HAL
open science

Calcul d'écoulements extérieurs incompressibles

Delphine Jennequin

To cite this version:

Delphine Jennequin. Calcul d'écoulements extérieurs incompressibles. Mathématiques [math]. Université du Littoral Côte d'Opale, 2005. Français. NNT : . tel-00011893

HAL Id: tel-00011893

https://theses.hal.science/tel-00011893

Submitted on 9 Mar 2006

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

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

Université du Littoral Cote d'Opale

Laboratoire de Mathématiques Pures et Appliquées

Numéro attribué par la bibliothèque : | | | | | | | | | | |

THÈSE

pour obtenir le grade de

Docteur de l'Université du Littoral Côte d'Opale

Spécialité : Mathématiques Appliquées

présentée et soutenue publiquement [1]

par

Delphine Jennequin

le 9 décembre 2005

Titre

Calculs d'écoulements extérieurs incompressibles

Composition du jury

Rapporteurs :	**François Alouges**	
	Jean-Pierre Croisille	
Examinateurs :	**Caterina Calgaro**	(co-directrice de thèse)
	Bruno Després	
	Paul Deuring	(directeur de thèse)
	Hassane Sadok	

[1] Le projet de soutenance a été approuvé par la commission des thèses de mathématiques des universités d'Artois, du Littoral Côte d'Opale, de Lille1 et Valenciennes-Hainaut-Cambresis.

Remerciements

Je tiens tout d'abord à remercier Caterina Calgaro et Paul Deuring qui m'ont encadrée pendant ces trois années de thèse. Ils m'ont guidée dans ce travail tout en respectant mes choix de recherche. Je les remercie pour leur disponibilité et leurs lectures attentives.

Je remercie ensuite François Alouges et Jean-Pierre Croisille qui ont accepté de rapporter sur ma thèse. Je leur suis reconnaissante de l'intérêt qu'ils ont porté à mon travail. J'adresse mes remerciements à Bruno Desprès et Hassane Sadok qui ont accepté de faire partie de mon jury.

Je n'oublie pas Thierry Dumont qui m'a "hébergée" sur les machines de calculs de Lyon. Sans sa gentillesse et sa disponibilité, cette thèse n'aurait pas pu voir le jour. Merci à Jacques Laminie pour ses conseils toujours enrichissants et également à Sylvain Faure, pour son accueil à Orsay.

Ce fut un grand plaisir pour moi de travailler avec Jean-Paul Chehab et Juanjo Nieto. Je remercie également chaleureusement toute l'équipe EDP de l'université de Lille, et plus particulièrement Pauline Lafitte, Thierry Goudon et Jean-François Coulombel. Merci à Thierry de m'avoir invitée à faire le CEMRACS 2003. Je suis très reconnaissante également envers Albert Cohen, Eric Sonnendrucker et Jean-Rodolphe Roche, au contact desquels j'ai beaucoup appris. Merci à Alexandre, Fredéric et Mourad qui rendent les déplacements toujours sympathiques, merci à Marcos pour nos discussions. Une pensée enfin pour les thésards du nord (Karim, Anas, Abdel, Siham, Chunjin, Youcef, Greg ...) et un grand merci pour les bons moments passés avec eux.

Et pour finir, merci à ceux qui m'ont supportée au quotidien pendant cette thèse, merci à Gilles, Sandrine, Mr Bean, merci à Virginie, Sébastien, Franck, merci aux amis de l'AVAN, merci à ma famille.

Table des matières

1 Introduction **9**

I Calculs d'écoulements extérieurs incompressibles **19**

2 Tour d'horizon **21**

2.1 Traitement de la non-linéarité 21

2.2 Le système d'Oseen 24

2.2.1 La méthode multigrille avec le lisseur de type Vanka. 25

2.2.2 Les méthodes de type Uzawa 25

2.2.3 Préconditionnement du problème complet 26

2.3 Préconditionner le problème de convection-diffusion 27

2.3.1 Méthodes multigrilles 27

2.3.2 Méthodes de décomposition de domaine 28

2.4 Préconditionner le complément de Schur 33

2.4.1 Pour des espaces d'éléments finis stables 33

2.4.2 Pour des espaces d'éléments finis non stables 35

2.5 Apport de la thèse 36

3 Problèmes de point-selle généralisés **39**

3.1 Cadre général 40

3.2 Préconditionnement 42

3.2.1 Définitions 42

3.2.2 Généralités et résultat principal 43

3.3 Preuve du théorème. 43

3.3.1 Résultats préliminaires 44
3.3.2 Estimation de la plus petite partie réelle des valeurs propres 46
3.3.3 Estimation de la plus grande partie réelle des valeurs propres 47
3.3.4 Estimation de la partie imaginaire des valeurs propres. 47

4 Application à Navier-Stokes 3D dans un cube 49

4.1 Un préconditionneur pour les équations de Navier-Stokes 49
4.1.1 Discrétisation des équations de Navier-Stokes 49
4.1.2 Condition inf-sup généralisée 53
4.1.3 Influence du préconditionneur 55
4.2 Résultats numériques 57
4.2.1 Présentation des tests 57
4.2.2 Calcul des valeurs propres 58
4.2.3 Performance du préconditionneur. 62
4.2.4 Itérations non-linéaires 64
4.2.5 Solutions de la cavité 3D 67
4.3 Conclusion 68

5 Application à Navier-Stokes 3D dans un domaine extérieur 73

5.1 Modèle et discrétisation 73
5.2 Propriété des formes bilinéaires 77
5.3 Préconditionnement pour le système d'Oseen 81
5.4 Choix des termes de stabilisation 83
5.5 Résultats numériques 84
5.5.1 Présentation des tests 84
5.5.2 Valeurs propres du problème préconditionné 85
5.5.3 Performances 89
5.5.4 Itérations non-linéaires 92
5.5.5 Représentation de la solution 95
5.6 Conclusion 95

6 Méthode de décomposition de domaine pour le problème extérieur 105

6.1 Partition du domaine et notations 105

6.2 Décomposition de domaine 106
6.2.1 Formulation variationnelle 106
6.2.2 Conditions aux limites à l'interface 108
6.2.3 Méthode FETI 110
6.3 Choix de l'algorithme FETI 112
6.3.1 Préconditionneur de Toselli pour le FETI-1 112
6.3.2 Préconditionneur simplifié pour le FETI-1 113
6.3.3 Préconditionnement pour le FETI-2 114
6.3.4 Performances connues pour la méthode FETI 115
6.4 Résultats numériques dans un domaine borné 116
6.4.1 Description de l'algorithme 116
6.4.2 Résultats dans un domaine fixe 117
6.4.3 Bilan des résultats numériques 121
6.5 Résultats numériques pour le problème extérieur 123
6.6 Conclusion 125

Perspectives **127**

Bibliographie **129**

II Une méthode adaptative pour le système de Vlasov-Poisson **141**

7 CEMRACS 2003 : An Adaptive Particle-In-Cell method using multi-resolution analysis **143**
7.1 Introduction 144
7.2 Density estimation by wavelet thresholding 145
7.3 Numerical schemes 149
7.3.1 The PIC method 149
7.3.2 The adaptive scheme (PICONU) 150
7.4 Numerical results 150
7.4.1 Comparison between NGP and W0. 152
7.4.2 Comparison between CIC and W1. 153
7.5 Concluding remarks and perspectives 153

Appendices **161**

A Définitions et rappels **161**

A.1 Résidu au sens des éléments finis 161

A.2 M-matrice 162

B Bibliothèques **163**

B.1 Fonctions LAPACK utilisées 163

B.1.1 Fonction `dhseqr_` 163

B.1.2 Fonction `dgeev_` 163

B.1.3 Fonction `dgeqrf_` 163

B.1.4 Fonction `dorgqr_` 164

B.2 Description de SparseLib++ 164

B.3 Description des classes du code pour la décomposition de domaine 165

C Mise En oeuvre **167**

C.1 Raffinement du maillage 167

C.2 Assemblage du problème de Navier-Stokes 167

C.3 Réordonnancement 169

Table des figures

1.1 Schéma de l'écoulement 9
1.2 Maillage exponentiel 11

2.1 Décomposition de domaines de Nataf et Rogier 29

4.1 Cas test a), résidu à chaque itération du FGMRES(30), $\mathcal{R}e = 100, h = 1/32$. . . 63
4.2 Résidu à l'intérieur des itérations internes du FGMRES pour chaque itération non-linéaire , Re=100 et h=1/32. 66
4.3 Cavité entraînée 3D sur la grille de taille $32 \times 32 \times 32$. 66
4.4 Profil de la première composante de la vitesse, $\mathcal{R}e = 100$. 68
4.5 Cavité entraînée 3D, Re=1, grille de taille $64 \times 64 \times 64$. 69
4.6 Cavité entraînée 3D, Re=10, grille de taille $64 \times 64 \times 64$. 70
4.7 Cavité entraînée 3D, Re=100, grille de taille $64 \times 64 \times 64$. 71
4.8 Cavité entraînée 3D, Re=200, grille de taille $64 \times 64 \times 64$. 72

5.1 Projection en 2D de notre maillage exponentiel 75
5.2 Spectre de $\boldsymbol{Q}^{-1}\boldsymbol{S}$ pour $R = 8$. 88
5.3 Spectre de $\boldsymbol{Q}^{-1}\boldsymbol{S}$ pour $\mathcal{R}e = 1$. 88
5.4 Spectre de $\boldsymbol{Q}^{-1}\boldsymbol{S}$ pour $\mathcal{R}e = 10$. 89
5.5 Nombre moyen d'itérations du FGMRES en fonction de R. 94
5.6 Champ de vitesse, $\mathcal{R}e = 1$ 96
5.7 Pression sur la sphère 97
5.8 Vorticité, $\mathcal{R}e = 1$ 98
5.9 Champ de vitesse, $\mathcal{R}e = 5$ 99
5.10 Vorticité, $\mathcal{R}e = 5$ 100

5.11 Champ de vitesse, $\mathcal{R}e = 10$ 101
5.12 Vorticité, $\mathcal{R}e = 10$ 102
5.13 Représentation des lignes de courant en 3D pour $\mathcal{R}e = 20$. 102
5.14 Isovaleurs de la vitesse dans la coupe (xOz) tracées dans la boule de rayon 16 . 103

6.1 Maillage dans le domaine fixe 117
6.2 résidu pour une convection fixe 119
6.3 Résidu dans le cas d'une convection tournante 120
6.4 Convection radiale 121

7.1 Landau damping with $\alpha = 0.1$. Charge density computed with 26000 particles, in solid line : the PICONU method (W0), in dashed line : the classical PIC (NGP). 151
7.2 Landau damping with $\alpha = 0.1$. Electrostatic field computed with 26000 particles, in solid line : the PICONU method(W0), in dashed line :the classical PIC (NGP). 155
7.3 Landau damping with $\alpha = 0.1, k = 0.5$. Electrostatic energy 156
7.4 Superposition of electrostatic energy with 26000 particles. In solid line : damping obtain with PICONU method (W0), In dashed line : classical PIC method (NGP). 156
7.5 linear Landau damping with classical PIC (CIC). 10000 particles, 128 cells. . . 157
7.6 Linear landau damping, PICONU (W1), 10000 particles 157

C.1 Raffinement d'un hexaèdre en 24 tétraèdres 168
C.2 Réordonnancement 169

Liste des tableaux

4.1 Valeurs propres pour la matrice d'Oden. Pas de stabilisation en vitesse. 60

4.2 Valeurs propres pour la matrice d'Oden. Stabilisation en vitesse de paramètres $A = P = 1$. 61

4.3 Valeurs propres pour la dernière itération non-linéaire dans la cavité entraînée. Pas de stabilisation en vitesse. 61

4.4 Nombre d'itération du GMRES/FGMRES 63

4.5 Nombre d'itérations du FGMRES préconditionné à droite, cas test a). 64

4.6 Nombre d'itérations non-linéaires (en roman) et nombre moyen d'itérations internes du FGMRES (en italique). 67

5.1 Valeurs propres du complément de Schur. $w_{h,R} = 0$. Pas de préconditionnement. En roman, la borne supérieure réelle. En italique, la borne inférieure réelle. 86

5.2 Valeurs propres de $\boldsymbol{Q}^{-1}\boldsymbol{S}$, $w_{h,R} = 0$. Pas de stabilisation en vitesse. En roman, la borne supérieure réelle. En italique, la borne inférieure réelle. 87

5.3 Valeurs propres de $\boldsymbol{Q}^{-1}\boldsymbol{S}$, convection polynomiale définie en a), stabilisation en vitesse de paramètres $A = P = 1$. En roman, la borne supérieure réelle. En italique, la borne inférieure réelle. 90

5.4 Nombre d'itérations du FGMRES pour le problème linéarisé 91

5.5 Résultats sur les itérations non-linéaires, cas test polynomial, $\mathcal{R}e = 1$ 93

5.6 Itérations non-linéaires dans le cas test physique 94

6.1 Taille de la maille 117

6.2 Nombre d'itération du GMRES, Convection fixe, pas de stabilisation en vitesse. 118

6.3 Nombre d'itération du GMRES, convection tournante, pas de stabilisation en vitesse. 119
6.4 Nombre d'itération du GMRES, convection radiale, pas de stabilisation en vitesse. 120
6.5 Avec un terme d'ordre 0 122
6.6 Ordre de grandeur 123
6.7 Méthode FETI pour le problème extérieur 124
6.8 Nombre d'itérations du FETI-1/FGMRES lors de la dernière itération non-linéaire 124

Chapitre 1

Introduction

Cette thèse traite de la résolution numérique des équations de Navier-Stokes incompressibles stationnaires dans un domaine extérieur. Ce mémoire est organisé de la façon suivante :

Première partie : Calculs d'écoulements extérieurs incompressibles stationnaires

Nous considérons une sphère se déplaçant sans rotation dans un fluide incompressible et sommes intéressés par l'écoulement de ce fluide en régime stationnaire. Comme nous ne

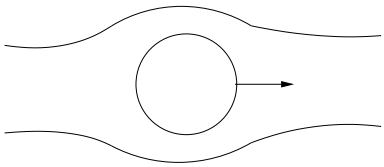


Fig. 1.1 – Schéma de l'écoulement

considérons qu'un seul objet, il est naturel de placer le centre du repère au centre de la sphère et de résoudre les équations régissant l'écoulement du fluide dans le domaine défini par le complémentaire de l'objet. Les équations modélisant cet écoulement sont les équations de Navier-Stokes avec une vitesse non nulle à l'infinie. Elle peuvent s'écrire sous forme du

système adimensionné suivant

$$\begin{cases} -\dfrac{1}{\mathcal{R}e}\Delta\tilde{u} \;+\; (\tilde{u}\cdot\nabla)\tilde{u} \;+\; \nabla\pi \;=\; \Phi & \text{dans } \Omega^c \\ \operatorname{div}\tilde{u} \;=\; 0 & \text{dans } \Omega^c, \\ \tilde{u}\,|\,\partial\Omega = 0, \quad \tilde{u}(x) \to e_1 \;\; (|x| \to \infty) \end{cases}$$

où le couple $(\tilde{u}, \pi)$ représente la solution en vitesse-pression, $\mathcal{R}e$ est le nombre de Reynolds et Ω est la boule centrée en l'origine et de rayon 1. Par un changement d'échelle sur la pression, i.e. en posant $\tilde{p} = \mathcal{R}e\pi$, on obtient le système suivant

$$\begin{cases} -\Delta\tilde{u} \;+\; \mathcal{R}e\,(\tilde{u}\cdot\nabla)\tilde{u} \;+\; \nabla\tilde{p} \;=\; \tilde{\Phi} & \text{dans } \Omega^c \\ \operatorname{div}\tilde{u} \;=\; 0 & \text{dans } \Omega^c, \\ \tilde{u}\,|\,\partial\Omega = 0, \quad \tilde{u}(\tilde{x}) \to e_1 \;\; (|\tilde{x}| \to \infty), \end{cases}$$

qui a été étudiée par Deuring dans [36]. Les deux principales difficultés de ce système sont le fait que le domaine ne soit pas borné et que la vitesse du fluide soit non nulle à l'infini. Du fait de la géométrie du domaine, nous utiliserons une méthode d'éléments finis associée à un maillage adapté que nous décrirons ultérieurement. Nous devons donc définir un domaine de travail en tronquant le domaine extérieur. Il est alors nécessaire de prescrire des conditions aux limites, notées ABC pour "Approximate Boundary Condition." Il existe beaucoup d'articles dans la littérature sur la détermination de conditions aux limites acceptables (voir la revue [102]) : de telles conditions doivent assurer au minimum que le problème soit bien posé et que l'erreur de troncature décroisse quand le diamètre du domaine de troncature augmente. La discrétisation de problèmes extérieurs tronqués nécessite un maillage particulier. Goldstein [51, 50] a introduit une variété de maillage, appelé **maillage exponentiel**, pour résoudre l'équation de Poisson en domaine extérieur. Ce type de maillage consiste à mailler par anneaux autour d'une sphère de façon à ce que le diamètre de la triangulation à l'intérieur d'un anneau croisse linéairement avec la distance au centre de l'obstacle (voir figure 1.2). Si on appelle R le rayon de la sphère de troncature, le nombre de degrés de liberté ne croît que logarithmiquement avec R. Quand on choisit de mailler homothétiquement entre chaque anneau et que le problème discrétisé est linéaire, ce type de maillage permet de n'assembler les matrices que sur le premier anneau, ce qui réduit énormément le stockage [6]. Le problème modèle considéré a été étudié théoriquement dans [36] : tout d'abord, par un changement de variable dont le nouveau couple d'inconnues se nomme (u, p), on modifie les équations de Navier-Stokes ci-dessus pour avoir

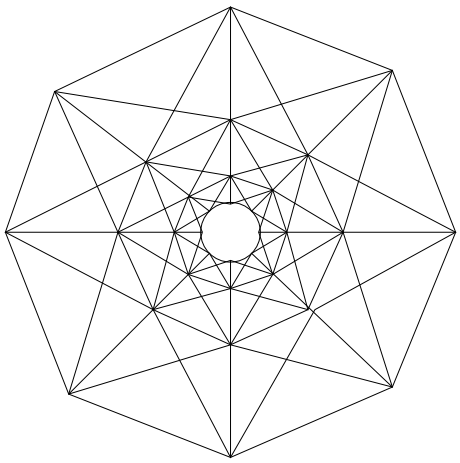


FIG. 1.2 – Maillage exponentiel

une vitesse nulle en l'infini [22][1]. Pour cela, il suffit d'ajouter un terme convectif aux équations de Navier-Stokes, c'est à dire que nous obtiendrons le système

$$\begin{cases} -\Delta u \;+\; \mathcal{R}e \;\cdot D_1 u \;+\; \mathcal{R}e \;\cdot (u\cdot\nabla)u \;+\; \nabla p \;=\; F & \text{dans} \quad \overline{\Omega}^c, \\ \operatorname{div} u \;=\; 0 & \text{dans} \quad \overline{\Omega}^c, \\ u\,|\,\partial\Omega \;=\; e_1, \quad u(x)\to 0 \;\; (|x|\to\infty), \end{cases} \tag{1.1}$$

où D_1 est l'opérateur de dérivation par rapport à la première variable. La décroissance de la solution loin de l'obstacle a été étudiée dans [36] et les auteurs en ont déduit des conditions aux limites acceptables. Nous utiliserons celles-ci et nous les associerons au maillage exponentiel de Goldstein [51, 50].

Nous devons choisir des éléments finis adaptés à la fois aux équations de Navier-Stokes et aux domaines extérieurs tronqués. Deuring [33] a montré que les P1+ bulles -P1 vérifient une condition inf-sup dont la constante est indépendante du rayon de troncature R. De plus, en trois dimension, on doit vraiment limiter le nombre de degrés de liberté pour des raisons de stockage et de temps de calcul. Rebollo [90] a proposé une variante de l'élimination des fonctions bulles qui est peu coûteuse et nous montrerons que cette technique, donnant du P1-P1 + stabilisation, conserve les propriétés des fonctions bulles dans le sens que la condition inf-sup discrète généralisée est vérifiée avec un coefficient indépendant de la taille de la maille et du rayon de troncature R. Supposons provisoirement que notre domaine de travail s'appelle

[1]Cette référence est un proceeding du GAMM 2003 écrit avec C. Calgaro et P. Deuring et publié dans PAMM.

Ω et appelons $\mathcal{T}_h$ une triangulation de Ω. On appelle V et Q les espaces de Hilbert dans lesquels vivent la vitesse et la pression, dont le couple est solution du problème de Navier-Stokes. Nous préciserons plus tard, en fonction des cas considérés, quels peuvent être ces espaces. On note $\| \|_V$ et $\| \|_Q$ leurs normes associées. On note ensuite V_h et Q_h les espaces d'éléments finis associés à la vitesse et la pression respectivement où l'indice h représente la taille de la maille, i.e.

$$V_h = \{u \in \mathcal{C}^0(\overline{\Omega})^3; u_{|K} \in \mathrm{P}_1(K)^3 \ \forall K \in \mathcal{T}_h\},$$

et

$$Q_h = \{p \in \mathcal{C}^0(\overline{\Omega}); p_{|K} \in \mathrm{P}_1(K) \ \forall K \in \mathcal{T}_h.\}$$

La formulation variationnelle de notre problème s'écrit alors, aux conditions aux limites près, sous la forme suivante : trouver $u_h \in V_h$ et $p_h \in Q_h$ tels que

$$\begin{cases} d(u_h, u_h, v_h) + b(v_h, p_h) = (f, v_h), & \forall v_h \in V_h \\ b(u_h, q_h) - c(p_h, q_h) = (g, q_h), & \forall q_h \in Q_h, \end{cases} \tag{1.2}$$

où les fonctions b, c, d sont continues sur $V \times Q$, $Q \times Q$ et $V \times V \times V$ respectivement, b et c sont des formes bilinéaires associées respectivement à l'opérateur de divergence et à la stabilisation en pression et d est la forme trilinéaire associée aux termes relatifs à l'opérateur D_1 et aux termes non linéaires.

Chapitre 2 : Tour d'Horizon

Nous présenterons dans le paragraphe 2.1 différentes méthodes pour résoudre ce type de problèmes non linéaires. Nous opterons finalement pour une méthode de quasi-Newton. Cela implique qu'à chaque étape, nous devons résoudre un **problème de point-selle généralisé** de la forme

$$\begin{cases} a(u_h, v_h) + b(v_h, p_h) = (f, v_h), & \forall v_h \in V_h, \\ b(u_h, q_h) - c(p_h, q_h) = (g, q_h), & \forall q_h \in Q_h, \end{cases} \tag{1.3}$$

où a est une linéarisation de d et dont la notation algébrique est

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix}. \tag{1.4}$$

La résolution des problèmes de point-selle en général a intéressé beaucoup d'auteurs et nous ferons un rapide tour d'horizon dans ce chapitre des méthodes proposées. Nous justifierons

le choix de notre méthode générale, qui consiste à appliquer un FGMRES(Flexible Generalized Minimum RESidual) sur la matrice totale préconditionnée par un préconditionneur triangulaire inférieur de la forme

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{Q} \end{pmatrix}, \tag{1.5}$$

où $\boldsymbol{Q}$ est la matrice de masse. Dans [26], Zhi Hao Cao démontre le résultat suivant :

Théorème 1.1 *Le résidu $r^{(k)}$ à la $k^{ième}$ itération du* GMRES *appliqué à une matrice $\boldsymbol{M}$ de taille $n \times n$ vérifie l'inégalité*

$$\|r^{(k)}\| \leq \left(1 - \tau(\boldsymbol{M})\tau(\boldsymbol{M}^{-1})\right)^{k/2} \|r^{(0)}\|,$$

où

$$\tau(\boldsymbol{M}) = \min_{y \in \mathbb{R}^n} \frac{(\boldsymbol{M}y, y)}{(y, y)},$$

et où $(\ ,\ )$ est le produit scalaire usuel de $\mathbb{R}^n$, et $\|\ \|$ la norme euclidienne associée. Pour le GMRES*(m) (*GMRES *avec redémarrage toutes les m itérations), cette inégalité devient*

$$\|r^{(k)}\| \leq \left(1 - \tau(\boldsymbol{M})\tau(\boldsymbol{M}^{-1})\right)^{km/2} \|r^{(0)}\|.$$

Le taux de convergence de notre méthode dépendra donc du spectre de la matrice

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{Q} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix},$$

c'est à dire que nous devrons étudier le problème généralisé aux valeurs propres

$$(\boldsymbol{E} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T)P = \lambda \boldsymbol{Q} P. \tag{1.6}$$

Nous allons donc traiter les points suivants : Nous validerons théoriquement et numériquement notre choix du préconditionneur (1.5) dans les chapitres 3 et 4. Ensuite nous adapterons ces résultats au problème de Navier-Stokes en domaine extérieur dans le chapitre 5 et nous étudierons numériquement dans le chapitre 6 l'apport d'une méthode de décomposition de domaine pour les problèmes extérieurs à l'intérieur de notre préconditionneur.

Chapitre 3 : Problèmes de point-selle généralisés

Dans ce chapitre, nous étudions le problème de point-selle (1.3). Nous supposons que a est uniformément continue et coercice dans $V_h \times V_h$ et nous notons $\mathcal{A}$ et α ses constantes de continuité et coercivité. Dans ces conditions, nous démontrerons le théorème d'existence des solutions discrètes suivant [23] [2] :

Théorème 1.2 *Supposons que les formes bilinéaires b et c satisfont la condition inf-sup généralisée*

$$\sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V} + c(q_h, q_h)^{1/2} \geq \gamma \|q_h\|_Q, \qquad \textit{pour tout } q_h \in Q_h, \tag{1.7}$$

où γ est une constante positive. Alors, le système (1.3) a une solution unique. De plus, si on appelle (u_h, p_h) sa solution, on obtient les inégalités suivantes

$$\|u_h\|_V \leq C_1\big(\|f\|_{V'} + \|g\|_{Q'}\big) \qquad \textit{et} \qquad \|p_h\|_Q \leq C_2\big(\|f\|_{V'} + \|g\|_{Q'}\big),$$

où $C_1 = C_1(\gamma, \mathcal{A}, \alpha)$ et $C_2 = C_2(\gamma, \mathcal{A}, \alpha)$ sont des constantes ne dépendant que de $\gamma, \mathcal{A}$ ou α.

Nous supposons maintenant que b et c sont uniformément continus dans $V_h \times Q_h$ et dans $Q_h \times Q_h$ respectivement et de constantes de continuité respectives $\mathcal{B}$ et $\boldsymbol{C}$. Nous donnons alors une estimation des bornes du spectre du problème préconditionné (1.6) via le théorème suivant :

Théorème 1.3 *Supposons que (1.7) soit réalisé. Alors les solutions du problème généralisé aux valeurs propres (1.6) sont incluses dans le domaine rectangulaire*

$$\left[\frac{\gamma^2}{2}\min(1, \alpha(\mathcal{A}_+^2 + \mathcal{A}_-^2)^{-1}), (\frac{\mathcal{B}^2}{\alpha} + \mathcal{C})\right] \times \left[-\frac{\mathcal{B}^2}{2\alpha}, \frac{\mathcal{B}^2}{2\alpha}\right],$$

où γ est défini dans (1.7), où les constantes $\mathcal{A}_+$ et $\mathcal{A}_-$ sont les bornes de continuité uniforme dans $V_h \times V_h$ des parties symétrique et anti-symétrique de a.

On notera que ces bornes sont indépendantes de la taille des espaces de discrétisation.

[2] Cette référence correspond à une publication écrite avec C. Calgaro et P. Deuring à paraître dans Numer. Meth. for PDE.

Chapitre 4 : Application à Navier-Stokes 3D dans un cube

Nous appliquons les résultats précédents aux équations de Navier-Stokes définies dans un domaine fixe Ω discrétisées par du P1-P1 avec la stabilisation de Rebollo [90]. Nous vérifions que cette discrétisation satisfait la condition inf-sup généralisée (1.7) pour $V = \left(H_0^1(\Omega)\right)^3$ et pour $Q = L^2(\Omega)$, puis nous exprimons les bornes de notre spectre en fonction du nombre de Reynolds : les valeurs propres du problème préconditionné sont contenues dans un rectangle de la forme

$$\left[\min\left(\frac{\gamma^2}{2}, C_1\mathcal{R}e^{-4}h^{-2}\right), C_2\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right], \tag{1.8}$$

où C_1, C_2 sont indépendantes de h et de $\mathcal{R}e$ et où γ est la constante introduite dans (1.7). De plus, si on omet la stabilisation en vitesse de type SUPG (Streamline Upwind Petrov-Galerkin [20]), les valeurs propres appartiennent au domaine rectangulaire

$$\left[\min\left(\frac{\gamma^2}{2}, C_1\mathcal{R}e^{-2}\right), C_2\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]. \tag{1.9}$$

Nous vérifierons numériquement les résultats ci-dessus et observerons que les résultats numériques sont meilleurs que les estimées théoriques. Nous testerons ensuite notre préconditionneur dans le cadre de la cavité entraînée 3D et montrerons que nous avons des résultats similaires aux résultats de référence de Jiang [62] et Shu [93].

Chapitre 5 : Application à Navier-Stokes 3D dans un domaine extérieur

Nous procéderons de la même façon que précédemment, mais nous devons en plus tenir compte d'un nouveau paramètre : le rayon de troncature. Nous introduirons des espaces à poids V et Q pour lesquels la condition inf-sup généralisée (1.7) est vérifiée avec un coefficient γ indépendant de R et de $\mathcal{R}e$. Le spectre du problème préconditionné est inclus dans un rectangle dont les bords varient avec $\mathcal{R}e$ et R : il peut s'écrire sous la forme

- $\left[C_1R^{-2}, C_2\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$, si $\mathcal{R}e$ est fixé et où C_1 et C_2 dépendent de $\mathcal{R}e$.
- $\left[C_1\mathcal{R}e^{-4}, C_2\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$, si R est fixé, et où C_1 et C_2 dépendent de R.

Ce dernier domaine se réduit à

$$\left[C_1\mathcal{R}e^{-2}, C_2\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right],$$

lorsque nous ne considérons pas la stabilisation en vitesse. Les résultats numériques donneront, comme dans le cas borné, des résultats meilleurs que ceux prévus théoriquement.

Chapitre 6 : Méthode de décomposition de domaine pour le problème extérieur

Le préconditionneur tel que nous l'avons défini nécessite de savoir inverser le bloc $\boldsymbol{A}$. Cette matrice correspond à la discrétisation d'un problème de type **convection-diffusion**. Nous utilisons pour cela une méthode de décomposition de domaine dans le but de préparer une parallélisation du code : cela nécessite de résoudre sur chaque sous-domaine un problème de convection-diffusion. Pour que ces problèmes soient bien posés, nous imposons des **conditions aux limites de Robin** sur les interfaces comme proposé par Achdou et al dans [2]. La continuité entre les interfaces est imposée via des **multiplicateurs de Lagrange**. Un telle méthode porte le nom de FETI (Finite Element Tearing and Interconnecting) et a été introduite initialement en mécanique des structures [43]. Il n'existe pas de résultats théoriques pour les problèmes non symétriques et nous nous baserons sur les résultats numériques de Toselli [100] : Ces résultats ont été obtenus pour le problème d'advection diffusion en dimension deux avec une décomposition en domaines rectangulaires. Nous verrons que l'adaptation aux problèmes tridimensionnels dans le domaine extérieur est assez difficile. Heureusement, cette méthode s'avère plus robuste quand nous considérons le problème de type convection-diffusion issu des itérations de quasi-Newton pour le problème de Navier-Stokes.

Deuxième partie : Une méthode adaptative pour le système de Vlasov-Poisson [3]

Ce chapitre est un travail indépendant du sujet de thèse qui a été réalisé lors du CEMRACS 2003 (du 21 Juillet au 29 Aout 2003). Il traite des plasmas de Vlasov, c'est à dire de gaz constitués de particules chargées dont les collisions peuvent être négligées. Cette situation est modélisée par le système de Vlasov-Poisson en dimension d

$$\begin{cases} \dfrac{\partial f}{\partial t}(x,v,t) + v \cdot \nabla_x f(x,v,t) + (E(x,t) \cdot \nabla_v) f(x,v,t) = 0, \quad (x,v,t) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}, \\ E = -\nabla_x \Phi \text{ avec } \triangle \Phi = \int_{\mathbb{R}^d} f dv - 1 \end{cases} \tag{1.10}$$

[3] Ce chapitre a été publié dans Numerical Methods for Hyperbolic and Kinetic Problems. Il a été écrit en collaboration avec J.P. Chehab, A. Cohen, J.J. Nieto , Ch. Roland et J.R. Roche.

où f est la fonction de distribution définie sur $\mathbb{R}^{2d+1}$ telle que $\forall t, \quad \int_{\mathbb{R}^{2d}} f(x,v,t)\,dxdv = 1$. Plus précisément, $f(x,v,t)$ est la probabilité de trouver une particule de position x et de vitesse v au temps t. E est le champ électrostatique : il est composé du champ induit par les particules chargées et du champ de confinement, car, sans champ extérieur, le modèle est dispersif. L'idée des méthodes PIC (Particle-In-Cell) est que la fonction de distribution est constante le long des courbes caractéristiques

$$\begin{cases} \dfrac{dX(t)}{dt} = V(t), \\ \dfrac{dV(t)}{dt} = E(X(t),t). \end{cases}$$

Les codes PIC sont des méthodes peu chères numériquement et classiquement utilisées pour simuler l'évolution des particules d'un plasma dont l'équation modèle est le système de Vlasov-Poisson : en effet, leur intérêt est de ne travailler que dans l'espace physique en suivant les particules le long des courbes caractéristiques alors que le système de Vlasov-Poisson est une équation aux dérivées partielles évolutive définie dans l'espace des phases (x,v) de dimension $2d$. Une des étapes des codes PIC consiste à construire la densité de charge des particules $\varrho(x,t)$ définie par $\varrho(x,t) = \int_{\mathbb{R}^d} f dv - 1$. Habituellement, on construit un histogramme par interpolation de type P_1 en utilisant une fenêtre fixe. Cette densité doit ensuite être interpolée sur un maillage régulier sur lequel nous devons résoudre l'équation de Poisson, d'où la nécessité d'obtenir la densité la plus lisse possible. Le problème de ce type de méthode est que, quand le nombre de particules par fenêtre est trop grand, on perd de la précision et on obtient une solution trop lisse. Par contre, quand ce nombre est trop petit, on approche la somme des Dirac. Le fait d'avoir une fenêtre fixe engendre souvent du bruit numérique, d'où l'idée d'utiliser des méthodes adaptatives. Nous avons donc remplacé l'interpolation P1 par une reconstruction d'ordre deux avec seuillage à partir d'un histogramme exprimé dans la base d'ondelettes de Haar sur une grille grossière. Cela permet d'avoir une méthode peu coûteuse mais dont le résultat est lisse. Nous testerons cette méthode dans le cas monodimensionnel avec des conditions aux limites périodiques en espace. Nous montrerons numériquement que notre méthode d'ondelettes, basée sur une interpolation de type P0, donne des résultats globalement comparables à la méthode classique avec une interpolation de type $P1$ [30]. Cependant, le cas monodimensionnel est spécifique : en effet, en une dimension, le noyau de Green n'a pas de singularité, donc le champ électromagnétique est naturellement lisse. Par contre, ce n'est plus le cas en dimension supérieure et nous avons de fortes raisons d'espérer que notre méthode

devienne dans ce cas plus efficace qu'une interpolation P1.

