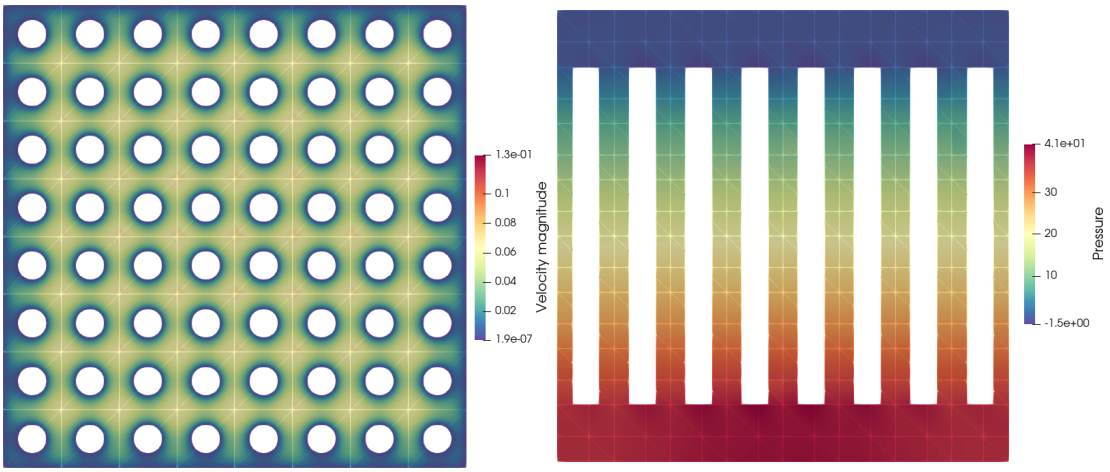


Figure 14.14: MsFEM approximation, $n = 2$, $H = 1/8$ (3072 coarse elements), $\varepsilon = 1/8$.



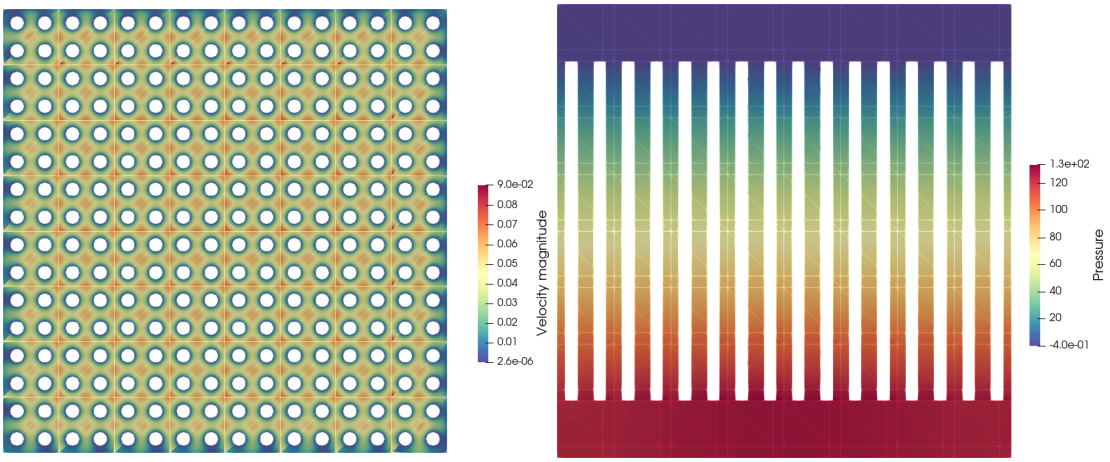
(a) Velocity magnitude (cut along the z axis). (b) Pressure field (cut along the y axis).

Figure 14.15: MsFEM approximation in a three dimensional domain (Stokes flow), $n = 1$, $H = 1/8$ (3072 coarse elements), $\varepsilon = 1/8$.



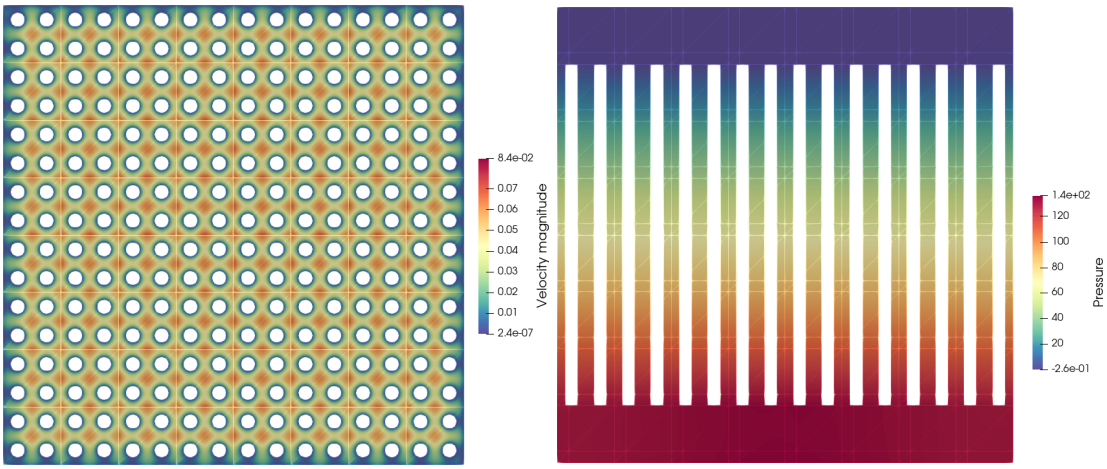
(a) Velocity magnitude (cut along the z axis). (b) Pressure field (cut along the y axis).

Figure 14.16: MsFEM approximation in a three dimensional domain (Stokes flow), $n = 2$, $H = 1/8$ (3072 coarse elements), $\varepsilon = 1/8$.



(a) Velocity magnitude (cut along the z axis). (b) Pressure field (cut along the y axis).

Figure 14.17: MsFEM approximation in a three dimensional domain (Stokes flow), $n = 1$, $H = 1/8$ (3072 coarse elements), $\varepsilon = 1/16$.



(a) Velocity magnitude (cut along the z axis). (b) Pressure field (cut along the y axis).

Figure 14.18: MsFEM approximation in a three dimensional domain (Stokes flow), $n = 2$, $H = 1/8$ (3072 coarse elements), $\varepsilon = 1/16$.

For this test, we do not have a reference solution. However, a visual inspection of Figures 14.15 to 14.18 allows us to be confident in the validity of the solutions. The comparison of the velocity magnitudes, allows to confirm that the MsFEM for $n = 2$ yields more accurate results. Indeed for $n = 1$ the velocity magnitude (Figures 14.15a and 14.17a) is not entirely symmetrical around the cylinders.

Remark 14.1. Three-dimensional simulations present greater challenges than their two-dimensional counterparts. Indeed, the implementation of MsFEM in three dimensions raises two computational issues. The first of these concerns the computation of the global problem, or the so-called "online step". Indeed, if the number of coarse elements (and thus the number of unknowns) in the coarse mesh is too high, the direct computation of the global problem requires a large amount of memory or is even intractable. Using the cluster Topaze, we have been unable to consider more than 3072 coarse elements. On a cluster with more memory per CPU, however, considering more coarse elements could be possible. The other limitation concerns the computations of the local problems. Indeed, if the local fine mesh contains too many elements, then a direct solver is no longer efficient and we reach out of memory. Here also, using a cluster with more memory per CPU or developing a more efficient solver for the local problems in three dimensions could allow for the consideration of finer local meshes. However, this last task is not straightforward (see Appendix E.1.2 for further details).

Chapter 15

Numerical results for MsFEM applied to the Oseen problem

In this chapter, we solve the Oseen problem (12.1) on a coarse grid using basis functions defined by the Oseen problem, i.e. the local problems (12.8) and (12.9), or by the Oseen adjoint problem, i.e. the local problems (12.17) and (12.18). We consider only two-dimensional simulations. We consider the channel flow test case presented in Section 13.3.2. We choose a viscosity $\nu = 5\times10^{-4}$ and the Oseen velocity as $\boldsymbol{U_0} = [y(1-x^2), -x(1-y^2)]^t$, (note that $\text{div}\,(\boldsymbol{U_0}) = 0$) which leads to a Reynolds number of 2000. This Oseen velocity is plotted in Figure 15.1.

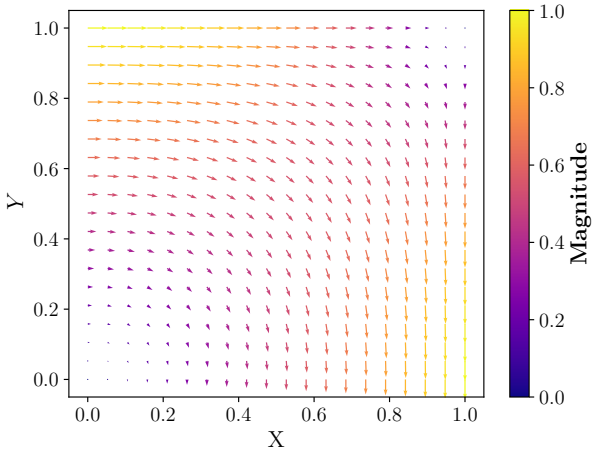


Figure 15.1: Oseen velocity field $\boldsymbol{U_0} = [y(1-x^2), -x(1-y^2)]^t$.

Remark 15.1. Numerical experiments done during this thesis have shown that the Petrov–Galerkin formulation of the Oseen problem leads to similar result as using a Galerkin formulation. We therefore propose not to proceed with this method. In addition, the Petrov–Galerkin method is twice as expensive as the Galerkin method since it requires to compute another local basis for the test functions.

Remark 15.2. In the numerical applications for the Oseen problem, we restrict ourselves to a viscosity of 5×10^{-4} and to an Oseen velocity of magnitude order of 1, due to the restriction on the mesh size imposed by the Péclet number. However, in view of the results, we can imagine that if we consider smaller fine mesh size, then high-order MsFEMs are able to deal with higher Reynolds number.

15.1 Convergence analysis in two-dimensional periodic domains

In this section, we perform the convergence analysis proposed in Section 13.3.3. Figures 15.2 to 15.4, show the relative errors between the reference solutions computed on reference meshes and the MsFEM approximations computed on coarse meshes for the case $n = 1$ and $n = 2$. In these figures, the different periods ε considered are represented by different colors. The high-order MsFEM for $n = 1$ is represented in solid line with square marks, while MsFEM for $n = 2$ is represented by dashed line with circle marks.

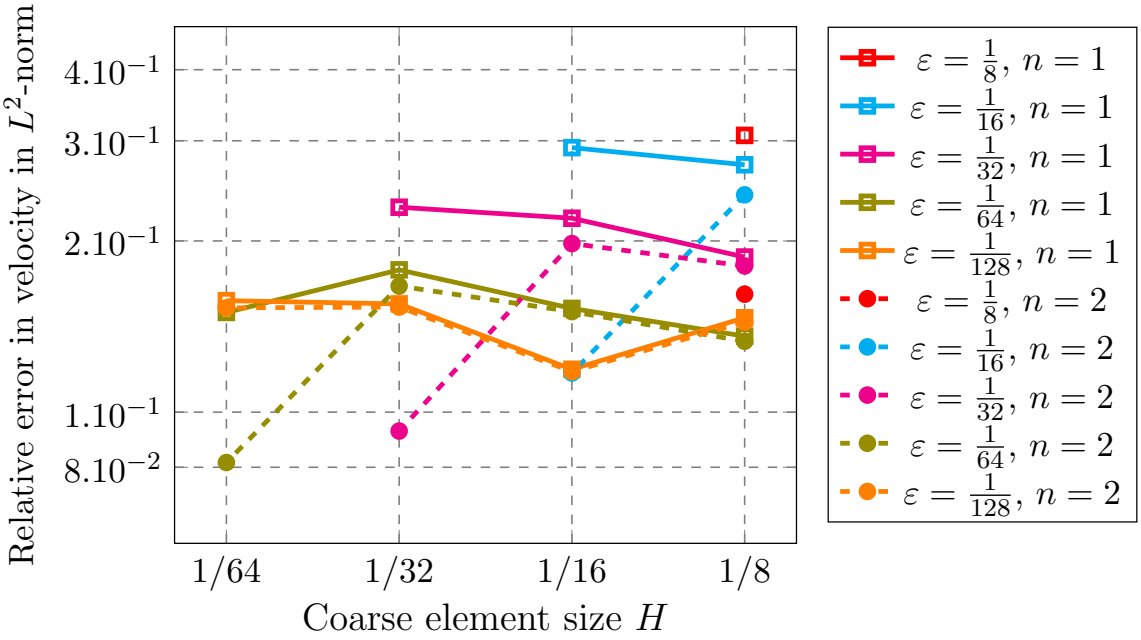


Figure 15.2: Relative errors between MsFEM approximations and reference solutions for velocity in L^2-norm (Oseen flow).

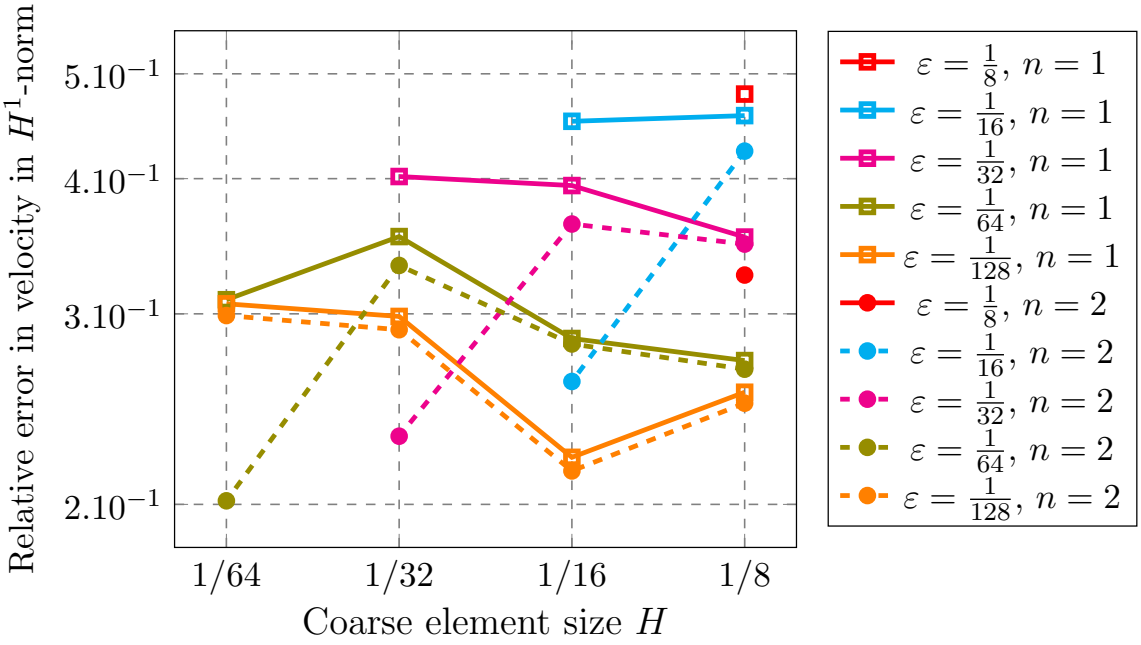


Figure 15.3: Relative errors between MsFEM approximations and reference solutions for velocity in H^1-norm (Oseen flow).

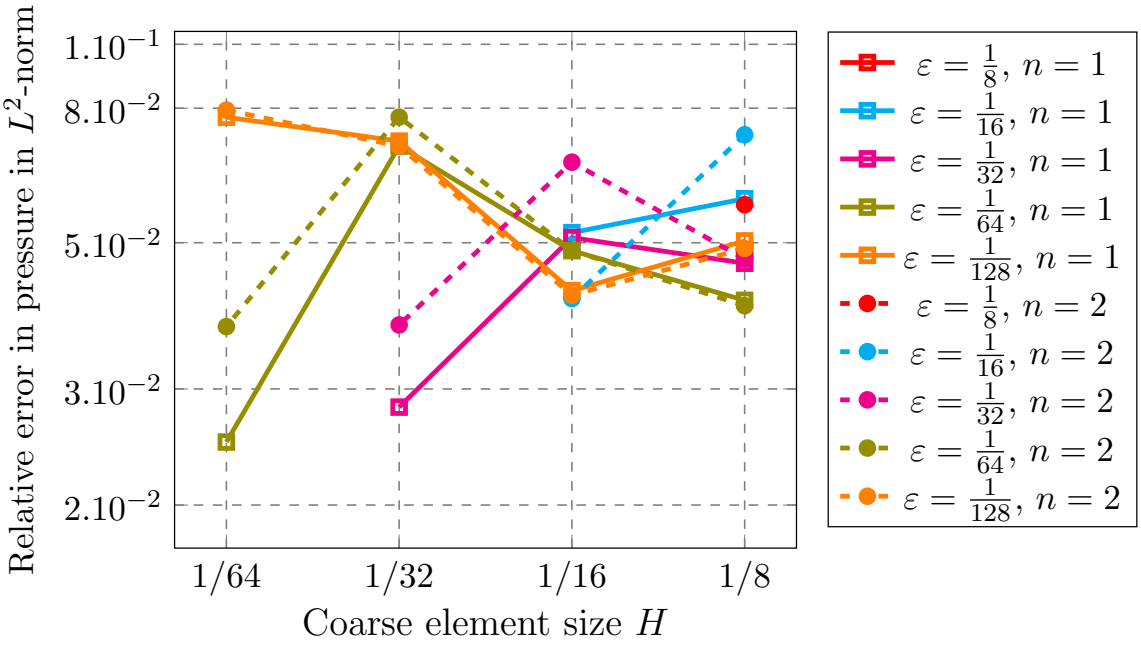


Figure 15.4: Relative errors between MsFEM approximations and reference solutions for pressure in L^2-norm (Oseen flow).

Contrary to the convergence analysis done for the Stokes problem in periodic cases in Section 14.2, Figures 15.2 and 15.3 emphasize the interest to increase the MsFEM order. Indeed, considering the MsFEM order $n = 2$, allows to improve significantly the velocity approximations. For example for the configuration ($\varepsilon = 1/32$, $H = 1/32$) or for the configuration ($\varepsilon = 1/16$, $H = 1/16$) the relative error is between twice and three times as big for $n = 1$ as for $n = 2$. What is more, for these configurations with $n = 1$, the errors do not decrease in the particular

cases where $H \approx \varepsilon$, i.e. when obstacles cut triangles only at their vertices. However, for the pressure approximations, Figure 15.4, the error behaviour is more complex. Indeed, it seems that the MsFEM for $n = 1$ leads to better approximations. We are not able to explain this phenomenon. In all cases, the relative errors are bounded. The relative errors in velocity in L^2-norm vary in a range approximately from 8×10^{-2} to 3×10^{-1}. The relative errors in velocity in H^1-norm vary in a range approximately from 2×10^{-1} to 5×10^{-1}. The relative errors in pressure in L^2-norm vary in a range approximately from 3×10^{-2} to 8×10^{-2}. These errors are typical errors for MsFEM approximations. These error behaviours emphasize the importance of choosing appropriate H, ε, and n values to minimise errors in the computational simulations.

Comparison with previous MsFEMs Figure 15.5 presents similar convergence curves as previously but by including the relative errors obtained with previous MsFEMs, i.e. the CR3 and the CR4 MsFEMs, for $\varepsilon = 1/64$, in order to emphasize the interest of considering high-order MsFEMs. Additional comparisons can be found in Figures F.19 to F.21 in Appendix for other periods ε.

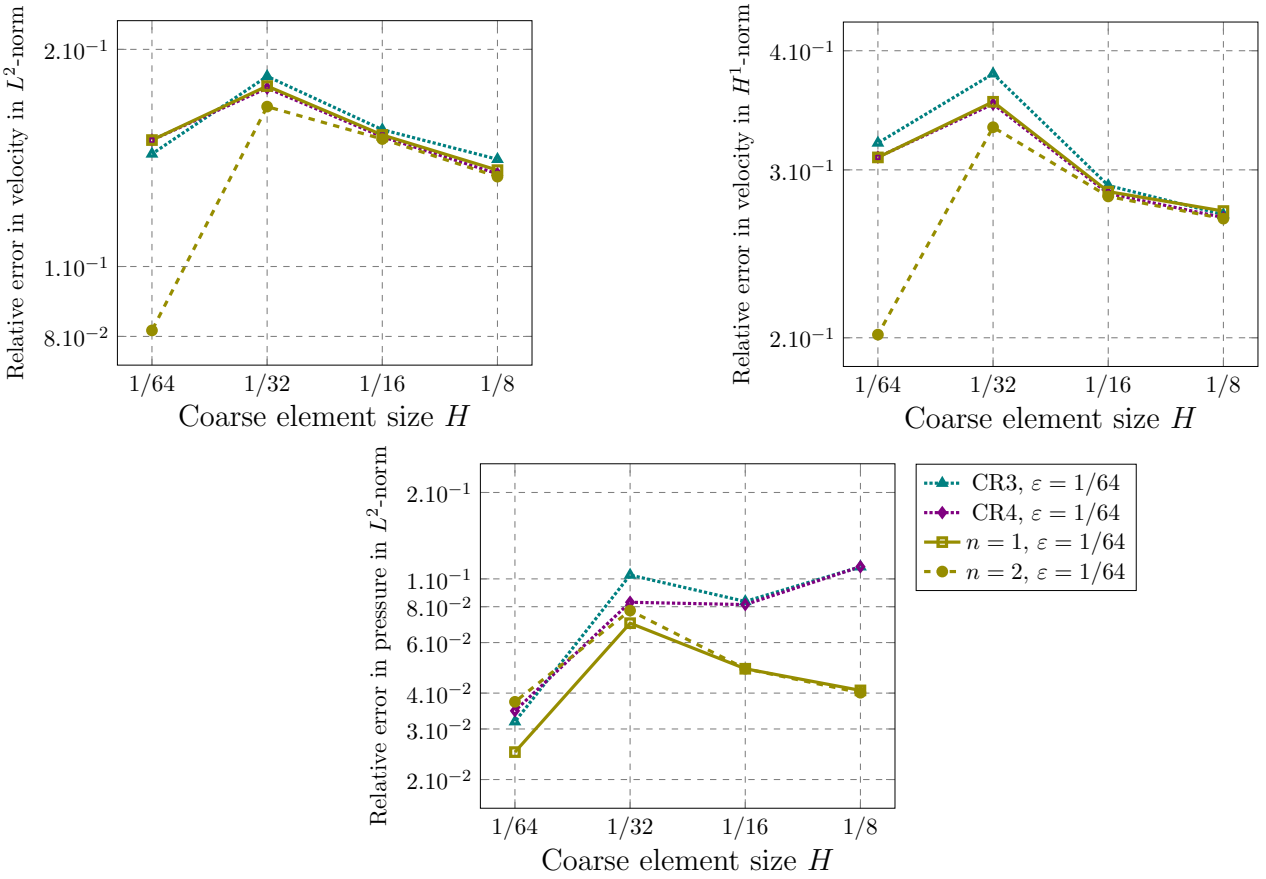
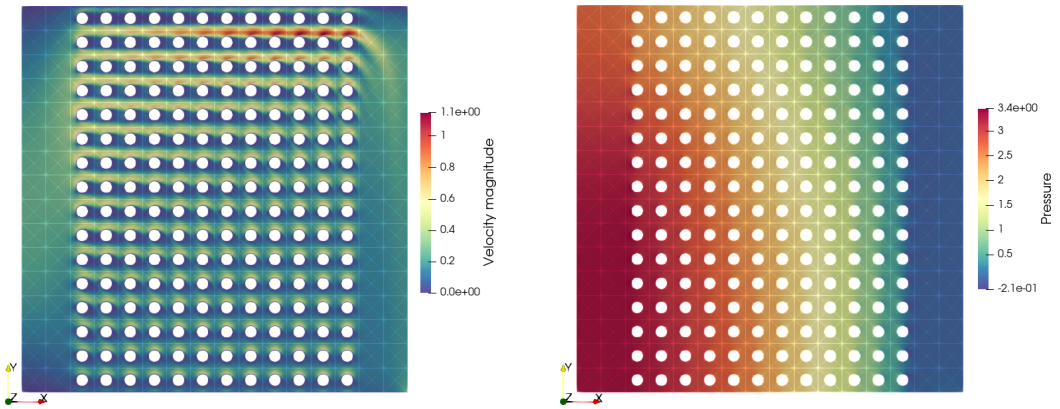


Figure 15.5: Relative errors between MsFEM approximations and the reference solution, $\varepsilon = 1/64$ (Oseen flow).

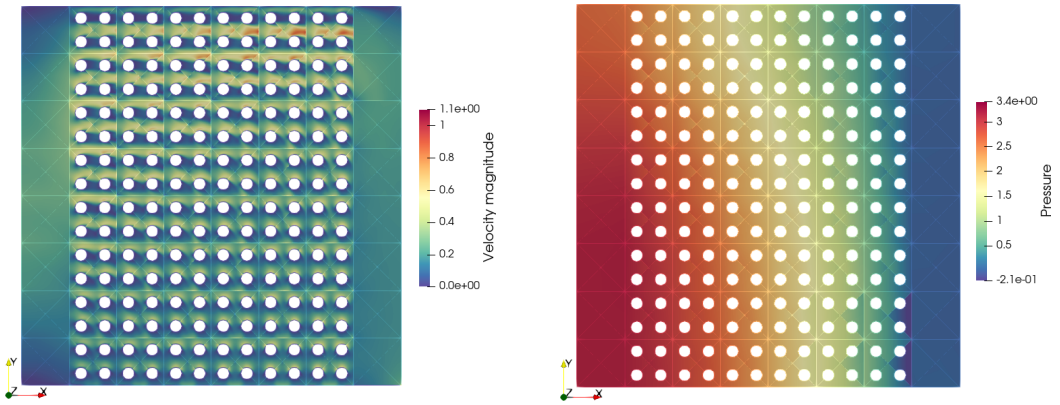
From Figures 15.5 and F.19 to F.21, same conclusions as for the convergence of MsFEMs for the Stokes flows can be drawn. In particular, the high-order MsFEMs outperform the previous MsFEMs, the CR3 and CR4 MsFEMs.

Some illustrative figures In what follows, we illustrate these errors with some numerical results. We focus on a periodic domain with $\varepsilon = 1/16$ for the ease of viewing.



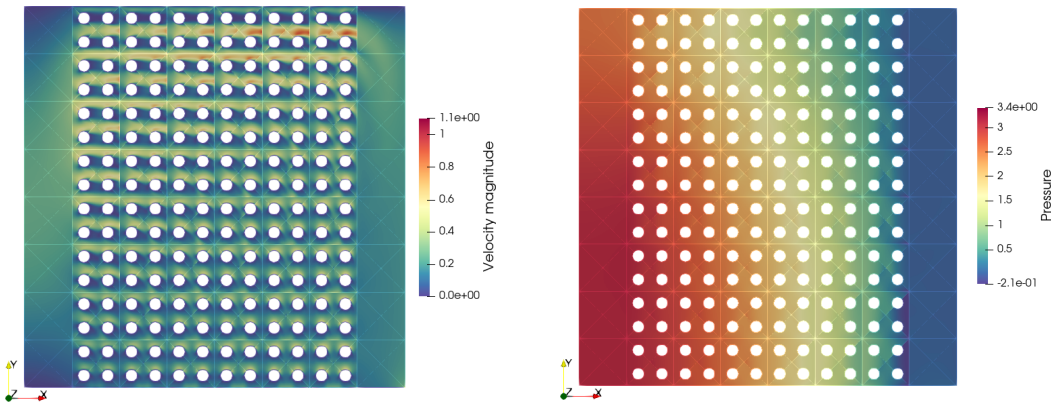
(a) Velocity magnitude. (b) Pressure field.

Figure 15.6: Reference solution (Oseen flow), $\varepsilon = 1/16$.



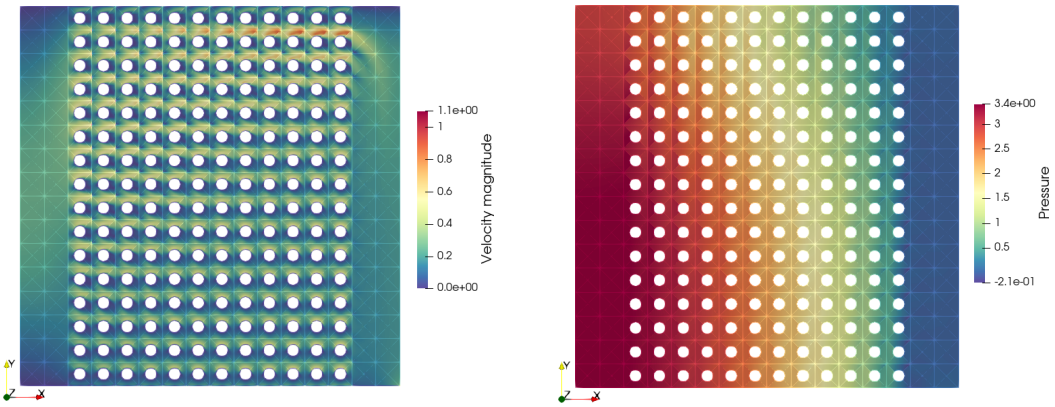
(a) Velocity magnitude. (b) Pressure field.

Figure 15.7: MsFEM approximation (Oseen flow), $n = 1$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/16$.



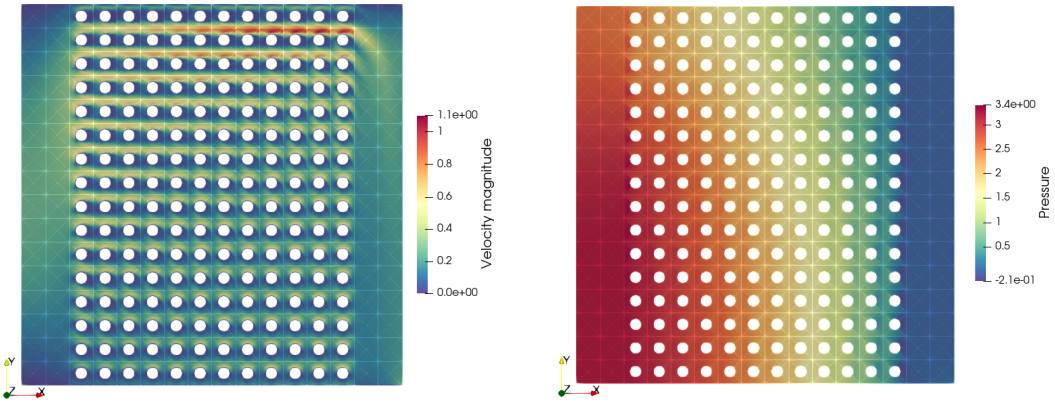
(a) Velocity magnitude. (b) Pressure field.

Figure 15.8: MsFEM approximation (Oseen flow), $n = 2$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/16$.



(a) Velocity magnitude. (b) Pressure field.

Figure 15.9: MsFEM approximation (Oseen flow), $n = 1$, $H = 1/16$ (1024 coarse elements), $\varepsilon = 1/16$.



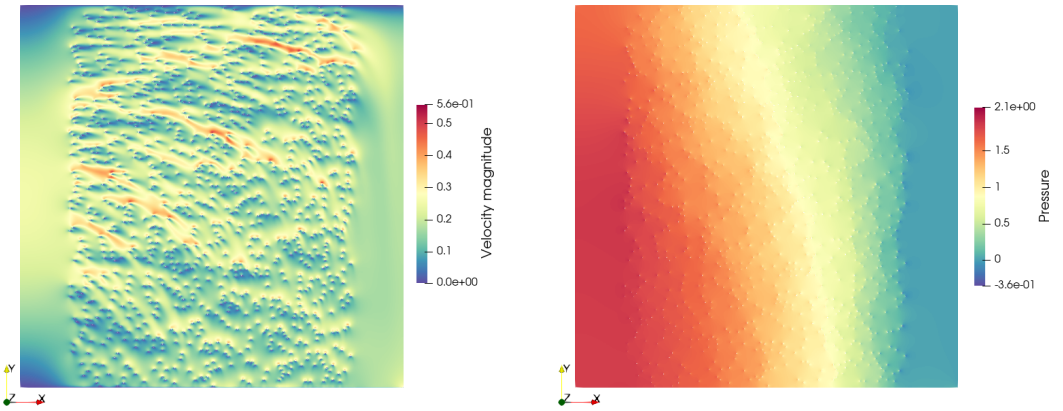
(a) Velocity magnitude. (b) Pressure field.

Figure 15.10: MsFEM approximation (Oseen flow), $n = 2$, $H = 1/16$ (1024 coarse elements), $\varepsilon = 1/16$.

A visual inspection of the solutions reveals that for $n = 1$, the physical solution lacks continuity when $H = 1/8$ (Figure 15.7) or $H = 1/16$ (Figure 15.9). Conversely, for $n = 2$ (Figures 15.7 and 15.9), the solutions are more continuous, and the variations of the velocity are captured more effectively. The case $n = 2$ and $H = 1/16$ provides the most accurate approximations in comparison to the reference solution (Figure 15.6), which is in agreement with the convergence analysis. In both cases, where $n = 1$ and $n = 2$, the pressure field is in agreement with the reference solution. For the sake of comparison, additional numerical applications, for other periods ε, are presented in Appendix F.2.2.