Ce travail est une première approche des techniques de débruitage à l'intérieur des codes PIC. Ce projet porté par E. Sonnendrucker intéresse le CEA et devrait se poursuivre dans les années qui viennent.

Première partie

Calculs d'écoulements extérieurs incompressibles

Chapitre 2

Tour d'horizon

Dans ce chapitre, nous présenterons un résumé des principales méthodes utilisées pour résoudre les équations de Navier-Stokes stationnaires. Nous expliquerons au fur et à mesure de ce tour d'horizon pourquoi telle méthode semble plus adaptée que telle autre au problème que nous devons résoudre. Ensuite, nous situerons notre travail par rapport à ce qui existe dans la littérature.

2.1 Traitement de la non-linéarité

On cherche une solution discrète au problème continu $A(u)u = f, u \in V$ et nous considérons pour cela le problème non linéaire discret suivant :

$$A_h(u_h)u_h = f_h,$$

où $u_h \in V_h$ et $A_h(u_h) : V_h \to V_h$ est un opérateur discret linéaire continu. On peut trouver la solution u_h comme limite des itérées u_h^n définies par le schéma suivant :

$$\begin{aligned} &u_h^{n-1} \text{ donné pour } n = 1, 2, \cdots \\ &u_h^n = u_h^{n-1} - [\widetilde{A}_h(u_h^{n-1})]^{-1} r^{n-1}, \end{aligned} \tag{2.1}$$

où $r^{n-1} = A(u_h^{n-1})u_h^{n-1} - f_h$ et où $\widetilde{A}_h$ est inversible. Nous proposerons plusieurs choix possibles pour $\widetilde{A}_h$ par la suite.

On appelle Jacobienne de l'opérateur différentiel A en u l'opérateur linéaire $J(u)$ défini par

$$\forall h, \quad A(u+h)(u+h) - A(u)u = J(u)(h) + \|h\|\varepsilon(h)$$

avec $\lim_{h\to 0} \varepsilon(h) = 0$. Si on choisit pour $\widetilde{A}_h(u_h^{n-1})$ une discrétisation de la jacobienne $J(u^{n-1})$ de A en u_h^{n-1}, alors cette méthode est exactement la méthode de Newton. On sait que, si $\tilde{A}_h$ est inversible en u_h^n et si u_h^0 est suffisamment proche de u_h, alors la convergence de cette méthode est quadratique, c'est à dire que

$$\forall n, \quad \|u_h^n - u_h\| \sim \|u_h^{n-1} - u_h\|^2.$$

Comme en pratique u_h^0 est souvent choisi de manière arbitraire, la convergence quadratique n'est obtenue qu'asymptotiquement et les premières itérations peuvent être très lentes. En plus, Si A_h et $\widetilde{A}_h$ ne sont pas calculés avec précision, l'ordre de la méthode dégénère. En effet, si A_h est remplacée par $A_h + \varepsilon$ et $\widetilde{A}_h$ par $\widetilde{A}_h + \eta$, où ε et η sont des opérateurs de petite norme symbolisant les imprécisions, alors l'erreur de la méthode de Newton décroît de la façon suivante [65] :

$$\|u_h^{n+1} - u_h\| \leq K\big(\|u_h^n - u_h\|^2 + \|\eta(u_h^n)\|\|u_h^n - u_h\| + \varepsilon(u_h^n)\big).$$

Si on choisit pour $\widetilde{A}_h(u_h^{n-1})$ une approximation discrète de la jacobienne de $A(u^{n-1})$, cela correspond aux méthodes de **quasi-Newton**. Les méthodes du point fixe sont un cas particulier des ces méthodes : elles sont obtenues quand on choisit pour approximation de la Jacobienne une linéarisation de A_h. En particulier, si on prend $\widetilde{A}_h(u_h^{n-1}) = A_h(u_h^{n-1})$, les itérées (2.1) se résument aux itérées de Picard

$$\begin{aligned} &u_h^{n-1} \text{ donné pour } n = 1, 2, \cdots \\ &\widetilde{A}_h(u_h^{n-1})u_h^n = f_h. \end{aligned} \tag{2.2}$$

Les équations de Navier-Stokes stationnaires définies sur un domaine Ω avec des conditions aux limites adéquates sur le bord du domaine $\partial\Omega$ s'écrivent

$$\begin{cases} A(u)u + \nabla q = g \text{ dans } \Omega \\ \nabla \cdot u = 0 \text{ dans } \Omega \end{cases}$$

avec

$$A(u)v := -\triangle v + \mathcal{R}e\, u \cdot \nabla v.$$

On considère ici le cas le plus général des équations de Navier-Stokes, et non pas la système qui nous intéresse. De plus, le terme additionnel lié à l'opérateur D_1 est linéaire. La première variable correspond à la vitesse du fluide et la seconde correspond à la pression.

Pour les équations de Navier-Stokes stationnaires, les choix sont donc les suivants :

– pour la méthode de Newton, on choisit comme opérateur $\widetilde{A_h}$ une discrétisation de l'opérateur différentiel $\widetilde{A}$ défini par

$$\widetilde{A}(u^{n-1})u = -\triangle u + \mathcal{R}e\ u^{n-1} \cdot \nabla u + \mathcal{R}e\ u \cdot \nabla u^{n-1}$$

– pour les itérations du point fixe de Picard (2.2), on choisit

$$\widetilde{A}(u^{n-1})u = -\triangle u + \mathcal{R}e\ u^{n-1} \cdot \nabla u.$$

On peut aussi envisager un shéma explicite en posant

$$\widetilde{A}(u^{n-1})u = -\triangle u,$$

où encore les shéma semi-implicites [32, 99]

$$\widetilde{A}(u^{n-1})u = -\triangle u + \mathcal{R}e\ w^{n-1} \cdot \nabla u,$$

avec

$$w^{n-1} = \omega u^n + (1-\omega)u^{n-1}, \quad 0 < \omega < 1.$$

Il est bien connu que les méthodes de Picard et de quasi-Newton ont un ordre de convergence inférieur à celui de la méthode de Newton : pour la méthode du point fixe, la convergence est linéaire, c'est à dire qu'il existe $C > 0$ tel que

$$\|u_h^{n+1} - u_h\| \leq C\|u_h^n - u_h\|,$$

tandis que pour les méthodes de quasi-Newton, cette convergence est au mieux super linéaire, c'est à dire que :

$$\lim_{n\to+\infty} \frac{\|u_h^{n+1} - u_h\|}{\|u_h^n - u_h\|} = 0.$$

On sait que le schéma explicite converge très lentement, et n'est jamais utilisé pour de tels problèmes. Le schéma implicite est parfois utilisé mais on peut accélérer de telles itérations avec un bon choix de $\widetilde{A}(u^{n-1})$ et en relaxant par un paramètre ω^{n-1} bien choisi. De cette manière, on peut récrire les itérées de quasi-Newton de la façon suivante :

$$\begin{aligned} &u_h^{n-1} \text{ donné pour } n = 1, 2, \cdots \\ &u_h^n = u_h^{n-1} - \omega^{n-1}[\widetilde{A}_h(u_h^{n-1})]^{-1} r^{n-1}. \end{aligned} \tag{2.3}$$

Pour notre problème de Navier-Stokes, un tel choix consistera à choisir pour $\widetilde{A}_h$ une discrétisation de $\widetilde{A}$ définie par

$$\widetilde{A}(u_h^{n-1})u = -\triangle u + \mathcal{R}e\ w^{n-1} \cdot \nabla u$$

où w^{n-1} est un vecteur dépendant de u^{n-1}. Ce type de modèle inclut les schémas semi-implicites [32] et la méthode nommée "The adaptive fixed point defect correction method" par Turek [103]. Si les paramètres de relaxation ω^n sont bien choisis, on remarque dans la pratique que le temps complet de calcul de cette méthode est inférieur au temps que l'on aurait pris en choisissant la méthode de Newton (voir [103]).

A chaque itération de la méthode de quasi-Newton (2.3) décrite ci-dessus, nous devons inverser un système linéaire. Ce problème peut être vu comme la discrétisation par éléments finis du problème linéaire continu d'inconnues u et p suivant :

$$\begin{aligned} -\triangle u + \mathcal{R}e\,(w\cdot\nabla)u + \nabla p &= f && \text{in } \Omega, \\ \nabla\cdot u &= 0 && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{2.4}$$

où w est une fonction donnée. Ce système est appelé **système d'Oseen** ou système de Stokes généralisé. La formulation variationnelle discrète des équations d'Oseen peut s'écrire sous la forme du problème de point-selle généralisé suivant : trouver $u_h \in V_h$ et $p_h \in Q_h$ tels que

$$\begin{cases} a(u_h, v_h) + b(v_h, p_h) = (f, v_h), & \forall v_h \in V_h, \\ b(u_h, q_h) - c(p_h, q_h) = (g, q_h), & \forall q_h \in Q_h, \end{cases} \tag{2.5}$$

où a est la forme linéaire associée à l'opérateur de convection-diffusion dans (2.4), b est associée à l'opérateur de divergence et c correspond à la stabilisation en pression.

2.2 Le système d'Oseen

La formulation matricielle du problème (2.5) s'écrit de la façon suivante :

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}. \tag{2.6}$$

La matrice $\boldsymbol{A}$ est non symétrique, définie et positive puisqu'elle est associée à l'opérateur de convection-diffusion avec addition d'une contrainte de divergence nulle sur la vitesse de convection. La matrice $\boldsymbol{E}$ correspond à la stabilisation en pression. Nous supposerons que cette matrice est semi-définie positive, ce qui est vérifié par les stabilisations de Brezzi-Pitkaranta [18], de Franca-Frey [48] et de Rebollo [90]. Nous considérons, dans ce paragraphe, que les problèmes de point-selle classiques ($\boldsymbol{E} = \boldsymbol{0}$) sont un cas particulier de (2.6). Plusieurs types de méthodes peuvent être utilisés pour résoudre de tels problèmes. Nous n'en citerons ici que quelques-unes. On pourra trouver une liste plus complète dans [88] et [103].

2.2.1 La méthode multigrille avec le lisseur de type Vanka.

Cette méthode consiste en une relaxation locale : elle agit directement sur chaque éléments comme itération interne d'une itération de Jacobi ou de Gauss-Seidel. Si on appelle $\mathcal{M}_K$ la sous-matrice de $\mathcal{M} = \left(\begin{smallmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{smallmatrix}\right)$ associée aux degrés de liberté de l'élément K, alors les itérations du lisseur de Vanka complet s'écrivent :

$$\begin{pmatrix} u \\ p \end{pmatrix}_K \mapsto \begin{pmatrix} u \\ p \end{pmatrix}_K + \mathcal{M}_K^{-1} \left(\begin{pmatrix} f \\ g \end{pmatrix} - \mathcal{M} \begin{pmatrix} u \\ p \end{pmatrix} \right)_K , \tag{2.7}$$

où l'indice K désigne la restriction aux noeuds de l'élément K. Si on remplace dans $\mathcal{M}_K$ la sous-matrice relative à la vitesse par sa diagonale, on obtient le lisseur de Vanka diagonal. John et Tobiska ont programmé les cas-test de référence en dimension 2 [63] et ont montré que le lisseur diagonal était très performant sur des machines parallèles. Plus récemment, d'autres variantes comme les lisseurs de Braess-Sarazin [14] ou, plus généralement, du type Schwarz ont été introduits. Les lisseurs de type Schwarz ont été adaptés aux problèmes de point selle symétriques par Schöberl et Zulehner [92]. Ce type de méthode est extrêmement performant quand il est associé aux itérations de quasi-Newton pour résoudre le problème non linéaire. Leur inconvénient est la difficulté d'implantation. Elles nécessitent de récupérer les matrices extraites et une bonne partie du temps de calcul consiste en l'extraction de ces sous-matrices. Si on n'assemble pas les matrices globales, le temps de calcul est consacré à l'assemblage des sous-matrices. Dans les deux cas, ces méthodes ne permettent qu'assez peu de profiter des bibliothèques optimisées de type LAPACK ou BLAS. Les auteurs utilisant ce type d'algorithmes se trouvent souvent dans les laboratoires les ayant développées et bénéficiant d'un code déjà existant.

2.2.2 Les méthodes de type Uzawa

Elles consistent à travailler sur l'espace de la pression grâce au complément de Schur. L'idée de départ est la suivante : la résolution de (2.6) est équivalente à celle de

$$\begin{cases} \boldsymbol{A}up = f - \boldsymbol{B}^T, \\ (\boldsymbol{E} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T)p = \boldsymbol{B}\boldsymbol{A}^{-1}f - g, \end{cases} \tag{2.8}$$

On note $\boldsymbol{S} := \boldsymbol{E} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$ le complément de Schur. Il suffit donc de résoudre le problème en pression pour connaître la vitesse. Il est bien évident que cette matrice $\boldsymbol{S}$ ne peut être

construite explicitement puisqu'il est inimaginable de calculer l'inverse de $\boldsymbol{A}$. Pour le moment, nous supposons disposer d'une méthode efficace permettant d'inverser $\boldsymbol{A}$: ce point sera traité ultérieurement. Dès lors, on peut connaître la solution en pression en utilisant n'importe quelle méthode itérative sur $\boldsymbol{S}$ (par exemple un Gradient Conjugué si $\boldsymbol{S}$ est symétrique ou un GMRES sinon). On peut aussi utiliser des itérations du type

$$p^n = p^{n-1} - \widetilde{\boldsymbol{S}}^{-1}(\boldsymbol{S}p^{n-1} - \varrho), \tag{2.9}$$

où $\varrho = \boldsymbol{B}^T\boldsymbol{A}^{-1}f - g$ et où $\widetilde{\boldsymbol{S}}$ est un préconditionneur de $\boldsymbol{S}$. Dans la même idée, une des méthodes les plus utilisées récemment est l'algorithme d'Uzawa inexact [25, 58] :

$$\begin{cases} u_{n+1} = u_n + \widetilde{\boldsymbol{A}}^{-1}[f - (\boldsymbol{A}u_n + \boldsymbol{B}p_n)], \\ p_{n+1} = p_n + \widetilde{\boldsymbol{S}}^{-1}(\boldsymbol{B}^T u_{n+1} - \boldsymbol{E}p_n - g), \end{cases} \tag{2.10}$$

où $\widetilde{\boldsymbol{A}}$ est un préconditionneur de $\boldsymbol{A}$. Nous voyons que le point crucial pour appliquer (2.9) ou (2.10) est de connaître un préconditionneur pour $\boldsymbol{S}$. Il existe principalement trois types d'approches pour choisir $\widetilde{\boldsymbol{S}}$:

- la première, issue des méthodes de type SIMPLE (Semi-Implicit Method for Pressure-Linked Equations), consiste à prendre $\widetilde{\boldsymbol{S}} = \boldsymbol{E} + \boldsymbol{B}\widetilde{\boldsymbol{A}}^{-1}\boldsymbol{B}^T$ où $\widetilde{\boldsymbol{A}}$ est une approximation de $\boldsymbol{A}$. On peut prendre en première approximation la diagonale de $\boldsymbol{A}$.
- la seconde approche correspond aux méthodes d'Uzawa les plus classiquement rencontrées : on prendra simplement pour $\widetilde{\boldsymbol{S}}$ un équivalent spectral du complément de Schur.
- dernièrement, on peut inverser exactement le complément de Schur sur des sous domaines de taille raisonnable Ω_i et on prendra $\widetilde{\boldsymbol{S}}^{-1} = \sum_i(\boldsymbol{E}_{|\Omega_i} + \boldsymbol{B}_{|\Omega_i}\boldsymbol{A}_{|\Omega_i}^{-1}\boldsymbol{B}_{|\Omega_i}^T)^{-1}$. Ces idées sont très proches de celles de Zulhener [92] pour les lisseurs de type Schwarz.

2.2.3 Préconditionnement du problème complet

L'autre méthode consiste à appliquer une méthode de résolution générale à la matrice complète. Bien sur, le problème principal est de trouver un bon préconditionneur de cette matrice. Les préconditionneurs proposés généralement triangulaires ou diagonaux du type $\begin{pmatrix} \widetilde{\boldsymbol{A}} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{S}} \end{pmatrix}$ ou $\begin{pmatrix} \widetilde{\boldsymbol{A}} & \boldsymbol{0} \\ \boldsymbol{B} & \widetilde{\boldsymbol{S}} \end{pmatrix}$. Cette fois encore, $\widetilde{\boldsymbol{A}}$ et $\widetilde{\boldsymbol{S}}$ sont des préconditionneurs de $\boldsymbol{A}$ et $\boldsymbol{S}$. Ce choix à l'intérieur des méthodes QMR et GMRES a été étudié par plusieurs auteurs. Nous ne citerons que les derniers articles concernant le GMRES. Dans [69], Klawonn et Starke établissent la convergence du GMRES quand $\boldsymbol{E} = \boldsymbol{0}$ et quand le préconditionneur de $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix}$ est $\begin{pmatrix} \widetilde{\boldsymbol{A}} & \boldsymbol{B}^T \\ \boldsymbol{0} & -\widetilde{\boldsymbol{S}} \end{pmatrix}$, où $\widetilde{\boldsymbol{A}}$

et $\widetilde{\boldsymbol{S}}$ sont des préconditionneurs particuliers de $\boldsymbol{A}$ et $\boldsymbol{S}$. Krzyzanowski [73] élargit cette analyse au cas où $\boldsymbol{E}$ est non nul et avec l'hypothèse que les espaces d'éléments finis sont stables au sens de Babuska-Brezzi. Dernièrement, Zhi-Hao Cao [26] a étudié le taux de convergence du GMRES appliqué au système (2.6) dans le cas où ce système est issu d'une discrétisation par éléments finis du même ordre. Il choisit comme préconditionneur $\begin{pmatrix} \widetilde{\boldsymbol{A}} & \boldsymbol{B}^T \\ \boldsymbol{0} & -\widetilde{\boldsymbol{S}} \end{pmatrix}$, où $\widetilde{\boldsymbol{S}}$ est un équivalent spectral du complément de Schur et a montré le taux de convergence du GMRES préconditionné était indépendant de h. Loghin [76] a montré également un résultat similaire dans le cadre d'un préconditionneur triangulaire par blocs non symétrique sous réserve que le pseudo-spectre du système préconditionné possède les mêmes propriétés d'indépendance vis à vis de h que le spectre de ce même système.

Beaucoup des méthodes citées ci-dessus peuvent se décomposer en deux sous-problèmes :

- trouver un préconditionneur pour la matrice $\boldsymbol{A}$ associée à l'opérateur de convection-diffusion.
- trouver un préconditionneur pour le complément de Schur $\boldsymbol{S}$.

Dans cette thèse, nous nous intéresserons à ces deux sous-problèmes. Le travail de préconditionnement sur le complément de Schur fera l'objet des chapitres 3, 4 et 5. Le préconditionnement du bloc convection-diffusion sera traité dans le chapitre 6.

2.3 Préconditionner le problème de convection-diffusion

Il existe une littérature très variée portant sur la résolution du problème de convection-diffusion. Nous la décomposerons en deux parties : les méthodes multigrilles et les méthodes de décomposition de domaine.

2.3.1 Méthodes multigrilles

Les méthodes multigrilles classiques (ou géométriques) sont les méthodes les plus performantes pour résoudre le problème de Poisson. Malheureusement, leur adaptation aux problèmes de convection-diffusion n'est pas aisée. Les difficultés viennent des lisseurs : on utilise classiquement les méthodes de Jacobi ou Gauss-Seidel. La méthode de Jacobi s'avère très rapidement insuffisante pour les problèmes non symétriques. La méthode de Gauss-Seidel peut s'utiliser, mais uniquement dans le cas d'une discrétisation de type 'upwind' associée à un

réordonnancement [66, 104], ce qui fait de la matrice de rigidité une M-matrice (A.2). De plus, Turek propose de n'utiliser Gauss-Seidel que pour un maillage dont le ratio est presque constant [103]. Il est également possible d'utiliser une factorisation incomplète, mais l'effet lissant de la factorisation n'est vraie qu'en dimension deux [97]. L'unique possibilité de travailler avec une méthode multigrille géométrique est d'utiliser des lisseurs plus compliqués, comme les 'relaxations en ligne' en dimension deux, dont l'idée est similaire à celle des lisseurs de type Vanka [15]. L'adaptation en trois dimensions de ces relaxations n'est pas évidentes. Plusieurs auteurs ont d'ailleurs abandonné les multigrilles pour se tourner vers les problèmes de décomposition de domaines [16]. Une autre approche consiste à utiliser des multigrilles algébriques [97, 13, 9] (à la différence des multigrilles géométriques), qui n'utilisent pas d'informations relatives au maillage. Ces méthodes sont très efficaces comme boîte noire, car elles opèrent indépendamment du problème sous-jacent. Étant donné que nous avons une géométrie très particulière, nous les avons écartées. Nous avons préféré nous tourner vers les méthodes de décomposition de domaines qui permettent d'exploiter les propriétés géométriques de notre problème.

2.3.2 Méthodes de décomposition de domaine

Nous nous intéresserons uniquement aux décompositions de domaine sans recouvrement, c'est à dire que les domaines considérés ont une intersection d'intérieur vide. Pour utiliser une méthode de décomposition de domaine, il faut d'abord choisir quelles conditions aux limites à l'interface peuvent être utilisées. Ensuite, nous devons choisir notre algorithme de résolution. Nous allons décrire deux méthodes : l'algorithme de Schwarz, qui est l'algorithme historique de toute décomposition de domaine, puis la méthode FETI ("Finite Element Tearing and Interconnecting") qui utilise les multiplicateurs de Lagrange pour imposer la continuité aux interfaces. Supposons que notre domaine de travail Ω se décompose en J sous-domaines $\Omega_1, \cdots, \Omega_J$ dont l'interface entre Ω_i et Ω_j se note $\Gamma_{i,j}$. Supposons que nous cherchons les solutions du problème linéaire continu suivant :

$$\begin{cases} \mathcal{L}(u) = f & \text{dans } \Omega, \\ \mathcal{C}(u) = g & \text{sur } \partial\Omega, \end{cases} \tag{2.11}$$

où $\mathcal{L}$ est l'opérateur continu de convection-diffusion défini par

$$\mathcal{L}(u) = -\nu\triangle u + w \cdot \nabla u, \tag{2.12}$$

et $\mathcal{C}$ est un opérateur différentiel défini sur le bord du domaine. Nous supposons que $\nabla \cdot w \in L^\infty(\Omega)$.

Algorithme de Schwarz Soit $k \geq 0$. L'itération k de l'algorithme de Schwarz ('additif') appliqué au problème (2.11) s'écrit de la façon suivante : Pour $i = 1, \cdots, J$,

$$\begin{cases} \mathcal{L}(u_i^{k+1}) = f & \text{dans } \Omega_i, \\ \mathcal{C}(u_i^{k+1}) = g & \text{sur } \partial\Omega \cap \partial\Omega_i, \\ \mathcal{B}(u_i^{k+1}) = \mathcal{B}(u_j^k) & \text{sur } \Gamma_{i,j}, \forall 1 \leq j \leq k, \end{cases} \tag{2.13}$$

où u_i^k est l'approximation de u restreinte au domaine Ω_i, et où $\mathcal{B}$ est un opérateur différentiel défini sur les interfaces $\Gamma_{i,j} := \partial\Omega_i \cap \partial\Omega_j, 1 \leq i, j \leq J$. Toute la difficulté de la méthode de Schwarz est de bien choisir ces conditions aux interfaces. Les conditions de Dirichlet ne suffisent pas à faire converger la méthode de Schwarz pour des domaines sans recouvrement.

Conditions aux limites à l'interface Il convient donc de définir des conditions aux limites pour lesquelles l'algorithme de décomposition de domaine converge et pour lesquelles les sous-problèmes sur chaque interface sont bien posés. Nataf et Rogier ont introduit dans

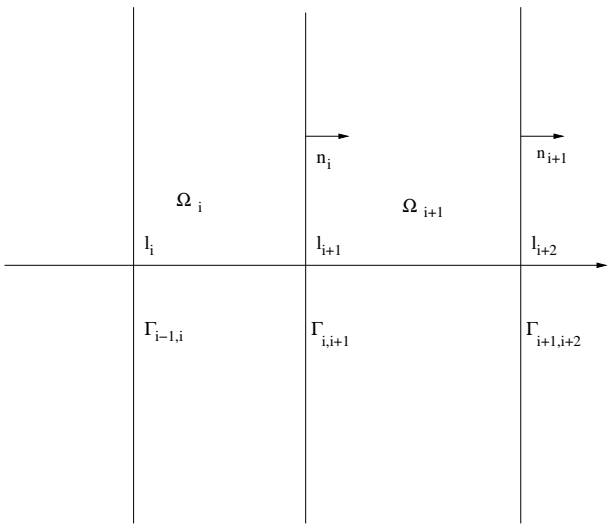


FIG. 2.1 – Décomposition de domaines de Nataf et Rogier

[81] des conditions aux limites de type Robin quand $\Omega = \mathbb{R}^2$ et quand les partitions sont les Ω_i définis par $\Omega_i =]l_i, l_{i+1}[\times \mathbb{R}$, et les interfaces sont les $\Gamma_{i,i+1} = \{l_{i+1}\} \times \mathbb{R}$ (voir figure 2.1). Ces conditions aux limites sont des approximations des opérateurs des conditions aux limites

artificielles introduites dans [55]. Dans le cas de notre problème de convection-diffusion stationnaire, les conditions aux limites de Nataf et Rogier [81] coïncident avec celles de Carlenzoli et Quarteroni [27]. Elles consistent à considérer les itérations suivantes pour $i = 1, \cdots, J$:

$$\begin{cases} \mathcal{L}(u_i^{k+1}) = f & \text{dans } \Omega_i \\ \mathcal{C}(u_i^{k+1}) = g & \text{sur } \partial\Omega \cap \partial\Omega_i, \\ \mathcal{B}_{i+1}(u_i^{k+1}) = B_{i+1}(u_i^k) & \text{sur } \Gamma_{i,i+1} \\ \mathcal{B}_i(u_i^{k+1}) = \mathcal{B}_i(u_{i-1}^{k+1}) & \text{sur } \Gamma_{i-1,i} \end{cases} \tag{2.14}$$

où $\mathcal{B}_{i+1}$ est l'opérateur d'interface défini sur $\Gamma_{i,i+1}$ par

$$\mathcal{B}_{i+1} = \nu \frac{\partial}{\partial n_{i+1}} - \min(w \cdot n_{i+1}, 0), \tag{2.15}$$

et où n_{i+1} la normale extérieure sortante à Ω_{i+1} restreinte à $\Gamma_{i,i+1}$. Ces conditions aux limites ne sont rien d'autre que le développement à l'ordre 0 des conditions aux limites artificielles exactes définies dans [81]. Pour plus de précision, on peut développer à l'ordre 2 ces conditions aux limites, mais certains termes explosent quand le flux est tangentiel à l'une des interfaces [60]. Dans [1, 61, 59], les auteurs ont développé une condition aux limites (la condition d'interface optimisée OO2) de la forme

$$\nu \frac{\partial}{\partial n} - c_1 + c_2 \frac{\partial}{\partial \tau} - c_3 \frac{\partial^2}{\partial \tau^2}, \tag{2.16}$$

où τ représente la direction tangentielle à l'interface. Le coefficient c_1 est donné, c_2 et c_3 sont choisis de manière à minimiser la taux de convergence de la méthode de Schwarz. L'adaptation de telles conditions aux limites en 3D est de loin non triviale, car cela nécessiterait de trouver 5 coefficients. Achdou et al [2] proposent d'utiliser directement les conditions aux limites de Robin

$$\mathcal{B}_{i+1} = \nu \frac{\partial}{\partial n_{i+1}} - w \cdot n_{i+1}, \tag{2.17}$$

et ce sont ces conditions qui ont été retenues par Toselli [100] pour adapter les méthodes FETI au problème de convection-diffusion.

Méthode FETI Le méthode FETI consiste à utiliser des multiplicateurs de Lagrange pour résoudre le problème aux interfaces de manière à assurer la continuité des solutions locales. Il existe plusieurs variantes de ces méthodes. On en distinguera 3 : FETI-1, FETI-2, et FETI-DP("Dual Primal FETI"). Nous allons décrire ces méthodes avec une approche matricielle.

Supposons qu'une discrétisation du problème (2.11) soit équivalente au problème matriciel suivant :

$$\boldsymbol{A}u = f. \tag{2.18}$$

Sur chaque sous-domaine, la discrétisation restreinte à $\Omega_i, i = 1, \cdots, J$ de notre formulation variationnelle avec des conditions aux limites suffisantes pour la convergence de l'algorithme de Schwarz nous donne le système suivant :

$$\boldsymbol{A}^{(i)}u^{(i)} = f^{(i)}. \tag{2.19}$$

Après une simple permutation des lignes et colonnes de $\boldsymbol{A}^{(i)}$, nous pouvons récrire notre problème avec les notations de [71] de la façon suivante

$$\begin{pmatrix} \boldsymbol{A}^{(i)}_{I,I} & \boldsymbol{A}^{(i)}_{I,\Gamma} \\ \boldsymbol{A}^{(i)}_{\Gamma,I} & \boldsymbol{A}^{(i)}_{\Gamma,\Gamma} \end{pmatrix} \begin{pmatrix} u^{(i)}_I \\ u^{(i)}_\Gamma \end{pmatrix} = \begin{pmatrix} f^{(i)}_I \\ f^{(i)}_\Gamma \end{pmatrix} \tag{2.20}$$

où l'indice I indique les numéros des noeuds intérieurs du sous-domaine Ω_i et aux numéros des noeuds du bord $\partial\Omega \cap \partial\Omega_i$ tandis que Γ correspond aux noeuds de toutes les interfaces $\Gamma_{i,j}$ pour tout j. Les ensembles d'indices I et Γ dépendent de i, nous omettons cependant les indices pour ne pas alourdir les notations (l'ensemble Γ n'est pas le même dans $\boldsymbol{A}^{(i)}_{\Gamma,\Gamma}$ que dans $\boldsymbol{A}^{(j)}_{\Gamma,\Gamma}$ pour tout $j \neq i$). Le problème (2.18) peut se résoudre grâce à l'équation à l'interface suivante

$$\boldsymbol{S}_\Gamma u_\Gamma = f^*_\Gamma \tag{2.21}$$

où

$$\boldsymbol{S}_\Gamma = \mathrm{diag}(\boldsymbol{S}^{(1)}_\Gamma, \cdots, \boldsymbol{S}^{(J)}_\Gamma), \quad u_\Gamma = (u^{(1)}_\Gamma, \cdots, u^{(J)}_\Gamma)^T, \quad f^*_\Gamma = (f^{(1),*}_\Gamma, \cdots, f^{(J),*}_\Gamma)^T, \tag{2.22}$$

et

$$\boldsymbol{S}^{(i)}_\Gamma := \boldsymbol{A}^{(i)}_{\Gamma,\Gamma} - \boldsymbol{A}^{(i)}_{\Gamma,I}(\boldsymbol{A}^{(i)}_{I,I})^{-1}\boldsymbol{A}^{(i)}_{I,\Gamma}, \quad f^{(i),*}_\Gamma := f^{(i)}_\Gamma - \boldsymbol{A}^{(i)}_{\Gamma,I}(\boldsymbol{A}^{(i)}_{I,I})^{-1}f^{(i)}_I. \tag{2.23}$$

Soit m le nombre de noeuds sur l'interface globale Γ et, pour chaque i, soit m_i le nombre de noeuds sur $\partial\Omega_i \cap \Gamma$. On appelle $\boldsymbol{C}_\Gamma$ la matrice de taille $m \times \sum_{i=1}^J m_i$ composée de 1, 0 et -1 qui définit la contrainte de continuité par

$$\boldsymbol{C}_\Gamma u_\Gamma = 0, \text{ avec } u_\Gamma = (u^{(1)}_\Gamma, \cdots, u^{(J)}_\Gamma)^T. \tag{2.24}$$

Cette matrice se construit simplement de la façon suivante : soit un noeud commun au deux sous-domaines $\overline{\Omega}_i$ et $\overline{\Omega}_j$. On note p sa numérotation pour le problème complet. Il possède

également deux numérotations locales pour les sous-problèmes i et j, c'est à dire que deux indices de u_Γ correspondent à ce noeud. Il suffit alors dans la ligne p de la matrice $\boldsymbol{C}_\Gamma$ d'attribuer les valeurs 1 et -1 pour les colonnes correspondant à ces deux indices et 0 ailleurs.

Dans le cas où $\boldsymbol{A}$ est symétrique, résoudre le problème (2.21) revient à minimiser l'énergie

$$E(u) = \frac{1}{2}(\boldsymbol{S}_\Gamma u, u) - \langle f_\Gamma^*, u\rangle \tag{2.25}$$

sous la contrainte (2.24), et on peut utiliser simplement les multiplicateurs de Lagrange pour résoudre ce problème de minimisation. Lorsque $\boldsymbol{A}$ n'est pas symétrique, on ne considère pas l'équation d'énergie mais, comme dans [100], on applique les multiplicateurs de Lagrange λ de la façon suivante

$$\begin{cases} \boldsymbol{S}_\Gamma u_\Gamma + \boldsymbol{C}_\Gamma^T \lambda = f_\Gamma^*, \\ \boldsymbol{C}_\Gamma u_\Gamma = 0. \end{cases} \tag{2.26}$$

Dans certains cas, il existe un domaine Ω_i sur lequel $\boldsymbol{S}_\Gamma^{(i)}$ n'est pas inversible. C'est le cas pour les conditions aux limites de Neumann sur un domaine n'ayant pas d'intersection avec $\partial\Omega$. Posons

$$\boldsymbol{G} = [\boldsymbol{C}_\Gamma^{(1)}\boldsymbol{R}_1, \cdots, \boldsymbol{C}_\Gamma^{(J)}\boldsymbol{R}_J],$$

où $\boldsymbol{R}_i$ est la matrice dont les colonnes sont les vecteurs de la base du noyau de $\boldsymbol{S}_\Gamma^{(i)}$. On considère alors le système

$$\begin{pmatrix} \boldsymbol{F} & -\boldsymbol{G} \\ -\boldsymbol{G}^T & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \lambda \\ \alpha \end{pmatrix} = \begin{pmatrix} d \\ e \end{pmatrix} \tag{2.27}$$

où

$$\boldsymbol{F} = \sum_{i=1}^{J} \boldsymbol{C}_\Gamma^{(i)} \boldsymbol{S}_\Gamma^{(i),\dagger} \boldsymbol{C}_\Gamma^{(i),T}, \quad d = \sum_{i=1}^{J} \boldsymbol{C}_\Gamma^{(i)} \boldsymbol{S}_\Gamma^{(i),\dagger} f_\Gamma^*, \quad e = [f_\Gamma^{(i),*,T}\boldsymbol{R}_i]^T,$$

avec $\boldsymbol{C}_\Gamma^{(i)}$ la sous matrice de $\boldsymbol{C}_\Gamma$ restreinte aux noeuds de $\partial\Omega_i$ et où $\boldsymbol{S}_\Gamma^{(i),\dagger}$ est le pseudo inverse de $\boldsymbol{S}_\Gamma^{(i)}$. On définit la projection

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T$$

parallèlement à $\mathrm{Im}(\boldsymbol{G})$ et on résout (2.27) par un GMRES ou un gradient conjugué projeté. L'espace $\mathrm{Im}(\boldsymbol{G})$ est appelé l'**espace grossier** et on doit résoudre un sous-problème sur cet espace lorsqu'on inverse $\boldsymbol{G}^T\boldsymbol{G}$.

Cette version correspond à la méthode FETI-1 telle qu'elle a été introduite par Farhat et Roux [47]. Cette méthode est extensible numériquement, dans le sens où, pour des problèmes symétriques d'ordre 2, nous avons les performances suivantes :

- si on fixe le nombre de sous domaines, le conditionnement de la méthode FETI croît asymptotiquement en $\log^2(1/h)$ où h est la taille de la maille,
- le nombre d'itérations de la méthode FETI décroît quand, pour h fixé, on augmente le nombre de sous-domaines,
- quand on fixe la taille H/h des sous-problèmes avec H le diamètre des sous-domaines et qu'on raffine, le conditionnement de la méthode FETI reste constant.

Le FETI classique n'est plus extensible lorsque l'on considère des problèmes elliptiques d'ordre 4. Farhat et Mandel [46] ont alors ajouté une contrainte de continuité et donc des multiplicateurs de Lagrange pour les coins (les noeuds appartenant à plus de deux sous-domaines). Cette variante est appelé la méthode FETI-2 et permet de retrouver l'extensibilité de la méthode. Ces méthodes ont ensuite été généralisées dans [44] sous le nom de la méthode FETI-DP (dual-primal).

L'essentiel de la bibliographie sur ces méthodes est relatif à des problèmes symétriques. Dans [100], Toselli généralise les méthodes FETI avec préconditionneur de type Neumann-Neumann au problème de convection-diffusion. Il définit par analogie des méthodes FETI-1 et FETI-2 associées à un préconditionneur de type Robin-Robin.

2.4 Préconditionner le complément de Schur

De nombreux travaux ont été réalisés pour calculer un équivalent spectral du complément de Schur. Dans un premier temps, nous recenserons les préconditionneurs valables pour des espaces d'éléments finis stables, c'est à dire des espaces vérifiant la condition inf-sup ou condition LBB(Ladyzhenskaya-Babuska-Brezzi) discrète suivante :

$$\sup_{v \in V_h} \frac{(p, \nabla \cdot v)}{\|v\|_{V_h}} \geq \gamma \|p\|_{Q_h}. \tag{2.28}$$

2.4.1 Pour des espaces d'éléments finis stables

La discrétisation des équations de Navier-Stokes par des espaces d'éléments finis stables donne lieu à des problèmes de point-selle (2.6) où $\boldsymbol{E} = \mathbf{0}$. Le préconditionnement de ce type de systèmes a donné lieu à beaucoup d'articles de la part de l'équipe d'Elman et Silvester principalement [39, 38, 40, 41, 42, 94, 105, 106, 95]. Beaucoup d'autres applications donnent lieu à des problèmes de point-selle. Nous ne citerons ici que le problème d'élasticité linéaire parce qu'il a donné lieu à des articles de référence sur le préconditionnement qui nous intéresse.

La discrétisation de ce problème donne un système du type (2.6) avec $\boldsymbol{E} \neq \boldsymbol{0}$ même dans des espaces d'éléments finis stables. Le travail sur ce sujet a été fait essentiellement par Axel Klawonn [69, 67, 68].

Problème de point selle standard Dans [41], Elman et Silvester montrent que la matrice de masse $\boldsymbol{Q}$ est un équivalent spectral du complément de Schur, c'est à dire que

$$\gamma_1 \leq \frac{p^T \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T p}{p^T \boldsymbol{Q} p} \leq \gamma_2$$

où γ_1 et γ_2 sont indépendantes de la taille de la maille. Cette relation découle exactement de la condition inf-sup (2.28). Dans [38], Elman propose de préconditionner $\boldsymbol{S} = \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$ par

$$\widetilde{\boldsymbol{S}} = (\boldsymbol{B}\boldsymbol{B}^T)(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^T)^{-1}(\boldsymbol{B}\boldsymbol{B}^T).$$

mais dans le cas d'un schéma aux différences finies de type MAC (Marker And Cell). Alors, $(\boldsymbol{B}\boldsymbol{B}^T)$ correspond à l'opérateur de Poisson sur l'espace de la pression. Pour les éléments finis, $(\boldsymbol{B}\boldsymbol{B}^T)$ doit être assemblée par produit de matrices creuses [53]. Plus récemment, dans [42], Elman, Silvester et Wathen proposent de prendre

$$\widetilde{\boldsymbol{S}} = \boldsymbol{L}_p \boldsymbol{A}_p^{-1} \boldsymbol{Q}_p$$

où $\boldsymbol{Q}_p$ est la matrice de masse, $\boldsymbol{A}_p$ est la matrice de convection-diffusion et $\boldsymbol{L}_p$ est la matrice du Laplacien avec des conditions aux limites de Neumann. L'indice p fait référence à l' espace de la pression alors que l'indice u (utilisé ci-dessous) fait référence à l'espace de la vitesse. L'idée de ce préconditionneur repose essentiellement sur deux arguments : le premier est heuristique et suppose qu'on peut intervertir des ordres de dérivation, c'est à dire que

$$\nabla(\mathcal{R}e\ w \cdot \nabla - \triangle)_p \sim (\mathcal{R}e\ w \cdot \nabla - \triangle)_u \nabla.$$

Des idées similaires avaient déjà été utilisées avec succès dans [21]. Le deuxième argument est l'inégalité inf-sup alternative de Bercovier et Pironneau [11] :

$$\sup_{v \in V_h} \frac{(v, \nabla p)}{\|v\|_{L^2}} \geq \beta \|\nabla p\|_{L^2}.$$

Ce préconditionneur nécessite à priori de réassembler les matrices du Laplacien et de convection-diffusion sur l'espace de la pression. Or, Graham and al (voir [53]) ont montré numériquement que dans le cas d'éléments finis Q2-Q1, on pouvait se contenter d'approcher les matrices $\boldsymbol{A}_p$

et $\boldsymbol{L}_p$ grâce à une projection par interpolation linéaire de l'espace des éléments finis Q_2 sur l'espace des éléments finis Q_1. Plus précisément, si on note Π cette projection, cela revient à considérer que

$$A_p \sim \Pi^T A_u \Pi.$$

Ce préconditionneur a donné de meilleurs résultats que les précédents dès que le nombre de Reynolds devient grand.

Problèmes de point-selle généralisés Dans [67], Klawonn a étudié l'influence des préconditionneurs triangulaires par blocs pour les problèmes de point-selle symétriques généralisés. Il démontre que si c est uniformément elliptique, que $\boldsymbol{E}$ est spectralement équivalent à $\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$, et que $\widetilde{\boldsymbol{A}}$ est spectralement équivalent à $\boldsymbol{A}$, alors la matrice triangulaire par bloc $\begin{pmatrix} \widetilde{\boldsymbol{A}} & \boldsymbol{B}^T \\ \boldsymbol{0} & -\boldsymbol{E} \end{pmatrix}$ est spectralement équivalente à la matrice $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix}$. Dans [68], il démontre que si la condition inf-sup discrète (2.28) est vérifiée et si la matrice de pénalité $\boldsymbol{E}$ définie dans (2.6) correspond à la matrice de masse, alors le préconditionneur diagonal par bloc $\begin{pmatrix} \widetilde{\boldsymbol{A}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_p \end{pmatrix}$ où $\widetilde{\boldsymbol{A}}$ est un équivalent spectral de $\boldsymbol{A}$ est un bon préconditionneur de (2.6) dans le sens où les valeurs propres du problème préconditionné ne dépendent pas du nombre d'inconnues.

2.4.2 Pour des espaces d'éléments finis non stables

Les premiers travaux sur les problèmes de point-selle généralisés ont été amenés par la résolution du problème de Stokes. Les stabilisations de Brezzi-Pitkaranta [18] et de Franca-Frey [48] ont permis la discrétisation du problème de Stokes par éléments finis de même ordre. Silvester et Wathen ont étudié dans [95] et [106] le comportement du GMRES dans le cas où le complément de Schur est Q_h-elliptique. Récemment, Kay, Loghin et Wathen [64] se sont intéressés au préconditionnement du complément de Schur pour des éléments finis non stables. Ils supposent que le domaine considéré est l'espace entier ($\mathbb{R}^2$ ou $\mathbb{R}^3$) et supposent avoir une convection constante. Ils considèrent l'opérateur différentiel discret $\mathcal{L}$ d'Oseen et, en appliquant la transformation de Fourier, déterminent $\mathcal{G}$ telle que $\widehat{\mathcal{G}} = \widehat{\mathcal{L}}^{-1}$. Ils en déduisent que le terme en pression est de la forme $(-\triangle + \mathcal{R}e\ w \cdot \nabla)\mathcal{G}_\triangle$ où $-\mathcal{G}_\triangle$ est la solution fondamentale du Laplacien et supposent que cette expression représente bien le complément de Schur. L'interprétation matricielle de cette relation est que l'on peut prendre comme préconditionneur

$$\widetilde{\boldsymbol{S}} = \boldsymbol{L}_p \boldsymbol{A}_p^{-1}.$$

Pour des raisons de 'continuité' avec les préconditionneurs connus pour Stokes, les auteurs feront le même choix que Silvester dans [94], c'est à dire

$$\widetilde{\boldsymbol{S}} = \boldsymbol{L}_p \boldsymbol{A}_p^{-1} \boldsymbol{Q}_p.$$

Cette démonstration reste assez formelle, puisque les hypothèses de départ (convection constante et problème défini sur l'espace tout entier) ne correspondent pas aux cas-tests qui nous intéressent. Elle montre cependant que la stabilité des espaces d'éléments finis n'est pas l'unique justification du préconditionneur de Silvester : cela peut se justifier par la signification des opérateurs différentiels eux-mêmes. On peut donc imaginer que ce préconditionneur est valable pour la partie $\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$ du complément de Schur, même si les espaces d'éléments finis ne sont pas stables. Cependant, on remarque que la stabilisation en pression est complètement ignorée dans la partie théorique de l'article de Kay, Loghin et Wathen [64]. Les tests numériques menés par les auteurs montrent que toutes les restrictions théoriques ne sont pas préjudiciables numériquement : les tests numériques en dimension deux confirment que le conditionnement du problème préconditionné discrétisé par des P1-P1 stabilisés ou des P2-P1 est borné indépendamment de h et que le nombre d'itérations nécessaires du GMRES reste raisonnable quand on augmente le nombre de Reynolds. Dans un travail non encore publié [76], Loghin montre, sous réserve que les solutions du problème de d'Oseen appartiennent à H^2, que le module du spectre du système préconditionné est indépendant de la taille de la maille. Dans ce travail, il tient compte du terme de stabilisation en vitesse. Les tests numériques en 2D laissent penser que les performances du GMRES diminuent proportionnellement à $\mathcal{R}e^{1/2}$ mais ce résultat n'est pas prouvé.

2.5 Apport de la thèse

Cette thèse donne un nouveau point de vue sur certains types de préconditionneurs par blocs pour les problèmes de point-selle généralisés. L'étude de la thèse a été menée dans le cas le plus général possible en explicitant les constantes, ce qui a permis de mettre en évidence l'influence de certains paramètres sur les performances du préconditionneur. L'explicitation des paramètres permet d'inclure dans cette étude les problèmes extérieurs tronqués, puisque les constantes de Poincaré deviennent apparentes. En particulier, nous avons mis l'accent sur l'apport de la stabilisation en pression : nous avons montré théoriquement le lien entre la condition inf-sup généralisée et notre classe de préconditionneurs. Cette partie n'avait pas

été traitée par l'équipe d'Elman et Silvester. Cette question, restée en suspens, a également intéressé d'autres auteurs, et Zhi-Hao Cao [26] s'est penché sur le même problème que nous, mais vu du point de vue de la convergence du GMRES. Ce préconditionneur a ensuite été testé numériquement dans des situations nouvelles : d'une part, tous nos tests ont été fait en dimension trois, ce qui n'apparaît quasiment jamais dans la littérature sur les problèmes de point-selle. Le code implanté en C++ a été entièrement développé pendant cette thèse à partir des bibliothèques Blas, Lapack et SparseLib++ et à partir d'une version du mailleur de Deuring qui a ensuite été modifiée. Nous avons d'abord validé notre solveur dans le cadre de la cavité entraînée tridimensionnelle en comparant nos résultats avec les résultats de référence. D'autre part, nous avons appliqué nos résultats aux domaines extérieurs avec maillage exponentiel : l'analyse et l'application n'avaient jamais été faites dans ce cas de figure. Enfin, dans le chapitre sur la partie décomposition de domaine, nous avons testé le FETI-1 en trois dimensions pour un maillage non standard : le traitement de ces deux difficultés constituent chacun une nouveauté. Nous avons montré l'influence du sens du vecteur d'advection par rapport aux sous-domaines, puis l'influence de la constante de Poincaré dans le cas des domaines tronqués.

Chapitre 3

Problèmes de point-selle généralisés

Comme nous l'avons vu dans l'introduction, la discrétisation et linéarisation des équations de Navier-Stokes mènent à des problèmes de point-selle. Nous définirons comme problème de point-selle tout problème pouvant s'écrire sous la forme d'un système linéaire par bloc 2×2 du type

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix}. \tag{3.1}$$

Or, les problèmes de point-selle apparaissent dans des situations très variées. Les problèmes de point-selle les plus classiquement rencontrés sont de la forme (3.1) avec $\boldsymbol{E} = \boldsymbol{0}$. On les obtient par exemple lorsqu'on résout des problèmes d'optimisation sous contrainte avec des multiplicateurs de Lagrange. Il existe cependant plusieurs applications menant à des problèmes de point-selle avec $\boldsymbol{E} \neq \boldsymbol{0}$: les problèmes d'élasticité linéaire, les problèmes de moindre carré pondérés et certaines méthodes de point intérieurs en optimisation (voir [80]). Dans ce chapitre, nous étudions l'action d'un préconditionneur triangulaire appliqué aux problèmes de point-selle généralisés. Nous nous intéressons particulièrement à l'estimation des valeurs propres du problème préconditionné et nous montrons que ces valeurs propres sont incluses dans un rectangle dont les bords sont indépendants de la taille des espaces de discrétisation. Les calculs sont menés dans un cadre très général de manière à pouvoir utiliser directement ces résultats pour d'autres applications.

3.1 Cadre général

Soient V et Q deux espaces de Hilbert dont les produits scalaires sont notés respectivement $(\ ,\ )_V$ et $(\ ,\ )_Q$. Leur normes associées sont naturellement notées $\|\ \|_V$ et $\|\ \|_Q$.
Nous notons V' et Q' les espaces duaux de V et Q respectivement. Soit

$$a : V \times V \longrightarrow \mathbb{R} \quad \text{et} \quad b : V \times Q \longrightarrow \mathbb{R}$$

deux formes bilinéaires telles que a est continue et coercive, et b est continue, i.e., il existe des constantes $\mathcal{A} > 0$, $\alpha > 0$ et $\mathcal{B} > 0$ tels que

$$\forall u, v \in V, \quad |a(u,v)| \leq \mathcal{A}\|u\|_V \|v\|_V, \tag{3.2}$$

$$\forall u \in V, \quad a(u,u) \geq \alpha \|u\|_V^2, \tag{3.3}$$

et

$$\forall u \in V,\ \forall p \in Q, \quad |b(u,p)| \leq \mathcal{B}\|u\|_V \|p\|_Q. \tag{3.4}$$

Soient V_h (respectivement Q_h) deux sous-espaces vectoriels de dimension finie de V (respectivement Q). Dans ce chapitre, h sert de paramètre formel pour les espaces vectoriels de dimension finie.
Soit c une forme bilinéaire symétrique définie sur $Q_h \times Q_h$ telle que

$$\forall q_h \in Q_h, \quad c(q_h, q_h) \geq 0, \tag{3.5}$$

et

$$\forall p_h, q_h \in Q_h, \quad |c(p_h, q_h)| \leq \mathcal{C}\|p_h\|_Q \|q_h\|_Q, \tag{3.6}$$

avec $\mathcal{C} > 0$ constante. Soient $\varrho \in V', \kappa \in Q'$. Nous considérons le problème de point-selle généralisé suivant :

$$\begin{array}{l} \text{Trouver } u_h \in V_h, p_h \in Q_h \text{ tels que} \\ \left\{ \begin{array}{ll} a(u_h, v_h) + b(v_h, p_h) = \varrho(v_h) & \forall v_h \in V_h, \\ b(u_h, q_h) - c(p_h, q_h) = \kappa(q_h) & \forall q_h \in Q_h. \end{array} \right. \end{array} \tag{3.7}$$

Nous allons démontrer dans un premier temps le théorème d'existence suivant :

Théorème 3.1 *Supposons que les formes bilinéaires b et c satisfont la condition inf-sup généralisée*

$$\sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V} + c(q_h, q_h)^{1/2} \geq \gamma \|q_h\|_Q, \quad \textit{pour tout } q_h \in Q_h, \tag{3.8}$$

où γ est une constante positive. Alors, le système (3.7) a une solution unique. De plus, si on appelle (u_h, p_h) sa solution, on obtient les inégalités suivantes

$$\|u_h\|_V \leq C_1\big(\|\varrho\|_{V'} + \|\kappa\|_{Q'}\big) \quad \text{et} \quad \|p_h\|_Q \leq C_2\big(\|\varrho\|_{V'} + \|\kappa\|_{Q'}\big),$$

où $C_1 = C_1(\gamma, \mathcal{A}, \alpha)$ et $C_2 = C_2(\gamma, \mathcal{A}, \alpha)$ sont des constantes ne dépendant que de $\gamma, \mathcal{A}$ ou α.

Preuve : Si on choisit $u_h = v_h$ et $q_h = p_h$ dans (3.7), et si on soustrait les deux équations dans (3.7), alors on obtient

$$a(u_h, u_h) + c(p_h, p_h) = \varrho(u_h) - \kappa(p_h).$$

Comme c est semi-définie positive et a est coercive, il s'ensuit que

$$\alpha\|u_h\|_V^2 \leq \|\varrho\|_{V'}\|u_h\|_V + \|\kappa\|_{Q'}\|p_h\|_Q,$$

ce qui implique que

$$\frac{\alpha}{2}\|u_h\|_V^2 \leq \frac{1}{2\alpha}\|\varrho\|_{V'}^2 + \|\kappa\|_{Q'}\|p_h\|_Q. \tag{3.9}$$

De plus, la condition inf-sup généralisée (3.8) impose que

$$\begin{aligned}
\gamma^2\|p_h\|_Q^2 &\leq 2c(p_h, p_h) + 2\left(\sup_{v_h \in V_h} \frac{b(v_h, p_h)}{\|v_h\|_V}\right)^2 \\
&= -2\kappa(p_h) + 2b(u_h, p_h) + 2\left(\sup_{v_h \in V_h} \frac{b(v_h, p_h)}{\|v_h\|_V}\right)^2 \\
&\leq -2\kappa(p_h) + 2\varrho(u_h) - 2a(u_h, u_h) + 2\left(\sup_{v_h \in V_h} \frac{\varrho(v_h) - a(u_h, v_h)}{\|v_h\|_V}\right)^2 \\
&\leq 2\|\kappa\|_{Q'}\|p_h\|_Q + 2\|\varrho\|_{V'}\|u_h\|_V + 2\mathcal{A}\|u_h\|_V^2 + 4\|\varrho\|_{V'}^2 + 4\mathcal{A}^2\|u_h\|_V^2,
\end{aligned}$$

d'où

$$\frac{\gamma^2}{2}\|p_h\|_Q^2 \leq \frac{2}{\gamma^2}\|\kappa\|_{Q'}^2 + 8\|\varrho\|_{V'}^2 + (1 + 2\mathcal{A} + 4\mathcal{A}^2)\|u_h\|_V^2. \tag{3.10}$$

En combinant (3.9) et (3.10), on obtient

$$\begin{aligned}
\frac{\gamma^2}{2}\|p_h\|_Q^2 &\leq \frac{2}{\gamma^2}\|\kappa\|_{Q'}^2 + 5\|\varrho\|_{V'}^2 + \frac{1 + 2\mathcal{A} + 4\mathcal{A}^2}{\alpha^2}\|\varrho\|_{V'} + \frac{2(1 + 2\mathcal{A} + 4\mathcal{A}^2)}{\alpha}\|\kappa\|_{Q'}\|p_h\|_Q \\
&\leq \frac{2}{\gamma^2}\|\kappa\|_{Q'}^2 + 5\|\varrho\|_{V'}^2 + \frac{1 + 2\mathcal{A} + 4\mathcal{A}^2}{\alpha^2}\|\varrho\|_{V'} + \frac{4(1 + 2\mathcal{A} + 4\mathcal{A}^2)^2}{\alpha^2\gamma^2}\|\kappa\|_{Q'}^2 + \frac{\gamma^2}{4}\|p_h\|_Q^2.
\end{aligned}$$

On peut conclure que

$$\forall p_h \in Q_h, \quad \|p_h\|_Q \leq C_1(\|\varrho\|_{V'} + \|\kappa\|_{Q'}), \tag{3.11}$$

avec $C_1 = \dfrac{4}{\gamma^2}\left(\dfrac{2}{\gamma^2} + 5 + \dfrac{1+2\mathcal{A}+4\mathcal{A}^2}{\alpha^2} + \dfrac{4(1+2\mathcal{A}+4\mathcal{A}^2)^2}{\alpha^2\gamma^2}\right)$. Maintenant, en insérant (3.11) dans (3.9), on obtient que

$$\forall u_h \in Q_h, \quad \|u_h\|_V \leq C_2(\|\varrho\|_{V'} + \|\kappa\|_{Q'}), \tag{3.12}$$

où $C_2 = C_2(\gamma, \mathcal{A}, \alpha)$ dépend seulement de $\gamma, \mathcal{A}$ ou α. L'unicité de u_h et p_h se déduit directement de (3.12) et (3.11).

□

3.2 Préconditionnement

3.2.1 Définitions

On appelle n la dimension de V_h et m la dimension de Q_h. Soit $(\varphi_i)_{1\leq i\leq n}$ (respectivement $(\psi_j)_{1\leq j\leq m}$) une base de V_h (respectivement Q_h). On définit les matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{E}, \boldsymbol{V}, \boldsymbol{Q}$ et les vecteurs F et G par

$$\boldsymbol{A}_{i,j} = a(\varphi_j, \varphi_i), \quad \boldsymbol{B}_{i,j} = b(\varphi_j, \psi_i), \quad \boldsymbol{E}_{i,j} = c(\psi_j, \psi_i),$$

$$\boldsymbol{V}_{i,j} = (\varphi_j, \varphi_i)_V, \quad \boldsymbol{Q}_{i,j} = (\psi_j, \psi_i)_Q,$$

$$F_i = f(\varphi_i), \quad G_i = \kappa(\psi_i).$$

Soit $(\ ,\ )$ le produit scalaire usuel de $\mathbb{R}^n$ (resp. $\mathbb{R}^m$), défini par $(U, V) = U^T V$ pour tout $U, V \in \mathbb{R}^n$ (resp. $\mathbb{R}^m$). D'où, si on a $u_h, v_h \in V_h$, $p_h, q_h \in Q_h$ et $U, V \in \mathbb{R}^n, P, Q \in \mathbb{R}^m$ tels que

$$u_h = \sum_{i=1}^{n} U_i\varphi_i, \ \ v_h = \sum_{i=1}^{n} V_i\varphi_i, \ \ p_h = \sum_{j=1}^{m} P_j\psi_j, \ \ q_h = \sum_{j=1}^{m} Q_j\psi_j, \tag{3.13}$$

alors on obtient les relations suivantes :

$$a(u_h, v_h) = (\boldsymbol{A}U, V), \quad b(u_h, p_h) = (\boldsymbol{B}U, P), \quad c(p_h, q_h) = (\boldsymbol{E}P, Q),$$

$$(p_h, q_h)_Q = (\boldsymbol{Q}P, Q), \quad (u_h, v_h)_V = (\boldsymbol{V}U, V),$$

$$\varrho(v_h) = (F, V), \quad \kappa(q_h) = (G, Q).$$

Avec les définitions précédentes de $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{E}, \boldsymbol{V}, \boldsymbol{Q}, F$ et G, le système (2.6) est équivalent au problème variationnel discret (3.7).

3.2.2 Généralités et résultat principal

En général, deux types de préconditionneurs par blocs sont utilisés pour les problèmes de point-selle : les préconditionneurs diagonaux ([39, 41, 42, 68, 106]) et les préconditionneurs triangulaires ([38, 67, 69, 73, 77]). Nous ne considérons ici que les préconditionneurs triangulaires parce qu'ils semblent plus efficaces. En effet, P. Krzyzanowski [73] a observé que les préconditionneurs triangulaires peuvent être deux fois plus rapide que les préconditionneurs diagonaux. Le préconditionneur de $\left(\begin{smallmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{smallmatrix}\right)$ que nous considérons ici est la matrice $\left(\begin{smallmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B} & -\boldsymbol{Q} \end{smallmatrix}\right)$. Dans ce cas, les valeurs propres du système préconditionné (2.6) sont solutions du problème généralisé aux valeurs propres

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B} & -\boldsymbol{Q} \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix}.$$

La valeur propre $\lambda = 1$ est solution de multiplicité m de ce problème ; Le sous espace propre associé est engendré par les vecteurs propres de la forme $\left(\begin{smallmatrix} U \\ 0 \end{smallmatrix}\right)$. Les autres valeurs propres sont solutions du problème généralisé aux valeurs propres

$$(\boldsymbol{E} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T)P = \lambda \boldsymbol{Q} P. \tag{3.14}$$

Nous allons prouver le théorème suivant :

Théorème 3.2 *Les solutions du problème généralisé aux valeurs propres (3.14) sont incluses dans le domaine rectangulaire*

$$\left[\frac{\gamma^2}{2}\min(1, \alpha(\mathcal{A}_+^2 + \mathcal{A}_-^2)^{-1}), (\frac{\mathcal{B}^2}{\alpha} + \mathcal{C})\right] \times \left[-\frac{\mathcal{B}^2}{2\alpha}, \frac{\mathcal{B}^2}{2\alpha}\right],$$

où $\gamma, \mathcal{B}, \mathcal{C}, \alpha$ sont introduits dans (3.8), (3.6), (3.4), (3.3), respectivement. Les constantes $\mathcal{A}_+$ et $\mathcal{A}_-$ sont les bornes de continuité des parties symétriques et anti-symétrique de a définies ultérieurement en (5.12) , (5.13).

3.3 Preuve du théorème.

Nous définissons les formes bilinéaires a_+ et a_- par

$$\forall u_h, v_h \in V_h, \quad a_+(u_h, v_h) = (a(u_h, v_h) + a(v_h, u_h))/2, \tag{3.15}$$

$$\forall u_h, v_h \in V_h, \quad a_-(u_h, v_h) = (a(u_h, v_h) - a(v_h, u_h))/2. \tag{3.16}$$

On remarque de a_+ et a_- sont respectivement les parties symétriques et anti-symétriques de la forme bilinéaire a. L'uniforme continuité de a implique l'uniforme continuité de a_+ et a_-, c'est à dire qu'il existe deux constants $\mathcal{A}_+ > 0$ et $\mathcal{A}_- > 0$, indépendantes de h, telles que

$$\forall u_h, v_h \in V_h, \quad |a_+(u_h, v_h)| \leq \mathcal{A}_+ \|u_h\|_V \|v_h\|_V, \tag{3.17}$$

$$\forall u_h, v_h \in V_h, \quad |a_-(u_h, v_h))| \leq \mathcal{A}_- \|u_h\|_V \|v_h\|_V. \tag{3.18}$$

De plus, a_+ hérite de la coercivité de a avec le même coefficient α. En utilisant la base $(\varphi_i)_{1\leq i\leq n}$ introduite dans la partie 3.2.1, nous pouvons définir les matrices $\boldsymbol{A}_+$ et $\boldsymbol{A}_-$ par

$$\forall U, V \in \mathbb{R}^n, \forall u_h, v_h \in V_h, \quad a_+(u_h, v_h) = (\boldsymbol{A}_+ U, V) \text{ et } a_-(u_h, v_h) = (\boldsymbol{A}_- U, V),$$

où U, V, u_h, v_h sont donnés comme en (3.13). On a évidemment que $\boldsymbol{A} = \boldsymbol{A}_+ + \boldsymbol{A}_-$. Posons

$$\boldsymbol{S} = \boldsymbol{E} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T, \ \boldsymbol{S}_+ = \boldsymbol{E} + \boldsymbol{B}(\frac{\boldsymbol{A}^{-1} + \boldsymbol{A}^{-T}}{2})\boldsymbol{B}^T, \ \boldsymbol{S}_- = \boldsymbol{B}(\frac{\boldsymbol{A}^{-1} - \boldsymbol{A}^{-T}}{2})\boldsymbol{B}^T.$$

Cela signifie que $\boldsymbol{S}_+$ est la partie symétrique et $\boldsymbol{S}_-$ la partie anti-symétrique de $\boldsymbol{S}$. Pour estimer les solutions du problème généralisé non symétrique aux valeurs propres,

$$\boldsymbol{S}P = \lambda \boldsymbol{Q}P,$$

nous allons utiliser le théorème de Bendixson (voir [96]), qui nous donne les estimations suivantes :

$$\min_{P\in \mathbb{C}^m} \frac{(P, \boldsymbol{S}_+P)}{(P, \boldsymbol{Q}P)} \leq Re(\lambda) \leq \max_{P\in \mathbb{C}^m} \frac{(P, \boldsymbol{S}_+P)}{(P, \boldsymbol{Q}P)}$$
$$|Im(\lambda)| \leq \max_{P\in \mathbb{C}^m} \frac{|(P, i\boldsymbol{S}_-P)|}{(P, \boldsymbol{Q}P)}.$$

Dans la suite, nous estimerons les quotients de Rayleigh $\dfrac{(P, \boldsymbol{S}_+P)}{(P, \boldsymbol{Q}P)}$ et $\dfrac{(P, \boldsymbol{S}_-P)}{(P, \boldsymbol{Q}P)}$ pour $P \in \mathbb{C}^m$.