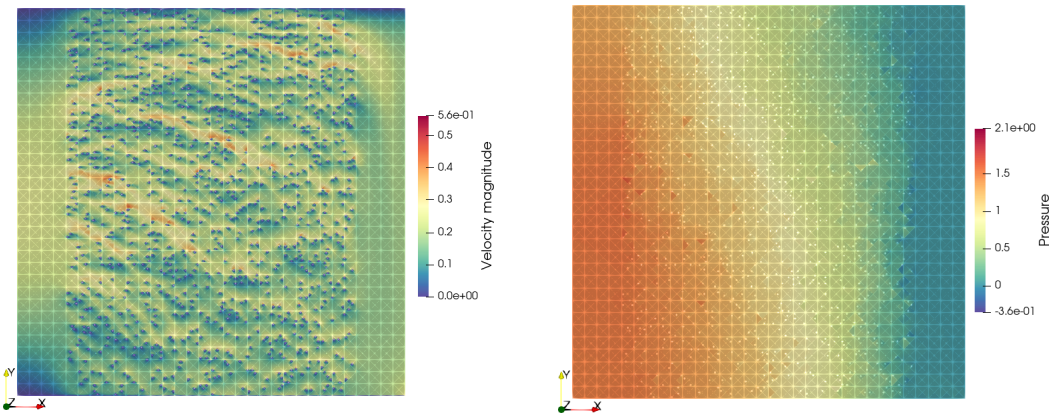
15.2 Simulations in a two-dimensional domain with ramdomly placed obstacles

In this section, we present simulations of the Oseen flows in the domain Figure 13.3a.



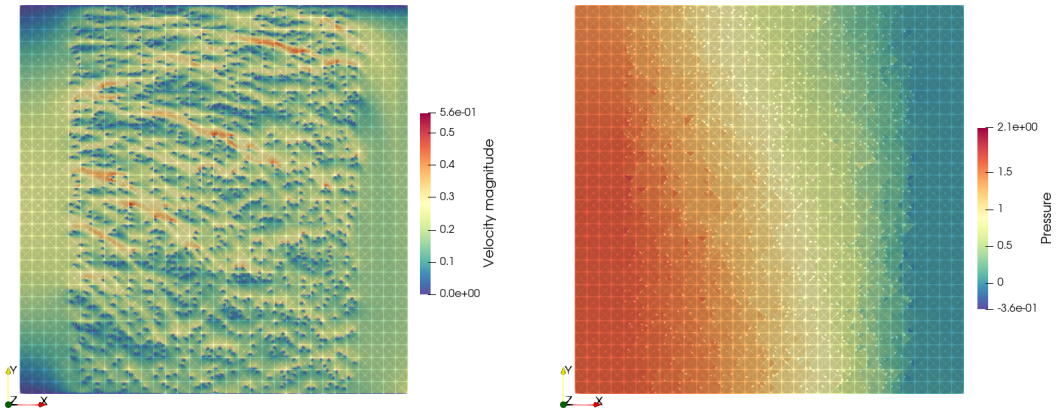
(a) Velocity magnitude. (b) Pressure field.

Figure 15.11: Reference solution (Oseen flow), domain with 1024 randomly placed obstacles.



(a) Velocity magnitude. (b) Pressure field.

Figure 15.12: MsFEM approximation (Oseen flow), $n = 1$, $H = 1/32$ (4096 coarse elements), domain with 1024 randomly placed obstacles.



(a) Velocity magnitude.

(b) Pressure field.

Figure 15.13: MsFEM approximation (Oseen flow), $n = 2$, $H = 1/32$ (4096 coarse elements), domain with 1024 randomly placed obstacles.

As for the Stokes problem, a visual inspection of the solutions reveals that the MsFEM for $n = 1$ (Figure 15.12) appears to underestimate the velocity magnitude in comparison to the reference solution (Figure 15.11). In contrast to the results of the convergence analysis conducted in periodic cases (see Figure 15.4), the MsFEM for $n = 2$ (Figure 14.13) yields a more precise approximation of pressure. For the sake of comparison, simulations of the Oseen flows in the domain Figure 13.3b are presented in Appendix F.2.2.

In conclusion, this part presents several numerical applications of the MsFEMs developed in this thesis. In all cases, the MsFEM approximations are in agreement with expectations and the relative errors obtained are typical for multi-scale approximations. In particular, the interest of considering high-order MsFEM and increasing the MsFEM order through numerical simulations of Oseen problems is demonstrated. It should be noted that the quantitative analysis of the numerical results is not straightforward. The first difficulty is obtaining reference solutions, as the considered meshes are large and the problems difficult to resolve due to the presence of perforations. The computation of reference solutions has necessitated the utilisation of domain decomposition methods and state-of-the-art solvers and preconditioners. Furthermore, the comparison between the reference solutions and the MsFEM approximations is not straightforward, as the solutions are not in the same finite element spaces. This necessitates the utilisation of interpolations, which also engenders the introduction of certain numerical errors. Furthermore, when the domains employed to compute the reference solutions (in the decomposition domains methods) are not the coarse triangles, but are generated by an alternative partitioner, this comparison could not be made due to difficulties in interpolating the solutions.

Part VI

Towards the Navier–Stokes problem

Chapter 16

A first approach to solve the Navier–Stokes problem and perspectives

Contents

In this chapter, we propose to solve the stationary Navier–Stokes problem using an MsFEM approach. We propose a strategy to solve the Navier–Stokes problem on a coarse mesh using well-adapted basis functions defined by Oseen problems.

16.1 Problem definition

The steady state Navier–Stokes problem with homogeneous Dirichlet boundary conditions in a perforated domain is to find the velocity $\boldsymbol{u_\varepsilon} : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\left\{\begin{array}{rcll} -\nu\Delta \boldsymbol{u_\varepsilon} + \boldsymbol{u_\varepsilon}\cdot\boldsymbol{\nabla}\boldsymbol{u_\varepsilon} + \boldsymbol{\nabla} p_\varepsilon &=& \boldsymbol{f} & \text{in } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u_\varepsilon} &=& 0 & \text{in } \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{on } \partial B^\varepsilon \cap \partial\Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &=& \boldsymbol{0} & \text{on } \partial\Omega \cap \partial\Omega^\varepsilon, \end{array}\right. \tag{16.1}$$

with $\boldsymbol{f}$ the applied force and $\nu > 0$ the viscosity. We introduce the velocity space

$$V = H_0^1(\Omega^\varepsilon)^d = \{\boldsymbol{u} \in H^1(\Omega^\varepsilon)^d \text{ such that } \boldsymbol{u}|_{\partial\Omega^\varepsilon} = 0\},$$

and the pressure space

$$M = L_0^2(\Omega^\varepsilon) = \{p \in L^2(\Omega^\varepsilon) \text{ such that } \int_{\Omega^\varepsilon} p = 0\}.$$

We note $X = V \times M$. For simplicity the fluid domain is assumed to be connected in order for the pressure to be uniquely defined in M. We introduce the following non-linear form

$$c^{NS}\left((\boldsymbol{u}, p), (\boldsymbol{v}, q)\right) = \int_{\Omega^\epsilon} \left(\nu \boldsymbol{\nabla}\boldsymbol{u} \cdot \boldsymbol{\nabla}\boldsymbol{v} + (\boldsymbol{u} \cdot \boldsymbol{\nabla})\boldsymbol{u} \cdot \boldsymbol{v} - p \operatorname{div} \boldsymbol{v} - q \operatorname{div} \boldsymbol{u}\right).$$

Then, a weak formulation of the Navier–Stokes problem (16.1) reads: find $(\boldsymbol{u_\varepsilon}, p_\varepsilon) \in X$ such that

$$c^{NS}((\boldsymbol{u_\varepsilon}, p_\varepsilon), (\boldsymbol{v}, q)) = \int_{\Omega^\varepsilon} \boldsymbol{f} \cdot \boldsymbol{v} \quad \forall (\boldsymbol{v}, q) \in X.$$

The analysis of non-linear variational problems is more difficult than the analysis of linear variational problem. We do not present this study in this thesis. The reader can found analysis of such problems in [84].

16.2 Numerical strategy

The main difference between the Navier–Stokes problem and the Stokes (7.1) or the Oseen (12.1) problems is the presence of non-linear terms. Consequently, iterative methods are required to solve the flows. Different methods, such as the Newton method or the pseudo-transient approach, are described in the literature. The aforementioned methodologies are all based on the linearisation of the Navier—Stokes equations. Given an initial guess, $\boldsymbol{u}^0 \in V$, a sequence of iterates $\{\boldsymbol{u}^k\}_{k=1,\dots,n} \in V$ is computed which is expected to converge towards the solution to the weak formulation. In most of these methods, the linearisation consists in considering explicitly the advection velocity, i.e. using the velocity at the previous iteration as advection velocity. This procedure leads to solve an Oseen type problem at each iteration of the iterative method.

The main paradigm of MsFEM is to build well-adapted basis functions to the considered problem. For example, as seen in Chapter 12, to solve the Oseen problem on a coarse grid, with a given Oseen velocity $\boldsymbol{U_0}$, the basis functions are solutions of local problems which are themselves Oseen type problems with the same Oseen velocity $\boldsymbol{U_0}$. However, as explained above, for solving iteratively the Navier–Stokes problems, at each iteration the global problem to be computed changes since the advection velocity changes. Consequently, to compute the Navier–Stokes problem on a coarse grid with MsFEM, a new basis has to be computed at each iteration, taking into account the new advection velocity.

Remark 16.1. One may attempt to resolve the Navier–Stokes problem by utilising a single fixed basis, derived either from the Oseen problem with a specified arbitrary Oseen velocity or from the Stokes problem (equivalent to the Oseen problem with a null advection velocity). Nevertheless, even if the iterative algorithm converges, the basis functions may not be adapted to the problem, resulting in inaccurate outcomes. Indeed, as soon as the advection velocity of the global problem differs significantly from the velocity used in the local problems, the basis functions become non-adapted to the problem.

In what follows, we propose a procedure to solve the stationary Navier–Stokes problem using a pseudo-transient strategy with adaptive basis. We consider a pseudo-transient problem and wait for it to converge in long time to the solution to the stationary problem. In other words, we solve the unsteady Navier–Stokes, possibly with large time steps, and wait until the solution no longer moves in time.

16.2.1 The unsteady Navier–Stokes problem

Instead of considering the stationary Navier–Stokes problem, we consider the unsteady Navier–Stokes problem. The unsteady Navier–Stokes problem with homogeneous Dirichlet boundary conditions in a perforated domain is to find the velocity $\boldsymbol{u}_\varepsilon : \Omega^\varepsilon \to \mathbb{R}^d$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\left\{\begin{array}{rcll} \partial_t \boldsymbol{u}_\varepsilon - \nu\Delta\boldsymbol{u}_\varepsilon + \boldsymbol{u}_\varepsilon\cdot\boldsymbol{\nabla}\boldsymbol{u}_\varepsilon + \boldsymbol{\nabla}p_\varepsilon &=& \boldsymbol{f} & \text{in } \Omega^\varepsilon\times(0,T),\\ \operatorname{div}\boldsymbol{u}_\varepsilon &=& 0 & \text{in } \Omega^\varepsilon\times(0,T),\\ \boldsymbol{u}_\varepsilon &=& \boldsymbol{0} & \text{on } \partial B^\varepsilon\cap\partial\Omega^\varepsilon\times(0,T),\\ \boldsymbol{u}_\varepsilon &=& \boldsymbol{0} & \text{on } \partial\Omega\cap\partial\Omega^\varepsilon\times(0,T),\\ \boldsymbol{u}_\varepsilon(\cdot,0) &=& \boldsymbol{u}^0 & \text{in } \Omega^\varepsilon, \end{array}\right. \tag{16.2}$$

with $\boldsymbol{f}$ the load and $\nu > 0$ the viscosity.

To compute the problem (16.2), we use a first order time integration, and we consider explicitly the advection velocity. Consequently, for $i > 0$, problem (16.2) becomes

$$\left\{\begin{array}{rcll} \dfrac{\boldsymbol{u}^{i+1}-\boldsymbol{u}^i}{\Delta t} - \nu\Delta\boldsymbol{u}^{i+1} + \boldsymbol{u}^i\cdot\boldsymbol{\nabla}\boldsymbol{u}^{i+1} + \boldsymbol{\nabla}p^{i+1} &=& \boldsymbol{f} & \text{in } \Omega^\varepsilon,\\ \operatorname{div}\boldsymbol{u}^{i+1} &=& 0 & \text{in } \Omega^\varepsilon,\\ \boldsymbol{u}^{i+1} &=& \boldsymbol{0} & \text{on } \partial B^\varepsilon\cap\partial\Omega^\varepsilon,\\ \boldsymbol{u}^{i+1} &=& \boldsymbol{0} & \text{on } \partial\Omega\cap\partial\Omega^\varepsilon, \end{array}\right. \tag{16.3}$$

where $\Delta t > 0$ is the time step. In particular the first equation of (16.3), can be rewritten as

$$\frac{\boldsymbol{u}^{i+1}}{\Delta t} - \nu\Delta\boldsymbol{u}^{i+1} + \boldsymbol{u}^i\cdot\boldsymbol{\nabla}\boldsymbol{u}^{i+1} + \boldsymbol{\nabla}p^{i+1} = \boldsymbol{f} + \frac{\boldsymbol{u}^i}{\Delta t},$$

which is a classical Oseen problem, with the Oseen velocity $\boldsymbol{u}^i$.

16.2.2 Discretized coarse scale problem

We note V_H the approximation space for the velocity and M_H the approximation space for the pressure. We recall that V_H is the space generated by the multi-scale basis functions (derived either from the Stokes or the Oseen problems) and M_H is the space of fully discontinuous polynomial of order n. The coarse-scale discretisation of the unsteady Navier–Stokes problem (16.3) reads at each iteration, for $i \geq 0$: find $\boldsymbol{u_H}^{i+1} \in V_H$ and $p_H^{i+1} \in M_H (:= \mathbb{P}_n^{dc})$ such that

$$\begin{cases} m_H(\boldsymbol{u_H}^{i+1}, \boldsymbol{v}) + a_H(\boldsymbol{u_H}^{i+1}, \boldsymbol{v}) + c_H(\boldsymbol{u_H}^{i}; \boldsymbol{u_H}^{i+1}, \boldsymbol{v}) + b_H(p_H^{i+1}, \boldsymbol{v}) \\ \qquad = F_H(\boldsymbol{v}) + m_H(\boldsymbol{u_H}^{i}, \boldsymbol{v}) \quad \forall \boldsymbol{v} \in V_H \\ \qquad b_H(q, \boldsymbol{u_H}^{i+1}) = 0 \quad \forall q \in M_H \end{cases} \tag{16.4}$$

where

$$m_H(\boldsymbol{u}, \boldsymbol{v}) = \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} \frac{1}{\Delta t} \boldsymbol{u} \cdot \boldsymbol{v},$$

$$a_H(\boldsymbol{u}, \boldsymbol{v}) = \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} \nu \boldsymbol{\nabla u} \cdot \boldsymbol{\nabla v},$$

$$c_H(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{v}) = \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} (\boldsymbol{w} \cdot \boldsymbol{\nabla}) \boldsymbol{u} \cdot \boldsymbol{v},$$

$$b_H(q, \boldsymbol{v}) = \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} q \operatorname{div} \boldsymbol{v},$$

$$F_H(\boldsymbol{v}) = \sum_{T \in \mathcal{T}_H} \int_{T \cap \Omega^\varepsilon} \boldsymbol{f} \cdot \boldsymbol{v}.$$

16.2.3 Algorithm

We note $\mathcal{B}^{St}(\nu)$ the MsFEM basis derived from the Stokes operator (see Chapter 7), computed with the viscosity ν. We note $\mathcal{B}^{Os}(\nu, \boldsymbol{u_0})$ the MsFEM basis derived from the Oseen operator (see Chapter 12), computed with the viscosity ν and the Oseen velocity $\boldsymbol{u_0}$. Then, the procedure to compute the Navier–Stokes problem on a coarse grid reads as follows.

Algorithm 16.2 (Pseudo-transient Navier–Stokes)**.**

- Set Δt;
- Set ${\boldsymbol{u_H}}^0 = \mathbf{0}$;
- Compute $\mathcal{B}^{St}(\nu)$;
- Find $({\boldsymbol{u_H}}^1, p_H^1)$ by computing (16.4), which is equivalent to a Stokes type problem due to the choice of ${\boldsymbol{u_H}}^0$, on the coarse grid using the basis $\mathcal{B}^{St}(\nu)$;
- for $(i = 1,\ i < niter,\ i = i + 1)$:
  - Compute $\mathcal{B}^{Os}(\nu, {\boldsymbol{u_H}}^i)$;
  - Find $({\boldsymbol{u_H}}^{i+1}, p_H^{i+1})$ by computing (16.4) on the coarse grid using the basis $\mathcal{B}^{Os}(\nu, {\boldsymbol{u_H}}^i)$;
  - Compute the residuals;
  - Break if convergence or divergence;

This procedure can be easily slightly improved. For example, if the residuals are small enough, i.e. if the solution does not vary a lot, the computation of a new basis may not be useful.

16.3 Numerical examples

We solve the Navier–Stokes problem (16.1) in a two-dimensional open channel $\Omega = [0 \leq x \leq 1, 0 \leq y \leq 1]$, where the heterogeneities are represented by solid obstacles. We choose the fine mesh size $h = 1 \times 10^{-3}$, so that the Péclet number is smaller than 1. We assign $\nu = 1 \times 10^{-3}$ and $\boldsymbol{f} = \mathbf{0}$. The parabolic inflow boundary condition $\boldsymbol{u} = y(1-y)\boldsymbol{e_1}$ is imposed at the inlet. The Neumann boundary condition $\nu \boldsymbol{\nabla u n} - p\boldsymbol{n} = \mathbf{0}$ is imposed at the outlet and the no-slip condition is applied on other boundaries. The Reynolds number is of 1000. Numerically, we define the residuals as

$$res = \frac{\|{\boldsymbol{u_H}}^{i+1} - {\boldsymbol{u_H}}^i\|_\infty}{\|{\boldsymbol{u_H}}^{i+1}\|_\infty} + \frac{\|p_H^{i+1} - p_H^i\|_\infty}{\|p_H^{i+1}\|_\infty},$$

and consider that the algorithm has converged when the residual becomes less than 1×10^{-8}. The time step Δt was set to 100 (this value has minimal impact on the results, numerical tests with a time step of 10 or 1000 yielded comparable outcomes). In the different simulations presented in the section, the algorithm converged in approximately 40 iterations. We present in Figure 16.1, the residuals for the MsFEM approximations Figure 16.3 for different time-steps.

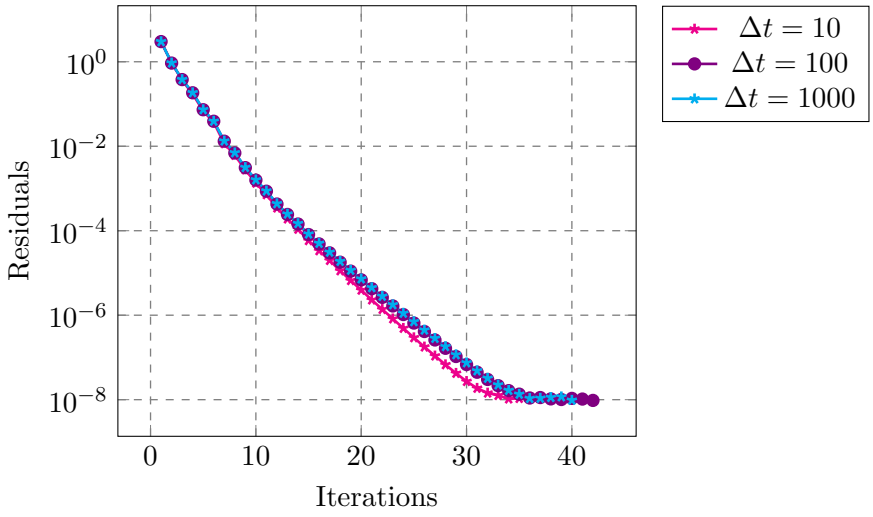
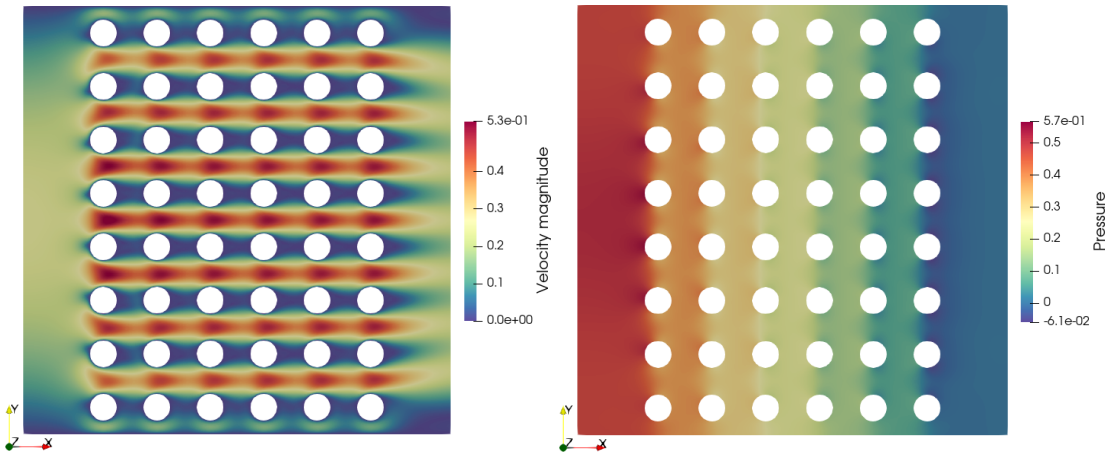
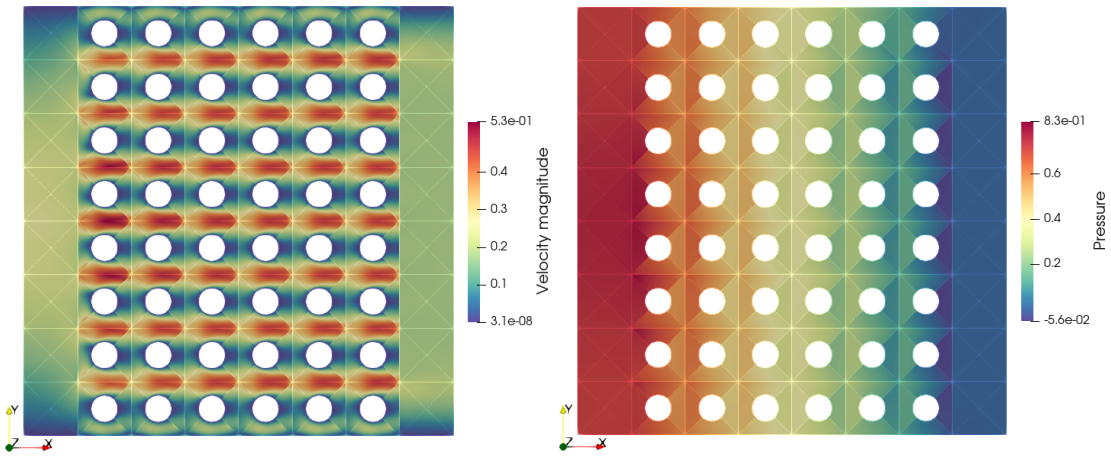


Figure 16.1: Convergence of the pseudo-transient Navier–Stokes algorithm.



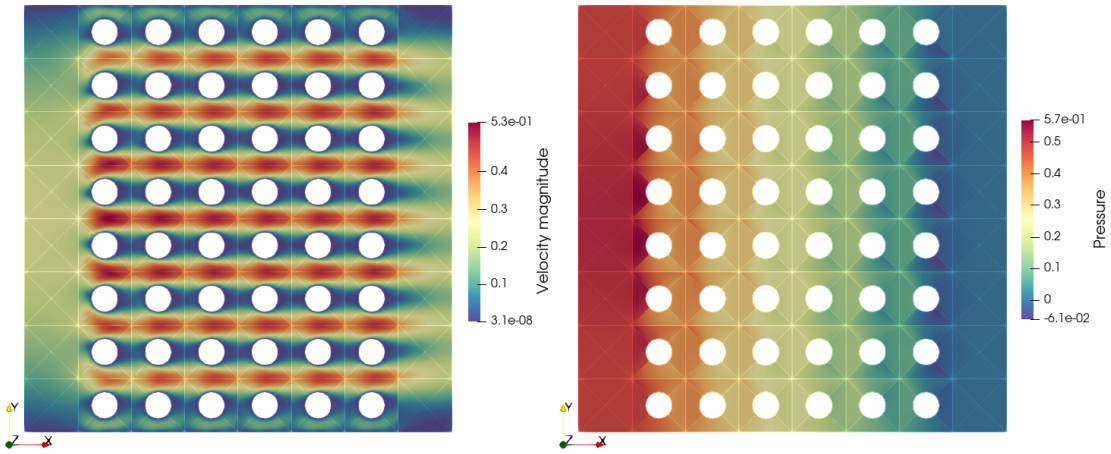
(a) Velocity magnitude. (b) Pressure field.

Figure 16.2: Reference solution, $\varepsilon = 1/8$.



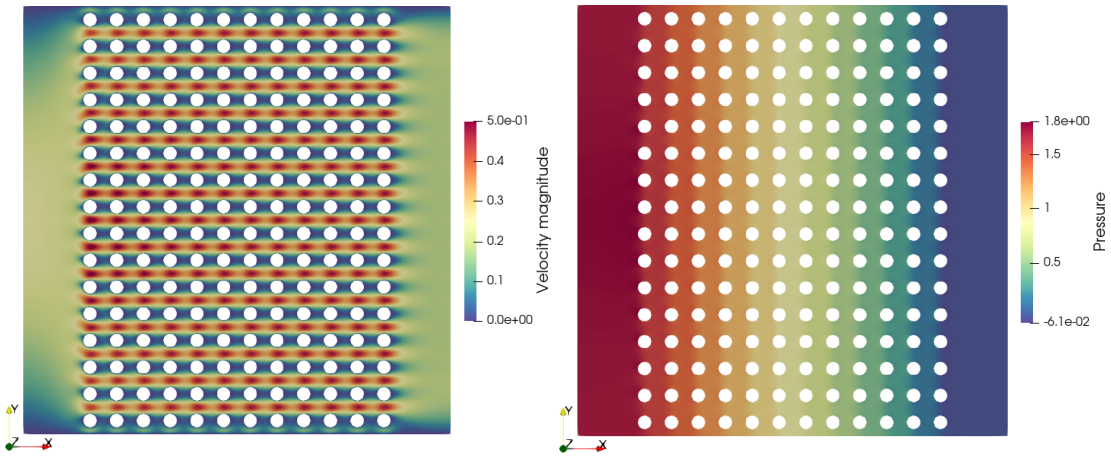
(a) Velocity magnitude. (b) Pressure field.

Figure 16.3: MsFEM approximation of the steady Navier–Stokes, $n = 1$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/8$.



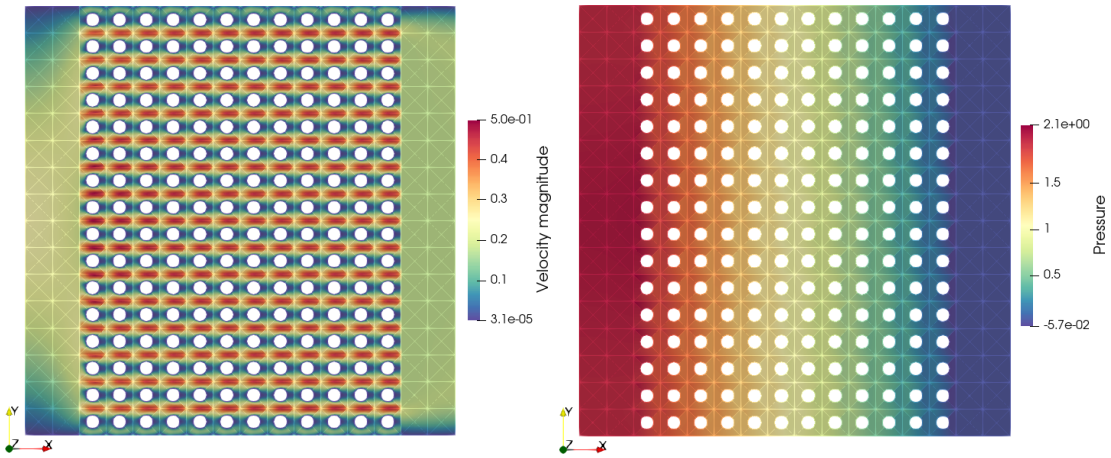
(a) Velocity magnitude. (b) Pressure field.

Figure 16.4: MsFEM approximation of the steady Navier–Stokes, $n = 2$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/8$.



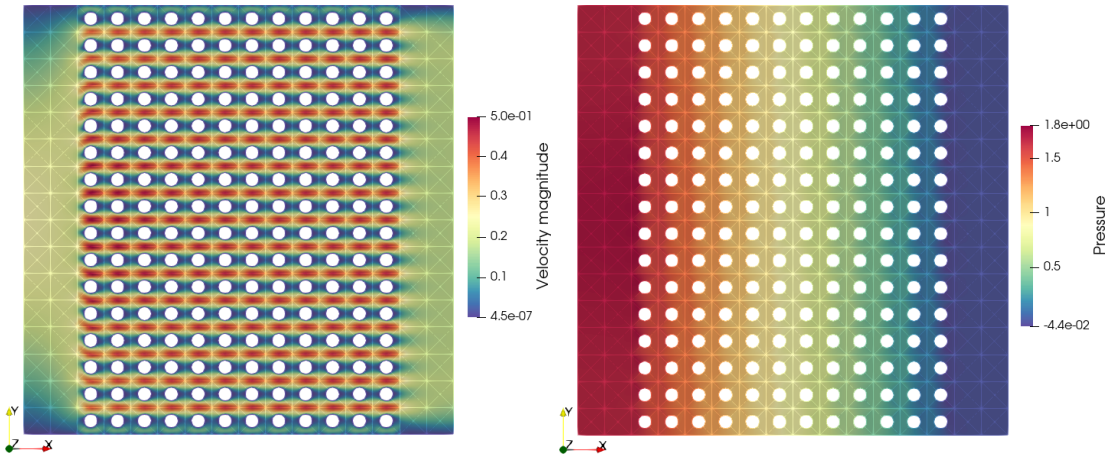
(a) Velocity magnitude. (b) Pressure field.

Figure 16.5: Reference solution, $\varepsilon = 1/16$.



(a) Velocity magnitude. (b) Pressure field.

Figure 16.6: MsFEM approximation of the steady Navier–Stokes, $n = 1$, $H = 1/16$ (1024 coarse elements), $\varepsilon = 1/16$.



(a) Velocity magnitude. (b) Pressure field.

Figure 16.7: MsFEM approximation of the steady Navier–Stokes, $n = 2$, $H = 1/16$ (1024 coarse elements), $\varepsilon = 1/16$.

First, Figures 16.3, 16.4, 16.6 and 16.7 allow to validate the methodology presented in Algorithm 16.2 to solve the unsteady Navier–Stokes problem. As expected, the velocity increases between obstacles. A visual inspection of the approximations allows to confirm that the MsFEM for $n = 2$ (Figures 16.4 and 16.7) yields to better physical solutions, in agreement with references solutions (Figures 16.2 and 16.5). Indeed, the approximations for $n = 2$ capture well the variations of the velocity magnitude on the one hand around the obstacles, and on the other hand at the exit of the congested area. This analysis clearly demonstrates the benefits of employing high-order MsFEM approaches.

Remark 16.3. The simulations presented in this section were conducted on a local machine. Indeed, the procedure outlined in Algorithm 16.2 is not suitable for use on a cluster due to the inherent limitations of the queueing system when submitting a job on a cluster. For instance, the job allowing the compututation of the basis functions must await its turn in the queue before being initiated.

16.4 Perspective to solve the Navier–Stokes problem: Generalized Multi-scale Finite Element Method (GMsFEM)

The strategy presented in Section 16.2, consisting in computing a new basis at each iteration can be too expensive and not tractable in practice. This might be a limitation of MsFEM for solving the Navier–Stokes problem or in general for solving non-linear problems or time dependent problems. One appealing approach to solve such problems is the Generalized Multi-scale Finite Element Method (GMsFEM), introduced in [69]. This method was originally presented for solving problems with input parameters, that we denote by θ in what follows. This method allows to solve global problems of the form $L_\theta(v) = 0$, in which θ is not known a priori when computing the basis functions. A typical example of such an input is the permeability in Darcy's equation. This method has been applied successfully in the literature for flow problems [54, 52, 55]. In what follows, we explain the concept of the Generalized Multi-scale Finite

Element Method (GMsFEM) and a possible way to apply it to solve the Navier–Stokes problem by considering the advective velocity as an input parameter.

16.4.1 Generic structure of GMsFEM

Similarly to MsFEM approach, GMsFEM is composed of two main steps: an offline step and an online step.

The offline step The offline stage consists first in building a snapshot space, noted V_H^{snap}. To build this space, we consider M_{snap} well-chosen inputs θ_k, for $k = 1, \ldots, M_{\text{snap}}$, discretising as best as possible the space in which the input parameter θ lies. Then, for each θ_k, we compute the corresponding multi-scale basis, noted $\mathcal{B}(\theta_k)$. Afterwards, the snapshot space V_H^{snap} corresponds to the space generated by the concatenation of all the basis $\mathcal{B}(\theta_k)$. To summarise, we have

$$\mathcal{B}^{\text{snap}} = \bigcup_{k=1}^{M_{\text{snap}}} \mathcal{B}(\theta_k), \quad V_H^{\text{snap}} = \text{span}\{\mathcal{B}^{\text{snap}}\}.$$

However, the basis $\mathcal{B}^{\text{snap}}$ contains a large number of functions. This leads to an expensive computation in the online stage. Consequently, it is possible to reduce the dimension of the local snapshot space using a spectral decomposition and by selecting high energy elements. As result, we obtain a new basis $\mathcal{B}^{\text{off}}$ which contains M_{off} functions, with $M_{\text{off}} < M_{\text{snap}}$. Each function in $\mathcal{B}^{\text{off}}$ is a linear combination of functions in $\mathcal{B}^{\text{snap}}$. The offline space, noted V_H^{off}, is the space generated by $\mathcal{B}^{\text{off}}$. To summarise we have

$$\mathcal{B}^{\text{off}} = \mathfrak{F}(\mathcal{B}^{\text{snap}}), \quad V_H^{\text{off}} = \text{span}\{\mathcal{B}^{\text{off}}\}.$$

with $\mathfrak{F}$ representing a generic application.