3.3.1 Résultats préliminaires

Démontrons d'abord que

$$\sup_{w_h\in V_h} \frac{b(w_h, p_h)}{\|w_h\|_V} = (P, \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^T P)^{1/2}. \tag{3.19}$$

En effet,

$$\sup_{w_h \in V_h} \frac{b(w_h, p_h)}{\|w_h\|_V} = \sup_{W \in \mathbb{R}^n} \frac{(\boldsymbol{B}W, P)}{(\boldsymbol{V}W, W)^{1/2}} = \sup_{W \in \mathbb{R}^n} \frac{(W, \boldsymbol{B}^T P)}{(\boldsymbol{V}^{1/2}W, \boldsymbol{V}^{1/2}W)^{1/2}}$$
$$= \sup_{\substack{Y = \boldsymbol{V}^{1/2}W \\ W \in \mathbb{R}^n}} \frac{(Y, \boldsymbol{V}^{-1/2}\boldsymbol{B}^T P)}{(Y, Y)^{1/2}} = (P, \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^T P)^{1/2},$$

puisque le maximum est atteint pour $Y = \boldsymbol{V}^{-1/2}\boldsymbol{B}^T P$. En utilisant la condition inf-sup généralisée, (3.8), nous en déduisons que

$$\gamma \|p_h\|_Q \leq \sup_{w_h \in V_h} \frac{b(w_h, p_h)}{\|w_h\|_V} + c(p_h, p_h)^{1/2},$$

dont la formulation matricielle est,

$$\gamma (P, \boldsymbol{Q}P)^{1/2} \leq (P, \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^T P)^{1/2} + (P, \boldsymbol{E}P)^{1/2},$$

c'est à dire

$$\frac{\gamma^2}{2}(P, \boldsymbol{Q}P) \leq \big(P, (\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^T + \boldsymbol{E})P\big). \tag{3.20}$$

Maintenant, nous allons étudier le spectre de $\boldsymbol{V}^{1/2}\boldsymbol{A}^{-1}\boldsymbol{V}^{1/2}$. Puisque

$$\forall u_h \in V_h, \quad \alpha(u_h, u_h)_V \leq a_+(u_h, u_h) \leq \mathcal{A}_+(u_h, u_h)_V,$$

nous avons

$$\forall U \in \mathbb{R}^n, \quad \alpha(U, \boldsymbol{V}U) \leq (U, \boldsymbol{A}_+ U) \leq \mathcal{A}_+(U, \boldsymbol{V}U).$$

Cela signifie que les valeurs propres de $\boldsymbol{V}^{-1/2}\boldsymbol{A}_+\boldsymbol{V}^{-1/2}$ sont contenues dans l'intervalle $[\alpha, \mathcal{A}_+]$, et dans ce cas, la matrice considérée est inversible et les valeurs propres de son inverse $\boldsymbol{V}^{1/2}\boldsymbol{A}_+^{-1}\boldsymbol{V}^{1/2}$ sont dans l'intervalle $[\mathcal{A}_+^{-1}, \alpha^{-1}]$. Cela revient à écrire

$$\forall U \in \mathbb{R}^n, \quad \mathcal{A}_+^{-1} \leq \frac{(U, \boldsymbol{A}_+^{-1}U)}{(U, \boldsymbol{V}^{-1}U)} \leq \alpha^{-1}. \tag{3.21}$$

Pour la forme bilinéaire anti-symétrique a_-, nous avons

$$\forall u_h, v_h \in V_h, \quad |a_-(u_h, v_h)| \leq \mathcal{A}_- \|u_h\|_V \|v_h\|_V.$$

Les valeurs propres de $\boldsymbol{V}^{-1/2}\boldsymbol{A}_-\boldsymbol{V}^{-1/2}$ sont imaginaires, et elles sont bornées en module par $\sup_{P \in \mathbb{C}^n} \left| \frac{(P, i\boldsymbol{A}_- P)}{(P, \boldsymbol{V}P)} \right|$. Or, pour tout $u_h, v_h \in V_h$, nous avons

$$|a_-(u_h + iv_h, u_h + iv_h)| = |2ia_-(u_h, v_h)| \leq 2\mathcal{A}_- \|u_h\|_V \|v_h\|_V$$
$$\leq \mathcal{A}_-(\|u_h\|_V^2 + \|v_h\|_V^2) = \mathcal{A}_-|(u_h + iv_h, u_h + iv_h)_V|,$$

donc les valeurs propres de $\boldsymbol{V}^{-1/2}\boldsymbol{A}_-\boldsymbol{V}^{-1/2}$ sont bornées en valeur absolue par $\mathcal{A}_-$. Cela implique que le spectre de $\boldsymbol{V}^{-1/2}\boldsymbol{A}\boldsymbol{V}^{-1/2}$ est contenu dans le rectangle $[\alpha, \mathcal{A}_+] \times [-\mathcal{A}_-, \mathcal{A}_-]$. En particulier, la matrice considérée est inversible. Soit $x+iy$ une valeur propre de $\boldsymbol{V}^{-1/2}\boldsymbol{A}\boldsymbol{V}^{-1/2}$. Alors la valeur propre correspondante $z := (x+iy)^{-1}$ de $\boldsymbol{V}^{1/2}\boldsymbol{A}^{-1}\boldsymbol{V}^{1/2}$ est donnée par

$$z = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}.$$

Comme $x \geq \alpha$, la partie imaginaire $\dfrac{y}{x^2+y^2}$ de z est bornée en valeur absolue par $2\alpha^{-1}$. La partie réelle $\frac{x}{x^2+y^2}$ de z est bornée inférieurement par α^{-1}, et bornée supérieurement par $\alpha(\mathcal{A}_+^2 + \mathcal{A}_-^2)^{-1}$. En utilisant la notation des quotients de Rayleigh, nous en déduisons les estimée suivantes pour les parties symétriques et anti-symétrique de $\boldsymbol{V}^{-1/2}\boldsymbol{A}\boldsymbol{V}^{-1/2}$:

$$\forall U \in \mathbb{C}^n, \quad \alpha(\mathcal{A}_+^2 + \mathcal{A}_-^2)^{-1} \leq \frac{(U, \dfrac{\boldsymbol{A}^{-1}+\boldsymbol{A}^{-T}}{2}U)}{(U, \boldsymbol{V}^{-1}U)} \leq \alpha^{-1}, \tag{3.22}$$

et

$$\forall U \in \mathbb{C}^n, \quad \left| \frac{(U, i\dfrac{\boldsymbol{A}^{-1}-\boldsymbol{A}^{-T}}{2}U)}{(U, \boldsymbol{V}^{-1}U)} \right| \leq \frac{1}{2\alpha}. \tag{3.23}$$

3.3.2 Estimation de la plus petite partie réelle des valeurs propres

Pour trouver la borne inférieure de $\frac{(P,\boldsymbol{S}_+P)}{(P,\boldsymbol{Q}P)}$, on doit d'abord remarquer que

$$\begin{aligned}
\frac{(P,\boldsymbol{S}_+P)}{(P,\boldsymbol{Q}P)} &= \frac{(P,\boldsymbol{E}P) + (P, \boldsymbol{B}\dfrac{\boldsymbol{A}^{-1}+\boldsymbol{A}^{-T}}{2}\boldsymbol{B}^TP)}{(P,\boldsymbol{Q}P)} \\
&= \frac{(P,\boldsymbol{E}P) + \dfrac{(P, \boldsymbol{B}\dfrac{\boldsymbol{A}^{-1}+\boldsymbol{A}^{-T}}{2}\boldsymbol{B}^TP)}{(P,\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^TP)} \times (P,\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^TP)}{(P,\boldsymbol{Q}P)} \\
&\geq \min\left(1, \inf_{P\in\mathbb{C}^m} \frac{(P, \boldsymbol{B}\dfrac{\boldsymbol{A}^{-1}+\boldsymbol{A}^{-T}}{2}\boldsymbol{B}^TP)}{(P,\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^TP)}\right) \frac{(P,(\boldsymbol{E}+\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^T)P)}{(P,\boldsymbol{Q}P)}.
\end{aligned}$$

L' inégalité (3.22) avec $U = \boldsymbol{B}^TP$ s'écrit

$$\inf_{P\in\mathbb{C}^m} \frac{(P, \boldsymbol{B}(\boldsymbol{A}^{-1}+\boldsymbol{A}^{-T})\boldsymbol{B}^TP)}{(P,\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^TP)} = \inf_{U=B^TP\in\mathbb{C}^n} \frac{(U,(\boldsymbol{A}^{-1}+\boldsymbol{A}^{-T})U)}{(U,\boldsymbol{V}^{-1}U)} \geq 2\alpha(\mathcal{A}_+^2+\mathcal{A}_-^2)^{-1},$$

d'où, en combinant avec (3.20), on obtient que

$$\frac{(P,\boldsymbol{S}_+P)}{(P,\boldsymbol{Q}P)} \geq \frac{\gamma^2}{2}\min(1, \alpha(\mathcal{A}_+^2+\mathcal{A}_-^2)^{-1}). \tag{3.24}$$

3.3.3 Estimation de la plus grande partie réelle des valeurs propres

Posons $U = \boldsymbol{A}^{-1}\boldsymbol{B}^T P$. Alors, en utilisant la coercivité de a et la continuité de b, nous obtenons

$$(U, \boldsymbol{V}U) \leq \alpha^{-1}(U, \boldsymbol{A}U) = \alpha^{-1}|(U, \boldsymbol{B}^T P)| \leq \alpha^{-1}\mathcal{B}(U, \boldsymbol{V}U)^{1/2}(P, \boldsymbol{Q}P)^{1/2}.$$

Par conséquent, $(U, \boldsymbol{V}U)^{1/2} \leq \frac{\mathcal{B}}{\alpha}(P, \boldsymbol{Q}P)^{1/2}$. Il est maintenant évident que

$$\left|(P, \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T P)\right| = |(P, \boldsymbol{B}U)| \leq \mathcal{B}(U, \boldsymbol{V}U)^{1/2}(P, \boldsymbol{Q}P)^{1/2} \leq \frac{\mathcal{B}^2}{\alpha}(P, \boldsymbol{Q}P),$$

d'où

$$\left|\frac{(P, (\boldsymbol{E} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T)P)}{(P, \boldsymbol{Q}P)}\right| \leq \mathcal{C} + \mathcal{B}^2\alpha^{-1}. \tag{3.25}$$

En procédant de manière similaire et en remplaçant $\boldsymbol{A}$ par $\boldsymbol{A}^T$, on obtient que

$$\left|\frac{(P, (\boldsymbol{E} + \boldsymbol{B}\boldsymbol{A}^{-T}\boldsymbol{B}^T)P)}{(P, \boldsymbol{Q}P)}\right| \leq \mathcal{C} + \mathcal{B}^2\alpha^{-1}. \tag{3.26}$$

On peut maintenant déduire de (3.25) et (3.26) que

$$\frac{(P, \boldsymbol{S}_+P)}{(P, \boldsymbol{Q}P)} \leq \mathcal{C} + \mathcal{B}^2\alpha^{-1}. \tag{3.27}$$

3.3.4 Estimation de la partie imaginaire des valeurs propres.

Par la continuité de b, et en posant $U = \boldsymbol{V}^{-1}\boldsymbol{B}^T P$, on a

$$|(\boldsymbol{B}^T P, U)| = |(P, \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^T P)| \leq \mathcal{B}(P, \boldsymbol{Q}P)^{1/2}(\boldsymbol{V}^{-1}\boldsymbol{B}^T P, \boldsymbol{B}^T P)^{1/2}.$$

De plus, on a

$$(P, \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^T P) \leq \mathcal{B}^2(P, \boldsymbol{Q}P), \tag{3.28}$$

et

$$\left|\frac{(P, i\boldsymbol{S}_-P)}{(P, \boldsymbol{Q}P)}\right| = \left|\frac{(P, i\boldsymbol{B}\dfrac{\boldsymbol{A}^{-1} - \boldsymbol{A}^{-T}}{2}\boldsymbol{B}^T P)}{(P, \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^{-T}P)}\right| \times \frac{(P, \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^{-T}P)}{(P, \boldsymbol{Q}P)}. \tag{3.29}$$

Par le changement de variable $U = \boldsymbol{B}^T P$, on remarque que

$$\left|\frac{(P, i\boldsymbol{B}\dfrac{\boldsymbol{A}^{-1} - \boldsymbol{A}^{-T}}{2}\boldsymbol{B}^T P)}{(P, \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^{-T}P)}\right| = \left|\frac{(U, i\dfrac{\boldsymbol{A}^{-1} - \boldsymbol{A}^{-T}}{2}U)}{(U, \boldsymbol{V}^{-1}U)}\right|.$$

En utilisant (3.23), (3.28) et (3.29), on peut maintenant conclure que

$$\left|\frac{(P, i\boldsymbol{S}_-P)}{(P, \boldsymbol{Q}P)}\right| \leq \frac{\mathcal{B}^2}{2\alpha}.$$

Chapitre 4

Application à Navier-Stokes 3D dans un cube

4.1 Un préconditionneur pour les équations de Navier-Stokes

4.1.1 Discrétisation des équations de Navier-Stokes

Nous nous intéressons à la discrétisation du système de Navier-Stokes stationnaire

$$\begin{aligned} -\triangle u + \mathcal{R}e\, u \cdot \nabla u + \nabla p &= f \qquad \text{in } \Omega, \\ \nabla \cdot u &= 0 \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega. \end{aligned} \tag{4.1}$$

Dans [90], Rebollo propose d'utiliser pour de telles équations des éléments finis de même ordre (en particulier des P1-P1), en ajoutant des termes de stabilisation adaptés. Nous allons d'abord décrire sa démarche.

Soit $\Omega \subset \mathbb{R}^3$ un ouvert polyédrique borné avec un bord Lipschitzien. On se place dans l'espace de Sobolev habituel $H_0^1(\Omega)$, et pour $p \in (1, \infty)$, l'espace $L^p(\Omega)$ usuel avec sa norme habituelle notée $\| \quad \|_p$. Posons $\boldsymbol{H}_0^1(\Omega) = \left(H_0^1(\Omega)\right)^3$ pour simplifier les notations. L'espace $\boldsymbol{H}_0^1(\Omega)$ est muni de sa norme usuelle, notée $| \quad |_1$, et définie par

$$\forall u \in \boldsymbol{H}_0^1(\Omega), \quad |\, u \,|_1 = \left(\int_\Omega |\nabla u|^2 \, dx \right)^{1/2}.$$

Soit $\boldsymbol{H}^{-1}(\Omega)$ l'espace dual de $\boldsymbol{H}_0^1(\Omega)$ et nous notons $(\cdot, \cdot)$ le produit scalaire de $L^2(\Omega)$. Nous définissons les formes bilinéaires b, d_1 et la forme trilinéaire d_2 en posant pour tout $u, v, w \in$

$\boldsymbol{H}_0^1(\Omega),\ q \in L^2(\Omega)$,

$$b(u,q) = \int_\Omega \nabla \cdot u\ q\ dx, \tag{4.2}$$

$$d_1(u,v) = \int_\Omega \nabla u \cdot \nabla v\, dx, \tag{4.3}$$

$$d_2(w,u,v) = \mathcal{R}e \int_\Omega (w \cdot \nabla) u \cdot v\ dx + \frac{\mathcal{R}e}{2} \int_\Omega (\nabla \cdot w)\, u \cdot v\ dx. \tag{4.4}$$

Le terme $\frac{1}{2}\mathcal{R}e \int_\Omega (\nabla \cdot w)\, u \cdot v\ dx$ est ajouté de manière à assurer l'anti-symétrie de $d_2(w,\,\cdot\,,\,\cdot\,)$ pour tout $w \in \boldsymbol{H}_0^1(\Omega)$.

Soit $h > 0$ et soit $\mathcal{T}_h$ une décomposition de Ω telle que tout élément $K \in \mathcal{T}_h$ est un tétraèdre dont le diamètre vérifie diam(K) $\leq$ h. Nous supposons de plus que toute les faces du tétraèdres $K \in \mathcal{T}_h$ sont, soit une face du bord $\partial\Omega$, soit une face d'un autre tétraèdre $K' \in \mathcal{T}_h$. On appelle ratio d'une triangulation le réel $\sigma \in (0,1)$ défini par

$$\sigma \leq \sup\{r > 0\ :\ B_r(x) \subset K \text{ pour } x \in K\}\,/\,h \quad \text{for } K \in \mathcal{T}_h,$$

où $B_r(x)$ représente la boule ouverte de centre x et de rayon r ($x \in \mathbb{R}^3, r > 0$). Les paramètres C_1, C_2, etc. représentent des constantes ne dépendant que de Ω et σ. Nous définissons les espaces d'éléments finis Y_h, V_h, Q_h par

$$Y_h = \{v_h \in \mathcal{C}^0(\overline{\Omega}) \text{ t. q. } v_{h|K} \in P_1, \forall K \in \mathcal{T}_h\},$$

$$V_h = \{v_h \in Y_h^3, v_{h|\partial\Omega} = 0\}, \quad Q_h = \{q_h \in Y_h, \int_\Omega q_h\, dx = 0\}.$$

Nous notons $\lambda_{1,K}, \cdots, \lambda_{4,K}$ les fonctions de base de l'espace d'éléments finis P1 sur $K \in \mathcal{T}_h$. On appelle fonction bulle relative à l'élément K la fonction définie par $\hat{b}_K = \lambda_{1,K} \cdots \lambda_{4,K}$. L'espace des fonctions bulles B_h est défini par

$$B_h = \left\{ v \in \mathcal{C}^0(\overline{\Omega}),\ \text{t. q.}\ \ v_{|K} \in \operatorname{span}(\hat{b}_K), \forall K \in \mathcal{T}_h \right\},$$

et on remarque que $B_h^3 \subset \boldsymbol{H}_0^1(\Omega)$. Soient $a_{1,h}, a_{2,h}$ deux formes bilinéaires sur $V_h \times V_h$, et nous supposons qu'il existe des constantes α_1, α_2, $B_{1,h}, B_2 > 0$ telle que,

$$\forall v_h, w_h \in V_h, \quad |a_{1,h}(v_h, w_h)| \leq B_{1,h}\,|\,v_h\,|_1\ |\,w_h\,|_1\,, \quad a_{1,h}(v_h, v_h) \geq \alpha_1\,|\,v_h\,|_1^2\,, \tag{4.5}$$

$$\forall v_h, w_h \in V_h, \quad |a_{2,h}(v_h, w_h)| \leq B_2\,|\,v_h\,|_1\ |\,w_h\,|_1\,, \quad a_{2,h}(v_h, v_h) \geq \alpha_2\,|\,v_h\,|_1^2\,. \tag{4.6}$$

La borne de continuité $B_{1,h}$ de la forme $a_{1,h}$ peut dépendre du paramètre h, tandis qu'il est essentiel de les constantes α_1, α_2 and B_2 en soient indépendantes. Nous définissons les opérateurs de Riesz $R_{1,h}, R_{2,h} : \boldsymbol{H}^{-1}(\Omega) \to \boldsymbol{H}_0^1(\Omega)$, appelés également opérateurs d'élimination

des fonctions bulles, de la façon suivante : pour $i \in \{1, 2\}, \varphi \in \boldsymbol{H}^{-1}(\Omega)$, $R_{i,h}\varphi$ est l'unique élément dans B_h^3 qui vérifie

$$a_{i,h}(R_{i,h}\varphi, v) = \varphi(v) \quad \forall v \in B_h^3;$$

On suppose que les produits scalaires $a_{i,h}$ sont de la forme

$$\forall i = 1, 2,\ a_{i,h}(u_h, v_h) = \sum_{K \in \mathcal{T}_h} C_{i,K}(\nabla u_h, \nabla v_h)_K,$$

où le symbole $(\ ,\ )_K$ représente le produit scalaire sur $L^2(K)$. Les opérateurs de Riesz peuvent alors s'écrire explicitement de la façon suivante : Si on note (e_1, e_2, e_3) la base canonique de $\mathbb{R}^3$, alors

$$\forall \varphi \in \boldsymbol{H}^{-1}(\Omega),\ R_{i,h}\varphi = \sum_{K \in \mathcal{T}_h} D_{i,K}^{-1} S_K(\varphi) \hat{b}_K$$

où

$$D_{i,K} = C_{i,K} \mid \hat{b}_K \mid_1^2 \quad \text{et} \quad S_K(\varphi) = \begin{pmatrix} \varphi(e_1 \hat{b}_K) \\ \varphi(e_2 \hat{b}_K) \\ \varphi(e_3 \hat{b}_K) \end{pmatrix}.$$

Soit $S : \boldsymbol{H}_0^1(\Omega) \times \boldsymbol{H}_0^1(\Omega) \to \boldsymbol{H}^{-1}(\Omega)$ défini par

$$S(u, v)(w) = d_2(u, v, w), \quad \forall u, v, w \in \boldsymbol{H}_0^1(\Omega). \tag{4.7}$$

Le problème variationnel proposé par Rebollo pour discrétiser les équations de Navier-Stokes (4.1) avec $f \in \boldsymbol{L^2}(\Omega)$, s'écrit

Trouver $(u_h, p_h) \in V_h \times Q_h$ tel que $\forall v_h \in V_h, \forall q_h \in Q_h$,

$$\begin{cases} d_1(u_h, v_h) + d_2(u_h, u_h, v_h) + a_{1,h}(R_{1,h}S(u_h, u_h), R_{1,h}S(u_h, v_h)) + b(v_h, p_h) = \displaystyle\int_\Omega f\, v_h\, dx, \\ b(u_h, q_h) - a_{2,h}(R_{2,h}(\nabla p_h), R_{2,h}(\nabla q_h)) = 0. \end{cases} \tag{4.8}$$

Quand on applique au problème (4.8) une méthode non-linéaire telle une méthode du point-fixe (Picard ou certaines méthodes de quasi-Newton), chaque itération revient à la résolution d'un problème linéaire du type

Trouver $(u_h, p_h) \in V_h \times Q_h$ tel que $\forall v_h \in V_h, \forall q_h \in Q_h$,

$$\begin{cases} d_1(u_h, v_h) + d_2(w_h, u_h, v_h) + a_{1,h}(R_{1,h}S(w_h, u_h), R_{1,h}S(w_h, v_h)) + b(v_h, p_h) = \varrho(v_h) \\ b(u_h, q_h) - a_{2,h}(R_{2,h}(\nabla p_h), R_{2,h}(\nabla q_h)) = \kappa(q_h), \end{cases} \tag{4.9}$$

où $w_h \in V_h, \varrho \in V_h', \kappa \in Q_h'$ sont déterminés par l'itération précédente de la méthode non-linéaire. Nous fixons w_h, ϱ, κ pour le reste du chapitre. Dans le cas des itérations de Picard, nous supposons naturellement que $\varrho = f$, et $\kappa = 0$.

Nous allons montrer que le problème variationnel (4.9) satisfait les hypothèses que nous avons formulées dans le paragraphe 3.1 pour le problème variationnel (3.7). Nous supposons que l'espace V correspond à l'espace $\boldsymbol{H}_0^1(\Omega)$ et pour Q nous choisissons l'espace des fonctions de $L^2(\Omega)$ qui sont à moyenne nulle. La forme b est définie en (4.2), tandis que c et a sont données par :

$$c(p_h, q_h) = a_{2,h}(R_{2,h}(\nabla p_h), R_{2,h}(\nabla q_h)), \quad \forall p_h, q_h \in Q_h,$$

$$a(u_h, v_h) = d_1(u_h, v_h) + d_2(w_h, u_h, v_h) + a_{1,h}(R_{1,h}S(w_h, u_h), R_{1,h}S(w_h, v_h)), \quad \forall u_h, v_h \in V_h.$$

Nous devons vérifier que les formes bilinéaires a, b et c vérifient les hypothèses (3.2), (3.3), (3.4), (3.5), (3.6), c'est à dire qu'elles sont uniformément continues sur leurs espaces respectifs, que a est coercive et que c est positive. D'abord, comme conséquence directe de l'inégalité de Hölder, nous obtenons l'inégalité suivante :

$$\forall u, v \in \boldsymbol{H}_0^1(\Omega), \quad |(w_h \cdot \nabla u, v)| \leq \| w_h \|_6 \; | u |_1 \; \| v \|_3 \, .$$

Les injections de Sobolev (voir [98]) nous donnent

$$\| v \|_3 \leq C_1 \, | v |_1 \, , \quad \| v \|_6 \leq C_2 \, | v |_1 \quad \forall v \in \boldsymbol{H}_0^1(\Omega), \tag{4.10}$$

alors, pour S défini par (4.7), on en déduit que

$$\forall u, v, w \in \boldsymbol{H}_0^1(\Omega), \quad |S(u, v)(w)| \leq C_2 \, | u |_1 \; | v |_1 \; \| w \|_3 \, . \tag{4.11}$$

De plus, en utilisant localement les injections de Sobolev sur chaque tétraèdre (voir le lemme [90, Lemma 4.1]), nous obtenons que :

$$\forall \hat{w}_h \in B_h^3, \;\; \| \hat{w}_h \|_3 \leq C_3 h^{1/2} \, | \hat{w}_h |_1 \, .$$

Nous pouvons donc en déduire une borne supérieure pour $a_{1,h}(R_{1,h}S(w_h, u_h), R_{1,h}S(w_h, v_h))$. En effet, d'après l'inégalité précédente, la définition de S en (4.7), et sa borne de continuité (4.11), nous obtenons :

$$\begin{aligned}
\forall v_h \in V_h, \quad & a_{1,h}(R_{1,h}S(w_h, u_h), R_{1,h}S(w_h, v_h)) = S(w_h, u_h)(R_{1,h}S(w_h, v_h)) \\
& \leq C_2 \mathcal{R}e \; | w_h |_1 \; | u_h |_1 \; \| R_{1,h}S(w_h, v_h) \|_3 \\
& \leq C_3 C_2 \mathcal{R}e \; | w_h |_1 \; | u_h |_1 \, h^{1/2} \, | R_{1,h}S(w_h, v_h) |_1 \\
& \leq C_3 C_2 \mathcal{R}e \; \alpha_1^{-1/2} h^{1/2} \, | w_h |_1 \; | u_h |_1 \, a_{1,h}(R_{1,h}S(w_h, v_h), R_{1,h}S(w_h, v_h))^{1/2}.
\end{aligned}$$

En prenant $v_h = u_h$ dans l'inégalité ci dessus, nous obtenons une borne supérieure pour $a_{1,h}(R_{1,h}S(w_h, v_h), R_{1,h}S(w_h, v_h))^{1/2}$. En appliquant ce résultat dans l'inéquation obtenue juste avant, nous pouvons conclure que

$$\forall u_h, v_h \in V_h, \quad a_{1,h}(R_{1,h}S(w_h, u_h), R_{1,h}S(w_h, v_h)) \leq C_4 \mathcal{R}e^2 \alpha_1^{-1} h \mid w_h \mid_1^2 \mid u_h \mid_1 \mid v_h \mid_1 .$$

Finalement, l'inégalité (3.2) est satisfaite avec

$$\mathcal{A} = 1 + C_4 \mathcal{R}e^2 h \alpha_1^{-1} \mid w_h \mid_1^2 + \mathcal{R}e\, C_1 \mid w_h \mid_1 . \tag{4.12}$$

De manière évidente, l'inégalité (3.4) est obtenue avec $\mathcal{B} = 1$. La borne de continuité de c s'obtient en procédant comme pour $a_{1,h}(R_{1,h}S(w_h, v_h), R_{1,h}S(w_h, v_h))^{1/2}$. En effet, pour tout $p_h, q_h \in Q_h$, on a

$$\begin{aligned} c(p_h, q_h) &= a_{2,h}(R_{2,h}(\nabla p_h), R_{2,h}(\nabla q_h)) = \langle \nabla p_h, R_{2,h}(\nabla q_h) \rangle \\ &\leq |(p_h, \nabla \cdot R_{2,h}(\nabla q_h))| \\ &\leq \| p_h \|_2 \mid R_{2,h}(\nabla q_h) \mid_1 . \end{aligned}$$

En utilisant la coercivité de $a_{2,h}$, on trouve que

$$\mid R_{2,h}(\nabla q_h) \mid_1^2 \leq \frac{1}{\alpha_2} |a_{2,h}(R_{2,h}(\nabla q_h), R_{2,h}(\nabla q_h))| \leq \frac{1}{\alpha_2} \| q_h \|_2 \mid R_{2,h}(\nabla q_h) \mid_1 ,$$

d'où

$$\forall q_h \in Q_h, \quad \mid R_{2,h}(\nabla q_h) \mid_1 \leq \frac{1}{\alpha_2} \| q_h \|_2 .$$

Nous pouvons donc choisir $\mathcal{C} = 1/\alpha_2$ dans (3.6). Comme

$$a(v_h, v_h) = \int_\Omega |\nabla v_h|^2 dx + a_{1,h}(R_{1,h}S(w_h, v_h), R_{1,h}S(w_h, v_h)) \geq |v_h|_1^2,$$

par (4.5), nous en déduisons que l'inégalité (3.3) est satisfaite pour $\alpha = 1$.

4.1.2 Condition inf-sup généralisée

Lemme 4.1 *Pour tout $p_h \in Q_h$, nous avons*

$$\sup_{z_h \in V_h} \frac{(\nabla \cdot z_h, p_h)}{\mid z_h \mid_1} + c(p_h, p_h)^{1/2} \geq \gamma \| p_h \|_2 ,$$

où la constante γ dépend seulement de Ω et du ratio σ.

Preuve Grâce aux propriétés de la triangulation $\mathcal{T}_h$, le couple d'éléments finis $(V_h \oplus B_h^3,\, Q_h)$, appelé mini-élément, satisfait la condition inf-sup discrète suivante [49] :

$$\forall p_h \in Q_h, \quad \sup_{\substack{z_h \in V_h \\ \tilde{z}_h \in B_h^3}} \frac{(\nabla \cdot (z_h + \tilde{z}_h), p_h)}{|\, z_h + \tilde{z}_h \,|_1} \geq C_5 \,\|\, p_h \,\|_2 \,. \tag{4.13}$$

Comme $\nabla p \in \boldsymbol{H}^{-1}(\Omega)$, et en utilisant la définition de l'opérateur de Riesz $R_{2,h}$ associé au produit scalaire $a_{2,h}$, on obtient

$$(\nabla \cdot \tilde{z}_h, p_h) = -\langle \nabla p_h, \tilde{z}_h \rangle = -a_{2,h}(R_{2h}(\nabla p_h), \tilde{z}_h).$$

De plus, comme $\tilde{z}_{h|K} \in H_0^1(K) \forall K \in \mathcal{T}_h$, on a $(\nabla v_h, \nabla \tilde{z}_h) = 0 \forall v_h \in V_h$. On peut alors conclure que

$$|\, z_h \,|_1 \,+\, |\, \tilde{z}_h \,|_1 \,\leq \sqrt{2} \;|\, z_h + \tilde{z}_h \,|_1 \,.$$

La relation précédente et l'inégalité (4.13) impliquent que

$$\begin{aligned} C_5 \,\|\, p_h \,\|_2 \;&\leq \sup_{\substack{z_h \in V_h \\ \tilde{z}_h \in B_h^3}} \frac{(\nabla \cdot (z_h + \tilde{z}_h), p_h)}{|\, z_h + \tilde{z}_h \,|_1} \leq \sup_{\substack{z_h \in V_h \\ \tilde{z}_h \in B_h^3}} \frac{(\nabla \cdot z_h, p_h)}{|\, z_h + \tilde{z}_h \,|_1} + \sup_{\substack{z_h \in V_h \\ \tilde{z}_h \in B_h^3}} \frac{(\nabla \cdot \tilde{z}_h, p_h)}{|\, z_h + \tilde{z}_h \,|_1} \\ &\leq \sqrt{2} \sup_{z_h \in V_h} \frac{(\nabla \cdot z_h, p_h)}{|\, z_h \,|_1} + \sqrt{2} \sup_{\tilde{z}_h \in B_h^3} \frac{(\nabla \cdot \tilde{z}_h, p_h)}{|\, \tilde{z}_h \,|_1}. \end{aligned}$$

De plus, d'après les hypothèses décrites en (4.6), on obtient

$$\begin{aligned} \sup_{\tilde{z}_h \in B_h^3} \frac{|(\nabla \cdot \tilde{z}_h, p_h)|}{|\, \tilde{z}_h \,|_1} &= \sup_{\tilde{z}_h \in B_h^3} \frac{\langle \nabla p_h, \tilde{z}_h \rangle}{|\, \tilde{z}_h \,|_1} = \sup_{\tilde{z}_h \in B_h^3} \frac{a_{2,h}(R_{2,h}(\nabla p_h), \tilde{z}_h)}{|\, \tilde{z}_h \,|_1} \\ &\leq B_2 \;|R_{2,h}(\nabla p_h)|_1 = \frac{B_2}{\sqrt{\alpha_2}} a_{2,h}(R_{2,h}(\nabla p_h), R_{2,h}(\nabla p_h))^{1/2}, \end{aligned}$$

d'où

$$\sup_{\tilde{z}_h \in B_h^3} \frac{|(\nabla \cdot \tilde{z}_h, p_h)|}{|\, \tilde{z}_h \,|_1} \leq \frac{B_2}{\sqrt{\alpha_2}} c(p_h, p_h)^{1/2}.$$

Le lemme 1 est donc vérifié avec

$$\gamma = \frac{C_5}{\sqrt{2} \max(1, \dfrac{B_2}{\sqrt{\alpha_2}})}. \quad \square$$

Comme les hypothèses décrites dans la partie 3.1 sont satisfaites, alors on peut appliquer le théorème 3.1. On en déduit l'existence et l'unicité des solutions du problème variationnel discret (4.9).

4.1.3 Influence du préconditionneur

Nous allons d'abord étudier les éventuelles dépendances en $\mathcal{R}e$ des constantes α_1, α_2 et B_2 définies en (4.5) et (4.6). Pour cela, nous allons préciser quels sont ces termes de stabilisation définis par Rebollo [90]. Les relations données par Rebollo ont dû être un peu modifiées car nous avons adimensionné des équations de Navier-Stokes. Commençons par la stabilisation en vitesse. Nous introduisons deux constantes numériques A et P et nous fixons un nombre $p \in (3, \infty)$. Pour tout $K \in \mathcal{T}_h$, on appelle nombre de Péclet le nombre Pe_K défini par

$$Pe_K = \mathcal{R}e \ \| w_{h|K} \|_p \, diam\, K.$$

On définit $\tau_{1,K}$ par

$$\tau_{1,K} = A \frac{diam\, K}{\mathcal{R}e \, \| w_{h|K} \|_p} \min(P, Pe_K),$$

et pour tout v_h, z_h,

$$a_{1,h}(u_h, v_h) = \sum_{K \in T_h} \frac{|K|}{840^2 \tau_{1,K} \, | \hat{b}_K |_1^2} (\nabla v_h, \nabla z_h)_K. \tag{4.14}$$

D'après Rebollo [90, Lemma 5.2], un tel produit scalaire $a_{1,h}$ est coercif et sa constante de coercivité α_1 est indépendante de $\mathcal{R}e$. Cela signifie que le terme de stabilisation en vitesse s'écrit explicitement de la façon suivante :

$$a_{1,h}(R_{1,h} S(w_h, u_h), R_{1,h} S(w_h, v_h)) = \mathcal{R}e^2 \sum_{K \in T_h} \tau_{1,K} (\bar{w}_K \cdot \nabla u_h, \bar{w}_K \cdot \nabla v_h)_K, \tag{4.15}$$

où $\bar{w}_K = \int_K w_h \, \hat{b}_K \, dx \, / \int_K \hat{b}_K \, dx$ pour tout $K \in \mathcal{T}_h$. Pour $a_{2,h}$, Rebollo propose de choisir

$$\forall v_h, z_h, \in Q_h, \quad a_{2,h}(v_h, z_h) = (\nabla v_h, \nabla z_h). \tag{4.16}$$

Cela signifie que $\alpha_2 = B_2 = 1$. L'élimination des fonctions bulles à travers cette forme bilinéaire nous donne la formulation explicite suivante pour la stabilisation en pression

$$a_{2,h}(R_{2,h} \nabla p_h, R_{2,h} \nabla q_h) = \sum_{K \in T_h} \frac{|K|}{840^2 \, | \hat{b}_K |_1^2} (\nabla p_h, \nabla q_h)_K.$$

Ce terme correspond exactement au bloc pression de l'élimination des fonctions bulles de Franca et Frey [48].

Maintenant, nous voulons étudier l'influence de $\mathcal{R}e$ et h sur le spectre du problème préconditionné proposé dans la partie 3.2. Avec notre choix de la forme bilinéaire a, on en déduit

que les formes a_+ et a_- introduites en (5.12) et (5.13), respectivement, sont données par

$$\forall u_h, v_h \in V_h, \quad a_+(u_h, v_h) = \int_\Omega \nabla u_h \cdot \nabla v_h \, dx + a_{1,h}(R_{1,h}(w_h \cdot \nabla u_h), R_{1,h}(w_h \cdot \nabla v_h)),$$
$$\forall u_h, v_h \in V_h, \quad a_-(u_h, v_h) = \mathcal{R}e \int_\Omega (w_h \cdot \nabla) \, u_h \cdot v_h \, dx + \frac{1}{2}\mathcal{R}e \int_\Omega (\nabla \cdot w_h) \, u_h \cdot v_h \, dx.$$

En utilisant les mêmes arguments que pour calculer $\mathcal{A}$ en (4.12), nous pouvons voir que les inégalités (3.17) et (3.18) sont satisfaites avec

$$\mathcal{A}_+ = 1 + C_4 \mathcal{R}e^2 h \alpha_1^{-1} \, | \, w_h \, |_1^2 \, , \quad \mathcal{A}_- = \mathcal{R}e \, C_1 \, | \, w_h \, |_1 \, .$$

En utilisant le théorème 3.2, on peut conclure que les valeurs propres du problème préconditionné (2.6) sont contenues dans le rectangle

$$\left[\frac{\gamma^2}{2} \min \Big(1, \big((1 + C_4 \mathcal{R}e^2 h \alpha_1^{-1} \, | \, w_h \, |_1^2 \,)^2 + \mathcal{R}e^2 C_1^2 \, | \, w_h \, |_1^2 \,) \big)^{-1} \Big) , (1 + 1/\alpha_2) \right] \quad \times \quad \left[-\frac{1}{2}, \frac{1}{2} \right] . \tag{4.17}$$

Quand on choisit les itérées de Picard comme méthode non-linéaire, la fonction w_h correspond à la partie vitesse v_h de la solution de (4.9) issue de la dernière itération non-linéaire. De plus, comme indiqué ci dessus, la fonctionnelle κ dans (4.9) est nulle et $\varrho = f$. On en déduit grâce au théorème 3.1 que l'inégalité (3.3) est satisfaite pour $\alpha = 1$ dans le cas des itérées de Picard, et on obtient que $| \, u_h \, |_1 \leq \| \, f \, \|_{-1}$. Cette remarque nous permet de supposer que $| \, w_h \, |_1 \leq \| \, f \, \|_{-1}$. Par conséquent, en utilisant le théorème 3.2 et les majorations de valeurs propres données par (4.17), on peut conclure qu'à chaque itération de Picard, les valeurs propres du problème préconditionné (2.6) sont incluses dans la domaine rectangulaire

$$\left[\frac{\gamma^2}{2} \min \Big(1, \big((1 + C_4 \, \| \, f \, \|_{-1} \, \mathcal{R}e^2 h \alpha_1^{-1})^2 + \mathcal{R}e^2 C_1^2 \, \| \, f \, \|_{-1}^2 \,) \big)^{-1} \Big) , (1 + 1/\alpha_2) \right] \times \left[-\frac{1}{2}, \frac{1}{2} \right] . \tag{4.18}$$

Comme seules les petites valeurs de h sont intéressantes et parce que $\operatorname{diam} K \leq \operatorname{diam} \Omega$ pour tout $K \in \mathcal{T}_h$, on peut supposer que $h \leq h_0$ pour un $h_0 > 0$ donné. Il s'ensuit qu'à chaque itération de Picard, les valeurs propres du système matriciel préconditionné (2.6) sont contenues dans le rectangle donné par (4.18), mais où h est remplacé par h_0. On obtient bien que les valeurs extrémales de ce rectangle ne dépendent pas de h. Introduisons une constante η ne dépendant que de $\| \, f \, \|_{-1}$, alors, pour des valeurs raisonnables de h et $\mathcal{R}e$, les valeurs propres considérées sont simplement contenues dans le rectangle

$$[C_6 \eta \mathcal{R}e^{-4} h^{-2}, C_7] \times [-C_8, C_8]. \tag{4.19}$$

De plus, si on omet la stabilisation en vitesse, les valeurs propres appartiennent au domaine rectangulaire

$$[C_9\eta\mathcal{R}e^{-2}, C_7] \times [-C_8, C_8]. \tag{4.20}$$

Notre préconditionneur ne permet bien sûr pas d'éliminer la dépendance en $\mathcal{R}e$ des valeurs propres, mais nous obtenons une dépendance du même ordre que celle de Silvester pour les éléments finis stables [41].

Nous choisirons numériquement la méthode de quasi-Newton proposée par Turek [103]. Dans ce cas, on n'a pas de majoration pour $|w_h|_1$ qui soit indépendante de h. Cependant, nous verrons dans les tests numériques que cela ne pose pas de problème.

4.2 Résultats numériques

4.2.1 Présentation des tests

Nous allons tester notre préconditionneur dans deux situations. Pour les deux cas, le domaine est

$$\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^3.$$

a) Nous choisissons comme premier test une généralisation du test d'Oden. Ce test est décrit dans [57] et [85] : La convection est égale à la fonction polynomiale w définie par :

$$w_x = \frac{1}{4}\frac{\partial^2\Phi}{\partial y\partial z}, \quad w_y = \frac{1}{4}\frac{\partial^2\Phi}{\partial x\partial z}, \quad \text{et} \quad w_z = \frac{-1}{2}\frac{\partial^2\Phi}{\partial x\partial y}, \tag{4.21}$$

où

$$\Phi(x,y,z) = (x-\frac{1}{2})^2(y-\frac{1}{2})^2(z-\frac{1}{2})^2.$$

Par construction, nous avons $\nabla\cdot w = 0$ et w satisfait les conditions de Dirichlet homogènes. La pression est déterminée par

$$p(x,y,z) = xyz.$$

Soit f défini par $f = -\triangle w + \mathcal{R}e\ (w\cdot\nabla)w + \nabla p$. Nous sommes intéressés par la résolution du système d'Oseen (2.4) d'inconnues u et p. Nous considérons le problème de point-selle généralisé (2.6) qui est la version algébrique du problème (2.4). Nous estimerons les valeurs propres extrémales du système préconditionné et étudierons les performances de notre préconditionneur à l'intérieur d'itérations du FGMRES.

b) Le second test est le test de référence appelé la 'cavité entraînée'. Les équations associées à ce test sont les équations de Navier-Stokes avec un terme source nul et satisfaisant les conditions de Dirichlet suivantes :

$$u(x,y,z) = (1,0,0)^T \quad \text{sur la face } z = \tfrac{1}{2},$$
$$u(x,y,z) = (0,0,0)^T \quad \text{sur les autres faces du cube.}$$

Par convention, nous imposerons que $p(0,0,0) = 0$. Les solutions de références ont été établies par Jiang [62] qui a résolu la cavité 3D d'inconnues vitesse, pression et vorticité. Dans la littérature, les tests ont souvent été poussé jusqu'à des nombres de Reynolds voisins de 1000. Les valeurs à partir desquelles il n'existe plus de solution stationnaire sont encore en discussion : les tests physiques menés dans [72] montrent qu'il y a des instabilité pour $\mathcal{R}e = 3200$, mais les auteurs pensent que des instabilités peuvent apparaître même pour des nombres de Reynolds de l'ordre de 2000. Nous ne dépasserons pas 200 dans nos tests : nous sommes donc sûrs qu'il existe une solution stationnaire à notre problème.

4.2.2 Calcul des valeurs propres

Nous allons calculer les valeurs propres de $\boldsymbol{S} = \boldsymbol{E} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$ et de $\boldsymbol{Q}^{-1}\boldsymbol{S}$ et les comparer pour différentes valeurs de h et $\mathcal{R}e$ dans le cas où la matrice $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix}$ est la forme algébrique de la formulation variationnelle du système d'Oseen (4.9). Il est bien évident que nous ne pouvons pas construire la matrice du complément de Schur. Nous devons utiliser une méthode itérative pour calculer les valeurs propres extrêmes. Nous avons choisi la méthode IRA ("Implicit Restarted Arnoldi") introduite par Lehoucq et Sorensen [74]. C'est l'algorithme qui est implanté dans la fonction EIGS de matlab. On note $(e_j)_{1\leq j\leq n}$ la base canonique de $\mathbb{R}^n$. Supposons que les fonctions `qr()` et `schur()` calculent la factorisation QR et la décomposition de Schur. Ces fonctions sont implantées dans LAPACK (nous fournirons plus de détails sur ces fonctions dans l'annexe B).

1. données initiales : f_0 ; tol, iter=0, maxiter, ritz= $[1, \cdots, 1]$;

2. tant que (iter<maxiter et $\|\text{ritz}\| > \text{tol}$)

3. $k_0 = 0$

(a) On calcule $m-k_0$ étapes de l'itération d'Arnoldi partant de l'étape k_0 et de vecteur initial f_{k_0}. On obtient $\boldsymbol{V}_m$ tel que $\boldsymbol{V}_m^T\boldsymbol{V}_m = \boldsymbol{I}_m$, $\boldsymbol{H}_m$ une matrice d'Hessenberg supérieure et f_m tel que

$$\boldsymbol{S}\boldsymbol{V}_m = \boldsymbol{V}_m\boldsymbol{H}_m + f_m e_m^T.$$

(b) On calcule des valeurs propres de $\boldsymbol{H}_m$ notées $(\lambda_j)_{1\leq j\leq m}$ et on sélectionne les $m-k$ valeurs propres non voulues de $\boldsymbol{H}_m$. (Si on cherche les valeurs propres de plus petite partie réelle, on sélectionne les $m-k$ valeurs propres ayant la partie réelle la plus grande.) Pour chaque valeur propre λ_j, on normalise le vecteur propre de façon à ce que sa plus grande composante (en module) soit réelle et on range cette valeur dans ritz(j).

(c) $\boldsymbol{Q}' = \boldsymbol{I}_m$

(d) pour $j = k, \cdots, m$

i. Si λ_j est réelle, on calcule $\boldsymbol{Q}'_j$ et $\boldsymbol{R}_j$ tel que $[\boldsymbol{Q}'_j, \boldsymbol{R}_j] = \texttt{qr}(H_m - \lambda_j I_m)$.
Sinon $[\boldsymbol{Q}'_j, \boldsymbol{R}_j] = \texttt{qr}((\boldsymbol{H}_m - \Re(\lambda_j)\boldsymbol{I}_m)^2 + \Im(\lambda_j)^2\boldsymbol{I}_m)$.

ii. $\boldsymbol{H}_m \leftarrow \boldsymbol{Q}'^T_j\boldsymbol{H}_m\boldsymbol{Q}'_j$ et $\boldsymbol{Q}' \leftarrow \boldsymbol{Q}'\boldsymbol{Q}'_j$.

fin pour

(e) ritz $\leftarrow$ ritz$(1:k)\|f_k\|$; $\quad \beta_k = \boldsymbol{H}_m(k+1,k)$; $\quad \sigma_k = \boldsymbol{Q}'(m,k)$.

(f) $f_k \leftarrow \beta_k\boldsymbol{V}(k+1,:) + \sigma_k f_m$.

(g) $\boldsymbol{V}(:,k) \leftarrow \boldsymbol{V}(:,m)\boldsymbol{Q}'(:,1:k); \boldsymbol{H}_k \leftarrow \boldsymbol{H}_m(1:k,1:k)$

(h) $k_0 = k$.

4. fin tant que.
Les k valeurs propres approchées de $\boldsymbol{S}$ selon le critère de sélection de l'étape 3b sont les valeurs propres de $\boldsymbol{H}_k$.

5. $[\boldsymbol{Q}', \boldsymbol{R}] = \texttt{schur}(\boldsymbol{H}_k)$; $\boldsymbol{V} \leftarrow \boldsymbol{V}(:,1:k)\boldsymbol{Q}'$. Le résidu est donné par la norme de $\boldsymbol{S}\boldsymbol{V} - \boldsymbol{V}\boldsymbol{R}$.

A l'intérieur de cet algorithme, nous inversons $\boldsymbol{A}$ par un GMRES préconditionné par ILU(0), et $\boldsymbol{Q}$ est inversé par un CG préconditionné par la diagonale de $\boldsymbol{Q}$.

Nous construisons d'abord la matrice d'Oseen correspondant au cas test a). Le terme convectif (4.21) est très lisse et ne correspond pas du tout à une solution physique, mais

ce test nous permet d'augmenter le nombre de Reynolds sans tenir compte des contraintes physiques. Cela nous permettra de tester le comportement des valeurs propres en fonction du nombre de Reynolds. Les tableaux 4.1 et 4.2 montrent les parties réelles des valeurs propres de plus grande partie réelle et de plus petite partie réelle respectivement. Les tests sont d'abord réalisés sans préconditionnement, puis avec. Dans le tableau 4.1, nous n'assemblons pas la stabilisation en vitesse, mais seulement le terme de stabilisation en pression correspondant à l'élimination des fonctions bulles défini en (5.41). Le tableau 4.2 contient les extrema de la partie réelle des valeurs propres quand, cette fois-ci, on ajoute la stabilisation en vitesse (de type SUPG) décrite en (5.41). Ensuite, nous étudions les valeurs propres de la matrice

		$\boldsymbol{E}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$		$\boldsymbol{Q}^{-1}(\boldsymbol{E}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T)$	
h	$\mathcal{R}e$	min($\Re(\lambda)$)	max($\Re(\lambda)$)	min($\Re(\lambda)$)	max($\Re(\lambda)$)
1/8	10000	3.5e-6	9.2e-3	0.0536	2.408
	50000	3.58e-5	5.3e-3	0.0536	1.793
	100000	3.57e-5	4.3e-3	0.0536	1.611
	200000	3.53e-5	3.5e-3	0.0536	1.442
1/16	10000	6.9e-6	1.3e-3	0.0536	2.384
	50000	6.2e-6	8.7e-4	0.0536	2.059
	100000	6.5e-6	7.6e-4	0.0536	1.978
	200000	6.5e-6	6.7e-4	0.0536	1.8939
1/32	10000	1.31e-6	1.75e-4	0.0536	2.396
	50000	1.22e-6	1.34e-4	0.0536	2.222
	100000	1.41 e-6	1.25e-4	0.0536	2.18
	200000	1.06e-6	1.1e-4	0.0536	2.139
1/64	10000			0.0536	2.41285
	50000			0.0536	2.31317
	100000			0.0536	2.29249

TAB. 4.1 – Valeurs propres pour la matrice d'Oden. Pas de stabilisation en vitesse.

d'Oseen pour le cas test physique b). Dans le cas de la cavité entraînée pour les valeurs de Reynolds que nous avons choisies, il n'a pas été nécessaire d'ajouter la stabilisation en vitesse pour obtenir la convergence. Nous prenons comme direction de convection w_h la solution obtenue à la dernière itération non-linéaire quand la méthode non-linéaire a convergé vers

			$\boldsymbol{E}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$		$\boldsymbol{Q}^{-1}(\boldsymbol{E}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T)$	
h	$\mathcal{R}e$	$\overline{Pe}_K$	min($\Re(\lambda)$)	max($\Re(\lambda)$)	min($\Re(\lambda)$)	max($\Re(\lambda)$)
1/8	10000	0.03	3.5716e-5	0.00877	0.0537	2.3389
	50000	0.15	3.4578e-5	0.00537	0.0537	1.8005
	100000	0.3	2.2332e-5	0.00420	0.0537	1.5993
	200000	0.6	2.8735e-5	0.00307	0.0393	1.3851
1/16	10000	0.004	4.6593e-6	0.00126	0.0537	2.3464
	50000	0.02	5.0141e-6	0.000895	0.0537	2.0598
	100000	0.04	4.6432e-6	0.000777	0.03041	1.9807
	200000	0.08	4.3475e-6	0.000656	0.02311	1.3957

TAB. 4.2 – Valeurs propres pour la matrice d'Oden. Stabilisation en vitesse de paramètres $A = P = 1$.

		$\boldsymbol{E}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$		$\boldsymbol{Q}^{-1}(\boldsymbol{E}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T)$	
h	$\mathcal{R}e$	min($\Re(\lambda)$)	max($\Re(\lambda)$)	min($\Re(\lambda)$)	max($\Re(\lambda)$)
1/8	1	0.000287	0.023	0.0535	0.803
	10	0.000287	0.022	0.0535	0.7939
	50	0.000287	0.021	0.0535	0.7855
	100	0.000287	0.020	0.0535	0.7791
	200	0.000287	0.019	0.0535	0.7768
1/16	1	3.08e-05	0.0044	0.0535	0.9699
	10	3.08e-05	0.0041	0.0535	0.9315
	50	3.08e-05	0.0037	0.0535	0.8713
	100	3.07e-05	0.0032	0.0535	0.815
	200	3.056e-05	0.0027	0.0535	0.7583

TAB. 4.3 – Valeurs propres pour la dernière itération non-linéaire dans la cavité entraînée. Pas de stabilisation en vitesse.

une solution stationnaire. Cette fois-ci, w_h est une bonne approximation d'un écoulement de Navier-Stokes. Par contre, nous ne pouvons considérer que des petites valeurs du nombre de Reynolds car les itérations non-linéaires doivent avoir convergées au préalable en un temps raisonnable. Les résultats sont présentés dans le tableau 4.3.

Nous observons que, quand nous n'avons pas assemblé la stabilisation en vitesse, les extrema de la partie réelle des valeurs propres sont indépendantes de la taille de la maille h conformément au théorème 3.2. Ce n'est bien sûr pas le cas pour le problème non préconditionné. Cependant, d'après l'estimation théorique (4.20), nous aurions dû observer une décroissance en fonction de Reynolds pour la plus petite partie réelle des valeurs propres. Il y a plusieurs raisons à cela. Quand on augmente le nombre de Reynolds, on observe une concentration de plus en plus importante des valeurs propres autour de la valeur propre de plus petite partie réelle. Or, la méthode IRA peine à trouver les valeurs propres de grande multiplicité. L'autre raison est donnée par Elman et al dans [42] : ils ont observé que quand ils augmentaient le nombre de Reynolds sans raffiner, ils obtenaient des valeurs propres artificiellement petites. Cependant, si nous anticipons et regardons les résultats sur la performance du préconditionneur dans le paragraphe 4.2.4, cela mène à penser que les valeurs propres du problème préconditionné restent dans une boîte de taille stable même sur des grilles fines. Quand on ajoute la stabilisation en vitesse, on observe cette fois-ci la décroissance de la partie réelle de la plus petite valeur propre de partie réelle quand h diminue. Cependant, cette décroissance est plus faible que celle prédite par la théorie, mais nous avions déjà constaté ce phénomène sans le terme de stabilisation en vitesse. Il est donc naturel qu'il en soit de même ici.

4.2.3 Performance du préconditionneur.

Nous allons vérifier dans ce paragraphe l'efficacité de notre préconditionneur pour la matrice par bloc obtenue quand on discrétise le système d'Oseen pour le cas test d'Oden 3D. Dans ce cas, nous devons résoudre uniquement un système linéaire. Nous appliquons notre préconditionneur à l'intérieur des itérations externes du FGMRES. Le FGMRES est une variante du GMRES introduite par Saad [91] qui est prévue pour des préconditionneurs qui varient à chaque itération. Typiquement, c'est le cas des méthodes itératives. Or, comme nous utilisons une méthode itérative pour préconditionner $\boldsymbol{A}$, le GMRES pourrait ne pas converger. Inverser $\left(\begin{smallmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{Q} \end{smallmatrix}\right)$ nécessite d'inverser $\boldsymbol{A}$ et $\boldsymbol{Q}$. Dans nos simulations, la matrice non-symétrique $\boldsymbol{A}$ est inversée par un GMRES préconditionné par ILU(0). Pour alléger les calculs, nous remplaçons $\boldsymbol{Q}$ par la matrice diagonale $\widehat{\boldsymbol{Q}}$ définie par $\widehat{\boldsymbol{Q}}(i,i) = \sum_{j=1}^{m} \boldsymbol{Q}(i,j)$ pour $i = 1, \cdots, J$ (autrement dit, $\widehat{\boldsymbol{Q}}$ est le "mass lumping" de $\boldsymbol{Q}$). Nous remarquons dans le tableau 4.4 que préconditionner dans un FGMRES revient à peu près à préconditionner à gauche dans un

GMRES. C'est la raison pour laquelle nous devons choisir le préconditionneur triangulaire inférieur.

h	1/2	1/4	1/8	1/16
Préconditionnement à gauche dans un GMRES	19	40	76	85
Préconditionnement triang inf dans FGMRES	16	41	70	84
Préconditionnement triang sup dans FGMRES	39	143	225	237

TAB. 4.4 – Nombre d'itération du GMRES/FGMRES

Nous traçons dans la figure 4.1 la norme euclidienne du résidu relatif pour des itérations du FGMRES préconditionné et non préconditionné.

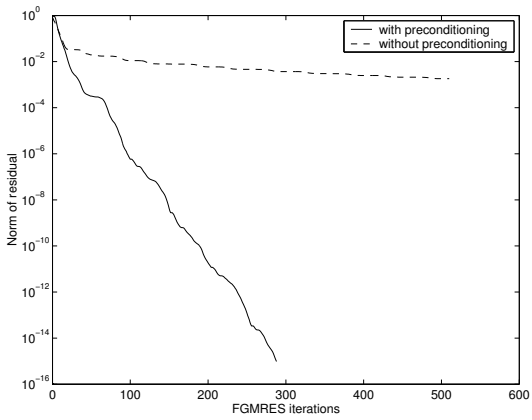


FIG. 4.1 – Cas test a), résidu à chaque itération du FGMRES(30), $\mathcal{R}e = 100, h = 1/32$.

Nous remarquons que la méthode FGMRES ne converge pas si nous ne préconditionnons pas. Il est en effet connu que les méthodes de types GMRES sont extrêmement robustes à condition qu'elles soient associées à un préconditionnement un minimum efficace. Nous observons également que notre préconditionneur décroît à peu près linéairement jusqu'à des valeurs du résidu très petites. Nous n'observons aucun phénomène de stagnation.

Nous présentons dans le tableau 4.5 le nombre d'itérations nécessaires à la résolution du système d'Oseen pour plusieurs valeurs de h et de $\mathcal{R}e$. Nous choisissons comme critère d'arrêt d'avoir la norme du résidu relatif plus petite de 10^{-6}. Nous appliquons un FGMRES avec redémarrage toutes les 30 itérations et partons d'une solution initiale toujours nulle. Les résultats de Zhi Hao Cao [26] sur la convergence du GMRES préconditionné montrent que le taux de décroissance du résidu dépend seulement des valeurs propres extrémales du

problème préconditionné. Les résultats numériques du tableau 4.5 confirment une nouvelle fois les résultats du théorème 3.2, et plus particulièrement, que le taux de convergence de notre système préconditionné à l'intérieur d'un FGMRES est indépendant de h et $\mathcal{R}e$ au moins pour de petites valeurs du nombre de Reynolds $\mathcal{R}e$.

h	1/8	1/16	1/32	1/64
$\mathcal{R}e = 1$	59	100	111	97
$\mathcal{R}e = 10$	64	101	111	97
$\mathcal{R}e = 50$	69	101	112	97
$\mathcal{R}e = 100$	70	101	113	97

TAB. 4.5 – Nombre d'itérations du FGMRES préconditionné à droite, cas test a).

4.2.4 Itérations non-linéaires

Les itérées de Picard convergent trop lentement pour être utilisées dans des problèmes comme celui de la cavité 3D. Nous préférons utiliser la méthode de quasi-Newton introduite par Turek [103] appelée 'adaptive fixed point defect correction method.' Une itération de cette méthode est constituée des 5 étapes suivantes :

On suppose U_{n-1}, ω_{n-2} et P_{n-1} donnés et on procède le la façon suivante :

1. On calcule le résidu non-linéaire
$$\begin{pmatrix} R_{n-1} \\ S_{n-1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}(U_{n-1})U_{n-1} + \boldsymbol{B}P_{n-1} - F \\ \boldsymbol{B}^T U_{n-1} + \boldsymbol{E}P_{n-1} - G \end{pmatrix}. \tag{4.22}$$
2. On résout
$$\begin{pmatrix} \boldsymbol{A}(U_{n-1}) & \boldsymbol{B} \\ \boldsymbol{B}^T & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} W \\ Z \end{pmatrix} = \begin{pmatrix} R_{n-1} \\ S_{n-1} \end{pmatrix}.$$
3. On assemble la matrice $\mathbf{M}_{n-1} = \boldsymbol{A}(U_{n-1} - \omega_{n-2}W)$.
4. On calcule le paramètre optimal ω_{n-1} par la formule
$$\omega_{n-1} = \frac{\left(\begin{pmatrix} \mathbf{M}_{n-1} & \boldsymbol{B} \\ \boldsymbol{B}^T & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} W \\ Z \end{pmatrix}, \begin{pmatrix} F \\ G \end{pmatrix} - \begin{pmatrix} \mathbf{M}_{n-1} & \boldsymbol{B} \\ \boldsymbol{B}^T & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} U_{n-1} \\ P_{n-1} \end{pmatrix} \right)_2}{\left(\begin{pmatrix} \mathbf{M}_{n-1} & \boldsymbol{B} \\ \boldsymbol{B}^T & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} W \\ Z \end{pmatrix}, \begin{pmatrix} \mathbf{M}_{n-1} & \boldsymbol{B} \\ \boldsymbol{B}^T & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} W \\ Z \end{pmatrix} \right)_2}.$$
où $(\ ,\ )_2$ est le produit scalaire euclidien.

5. On calcule U_n, P_n par

$$\begin{pmatrix} U_n \\ P_n \end{pmatrix} = \begin{pmatrix} U_{n-1} \\ P_{n-1} \end{pmatrix} - \omega_{n-1} \begin{pmatrix} W \\ Z \end{pmatrix}.$$

La norme du résidu non-linéaire dépend de la discrétisation, et donc de la finesse du maillage. On note $\boldsymbol{Q}_{ml,h}$ la matrice diagonale obtenue par un "mass lumping" de la matrice de masse sur l'espace vitesse-pression. On prendra comme résidu

$$res = \boldsymbol{Q}_{ml,h}^{-1} \begin{pmatrix} R_{n-1} \\ S_{n-1} \end{pmatrix}$$

où $\begin{pmatrix} R_{n-1} \\ S_{n-1} \end{pmatrix}$ est défini dans (4.22). On arrêtera les itérations lorsque :

$$\|res\| < \mathrm{tol}_{nonlin}. \tag{4.23}$$

Les itérations linéaires internes sont résolues par un FGMRES comme décrit dans le paragraphe 4.2.3. Nous arrêtons la résolution quand la norme euclidienne du résidu relatif est plus petite que 10^{-1}. Choisir une tolérance plus petite ne réduirait pas efficacement le nombre d'itérations non-linéaires mais ne ferait qu'augmenter inutilement le temps de calcul de la méthode globale. Nous arrêtons les itérations non-linéaires quand la norme L^2 du résidu au sens des élément finis (voir annexe A.1 ou [103]) est plus petite que 10^{-4}. Nous démarrons de la solution initiale $u = 0, p = 0$ pour résoudre les équations de Navier-Stokes avec $\mathcal{R}e = 1$. Ensuite, nous récupérons ce résultat pour résoudre le problème avec $\mathcal{R}e = 10$, puis le résultat de $\mathcal{R}e = 10$ pour résoudre $\mathcal{R}e = 100$ et ainsi de suite. Tous ces tests ont été réalisés sur la grille de taille $32 \times 32 \times 32$.

Dans la figure 4.2, nous traçons le résidu relatif du FGMRES préconditionné à chaque itération non-linéaire. La figure 4.3(a) montre que la norme Euclidienne du résidu non-linéaire $\binom{R_{n-1}}{S_{n-1}}$ et la figure 4.3(b) montre le nombre d'itérations du FGMRES réalisées à chaque itération non-linéaire. On pourrait penser d'après la figure 4.2, qu'il faut plus d'itérations du FGMRES pour les dernières itérées non-linéaires mais la figure 4.3(b) indique que ce nombre reste borné. On voit nettement que les itérations non-linéaires convergent très lentement vers la solution stationnaire (voir 4.3(a).) Il n'existe pas de méthode "miracle" qui donne rapidement la solution : c'est pourquoi on considère souvent la solution stationnaire comme limite à l'équilibre des solutions évolutives.

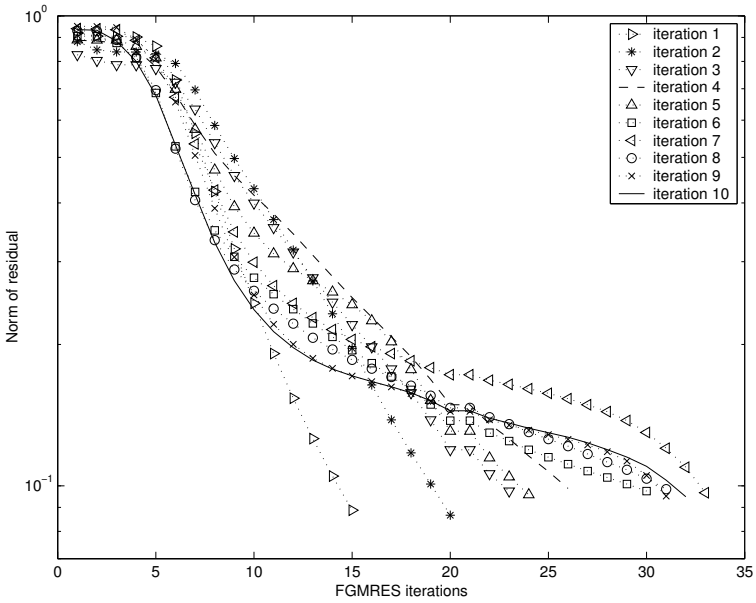
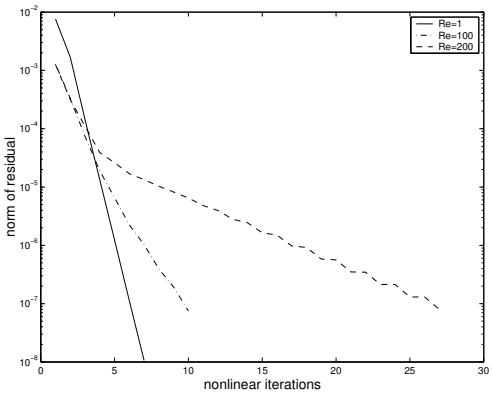
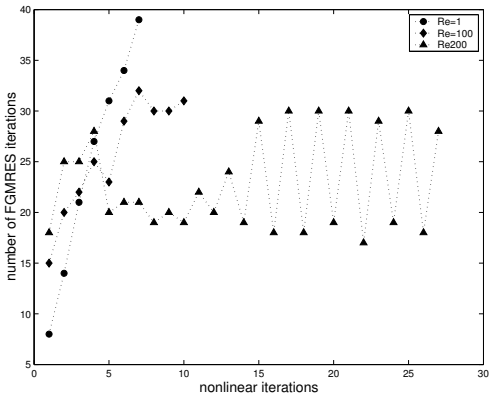


FIG. 4.2 – Résidu à l'intérieur des itérations internes du FGMRES pour chaque itération non-linéaire , Re=100 et h=1/32.



(a) Résidu non-linéaire



(b) Nombre d'itérations internes du FGMRES

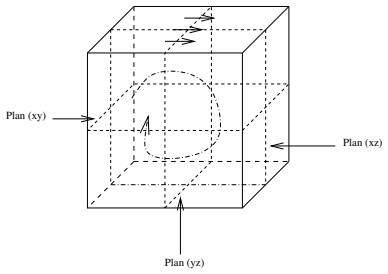
FIG. 4.3 – Cavité entraînée 3D sur la grille de taille $32 \times 32 \times 32$.

h	1/8	1/16	1/32	1/64
$\mathcal{R}e = 10$	5-*13*	5-*19*	5-*22*	5-*23*
$\mathcal{R}e = 100$	9-*12*	10-*19*	10-*25*	12-*28*
$\mathcal{R}e = 200$	10-*13*	24-*18*	27-*22*	48-*26*

TAB. 4.6 – Nombre d'itérations non-linéaires (en roman) et nombre moyen d'itérations internes du FGMRES (en italique).

4.2.5 Solutions de la cavité 3D

Afin d'étudier qualitativement notre méthode de résolution, nous traçons le champ de vecteurs et les iso-valeurs de la pression et vorticité de notre fluide dans les plans xy, xz et yz correspondant aux équations $z = 0, y = 0$ et $x = 0$.



En comparant nos résultats avec les solutions de références obtenues dans [62] et [93], on observe que nos résultats sont comparables, excepté pour la vorticité où apparaissent quelques différences au voisinage du bord de la cavité pour les plans yz et xy. Rappelons que Jiang et al avaient considéré la vorticité comme une inconnue de leur problème de Navier-Stokes alors que dans notre cas, nous devons la déduire de notre champs de vitesse. Ces différences semblent dues au fait que la grille n'est pas assez fine. Nous avons utilisé un maillage régulier, et non pas un maillage de type Tchebyshev (qui a des mailles plus petites quand on est proche du bord) comme c'est souvent le cas pour cette simulation. De plus, quand on trace le profil de la première composante de la vitesse (voir la figure 4.4) pour plusieurs valeurs de h, on s'aperçoit que notre profil se rapproche de celui de Jiang [62] quand h diminue mais que nous n'avons pas encore assez de précision. Olshanskii [86] a obtenu des écarts semblables pour des éléments finis en 2D appliqués à Navier-Stokes en vitesse-pression. Nous arrêtons nos calculs à une grille de taille $64 \times 64 \times 64$ (environ 1 million d'inconnues) à cause des limitations de mémoire vive de nos moyens informatiques.