Online step At this stage, we have at disposal, a multi-scale basis adapted to a large range of input parameters. Then, in the online step, for a given input parameter, θ_{on}, of the global problem, we compute the corresponding local space, called V_H^{on}. This space is obtained by computing M_{on} functions, with $M_{\text{on}} \leq M_{\text{off}}$, obtained by linear combinations of functions in $\mathcal{B}^{\text{off}}$. The coefficients in the linear combination are given by well-chosen eigenvalue problems involving the input parameter θ_{on}. In other words, we choose the basis functions the more adapted to the input parameter θ_{on}. To summarise we have

$$\mathcal{B}^{\text{on}} = \mathfrak{F}(\theta_{on}; \mathcal{B}^{\text{off}}), \quad V_H^{\text{on}} = \text{span}\{\mathcal{B}^{\text{on}}\}.$$

16.4.2 Application to the Navier–Stokes problem

For solving the Navier–Stokes problem, we can consider the advection velocity as an input parameter. The offline space will be obtained by computing local basis functions derived from the Oseen problem for a large range of well-chosen Oseen velocities ${\boldsymbol{u_0}}^k$. Then, applying the procedure described above, one needs at each iteration of the iterative method to determine a new basis solving local eigenvalue problems and doing linear combinations. This procedure

can be less expensive than computing a new basis at each iteration. Moreover, this approach is more in accordance with the MsFEM paradigm, which comprises an offline phase and an online phase. We did not consider this strategy in this thesis, since it is out of the scope of the work, but it is a path for future development.

Appendix A

Appendix of Chapter 5: Strategy to find the unisolvent basis functions in three dimensions and examples

Contents

In Chapter 5, we have introduced a family of scalar non-conforming finite elements of order two and three with respect to the H^1-norm on tetrahedra. In this appendix, we explain the strategy that we have used to find these finite elements. In addition, we give the list of the suitable basis Σ_3 to complete the polynomial space $\mathbb{P}_2$. This appendix is also the appendix of [29].

A.1 Strategy to find the unisolvent basis functions

We propose to find numerically the additional functions of $\mathbb{P}_{n+2}$ needed to complete the space $\mathbb{P}_{n+1}$ while ensuring that the generated space V_{n+1} is unisolvent (in the sense of Definition 5.3) with respect to the set $\mathcal{N}_{n+1}$ of degrees of freedom. First, we introduce the following lemma.

Lemma A.1. Let K be a tetrahedron of vertices A, B, C, D, and λ_A, λ_B, λ_C, λ_D, the associated barycentric coordinates. Then, the family $\left\{(\lambda_A^\alpha \lambda_B^\beta \lambda_C^\gamma \lambda_D^\delta) \text{ such that } (\alpha,\beta,\gamma,\delta) \in [\![0,n]\!]^4 \text{ and } \alpha+\beta+\gamma+\delta = n\right\}$ is a basis of the space $\mathbb{P}_n(K)$.

Proof. The proof consists of two steps. In the first one, we show that the family $\left\{(\lambda_A^\alpha \lambda_B^\beta \lambda_C^\gamma \lambda_D^\delta) \text{ such that } (\alpha,\beta,\gamma,\delta) \in [\![0,n]\!]^4 \text{ and } \alpha+\beta+\gamma+\delta = n\right\}$ is linearly independent. In the second one, we show that this family has the same dimension as the space $\mathbb{P}_n(K)$.

Step 1 Let us call (ABC) the face delimited by the vertices A, B and C, and (AB) the edge delimited by the vertices A and B. We define similarly the other faces and edges of the

tetrahedron K. Let us show that the family $(\lambda_A^{\alpha_i}\lambda_B^{\beta_i}\lambda_C^{\gamma_i}\lambda_D^{\delta_i})$ with $\alpha_i+\beta_i+\gamma_i+\delta_i=n$ is linearly independent. We define the polynomial P as

$$P=\sum_i a_i\lambda_A^{\alpha_i}\lambda_B^{\beta_i}\lambda_C^{\gamma_i}\lambda_D^{\delta_i}$$

with some real scalars a_i. The objective is to show that $P=0$ implies that all the coefficients a_i are equal to 0. Below, we give the general procedure to show this implication.

First, we consider the restriction of P on the edge (AB), i.e. on the edge on which $\lambda_C=0$ and $\lambda_D=0$. Consequently, P is reduced to

$$P|_{(AB)}=\sum_i a_i\lambda_A^{\alpha_i}\lambda_B^{\beta_i}=0 \tag{A.1}$$

with $\alpha_i+\beta_i=n$. Now, given that on (AB), λ_B verifies $\lambda_B=1-\lambda_A$, the family $(\lambda_A^{\alpha_i}\lambda_B^{\beta_i})_i$ is a family of Bernstein polynomials which is known to be a basis of the space $\mathbb{P}_n(AB)$ (polynomials of order n on the edge (AB)). Consequently, all the coefficients a_i in (A.1) are null. We repeat the same procedure on all the other edges of K, which allows to show that all the coefficient a_i associated with monomials involving at most two barycentric coordinates are null.

At this stage, it remains in P only monomials involving at least three barycentric coordinates. We consider the restriction of P on the face (ABC), i.e. on the face on which $\lambda_D=0$. Consequently, P is reduced to

$$P|_{(ABC)}=\sum_i a_i\lambda_A^{\alpha_i}\lambda_B^{\beta_i}\lambda_C^{\gamma_i}=0$$

with $\alpha_i+\beta_i+\gamma_i=n$ and $\alpha_i\neq 0$, $\beta_i\neq 0$ and $\gamma_i\neq 0$. Consequently, it is possible to factorize P by the term $\lambda_A\lambda_B\lambda_C$ which leads to

$$P|_{(ABC)}=\lambda_A\lambda_B\lambda_C\sum_i a_i\lambda_A^{\alpha'_i}\lambda_B^{\beta'_i}\lambda_C^{\gamma'_i}=0$$

with $\alpha'_i+\beta'_i+\gamma'_i=n-3$. The polynomial P vanishing on (ABC) and $\lambda_A\lambda_B\lambda_C$ not being the null polynomial on (ABC), it follows that

$$Q=\sum_i a_i\lambda_A^{\alpha'_i}\lambda_B^{\beta'_i}\lambda_C^{\gamma'_i}=0.$$

It should be noted that in Q some α'_i, β'_i or γ'_i are equal to zero. Consequently, it is possible to resctrict Q to edges of (ABC) to show that the coefficients associated with monomials involving at most two barycentric coordinates are null. Afterwards, if necessary, it is one again possible to factorize Q by $\lambda_A\lambda_B\lambda_C$ and we repeat the procedure until we have shown that all the coefficients a_i in the polynomial Q are null. We repeat the same procedure on all the other faces of K, which allows to show that all the coefficient a_i associated with monomials involving at most three barycentric coordinates are null.

At this stage, it remains in P only monomials involving exactly four barycentric coordinates,

i.e. for $n \geq 4$,

$$P = \sum_i a_i \lambda_A^{\alpha_i} \lambda_B^{\beta_i} \lambda_C^{\gamma_i} \lambda_D^{\delta_i} = 0,$$

with $\alpha_i \neq 0$, $\beta_i \neq 0$, $\gamma_i \neq 0$ and $\delta_i \neq 0$ and $\alpha_i + \beta_i + \gamma_i + \delta_i = n$. Consequently, it is possible to factorize P by $\lambda_A \lambda_B \lambda_C \lambda_D$ which leads to

$$P = \lambda_A \lambda_B \lambda_C \lambda_D \sum_i a_i \lambda_A^{\alpha'_i} \lambda_B^{\beta'_i} \lambda_C^{\gamma'_i} \lambda_D^{\delta'_i} = 0$$

with $\alpha'_i + \beta'_i + \gamma'_i + \delta'_i = n - 4$. The polynomial P vanishing on K and $\lambda_A \lambda_B \lambda_C \lambda_D$ not being the null polynomial on K, it follows that

$$Q = \sum_i a_i \lambda_A^{\alpha'_i} \lambda_B^{\beta'_i} \lambda_C^{\gamma'_i} \lambda_D^{\delta'_i} = 0.$$

It should be noted that in Q some α'_i, β'_i, γ'_i or δ'_i are equal to zero. Consequently, we can repeat the above procedure (restriction of Q to faces, then to edges) to show that all coefficients a_i in Q are null. We have thus shown that all coefficients a_i associated with monomials involving four barycentric coordinates are null.

At the end of this general procedure, we have showed that all the coefficient a_i in P are null and consequently, that the family $\left\{(\lambda_A^\alpha \lambda_B^\beta \lambda_C^\gamma \lambda_D^\delta) \text{ such that } (\alpha, \beta, \gamma, \delta) \in [\![0, n]\!]^4 \text{ and } \alpha + \beta + \gamma + \delta = n\right\}$ is linearly independent.

Step 2 In the second step, let us consider the following set

$$\mathcal{E}_n^p = \left\{ (\alpha_1, \ldots, \alpha_p) \in [\![0, n]\!]^p \text{ such that } \sum_{i=1}^p \alpha_i = n \right\}.$$

A basis of $\mathbb{P}_n^d$ (polynomial of order n with d variables) is given by $\left(\prod_{i=1}^d x_i^{\alpha_i}\right)$ with $\sum_{i=1}^d \alpha_i \leq n$. Here, we have $\left(\prod_{i=1}^{d+1} \lambda_i^{\alpha_i}\right)$ with $\sum_{i=1}^{d+1} \alpha_i = n$. By noting that $\sum_{i=1}^d \alpha_i \leq n$ is equivalent to: there exists $c \leq n$ such that $\sum_{i=1}^d \alpha_i = n - c$, then it is clear that

$$\text{card}(\mathcal{E}_n^{d+1}) = \dim(\mathbb{P}_n^d) = \binom{n+d}{n}. \tag{A.2}$$

And in particular,

$$\text{card}(\mathcal{E}_n^4) = \frac{(n+1)(n+2)(n+3)}{6} = \dim(\mathbb{P}_n^3). \tag{A.3}$$

Below we prove the last equality of (A.2). First, we note that

$$\text{card}(\mathcal{E}_n^p) = \sum_{i=0}^n \text{card}(\mathcal{E}_i^{p-1}).$$

By noting that

$$\text{card}(\mathcal{E}_n^1) = 1,$$

then, it follows that

$$\begin{aligned}
\text{card}(\mathcal{E}_n^2) &= n+1, \\
\text{card}(\mathcal{E}_n^3) &= \frac{(n+1)(n+2)}{2}, \\
\text{card}(\mathcal{E}_n^4) &= \frac{(n+1)(n+2)(n+3)}{6}.
\end{aligned}$$

Consequently $\text{card}(\mathcal{E}_n^4)$ is exactly the dimension of $\mathbb{P}_n^3$. More generally, we show that

$$\text{card}(\mathcal{E}_n^{d+1}) = \binom{n+d}{n} = \frac{(n+d)!}{n!d!} = \dim(\mathbb{P}_n^d).$$

Combining the step 1 and the step 2 (in particular equality (A.3)) allows to conclude the proof. □

Below, we explain the main steps of the algorithm to find suitable additional functions.

1. **Step 1**: In order to complete the space $\mathbb{P}_{n+1}$, we generate all functions of degree $n+2$ that can be written as $\lambda_1^i\lambda_2^j\lambda_3^k\lambda_4^l$ with $i+j+k+l = n+2$ (see Lemma A.1). We store them in the list $AllFunctionsInSigma$.

2. **Step 2**: For all possible combinations of $n(n+2)$ functions among the list $AllFunctionsInSigma$ (for all possible Σ_{n+2}):

   - **Step 2.1**: We generate the corresponding space V_{n+1} by concatenating a basis of $\mathbb{P}_{n+1}$ with the considered basis of Σ_{n+2}.
   - **Step 2.2**: We test the unisolvence of the considered basis of V_{n+1}. To verify the unisolvence of the set of functions in V_{n+1} according to the set of degrees of freedom, we assemble the following matrix

     $$\mathcal{M}_{ij} = N_i(\varphi_j) \quad i,\ j \geq 0,$$

     for N_i in $\mathcal{N}_{n+1}$ and $\varphi_j \in V_{n+1}$. Then, we compute the rank of the matrix $\mathcal{M}$ (in this work, we use the NumPy linear algebra function `matrix_rank` [3]). If the matrix $\mathcal{M}$ is of full rank, i.e. invertible, the set of functions (φ_j) is unisolvent with respect to the set of degrees of freedom $\mathcal{N}_{n+1}$ (see Definition 5.3).

Results for the case $n = 1$ We discuss about the output of the algorithm for the case $n = 1$.

- At the first step of the algorithm, we generate the list of the 20 functions of degree three.
- Then, by considering 3 functions among 20, we obtain 1140 potential spaces Σ_3.
- By testing the unisolvence of the resulting space V_2, we find that 128 spaces Σ_3 are suitable to complete $\mathbb{P}_2$. These subspaces are presented in Appendix A.2.

Results for the case $n = 2$ We discuss about the output of the algorithm for the case $n = 2$.

- At the first step of the algorithm, we generate the list of 35 functions of degree 4.
- Then, by considering 8 functions among 35, we obtain 23,535,820 potential spaces Σ_4.
- By testing the unisolvence of the resulting space V_3, we find that 2,368,236 spaces Σ_4 are suitable to complete $\mathbb{P}_3$.

Conjecture for the case $n = 3$ For the case $n = 3$, we follow the same procedure. We generate first the list of the 56 basis functions of $\mathbb{P}_5(K)$. However, at this stage by considering 15 functions among the 56 functions to form the subspace Σ_4, the number of possible combinations is higher than 16,000 billion. Obviously, we cannot test the unisolvence of all resulting space V_4. We test randomly millions of possible space V_4, but as of today, no suitable space Σ_5 has been found. The existence (or not) of such basis is still open.

A.2 Suitable basis of Σ_3

In this section, we present the output of the procedure described in Appendix A.1 for the case $n = 1$, i.e. all the basis of the space Σ_3 that can be used to enrich $\mathbb{P}_2$ to build the space V_2.

1. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_4^2,\ \lambda_2^1\lambda_3^2)$
2. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_4^2,\ \lambda_2^2\lambda_3^1)$
3. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_3^2)$
4. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_2^2)$
5. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_4^2,\ \lambda_1^2\lambda_3^1)$
6. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_4^2,\ \lambda_1^2\lambda_2^1)$
7. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_2^2\lambda_4^1)$
8. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_4^2)$
9. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_2^2)$
10. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
11. $(\lambda_3^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_2^1)$
12. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_4^1,\ \lambda_2^2\lambda_3^1)$
13. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_3^2)$
14. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_2^2)$
15. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
16. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_2^1)$
17. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_4^2)$
18. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_2^2)$
19. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_4^1)$
20. $(\lambda_3^1\lambda_4^2,\ \lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_2^1)$
21. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_3^2)$
22. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_2^2)$
23. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_3^1)$
24. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_2^1)$
25. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^1\lambda_2^2)$
26. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
27. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_2^1)$
28. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_4^1)$
29. $(\lambda_3^1\lambda_4^2,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_3^1)$
30. $(\lambda_3^1\lambda_4^2,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
31. $(\lambda_3^1\lambda_4^2,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_2^1)$
32. $(\lambda_3^1\lambda_4^2,\ \lambda_1^2\lambda_3^1,\ \lambda_1^2\lambda_2^1)$
33. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_4^2,\ \lambda_2^1\lambda_3^2)$
34. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_4^2,\ \lambda_2^2\lambda_3^1)$
35. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_3^2)$
36. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_2^2)$
37. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_4^2,\ \lambda_1^2\lambda_3^1)$
38. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_4^2,\ \lambda_1^2\lambda_2^1)$
39. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_3^2,\ \lambda_2^2\lambda_4^1)$
40. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_4^2)$
41. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_2^2)$
42. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
43. $(\lambda_3^2\lambda_4^1,\ \lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_2^1)$
44. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_4^1,\ \lambda_2^2\lambda_3^1)$
45. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_3^2)$
46. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_2^2)$
47. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
48. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_2^1)$
49. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_4^2)$
50. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_2^2)$
51. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_4^1)$
52. $(\lambda_3^2\lambda_4^1,\ \lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_2^1)$
53. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_3^2)$
54. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_2^2)$

55. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_3^1)$
56. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_2^1)$
57. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^1\lambda_2^2)$
58. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
59. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_2^1)$
60. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_4^1)$
61. $(\lambda_3^2\lambda_4^1,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_3^1)$
62. $(\lambda_3^2\lambda_4^1,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
63. $(\lambda_3^2\lambda_4^1,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_2^1)$
64. $(\lambda_3^2\lambda_4^1,\ \lambda_1^2\lambda_3^1,\ \lambda_1^2\lambda_2^1)$
65. $(\lambda_2^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_4^2)$
66. $(\lambda_2^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_3^2)$
67. $(\lambda_2^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
68. $(\lambda_2^1\lambda_4^2,\ \lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_3^1)$
69. $(\lambda_2^1\lambda_4^2,\ \lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_4^2)$
70. $(\lambda_2^1\lambda_4^2,\ \lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_3^2)$
71. $(\lambda_2^1\lambda_4^2,\ \lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_4^1)$
72. $(\lambda_2^1\lambda_4^2,\ \lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_3^1)$
73. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_3^2)$
74. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_2^2)$
75. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_3^1)$
76. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_2^1)$
77. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^1\lambda_2^2)$
78. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
79. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_2^1)$
80. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_4^1)$
81. $(\lambda_2^1\lambda_4^2,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_3^1)$
82. $(\lambda_2^1\lambda_4^2,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
83. $(\lambda_2^1\lambda_4^2,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_2^1)$
84. $(\lambda_2^1\lambda_4^2,\ \lambda_1^2\lambda_3^1,\ \lambda_1^2\lambda_2^1)$
85. $(\lambda_2^1\lambda_3^2,\ \lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_4^2)$
86. $(\lambda_2^1\lambda_3^2,\ \lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_3^2)$
87. $(\lambda_2^1\lambda_3^2,\ \lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_4^1)$
88. $(\lambda_2^1\lambda_3^2,\ \lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
89. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_3^2)$
90. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_2^2)$
91. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_3^1)$
92. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_2^1)$
93. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^1\lambda_2^2)$
94. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
95. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_2^1)$
96. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_4^1)$
97. $(\lambda_2^1\lambda_3^2,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_3^1)$
98. $(\lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
99. $(\lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_2^1)$
100. $(\lambda_2^1\lambda_3^2,\ \lambda_1^2\lambda_3^1,\ \lambda_1^2\lambda_2^1)$
101. $(\lambda_2^2\lambda_4^1,\ \lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_4^2)$
102. $(\lambda_2^2\lambda_4^1,\ \lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_3^2)$
103. $(\lambda_2^2\lambda_4^1,\ \lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_4^1)$
104. $(\lambda_2^2\lambda_4^1,\ \lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_3^1)$
105. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_3^2)$
106. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_2^2)$
107. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_3^1)$
108. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_2^1)$
109. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^1\lambda_2^2)$
110. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
111. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_2^1)$
112. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_4^1)$
113. $(\lambda_2^2\lambda_4^1,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_3^1)$
114. $(\lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
115. $(\lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_2^1)$
116. $(\lambda_2^2\lambda_4^1,\ \lambda_1^2\lambda_3^1,\ \lambda_1^2\lambda_2^1)$
117. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_3^2)$
118. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^1\lambda_2^2)$
119. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_3^1)$
120. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_4^2,\ \lambda_1^2\lambda_2^1)$
121. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^1\lambda_2^2)$
122. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_4^1)$
123. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_3^2,\ \lambda_1^2\lambda_2^1)$
124. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_4^1)$
125. $(\lambda_2^2\lambda_3^1,\ \lambda_1^1\lambda_2^2,\ \lambda_1^2\lambda_3^1)$
126. $(\lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_3^1)$
127. $(\lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_4^1,\ \lambda_1^2\lambda_2^1)$
128. $(\lambda_2^2\lambda_3^1,\ \lambda_1^2\lambda_3^1,\ \lambda_1^2\lambda_2^1)$

Appendix B

Appendix of Chapter 6: Extension in case of connected obstacles which cut the border of the domain

Contents

In Chapter 6, we have considered that the obstacles do not touch the border of Ω for simplicity of presentation. However, the proof of Theorem 6.9 can be extended in the case in which the obstacles touch the border of Ω. This configuration may appear only when the obstacles are connected. We give in this chapter the main outlines of the proof. One of the main difference compared to the work presented in Chapter 6 is the definition of the domain Ω^ε when the obstacles are connected. Then the proof of Theorem 6.9 follows the same lines that those presented in Chapter 6 but with a different definition of the velocity estimator $\boldsymbol{u}_{\boldsymbol{\varepsilon},\mathbf{3}}$.

B.1 Definition of the open set Ω^ε in the case of connected obstacles which touch the border of Ω

Let Ω be a bounded and connected open set of $\mathbb{R}^d$ ($d \geq 2$). The set Ω is covered with a regular square mesh of size ε, each cell being a cube Y_i^ε. We define for any $i \in \mathbb{Z}^d$:

$$Y_i = Y + i, \quad Y_{i,F} = Y_F + i, \quad Y_{i,S} = Y_S + i,$$

$$Y_i^\varepsilon = \varepsilon(Y + i), \quad Y_{i,F}^\varepsilon = \varepsilon(Y_F + i), \quad Y_{i,S}^\varepsilon = \varepsilon(Y_F + i).$$

Define the set $\mathcal{I} = \{i \in \mathbb{Z}^d \ s.t. \ Y_i^\varepsilon \subset \Omega\}$ and the periodic set of perforations inside Ω as

$$B^\varepsilon = \bigcup_{i \in \mathcal{I}} Y_{i,S}^\varepsilon.$$

Finally, the fluid part Ω^ε of the perforated medium is defined by $\Omega^\varepsilon = \Omega \setminus B^\varepsilon$, i.e.

$$\Omega^\varepsilon = \Omega \setminus \bigcup_{i\in\mathcal{I}} Y^\varepsilon_{i,S} = \Omega \cap \bigcup_{i\in\mathcal{I}} Y^\varepsilon_{i,F}.$$

Remark that only entire obstacles $Y^\varepsilon_{i,S}$ are removed from Ω or, in other words, that no obstacles are cut by the boundary $\partial\Omega$.

Assumption B.1. If the solid part Y_S touches the boundary of Y, meaning that obstacles are connected, then Ω is assumed to be a parallelepiped, namely a polygonal open set with flat boundary faces which are each orthogonal to one of the unit canonical basis vector $\boldsymbol{e_k}$, $1 \le k \le d$, of $\mathbb{R}^d$ and that Ω has commensurable sizes which guarantees the existence of a sequence $(\varepsilon_n)_{n\in\mathbb{N}} \to 0$ such that $\overline{\Omega} = \bigcup_{i\in\mathcal{I}} \overline{Y_i^{\varepsilon_n}}$. In other words, Ω is the union of entire cells and its boundary $\partial\Omega$ does not cut any cell.

Remark B.2. Assumption B.1 is possible only for space dimensions $d \ge 3$ because in 2-d E_F and E_S cannot be connected simultaneously. The general case of any smooth domain in Assumption B.1 is much more tricky and we refer to [9] for more details. Indeed, when the solid part Y_S touches the boundary of Y and Ω is not the union of entire cells $\bigcup_{i\in\mathcal{I}} Y_i^\varepsilon$, there may be some pathological cut obstacles near the boundary $\partial\Omega$. Therefore, only local results, away from the boundary, can be obtained. For the sake of simplicity, this general case is not discussed here.

Although the fluid domain E_F is connected, Assumption B.1 does not ensure that the open set Ω^ε is connected (which allows us to define a unique pressure in Ω^ε, up to a single additive constant). Therefore, if Ω^ε is not connected, we remove the small connected components of Ω^ε touching the boundary $\partial\Omega$. For simplicity, we keep the notation Ω^ε for its main connected components. In other words, we now assume that Ω^ε is connected. Assumption B.1 allows us to treat the realistic case of a connected solid part and a diphasic boundary of the porous medium, i.e. $\partial\Omega \cap \overline{\Omega^\varepsilon} \neq \emptyset$ and $\partial\Omega \cap \overline{B^\varepsilon} \neq \emptyset$.

B.2 Proof of Theorem 6.9 in the case of connected obstacles which touch the border of Ω

In such a case, contrary to the work done in Chapter 6, it is not possible to build a cut-off function η^ε which varies in the ε-vicinity O^ε of $\partial\Omega$ and does not see any solid obstacles. The difficulty is that the second line of the previous velocity estimator (6.33) does not vanish on ∂B^ε in O^ε. To overcome this difficulty, we use the restriction operator R_ε defined in Lemma 6.15

and modify (6.33) as follows

$$\begin{aligned}\boldsymbol{u_{\varepsilon,3}}(x) =& R_\varepsilon\Big\{\frac{\varepsilon^2}{\nu}|Y_F|\langle\boldsymbol{\omega_i}\rangle\left(f_i(x)-\frac{\partial p^*}{\partial x_i}(x)\right)-\varepsilon^2\boldsymbol{v_{b\varepsilon}}\Big\}\\ &+R_\varepsilon\Big\{\frac{\varepsilon^4}{\nu}(-\Delta+\boldsymbol{\nabla}\,\mathrm{div}\,)\left(\eta^\varepsilon(x)\boldsymbol{\psi_i}\left(\frac{x}{\varepsilon}\right)\right)\Big\}\left(f_i(x)-\frac{\partial p^*}{\partial x_i}(x)\right)\\ &+\frac{\varepsilon^3}{\nu}\boldsymbol{\gamma_{ij}}\left(\frac{x}{\varepsilon}\right)\eta^\varepsilon(x)\frac{\partial}{\partial x_j}\left(f_i(x)-\frac{\partial p^*}{\partial x_i}(x)\right)-\boldsymbol{v_{d\varepsilon}}.\end{aligned}\tag{B.1}$$

The pressure estimator (6.34) is unchanged.

In particular, the next two lemmas show that the restriction operator is almost the identity when applied to functions which vanish on the solid obstacles outside from the ε-vicinity O^ε of $\partial\Omega$.

Lemma B.3. Let Q_ε be the orthogonal projection operator on $H_0^1(\Omega^\varepsilon)^d$, defined by (6.4.2). Let $\boldsymbol{u}\in\left(H_0^1(\Omega)\cap C^1(\Omega)\right)^d$ be such that $\boldsymbol{u}\equiv 0$ in $B^\varepsilon\setminus O^\varepsilon$. Then, there exists a constant C, which does not depend on ε, such that

$$\|\boldsymbol{\nabla}(Q_\varepsilon\boldsymbol{u}-\boldsymbol{u})\|_{L^2(\Omega^\varepsilon)}\le C\sqrt{\varepsilon}\|\boldsymbol{\nabla}\boldsymbol{u}\|_{L^\infty(\Omega)}.$$

Proof. Introduce a new cut-off function $\tilde\eta^\varepsilon$, in the same spirit as in Definition 6.25. Recall that $O^\varepsilon:=\{x\in\Omega:\mathrm{dist}(x,\partial\Omega)\le c\varepsilon\}$. The function $\tilde\eta^\varepsilon\in C^\infty(\overline{\Omega})$ is chosen such that $\tilde\eta^\varepsilon$ and all its derivatives vanish in O^ε, while $\tilde\eta^\varepsilon(x)=1$ on $\Omega\setminus O^{2\varepsilon}$ and

$$\|1-\tilde\eta^\varepsilon\|_{L^2(\Omega)}\le C\sqrt{\varepsilon};\quad \|\tilde\eta^\varepsilon\|_{L^\infty(\Omega)}=1;\quad \|\boldsymbol{\nabla}\tilde\eta^\varepsilon\|_{L^\infty(\Omega)}\le\frac{C}{\varepsilon}.$$

In the minimization property (6.24) we choose the test function $\boldsymbol{v}=\tilde\eta^\varepsilon\boldsymbol{u}$, implying that

$$\|\boldsymbol{\nabla}(Q_\varepsilon\boldsymbol{u}-\boldsymbol{u})\|_{L^2(\Omega^\varepsilon)}\le\|\boldsymbol{\nabla}(\tilde\eta^\varepsilon\boldsymbol{u}-\boldsymbol{u})\|_{L^2(\Omega^\varepsilon)}\le\|(\tilde\eta^\varepsilon-1)\boldsymbol{\nabla}\boldsymbol{u}\|_{L^2(O^{2\varepsilon})}+\|\boldsymbol{u}\boldsymbol{\nabla}\tilde\eta^\varepsilon\|_{L^2(O^{2\varepsilon})}.$$

It is easy to check that $\|(\tilde\eta^\varepsilon-1)\boldsymbol{\nabla}\boldsymbol{u}\|_{L^2(O^{2\varepsilon})}\le C\sqrt{\varepsilon}\|\boldsymbol{\nabla}\boldsymbol{u}\|_{L^\infty(\Omega)}$. Using for $\boldsymbol{u}$ the same type of Poincaré-type argument as in Lemma 6.21, one obtains that $\|\boldsymbol{u}\boldsymbol{\nabla}\tilde\eta^\varepsilon\|_{L^2(O^{2\varepsilon})}\le C\sqrt{\varepsilon}\|\boldsymbol{\nabla}\boldsymbol{u}\|_{L^\infty(\Omega)}$, thus proving the lemma. □

Lemma B.4. Let R_ε be the linear operator on $H_0^1(\Omega^\varepsilon)^d$, defined by Lemma 6.15. Let $\boldsymbol{u}\in\left(H_0^1(\Omega)\cap C^1(\Omega)\right)^d$ be such that $\boldsymbol{u}\equiv 0$ in $B^\varepsilon\setminus O^\varepsilon$. Then, there exists a constant C, which does not depend on ε, such that

$$\|R_\varepsilon\boldsymbol{u}-\boldsymbol{u}\|_{L^2(\Omega^\varepsilon)}+\varepsilon\|\boldsymbol{\nabla}(R_\varepsilon\boldsymbol{u}-\boldsymbol{u})\|_{L^2(\Omega^\varepsilon)}\le C\varepsilon^{3/2}\|\boldsymbol{\nabla}\boldsymbol{u}\|_{L^\infty(\Omega)}.\tag{B.2}$$

Proof. In view of system (6.25), if $\boldsymbol{u}\in H^1(Y)^d$ satisfies $\boldsymbol{u}\equiv 0$ in Y_S, we deduce that $R_i\boldsymbol{u}-\boldsymbol{u}$ satisfies a homogeneous Stokes equation with a boundary condition on ∂Y depending on $Q_i\boldsymbol{u}-\boldsymbol{u}$,

which implies that there exists a constant C such that

$$\|R_i\boldsymbol{u}-\boldsymbol{u}\|^2_{L^2(Y_F)}+\|\boldsymbol{\nabla}(R_i\boldsymbol{u}-\boldsymbol{u})\|^2_{L^2(Y_F)}\leq C\|\boldsymbol{\nabla}(Q_i\boldsymbol{u}-\boldsymbol{u})\|^2_{L^2(Y_F)},$$

where we applied Poincaré inequality to $(Q_i\boldsymbol{u}-\boldsymbol{u})$ which vanishes in Y_S. After scaling and summation on all cells Y_i^ε, recalling that Q_ε and Q_i are related by (6.24), and using Lemma B.3, we deduce (B.2). □

Then using the estimator (B.1), and Lemmas B.3 and B.4, we are able to extend Lemma 6.27 to the case of connected obstacles which touch the border.

Lemma B.5. There exists $\boldsymbol{v_{d\varepsilon}}\in H^1_0(\Omega^\varepsilon)^d$, such that the velocity estimator $\boldsymbol{u_{\varepsilon,3}}$, defined by (B.1), satisfies

$$\boldsymbol{u_{\varepsilon,3}}=\varepsilon^2\boldsymbol{u_2}+\varepsilon^3\boldsymbol{u_3}-\boldsymbol{v_{d\varepsilon}}+\boldsymbol{r_\varepsilon}\quad\text{in }\Omega\setminus O^\varepsilon$$

where $\boldsymbol{r_\varepsilon}$ is a small remainder term satisfying

$$\|\boldsymbol{r_\varepsilon}\|_{L^2(\Omega^\varepsilon)}+\varepsilon\|\boldsymbol{\nabla r_\varepsilon}\|_{L^2(\Omega^\varepsilon)}\leq C\varepsilon^{3/2}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{C^{1,\alpha}(\overline{\Omega})},$$

$$\boldsymbol{u_{\varepsilon,3}}=\boldsymbol{0}\quad\text{on }\partial\Omega,$$

$$\boldsymbol{u_{\varepsilon,3}}=\boldsymbol{0}\quad\text{on }\partial B^\varepsilon,$$

$$\operatorname{div}\boldsymbol{u_{\varepsilon,3}}=0\quad\text{in }\Omega$$

and

$$|\boldsymbol{v_{d\varepsilon}}|_{H^1(\Omega^\varepsilon)}\leq C\varepsilon^{\frac{3}{2}}\|\boldsymbol{f}-\boldsymbol{\nabla}p^*\|_{H^2(\Omega)\cap C^{1,\alpha}(\overline{\Omega})}$$

with C independent of ε.

Then Lemma 6.28 remains true with a little more technical proof, and the rest of the proof follows the same lines.

Appendix C

Appendix of Chapter 8: Some results about quadrature rules

Contents

This appendix is dedicated to the proof of some results concerning quadrature rules. In particular, in Appendix C.1 we prove an exact quadrature rule for quadratic functions on triangles and in Appendix C.2 we explain why we cannot show the discrete well-posedness of the local problems for higher order MsFEM with usual pairs of finite elements.