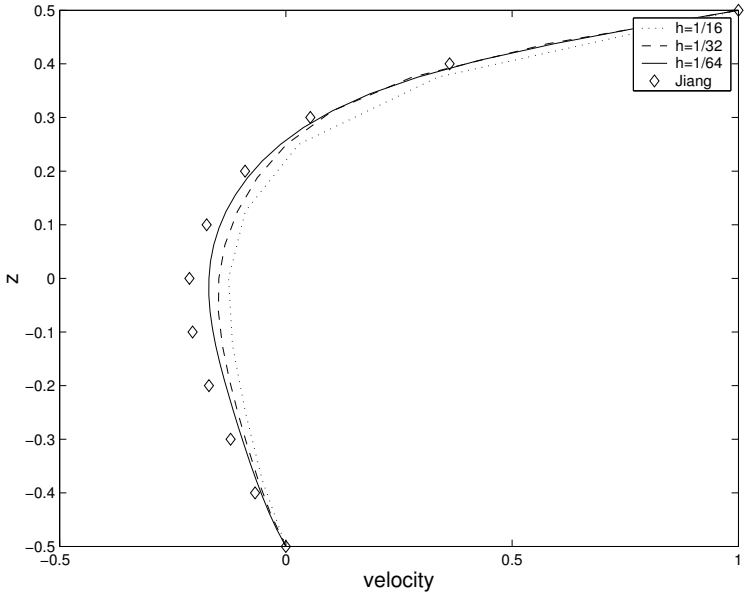
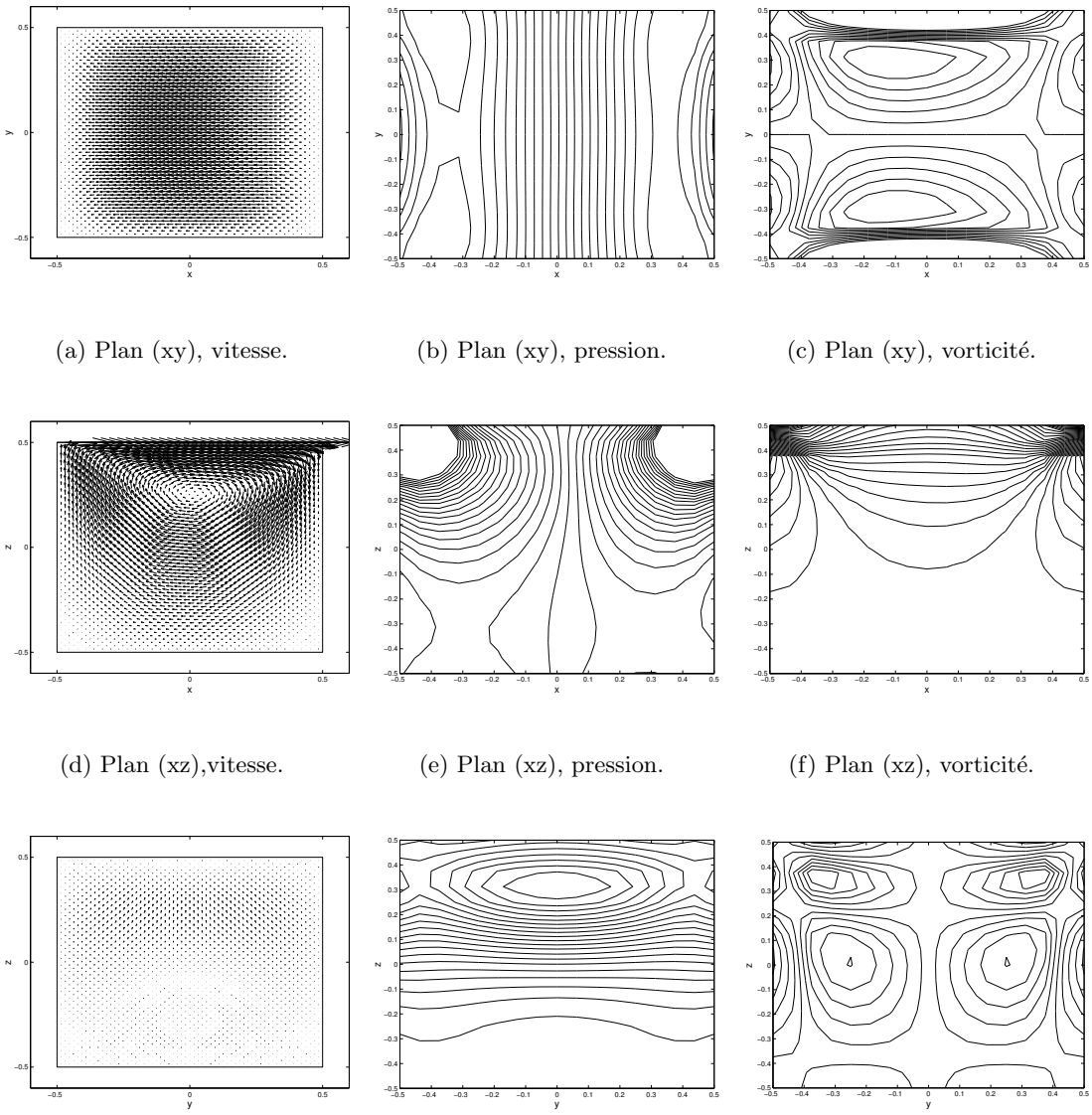


FIG. 4.4 – Profil de la première composante de la vitesse, $\mathcal{R}e = 100$.

4.3 Conclusion

Nous pouvons déjà tirer quelques conclusions sur l'utilisation de ce préconditionneur pour résoudre les problèmes d'écoulement de Navier-Stokes en 3D discrétisés par des éléments finis non-stables. Nous avons vérifié que notre préconditionneur par blocs est spectralement équivalent à la matrice de point-selle obtenue par discrétisation au sens de Rebollo [90] des équations d'Oseen, ce qui signifie que le conditionnement du problème préconditionné est indépendant de h. Cela revient à dire que le taux de convergence du GMRES ne dépend pas du nombre d'inconnues. De plus, le dépendance en Reynolds n'intervient pas pour les petites valeurs du nombre de Reynolds permettant d'obtenir une solution stationnaire. Ces résultats sont au moins aussi bons que ceux connus dans la littérature pour des éléments finis stables. Comme les $P1 - P1$ nécessitent beaucoup moins d'inconnues, cette méthode semble prometteuse pour être appliquée sur des grilles beaucoup plus fines en ajoutant de la parallélisation dans le code.

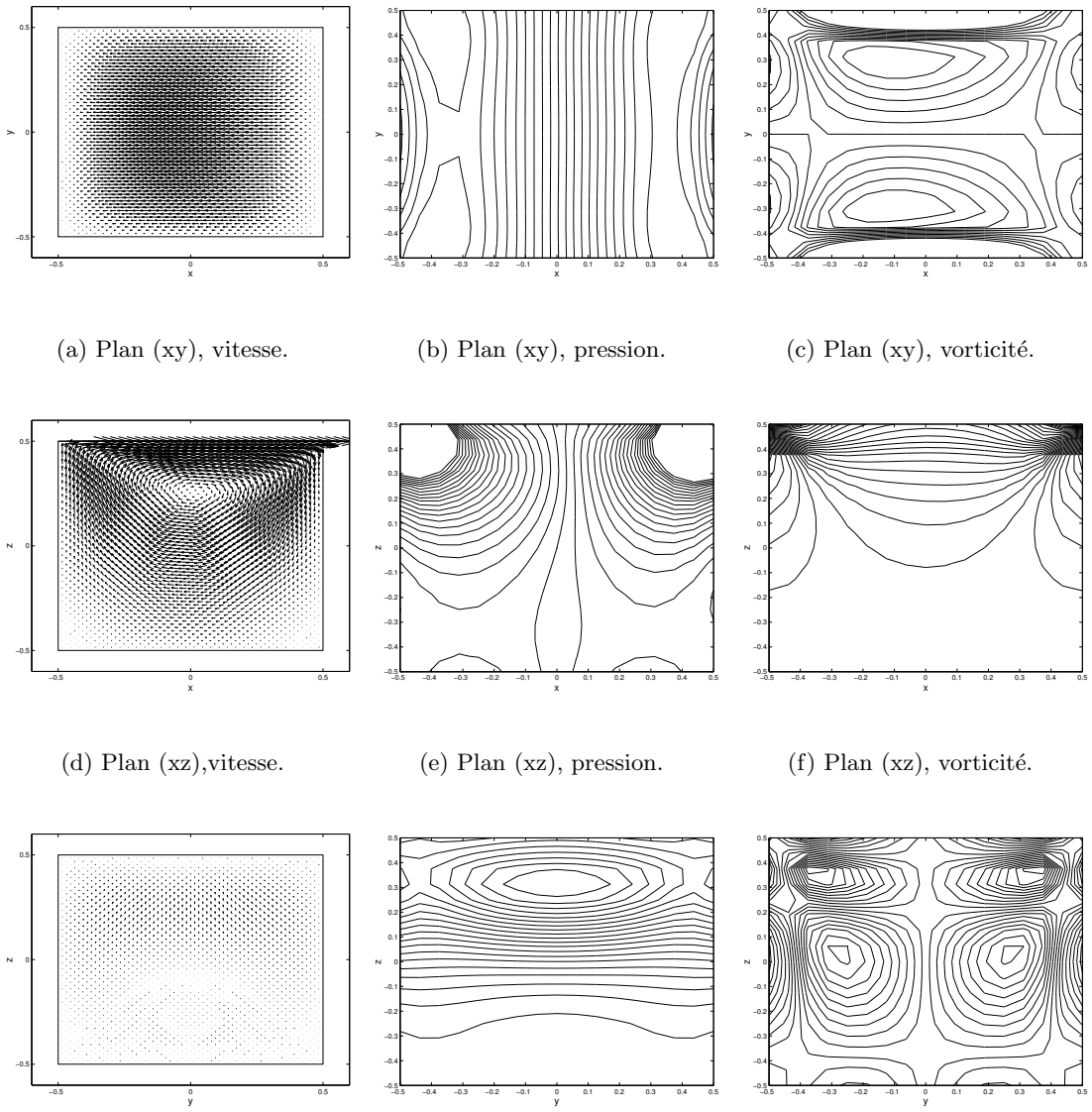


(a) Plan (xy), vitesse. (b) Plan (xy), pression. (c) Plan (xy), vorticité.

(d) Plan (xz),vitesse. (e) Plan (xz), pression. (f) Plan (xz), vorticité.

(g) Plan (yz), vitesse. (h) Plan (yz), pression. (i) Plan (yz), vorticité.

FIG. 4.5 – Cavité entraînée 3D, Re=1, grille de taille $64 \times 64 \times 64$.

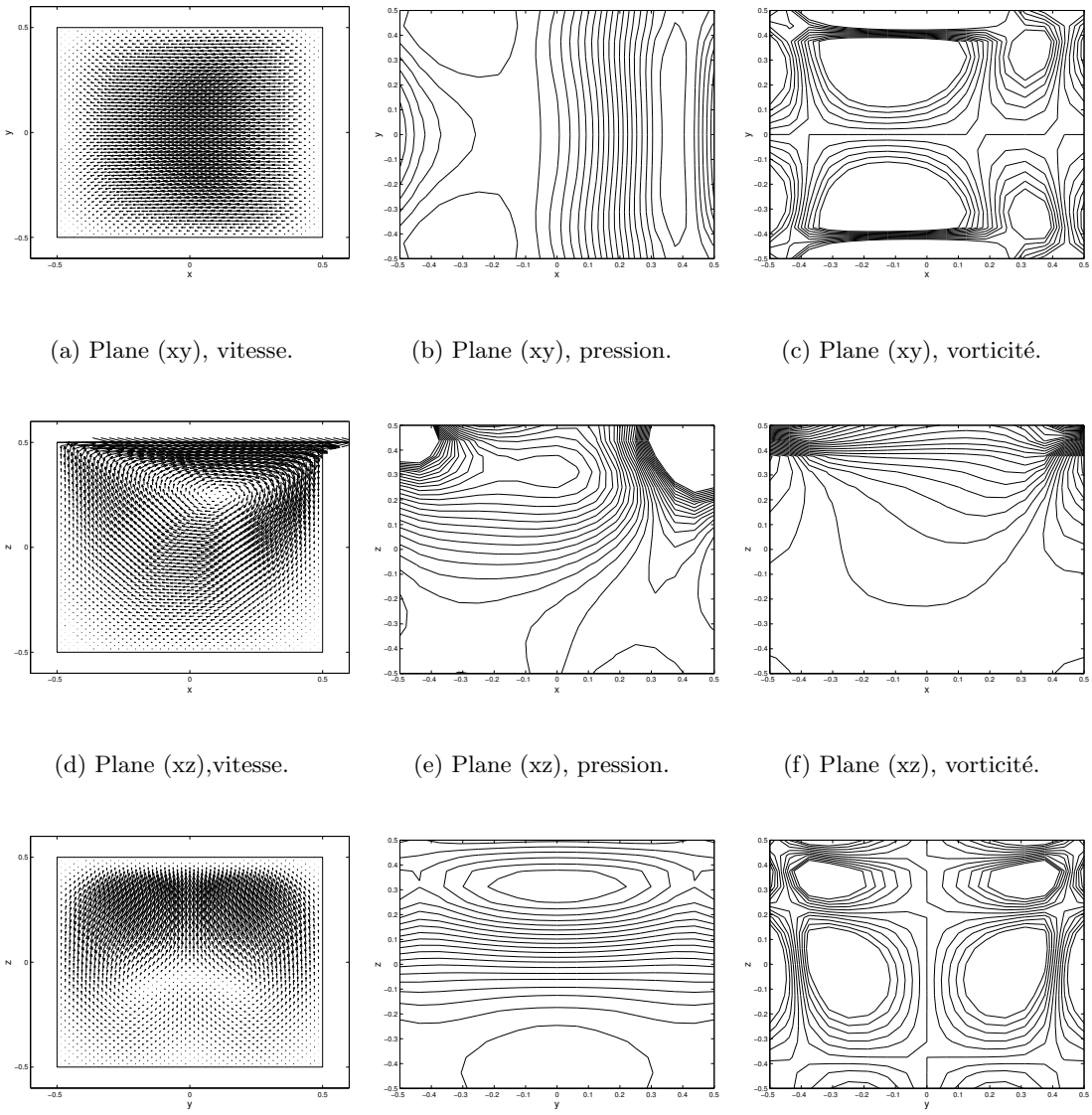


(a) Plan (xy), vitesse. (b) Plan (xy), pression. (c) Plan (xy), vorticité.

(d) Plan (xz),vitesse. (e) Plan (xz), pression. (f) Plan (xz), vorticité.

(g) Plan (yz), vitesse. (h) Plan (yz), pression. (i) Plan (yz), vorticité.

FIG. 4.6 – Cavité entraînée 3D, Re=10, grille de taille $64 \times 64 \times 64$.

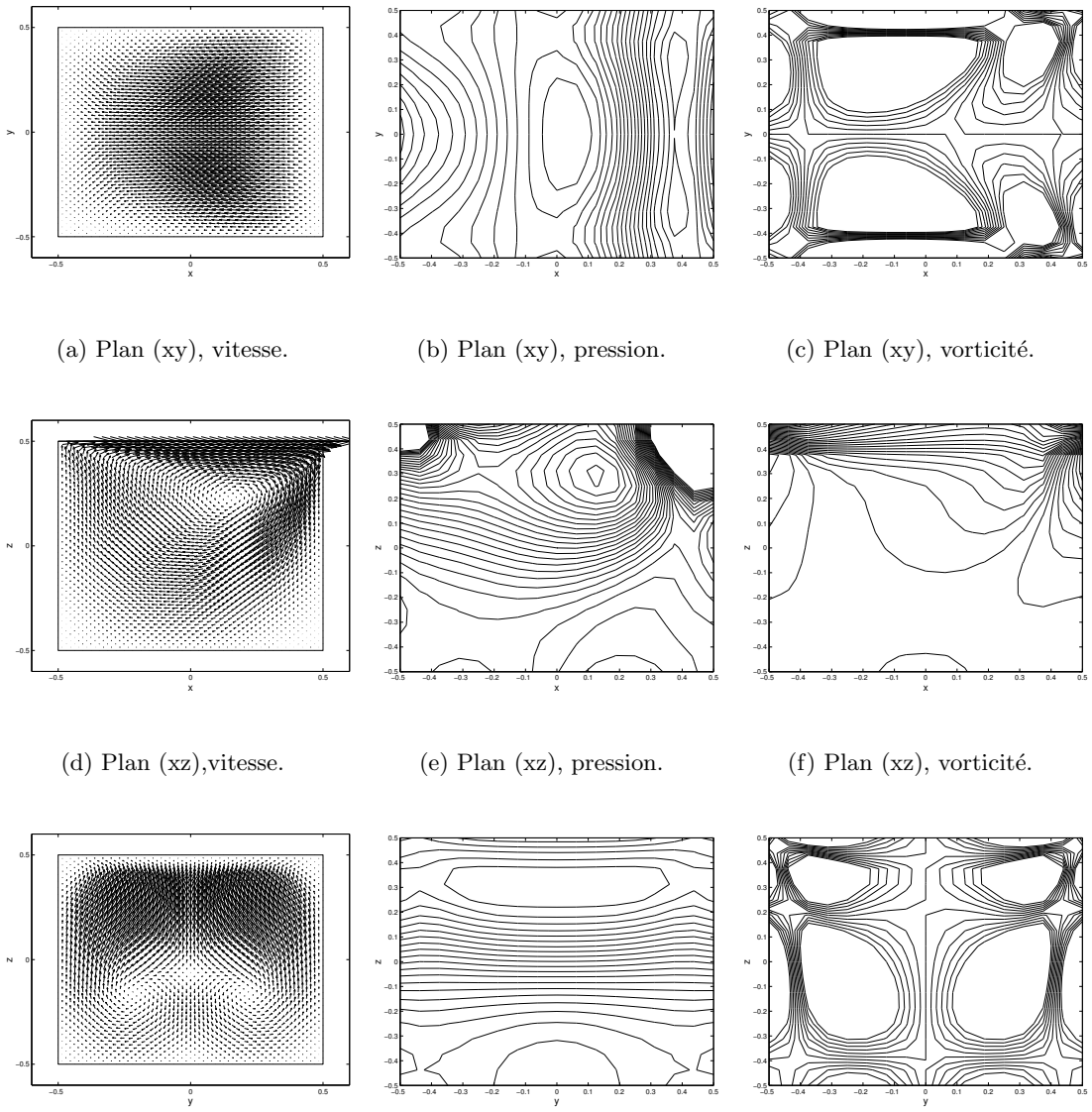


(a) Plane (xy), vitesse. (b) Plane (xy), pression. (c) Plane (xy), vorticité.

(d) Plane (xz),vitesse. (e) Plane (xz), pression. (f) Plane (xz), vorticité.

(g) Plane (yz), vitesse. (h) Plane (yz), pression. (i) Plane (yz), vorticité.

FIG. 4.7 – Cavité entraînée 3D, Re=100, grille de taille $64 \times 64 \times 64$.



(a) Plan (xy), vitesse. (b) Plan (xy), pression. (c) Plan (xy), vorticité.

(d) Plan (xz),vitesse. (e) Plan (xz), pression. (f) Plan (xz), vorticité.

(g) Plan (yz), vitesse. (h) Plan (yz), pression. (i) Plan (yz), vorticité.

FIG. 4.8 – Cavité entraînée 3D, Re=200, grille de taille $64 \times 64 \times 64$.

Chapitre 5

Application à Navier-Stokes 3D dans un domaine extérieur

5.1 Modèle et discrétisation

Nous considérons l'écoulement d'un corps solide dans un fluide. Le corps est représenté par l'ouvert borné polyédrique $\hat{\Omega} \subset \mathbb{R}^3$. Nous prenons comme origine le barycentre de notre solide et nous négligeons les mouvements de rotation du solide. Nous pouvons alors simplement modéliser l'écoulement de notre fluide par les équations de Navier-Stokes avec des conditions de Dirichlet homogènes sur le bord de l'obstacle $\partial\hat{\Omega}$ et une vitesse non nulle à l'infini. Ces équations s'écrivent de la façon suivante :

$$\left\{ \begin{array}{l} -\nu\Delta v \ + \ (v\cdot\nabla)v \ + \ \nabla\pi \ = \ \Phi, \quad \operatorname{div} v \ = \ 0 \quad \text{dans}\, \hat{\Omega}^c = \mathbb{R}^3 \backslash \overline{\hat{\Omega}}, \\ \quad v \,|\, \partial\hat{\Omega} = 0, \quad v(x) \to \kappa\cdot e_1 \ \ (|x| \to \infty), \end{array} \right. \tag{5.1}$$

où $e_1 = (1,0,0)$, ν est la viscosité cinématique et $\kappa \in (0,\infty)$ est la vitesse à l'infini. Nous supposons que le terme source $\Phi : \mathbb{R}^3\backslash\overline{U} \mapsto \mathbb{R}^3$ est donné. Nous choisissons comme vitesse caractéristique L le rayon de l'obstacle et nous prenons comme vitesse de référence la vitesse à l'infini κ. Posons $v(x) = \kappa\tilde{u}(\tilde{x}) = \kappa\tilde{u}(x/L)$. En adimensionnant l'équation (5.1), nous obtenons le système suivant :

$$\left\{ \begin{array}{ll} -\Delta\tilde{u} \ + \ \mathcal{R}e\,(\tilde{u}\cdot\nabla)\tilde{u} \ + \ \nabla\tilde{p} \ = \ \tilde{\Phi} & \text{dans}\, \Omega^c \\ \operatorname{div}\tilde{u} \ = \ 0 & \text{dans}\, \Omega^c, \\ \tilde{u} \,|\, \partial\Omega = 0, \quad \tilde{u}(\tilde{x}) \to e_1 \ \ (|\tilde{x}| \to \infty) & \end{array} \right. \tag{5.2}$$

où le domaine Ω est l'homothétique de $\hat{\Omega}$ de rayon égal à 1, $\mathcal{R}e = \frac{\kappa L}{\nu}$ et $\tilde{p} = \frac{L}{\kappa^2}\mathcal{R}e\,\pi$. Posons $x := \tilde{x}$ pour simplifier. Si $\tilde{u}$ est solution de (5.2), alors (u,p) avec $u = \tilde{u} - e_1$ et $p = \tilde{p}$ est solution du système suivant :

$$\begin{cases} -\Delta u \;+\; \mathcal{R}e\;\; D_1 u \;+\; \mathcal{R}e\,(u\cdot\nabla)u \;+\; \nabla p \;=\; F & \text{dans} \;\; \overline{\Omega}^c, \\ \operatorname{div} u \;=\; 0 & \text{dans} \;\; \overline{\Omega}^c, \\ u\,|\,\partial\Omega \;=\; -e_1, \quad u(x)\to 0 \;\; (|x|\to\infty), \end{cases} \tag{5.3}$$

Nous introduisons maintenant le maillage exponentiel introduit par Goldstein [50]. Nous choisissons $s \in]1,\infty[$ et nous supposons que $\Omega_s := B(0,s)\setminus\overline{\Omega}$ est le voisinage de Ω dans lequel nous voulons avoir une approximation précise de l'écoulement extérieur (u,p) solution de (5.3). Choisissons $h \in (0,s/2)$, $R \in (2s,\infty)$. Nous considérons le domaine extérieur $\Omega_{h,R}$ tronqué par le polyèdre nommé $\Gamma_{h,R}$ et tel que

$$B(0,R(1-h^2/s^2)^{1/2})\setminus\overline{\Omega} \subset \Gamma_{h,R} \subset B(0,R)\setminus\overline{\Omega}.$$

Nous voulons de plus que les noeuds du polyèdre de troncature $\Gamma_{h,R}$ appartiennent à la sphère $B(0,R)$. Nous introduisons une décomposition $\mathcal{T}_{h,R} = (K_l)_{1\le l\le k} = (K_l^{h,R})_{1\le l\le k_{h,R}}$ adaptée à $\Omega_{h,R}$ et dont les éléments K_l sont des tétraèdres. Nous introduisons une partitions en anneaux $(U_j)_{1\le j\le J}$ de la façon suivante :

$$U_0 := \Omega_s, \quad U_j := B(0,2^j\cdot s)\setminus B(0,2^{j-1}\cdot s), \forall 1\le j\le N, \tag{5.4}$$

et nous supposons que le diamètre de chaque tétraèdre K_l avec $K_l \cap U_j \neq \emptyset$ est d'ordre $2^j h$ pour tout $j \in \{1,\cdots,J\}$. Nous imposons de plus à cette triangulation que le maillage de chaque anneau se déduise du précédent par une homothétie. Cette décomposition entraîne que le nombre de noeuds de $\mathcal{T}_{h,R}$ croît seulement en $\log(R)$ quand R augmente (voir [50]). Appelons h_K le diamètre de K pour tout $K \in T_{h,R}$. La régularité σ_K de K est définie par

$$\sigma_K^{-1} = \sup\{\text{ diam }(B);\ \text{B est une boule incluse dans } K\}/h_K.$$

Nous supposons que cette triangulation est régulière (voir [49, Appendix A.1]) ce qui signifie qu'il existe une constante σ telle que

$$\forall K \in \mathcal{T}_{h,R}, \quad \sigma_K \le \sigma.$$

Nous prescrivons sur le bord de troncature $\Gamma_{h,R}$ la condition aux limites artificielles introduite

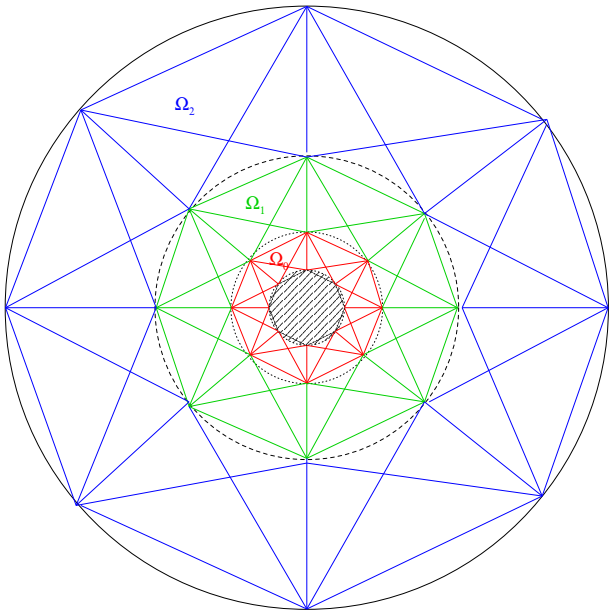


FIG. 5.1 – Projection en 2D de notre maillage exponentiel

dans [36] et définie par

$$\sum_{j=1}^{3}(D_j v_k - \delta_{jk}\cdot p - \frac{\mathcal{R}e}{2}\cdot v_j\cdot v_k)(x)\cdot n_j^{h,R}(x) \;+\; \big(R^{-1}+\frac{\mathcal{R}e}{2}\cdot(1-n_1^{h,R}(x))\big)\cdot v_k(x) \;=\; 0$$

pour $1 \leq k \leq 3,\ x \in \Gamma_{h,R}$, où $n^{h,R}$ est la normale extérieure sortante à $\Omega_{h,R}$. Nous définissons la norme $\|\ \ \|^{(h,R)}$ par

$$\forall v_{h,R} \in H^1(\Omega_{h,R})^3, \qquad \| v_{h,R} \|^{(h,R)} := \Big(\| \nabla v_{h,R} \|_2^2 \; + R^{-1} \| v_{h,R|\Gamma_{h,R}} \|_2^2 \Big)^{1/2}, \tag{5.5}$$

et le produit scalaire associé s'écrit

$$\forall u_{h,R}, v_{h,R} \in H^1(\Omega_{h,R})^3, \quad (u_{h,R}, v_{h,R})_{h,R} := (\nabla u_{h,R}, \nabla v_{h,R}) + R^{-1}(u_{h,R|\Gamma_{h,R}}, v_{h,R|\Gamma_{h,R}}). \tag{5.6}$$

Nous allons utiliser comme au chapitre précédent la discrétisation de Rebollo [90]. Nous allons redéfinir les espaces d'éléments finis et les formes bilinéaires dans le cas d'un domaine extérieur tronqué. Nous définissons les espaces d'éléments finis par

$$V_{h,R} = \{v \in C^0(\overline{\Omega_{h,R}}) : v|K_l \in P_1(K_l)^3 \;\forall\, K_l \in \mathcal{T}_{h,R}\}, \tag{5.7}$$

$$Q_{h,R} = \{p \in C^0(\overline{\Omega_{h,R}}) : p|K_l \in P_1(K_l) \;\forall K_l \in \mathcal{T}_{h,R}\}. \tag{5.8}$$

L'espace des fonctions bulles est noté

$$B_{h,R} = \left\{ v \in C^0(\overline{\Omega_{h,R}}),\ \text{tel que}\ \ v_{|K} \in \operatorname{span}(\hat{b}_K), \forall K \in \mathcal{T}_h \right\}.$$

Les produits scalaires $a_{1,h,R}$ et $a_{2,h,R}$ intervenant pour les termes de stabilisation vérifient : $\forall h > 0, R > 0$, il existe des constantes α_1, α_2, B_2 indépendantes de h et R et une constante $B_{1,h,R} > 0$ telle que, $\forall v_{h,R}, w_{h,R} \in V_{h,R}$, nous avons

$$|a_{1,h,R}(v_{h,R}, w_{h,R})| \leq B_{1,h,R} \,|\, v_{h,R} \,|_1 \,|\, w_{h,R} \,|_1 \,, \quad a_{1,h,R}(v_{h,R}, v_{h,R}) \geq \alpha_1 \,|\, v_{h,R} \,|_1^2 \,, \tag{5.9}$$

et

$$|a_{2,h,R}(v_{h,R}, w_{h,R})| \leq B_2 \,|\, v_{h,R} \,|_1 \,|\, w_{h,R} \,|_1 \,, \quad a_{2,h,R}(v_{h,R}, v_{h,R}) \geq \alpha_2 \,|\, v_{h,R} \,|_1^2 \,. \tag{5.10}$$

Nous remarquons que les notions d'uniforme continuité et coercivité en h pour le domaine borné s'étendent en une continuité et coercivité uniforme en h et R pour notre problème extérieur tronqué. Comme précédemment, nous définissons les opérateurs d'élimination des fonctions bulles par

$$R_{i,h,R} : H^{-1}(\Omega_{h,R}) \to H_0^1(\Omega_{h,R}),$$

tel que $\forall g \in H^{-1}(\Omega_R)$, $R_{i,h}g$ est l'unique élément de $B^3_{h,R}$ qui vérifie

$$a_{i,h,R}(R_{i,h,R}g, v) = \langle g, v \rangle \quad \forall v \in B^3_{h,R}.$$

Soit $S : H^1(\Omega_{h,R}) \times H^1(\Omega_{h,R}) \to H^{-1}(\Omega_{h,R})$ l'opérateur bilinéaire continu défini par

$$\begin{aligned}
&\forall u_{h,R}, w_{h,R}, v_{h,R}, \quad \langle S(u_{h,R}, v_{h,R}), w_{h,R} \rangle = \mathcal{R}e \int_{\Omega_{h,R}} D_1 v_{h,R} \cdot w_{h,R} \, dx - \frac{\mathcal{R}e}{2} \int_{\Gamma_{h,R}} v_{h,R} \, w_{h,R} n_1^{h,R} \, dS \\
&+\mathcal{R}e \int_{\Omega_{h,R}} \left((u_{h,R} \cdot \nabla) v_{h,R} \cdot w_{h,R} + \frac{1}{2} (\operatorname{div} u_{h,R}) \, v_{h,R} \cdot w_{h,R} \right) dx \\
&-\frac{\mathcal{R}e}{2} \int_{\Gamma_{h,R}} (u_{h,R} \cdot n^{(h,R)}) v_{h,R} \cdot w_{h,R} \, dS.
\end{aligned} \tag{5.11}$$

Nous introduisons les formes bilinéaires et trilinéaires suivantes :

$$\begin{aligned}
&\forall u_{h,R}, v_{h,R}, w_{h,R} \in V_{h,R}, \quad a^{(+)}_{h,R}(w_{h,R})(u_{h,R}, v_{h,R}) := \int_{\Omega_{h,R}} (\nabla u_{h,R} \cdot \nabla v_{h,R}) \, dx \\
&+ \int_{\Gamma_{h,R}} (\frac{1}{R} + \frac{\mathcal{R}e}{2}) u_{h,R} v_{h,R} \, dS + a_{1,h,R}\big(R_{1,h,R} S(w_{h,R}, u_{h,R}), R_{1,h,R} S(w_{h,R}, v_{h,R})\big),
\end{aligned} \tag{5.12}$$

$$\forall u_{h,R}, v_{h,R}, w_{h,R} \in V_{h,R}, \quad a^{(-)}_{h,R}(u_{h,R})(v_{h,R}, w_{h,R}) := \langle S(u_{h,R}, v_{h,R}), w_{h,R} \rangle, \tag{5.13}$$

$$\forall u_{h,R} \in V_{h,R}, \quad \forall p_{h,R} \in Q_{h,R}, \quad b_{h,R}(u_{h,R}, p_{h,R}) = \int_{\Omega_{h,R}} p_{h,R} \nabla \cdot w_{h,R}\, dx, \tag{5.14}$$

et

$$\forall p_{h,R}, q_{h,R} \in Q_{h,R}, \quad c_{h,R}(p_{h,R}, q_{h,R}) = a_{2,h,R}(R_{2,h,R}\nabla p_{h,R}, R_{2,h,R}\nabla q_{h,R}). \tag{5.15}$$

Nous considérons, comme discrétisation dans $\Omega_{h,R}$ du problème (5.3) le problème variationnel non-linéaire suivant :

$$\begin{cases} \text{Trouver} \quad u_{h,R} \in V_{h,R}, \;\; p_{h,R} \in Q_{h,R} \;\; \text{avec} \;\; u_{h,R|\partial\Omega} = -e_1 \text{ tel que} \\ \forall\, v_{h,R} \in V_{h,R} \text{ avec } v_{h,R|\partial\Omega} = 0, \quad \forall\, q_{h,R} \in Q_{h,R}, \\ a^{(+)}_{h,R}(u_{h,R})(u_{h,R}, v_{h,R}) + a^{(-)}_{h,R}(u_{h,R})(u_{h,R}, v_{h,R}) - b_{h,R}(v_{h,R}, p_{h,R}) = l(v_{h,R}), \\ -b_{h,R}(u_{h,R}, q_{h,R}) - c_{h,R}(p_{h,R}, q_{h,R}) = 0. \end{cases} \tag{5.16}$$

Nous résolvons la non-linéarité de (5.16) par une méthode de quasi-Newton, ce qui signifie que nous devons résoudre à chaque itération le problème linéaire de type Oseen avec des conditions aux limites de Dirichlet homogène suivant :

$$\text{Pour } w_{h,R} \in V_{h,R} \text{ donné tel que } w_{h,R|\partial\Omega} = -e_1, \tag{5.17}$$

Trouver $(u_{h,R}, p_{h,R}) \in V_{h,R} \times Q_{h,R}, u_{h,R|\partial\Omega} = 0$ tel que :

$$\begin{cases} \forall (v_{h,R}, q_{h,R}) \in V_{h,R} \times Q_{h,R}, \quad v_{h,R|\partial\Omega} = 0 \\ a^{(+)}_{h,R}(w_{h,R})(u_{h,R}, v_{h,R}) + a^{(-)}_{h,R}(w_{h,R})(u_{h,R}, v_{h,R}) - b_{h,R}(v_{h,R}, p_{h,R}) = \langle f, v_{h,R} \rangle \\ -b_{h,R}(u_{h,R}, q_{h,R}) - c_{h,R}(p_{h,R}, q_{h,R}) = \langle g, p_{h,R} \rangle \end{cases} \tag{5.18}$$

Dans la méthode de quasi-Newton que nous utilisons, nous devons résoudre ce problème d'Oseen avec des conditions de Dirichlet homogènes. Les second membres f et g correspondent au résidu non-linéaire.

5.2 Propriété des formes bilinéaires

Nous devons recalculer les constantes de continuité et de coercivité dans le cas d'un domaine tronqué dépendant du rayon de troncature R. Nous rappelons d'abord les inégalités de Sobolev : Il existe deux constantes $C_1 > 0$ et $C_2 > 0$ telles que $\forall h > 0, \forall R > 0$, nous avons

$$\forall w_{h,R} \in H^1(\Omega_{h,R}), \quad \| w_{h,R} \|_6 \leq C_1 \| w_{h,R} \|^{(h,R)}, \tag{5.19}$$

et

$$\forall w_{h,R} \in H^1(\Omega_{h,R}), \quad \| w_{h,R} \|_2 \leq C_2 R \| w_{h,R} \|^{(h,R)}. \tag{5.20}$$

Nous en déduisons, en utilisant l'inégalité d'interpolation, que

$$\forall w_{h,R} \in H^1(\Omega_{h,R}), \quad \| w_{h,R} \|_3 \leq C_2 R^{1/2} \| w_{h,R} \|^{(h,R)}. \tag{5.21}$$

Par intégration par partie, nous observons que

$$\begin{aligned} &\mathcal{R}e \int_{\Omega_{h,R}} ((u_{h,R} \cdot \nabla) v_{h,R} \cdot w_{h,R} + \frac{1}{2} (\operatorname{div} u_{h,R})\, v_{h,R} \cdot w_{h,R}\,)\, dx \\ &-\frac{\mathcal{R}e}{2} \int_{\Gamma_{h,R}} (u_{h,R} \cdot n^{(h,R)}) v_{h,R} \cdot w_{h,R}\, dS = \\ &\frac{\mathcal{R}e}{2} \int_{\Omega_{h,R}} (u_{h,R} \cdot \nabla v_{h,R}) w_{h,R} - (u_{h,R} \cdot \nabla w_{h,R}) v_{h,R}\, dx. \end{aligned} \tag{5.22}$$

D'où on en déduit que $\forall u_{h,R}, v_{h,R}, w_{h,R} \in V_{h,R}$,

$$\begin{aligned} |a_{h,R}^{(-)}(u_{h,R})(v_{h,R}, w_{h,R})| &\leq \mathcal{R}e \ \| D_1 v_{h,R} \|_2 \ \| w_{h,R} \|_2 + \frac{\mathcal{R}e}{2} \| v_{h,R_{|\Gamma_{h,R}}} \|_2 \ \| w_{h,R_{|\Gamma_{h,R}}} \|_2 \\ &+ \frac{\mathcal{R}e}{2} \| u_{h,R} \|_3 \ \| \nabla v_{h,R} \|_2 \ \| w_{h,R} \|_3 + \frac{\mathcal{R}e}{2} \| u_{h,R} \|_3 \ \| \nabla w_{h,R} \|_2 \ \| v_{h,R} \|_3 \\ &\leq C\mathcal{R}e\, R \| v_{h,R} \|^{(h,R)} \| w_{h,R} \|^{(h,R)} + C\mathcal{R}e\, R^{1/2} \| u_{h,R} \|^{(h,R)} \| v_{h,R} \|^{(h,R)} \| w_{h,R} \|^{(h,R)}. \end{aligned} \tag{5.23}$$

Lemme 5.1 *Il existe une constante $C > 0$ telle que*

$$\forall \hat{w}_{h,R} \in B^3_{h,R}, \quad \| \hat{w}_{h,R} \|_2 \leq C\,(hR) \| \nabla \hat{w}_{h,R} \|_2 \,.$$

Preuve : Soit K^* l'élément de référence. Pour tout $K \in \mathcal{T}_{h,R}$, il existe une application linéaire $F_K : K^* \to K$ définie par $F_K(x^*) = B_K x^* + b_K = x$. Nous notons w^* la fonction de $H^1_0(K^*)^3$ définie par $w^*(x^*) = \hat{w}_{h,R}(x)$. Nous avons

$$\| \hat{w}_{h,R} \|_2^2 = \sum_{K \in \mathcal{T}_{h,R}} \| \hat{w}_{h,R} \|_{2,K}^2 = \sum_{K \in \mathcal{T}_{h,R}} |\det(B_K)| \, \| w^* \|_{2,K^*}^2 \,.$$

Comme $w^* \in H^1_0(K^*)^3$, nous pouvons utiliser l'inégalité de Poincaré dans K^*, i.e. il existe une constante C^* telle que

$$\| w^* \|_{2,K^*} \leq C^* \| \nabla w^* \|_{2,K^*} \,.$$

Cela implique que

$$\| \hat{w}_{h,R} \|_2^2 \leq C^* \sum_{K \in \mathcal{T}_{h,R}} |\det(B_K)| \, \| \nabla w^* \|_{2,K^*}^2 \leq C^* \sum_{K \in \mathcal{T}_{h,R}} \|B_K\|^2 \, \| \nabla w^* \|_{2,K^*}^2 \,,$$

où $\|B_K\| := \sqrt{(B_K^T B_K)}$ est la norme Frobenius de la matrice B_K. La valeur de $\|B_K\|$ pour $K \in \mathcal{T}_{h,R} \cap U_j$ est d'ordre $2^j h$, d'où $\|B_K\| \leq C(hR)$ pour tout $K \in \mathcal{T}_{h,R}$. $\square$

Remarque 5.1 *Nous pouvons déduire, en utilisant l'inégalité d'interpolation, qu'il existe une constante $C > 0$ telle que, pour que $\hat{w}_{h,R} \in B^3_{h,R}$, nous avons*

$$\| \hat{w}_{h,R} \|_3 \ \leq C(hR)^{1/2} \| \nabla \hat{w}_{h,R} \|_2 \ .$$

Lemme 5.2

$$\begin{aligned} & a_{1,h,R}(R_{1,h}S(w_{h,R}, u_{h,R}), R_{1,h}S(w_{h,R}, v_{h,R})) \\ & \qquad \leq C\alpha_1^{-1}\mathcal{R}e^2\big((hR) + (hR)^{1/2} \| w_{h,R} \|^{(h,R)}\big)^2 \| u_{h,R} \|^{(h,R)} \| v_{h,R} \|^{(h,R)}. \end{aligned}$$

Preuve :

$$\begin{aligned} & |a_{1h}(R_{1,h}S(w_{h,R}, u_{h,R}), R_{1,h}S(w_{h,R}, v_{h,R}))| = |S(w_{h,R}, u_{h,R})(R_{1,h}S(w_{h,R}, v_{h,R}))| \\ & = \Bigg| \mathcal{R}e \int_{\Omega_{h,R}} D_1 u_{h,R} R_{1,h}S(w_{h,R}, v_{h,R})\, dx \\ & \quad + \mathcal{R}e \int_{\Omega_{h,R}} \big((w_{h,R} \cdot \nabla) u_{h,R} R_{1,h}S(w, v)\, dx + \frac{1}{2}(\nabla \cdot w_{h,R}) u_{h,R} R_{1,h}S(w_{h,R}, v_{h,R})\big) \Bigg| \\ & \leq C\mathcal{R}e \ \| D_1 u_{h,R} \|_2 \ \| R_{1,h}S(w_{h,R}, v_{h,R}) \|_2 \\ & \quad + C\mathcal{R}e \ \big(\| w_{h,R} \|_6 \ \| \nabla u_{h,R} \|_2 \ \| R_{1,h}S(w_{h,R}, v_{h,R}) \|_3 \\ & \quad + \ \| \nabla \cdot w_{h,R} \|_2 \ \| u_{h,R} \|_6 \ \| R_{1,h}S(w_{h,R}, v_{h,R}) \|_3 \ \big) \\ & \leq C\mathcal{R}e \ \| u_{h,R} \|^{(h,R)} \| R_{1,h}S(w_{h,R}, v_{h,R}) \|_2 \\ & \quad + C\mathcal{R}e \ \| u_{h,R} \|^{(h,R)} \| w_{h,R} \|^{(h,R)} \| R_{1,h}S(w_{h,R}, v_{h,R}) \|_3 \ . \end{aligned}$$

En appliquant le lemme 5.1 et la remarque 5.1, nous obtenons :

$$\begin{aligned} & |a_{1h}(R_{1,h}S(w_{h,R}, u_{h,R}), R_{1,h}S(w_{h,R}, v_{h,R}))| \\ & \leq C\mathcal{R}e\big((hR) \| u_{h,R} \|^{(h,R)} + (hR)^{1/2} \| u_{h,R} \|^{(h,R)} \| w_{h,R} \|^{(h,R)}\big) \| \nabla R_{1,h}S(w_{h,R}, v_{h,R}) \|_2 \ . \end{aligned} \tag{5.24}$$

En utilisant la coercivité de la forme bilinéaire a_{1h} et le résultat précédent (5.24), on en déduit :

$$\begin{aligned} & \| \nabla R_{1,h}S(w_{h,R}, v_{h,R}) \|_2^2 \ \leq \alpha_1^{-1} a_{1h}(R_{1,h}S(w_{h,R}, v_{h,R}), R_{1,h}S(w_{h,R}, v_{h,R})) \\ & \leq C\alpha_1^{-1}\mathcal{R}e\big((hR) \| v_{h,R} \|^{(h,R)} + (hR)^{1/2} \| v_{h,R} \|^{(h,R)} \| w_{h,R} \|^{(h,R)}\big) \| \nabla R_{1,h}S(w_{h,R}, v_{h,R}) \|_2 \ . \end{aligned}$$

Par conséquent, nous obtenons l'estimation suivante pour $\| \nabla R_{1,h} S(w_{h,R}, v_{h,R}) \|_2$:

$$\| \nabla R_{1,h} S(w_{h,R}, v_{h,R}) \|_2 \leq C\alpha_1^{-1}\mathcal{R}e\big((hR) \| v_{h,R} \|^{(h,R)} + (hR)^{1/2} \| v_{h,R} \|^{(h,R)} \| w_{h,R} \|^{(h,R)}\big). \tag{5.25}$$

Injectant (5.25) dans (5.24), nous trouvons la relation

$$\begin{aligned} & a_{1,h,R}(R_{1,h}S(w_{h,R}, u_{h,R}), R_{1,h}(w_{h,R}, v_{h,R})) \\ & \leq C\alpha_1^{-1}\mathcal{R}e^2 \left((hR) + (hR)^{1/2} \| w_{h,R} \|^{(h,R)}\right)^2 \| u_{h,R} \|^{(h,R)} \| v_{h,R} \|^{(h,R)}. \quad \square \end{aligned}$$

Lemme 5.3 *Il existe une constante $\mathcal{C} > 0$ indépendante de h, R et $\mathcal{R}e$ telle que*

$$\forall p_{h,R}, q_{h,R} \in Q_{h,R}, \quad c_{h,R}(p_{h,R}, q_{h,R}) \leq \mathcal{C} \| p_{h,R} \|_2 \| q_{h,R} \|_2 . \tag{5.26}$$

Preuve : Pour tout $p_{h,R}, q_{h,R} \in Q_{h,R}$,

$$|a_{2h}(R_{2,h}\nabla p_{h,R}, R_{2,h}\nabla q_{h,R})| = |(p_{h,R}, \nabla \cdot (R_{2,h}\nabla q_{h,R}))| \leq \| p_{h,R} \|_2 \| \nabla R_{2,h}(\nabla q_{h,R}) \|_2 .$$

Comme

$$\| \nabla R_{2,h}(\nabla q_{h,R}) \|_2^2 \leq \alpha_2^{-1} a_{2h}(R_{2,h}\nabla p_{h,R}, R_{2,h}\nabla q_{h,R}) \leq \alpha_2^{-1} \| q_{h,R} \|_2 \| \nabla R_{2,h}(\nabla q_{h,R}) \|_2 ,$$

on a

$$\| \nabla R_{2,h}(\nabla q_{h,R}) \|_2 \leq \alpha_2^{-1} \| q_{h,R} \|_2 ,$$

et

$$|a_{2h}(R_{2,h}\nabla p_{h,R}, R_{2,h}\nabla q_{h,R})| \leq \alpha_2^{-1} \| q_{h,R} \|_2 \| p_{h,R} \|_2 . \quad \square$$

En résumé, nous avons

$$\begin{aligned} & \forall u_{h,R}, v_{h,R}, w_{h,R}, \quad a_{h,R}^{(+)}(w_{h,R})(u_{h,R}, v_{h,R}) \leq \mathcal{A}_+ \| u_{h,R} \|^{(h,R)} \| v_{h,R} \|^{(h,R)}, \\ & \text{où } \mathcal{A}_+ = C\left(1 + \mathcal{R}e\, R + \alpha_1^{-1}\mathcal{R}e^2\big((hR) + (hR)^{1/2} \| w_{h,R} \|^{(h,R)}\big)^2\right). \end{aligned} \tag{5.27}$$

Comme $a_{1,h,R}(R_{1,h}S(w_{h,R}, u_{h,R}), R_{1,h}S(w_{h,R}, u_{h,R})) \geq 0$, nous avons

$$\forall u_{h,R}, w_{h,R} \in V_{h,R}, \quad a_{h,R}^{(+)}(w_{h,R})(u_{h,R}, u_{h,R}) \geq \alpha \left(\| u_{h,R} \|^{(h,R)}\right)^2 , \tag{5.28}$$

pour $\alpha = 1$.

$$\forall u_{h,R}, v_{h,R}, w_{h,R} \in V_{h,R}, \quad a_{h,R}^{(-)}(w_{h,R})(u_{h,R}, v_{h,R}) \leq \mathcal{A}_- \| u_{h,R} \|^{(h,R)} \| v_{h,R} \|^{(h,R)},$$
$$\text{où } \mathcal{A}_- = C\mathcal{R}e \left(R + R^{1/2} \| w_{h,R} \|^{(h,R)}\right). \quad (5.29)$$

Il existe une constante $\mathcal{B} > 0$ pour que

$$\forall u_{h,R} \in V_{h,R}, p_{h,R} \in Q_{h,R}, \quad b_{h,R}(u_{h,R}, p_{h,R}) \leq \mathcal{B} \| p_{h,R} \|_2 \ \| u_{h,R} \|^{(h,R)}. \quad (5.30)$$

5.3 Préconditionnement pour le système d'Oseen

Nous considérons le système d'Oseen (5.18) où $w_{h,R}$ est une fonction donnée dans $V_{h,R}$. Nous notons $n_{h,R}$ la dimension de l'espace d'éléments finis associé à la vitesse $V_{h,R}$ et $m_{h,R}$ la dimension de l'espace d'éléments finis associé à la pression $Q_{h,R}$. Nous appelons $(\varphi_i)_{1 \leq i \leq n_{h,R}}$ et $(\psi_j)_{1 \leq j \leq m_{h,R}}$ les fonctions de base de $V_{h,R}$ et $Q_{h,R}$. Nous associons les fonctions $u_{h,R}, v_{h,R} \in V_{h,R}$, $p_{h,R}, q_{h,R} \in Q_{h,R}$ avec les vecteurs $U, V \in \mathbb{R}^{n_{h,R}}, P, Q \in \mathbb{R}^{m_{h,R}}$ tel que $u_{h,R} = \sum_{i=1}^{n_{h,R}} U_i \varphi_i$, $v_{h,R} = \sum_{i=1}^{n_{h,R}} V_i \varphi_i$, $p_{h,R} = \sum_{j=1}^{m_{h,R}} P_i \psi_j$ et $q_{h,R} = \sum_{j=1}^{m_{h,R}} Q_i \psi_j$. Soient les matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{E}, \boldsymbol{V}, \boldsymbol{Q}$ et soient les vecteurs F et G définis par :

$$(\boldsymbol{A}_+)_{i,j} = a_{h,R}^{(+)}(w_{h,R})(\varphi_j, \varphi_i), \quad (\boldsymbol{A}_-)_{i,j} = a_{h,R}^{(-)}(w_{h,R})(\varphi_j, \varphi_i), \quad \boldsymbol{B}_{i,j} = b_{h,R}(\varphi_j, \psi_i),$$

$$\boldsymbol{E}_{i,j} = c_{h,R}(\psi_j, \psi_i), \quad \boldsymbol{V}_{i,j} = (\varphi_j, \varphi_i)_{h,R}, \quad \boldsymbol{Q}_{i,j} = (\psi_j, \psi_i)_Q$$

$$F_i = \langle f, \varphi_i \rangle, \quad G_i = \langle g, \psi_i \rangle.$$

Nous notons $\boldsymbol{A} := \boldsymbol{A}_+ + \boldsymbol{A}_-$ pour la matrice de type convection diffusion. La formulation matricielle de (5.18) peut s'écrire sous la forme du système linéaire

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix} \quad (5.31)$$

où $\boldsymbol{E}$ est symétrique définie positive. Nous avons étudié dans le chapitre 3 l'influence du préconditionneur

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B}^T & \boldsymbol{Q} \end{pmatrix} \quad (5.32)$$

pour de tels problèmes de point-selle généralisés. Nous allons appliquer ces résultats à notre problème tronqué. Nous commençons par prouver que la condition inf-sup généralisée est vérifiée.

Lemme 5.4 *Il existe une constante $\gamma > 0$ telle que*

$$\sup_{v_{h,R}\in V_{h,R}} \frac{b(v_{h,R}, q_{h,R})}{\| v_{h,R} \|^{(h,R)}} + c(q_{h,R}, q_{h,R})^{1/2} \geq \gamma \| q_{h,R} \|_2 . \tag{5.33}$$

Preuve : Nous savons que l'inégalité de Babuska-Brezzi pour les P1+bulles -P1 est satisfaite pour notre maillage exponentielle avec une constante indépendante de h et R (voir [33]). Cela signifie qu'il existe une constante $\beta > 0$ telle que $\forall h > 0$ suffisamment petit, $\forall R > 0$ assez grand, on a :

$$\forall q_{h,R} \in Q_{h,R}, \qquad \sup_{\substack{v_{h,R}\in V_{h,R} \\ \hat{v}_{h,R}\in B^3_{h,R}}} \frac{b(v_{h,R} + \hat{v}_{h,R}, q_{h,R})}{\| v_{h,R} + \hat{v}_{h,R} \|^{(h,R)}} \geq \beta \| q_{h,R} \|_2 . \tag{5.34}$$

Nous savons déjà par Rebollo [90] que pour tout $v_{h,R} \in V_{h,R}$, pour tout $\hat{v}_{h,R} \in \boldsymbol{B}^3_{h,R}$, nous avons

$$|v_{h,R} + \hat{v}_{h,R}|_1^2 \leq 2 \left(|v_{h,R}|_1^2 + |\hat{v}_{h,R}|_1^2 \right) .$$

Comme $\hat{v}_{h,R} = 0$ sur $\Gamma_{h,R}$, nous en obtenons l'inégalité suivante :

$$\forall v_{h,R} \in V_{h,R}, \forall \hat{v}_{h,R} \in B^3_{h,R}, \qquad \| v_{h,R} + \hat{v}_{h,R} \|^{(h,R)} \leq \sqrt{2} \Big(\| v_{h,R} \|^{(h,R)} + \| \hat{v}_{h,R} \|^{(h,R)} \Big) . \tag{5.35}$$

La suite de la preuve est similaire à celle du lemme 4.1 où $| \ |_1$ est remplacé par $\| \ \|^{(h,R)}$. $\square$

Nous avons démontré dans le chapitre 4 le résultat suivant (voir théorème 3.2) : Les valeurs propres du problème (5.31) préconditionné par (5.32) sont incluses dans le domaine rectangulaire

$$\left[\frac{\gamma^2}{2} \min(1, \alpha(\mathcal{A}_+^2 + \mathcal{A}_-^2)^{-1}), (\frac{\mathcal{B}^2}{\alpha} + \mathcal{C}) \right] \times \left[-\frac{\mathcal{B}^2}{2\alpha}, \frac{\mathcal{B}^2}{2\alpha} \right] , \tag{5.36}$$

où γ est la constante de l'inégalité LBB définie dans (5.33) et $\mathcal{A}_+$, $\mathcal{A}_-, \mathcal{B}, \mathcal{C}, \alpha$ sont les constantes de coercivité et continuité des formes bilinéaires a, b, c définie en (5.27), (5.29),(5.30), (5.26) et (5.28). Nous explicitons ces bornes en supposant qu'il existe une constante μ telle que $hR \leq \mu$. En effet, cette supposition apparaît naturellement quand on calcule l'erreur de discrétisation. Nous déduisons de (5.36) qu'il existe quatre constantes C_1, C_2, C_3 et C_4 indépendantes de h, R et $\mathcal{R}e$ telles que les valeurs propres du problème préconditionné soient

incluses dans le rectangle

$$\left[C_1 \min\left(1, m(R, \mathcal{R}e, w, \mu)^{-1}\right), C_4\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$$

$$\text{avec} \quad m(R, \mathcal{R}e, w, \mu) = C_2 \left(1 + \mathcal{R}e\ R + \mathcal{R}e^2 \left(\mu + \mu^{1/2} \parallel w \parallel^{(h,R)}\right)^2\right)^2 + C_3 \mathcal{R}e^2 \left(R + R^{1/2} \parallel w \parallel^{(h,R)}\right)^2. \quad (5.37)$$

Cela signifie que les valeurs propres sont indépendantes de la taille de la maille, mais dépendent du rayon de troncature R.

Nous souhaitons prévoir les décroissances asymptotiques en fonction de $\mathcal{R}e$ et R pour ce rectangle. Supposons que $\exists \eta > 0$ tel que $\parallel w \parallel^{(h,R)} < \eta$. Alors, si nous fixons $\mathcal{R}e$ et si R est suffisamment grand, le domaine est de la forme

$$\left[\tilde{C}_1 \eta R^{-2}, \tilde{C}_2\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right], \quad (5.38)$$

où $\tilde{C}_1$ et $\tilde{C}_2$ dépendent de $\mathcal{R}e$. Si nous fixons R, alors pour $\mathcal{R}e$ assez grand, nous obtenons un rectangle du type

$$\left[\bar{C}_1 \eta \mathcal{R}e^{-4}, \bar{C}_2\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right], \quad (5.39)$$

où $\bar{C}_1$ et $\bar{C}_2$ dépendent de R. Ce dernier rectangle se réduit à

$$\left[\bar{C}_1 \eta \mathcal{R}e^{-2}, \bar{C}_2\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right], \quad (5.40)$$

lorsque nous ne considérons pas la stabilisation en vitesse.

5.4 Choix des termes de stabilisation

Nous allons vérifier que les expressions données par Rebollo [90] pour calculer les termes de stabilisation vérifient les relations (5.9) et (5.10). Rappelons que nous avons modifié ces expressions dans le chapitre 4 pour qu'elles soient adaptées aux équations de Navier-Stokes adimensionnées. Soit A et P deux constantes numériques et soit $p \in (3, \infty)$. Pour $K \in \mathcal{T}_h$, le nombre de Péclet est défini de la façon suivante :

$$Pe_K = \mathcal{R}e \ \parallel (e_1 + w_h)|_K \parallel_p \, diam\, K$$

Nous avons alors que pour tout v_h, z_h,

$$a_{1,h}(v_{h,R}, z_{h,R}) = \sum_{K \in T_{h,R}} \frac{Vol(K)}{840^2 \tau_{1,K} \mid \hat{b}_K \mid_1^2} (\nabla v_{h,R}, \nabla z_{h,R})_K, \quad (5.41)$$

où le symbole $(\ ,\ )_K$ représente le produit scalaire de $L^2(K)$ et

$$\tau_{1,K} = A\frac{h_K}{\mathcal{R}e\ \ \|\,(e_1 + w_{h,R})|_K\,\|_p}\min\,(P, Pe_k).$$

On voit que Ah_K^2 est une borne supérieure de $\tau_{1,K}$. Pour la borne inférieure, si $P \geq Pe_K$, alors $\tau_{1,K} \geq Ah_k^2$. Si $P \leq Pe_K$,

$$\tau_{1,K} \geq A\,\frac{h_K}{\mathcal{R}e}\,\frac{1}{\|\,w_{h,R} + e_1\,\|_p}.$$

Nous pouvons supposer que $h < 1$ puisque le diamètre du premier anneau est égal à 1. Pour $K \in U_j$, $h_K \equiv 2^j h$, d'où nous obtenons que

$$\tau_{1,K} \geq A\,\frac{h_K^2}{2^J\mathcal{R}e}\,\frac{1}{\|\,w_{h,R} + e_1\,\|_p} \geq AS\frac{h_K^2}{R\mathcal{R}e}\,\frac{1}{\|\,w_{h,R} + e_1\,\|_p}.$$

D'après [90, Lemma 5.2], nous en déduisons qu'il existe deux constantes η_1 et η_2 ne dépendant que de la régularité σ de la triangulation telle que

$$\alpha_1 = \frac{\eta_2}{840^2 A} \quad \text{et} \quad B_{1,h,R} = \frac{\eta_1}{840^2 A}\max(1, \mathcal{R}e\ R\,\|\,w_{h,R} + e_1\,\|_p\,). \tag{5.42}$$

5.5 Résultats numériques

5.5.1 Présentation des tests

Nous ferons les tests dans deux situations :

a) Cas test polynomial.

On choisit comme fonction test

$$\begin{aligned} u_1(x) &= (y - z)\Lambda(x) \\ u_2(x) &= (z - x)\Lambda(x) \\ u_3(x) &= (x - y)\Lambda(x) \\ \pi(x) &= 10(|x|^{-3} - 2|x|^{-4} + |x|^{-5}). \end{aligned} \tag{5.43}$$

avec

$$\Lambda(x) = 40|x|^{-3}(1 - 3|x|^{-1} + 3|x|^{-2} - |x|^{-3}).$$

Si

$$-\triangle u + \mathcal{R}e\ D_1 u + \mathcal{R}e' u \cdot \nabla u + \nabla p = f,$$

alors f s'écrit :

$$\begin{aligned}
f_0(x) &= -(y-z)\Gamma_4(x) + \mathcal{R}e\ (y-z)x\Gamma_3(x) + \mathcal{R}e'\Lambda(x)^2(z+y-2x) + x\Gamma_2(x),\\
f_1(x) &= -(z-x)\Gamma_4(x) + \mathcal{R}e\ (z-x)x\Gamma_3(x) - \mathcal{R}e\ \Lambda(x) + \mathcal{R}e'\Lambda(x)^2(x+z-2y) + y\Gamma_2(x),\\
f_3(x) &= -(x-y)\Gamma_4(x) + \mathcal{R}e\ (x-y)x\Gamma_3(x) + \mathcal{R}e\Lambda(x) + \mathcal{R}e'\Lambda(x)^2(x+y-2z) + z\Gamma_2(x),\\
f_4(x) &= 0,
\end{aligned}$$

où

$$\Gamma_2(x) = 10|x|^{-5}(-3+8|x|^{-1}-5|x|^{-2}),$$

$$\Gamma_4(x) = 240|x|^{-6}(-2+5|x|^{-1}-3|x|^{-2}),$$

$$\Gamma_3(x) = 120|x|^{-5}(-1+4|x|^{-1}-5|x|^{-2}+2|x|^{-3}).$$

Cette fonction-test vérifie les conditions de Dirichlet homogènes et est construite de façon à ce que $\nabla \cdot (u_1, u_2, u_3)^T = 0$. Nous introduisons un paramètre supplémentaire $\mathcal{R}e'$ de manière à pouvoir inclure le cas $\mathcal{R}e' = 0$, correspondant au problème d'Oseen

$$-\triangle u + \mathcal{R}e\ D_1 u + \nabla p = f.$$

b) Cas test physique

Nous choisirons comme second membre $f = 0$. Cela veut dire que le fluide n'est soumis à aucun champ de force extérieur et que nous négligeons la pesanteur. Nous prenons comme condition de Dirichlet $u_{|\partial\Omega} = -e_1$.

Remarque 5.2 *Dans ce chapitre, les matrices ont été assemblées sur le domaine entier. Nous pouvons donc utiliser les méthodes d'algèbres linéaires usuelles. L'utilisation du maillage exponentiel pour réduire le stockage sera discuté par la suite.*

5.5.2 Valeurs propres du problème préconditionné

Nous nous plaçons d'abord dans le cas test a) avec $\mathcal{R}e' = 0$. Nous calculons les plus petites et les plus grandes parties réelles des valeurs propres du problème non préconditionné puis préconditionné. Nous n'assemblons pas dans un premier temps la stabilisation en vitesse. Ce test nous permet de connaître exactement l'influence sur les valeurs propres du terme $e_1 \cdot \nabla$ en fonction du rayon R et du nombre de Reynolds $\mathcal{R}e$. Dans le tableau 5.1, la première valeur représente la partie réelle de la valeur propre de plus grande partie réelle. La seconde valeur (en italique) correspond à la partie réelle de la valeur propre de plus petite partie réelle. On

	R=4	R=8	R=16	R=32
$h = 1/2, \quad \mathcal{R}e = 1$	2.39068	13.05051	71.09987	407.19287
	0.01142	*0.01136*	*0.01136*	*0.01136*
$h = 1/2, \quad \mathcal{R}e = 10$	0.80746	3.87074	26.54404	201.18800
	0.01136	*0.01134*	*0.01134*	*0.01134*
$h = 1/4, \quad \mathcal{R}e = 1$	1.43347	4.98361	35.00725	238.03966
	0.06355	*0.0014056*	*0.0014057*	*0.0015853*
$h = 1/4, \quad \mathcal{R}e = 10$	0.42256	2.560515	14.669986	82.443984
	0.06126	*0.0014021*	*0.0014023*	*0.0014948*

TAB. 5.1 – Valeurs propres du complément de Schur. $w_{h,R} = 0$. Pas de préconditionnement. En roman, la borne supérieure réelle. En italique, la borne inférieure réelle.

observe que la borne supérieure réelle du spectre croît à peu près exponentiellement avec R et les bornes réelles du spectre décroissent avec h. La borne supérieure dépend également fortement du nombre de Reynolds $\mathcal{R}e$. Nous ne poursuivons pas les calculs, car le très mauvais conditionnement du complément de Schur fait diverger la méthode d'Arnoldi.

Au vue du tableau 5.2, on constate que quand la convection est faible (faible nombre de Reynolds ou faible rayon), on n'observe aucune décroissance de la plus petite valeur propre (par ex pour $R = 4$ ou pour $\mathcal{R}e = 1$.) Par contre, si nous regardons les données pour $\mathcal{R}e = 10$ et $R = 32$, nous voyons que la plus petite valeur propre décroît lors des trois premiers raffinements mais que cette décroissance cesse à $h = 1/16$. Pour $R = 16$, le phénomène est moins net mais dans tous les cas, cette décroissance tend à stagner quand on continue à raffiner. Cela ne contredit pas les inégalités théoriques : En effet, pour de grandes valeurs de h, les inégalités inverses sont de meilleures inégalités que celles utilisant les bornes de continuité et on pourrait prédire théoriquement une décroissance selon h jusqu'à un certain seuil.