C.1 Exact quadrature rule for quadratic function on a triangle

Lemma C.1. Let K be a non degenerated triangle. Let A^1, A^2 and A^3 be the three vertices of K and $A^{1,2}$ the middle of A^1A^2, $A^{1,3}$, the middle of A^1A^3, and $A^{2,3}$ the middle of A^2A^3. Let Q be a quadratic function, then the following exact quadrature rule holds

$$\int_K Q(\boldsymbol{x})\mathrm{d}x = \frac{|K|}{3}\left(Q(A^{1,2}) + Q(A^{1,3}) + Q(A^{2,3})\right).$$

In this section, we prove Lemma C.1. The proof is done in two steps. In the first step, we determine the Lagrange polynomials associated to the Lagrange interpolation points. In the second step, we derive the quadrature rule using these polynomials.

C.1.1 Definition of the Lagrange polynomials

Let us consider a triangle K with the barycentric coordinates λ_1, λ_2 and λ_3. We consider the three vertices A^1 $(1,0,0)$, A^2 $(0,1,0)$ and A^3 $(0,0,1)$, localized with the barycentric coordinates and let $A^{1,2}$ $(\frac{1}{2},\frac{1}{2},0)$ the middle of A^1A^2, $A^{1,3}$ $(\frac{1}{2},0,\frac{1}{2})$, the middle of A^1A^3, and $A^{2,3}$ $(0,\frac{1}{2},\frac{1}{2})$ the middle of A^2A^3. It is well-known that the family $\{\lambda_1^2,\ \lambda_2^2,\ \lambda_3^2,\ \lambda_1\lambda_2,\ \lambda_1\lambda_3,\ \lambda_2\lambda_3\}$ is a basis of $\mathbb{P}_2$. Now, we propose to find the Lagrange polynomials associated with the Lagrange interpolation points, i.e. the three vertices A^i and the three middle of edges $A^{i,j}$. First, we note that everything is invariant by permutation of the vertices, this is sufficient to construct for example the Lagrange polynomial corresponding to A^1 and that corresponding to $A^{1,2}$.

Let us start with the polynomial p_1 associated with the node A^1. We need a polynomial $p_1 \in \mathbb{P}_2$ such that $p_1(A^1) = 1$ and p_1 vanishes at all the other vertices. First, p_1 is of degree at most two on the line (A^2, A^3) with three roots A^2, A^3 and $A^{2,3}$, thus it vanishes on (A^2, A^3). The equation of the straight line (A^2, A^3) is $\lambda_1 = 0$, hence p_1 is divisible by λ_1. There exist a polynomial q such that $p_1 = q\lambda_1$. Given that λ_1 is of degree one, q is of degree at most one. Moreover since $\lambda_1(A^{1,2}) = \lambda_1(A^{3,1}) = \frac{1}{2} = \frac{1}{2} \neq 0$ we have $q(A^{1,2}) = q(A^{1,3}) = 0$. Thus q vanishes along the line $(A^{1,2}, A^{1,3})$ of equation $\lambda_1 - \frac{1}{2} = 0$. Thus q is divisible by $\lambda_1 - \frac{1}{2}$, so that $q = c(\lambda_1 - \frac{1}{2})$ with c a constant. Finally, given that $p_1(A^1) = 1$, we have $1 = \frac{C}{2}$, hence $p_1 = \lambda_1(2\lambda_1 - 1)$. Finally, we have

$$p_1 = \lambda_1(2\lambda_1 - 1), \quad p_2 = \lambda_2(2\lambda_2 - 1), \quad p_3 = \lambda_3(2\lambda_3 - 1).$$

Now, we consider the polynomial $p_{1,2}$ associated with the node $A^{1,2}$. The polynomial vanishes on both lines (A^1, A^3) and (A^2, A^3) hence $p_{1,2} = c\lambda_1\lambda_2$ where c is a constant. Using $p_{1,2}(A^{1,2}) = 1$ gives $c = 4$. Finally, we have

$$p_{1,2} = 4\lambda_1\lambda_2, \quad p_{1,3} = 4\lambda_1\lambda_3, \quad p_{2,3} = 4\lambda_2\lambda_3.$$

C.1.2 Derivation of the quadrature formula

In this part, we denote the vertices of the triangle K with indices $1-3$ and the middle point of the edges with indices $4-6$. We call p_i the Lagrange polynomial associated with the Lagrange interpolation point i. First, every quadratic function can be uniquely expressed in the Lagrange basis as

$$Q(\boldsymbol{x}) = \sum_{i=1}^{6} \alpha_i p_i(\boldsymbol{x}).$$

Since the p_i is the Lagrange basis functions, we have

$$Q(\boldsymbol{r_j}) = \sum_{i=1}^{6} \alpha_i p_i(\boldsymbol{r_j}) = \sum_{i=1}^{6} \alpha_i \delta_{i,j} = \alpha_j, \tag{C.1}$$

with $\boldsymbol{r_j}$ the Lagrange interpolation point associated with Lagrange polynomial p_j. We can thus rewrite (C.1) as

$$Q(\boldsymbol{x}) = \sum_{i=1}^{6} Q(\boldsymbol{r_i}) p_i(\boldsymbol{x}).$$

Then, the integration over the triangle K reads

$$\int_K Q(\boldsymbol{x})\mathrm{d}x = \sum_{i=1}^{6} Q(r_i) \int_K p_i(\boldsymbol{x})\mathrm{d}x = \sum_{i=1}^{6} f(\boldsymbol{r_i})\omega_i.$$

We have seen above that the $p_i = \lambda_i(2\lambda_i - 1)$ and the $p_{i,j} = 4\lambda_i\lambda_j$. Thus by symmetry we have $\omega_1 = \omega_2 = \omega_3$ and $\omega_4 = \omega_5 = \omega_6$, and it is sufficient to compute

$$\omega_i = \int_K \lambda_i(2\lambda_i - 1)\mathrm{d}x = 0,$$

and

$$\omega_{i,j} = \int_K 4\lambda_i\lambda_j\mathrm{d}x = \frac{|K|}{3},$$

using the following exact integration formula for product of barycentric coordinates on polyhedron [163],

$$\int_K \prod_{i=1}^{d+1} \lambda_i^{\alpha_i} = |K| d! \frac{\sum_{i=1}^{d+1} \alpha_i!}{\left(\sum_{i=1}^{d+1} \alpha_i + d\right)!}.$$

Finally, we have

$$\int_K Q(\boldsymbol{x})\mathrm{d}x = \frac{|K|}{3}\left(Q(A^{1,2}) + Q(A^{1,3}) + Q(A^{2,3})\right).$$

C.2 High order quadrature rules on triangles

In this section, we explain more in details why we cannot prove the well-posedness of the discrete local problems with Taylor–Hood finite elements. This section is based of the results about Gauss-Lobato quadrature rules presented in [170, 94]. Indeed, the discretization of the local problems leads to integrals of type

$$\int_T \operatorname{div} \boldsymbol{v_h}\ p_h = \int_T \lambda\ \boldsymbol{\varphi_T} \cdot \boldsymbol{v_h}$$

with $\boldsymbol{\varphi_T} \in (\mathbb{P}_{n-1}(T))^d$. Now, if we choose $\boldsymbol{v_h} \in \mathbb{P}_{n+1}(T)$ and $p_h \in \mathbb{P}_n(T)$, it follows that the terms $\operatorname{div} \boldsymbol{v_h}\ p_h$ and $\boldsymbol{\varphi_T} \cdot \boldsymbol{v_h}$ are polynomials of order $2n$. Therefore, to show the discrete well-posedness of the local problems, the objective is to find a quadrature formula for polynomials of degree $2n$, i.e. for polynomials in $\mathbb{P}_{2n}(T)$, using the Lagrange points of polynomials of degree $n+1$, i.e. in $\mathbb{P}_{n+1}(T)$.

Remark C.2. In two dimensions on triangles, we have seen that if $n = 1$, such a quadrature rule exists, since this is equivalent to integrate a quadratic polynomial with the Lagrange points of polynomials of degree 2 (see Lemma C.1), hence the well-posedness of the local problems with the pair of finite element $\mathbb{P}_2 - \mathbb{P}_1$ for the two-dimensional MsFEM in the case $n = 1$.

C.2.1 Theoreticals results

[94] shows that it is impossible to integrate polynomials of degree $(2n+1)$ with $\frac{(n+2)(n+3)}{2}$ points. By recalling that $\dim(\mathbb{P}_n(T)) = \frac{(n+1)(n+2)}{2}$, it is consequently impossible to integrate polynomials of degree $(2n+1)$ with the Lagrange interpolation points of polynomials in $\mathbb{P}_{n+1}(T)$. Nevertheless, in the present case, we need to integrate polynomials of degree $2n$ with $\frac{(n+2)(n+3)}{2}$ points. In what follows, we check if such a quadrature rule can exist.

First, we recall that the structure of the space $\mathbb{P}_{n+1}(T)$ is as follows: three points at the vertices, n points on the edges and $\frac{n(n-1)}{2}$ points in the element. [170] proposes a method to build quadrature rules of the form

$$\int_\Omega Q(x,y)\mathrm{d}x\mathrm{d}y = \sum_{k=1}^{N} \lambda_k Q(x_k, y_k), \tag{C.2}$$

with Q a polynomial. [170] gives also the lower bound N for such a quadrature rule, i.e.

$$N \geq \begin{cases} \frac{(n+1)(n+2)}{2} + \lfloor \frac{(n+1)}{2} \rfloor & \text{if } \deg(Q) = 2n+1, \\ \frac{(n+1)(n+2)}{2} & \text{if } \deg(Q) = 2n. \end{cases}$$

It should be noted that the minimal quadrature rule could require more points that the lower bound N. We re-write (C.2) as

$$\begin{aligned}\int_\Omega Q(x,y)\mathrm{d}x\mathrm{d}y = &\sum_{k=1}^{N_K} \lambda_{k,0} Q(x_{k,0}, y_{k,0}) + \sum_{k=1}^{N_{E_1}} \lambda_{k,1} Q(x_{k,1}, 0) + \sum_{k=1}^{N_{E_2}} \lambda_{k,2} Q(0, y_{k,2}) \\ &+ \sum_{k=1}^{N_{E_3}} \lambda_{k,3} Q(x_{k,0}, 1-x_{k,3}) + \mu_0 f(0,0) + \mu_1 f(1,0) + \mu_2 f(0,1)\end{aligned}$$

with $N = N_K + N_{E_1} + N_{E_2} + N_{E_3} + 3$ where N_K is the number of integration points in the element K, N_{E_i} the number of integration points on each edge E_i. According to [94], such a quadrature rule does not exist if $\deg(Q) = 2n+1$, $N_{E_1} = N_{E_2} = N_{E_3}$ and $N_K = \frac{(n-1)n}{2}$ (it is impossible to integrate a polynomial Q of degree $(2n+1)$ with the Lagrange interpolation points of polynomials in $\mathbb{P}_{n+1}(T)$). The main result of [170] is that if such a quadrature rule exists (a necessary condition but non sufficient) with $\deg(Q) = 2n-1$ or $\deg(Q) = 2n$, then

$$N_K \geq \begin{cases} \frac{n(n-1)}{2} & \text{if } \deg(Q) = 2n-1, \\ \frac{n(n-1)}{2} + \lfloor \frac{(n-1)}{2} \rfloor & \text{if } \deg(Q) = 2n, \end{cases}$$

and

$$N_K + N_{E_i} \geq \begin{cases} \frac{n(n-1)}{2} + \lfloor \frac{(n-1)}{2} \rfloor & \text{if } \deg(Q) = 2n-1, \\ \frac{n(n+1)}{2} & \text{if } \deg(Q) = 2n. \end{cases}$$

In the following section, give some examples to illustrate this rule.

C.2.2 Some illustrative examples

First, let us consider $n = 2$, then if a quadrature rule, allowing to integrate polynomials of degree 3 or 4 with the Lagrange interpolation points of $\mathbb{P}_3$, exists, then N_K and N_{E_i} should verify

$$N_K \geq \begin{cases} 1 & \text{if } \deg(Q) = 3, \\ 1 & \text{if } \deg(Q) = 4, \end{cases}$$

and

$$N_K + N_{E_i} \geq \begin{cases} 1 + \lfloor \frac{1}{2} \rfloor = 1 & \text{if } \deg(Q) = 3, \\ \frac{2\times 3}{2} = 3 & \text{if } \deg(Q) = 4. \end{cases}$$

For the space $\mathbb{P}_3$, $N_K = 1$ and $N_{E_i} = 2$, then such a quadrature rule to integrate polynomials in $\mathbb{P}_4$ with the Lagrange interpolation points of $\mathbb{P}_3$ may exists.

Let us consider now $n = 3$, then if a quadrature rule, allowing to integrate polynomials of degree 5 or 6 with the Lagrange interpolation points of $\mathbb{P}_4$, exists then N_K and N_{E_i} should verify

$$N_K \geq \begin{cases} 3 & \text{if } \deg(Q) = 5, \\ 3 + 1 = 4 & \text{if } \deg(Q) = 6, \end{cases}$$

and

$$N_K + N_{E_i} \geq \begin{cases} 3 + \lfloor \frac{2}{2} \rfloor = 4 & \text{if } \deg(Q) = 5, \\ 6 & \text{if } \deg(Q) = 6. \end{cases}$$

For the space $\mathbb{P}_4$, $N_K = 3$ and $N_{E_i} = 3$. Then, such a quadrature rule to integrate polynomials in $\mathbb{P}_5$ with the Lagrange interpolation points of $\mathbb{P}_4$ may exists but a quadrature rule to integrate polynomials in $\mathbb{P}_6$ with the Lagrange interpolation points of $\mathbb{P}_4$ may not exist.

Consequently, in view of the results presented in this section, it may be possible to show that the local problems for the high-order MsFEM in the case $n = 2$ in two dimensions is well-posed with the pair of finite element $\mathbb{P}_3 - \mathbb{P}_2$. However, such a quadrature rule has not been found and the existence of such a quadrature rule is not guaranteed. Besides, we not pursue in this way to prove the discrete well-posedness since for higher order, let say $n = 3$, if we choose $\boldsymbol{v_h} \in \mathbb{P}_4$ and $p_h \in \mathbb{P}_3$, it follows that the quantities $\boldsymbol{\varphi_T} \cdot \boldsymbol{v_h}$ are polynomial of degree 6, and we have just seen that we cannot integrate a polynomial of degree 6 with the Lagrange interpolation points of polynomials of degree 4.

Appendix D

Preconditioning strategy for the computation of reference solutions in two dimensions

Contents

Note: We would like to thank Pierre Jolivet for his help with the development of solvers for the reference solutions.

For the sake of comparison with the MsFEM approximations, in absence of analytical solutions, we compute reference solutions on reference meshes. In this appendix, we explain the computation of the reference solutions for the Stokes and the Oseen problems in two-dimensional heterogeneous domains.

The finite element discretisation of these problems leads to linear systems of equations in the form

$$\mathcal{A}x = b, \tag{D.1}$$

where $\mathcal{A} \in \mathbb{R}^{n\times n}$ is sparse, b is the right-hand side, and x the sought solution. Given that the fine meshes considered, consist of millions of fine elements, the sparse matrix $\mathcal{A}$ resulting from the discretisation of these problems is very large. A first class of method to solve such

problems are methods based on direct factorisation of the sparse matrix such as the Cholesky or LU factorisations. However, these direct methods, although robust and reliable, are mainly sequential, and require consequently a very large amount of memory resources, often intractable. Another popular class of methods is the iterative methods, relying on Krylov subspaces. These methods require only a sparse matrix vector multiplication per iteration and they are highly parallelizable. Among these methods, we can cite for example the Generalized Minimal Residual (GMRES) [152], the Conjugate Gradient (CG) [129] or the Minimal Residual (MINRES) method [142]. However, the convergence of these methods relies heavily on the properties of the matrix $\mathcal{A}$. For example, if $\mathcal{A}$ is symmetric positive definite (SPD), the $\mathcal{A}$-norm of the error at the kth iteration of CG algorithm, satisfies the following inequality

$$\|\mathcal{A}^{-1}b - x_k\|_{\mathcal{A}} \leq 2\|\mathcal{A}^{-1}b - x_0\|_{\mathcal{A}} \left(\frac{\sqrt{\kappa_2(\mathcal{A})} - 1}{\sqrt{\kappa_2(\mathcal{A})} + 1} \right)^k$$

where $\kappa_2(\mathcal{A})$, the condition number of $\mathcal{A}$, i.e. the ratio between largest and smallest eigenvalues of $\mathcal{A}$.

In particular, in the cases considered in this thesis, the discretisation of the Stokes or the Oseen problems in highly heterogeneous domains with lots of obstacles, leads to very ill-conditioned matrix, i.e. matrix with high condition number. Consequently, applying iterative solvers directly on system (D.1) is not efficient to solve these problems. To overcome this limitation, we use a preconditioning strategy, which consists to multiply both sides of (D.1), for example from the left by a matrix M^{-1}, such that the condition number of the resulting matrix $M^{-1}\mathcal{A}$ is small.

In what follows, we explain the preconditioning strategy to compute the reference solutions for the Stokes and the Oseen problems in two-dimensional heterogeneous domains. This preconditioning strategy is based mainly on two level Schwarz type Domain Decomposition (DD) associated with the GenEO (Generalized eigenvalue problem in the overlap) method.

D.1 Two level Schwarz type Domain Decomposition (DD) with GenEO for symmetric positive definite (SPD) matrices

Domain decomposition methods [65] are one of the dominant paradigms for defining efficient and robust preconditioners in modern large-scale applications. Besides, it is well-known that one level method scales poorly when the number of subdomains used increases. That is why, in this section we present a two level Schwarz type Domain Decomposition (DD) [65] associated with the GenEO (Generalized eigenvalue problem in the overlap) method [157, 115]. These types of domain decomposition methods are based on the two level method, the overlapping additive Schwarz preconditioner, which consists in a coarse space and a formulation of how this coarse space is incorporated into the domain decomposition method. Here, the coarse space is derived using the GenEO method, taking advantage that we deal with positive definite matrices, the GenEO methods having been developed for SPD matrices. The GenEO coarse space is constructed automatically solving a generalized eigenproblem on each subdomain.

Let $\mathcal{Z}$ be a symmetric positive definite (SPD) matrix of size $n \times n$. Then, the preconditioning

of $\mathcal{Z}$ using a two level Schwarz type Domain Decomposition (DD) method reads (with an additive correction)

$$M^{-1}_{\mathcal{Z},\text{additive}} = R_H^T \mathcal{Z}_H^{-1} R_H + \sum_{j=1}^{N} R_j^T \mathcal{Z}_j^{-1} R_j,$$

where

$$\mathcal{Z}_H = R_H \mathcal{Z} R_H^T \text{ and } \mathcal{Z}_j = R_j \mathcal{Z} R_j^T.$$

N is the number of subdomains, R_H is a full rank matrix of size $\dim(V_H) \times n$ where V_H (the coarse space) denotes the space spanned by the columns of R_H^T. However, more numerically efficient corrections can be chosen, such as the deflated or balanced corrections defined respectively by

$$M^{-1}_{\mathcal{Z},\text{deflated}} = R_H^T \mathcal{Z}_H^{-1} R_H + \left(\sum_{j=1}^{N} R_j^T \mathcal{Z}_j^{-1} R_j \right) (I - \mathcal{Z} R_H^T \mathcal{Z}_H^{-1} R_H), \tag{D.2}$$

$$M^{-1}_{\mathcal{Z},\text{balanced}} = R_H^T \mathcal{Z}_H^{-1} R_H + (I - \mathcal{Z} R_H^T \mathcal{Z}_H^{-1} R_H) \left(\sum_{j=1}^{N} R_j^T \mathcal{Z}_j^{-1} R_j \right) (I - \mathcal{Z} R_H^T \mathcal{Z}_H^{-1} R_H). \tag{D.3}$$

The action of each $\{R_j \mathcal{Z} R_j^T\}_i^{-1}$ can be parametrized through a local KSP. In order to construct the coarse space, V_H, we use local generalised eigenvalue problems in each subdomains allowing to select suitable coarse vectors that satisfy certain local stability estimates. Indeed, using the GenEO method [157, 115], it is possible to fix in advance two constants $0 < \lambda_m < 1 \leq \lambda_M$ and then, build a coarse space V_H such that $M_{\mathcal{Z}}^{-1}$ is spectrally equivalent to $\mathcal{Z}^{-1}$

$$\frac{1}{\lambda_M} M_{\mathcal{Z}}^{-1} \leq \mathcal{Z}^{-1} \leq \frac{1}{\lambda_m} M_{\mathcal{Z}}^{-1}. \tag{D.4}$$

The dimension of the coarse space V_H is proportional to the number of subdomains N. For each subdomain $1 \leq i \leq N$, let D_i be a non negative diagonal matrix that defines a discrete partition of unity, i.e.

$$\sum_{i=1}^{N} R_i^T D_i R_i = I$$

and $\mathcal{Z}_i^{Neu}$ be a symmetric semi-definite positive matrix such that for the maximum multiplicity of the intersection of subdomains denoted k_1, we have

$$\sum_{i=1}^{N} R_i^T \mathcal{Z}_i^{Neu} R_i \leq k_1 \mathcal{Z}.$$

To build the Geneo coarse space, the eigenvalue problem to be solved, local to each subdomain, reads as follows. Find $(\lambda_{i_k}, V_{i_k}) \in \mathbb{R} \times \mathbb{R}^{rank(R_i)}$ solution to

$$(D_i R_i \mathcal{Z} R_i^T D_i) V_{i_k} = \lambda_{i_k} \mathcal{Z}_i^{Neu} V_{i_k}. \tag{D.5}$$

Let $\tau > 0$ be a positive threshold, the coarse space is the vector space spanned by the vector $R_i^T D_i V_{i_k}$, $\forall \lambda_{i_k} > \tau$. Then, inequality (D.4) holds with $\lambda_m = (1 + k_1 \tau)^{-1}$ and $\lambda_M = k_0$ where k_0

is the maximal number of neighbours.

D.2 Computation of Stokes reference solutions in two dimensions

D.2.1 Problem definition

The steady state Stokes problem with homogeneous Dirichlet boundary conditions in a perforated domain is to find the velocity $\boldsymbol{u_\varepsilon} : \Omega^\varepsilon \to \mathbb{R}^2$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\begin{cases} -\nu\Delta\boldsymbol{u_\varepsilon} + \boldsymbol{\nabla} p_\varepsilon &= \boldsymbol{f} \quad \text{in } \Omega^\varepsilon, \\ \operatorname{div} \boldsymbol{u_\varepsilon} &= 0 \quad \text{in } \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &= \boldsymbol{0} \quad \text{on } \partial B^\varepsilon \cap \partial\Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &= \boldsymbol{0} \quad \text{on } \partial\Omega \cap \partial\Omega^\varepsilon, \end{cases} \tag{D.6}$$

with $\boldsymbol{f}$ the applied force and $\nu > 0$ the viscosity. The analysis of the Stokes problem has been extensively presented through different sections of this thesis. The readers may refer to Chapter 4, Chapter 5 or Chapter 7.

D.2.2 Efficient preconditioning strategy for the Stokes equations

D.2.3 Stokes problem in matrix form

It is well-known that the discretisation of the Stokes problem (D.6) leads to a linear system featuring a two-by-two block matrix. Let m and n be two integers with $m < n$. Let A be a sparse $n \times n$ matrix, B be a sparse $n \times m$ full-rank matrix of constraints, and C an $m \times m$ matrix (in particular $C = 0$ is allowed). From this, a saddle point matrix is built as

$$\mathcal{A} = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}.$$

We also define

$$b = \begin{pmatrix} f \\ g \end{pmatrix},$$

where f is the source term and g is the value of the divergence (g is usually equal to 0 for the Stokes problem since we consider a divergence-free velocity). The objective is thus to efficiently solve the following linear system

$$\mathcal{A}x = b, \tag{D.7}$$

where x is the vector of unknowns. In the case of the Stokes system, $x = \begin{pmatrix} u \\ p \end{pmatrix}$, with u the velocity unknowns and p the pressure unknowns. Below, we introduce a preconditioning strategy based on a decomposition of the problem by blocks with respect to the physics, using the fieldsplit strategy introduced in [42]. In particular, we use the same strategy as presented in [138] for saddle point problems, taking the advantage of the two by two blocks structure of the Stokes

problem. Indeed, the decomposition

$$\mathcal{A} = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ BA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & -(C + BA^{-1}B^T) \end{pmatrix} \begin{pmatrix} I & A^{-1}B^T \\ 0 & I \end{pmatrix}$$

shows that solving the system with $\mathcal{A}$ can be performed by solving sequentially linear systems with A and one with the Schur complement $S = C + BA^{-1}B^T$. Consequently, we need a preconditioner for A and a preconditioner for S. The development of efficient preconditioners for A and for S, called respectively M_A and N_S, is the main topic of what follows. These preconditioners are based on the methods presented in Appendix D.1. Then, we combine M_A and N_S to define a compact way to precondition the parallel saddle point problem.

D.2.3.1 Preconditioning of the matrix A and S

The preconditioners for A and S, denoted respectively by M_A and N_S, are based on the two level additive Schwarz method as presented in Appendix D.1. The matrix A is preconditioned considering $\mathcal{Z} = A$ in Appendix D.1. The development of a preconditioner for the Schur complement S, defined by

$$S = C + BA^{-1}B^T$$

is more technical. We do not describe all the procedure. As presented in [138], the final preconditioner N_S for S reads

$$N_S = S_0 + M_{S_1}$$

where S_0 is a coarse correction and M_{S_1} is a preconditioner based on two-level domain decomposition, whose expression can be found in [138].

D.2.4 Implementation details

The Stokes problem is discretized using classical Taylor–Hood finite element (the two components of the velocity are discretized with $\mathbb{P}_2$ finite element and the pressure is discretized with $\mathbb{P}_1$ finite element) and solved with the open-source parallel finite element software FreeFEM [93]. The domain decomposition algorithm based on GenEO, presented above, is implemented in PETSc [27] and in particular in the library HPDDM [116], interfaced in FreeFEM. As is usual in domain decomposition methods, we assign one subdomain per MPI process. For the ease of post-processing (comparison with MsFEM approximations), we use the same partitioning as used in the MsFEM approach, i.e. a subdomain corresponds to a coarse element (a coarse triangle in two dimensions or a coarse tetrahedron in three dimensions). If required, each coarse element is partitioned again in subdomains using an automatic graph partitioner such as METIS [118] or in contrary, several coarse elements can be glued together to form a new larger subdomain. In practice, in order to use GenEO from PETSc, the auxiliary operators $\{A_i^{Neu}\}_{1 \leq i \leq N}$ (see (D.5), also known as Neumann matrices, representing the discretization of the problem on the extended local subdomain with overlap, endowed with natural boundary conditions) and the corresponding local to global map of degrees of freedom have to be provided. However, in case where the matrix is constructed using the PETSc discretisation infrastructure (as in this work), the relevant information are automatically passed from FreeFEM to PETSc. The GenEO

eigenvalue problems are solved with SLEPc [100]. The default number of eigenvectors `-esp_nev` used to build the coarse space is 20. The dimension of the coarse space is consequently, by default, $20 \times N$ (with N the number of subdomains) under the assumption that all concurrent SLEPc solves converge. The default coarse correction is the deflated ones (D.2). The precision of the computations depends on the relative tolerance, `-ksp_rtol`, based on the L_2-norm of the residuals. Convergence is detected at iteration k if

$$\|r_k\|_2 > \max(rtol\|b\|_2, atol)$$

where $r_k = b - Ax_k$. By defaut, $rtol = 10^{-5}$ and $atol = 10^{-50}$.

D.2.4.1 Final structure of the solver

Now, we summarise the proposed methodology for solving iteratively (D.7) and detail the associated PETSc instructions for the sake of reproducibility. We use the Flexible Generalized Minimal Residual algorithm (FGMRES, `-ksp_type fgmres`) [151], which iteratively solves (D.7) using full Schur complement as outer preconditioner on the right, taking advantage that FGMRES allows the preconditioner to change at each iteration. Indeed, the inner Krylov solver for both physical blocks, performs only a few iterations (the relative tolerance on this solver is high), which justifies the use of the flexible GMRES algorithm as outer Krylov solver. The first options to initialise this solver are:

```
-ksp_type fgmres
-ksp_rtol 1.0E-10
-pc_type fieldsplit
-pc_type_fieldsplit schur
-pc_type_fieldsplit_schur_precondition self
```

The linear solves require inner velocity and pressure solves, involving respectively matrices A and S, which are themselves solved iteratively using the preconditioners M_A and N_S defined above.

The options used to implement the preconditioner N_S for the Schur complement are:

```
-prefix_push fieldsplit_1_
-ksp_rtol 1.0E-1
-ksp_type fgmres
-pc_type hpddm
-pc_hpddm_has_neumann
-pc_hpddm_ksp_rtol 1.0E-2
-pc_hpddm_schur_precondition geneo
-prefix_push pc_hpddm_levels_1_
-eps_nev 30
-eps_threshold 0.4
-st_share_sub_ksp
-eps_gen_non_hermitian
-sub_pc_factor_mat_solver_type mumps
-sub_pc_type lu
-eps_tol 1.0E-2
-prefix_pop
-prefix_push pc_hpddm_coarse
-correction balanced
-pc_type cholesky
-pc_factor_mat_solver_type mumps
-mat_filter 1.0E-15
-prefix_pop
-prefix_pop
```

The inner Krylov solver for the Schur complement block is FGMRES (`-ksp_type fgmres`). The relative tolerance (`-ksp_rtol`) of the outer Krylov method is set to 10^{-1} (only a few iterations are required). The threshold τ_{S_1} (`-eps_threshold`) for selecting the local eigenvectors involved in the coarse space is set to 0.4 and the maximum number of eigenvectors (`-eps_nev`) is set to 30. Local subdomain matrices are factorised by the sparse direct solver MUMPS [15]. The tolerance for the resolution of the eigenvalue problems (`-eps_tol`) is set to 10^{-2}. The coarse problem is solved using an exact Cholesky factorisation performed by MUMPS. For the coarse correction, we use a balanced coarse scale correction as defined in (D.3). The coarse matrix is filtered with a threshold set to 10^{-15} to reduced the amount of memory used.

The options used to implement the preconditioner M_A^{-1} for the matrix A are:

```
-prefix_push fieldsplit_0_
-ksp_rtol 1.0E-4
-pc_type hpddm
-ksp_pc_side right
-pc_hpddm_has_neumann
-pc_hpddm_define_subdomains
-prefix_push pc_hpddm_levels_1_
-eps_nev 60
-eps_threshold 0.2
-st_share_sub_ksp
-sub_pc_factor_mat_solver_type mumps
-sub_pc_type lu
-prefix_pop
-prefix_push pc_hpddm_coarse_
-pc_type cholesky
-pc_factor_mat_solver_type mumps
-mat_filter 1.0E-8
-mat_mumps_icntl_35 1
-mat_mumps_cntl_7 1.0E-7
-prefix_pop
-prefix_pop
```

The inner Krylov solver for the physical blocks is FGMRES (`-ksp_type fgmres`). The relative tolerance (`-ksp_rtol`) of the outer Krylov method is set to 10^{-4} (only a few iterations are required). We set the threshold τ_A (`-eps_threshold`) for selecting the local eigenvectors involving in the coarse space to 0.2 and the number maximum of eigenvectors (`-eps_threshold`) is set to 60. Local subdomain matrices are factorised by the sparse direct solver MUMPS. The coarse problem is solved using an exact Cholesky factorisation performed by MUMPS. The MUMPS parameter (`ICNTL(35)`) is set to 1, which allows an automatic choice of block low-rank (BLR)

[14] by the software (the block low-rank (BLR) allows for memory gains by storing low-rank factors). The accuracy of the low-rank approximation is itself controlled by MUMPS parameter (`CNTL(7)`), which is set to 10^{-7}. The coarse matrix is filtered with a threshold set to 10^{-8} to reduced the amount memory used.

D.3 Computation of Oseen reference solutions in two dimensions

D.3.1 Problem definition

The steady state Oseen problem with homogeneous Dirichlet boundary conditions in a perforated domain is to find the velocity $\boldsymbol{u_\varepsilon} : \Omega^\varepsilon \to \mathbb{R}^2$ and the pressure $p_\varepsilon : \Omega^\varepsilon \to \mathbb{R}$ solution to

$$\begin{cases} -\nu\Delta\boldsymbol{u_\varepsilon} + (\boldsymbol{U_0}\cdot\boldsymbol{\nabla})\boldsymbol{u_\varepsilon} + \boldsymbol{\nabla}p_\varepsilon &= \boldsymbol{f} \quad \text{in } \Omega^\varepsilon, \\ \operatorname{div}\boldsymbol{u_\varepsilon} &= 0 \quad \text{in } \Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &= \boldsymbol{0} \quad \text{on } \partial B^\varepsilon \cap \partial\Omega^\varepsilon, \\ \boldsymbol{u_\varepsilon} &= \boldsymbol{0} \quad \text{on } \partial\Omega \cap \partial\Omega^\varepsilon, \end{cases} \tag{D.8}$$

with $\boldsymbol{f}$ the applied force, $\nu > 0$ the viscosity and $\boldsymbol{U_0}$ the known Oseen velocity. The analysis of the Stokes problem have been extensively presented through different sections of this thesis. The readers may refer to Chapter 12.

D.3.2 Preconditioning strategy

The discretisation of the Oseen problem leads to a saddle point matrix as in the case of the Stokes problem, i.e.

$$\mathcal{A}x = b, \tag{D.9}$$

where $\mathcal{A}$ is in the form of

$$\mathcal{A} = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}.$$

Consequently, we precondition the resulting system, using the same fieldsplit strategy based on the Schur complement, as presented above for the Stokes problem. However, in this case, the matrix A is no longer symmetric, and in particular the problem becomes more and more non-symmetric as the viscosity ν is decreasing. Consequently, we can no longer use the GenEO strategy to precondition the matrix A. Instead, for preconditioning the matrix A we used the strategy introduced in [8].

D.3.3 Final structure of the solver

We use the Flexible Generalized Minimal Residual algorithm (FGMRES, `-ksp_type fgmres`), which iteratively solves (D.9) using full Schur complement as outer preconditioner on the right, taking advantage that FGMRES allows the preconditioner to change at each iteration. The first options to initialize this solver are:

```
-ksp_type fgmres
-ksp_rtol 1.0E-10
-pc_type fieldsplit
-pc_type_fieldsplit schur
-pc_type_fieldsplit_schur_precondition self
```

For preconditioning the Schur complement, we use the same strategy as for the Stokes problem. The options used to implement the preconditioner for the Schur complement are:

```
-prefix_push fieldsplit_1_
-ksp_rtol 1.0E-1
-ksp_type fgmres
-ksp_max_it 100
-pc_type hpddm
-pc_hpddm_has_neumann
-pc_hpddm_ksp_type fgmres
-prefix_push pc_hpddm_levels_1_
-eps_nev 40
-eps_threshold 0.6
-st_share_sub_ksp
-eps_gen_non_hermitian
-sub_pc_factor_mat_solver_type mumps
-sub_mat_mumps_icntl_14 400
-sub_pc_type lu
-eps_tol 1.0E-2
-prefix_pop
-prefix_push pc_hpddm_coarse_
-correction balanced
-pc_type lu
-mat_type baij
-pc_factor_mat_solver_type mumps
-mat_mumps_icntl_14 400
-p 4
-mat_mumps_icntl_24 1
-prefix_pop
-prefix_pop
```

The inner Krylov solver for the physical blocks is FGMRES (`-ksp_type fgmres`). The relative tolerance (`-ksp_rtol`) of the outer Krylov method is set to 10^{-1}. We set the maximum of iterations (`-ksp_max_it`) to 100. The threshold τ_{S_1} (`-esp_threshold`) for selecting the local eigenvectors involved in the coarse space is set to 0.6 and the number maximum of eigenvectors (`-esp_nev`) is set to 40. Local subdomain matrices are factorised by the sparse direct solver MUMPS. We use the MUMPS parameter (`ICNTL(14)`), set to 400, which allows to control the percentage increase in the estimated working space. The coarse problem is solved using an exact Cholesky factorisation performed with MUMPS. We used a balanced coarse scale correction as defined in (D.3). We use the option `-pc_hpddm_coarse_p` to provide the size of the subcommunicator that will be used to remap the coarse operator (with the default value of 1 the coarse matrix is centralised on a single process). Here, we use a value of 4. We use the MUMPS parameter `ICNTL(24)`, set to 1, which allows to detect null pivot row.

For preconditioning the matrix A, which is no longer SPD, we use the approach described in [8] which proposes a fully algebraic two-level additive Schwarz preconditioner for general sparse large-scale matrices. The coarse space is constructed based on the approximation of two local subspaces in each subdomain. These subspaces are obtained by approximating a number of eigenvectors corresponding to dominant eigenvalues of two generalised eigenvalue problems. The number of eigenvectors can be chosen to control the condition number. For general sparse matrices, the coarse space is constructed by approximating the image of a local

operator that can be defined from information in the coefficient matrix. For the theoretical analysis of this preconditioner the readers can refer to [8]. We present below the parameters allowing to precondition the matrix A with this preconditioner, for the sake of reproducibility.

```
-prefix_push fieldsplit_0_
-ksp_rtol 1.0E-4
-ksp_max_it 100
-pc_type hpddm
-ksp_converged_reason
-ksp_type fgmres
-ksp_pc_side right
-pc_hpddm_block_splitting
-pc_hpddm_define_subdomains
-pc_hpddm_levels_1_svd_relative_threshold 1E-6
-pc_hpddm_levels_1_svd_nsv 200
-pc_hpddm_levels_1_eps_gen_non_hermitian
-pc_hpddm_harmonic_overlap 1
-prefix_push pc_hpddm_levels_1_
-eps_gen_non_hermitian
-st_share_sub_ksp
-sub_pc_factor_mat_solver_type
mumps
-sub_mat_mumps_icntl_14 400
-sub_pc_type lu
-prefix_pop
-prefix_push pc_hpddm_coarse_
-pc_type lu
-correction deflated
-pc_factor_mat_solver_type mumps
-mat_type baij
-mat_mumps_icntl_14 400
-p 8
-mat_mumps_icntl_24 1
-mat_filter 1.0E-8
-prefix_pop
-prefix_pop
```

Appendix E

Computation of MsFEM local and coarse problems

Contents

Note: We would like to thank Pierre Jolivet for his help with the development of parallelism and solvers for the local problems.

This appendix is dedicated to give details about the computation of the local and global problems involved in the MsFEM.

E.1 Computation of the local problems

The different local problems ((8.1), (8.2), (12.8), (12.9), (12.17) and (12.18)) are discretized and solved in FreeFEM [93]. Each bi-linear involved in the local problems are assembled in matrix form, and gathered in a PETSc nested matrix, resulting in a matrix of type (8.10). The matrices inside the nested matrix can be sparse matrix (`Mat`), dense matrix (`real[int, int]`) or vector (`real[int]`). We assemble a multiple right-hand side, so that the computation of all the basis functions are managed directly by PETSc (instead of doing a `for` loop to solve one after the other the local problems for each right-hand side).

E.1.1 Parallelism for the local problems

To compute the local problems, we use two levels of parallelism. The first level of parallelism is a parallelism between each coarse elements (without communication), given that all coarse element are fully decoupled (the computation of the basis functions are local). One process (consisting of N_{loc} CPUs, with $N_{loc} \geq 1$) is in charge of N coarse elements, with $N \geq 1$. Each process will treat each coarse element allocated to it sequentially (one after the other).

The second level of parallelism is a parallelism inside each coarse element. Indeed, if the local problems are too big, a parallel computation is required. By setting, $N_{loc} > 1$, i.e. the process is composed of more than one CPUs, the computation of the local problems inside each coarse element is done in parallel. For this intra-parallelism, each coarse element is partitioned first in two subdomains: a domain D_{out} which contains the border of the coarse element (useful for computing the Lagrange multipliers on the faces), in yellow on Figure E.1, and a domain D_{in} which does not touch the border of the coarse element. Then, the subdomain D_{in} is decomposed in $N_{loc} - 1$ subdomains using classical partitioner such as METIS (see for example Figure E.1).

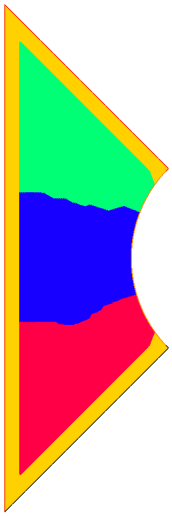


Figure E.1: Example of partitioning of a coarse triangle in four subdomains.

E.1.2 Preconditioning strategy to solve three-dimensional local problems

For the two-dimensional cases, the local problems are solved with a direct solver. However, for three-dimensional problem a direct solver cannot be longer used due to the size of the problem. However, preconditioning efficiently the local problems is very difficult due to the complexity of the problems involving Lagrange multipliers. We use a pseudo-block FGMRES as implemented in HPDDM to thread multiple right-hand sides simultaneously [117]. The parameters used to implement this solver are given below:

```
-ksp_type hpddm
-ksp_hpddm_variant flexible
-ksp_hpddm_type gmres
-ksp_gmres_restart 5
-ksp_matsolve_batch_size 5
-ksp_rtol 1E-10
-mat_mumps_icntl_29 2
-mat_mumps_icntl_28 2
-pc_type cholesky
-pc_factor_mat_solver_type mumps
-pc_precision single
```

The Krylov solver is the flexible variant of GMRES (`-ksp_hpddm_type gmres`). The relative tolerance (`-ksp_rtol`) of the Krylov method is set to 10^{-10}. We restart the Krylov method each 5 iterations (`-ksp_gmres_restart`) which allows to save memory. Similarly, with the aim to save memory, we limit the number of right-hand sides computed simultaneously to 5 (`-ksp_matsolve_batch_size 5`). As the problem is symmetric, the problem is factorised, using single precision (instead of a double precision to save memory), with an exact Cholesky factorisation performed by MUMPS. We use the MUMPS parameters `ICNTL(28)` and `ICNTL(29)` which determine respectively whether a sequential or parallel computation of the ordering (to reduce permutation) is performed and the parallel ordering tool to be used. We set the parameter `ICNTL(28)` to 2 allowing a parallel ordering and parallel symbolic factorisation and we set the parameter `ICNTL(29)` to 2 which allows to use ParMETIS to reorder the input matrix, if available.

Remark E.1. It may be thought that preconditioning the system using a fieldsplit strategy, i.e. developing a preconditioner for each physical block is also possible. However, for the moment, the fieldsplit preconditioning in FreeFEM does not handle multiple right-hand sides (a block of right-hand sides). Consequently, applying the fieldsplit strategy will require to apply the preconditioner right-hand side by right-hand side which is clearly not efficient. On the contrary, with the proposed methodology, the set up is more expensive but the preconditioner is applied once on a block of right-hand sides, which is much more efficient.

E.1.3 Procedure for solving the local problems

Algorithm 1: Solving the local problems.

$N_{CPU} \leftarrow$ number of CPUs used ;
$N_{loc} \leftarrow$ number of CPUs allocated to each sub-process ;
Initialise `commGlobal` ; `// global MPI communicator of size` N_{CPU}
Initialise `commLocal` ; `// MPI communicator for each sub-process of size` N_{loc}
Initialise $\mathbb{A}_{glob}$ and $\mathbb{B}_{glob}$; `// Global stiffness and velocity-pressure matrices`
Initialise $\mathbb{A}_{tmp}$ and $\mathbb{B}_{tmp}$; `// Temporary matrices to add local contributions`
Ensure: $N_{loc} | N_{CPU}$;
foreach *process* P_j *(composed of* N_{loc} *CPUs)* **do**
  foreach *coarse element* T *allocated to the process* P_j **do**
    Load the local fine mesh $\mathcal{T}_h(T)$;
    if $N_{loc} > 1$ **then**
      Partition $\mathcal{T}_h(T)$;
    end
    Assemble the local problems under matrix form $\mathcal{A}_{loc}$;
    Assemble the multiple right-hand sides $\mathcal{F}$;
    Solve $\mathcal{A}_{loc}\phi = \mathcal{F}$ on N_{loc} CPUs ;
    if $N_{loc} > 1$ **then**
      Reconstruct the basis functions on a non-partitioned mesh $\mathcal{T}_h(T)$;
    end
    if *mpiRank(commLocal)=0* **then**
      Store the basis functions ;
    end
    `/* Assemble matrices locally */`
    if *mpiRank(commLocal)=0* **then**
      Compute $\mathbb{A}_T$; `// Local stiffness matrix`
      Compute $\mathbb{B}_T$; `// Local matrix velocity-pressure`
      $\mathbb{A}_{tmp} += \mathbb{A}_T$; `// Add local contributions`
      $\mathbb{B}_{tmp} += \mathbb{B}_T$;
    end
  end
end
`/* Gather the contributions of all the processes */`
`MPIAllReduce`($\mathbb{A}_{tmp}$, $\mathbb{A}_{glob}$, `commGlobal`, `mpiSUM`) ;
`MPIAllReduce`($\mathbb{B}_{tmp}$, $\mathbb{B}_{glob}$, `commGlobal`, `mpiSUM`) ;
if *mpiRank(commGlobal)=0* **then**
  Store $\mathbb{A}_{glob}$;
  Store $\mathbb{B}_{glob}$;
end

E.2 Computation of the coarse problem

During the online stage, the resolution of the global problem, the assembly of the right-hand side and the penalisation of the stiffness matrix can be done in parallel (see Algorithm 2). The resolution of the resulting linear system is done with a direct solver, since this problem is quite small (see Algorithm 3). The reconstruction of the solution on the fine mesh can also be done in parallel (see Algorithm 4). During this online stage, the number of CPUs used is quite small. Using more CPUs will decrease the time to compute the right-hand side, penalise the stiffness matrix and reconstruct the solution, however in any case the coarse problem is solved on one CPU.

Remark E.2. This part emphasises another advantages of MsFEM: no complex preconditioning strategies have to be developed to have an approximation of the solutions.

Algorithm 2: Boundary conditions and Load of the coarse problem.

$N_{CPU} \leftarrow$ number of CPUs used ; // A quite small number
$\Gamma_D \leftarrow$ Dirichlet boundary ;
$\boldsymbol{g} \leftarrow$ Dirichlet boundary conditions ;
$\boldsymbol{f} \leftarrow$ Source term ;
tgv $\leftarrow$ big value for the penalization method ; // 10^{30}
Load $\mathbb{A}_{glob}$ and $\mathbb{B}_{glob}$;
Initialise $\mathbb{F}_{tmp}$ and $\mathbb{F}_{glob}$; // temporary and global right-hand side
Initialise $\mathbb{A}_{tmp}$ and $\mathbb{A}_{penal}$; // temporary matrices to apply the penalization method in parallel
foreach *process* P_j *(composed of 1 CPU)* **do**
  foreach *coarse element* T *allocated to the process* P_j **do**
    Load the local fine mesh $\mathcal{T}_h(T)$;
    for *each basis functions* $\boldsymbol{\varphi_i}$ **do**
      I $\leftarrow$ global numbering of the d.o.f. associated with $\boldsymbol{\varphi_i}$; // Face and Element basis functions
      b $\leftarrow \int_T \boldsymbol{f} \cdot \boldsymbol{\varphi_i}$;
      $\mathbb{F}_{tmp}(I)$ += b ;
    end
    foreach F *in* $\mathcal{F}_T$ **do**
      if $F \subset \Gamma_D$ **then**
        foreach *weighting functions* $\boldsymbol{\omega_{F,i}}$ **do**
          I $\leftarrow$ global numbering of the d.o.f. corresponding to the basis functions $\boldsymbol{\phi_{F,i}}$; // Face basis functions
          penal $\leftarrow \int_F \boldsymbol{\phi_{F,i}} \cdot \boldsymbol{\omega_{F,i}}$ * tgv ;
          bound $\leftarrow \int_F \boldsymbol{g} \cdot \boldsymbol{\omega_{F,i}}$ * tgv ;
          $\mathbb{A}_{tmp}(I, I)$+= penal ;
          $\mathbb{F}_{tmp}(I)$ += bound ;
        end
      end
    end
  end
end
/* Gather the contributions of all processus */
`MPIAllReduce`($\mathbb{A}_{tmp}$, $\mathbb{A}_{penal}$, `commGlobal`, `mpiSUM`) ;
`MPIAllReduce`($\mathbb{F}_{tmp}$, $\mathbb{F}_{glob}$, `commGlobal`, `mpiSUM`) ;
$\mathbb{A}_{glob} += \mathbb{A}_{penal}$; // Apply the penalizations to the global stiffness matrix

Algorithm 3: Solving the coarse problem.

Define $c_H(p,q) = -\int_T \eta\ p * q$ // $\eta = 10^{-8}$
Compute $\mathbb{C}_{glob} = c_H(M_H, M_H)$ // To help the solving
Assemble $\mathbb{N}_{glob} = \begin{pmatrix} \mathbb{A}_{glob} & \mathbb{A}_{glob} \\ \mathbb{B}'_{glob} & \mathbb{C}_{glob} \end{pmatrix}$
Set $\mathbb{F} = [\mathbb{F}_{glob}, \mathbf{0}]^t$
Solve $\mathbb{N}_{glob} X_{glob} = \mathbb{F}$

Algorithm 4: Reconstruction of the local solutions.

$N_{CPU} \leftarrow$ number of CPUs used ;
$X_{glob} \leftarrow$ global solution ;
Initialise $\boldsymbol{u_{loc}} = \mathbf{0}$; // Local velocity
Initialise $p_{loc} = 0$; // Local pressure
foreach *process* P_j *(composed of 1 CPU)* **do**
  foreach *coarse element* T *allocated to the process* P_j **do**
    Load the local fine mesh $\mathcal{T}_h(T)$;
    Load the local basis functions ;
    foreach *basis functions* $\boldsymbol{\varphi_i}$ *and the associated* π_i **do**
      $I \leftarrow$ global numbering of the d.o.f. associated with $\boldsymbol{\varphi_i}$; // Face and Element basis functions
      $\boldsymbol{u_{loc}} += X_{glob}[I] * \boldsymbol{\varphi_i}$;
      $p_{loc} += X_{glob}[I] * \pi_i$;
    end
    $p_{loc} += \overline{p}_H$; // Add the average global pressure
    Store $\boldsymbol{u_{loc}}$ and p_{loc} ;
  end
end

E.3 Details about the MsFEM implementations

In this thesis, for the ease of implementation, the coarse elements are triangles in two dimensions and tetrahedra in three dimensions. In this section, we explain the construction of two ingredients of MsFEM implementations: the barycentric coordinates and the tangential vectors on the faces of the coarse elements. They are are useful to build polynomials on faces of the coarse elements.

E.3.1 Construction of barycentric coordinates on faces in two and three dimensions

E.3.1.1 Construction of barycentric coordinates on faces in two dimensions

Given a non degenerated triangle (ABC), we consider the edge linking the vertices A and B. The barycentric coordinate associated with the vertex $A(x_A, y_A)$ is the function λ_A verifying $\lambda_A(A) = 1$, $\lambda_A(B) = 0$, and which is null on the line normal to the edge (AB) and passing

through the point $B(x_B, y_B)$. This hyperplane is therefore uniquely defined: three parameters to be determined and three constraints. Indeed, the equation of the barycentric coordinates in two dimensions are of type

$$\lambda(x, y) = ax + by + c.$$

In two dimensions, the two barycentric coordinates on the edge (AB) reads

$$\lambda_A(x, y) = \frac{(x - x_B)(x_A - x_B) + (y - y_B)(y_A - y_B)}{L^2},$$

$$\lambda_B(x, y) = \frac{(x - x_A)(x_B - x_A) + (y - y_A)(y_B - y_A)}{L^2},$$

where L is the length of the segment given by

$$L = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}.$$

E.3.1.2 Construction of barycentric coordinates in three dimensions

Given a non degenerated tetrahedron $(ABCD)$, we consider the face defined by the three points $A(x_A, y_A, z_A)$, $B(x_B, y_B, z_B)$, $C(x_C, y_C, z_C)$. The barycentric coordinate associated with the point A is the function λ_A verifying $\lambda_A(A) = 1$, $\lambda_A(B) = 0$, $\lambda_A(C) = 0$ and that is null on the plane orthogonal to the face and passing through the points B and C. This hyperplane is therefore uniquely defined: four parameters to be determined and four constraints. Indeed, the equation of the barycentric coordinates in three dimensions is of type

$$\lambda(x, y, z) = ax + by + cz + d.$$

In three dimensions, the three barycentric coordinates associated with the face (ABC) read

$$\lambda_A(x, y, z) = \frac{a_0(x - x_A) + b_0(y - y_A) + c_0(z - z_A)}{2S^2} + 1,$$

$$\lambda_B(x, y, z) = \frac{a_1(x - x_B) + b_1(y - y_B) + c_1(z - z_B)}{2S^2} + 1,$$

$$\lambda_C(x, y, z) = \frac{a_2(x - x_C) + b_2(y - y_C) + c_2(z - z_C)}{2S^2} + 1,$$

where S is the surface of a triangle ABC in three dimensions, given by

$$S = \frac{1}{2}\sqrt{\left(\det\begin{pmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{pmatrix}\right)^2 + \left(\det\begin{pmatrix} y_A & y_B & y_C \\ z_A & z_B & z_C \\ 1 & 1 & 1 \end{pmatrix}\right)^2 + \left(\det\begin{pmatrix} z_A & z_B & z_C \\ x_A & x_B & x_C \\ 1 & 1 & 1 \end{pmatrix}\right)^2}$$

and the constants a_i, b_i, c_i, with $i = 0, 1, 2$, are chosen such that the properties, mentioned above, of the barycentric coordinates are satisfied.

E.3.2 Construction of tangential vectors in faces in two and three dimensions

In FreeFEM, we have only access to the normal vector $\boldsymbol{n}$. We explain below, how to construct, from the normal vector $\boldsymbol{n}$, the tangential vector $\boldsymbol{t}$ in two dimensions and the tangential and binormal vectors ($\boldsymbol{t}$, $\boldsymbol{b}$) in three dimensions.

E.3.2.1 Tangential vector on edge in two dimensions

In two dimensions, building the tangential vector from the normal vector is easy: it suffices to swap the vector components and invert the sign of one of the entries. Then, given an arbitrary normal vector $\boldsymbol{n} \in \mathbb{R}^2 \setminus \{\boldsymbol{0}\}$ the orthogonal tangential vector $\boldsymbol{t}$ is obtained with

$$\boldsymbol{t} = [-n_y \; n_x]^t \tag{E.1}$$

where n_x and n_y are respectively the first and the second components of the normal vector $\boldsymbol{n}$.

E.3.2.2 Tangential vectors on faces in three dimensions

In three dimensions, the construction of the tangential vectors is more complex. The naive approach is to consider a non-colinear vector $\boldsymbol{v}$ to the normal vector $\boldsymbol{n}$, whose cross-product would provide an orthogonal vector $\boldsymbol{t}$. By applying a cross-product between $\boldsymbol{n}$ and $\boldsymbol{t}$ a second vector, $\boldsymbol{b}$, of the basis is obtained. However this approach is time consuming. Instead, in order to compute the tangential and binormal vectors on faces, from the normal vector, we use the approach proposed by [124]. In this work, the authors propose an efficient way to determine the tangential and binormal vectors, from the normal vector, based on the Householder transformation (they compare the efficiency of this method with other well-known methods such as Eberly, Square Plate or Projection Matrix methods). Below, we present the peudo-code for obtaining the unit vectors based on the Householder (HH) orthogonalization.

Algorithm E.3 (HH orthogonalization)**.**

- Evaluate the sign the unit's vector first component, i.e. $\text{sign}(n_x)$;
- Determine the tangent vector vector with the following simplified expression:

$$\boldsymbol{t} = \begin{cases} \left[-n_y \quad 1 - \dfrac{n_y^2}{n_x+1} \quad -\dfrac{n_y n_z}{n_x+1}\right]^t, & n_x \geq 0, \\ \left[n_y \quad 1 + \dfrac{n_y^2}{n_x-1} \quad -\dfrac{n_y n_z}{n_x-1}\right]^t, & n_x < 0, \end{cases}$$

- Determine the binormal vector with the following simplified expression:

$$\boldsymbol{b} = \begin{cases} \left[-n_z \quad 1 - \dfrac{n_y n_z}{n_x+1} \quad -\dfrac{n_z^2}{n_x+1}\right]^t, & n_x \geq 0, \\ \left[n_z \quad 1 + \dfrac{n_y n_z}{n_x-1} \quad 1 + \dfrac{n_z^2}{n_z-1}\right]^t, & n_x < 0. \end{cases}$$

E.3.3 Definition of the weighting functions in triangles and tetrahedra

E.3.3.1 Weighting functions for two dimensional MsFEM

The velocity weighting functions associated with the edge F are defined by:

$$\begin{aligned} \boldsymbol{\omega_{F,1}} &= 1\boldsymbol{n}, & \boldsymbol{\omega_{F,2}} &= 1\boldsymbol{t}, \\ \boldsymbol{\omega_{F,3}} &= (\lambda_1 - \tfrac{1}{2})\boldsymbol{n}, & \boldsymbol{\omega_{F,4}} &= (\lambda_1 - \tfrac{1}{2})\boldsymbol{t}, \\ \boldsymbol{\omega_{F,5}} &= (\lambda_1\lambda_2 - \tfrac{1}{6})\boldsymbol{n}, & \boldsymbol{\omega_{F,5}} &= (\lambda_1\lambda_2 - \tfrac{1}{6})\boldsymbol{t}. \end{aligned}$$

where λ_1, λ_2 are the two barycentric coordinates associated with the edge F. The vectors $\boldsymbol{n}$, $\boldsymbol{t}$ are respectively the normal and the tangential vectors (defined in (E.1)) to the edge F.

The velocity weighting functions associated with the element T are defined by:

$$\begin{aligned} \boldsymbol{\varphi_{T,1}} &= \boldsymbol{e_1}, & \boldsymbol{\varphi_{T,2}} &= \boldsymbol{e_2}, \\ \boldsymbol{\varphi_{T,3}} &= (x - x_b)\boldsymbol{e_1}, & \boldsymbol{\varphi_{T,4}} &= (x - x_b)\boldsymbol{e_2}, \\ \boldsymbol{\varphi_{T,5}} &= (y - y_b)\boldsymbol{e_1}, & \boldsymbol{\varphi_{T,6}} &= (y - y_b)\boldsymbol{e_2}. \end{aligned}$$

where x_b, y_b are the two components of the barycenter of the element T, and $(\boldsymbol{e_1}, \boldsymbol{e_2})$ are the canonical basis of $\mathbb{R}^2$.

The pressure weighting functions associated with the element T are defined by:

$$\begin{aligned} \varpi_{T,1} &= 1, & \varpi_{T,2} &= (x - x_b), & \varpi_{T,3} &= (y - y_b), \\ \varpi_{T,4} &= (x - x_b)(y - y_b), & \varpi_{T,5} &= (x - x_b)^2, & \varpi_{T,6} &= (y - y_b)^2. \end{aligned}$$

where x_b, y_b are the two components of the barycenter of the element T.

E.3.3.2 Weighting functions for three dimensional MsFEM

The velocity weighting functions associated with the face F are defined by:

$$
\begin{array}{lll}
\boldsymbol{\omega_{F,1}} = \frac{1}{2}\boldsymbol{n}, & \boldsymbol{\omega_{F,2}} = \frac{1}{2}\boldsymbol{t}, & \boldsymbol{\omega_{F,3}} = \frac{1}{2}\boldsymbol{b}, \\
\boldsymbol{\omega_{F,4}} = (\lambda_1 - \frac{1}{3})\boldsymbol{n}, & \boldsymbol{\omega_{F,5}} = (\lambda_1 - \frac{1}{3})\boldsymbol{t}, & \boldsymbol{\omega_{F,6}} = (\lambda_1 - \frac{1}{3})\boldsymbol{b}, \\
\boldsymbol{\omega_{F,7}} = (\lambda_2 - \frac{1}{3})\boldsymbol{n}, & \boldsymbol{\omega_{F,8}} = (\lambda_2 - \frac{1}{3})\boldsymbol{t}, & \boldsymbol{\omega_{F,9}} = (\lambda_2 - \frac{1}{3})\boldsymbol{b}, \\
\boldsymbol{\omega_{F,10}} = (\lambda_1\lambda_2 - \frac{1}{12})\boldsymbol{n}, & \boldsymbol{\omega_{F,11}} = (\lambda_1\lambda_2 - \frac{1}{12})\boldsymbol{t}, & \boldsymbol{\omega_{F,12}} = (\lambda_1\lambda_2 - \frac{1}{12})\boldsymbol{b}, \\
\boldsymbol{\omega_{F,13}} = (\lambda_1\lambda_3 - \frac{1}{12})\boldsymbol{n}, & \boldsymbol{\omega_{F,14}} = (\lambda_1\lambda_3 - \frac{1}{12})\boldsymbol{t}, & \boldsymbol{\omega_{F,15}} = (\lambda_1\lambda_3 - \frac{1}{12})\boldsymbol{b}, \\
\boldsymbol{\omega_{F,16}} = (\lambda_2\lambda_3 - \frac{1}{12})\boldsymbol{n}, & \boldsymbol{\omega_{F,17}} = (\lambda_2\lambda_3 - \frac{1}{12})\boldsymbol{t}, & \boldsymbol{\omega_{F,18}} = (\lambda_2\lambda_3 - \frac{1}{12})\boldsymbol{b},
\end{array}
$$

where λ_1, λ_2 and λ_3 are the three barycentric coordinates associated with the face F. The vectors $\boldsymbol{n}$, $\boldsymbol{t}$ and $\boldsymbol{b}$ are respectively the normal, the tangential and the binormal vectors (defined in Algorithm E.3) to the face F.

The velocity weighting functions associated with the element T are defined by:

$$
\begin{array}{lll}
\boldsymbol{\varphi_{T,1}} = \boldsymbol{e_1}, & \boldsymbol{\varphi_{T,2}} = \boldsymbol{e_2}, & \boldsymbol{\varphi_{T,3}} = \boldsymbol{e_3}, \\
\boldsymbol{\varphi_{T,4}} = (x - x_b)\boldsymbol{e_1}, & \boldsymbol{\varphi_{T,5}} = (x - x_b)\boldsymbol{e_2}, & \boldsymbol{\varphi_{T,6}} = (x - x_b)\boldsymbol{e_3}, \\
\boldsymbol{\varphi_{T,7}} = (y - y_b)\boldsymbol{e_1}, & \boldsymbol{\varphi_{T,8}} = (y - y_b)\boldsymbol{e_2}, & \boldsymbol{\varphi_{T,9}} = (y - y_b)\boldsymbol{e_3}, \\
\boldsymbol{\varphi_{T,10}} = (z - z_b)\boldsymbol{e_1}, & \boldsymbol{\varphi_{T,11}} = (z - z_b)\boldsymbol{e_2}, & \boldsymbol{\varphi_{T,12}} = (z - z_b)\boldsymbol{e_3}
\end{array}
$$

where x_b, y_b and z_b are the three components of the barycenter of the element T, and ($\boldsymbol{e_1}$, $\boldsymbol{e_2}$, $\boldsymbol{e_3}$) are the canonical basis of $\mathbb{R}^3$.

The pressure weighting functions associated with the element T are defined by:

$$
\begin{array}{lll}
\varpi_{T,1} = 1, & \varpi_{T,2} = (x - x_b), & \varpi_{T,3} = (y - y_b), \\
\varpi_{T,4} = (x - x_b)(y - y_b), & \varpi_{T,5} = (x - x_b)(z - z_b), & \varpi_{T,6} = (y - y_b)(z - z_b), \\
\varpi_{T,7} = (x - x_b)^2, & \varpi_{T,8} = (y - y_b)^2, & \varpi_{T,9} = (z - z_b)^2.
\end{array}
$$

where x_b, y_b and z_b are the three components of the barycenter of the element T.

Appendix F

Some additional numerical results

F.1 Additional numerical results for Stokes flows

This section is the Appendix of Chapter 14. As in Chapter 14, we solve the Stokes problem (7.1) on a coarse grid using basis functions defined by the Stokes equations, i.e. the local problems (8.1) and (8.2). We set $\nu = 1$. For the two-dimensional simulations, we consider the channel flow test case presented in Section 13.3.2.

F.1.1 Comparison between high-order MsFEMs and previous MsFEMs

In this section, we perform the convergence analysis proposed in Section 13.3.3. Figures F.1 to F.3 present the computed relative errors between the reference solutions computed on reference meshes and the MsFEM approximations computed on coarse meshes for high-order MsFEMs and previous MsFEMs developed in the literature.

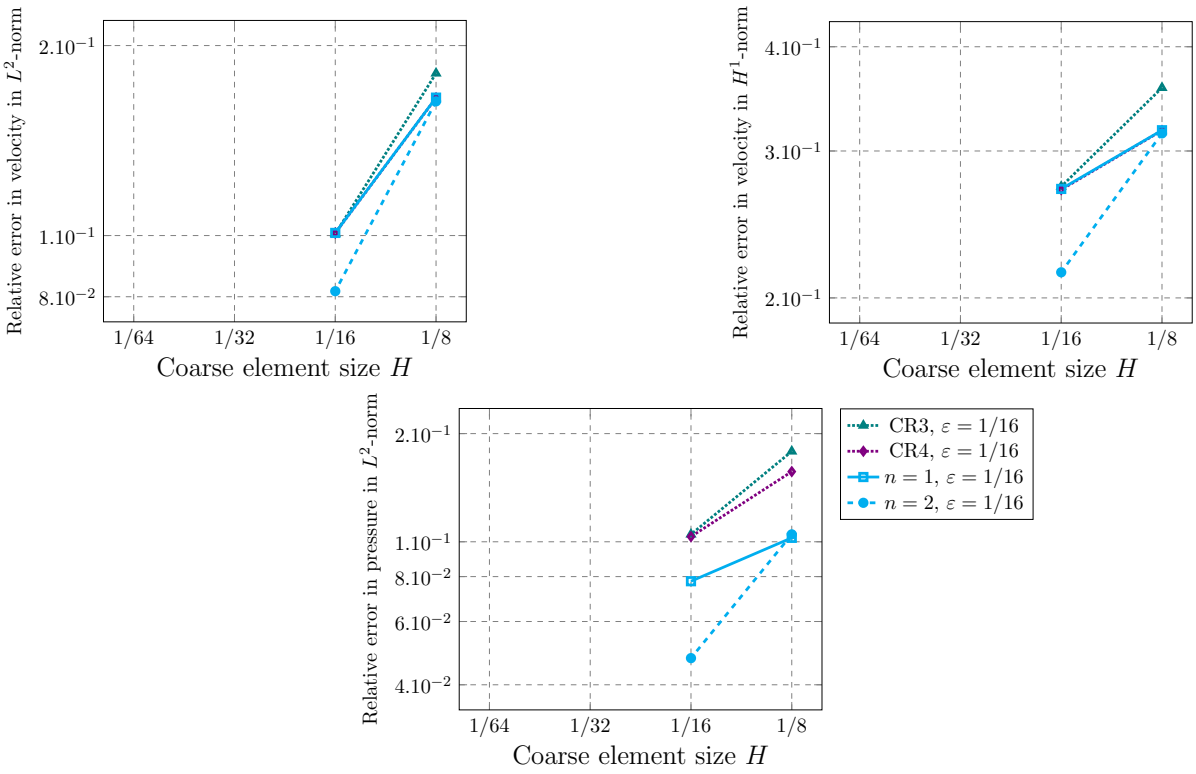


Figure F.1: Relative errors between MsFEM approximations and the reference solution, $\varepsilon = 1/16$ (Stokes flow).

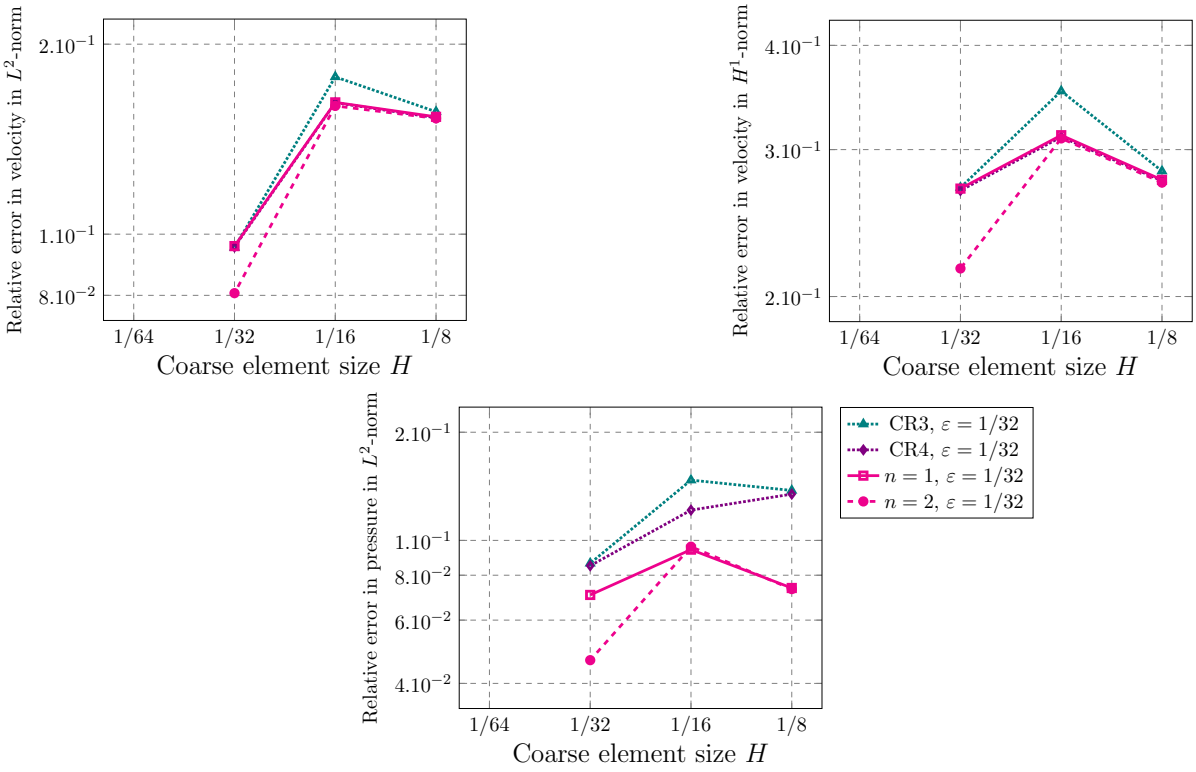


Figure F.2: Relative errors between MsFEM approximations and the reference solution, $\varepsilon = 1/32$ (Stokes flow).

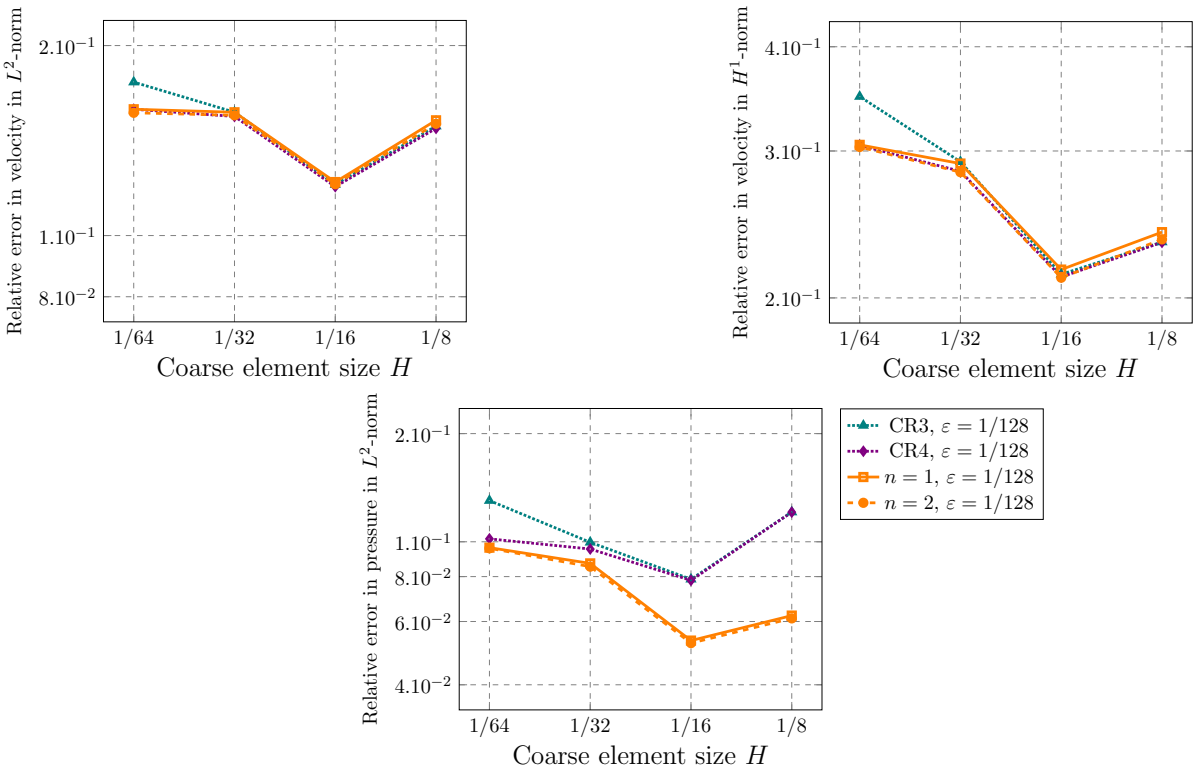
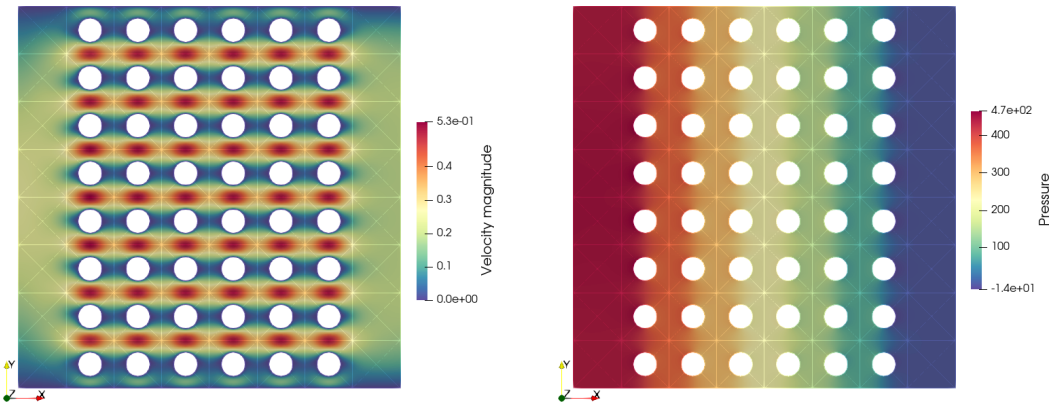


Figure F.3: Relative errors between MsFEM approximations and the reference solution, $\varepsilon = 1/128$ (Stokes flow).

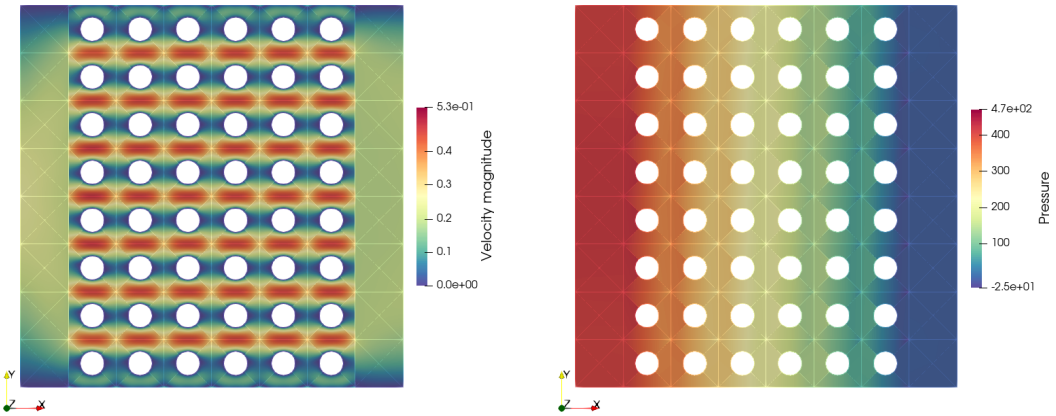
F.1.2 Numerical simulations of Stokes flows

F.1.2.1 Stokes flow in a periodic domain $\varepsilon = 1/8$



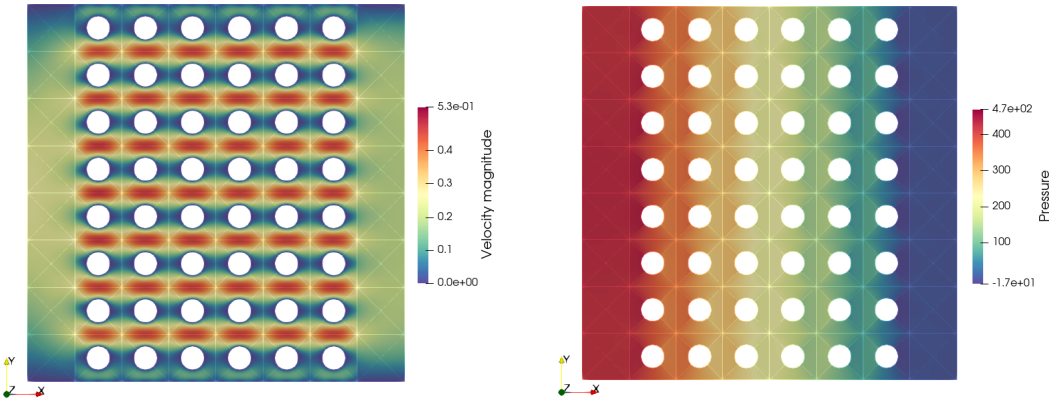
(a) Velocity magnitude. (b) Pressure field.

Figure F.4: Reference solution (Stokes flow), $\varepsilon = 1/8$.



(a) Velocity magnitude. (b) Pressure field.

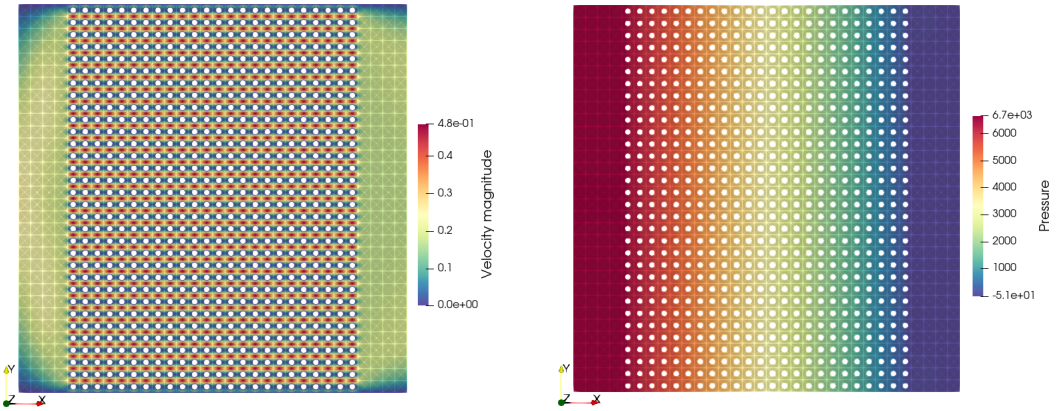
Figure F.5: MsFEM approximation (Stokes flow), $n = 1$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/8$.



(a) Velocity magnitude. (b) Pressure field.

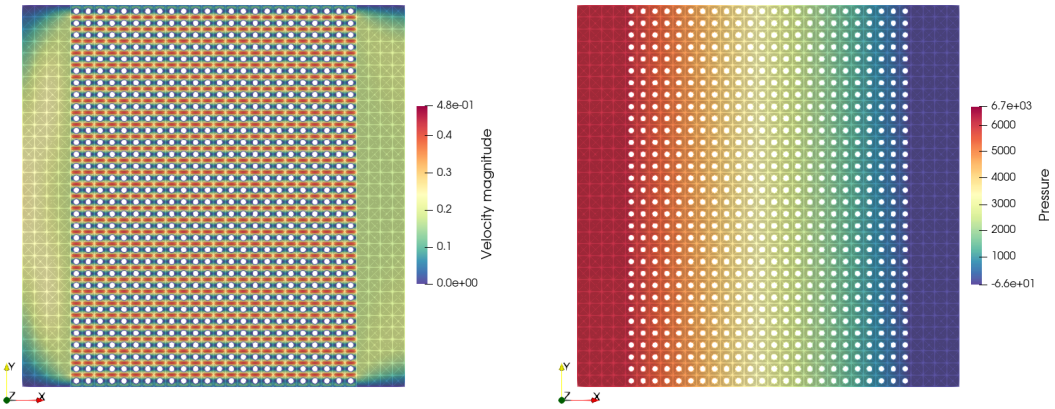
Figure F.6: MsFEM approximation (Stokes flow), $n = 2$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/8$.

F.1.2.2 Stokes flow in a periodic domain $\varepsilon = 1/32$



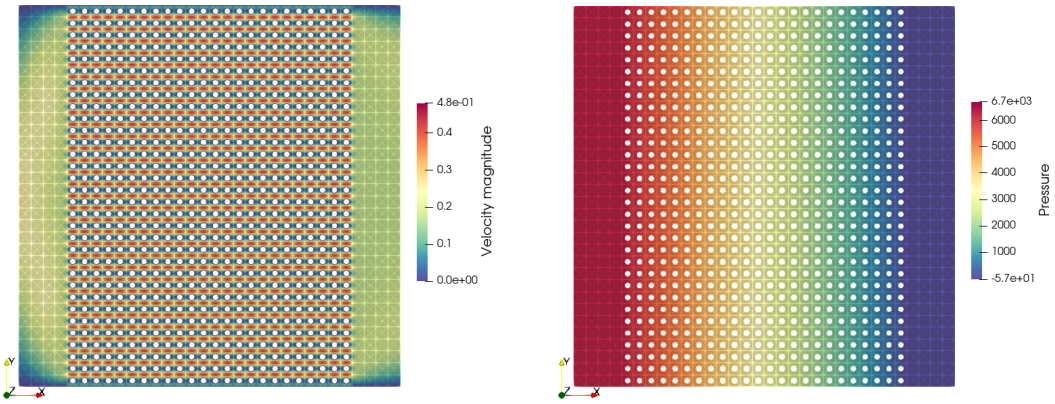
(a) Velocity magnitude. (b) Pressure field.

Figure F.7: Reference solution (Stokes flow), $\varepsilon = 1/32$.



(a) Velocity magnitude. (b) Pressure field.

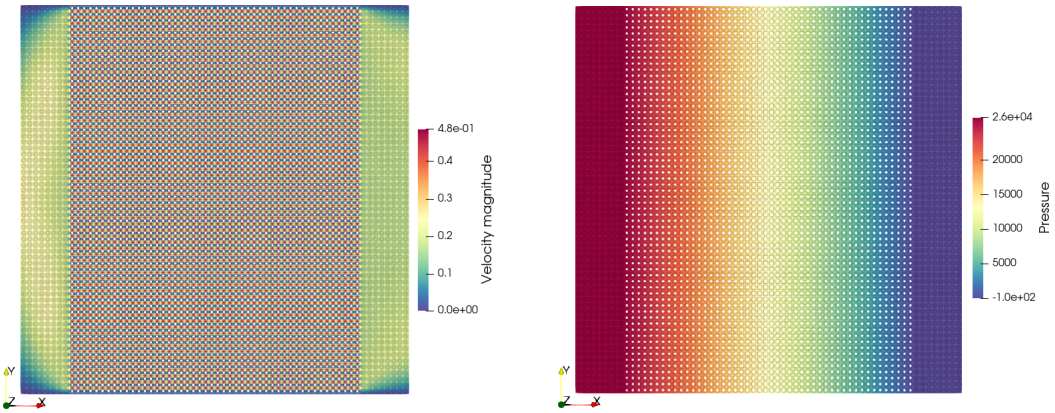
Figure F.8: MsFEM approximation (Stokes flow), $n = 1$, $H = 1/32$ (4096 coarse elements), $\varepsilon = 1/32$.



(a) Velocity magnitude. (b) Pressure field.

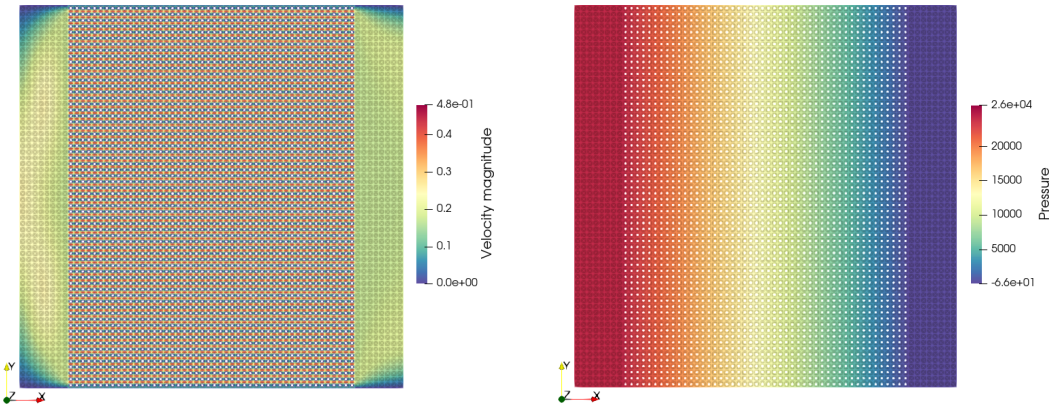
Figure F.9: MsFEM approximation (Stokes flow), $n = 2$, $H = 1/32$ (4096 coarse elements), $\varepsilon = 1/32$.

F.1.2.3 Stokes flow in a periodic domain $\varepsilon = 1/64$



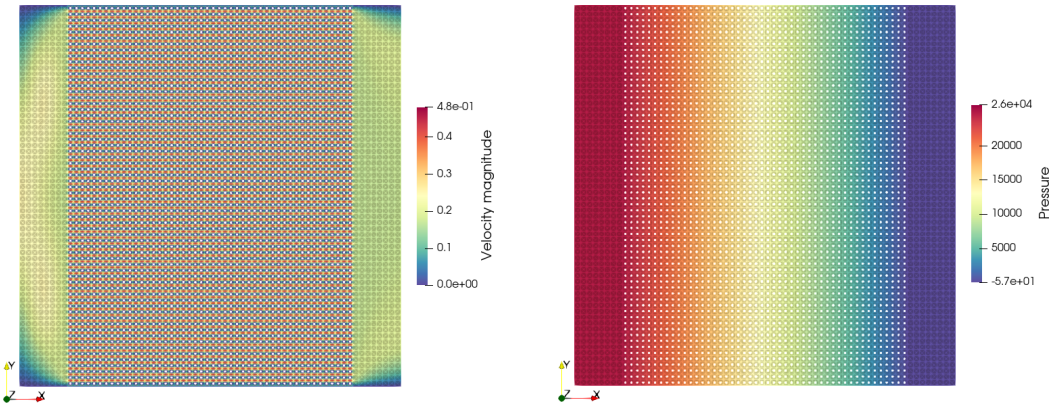
(a) Velocity magnitude. (b) Pressure field.

Figure F.10: Reference solution (Stokes flow), $\varepsilon = 1/64$.



(a) Velocity magnitude. (b) Pressure field.

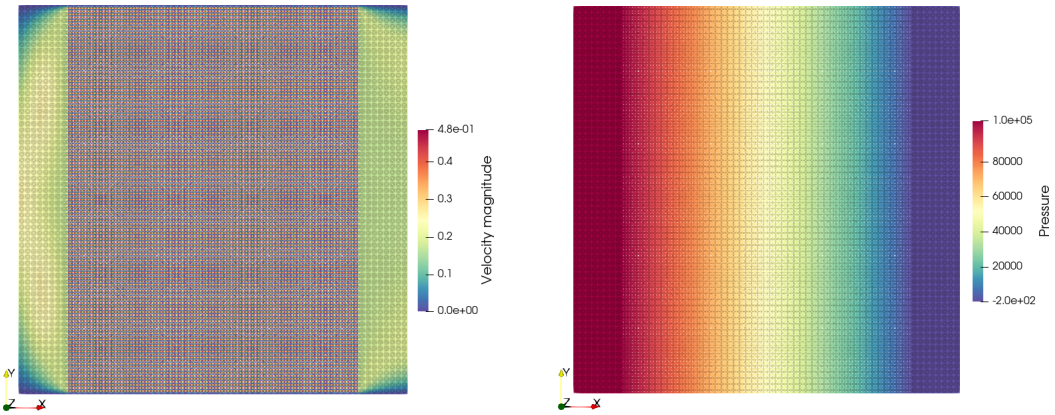
Figure F.11: MsFEM approximation (Stokes flow), $n = 1$, $H = 1/64$ (16384 coarse elements), $\varepsilon = 1/64$.



(a) Velocity magnitude. (b) Pressure field.

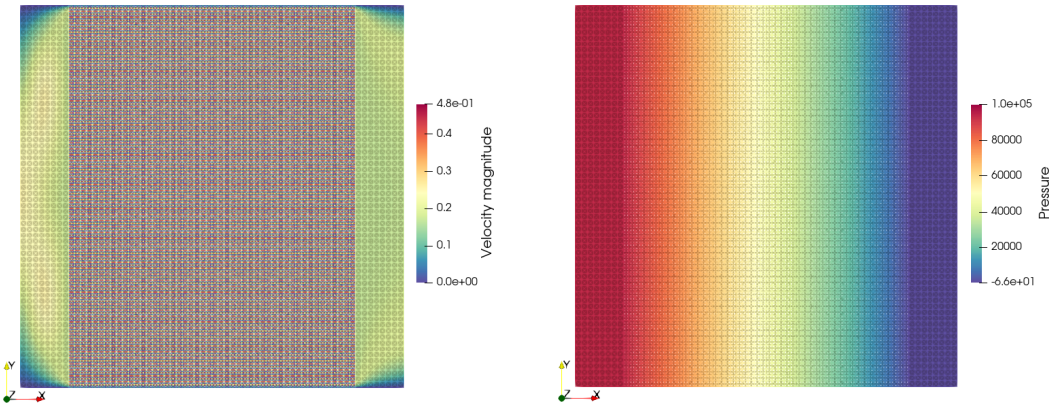
Figure F.12: MsFEM approximation (Stokes flow), $n = 2$, $H = 1/64$ (16384 coarse elements), $\varepsilon = 1/64$.

F.1.2.4 Stokes flow in a periodic domain $\varepsilon = 1/128$



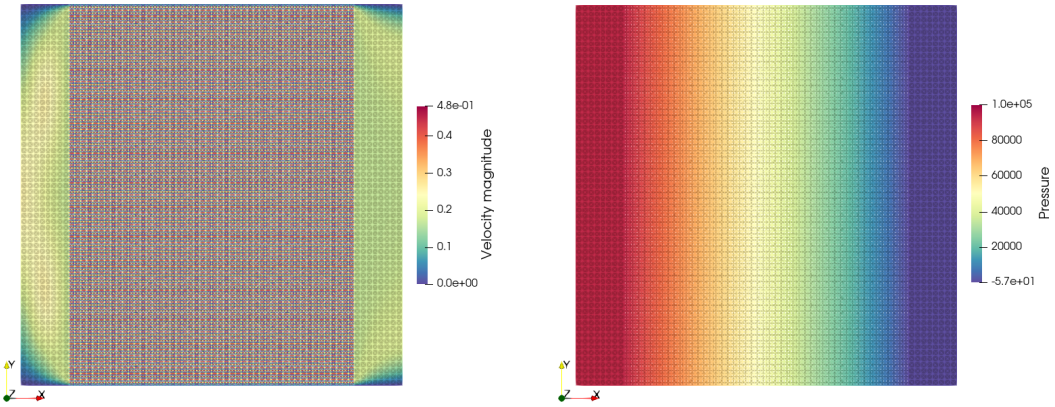
(a) Velocity magnitude. (b) Pressure field.

Figure F.13: Reference solution (Stokes flow), $\varepsilon = 1/128$.



(a) Velocity magnitude. (b) Pressure field.

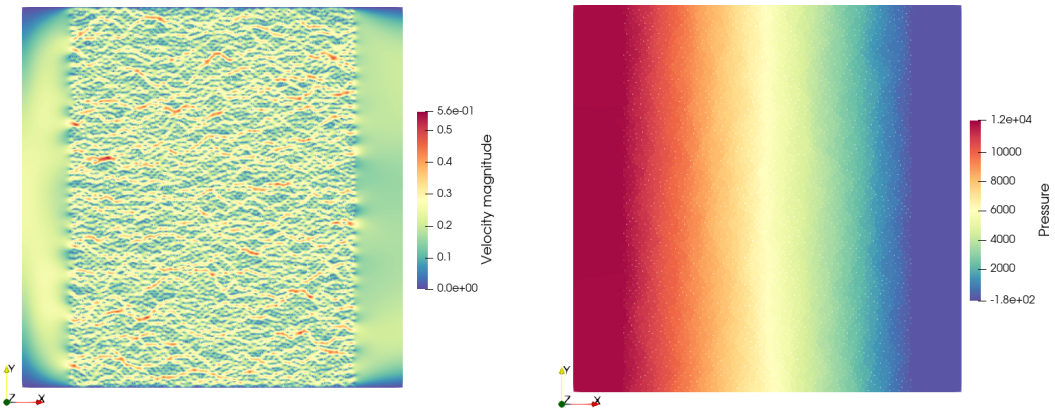
Figure F.14: MsFEM approximation (Stokes flow), $n = 1$, $H = 1/64$ (16384 coarse elements), $\varepsilon = 1/128$.



(a) Velocity magnitude. (b) Pressure field.

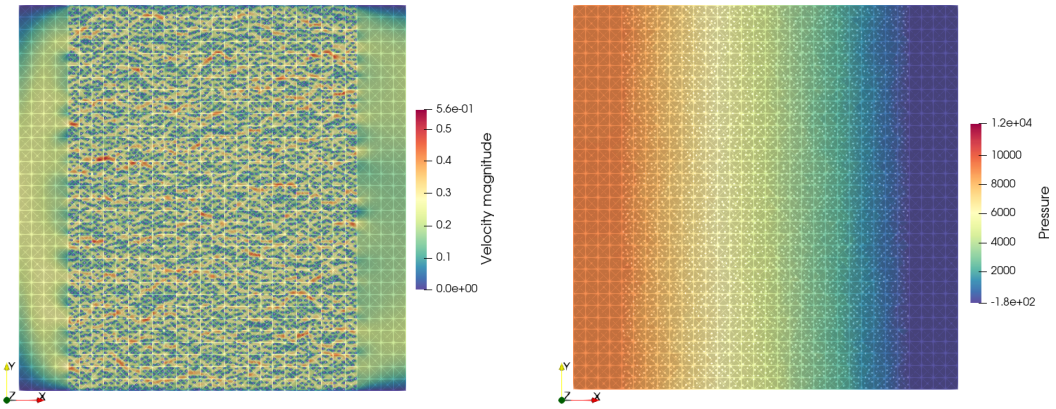
Figure F.15: MsFEM approximation (Stokes flow), $n = 2$, $H = 1/64$ (16384 coarse elements), $\varepsilon = 1/128$.

F.1.2.5 Stokes flow in domain with 4096 randomly placed obstacles



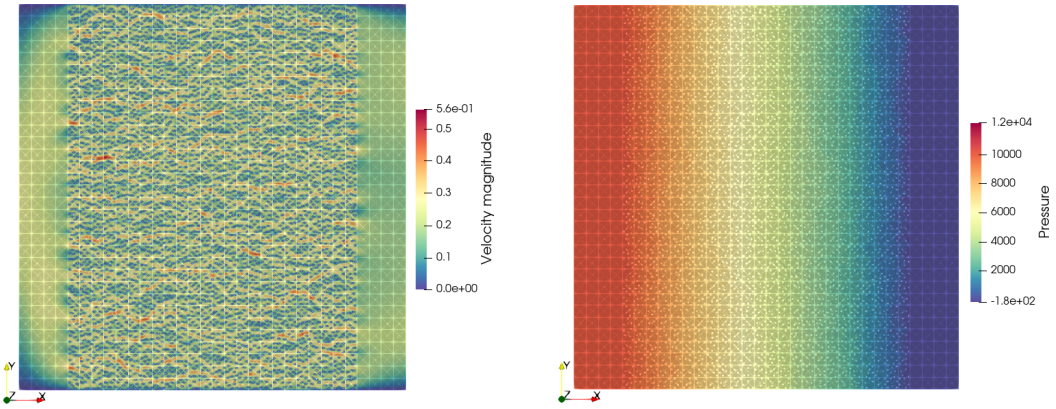
(a) Velocity magnitude. (b) Pressure field.

Figure F.16: Reference solution (Stokes flow), domain with 4096 randomly placed obstacles.



(a) Velocity magnitude. (b) Pressure field.

Figure F.17: MsFEM approximation (Stokes flow), $n = 1$, $H = 1/32$ (4096 coarse elements), domain with 4096 randomly placed obstacles.



(a) Velocity magnitude. (b) Pressure field.

Figure F.18: MsFEM approximation (Stokes flow), $n = 2$, $H = 1/32$(4096 coarse elements), domain with 4096 randomly placed obstacles.

F.2 Additional numerical results for Oseen flows

This section is the appendix of Chapter 15. As in Chapter 15, we solve the Oseen problem (12.1) on a coarse grid using basis functions defined by the Oseen equations, i.e. the local problems (12.8) and (12.9). In this section we consider only two-dimensional simulations. We consider the channel flow test case presented in Section 13.3.2. We choose a viscosity $\nu = 5 \times 10^{-4}$ and the Oseen velocity as $\boldsymbol{U_0} = [y(1-x^2), -x(1-y^2)]^t$ (note that $\text{div}\,(\boldsymbol{U_0}) = 0$), which leads to a Reynolds number of 2000.

F.2.1 Comparison between high-order MsFEMs and previous MsFEMs

In this section, we perform the convergence analysis proposed in Section 13.3.3. Figures F.19 to F.21 present the computed relative errors between the reference solutions computed on reference meshes and the MsFEM approximations computed on coarse meshes for high-order MsFEMs and previous MsFEMs developed in the literature.

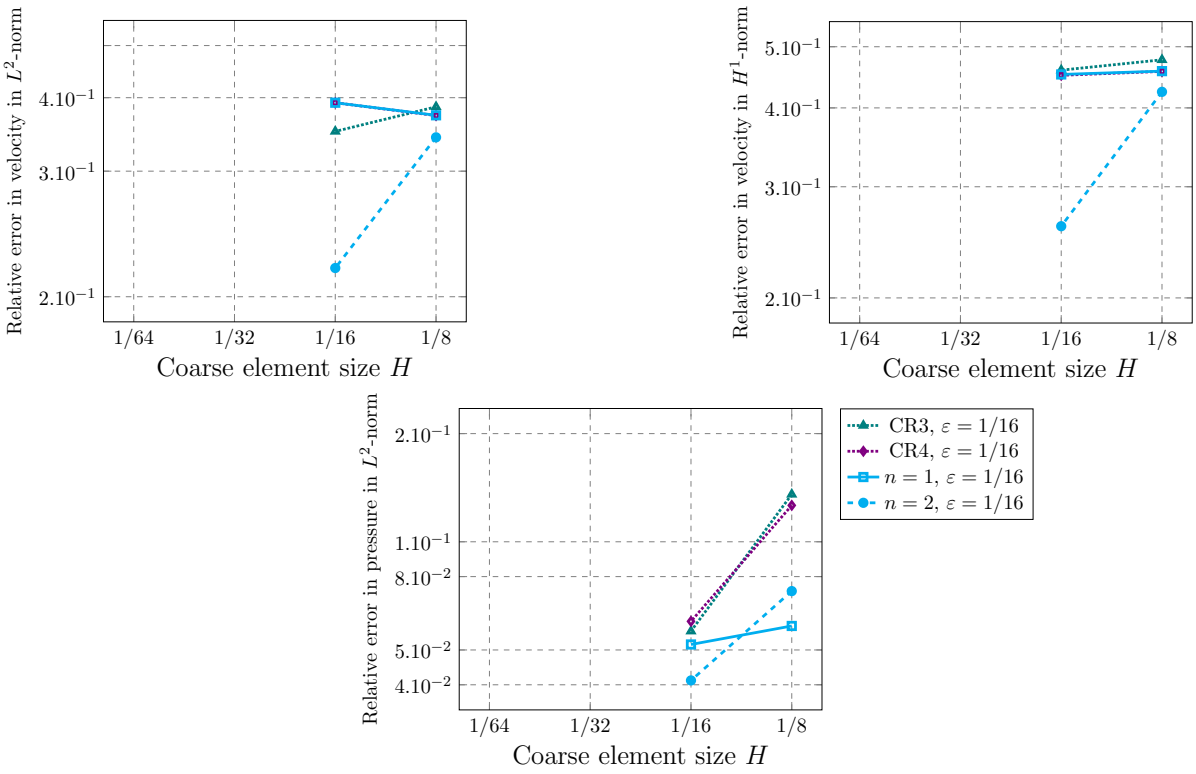


Figure F.19: Relative errors between MsFEM approximations and the reference solution, $\varepsilon = 1/16$ (Oseen flow).

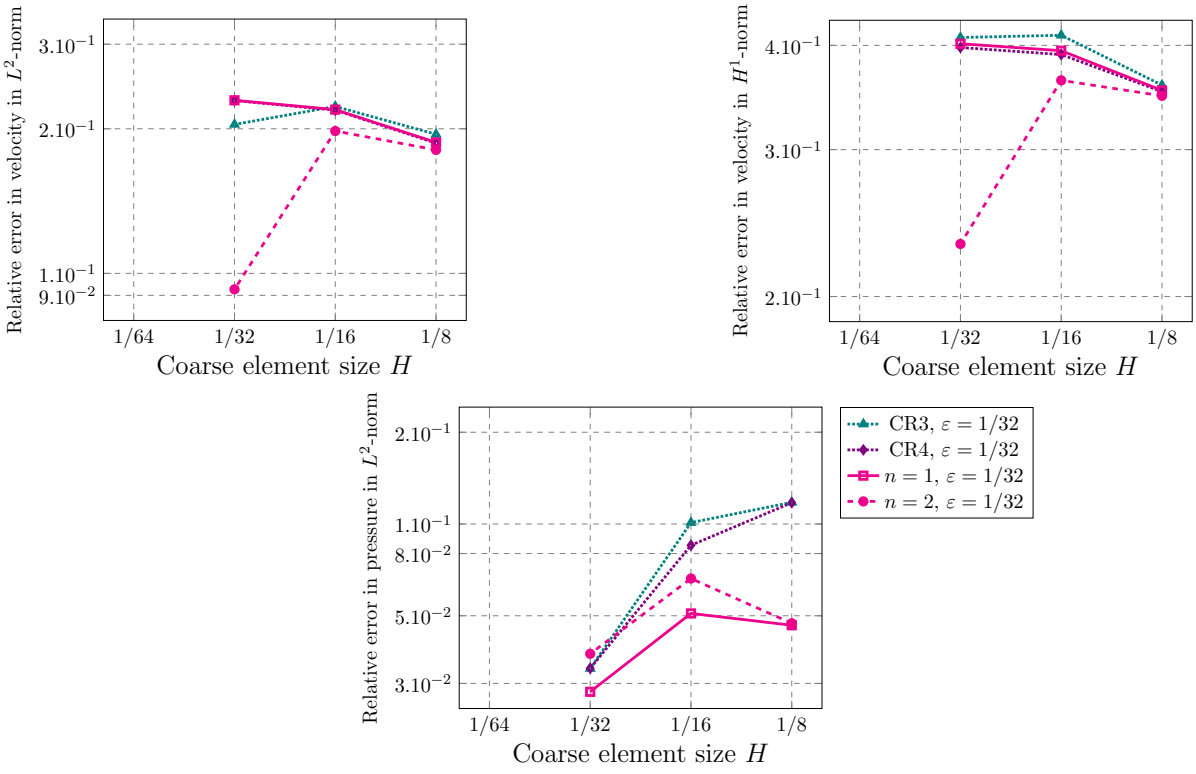


Figure F.20: Relative errors between MsFEM approximations and the reference solution, $\varepsilon = 1/32$ (Oseen flow).

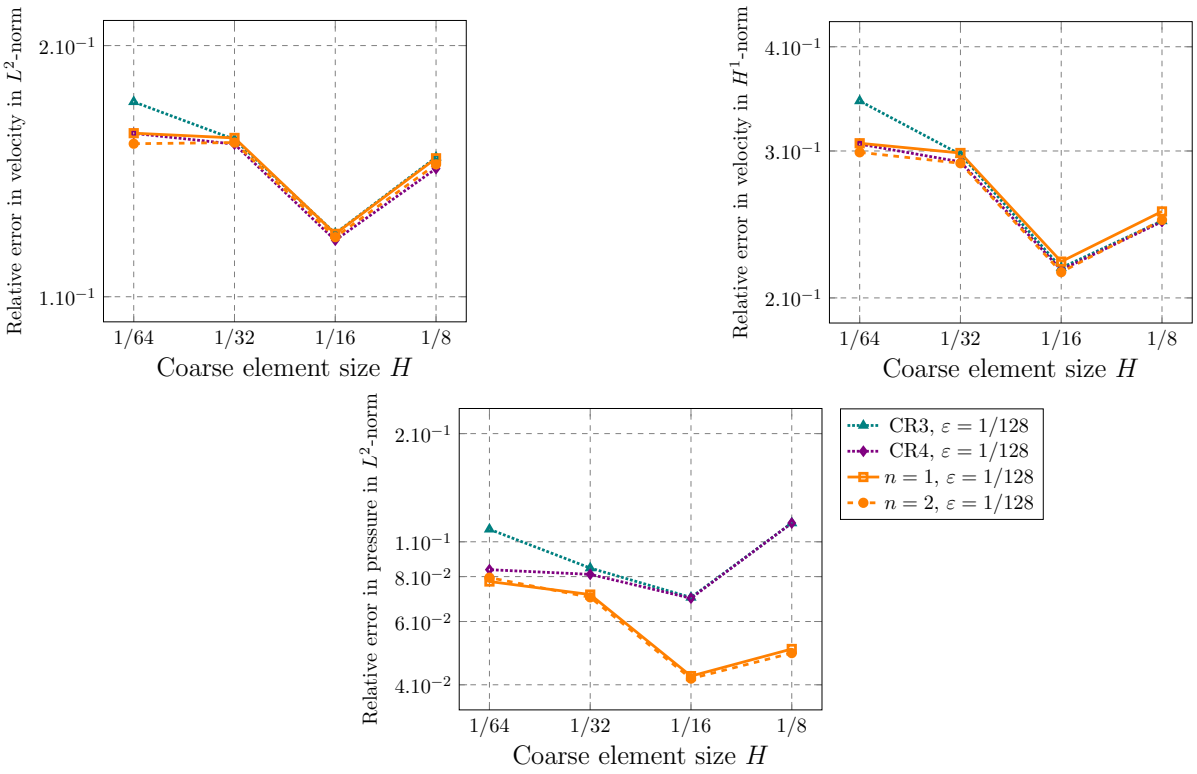
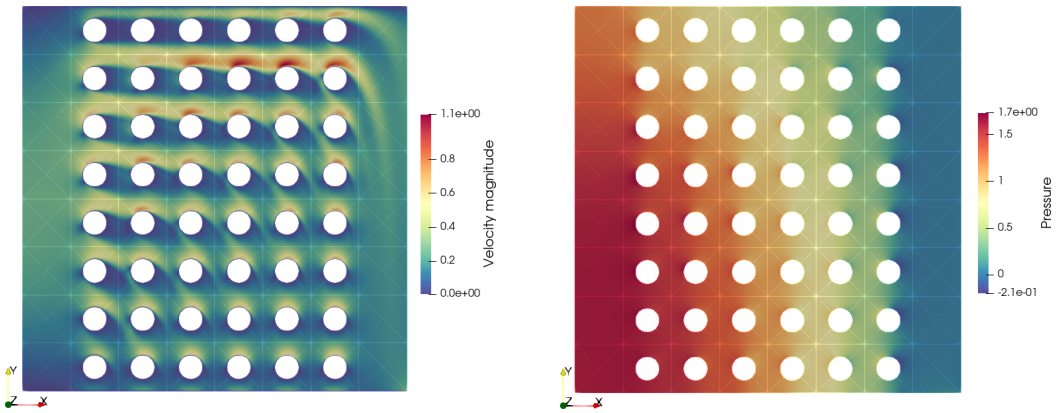


Figure F.21: Relative errors between MsFEM approximations and the reference solution, $\varepsilon = 1/128$ (Oseen flow).

F.2.2 Numerical simulations of Oseen flows

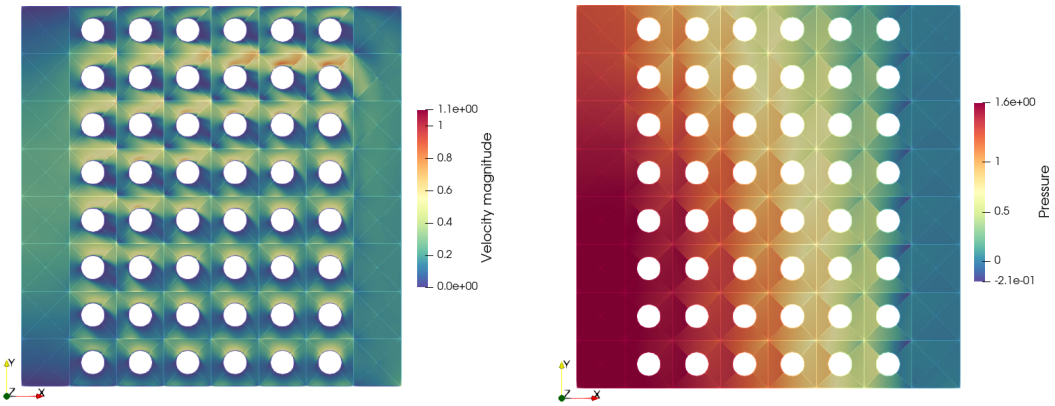
F.2.2.1 Oseen flow in a periodic domain $\varepsilon = 1/8$



(a) Velocity magnitude.

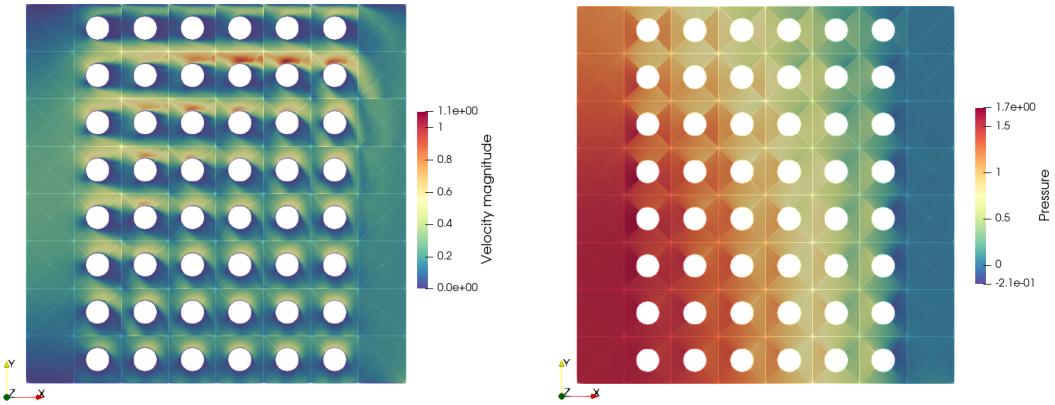
(b) Pressure field.

Figure F.22: Reference solution (Oseen flow), $\varepsilon = 1/8$.



(a) Velocity magnitude. (b) Pressure field.

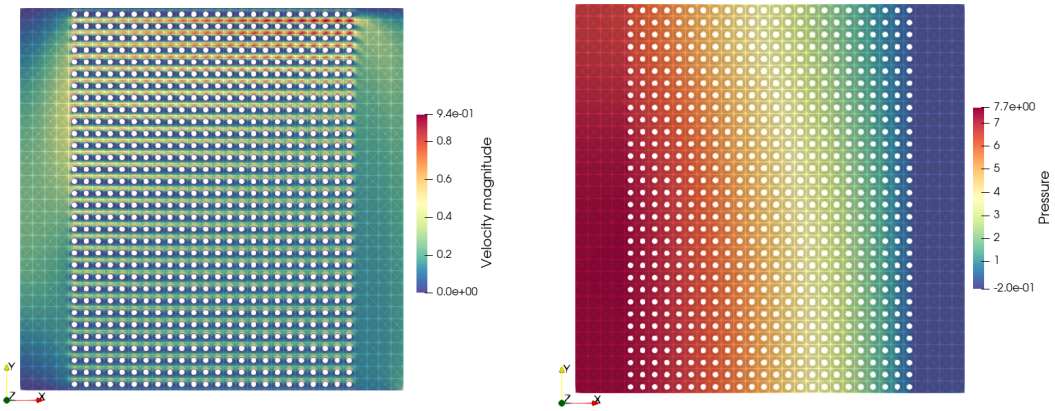
Figure F.23: MsFEM approximation (Oseen flow), $n = 1$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/8$.



(a) Velocity magnitude. (b) Pressure field.

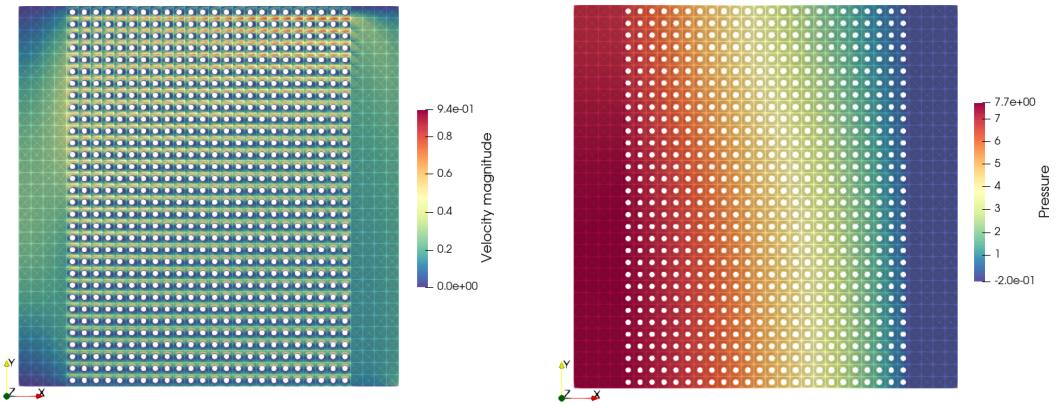
Figure F.24: MsFEM approximation (Oseen flow), $n = 2$, $H = 1/8$ (256 coarse elements), $\varepsilon = 1/8$.

F.2.2.2 Stokes flow in a periodic domain $\varepsilon = 1/32$



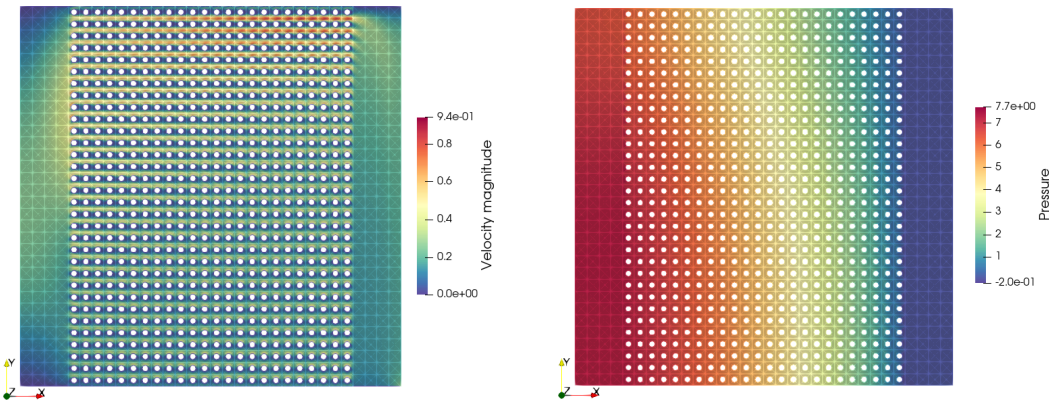
(a) Velocity magnitude. (b) Pressure field.

Figure F.25: Reference solution (Oseen flow), $\varepsilon = 1/32$.



(a) Velocity magnitude. (b) Pressure field.

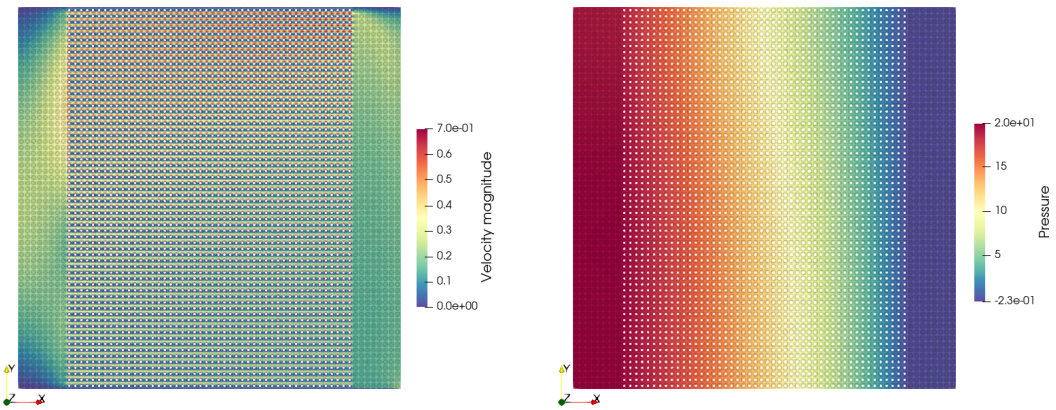
Figure F.26: MsFEM approximation (Oseen flow), $n = 1$, $H = 1/32$ (4096 coarse elements), $\varepsilon = 1/32$.



(a) Velocity magnitude. (b) Pressure field.

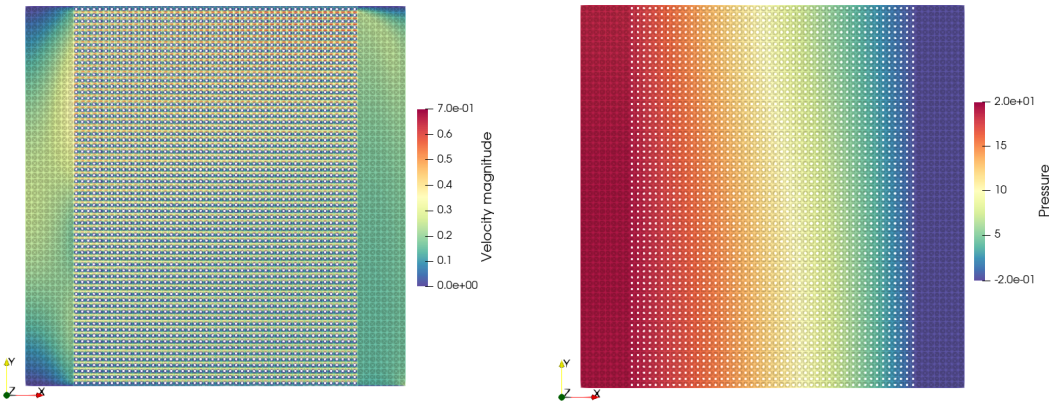
Figure F.27: MsFEM approximation (Oseen flow), $n = 2$, $H = 1/32$ (4096 coarse elements), $\varepsilon = 1/32$.

F.2.2.3 Oseen flow in a periodic domain $\varepsilon = 1/64$



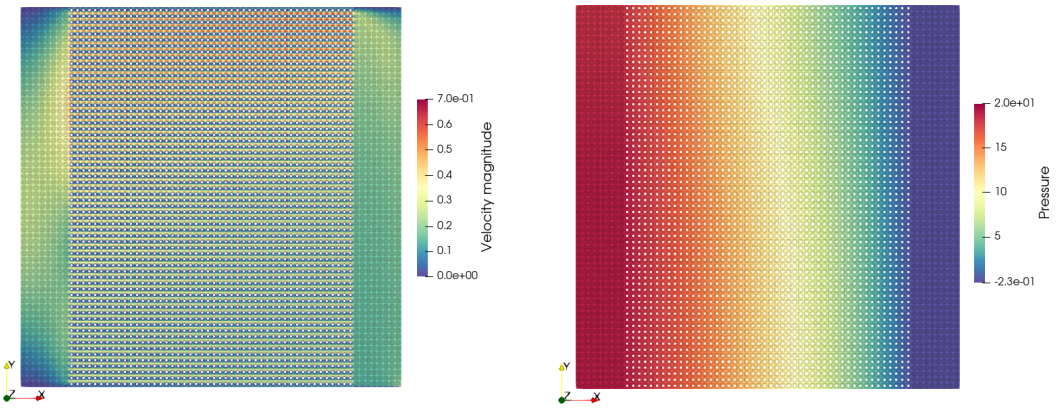
(a) Velocity magnitude. (b) Pressure field.

Figure F.28: Reference solution (Oseen flow), $\varepsilon = 1/64$.



(a) Velocity magnitude. (b) Pressure field.

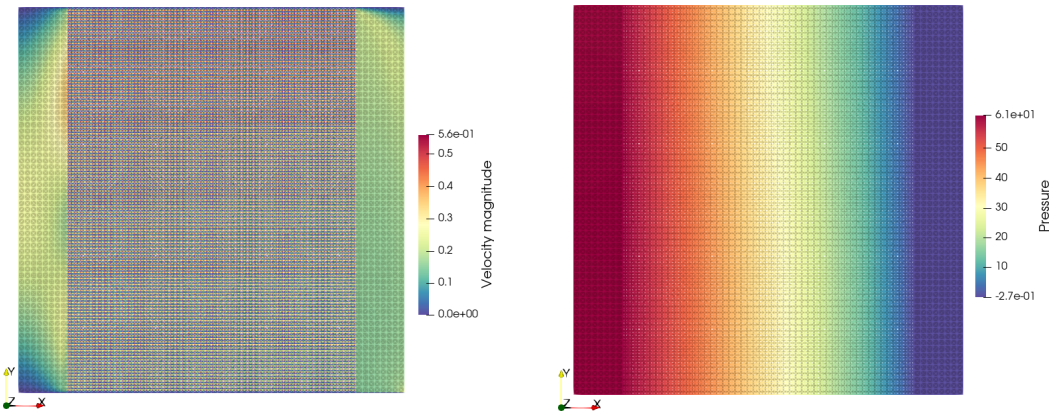
Figure F.29: MsFEM approximation (Oseen flow), $n = 1$, $H = 1/64$ (16384 coarse elements), $\varepsilon = 1/64$.



(a) Velocity magnitude. (b) Pressure field.

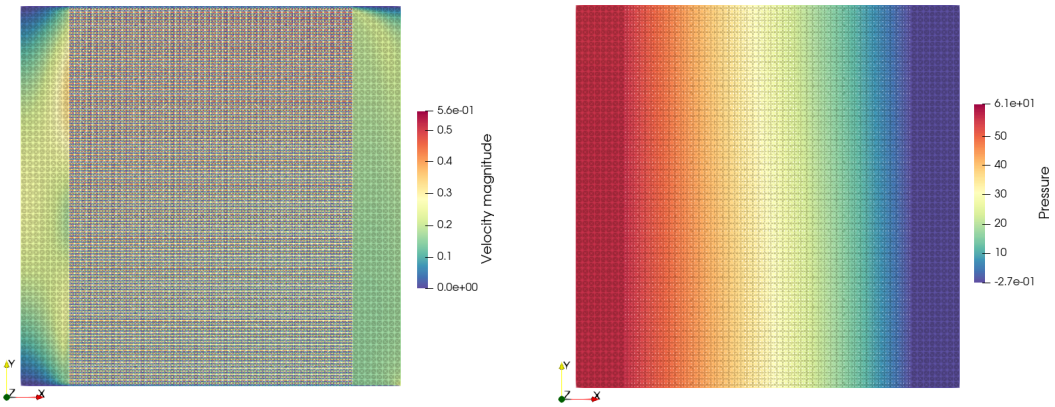
Figure F.30: MsFEM approximation (Oseen flow), $n = 2$, $H = 1/64$ (16384 coarse elements), $\varepsilon = 1/64$.

F.2.2.4 Oseen flow in a periodic domain $\varepsilon = 1/128$



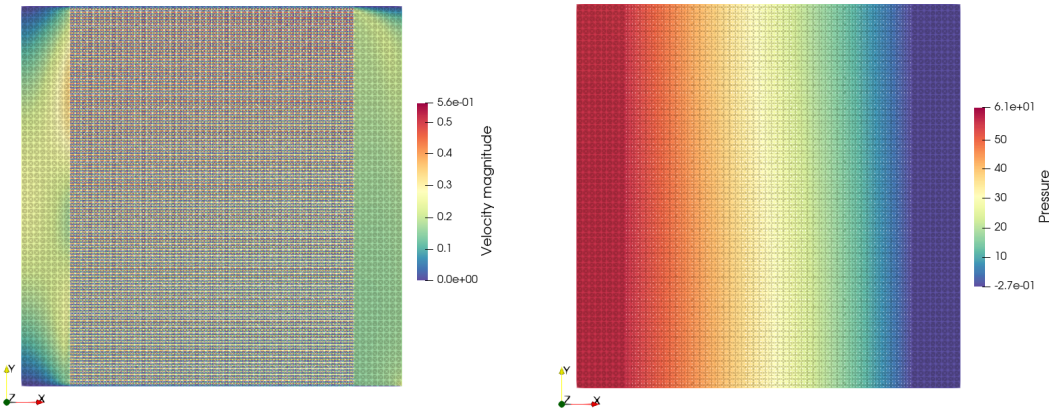
(a) Velocity magnitude. (b) Pressure field.

Figure F.31: Reference solution (Oseen flow), $\varepsilon = 1/128$.



(a) Velocity magnitude. (b) Pressure field.

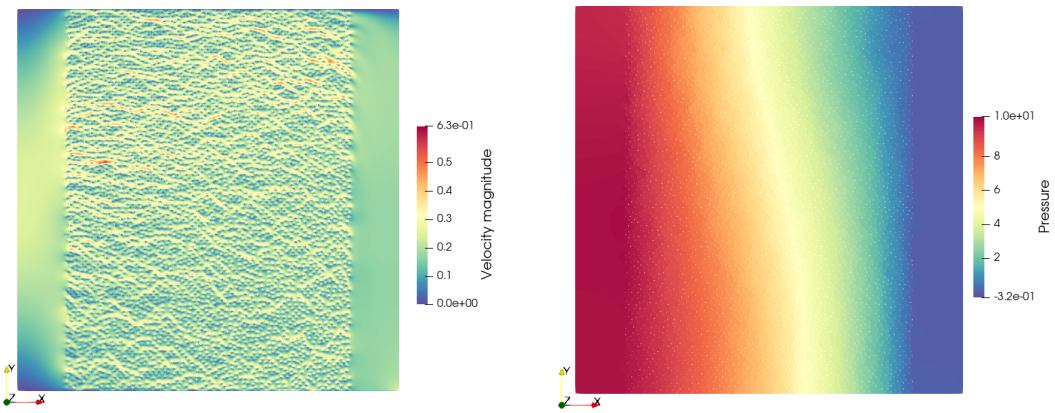
Figure F.32: MsFEM approximation (Oseen flow), $n = 1$, $H = 1/64$ (16384 coarse elements), $\varepsilon = 1/128$.



(a) Velocity magnitude. (b) Pressure field.

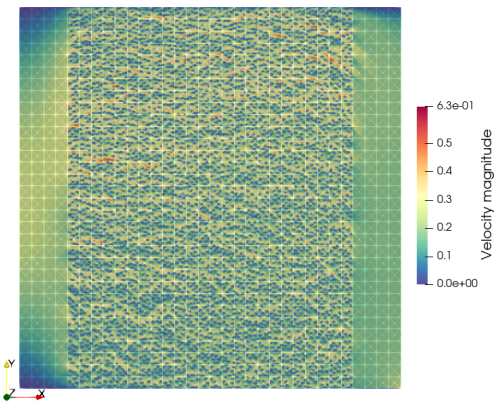
Figure F.33: MsFEM approximation (Oseen flow), $n = 2$, $H = 1/64$ (16384 coarse elements), $\varepsilon = 1/128$.

F.2.2.5 Oseen flow in a domain with 4096 randomly placed obstacles

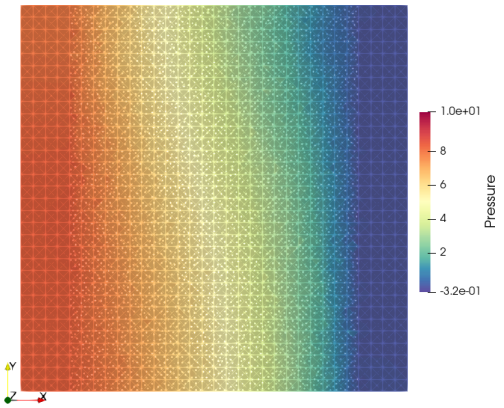


(a) Velocity magnitude. (b) Pressure field.

Figure F.34: Reference solution (Oseen flow), domain with 4096 randomly placed obstacles.

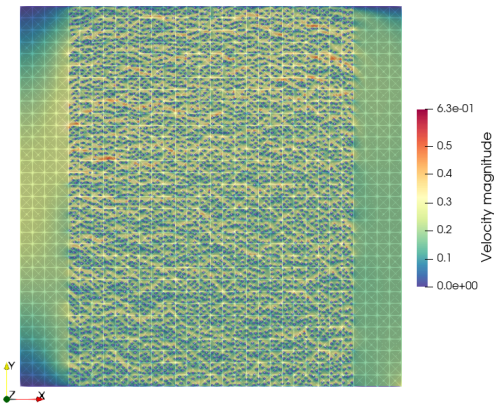


(a) Velocity magnitude.

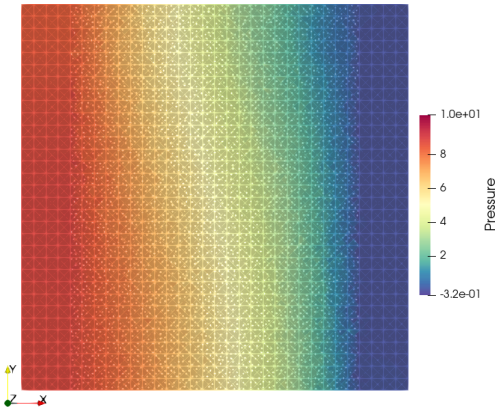


(b) Pressure field.

Figure F.35: MsFEM approximation (Oseen flow), $n = 1$, $H = 1/32$ (4096 coarse elements), domain with 4096 randomly placed obstacles.



(a) Velocity magnitude.



(b) Pressure field.

Figure F.36: MsFEM approximation (Oseen flow), $n = 2$, $H = 1/32$ (4096 coarse elements), domain with 4096 randomly placed obstacles.

Bibliography

[1] NRC, https://www.nrc.gov/.

[2] Nuclear Power Reactors - World Nuclear Association, https://world-nuclear.org/.

[3] numpy.linalg.matrix_rank — NumPy v1.26 Manual.

[4] Salome Platform - The open-source platform for numerical simulation.

[5] A. Abdulle and O. Budáč. An Adaptive Finite Element Heterogeneous Multiscale Method for Stokes Flow in Porous Media. *Multiscale Modeling & Simulation*, 13(1):256–290, 2015.

[6] A. Abdulle and P. Henning. Localized Orthogonal Decomposition method for the wave equation with a continuum of scales. *Mathematics of Computation*, 86(304):549–587, 2017.

[7] A. Abdulle and C. Schwab. Heterogeneous Multiscale FEM for diffusion problems on rough surfaces. *Multiscale Modeling and Simulation*, 3(1):195–220, 2004.

[8] H. Al Daas, P. Jolivet, F. Nataf, and P.-H Tournier. A robust two-level schwarz preconditioner for sparse matrices. *arXiv preprint arXiv:12401.03915*, 2024.

[9] G. Allaire. Homogenization of the Stokes flow in a connected porous medium. *Asymptotic Analysis*, 2(3):203–222, 1989.

[10] G. Allaire. A brief introduction to homogenization and miscellaneous applications. *ESAIM: Proceedings*, 37:1–49, 2012.

[11] G. Allaire and M. Amar. Boundary layer tails in periodic homogenization. *ESAIM, Control Optim. Calc. Var.*, 4:209–243, 1999.

[12] G. Allaire and R. Brizzi. A Multiscale Finite Element Method for Numerical Homogenization. *Multiscale Modeling & Simulation*, 4(3):790–812, 2005.

[13] R. Altmann, P. Henning, and D. Peterseim. Numerical homogenization beyond scale separation. *Acta Numerica*, 30:1–86, 2021.

[14] P.R. Amestoy, A. Buttari, J.-Y. L'Excellent, and T. Mary. Performance and Scalability of the Block Low-Rank Multifrontal Factorization on Multicore Architectures. *ACM Transactions on Mathematical Software*, 45(1):1–26, 2019.

[15] P.R. Amestoy, I.S. Duff, J.-Y. L'Excellent, and J. Koster. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal on Matrix Analysis and Applications*, 23(1):15–41, 2001.

[16] P.-E. Angeli, U. Bieder, and G. Fauchet. Overview of the TrioCFD code: Main features, V&V procedures and typical applications to nuclear engineering. In *NURETH-16, Chicago*, 2015.

[17] P.-E. Angeli, M.-A. Puscas, G. Fauchet, and A. Cartalade. FVCA8 Benchmark for the Stokes and Navier–Stokes Equations with the TrioCFD Code—Benchmark Session. In *Finite Volumes for Complex Applications VIII - Methods and Theoretical Aspects*, volume 199, pages 181–202, 2017.

[18] R. Araya, C. Cárcamo, A.H. Poza, and F. Valentin. An Adaptive Multiscale Hybrid-Mixed Method for the Oseen equations. *Advances in Computational Mathematics*, 47(1):15, 2021.

[19] R. Araya, C. Harder, D. Paredes, and F. Valentin. Multiscale Hybrid-Mixed Method. *SIAM Journal on Numerical Analysis*, 51(6):3505–3531, 2013.

[20] R. Araya, C. Harder, A.H. Poza, and F. Valentin. Multiscale Hybrid-Mixed Method for the Stokes and Brinkman equations - The method. *Computer Methods in Applied Mechanics and Engineering*, 324:29–53, 2017.

[21] R. Araya, R. Rebolledo, and F. Valentin. On a multiscale *a posteriori* error estimator for the stokes and Brinkman equations. *IMA Journal of Numerical Analysis*, 41(1):344–380, 2021.

[22] S. Armstrong, T. Kuusi, and J.-C. Mourrat. *Quantitative Stochastic Homogenization and Large-Scale Regularity*, volume 352 of *Grundlehren der mathematischen Wissenschaften*. Springer International Publishing, Cham, 2019.

[23] I. Babuška. The Finite Element Method with Lagrangian multipliers. *Numerische Mathematik*, 20(3):179–192, 1973.

[24] I. Babuška, G. Caloz, and J.E. Osborn. Special Finite Element Methods for a class of second order elliptic problems with rough coefficients. *SIAM Journal on Numerical Analysis*, 31(4):945–981, 1994.

[25] M.A. Badri. medio: An Open-source Library Bridging Finite Element Kernels and SALOME for Enhanced Mesh Data Interoperability, 2023. original-date: 2023-07-10T08:33:34Z.

[26] N. Bakhvalov and G. Panasenko. *Homogenisation: Averaging Processes in Periodic Media*, volume 36 of *Mathematics and its Applications*. Springer Netherlands, Dordrecht, 1989.

[27] S. Balay, S. Abhyankar, M.F. Adams, S. Benson, J. Brown, P. Brune, K. Buschelman, E. Constantinescu, L. Dalcin, A. Dener, V. Eijkhout, J. Faibussowitsch, W.D. Gropp, V. Hapla, T. Isaac, P. Jolivet, D. Karpeev, D. Kaushik, M.G. Knepley, F. Kong,

S. Kruger, D.A May, L.C. McInnes, R.T. Mills, L. Mitchell, T. Munson, J.E. Roman, K. Rupp, P. Sanan, J. Sarich, B.F. Smith, S. Zampini, H. Zhang, H. Zhang, and J. Zhang. PETSc/TAO users manual. Technical Report ANL-21/39 - Revision 3.20, Argonne National Laboratory, 2023.

[28] S. Balay, S. Abhyankar, M.F. Adams, S. Benson, J. Brown, P. Brune, K. Buschelman, E.M. Constantinescu, L. Dalcin, A. Dener, V. Eijkhout, J. Faibussowitsch, W.D. Gropp, V. Hapla, T. Isaac, P. Jolivet, D. Karpeev, D. Kaushik, M.G. Knepley, F. Kong, S. Kruger, D.A. May, L.C. McInnes, R.T. Mills, L. Mitchell, T. Munson, J.E Roman, K. Rupp, P. Sanan, J. Sarich, B.F. Smith, S. Zampini, H. Zhang, H. Zhang, and J. Zhang. PETSc Web page. https://petsc.org/, 2023.

[29] L. Balazi, G. Allaire, P. Jolivet, and P. Omnes. Inf-sup stable non-conforming finite elements on tetrahedra with second and third order accuracy. https://hal.science/hal-04541809, 2024.

[30] L. Balazi, G. Allaire, and P. Omnes. Sharp convergence rates for the homogenization of the Stokes equations in a perforated domain. *Discrete and Continuous Dynamical Systems - B*, 2024.

[31] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*. Providence, RI: AMS Chelsea Publishing, reprint of the 1978 original with corrections and bibliographical additions edition, 2011.

[32] U. Bieder, G. Fauchet, C. Calvin, and P. Ledac. High performance Large Eddy Simulation of turbulent flows around PWR mixing grids. In *Proceedings of 16th IEEE International Conference on High Performance Computing and Communications (HPCC 2014)*, 2014.

[33] R. Biezemans. *Multiscale methods : non-intrusive implementation, advection-dominated problems and related topics*. PhD Thesis, École des Ponts ParisTech, 2023.

[34] D. Boffi. Three-Dimensional Finite Element Methods for the Stokes Problem. *SIAM Journal on Numerical Analysis*, 34(2):664–670, 1997.

[35] D. Boffi, F. Brezzi, and M. Fortin. Finite elements for the Stokes Problem. In *Mixed Finite Elements, Compatibility Conditions, and Applications*, volume 1939, pages 45–100. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[36] M. Boulakia, C. Grandmont, F. Lespagnol, and P. Zunino. Reduced models for the Poisson problem in perforated domains. working paper or preprint, 2021.

[37] M. Braack, E. Burman, V. John, and G. Lube. Stabilized Finite Element Methods for the generalized Oseen problem. *Computer Methods in Applied Mechanics and Engineering*, 196(4):853–866, 2007.

[38] S.C. Brenner, J.C. Garay, and L.-Y. Sung. Multiscale Finite Element Methods for an Elliptic Optimal Control Problem with Rough Coefficients. *Journal of Scientific Computing*, 91(3):76, 2022.

[39] S.C. Brenner, J.C. Garay, and L.-Y. Sung. A Multiscale Finite Element Method for an Elliptic Distributed Optimal Control Problem with Rough Coefficients and Control Constraints, 2023. arXiv:2309.16062 [cs, math].

[40] F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from lagrangian multipliers. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 8(R2):129–151, 1974.

[41] J. Brown, M.G. Knepley, D.A. May, L.C. McInnes, and B. Smith. Composable linear solvers for multiphysics. In *2012 11th International Symposium on Parallel and Distributed Computing*, pages 55–62, 2012.

[42] J. Brown, M.G. Knepley, D.A. May, L.C. McInnes, and B. Smith. Composable Linear Solvers for Multiphysics. In *2012 11th International Symposium on Parallel and Distributed Computing*, pages 55–62, Munich, Germany, 2012. IEEE.

[43] X.-C. Cai and M. Sarkis. A restricted additive schwarz preconditioner for general sparse linear systems. *SIAM Journal on Scientific Computing*, 21(2):792–797, 1999.

[44] L. Cattabriga. Su un problema al contorno relativo al sistema di equazioni di Stokes. *Rend. Semin. Mat. Univ. Padova*, 31:308–340, 1961.

[45] CEA. TRIOCFD, https://triocfd.cea.fr, 2023.

[46] L. Chamoin and F. Legoll. A posteriori error estimation and adaptive strategy for the control of MsFEM computations. *Computer Methods in Applied Mechanics and Engineering*, 336:1–38, 2018.

[47] T. Chaumont-Frelet and F. Valentin. A Multiscale Hybrid-Mixed Method for the Helmholtz equation in heterogeneous domains. *SIAM Journal on Numerical Analysis*, 58(2):1029–1067, 2020.

[48] G.A. Chechkin, A.L. Piatnitski, and A.S. Shamaev. *Homogenization. Methods and applications. Transl. from the Russian by Tamara Rozhkovskaya*, volume 234 of *Transl. Math. Monogr.* Providence, RI: American Mathematical Society (AMS), 2007.

[49] L. Chen. A simple construction of a Fortin operator for the two dimensional Taylor–Hood element. *Computers & Mathematics with Applications*, 68(10):1368–1373, 2014.

[50] W. Chen, J. Hu, and M. Zhang. Nonconforming Finite Element Methods of order two and order three for the Stokes flow in three dimensions, 2022. arXiv:2212.11748 [cs, math].

[51] Y. Chen, L.J. Durlofsky, M. Gerritsen, and X.H. Wen. A coupled local–global upscaling approach for simulating flow in highly heterogeneous formations. *Advances in Water Resources*, 26(10):1041–1060, 2003.

[52] E. Chung, Y. Efendiev, W. Leung, and J. Ren. Multiscale Simulations for Coupled Flow and Transport Using the Generalized Multiscale Finite Element Method. *Computation*, 3(4):670–686, 2015.

[53] E.T. Chung, Y. Efendiev, W.T. Leung, M. Vasilyeva, and Y. Wang. Non-local multi-continua upscaling for flows in heterogeneous fractured media. *Journal of Computational Physics*, 372:22–34, 2018.

[54] E.T. Chung, Y. Efendiev, and G. Li. An adaptive GMsFEM for high-contrast flow problems. *Journal of Computational Physics*, 273:54–76, 2014.

[55] E.T. Chung, U. Kalachikova, M. Vasilyeva, and V. Alekseev. Generalized Multiscale Discontinuous Galerkin Method for convection–diffusion equation in perforated media. *Mathematics and Computers in Simulation*, 193:666–688, 2022.

[56] A.G. Churbanov, N.G. Churbanova, and M.A. Trapeznikova. Coupled Prediction of Flows in Domains Containing a Porous Medium and a Free Stream. *Mathematical Models and Computer Simulations*, 15(4):643–653, 2023.

[57] D. Cioranescu and P. Donato. *An introduction to homogenization*. Number 17 in Oxford lecture series in mathematics and its applications. Oxford University Press, Oxford ; New York, 1999.

[58] D. Cioranescu and J.S.J. Paulin. Homogenization in open sets with holes. *Journal of Mathematical Analysis and Applications*, 71(2):590–607, 1979.

[59] I. Clifford, M. Pecchia, R. Puragliesi, A. Vasiliev, and H. Ferroukhi. On the characteristics of the flow and heat transfer in the core bypass region of a PWR. *Nuclear Engineering and Design*, 330:117–128, 2018.

[60] B. Cockburn, D.A. Di Pietro, and A. Ern. Bridging the Hybrid High-Order and hybridizable discontinuous Galerkin methods. *ESAIM: Mathematical Modelling and Numerical Analysis*, 50(3):635 – 650, 2016.

[61] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming Finite Element Methods for solving the stationary Stokes equations I. *Revue française d'automatique informatique recherche opérationnelle. Mathématique*, 7(R3):33–75, 1973.

[62] B. Dacorogna. Existence and regularity of solutions of $d\omega = f$ with Dirichlet boundary conditions. In *Nonlinear problems in mathematical physics and related topics I. In honor of Professor O. A. Ladyzhenskaya*, pages 67–82. New York, NY: Kluwer Academic/Plenum Publishers, 2002.

[63] P. Degond, A. Lozinski, B.P. Muljadi, and J. Narski. Crouzeix–Raviart MsFEM with Bubble Functions for Diffusion and Advection-Diffusion in Perforated Media. *Communications in Computational Physics*, 17(4):887–907, 2015.

[64] D.A. Di Pietro and R. Tittarelli. An Introduction to Hybrid High-Order Methods. *SEMA SIMAI Springer Series*, 15:75 – 128, 2018.

[65] V. Dolean, P. Jolivet, and F. Nataf. *An Introduction to Domain Decomposition Methods*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, 2015.

[66] W. E and B. Engquist. The Heterogeneous Multi-scale Method for homogenization problems. *Lecture Notes in Computational Science and Engineering*, 44:89–110, 2005.

[67] W. E, B. Engquist, and Z. Huang. Heterogeneous Multiscale Method: A general methodology for multiscale modeling. *Physical Review B - Condensed Matter and Materials Physics*, 67(9):921011–921014, 2003.

[68] W. E, P. Ming, and P. Zhang. Analysis of the Heterogeneous Multiscale Method for elliptic homogenization problems. *Journal of the American Mathematical Society*, 18(1):121–156, 2005.

[69] Y. Efendiev, J. Galvis, and T.Y. Hou. Generalized Multiscale Finite Element Methods (GMsFEM). *Journal of Computational Physics*, 251:116–135, 2013.

[70] Y. Efendiev and T.Y. Hou. *Multiscale Finite Element Methods: theory and applications.* Number 4 in Surveys and tutorials in the applied mathematical sciences. Springer, New York, NY, 2009.

[71] Y. Efendiev, T.Y. Hou, and X.-H. Wu. Convergence of a Nonconforming Multiscale Finite Element Method. *SIAM Journal on Numerical Analysis*, 37(3):888–910, 2000.

[72] D. Elfverson, E.H. Georgoulis, and A. Målqvist. An Adaptive Discontinuous Galerkin Multiscale Method for Elliptic Problems. *Multiscale Modeling & Simulation*, 11(3):747–765, 2013.

[73] D. Elfverson, E.H. Georgoulis, A. Målqvist, and D. Peterseim. Convergence of a Discontinuous Galerkin Multiscale Method. *SIAM Journal on Numerical Analysis*, 51(6):3351–3372, 2013.

[74] H.C. Elman, D.J. Silvester, and A.J. Wathen. *Finite elements and fast iterative solvers. With applications in incompressible fluid dynamics.* Numer. Math. Sci. Comput. Oxford: Oxford University Press, 2nd ed. edition, 2014.

[75] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*, volume 159 of *Applied Mathematical Sciences.* Springer New York, New York, NY, 2004.

[76] A. Ern and J.-L. Guermond. *Finite elements. 1: Approximation and interpolation.* Number Volume 72 in Texts in applied mathematics. Springer, Cham, Switzerland, 2021.

[77] R. Ewing, O. Iliev, R. Lazarov, I. Rybak, and J. Willems. A Simplified Method for Upscaling Composite Materials with High Contrast of the Conductivity. *SIAM Journal on Scientific Computing*, 31(4):2568–2586, 2009.

[78] Q. Feng. *Development of a multiscale finite element method for incompressible flows in heterogeneous media.* PhD Thesis, Université Paris Saclay (COmUE), 2019.

[79] Q. Feng, G. Allaire, and P. Omnes. Enriched Nonconforming Multiscale Finite Element Method for Stokes Flows in Heterogeneous Media Based on High-order Weighting Functions. *Multiscale Modeling & Simulation*, pages 462–492, 2022.

[80] M. Fortin. An analysis of the convergence of mixed Finite Element Methods. *RAIRO. Analyse numérique*, 11(4):341–354, 1977.

[81] D. Gérard-Varet and N. Masmoudi. Homogenization in polygonal domains. *Journal of the European Mathematical Society*, 013(5):1477–1503, 2011.

[82] D. Gérard-Varet and N. Masmoudi. Homogenization and boundary layers. *Acta Mathematica*, 209(1):133–178, 2012.

[83] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*, volume 224 of *Classics in Mathematics*. Springer, Berlin, Heidelberg, 2001.

[84] V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes equations. Theory and algorithms.*, volume 5 of *Springer Ser. Comput. Math.* Springer, Cham, (Extended version of the 1979 publ.) edition, 1986.

[85] V. Girault, D. Vassilev, and I. Yotov. Mortar Multiscale Finite Element Methods for Stokes–Darcy flows. *Numerische Mathematik*, 127(1):93–165, 2014.

[86] P. Grisvard. *Elliptic problems in nonsmooth domains*. Number 69 in Classics in applied mathematics. Society for Industrial and Applied Mathematics, Philadelphia, siam ed edition, 2011.

[87] H. Hajibeygi and P. Jenny. A General Multiscale Finite-Volume Method for compressible multiphase flow in porous media. In *ECMOR XI - 11th European Conference on the Mathematics of Oil Recovery*, 2008.

[88] H. Hajibeygi and P. Jenny. Multiscale Finite-Volume Method for parabolic problems arising from compressible multiphase flow in porous media. *Journal of Computational Physics*, 228(14):5129–5147, 2009.

[89] K. Hanjalić and B.E. Launder. A Reynolds stress model of turbulence and its application to thin shear flows. *Journal of Fluid Mechanics*, 52(4):609–638, 1972.

[90] A. Hannukainen, R. Stenberg, and M. Vohralík. A unified framework for a posteriori error estimation for the Stokes problem. *Numerische Mathematik*, 122(4):725–769, 2012.

[91] C. Harder, D. Paredes, and F. Valentin. A family of Multiscale Hybrid-Mixed Finite Element Methods for the Darcy equation with rough coefficients. *Journal of Computational Physics*, 245:107–130, 2013.

[92] C. Harder, D. Paredes, and F. Valentin. On a Multiscale Hybrid-Mixed Method for advective-reactive dominated problems with heterogeneous coefficients. *Multiscale Modeling and Simulation*, 13(2):491–518, 2015.

[93] F. Hecht. New development in FreeFEM++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[94] B. Helenbrook. On the Existence of Explicit *hp*-Finite Element Methods using Gauss–Lobatto Integration on the Triangle. *SIAM J. Numerical Analysis*, 47:1304–1318, 2009.

[95] F. Hellman and A. Målqvist. Numerical homogenization of elliptic PDEs with similar coefficients. *Multiscale Modeling and Simulation*, 17(2):650–674, 2019.

[96] J.-P. Hennart, J. Jaffre, and J.E. Roberts. A constructive method for deriving finite elements of nodal type. *Numerische Mathematik*, 53(6):701–738, 1988.

[97] P. Henning and A. Målqvist. Localized Orthogonal Decomposition techniques for boundary value problems. *SIAM Journal on Scientific Computing*, 36(4):A1609–A1634, 2014.

[98] P. Henning, A. Målqvist, and D. Peterseim. A localized Orthogonal Decomposition method for semi-linear elliptic problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(5):1331–1349, 2014.

[99] P. Henning, M. Ohlberger, and B. Schweizer. An Adaptive Multiscale Finite Element Method. *SIAM Journal on Multiscale Modeling and Simulation*, 12:1078, 2014.

[100] V. Hernandez, E.R. Jose, and V. Vidal. SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Trans. Math. Software*, 31(3):351–362, 2005.

[101] J.S. Hesthaven, S. Zhang, and X. Zhu. High-Order Multiscale Finite Element Method for Elliptic Problems. *Multiscale Modeling & Simulation*, 12(2):650–666, 2014.

[102] U. Hornung. *Homogenization and porous media*, volume 6 of *Interdiscip. Appl. Math.* New York, NY: Springer, 1997.

[103] T. Hou, X.-H. Wu, and Z. Cai. Convergence of a Multiscale Finite Element Method for Elliptic Problems with rapidly oscillating coefficients. *Mathematics of Computation*, 68(227):913–943, 1999.

[104] T.Y. Hou and X.-H. Wu. A Multiscale Finite Element Method for Elliptic Problems in Composite Materials and Porous Media. *Journal of Computational Physics*, 134(1):169–189, 1997.

[105] 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. *Communications in Mathematical Sciences*, 2(2):185–205, 2004.

[106] T.J.R. Hughes. Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods. *Computer Methods in Applied Mechanics and Engineering*, 127(1):387–401, 1995.

[107] T.J.R. Hughes. Multiscale phenomena: Green's functions, the dirichlet-to-neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods. *Computer Methods in Applied Mechanics and Engineering*, 127(1):387–401, 1995.

[108] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, and J.-B. Quincy. The Variational Multiscale Method-a paradigm for computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 166(1):3–24, 1998.

[109] O. Iliev, Z. Lakdawala, and V. Starikovicius. On a numerical subgrid upscaling algorithm for Stokes–Brinkman equations. *Computers & Mathematics with Applications*, 65(3):435–448, 2013.

[110] O. Iliev, R. Lazarov, and J. Willems. Fast numerical upscaling of heat equation for fibrous materials. *Computing and Visualization in Science*, 13(6):275–285, 2010.

[111] G. Jankowiak and A. Lozinski. Non-conforming Multiscale Finite Element Method for Stokes flows in heterogeneous media. part II: Error estimates for periodic microstructure. *Discrete and Continuous Dynamical Systems - B*, 29(5):2298–2332, 2024.

[112] P. Jenny, S.H. Lee, and H.A. Tchelepi. Multi-scale Finite-Volume Method for elliptic problems in subsurface flow simulation. *Journal of Computational Physics*, 187(1):47–67, 2003.

[113] P. Jenny, S.H. Lee, and H.A. Tchelepi. Adaptive Multiscale Finite-Volume Method for multiphase flow and transport in porous media. *Multiscale Modeling and Simulation*, 3(1):50–64, 2004.

[114] P. Jenny, S.H. Lee, and H.A. Tchelepi. Adaptive fully implicit Multi-scale Finite-Volume Method for multi-phase flow and transport in heterogeneous porous media. *Journal of Computational Physics*, 217(2):627–641, 2006.

[115] P. Jolivet, F. Hecht, F. Nataf, and C. Prud'Homme. Scalable Domain Decomposition Preconditioners for Heterogeneous Elliptic Problems. In *SC13: International Conference for High Performance Computing, Networking, Storage and Analysis*, page 80:1. ACM, 2013.

[116] P. Jolivet, J.E. Roman, and S. Zampini. KSPHPDDM and PCHPDDM: Extending PETSc with advanced Krylov methods and robust multilevel overlapping Schwarz preconditioners. *Computers & Mathematics with Applications*, 84:277–295, 2021.

[117] P. Jolivet and P.-H Tournier. Block Iterative Methods and Recycling for Improved Scalability of Linear Solvers. In *SC16: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 190–203, Salt Lake City, UT, USA, 2016. IEEE.

[118] G. Karypis and V. Kumar. *METIS: A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices*, 1998.

[119] S. Lanteri, D. Paredes, C. Scheid, and F. Valentin. The Multiscale Hybrid-Mixed Method for the Maxwell equations in heterogeneous media. *Multiscale Modeling and Simulation*, 16(4):1648–1683, 2018.

[120] D. Lasseux and F.J. Valdés-Parada. A macroscopic model for immiscible two-phase flow in porous media. *Journal of Fluid Mechanics*, 944:A43, 2022.

[121] C. Le Bris, F. Legoll, and A. Lozinski. MsFEM à la Crouzeix-Raviart for Highly Oscillatory Elliptic Problems. *Chinese Annals of Mathematics, Series B*, 34(1):113–138, 2013.

[122] C. Le Bris, F. Legoll, and A. Lozinski. An MsFEM Type Approach for Perforated Domains. *Multiscale Modeling & Simulation*, 12(3):1046–1077, 2014.

[123] F. Legoll, P.-L. Rothé, C. Le Bris, and U. Hetmaniuk. An MsFEM Approach Enriched Using Legendre Polynomials. *Multiscale Modeling & Simulation*, 20(2):798–834, 2022.

[124] D.S. Lopes, M.T. Silva, and J.A. Ambrósio. Tangent vectors to a 3-D surface normal: A geometric tool to find orthogonal vectors based on the Householder transformation. *Computer-Aided Design*, 45(3):683–694, 2013.

[125] G. Lube and G. Rapin. Residual-based stabilized higher-order FEM for a generalized Oseen problem. *Mathematical Models and Methods in Applied Sciences*, 16(7):949–966, 2006.

[126] I. Lunati and P. Jenny. Multiscale Finite-Volume Method for compressible multiphase flow in porous media. *Journal of Computational Physics*, 216(2):616–636, 2006.

[127] I. Lunati and P. Jenny. Treating highly anisotropic subsurface flow with the Multiscale Finite-Volume Method. *Multiscale Modeling and Simulation*, 6(1):308–318, 2007.

[128] F. Madiot. *Multiscale finite element methods for advection-diffusion problems*. PhD Thesis, Université Paris-Est, 2016.

[129] R.H. Magnus and S. Eduard. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49:409–435, 1952.

[130] E. Marusic-Paloka and A. Mikelić. An Error Estimate for Correctors in the Homogenization of the Stokes and the Navier-Stokes Equations in a Porous Medium. *Boll. Un. Mat. Ital. A (7)*, 10:661–671, 1996.

[131] G. Matthies. Inf-sup stable nonconforming finite elements of higher order on quadrilaterals and hexahedra. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(5):855–874, September 2007.

[132] G. Matthies and L. Tobiska. Inf-sup stable non-conforming finite elements of arbitrary order on triangles. *Numerische Mathematik*, 102(2):293–309, 2005.

[133] J.M. Melenk and I. Babuška. The partition of unity Finite Element Method: Basic theory and applications. *Computer Methods in Applied Mechanics and Engineering*, 139(1):289–314, 1996.

[134] S. Moskow and M. Vogelius. First-order corrections to the homogenised eigenvalues of a periodic composite medium. A convergence proof. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 127(6):1263–1299, 1997.

[135] B.P. Muljadi. Multiscale Method for Oseen Problem in Porous Media with Non-periodic Grain Patterns. *Transport in Porous Media*, 116(1):1–18, 2017.

[136] B.P. Muljadi, J. Narski, A. Lozinski, and P. Degond. Nonconforming Multiscale Finite Element Method for Stokes Flows in Heterogeneous Media. Part I: Methodologies and Numerical Experiments. *Multiscale Modeling & Simulation*, 13(4):1146–1172, 2015.

[137] S. Müller and Y. Stiriba. Fully adaptive multiscale schemes for conservation laws employing locally varying time stepping. *Journal of Scientific Computing*, 30(3):493–531, 2007.

[138] F. Nataf and P.-H. Tournier. A GenEO Domain Decomposition method for Saddle Point problems. *Comptes Rendus. Mécanique*, 351(S1):1–18, 2023.

[139] M. Neuss-Radu. A result on the decay of the boundary layers in the homogenization theory. *Asymptotic Anal.*, 23(3-4):313–328, 2000.

[140] M. Ohlberger and B. Verfürth. Localized Orthogonal Decomposition for two-scale Helmholtz-type problems. *AIMS Mathematics*, 2(3):458–478, 2017.

[141] F. Ouaki. *Etude de schémas multi-échelles pour la simulation de réservoir*. PhD Thesis, Palaiseau, Ecole polytechnique, 2013.

[142] C.C. Paige and M.A. Saunders. Solution of Sparse Indefinite Systems of Linear Equations. *SIAM Journal on Numerical Analysis*, 12(4):617–629, 1975.

[143] D. Paredes, F. Valentin, and H.M. Versieux. On the robustness of Multiscale Hybrid-Mixed Methods. *Mathematics of Computation*, 86(304):525–548, 2017.

[144] B. Parisse and D. De Graeve. Giac/Xcas, logiciel libre de calcul formel.

[145] PETSc. Guide to the Stokes Equations using Finite Elements — PETSc v3.19.1-422-gcb867a2c0e7 documentation.

[146] A. Plaza and M.C. Rivara. Average adjacencies for tetrahedral skeleton-regular partitions. *Journal of Computational and Applied Mathematics*, 177(1):141–158, 2005.

[147] M. Puscas, D. Panunzio, and R. Lagrange. FSI - vibrations of immersed cylinders. Simulations with the engineering open-source code TrioCFD. Test cases and experimental comparisons. *Comptes Rendus. Mécanique*, 350, 2022.

[148] A. Quarteroni and A. Valli. *Numerical approximation of partial differential equations: with 17 tables.* Number 23 in Springer series in computational mathematics. Springer, Berlin Heidelberg, 2008.

[149] M. Quintard and S. Whitaker. Two-phase flow in heterogeneous porous media: the method of large-scale averaging. *Transport in Porous Media*, 3(4):357–413, 1988.

[150] W. Ren and W. E. Heterogeneous Multiscale Method for the modeling of complex fluids and microfluidics. *Journal of Computational Physics*, 204(1):1–26, 2005.

[151] Y. Saad. A Flexible Inner-Outer Preconditioned GMRES algorithm. *SIAM Journal on Scientific Computing*, 14(2):461–469, 1993.

[152] Y. Saad and M.H. Schultz. GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986.

[153] E. Saikali, P. Ledac, A. Bruneton, A. Khizar, C. Bourcier, G. Bernard-Michel, E. Adam, and D. Houssin-Agbomson. Numerical modeling of a moderate hydrogen leakage in a typical two-vented fuel cell configuration. In *International Conference on Hydrogen Safety*, 2021.

[154] E. Sanchez-Palencia. *Non-homogeneous media and vibration theory*. Number 127 in Lecture notes in physics. Springer, Berlin, 1980.

[155] S. Sauter and C. Torres. On the inf-sup stability of Crouzeix–Raviart Stokes elements in 3D. *Mathematics of Computation*, 92(341):1033–1059, 2023.

[156] Z. Shen. Sharp convergence rates for Darcy's law. *Communications in Partial Differential Equations*, 47(6):1098–1123, 2022.

[157] N. Spillane, V. Dolean, P. Hauret, F. Nataf, C. Pechstein, and R. Scheichl. Abstract robust coarse spaces for systems of PDEs via generalized eigenproblems in the overlaps. *Numerische Mathematik*, 126(4):741–770, 2014.

[158] R. Stenberg. On some three-dimensional finite elements for incompressible media. *Computer Methods in Applied Mechanics and Engineering*, 63(3):261–269, August 1987.

[159] A.H. Stroud and D. Secrest. *Gaussian Quadrature Formulas: By A.H. Stroud and Don Secrest*. Prentice-Hall series in automatic computation. Prentice-Hall, 1966.

[160] L. Tartar. Convergence of the homogenization process. *Appendix of [154]*, 1980.

[161] L. Tartar. *The General Theory of Homogenization: A Personalized Introduction*, volume 7 of *Lecture Notes of the Unione Matematica Italiana*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[162] R. Temam. *Navier-Stokes equations: theory and numerical analysis*. Elsevier, 2016.

[163] F. J. Vermolen and A. Segal. On an integration rule for products of barycentric coordinates over simplexes in Rn. *Journal of Computational and Applied Mathematics*, 330:289–294, 2018.

[164] V.V. Žikov, S.M. Kozlov, and O.A. Olejnik. *Homogenization of differential operators and integral functionals*. Springer, Berlin Heidelberg, 1994.

[165] Li Wang, Qiang Xu, and Zhifei Zhang. Corrector estimates and homogenization error of unsteady flow ruled by Darcy's law. *arXiv 2202.04826*, 2022.

[166] E. Weinan and B. Engquist. The Heterogenous Multiscale Methods. *Communications in Mathematical Sciences*, 1(1):87–132, 2003.

[167] D.C. Wilcox. Formulation of the $k - \omega$ Turbulence Model Revisited. *AIAA Journal*, 46(11):2823–2838, 2008.

[168] C. Wolfsteiner, S.H. Lee, and H.A. Tchelepi. Well modeling in the Multiscale Finite Volume Method for subsurface flow simulation. *Multiscale Modeling and Simulation*, 5(3):900–917, 2006.

[169] X.H. Wu, Y. Efendiev, and T.Y. Hou. Analysis of upscaling absolute permeability. *Discrete & Continuous Dynamical Systems - B*, 2(2):185–204, 2002.

[170] Y. Xu. On Gauss–Lobatto Integration on the Triangle. *SIAM Journal on Numerical Analysis*, 49(2):541–548, 2011.

ECOLE
DOCTORALE
DE MATHEMATIQUES
HADAMARD

Titre : Méthode des éléments finis multi-échelles pour les écoulements incompressibles dans des milieux hétérogènes : Implémentation et Analyse de convergence.

Mots clés : Méthode des Éléments Finis Multi-échelles, Élément de Crouzeix–Raviart, Équations de Navier–Stokes, Milieux hétérogènes, Analyse Numérique, Calcul Haute Performance.

Résumé : Cette thèse porte sur l'application d'une méthode d'éléments finis multi-échelles (MsFEM) pour résoudre les écoulements incompressibles dans des milieux multi-échelles. En effet, la simulation de l'écoulement dans un milieu multi-échelle comportant de nombreux obstacles, tel que le cœur d'un réacteur nucléaire, est un défi de taille. Afin de capturer avec précision les échelles les plus fines de l'écoulement, il est nécessaire d'utiliser un maillage très fin. Cependant, cela conduit souvent à des simulations difficiles à réaliser en raison du manque de ressources informatiques. Pour remédier à cette limitation, cette thèse développe une MsFEM non-conforme enrichie pour résoudre les écoulements visqueux incompressibles dans des milieux hétérogènes, basée sur la méthode classique des éléments finis non conformes de Crouzeix-Raviart avec des fonctions de poids d'ordre élevé. La MsFEM utilise un maillage grossier sur lequel de nouvelles fonctions de base sont définies. Ces fonctions ne sont pas les fonctions de base polynomiales classiques des éléments finis, mais résolvent les équations de la mécanique des fluides sur les éléments du maillage grossier. Ces fonctions sont elles-mêmes approximées numériquement sur un maillage fin, en tenant compte de tous les détails géométriques, ce qui confère à cette méthode son aspect multi-échelle. Une étude théorique de la MsFEM proposée est menée aux niveaux continu et discret. Tout d'abord, le caractère bien posé des problèmes locaux discrets impliqués dans la MsFEM a été démontré à l'aide de nouvelles familles d'éléments finis. Pour ce faire, une nouvelle famille d'éléments finis non conformes en trois dimensions sur les tétraèdres a été développée. En outre, la première estimation d'erreur pour l'approximation du problème de Stokes dans des milieux perforés périodiques à l'aide de cette MSFEM est établie, démontrant sa convergence. Cette estimation est basée sur la théorie de l'homogénéisation du problème de Stokes dans les domaines périodiques et sur la théorie usuelle des éléments finis. Au niveau numérique, la MsFEM pour résoudre les problèmes de Stokes et d'Oseen a été implémenté dans un cadre massivement parallèle dans FreeFEM. En outre, une méthodologie pour résoudre le problème de Navier-Stokes est fournie.

Title : Multi-scale Finite Element Method for incompressible flows in heterogeneous media : Implementation and Convergence analysis.

Keywords : Multi-scale Finite Element Method, Crouzeix–Raviart element, Navier–Stokes equations, Heterogeneous media, Numerical Analysis, High Performance Computing.

Abstract : This thesis is concerned with the application of a Multi-scale Finite Element Method (MsFEM) to solve incompressible flow in multi-scale media. Indeed, simulating the flow in a multi-scale media with numerous obstacles, such as nuclear reactor cores, is a highly challenging endeavour. In order to accurately capture the finest scales of the flow, it is necessary to use a very fine mesh. However, this often leads to intractable simulations due to the lack of computational resources. To address this limitation, this thesis develops an enriched non-conforming MsFEM to solve viscous incompressible flows in heterogeneous media, based on the classical non-conforming Crouzeix-Raviart finite element method with high-order weighting functions. The MsFEM employs a coarse mesh on which new basis functions are defined. These functions are not the classical polynomial basis functions of finite elements, but rather solve fluid mechanics equations on the elements of the coarse mesh. These functions are themselves numerically approximated on a fine mesh, taking into account all the geometric details, which gives the multi-scale aspect of this method. A theoretical investigation of the proposed MsFEM is conducted at both the continuous and discrete levels. Firstly, the well-posedness of the discrete local problems involved in the MsFEM was demonstrated using new families of finite elements. To achieve this, a novel non-conforming finite element family in three dimensions on tetrahedra was developed. Furthermore, the first error estimate for the approximation of the Stokes problem in periodic perforated media using this MSFEM is derived, demonstrating its convergence. This is based on homogenization theory of the Stokes problem in periodic domains and on usual finite element theory. At the numerical level, the MsFEM to solve the Stokes and the Oseen problems in two and three dimensions is implemented in a massively parallel framework in FreeFEM. Furthermore, a methodology to solve the Navier–Stokes problem is provided.

Institut Polytechnique de Paris
91120 Palaiseau, France