Le décroissance en R et $\mathcal{R}e$ est nettement plus faible que celle annoncée par les estimées théoriques. Nous avions déjà constaté cela lors du test dans un cube. Pour mieux comprendre le comportement des valeurs propres, nous avons tracé le spectre de $\boldsymbol{Q}^{-1}\boldsymbol{S}$ pour différentes valeurs de R et $\mathcal{R}e$ quand $h = 1/4$. Nous observons dans la figure 5.2 qu'une augmentation du nombre de Reynolds $\mathcal{R}e$ joue simultanément sur la partie réelle et la partie imaginaire des valeurs propres. Par contre, R agit différemment sur le spectre : on voit dans la figure 5.3 que

	R=4	R=8	R=16	R=32
$h = 1/2, \quad \mathcal{R}e = 1$	1.433473	1.434075	1.4340454	1.4340586
	0.063548	*0.063357*	*0.062993*	*0.061958*
$h = 1/2, \quad \mathcal{R}e = 5$	1.395297	1.396560	1.396773	1.3968181
	0.0628294	*0.0613446*	*0.057675*	*0.04742464*
$h = 1/2, \quad \mathcal{R}e = 10$	1.379707	1.381793	1.3820807	1.382141
	0.061256	*0.057675*	*0.0482294*	*0.034518*
$h = 1/4, \quad \mathcal{R}e = 1$	1.27322	1.2732836	1.2732997	1.27329999
	0.06918	*0.069618*	*0.069804*	*0.06988896*
$h = 1/4, \quad \mathcal{R}e = 5$	1.0572641	1.057342	1.05721083	1.057198
	0.0696944	*0.0698615*	*0.0442464*	*0.0286589*
$h = 1/4, \quad \mathcal{R}e = 10$	0.93265	0.933576	0.934647	0.934605
	0.069777	*0.046009*	*0.0278978*	*0.0217531*
$h = 1/8, \quad \mathcal{R}e = 1$	1.1194181	1.1942614	1.19426134	1.194213
	0.0697365	*0.0697829*	*0.0698052*	*0.0698393*
$h = 1/8, \quad \mathcal{R}e = 5$	1.0208700	1.0208712	1.0208769	1.0208773
	0.0697584	*0.0698143*	*0.0379934*	*0.0236745*
$h = 1/8, \quad \mathcal{R}e = 10$	0.888447	0.888533	0.8885484	0.888555
	0.0697829	*0.03844003*	*0.022875*	*0.0171925*
$h = 1/16, \quad \mathcal{R}e = 1$	1.106324	1.1063514	1.1063513	1.1063513
	0.068540	*0.0685579*	*0.06856596*	*0.0685715*
$h = 1/16, \quad \mathcal{R}e = 5$	1.0076067	1.0076067	1.0076067	1.0076067
	0.0685418	*0.0685594*	*0.0366970*	*0.0227263*
$h = 1/16, \quad \mathcal{R}e = 10$	0.09176969	0.09176969	0.0917697	0.091769692
	0.06854277	*0.036339*	*0.0218450*	*0.01837461*

TAB. 5.2 – Valeurs propres de $\boldsymbol{Q}^{-1}\boldsymbol{S}$, $w_{h,R} = 0$. Pas de stabilisation en vitesse. En roman, la borne supérieure réelle. En italique, la borne inférieure réelle.

l'amplitude de la partie imaginaire des valeurs propres reste inchangée. Par contre, le nombre de valeurs propres de petite partie réelle augmente avec R. Il est tout à fait normal que R et $\mathcal{R}e$ agissent différemment sur les valeurs propres puisque changer R ne fait plus ou moins que changer la discrétisation tronquée du même opérateur continu.

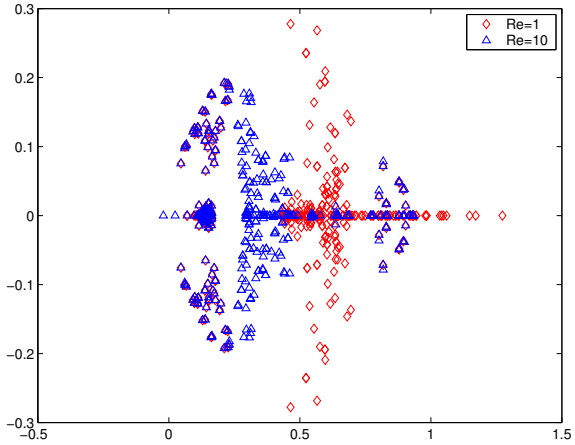


FIG. 5.2 – Spectre de $\boldsymbol{Q}^{-1}\boldsymbol{S}$ pour $R = 8$.

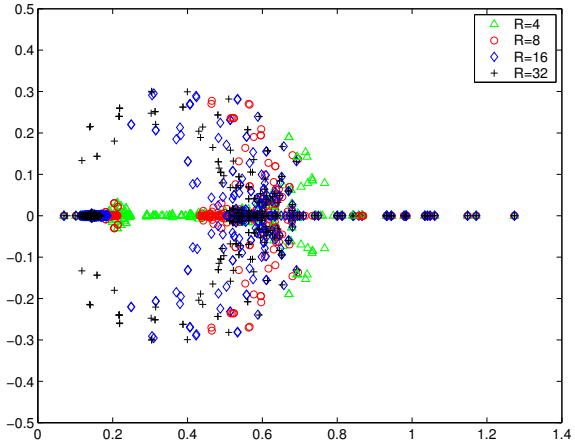


FIG. 5.3 – Spectre de $\boldsymbol{Q}^{-1}\boldsymbol{S}$ pour $\mathcal{R}e = 1$.

Nous choisissons ensuite $\mathcal{R}e' = \mathcal{R}e$ et nous assemblons la matrice avec pour terme convectif le vecteur solution u défini en (5.43). Cette fois-ci, on assemble la stabilisation en vitesse avec pour paramètres $A = P = 1$. On observe dans le tableau 5.3 que la borne réelle inférieure du spectre décroît avec $\mathcal{R}e$ quand h devient petit. On a déjà noté ce phénomène dans le cas

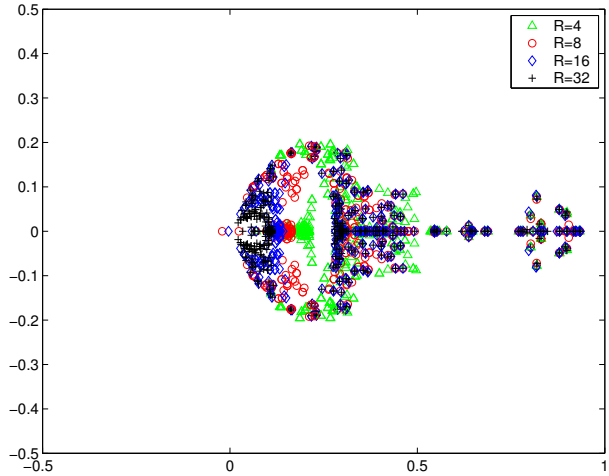


FIG. 5.4 – Spectre de $\boldsymbol{Q}^{-1}\boldsymbol{S}$ pour $\mathcal{R}e = 10$.

borné et l'augmentation de la dépendance en Reynolds est prévue par la théorie.

5.5.3 Performances

On fait sur le bloc matriciel de type convection-diffusion un GMRES(30) préconditionné par ILU(0) jusqu'à ce que le résidu relatif soit inférieur à 10^{-15}. La résolution très précise du problème de convection diffusion permet d'avoir des résultats indépendants du mauvais conditionnement de cet opérateur. La matrice $\boldsymbol{Q}$ est remplacée par un 'mass lumping' de $\boldsymbol{Q}$. On n'assemble pas dans un premier temps la stabilisation en vitesse. On s'occupe tout d'abord du problème linéaire : on prend $\mathcal{R}e' = 0$. Sur la matrice complète, on applique un FGMRES sans redémarrage que l'on résout jusqu'à un résidu de l'ordre de 10^{-5}. Le fait de ne pas redémarrer permet une meilleure comparaison des résultats : en effet, le nombre d'itérations du GMRES redémarré dépend également du nombre de redémarrage. Le tableau 5.4 nous donne le nombre d'itérations du FGMRES nécessaires pour obtenir la convergence. L'appellation second membre polynomial correspond au cas test a). La dénomination second membre nul correspond à $w_{h,R} = 0$, $f = 0$ et où le problème d'Oseen est résolu avec les conditions de Dirichlet inhomogènes $u = -e_1$ sur $\partial\Omega$.

On constate dans les trois premières colonnes du tableau 5.4 que le nombre d'itérations du FGMRES croît avec le nombre de raffinements. Ce phénomène est dû au fait que le conditionnement du second membre croît avec $\mathcal{R}e$ et R. En effet, si nous regardons les trois dernières colonnes du tableau 5.4, on voit que cette fois-ci le nombre d'itérations tend à

	R=4	R=8	R=16	R=32
$h = 1/2, \quad \mathcal{R}e = 1$	1.41135	1.411135	1.411112	1.411126
	0.062800	*0.062748*	*0.062740*	*0.061885*
$h = 1/2, \quad \mathcal{R}e = 5$	1.36556	1.365648	1.365650	1.36548
	0.059303	*0.058197*	*0.056417*	*0.046989*
$h = 1/2, \quad \mathcal{R}e = 10$	1.37526	1.375736	1.375704	1.375695
	0.058401	*0.053160*	*0.049435*	*0.034597*
$h = 1/4, \quad \mathcal{R}e = 1$	1.248038	1.247578	1.247549	1.247549
	0.069733	*0.069968*	*0.699687*	*0.069967*
$h = 1/4, \quad \mathcal{R}e = 5$	1.013431	1.013309	1.013308	1.013312
	0.072150	*0.070182*	*0.047684*	*0.034597*
$h = 1/4, \quad \mathcal{R}e = 10$	0.989509	0.989459	0.989460	0.989480
	0.057357	*0.034037*	*0.027707*	*0.021960*
$h = 1/8, \quad \mathcal{R}e = 1$	1.188877	1.188412	1.188390	1.188390
	0.069884	*0.070112*	*0.070007*	*0.069911*
$h = 1/8, \quad \mathcal{R}e = 5$	1.042867	1.042867	1.042867	1.042867
	0.039913	*0.038175*	*0.038167*	*0.024657*
$h = 1/8, \quad \mathcal{R}e = 10$	0.921047	0.899542	0.899547	0.899540
	0.029208	*0.028095*	*0.020501*	*0.017386*

TAB. 5.3 – Valeurs propres de $\boldsymbol{Q}^{-1}\boldsymbol{S}$, convection polynomiale définie en a), stabilisation en vitesse de paramètres $A = P = 1$. En roman, la borne supérieure réelle. En italique, la borne inférieure réelle.

			Second membre polynomial			Second membre nul		
h	R	Taille	$\mathcal{R}e = 1$	$\mathcal{R}e = 5$	$\mathcal{R}e = 10$	$\mathcal{R}e = 1$	$\mathcal{R}e = 5$	$\mathcal{R}e = 10$
1/2	4	312	18	18	18	27	25	24
1/2	8	440	19	20	20	27	27	26
1/2	16	568	23	24	24	29	30	33
1/2	32	696	27	28	30	32	37	37
1/2	64	824	31	34	48	37	42	40
1/4	4	2152	25	27	28	36	37	37
1/4	8	3128	29	33	33	38	48	50
1/4	16	4104	35	41	40	47	59	68
1/4	32	5080	43	49	51	52	75	79
1/4	64	6056	56	63	103	64	89	87
1/8	4	16200	30	38	42	42	75	64
1/8	8	23912	37	47	56	51	78	88
1/8	16	31624	46	66	75	76	105	117
1/8	32	39336	62	88	99	78	126	159
1/8	64	47048	83	121	177	99	194	198
1/16	4	131522	35	48	62	40	56	70
1/16	8	167592	42	64	84	45	73	95
1/16	16	249096	56	95	118	55	99	142
1/16	32	310600	83	140	156	74	148	189
1/16	64	372104	123	197	231	109	201	261

Tab. 5.4 – Nombre d'itérations du FGMRES pour le problème linéarisé

s'arrêter de croître. Dans les cas où $\mathcal{R}e$ et R sont petits, la borne supérieure semble atteinte pour $h = 1/8$. Par contre, il faudrait raffiner davantage pour les grandes valeurs de $\mathcal{R}e$ et R. N'oublions pas qu'augmenter R sans raffiner entraîne que nous avons des tailles de maille très grandes dans les derniers anneaux.

5.5.4 Itérations non-linéaires

Nous choisissons la méthode de quasi Newton proposée pas Turek [103] : Le critère d'arrêt des itérations non-linéaires est (4.23) avec $\text{tol}_{nonlin} = 10^{-3}$. Pour chaque résolution du système linéarisé, nous stoppons les itérations lorsque le résidu relatif est inférieur à 10^{-1}, c'est à dire un gain d'une décimale. Ev (resp Ep) est l'erreur calculée sur un voisinage du premier anneau pour la vitesse (resp la pression). D'après Deuring [35], l'erreur est donnée par la formule suivante :

$$Ev + Ep \leq C(h \ln(R/s) + h^t + R^{-1})$$

où $t \in]0, 1]$ dépend de la régularité de la vitesse et la pression au voisinage de l'obstacle. Nous rappelons que s est la rayon de la boule dans laquelle nous souhaitons avoir une bonne approximation du fluide. Cette inégalité est vraie sous des hypothèses de petitesse sur $\mathcal{R}e$ et à condition que $hR \leq \mu$. Les erreurs Ev et Ep sont calculées avec la norme L^2 discrète, c'est à dire la norme l^2 divisée par la racine carrée du nombre d'inconnues. Nous pouvons nous permettre cette approximation car le maillage est quasiment uniforme sur le premier anneau. it-nl est le nombre d'itérations non-linéaires et it-fgmres est le nombre moyen d'itérations du FGMRES, c'est à dire le nombre total d'itérations du FGMRES divisé par le nombre d'itérations non-linéaires. Il n'y a pas de stabilisation en vitesse et on choisit $\mathcal{R}e = 1$. Les résultats sur la décroissance de l'erreur sont conformes à ceux présentés dans [34] et [35]. On s'aperçoit que le nombre d'itérations non linéaires croît très faiblement avec le rayon et la taille de la maille. Pour le nombre d'itérations du FGMRES, nous avons reporté les résultats dans la figure 5.5. On voit que la croissance du nombre d'itérations en fonction de R et en fonction du nombre de raffinement N tend à se stabiliser.

Afin d'étudier le comportement en fonction du nombre de Reynolds, nous abandonnons le cas test polynomial pour nous placer dans le cadre du cas test physique b). Les résultats sont présentés dans le tableau 5.6. La valeur en roman représente le nombre d'itérations non-linéaires et la valeur en italique le nombre moyen d'itération du FGMRES. Comme lors du cas linéaire, le nombre d'itérations du FGMRES croît avec le nombre de raffinements N puis

		R=4	R=8	R=16	R=32	R=64
h=1/2	Taille	312	440	568	699	952
	Ev	0.54	0.36	0.28	0.26	0.25
	Ep	2.40	2.13	1.52	1.31	1.25
	it-nl	6	7	8	9	9
	it-fgmres	13.8	16.4	18.25	22	26.6
h=1/4	Taille	2 152	3 128	4 104	5 080	7 032
	Ev	0.45	0.23	0.13	0.09	0.08
	Ep	1.84	1.16	0.64	0.48	0.45
	it-nl	7	9	10	11	11
	it-fgmres	25.7	29.2	32.4	45.3	62.6
h=1/8	Taille	16 200	23 912	31 624	39 336	47 048
	Ev	0.41	0.19	0.08	0.035	0.022
	Ep	1.51	0.78	0.31	0.19	0.16
	it-nl	7	10	11	11	11
	it-fgmres	32.3	41.5	50.2	70.2	93.72
h=1/16	Taille	131 522	167 592	249 096	310 600	372 014
	Ev	0.41	0.19	0.07	0.024	0.013
	Ep	1.40	0.67	0.23	0.13	0.14
	it-nl	7	9	10	10	10
	it-fgmres	39.1	46.5	58.4	82.7	104.7

TAB. 5.5 – Résultats sur les itérations non-linéaires, cas test polynomial, $\mathcal{R}e = 1$

stagne pour les cas les plus favorables ($\mathcal{R}e = 1$ et $R = 4, 8$ ou 16.) Pour les autres cas, il faudrait certainement raffiner davantage pour observer le même phénomène. Nous pouvons également constater la robustesse de la méthode non-linéaire par rapport aux paramètres h et R. Cette robustesse fait reporter les difficultés sur le problème non-symétrique linéaire, puisque le spectre du problème préconditionné dépend de la convection. Ceci explique qu'il nous faut un nombre d'itérations assez grands pour les cas les moins favorables.

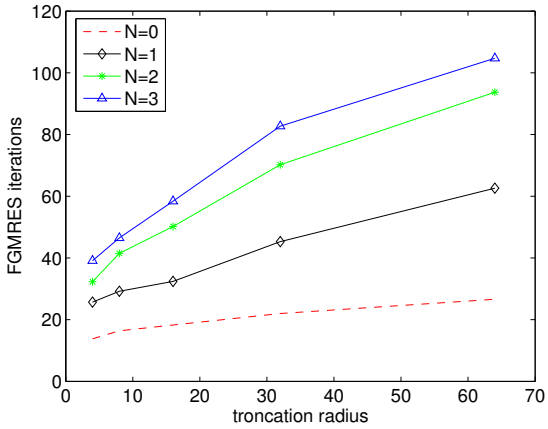


FIG. 5.5 – Nombre moyen d'itérations du FGMRES en fonction de R.

$\mathcal{R}e$	h	$R = 4$	$R = 8$	$R = 16$	$R = 32$	$R = 64$
1	1/2	5 - *15.4*	5 - *17.4*	5 - *20.8*	5 - *24.8*	5 - *33.2*
	1/4	5 - *33.8*	5 - *42*	5 - *54.6*	5 - *82.8*	5 - *137.2*
	1/8	5 - *52*	5 - *60.8*	5 - *93.6*	5 - *135.2*	5 - *257.8*
	1/16	5 - *59.6*	5 - *71*	5 - *98.2*	5 - *229.5*	5 - *356.6*
10	1/2	6 - *14.6*	6 - *18.7*	6 - *23.7*	6 - *27.3*	6 - *30.2*
	1/4	6 - *36.2*	6 - *46.8*	6 - *65.7*	6 - *78.5*	6 - *88.8*
	1/8	7 - *92.7*	7 - *163.4*	7 - *203.8*	6 - *208.3*	6 - *235.5*
	1/16	7 - *135.8*	7 - *254.8*	7 - *357.5*	7 - *412.1*	7- *581.1*

TAB. 5.6 – Itérations non-linéaires dans le cas test physique

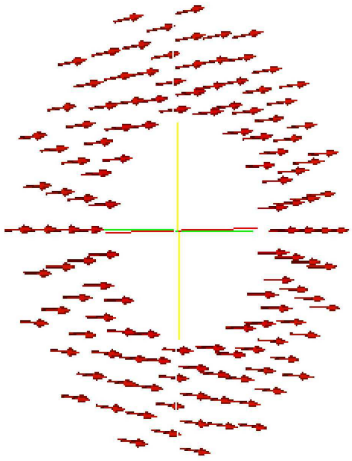
5.5.5 Représentation de la solution

On représente la solution du cas test physique. Les figures sont représentées à l'aide du logiciel ParaView [1]. Nous traçons les solutions du système adimensionné (5.2). Nous représentons le champ de vitesse sur le premier anneau dans les trois coupes usuelles du plan. La coupe (xOy) est de profil afin de pouvoir visualiser correctement le champ de vecteur. On trace ensuite les isovaleurs de la pression sur la sphère. Les figures relatives à le pression sont à vue d'oeil les mêmes pour les différents Reynolds. Nous ne tracerons donc que la figure obtenue avec $\mathcal{R}e = 1$. La vorticité est tracée également sous forme d'isovaleurs dans les trois coupes. Le graphique 5.13 représente les lignes de courant de la solution pour $\mathcal{R}e = 20$ et la figure 5.14 représente les isovaleurs pour la norme euclidienne de la vitesse. On voit très nettement ici la traînée pour différents nombres de Reynolds. Nous ne considérons pas de valeurs plus grandes du nombre de Reynolds. En effet, les itérations non-linéaires ne convergent plus (au moins en un temps raisonnable) pour $\mathcal{R}e = 50$.

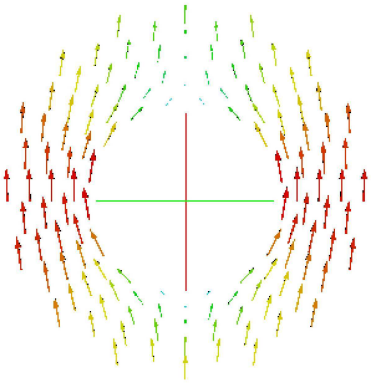
5.6 Conclusion

Notre préconditionneur s'adapte bien au problème extérieur considéré. En particulier, les itérations non linéaires semblent indépendantes du rayon de troncature et de la taille de la maille. Les résultats numériques concernant le préconditionneur sont meilleurs que ceux prévus théoriquement et donnent de bons résultats pour des nombres de Reynolds assez petits. Considérant que les itérations non-linéaires ne convergent plus pour un nombre de Reynolds égal à 50, nous pouvons considérer ces résultats comme satisfaisants.

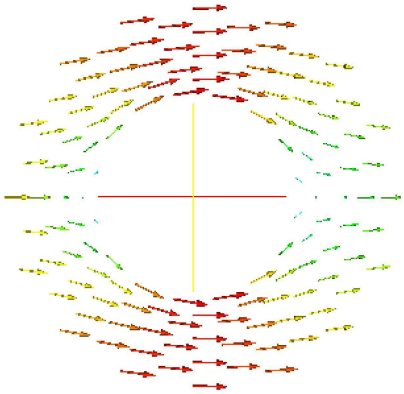
[1] ParaView est une interface graphique pour visualiser les données en deux et trois dimensions. Ce logiciel libre utilise les bibliothèques graphiques VTK. Pour plus d'information, voir http ://www.paraview.org



(a) coupe (xOy)



(b) coupe (yOz)



(c) coupe (xOz)

FIG. 5.6 – Champ de vitesse, $\mathcal{R}e = 1$

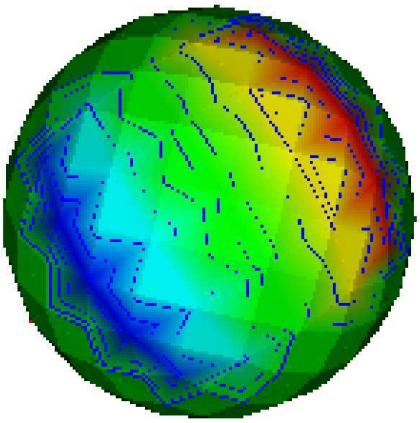
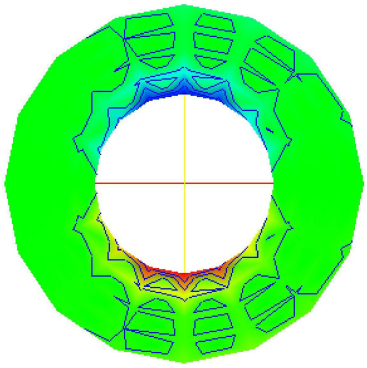
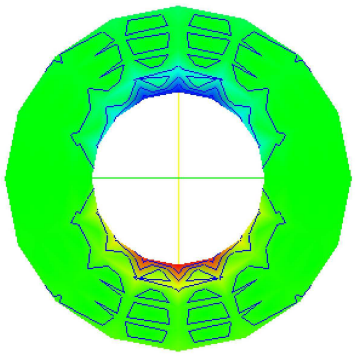


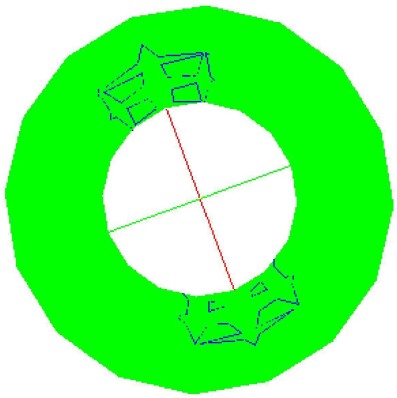
Fig. 5.7 – Pression sur la sphère



(a) coupe (xOy)

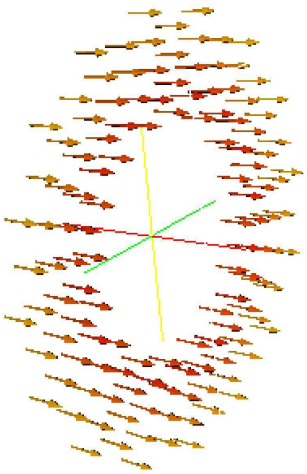


(b) coupe (yOz)

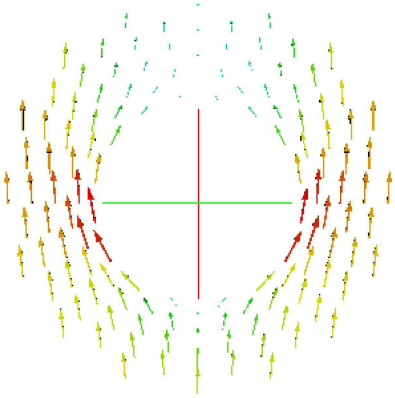


(c) oupe (xOz)

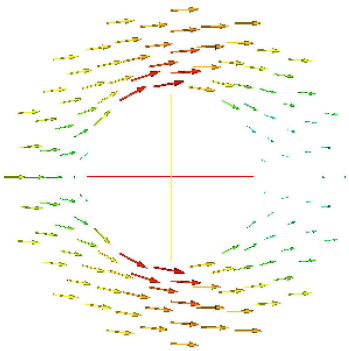
FIG. 5.8 – Vorticité, $\mathcal{R}e = 1$



(a) coupe (xOy)

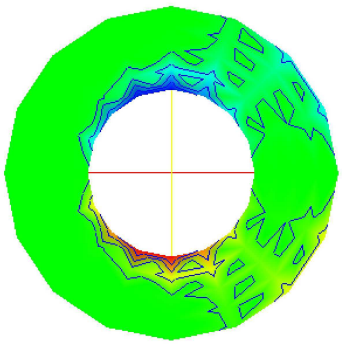


(b) coupe (yOz)

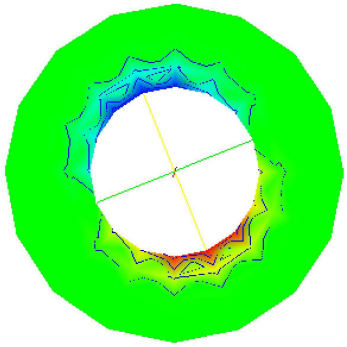


(c) coupe (xOz)

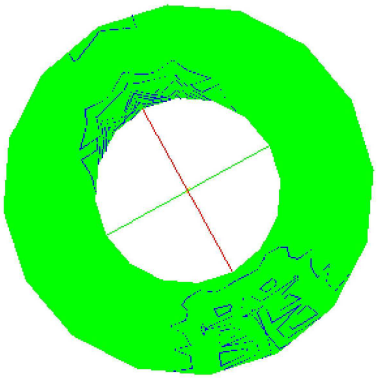
FIG. 5.9 – Champ de vitesse, $\mathcal{R}e = 5$



(a) coupe (xOy)

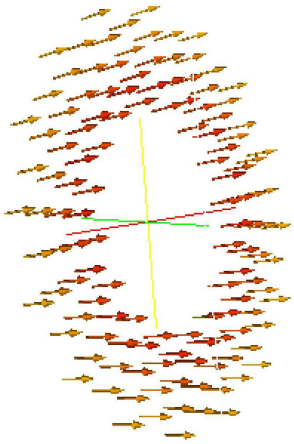


(b) coupe (yOz)

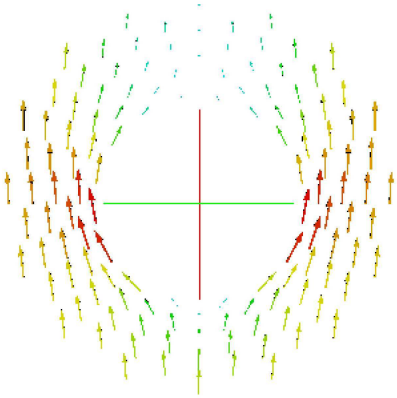


(c) oupe (xOz)

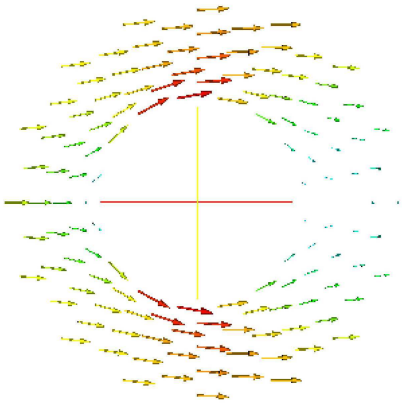
FIG. 5.10 – Vorticité, $\mathcal{R}e = 5$



(a) coupe (xOy)

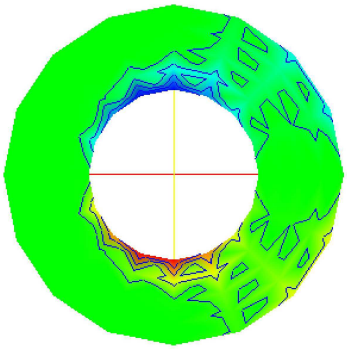


(b) coupe (yOz)

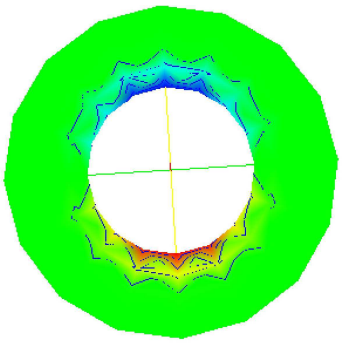


(c) coupe (xOz)

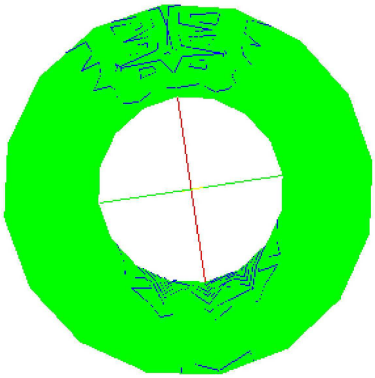
FIG. 5.11 – Champ de vitesse, $\mathcal{R}e = 10$



(a) coupe (xOy)



(b) coupe (yOz)



(c) coupe (xOz)

FIG. 5.12 – Vorticité, $\mathcal{R}e = 10$

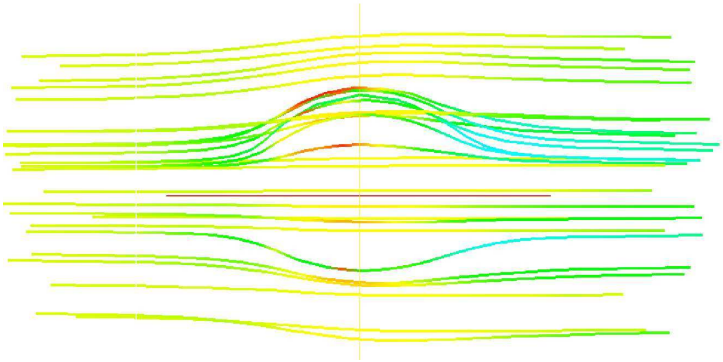
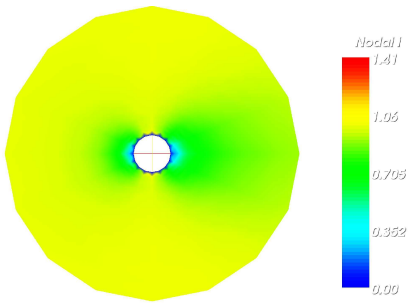
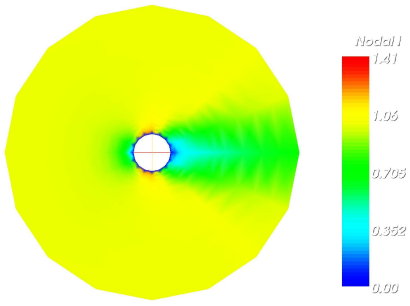


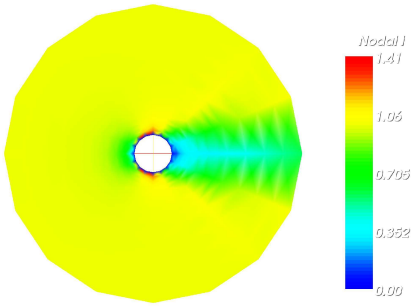
FIG. 5.13 – Représentation des lignes de courant en 3D pour $\mathcal{R}e = 20$.



(a) Re=1



(b) Re=10



(c) Re=20

FIG. 5.14 – Isovaleurs de la vitesse dans la coupe (xOz) tracées dans la boule de rayon 16

Chapitre 6

Méthode de décomposition de domaine pour le problème extérieur

Dans ce chapitre, nous allons proposer une méthode de décomposition de domaines sans recouvrement pour le problème de convection-diffusion dans un domaine extérieur.

6.1 Partition du domaine et notations

La triangulation introduite dans le chapitre 5 est basée sur la décomposition du domaine en anneaux $(U_j)_{0\leq j\leq J}$ définis en (5.4) sur lesquels le diamètre des tétraèdres de la triangulation $\mathcal{T}_{h,R}$ inclus dans U_j est d'ordre $2^j \cdot h$ pour tout $j = 0 \cdots J$ (voir figure 5.1). Nous prenons comme obstacle la sphère $\Omega := B(O,1)$. Nous appelons Γ_0 l'approximation par un polyèdre du bord de l'obstacle $\partial\Omega$, de façon à ce que tous les sommets de Γ_0 appartiennent à Ω. Nous construisons alors les polyèdres Γ_j par

$$\Gamma_j = (2^j \cdot S)\Gamma_0. \tag{6.1}$$

Le domaine Ω_j est le domaine inclus entre les polyèdres Γ_j et Γ_{j+1} et on note n_{j+1} la normale extérieure sortante à $\partial\Omega_j \setminus \Gamma_j$. Il est évident que Ω_j, Γ_j et n_{j+1} dépendent de h mais nous ne faisons pas apparaître l'indice pour alléger les notations. Nous avons donc, en faisant le lien avec les notations du chapitre 5, que le bord externe $\Gamma_{h,R}$ de notre domaine correspond à Γ_{J+1}, d'où $n_{J+1} = n^{h,R}$. Nous introduisons une triangulation régulière $\mathcal{T}_h^0$ du domaine Ω_0 et pour tout $1 \leq i \leq J$, nous déduisons de $\mathcal{T}_h^0$ une triangulation de Ω_i par l'homothétie de

rapport 2^i et de centre O. La triangulation ainsi obtenue est régulière et sa régularité est indépendante de R.

Le domaine $\Omega_{h,R} = \bigcup_{i=0}^{J} \Omega_i$, défini comme union sans recouvrement des Ω_i semble bien adapté aux méthodes de décomposition de domaine, puisqu'une telle partition assure une charge équilibrée (en terme de données) sur chaque processeur lorsqu'on parallélise le code de calcul. Chaque sous-domaine Ω_i contribue au plus à deux interfaces, ce qui simplifie les communications. De plus, sur chaque sous-domaine, les mailles ont un diamètre de même ordre, ce qui devrait simplifier la résolution des sous problèmes locaux. Nous appelons Γ la réunion de toutes les interfaces, c'est à dire

$$\Gamma = \bigcup_{j=1}^{J} \Gamma_j. \tag{6.2}$$

6.2 Décomposition de domaine

6.2.1 Formulation variationnelle

Nous allons commencer par définir les espaces de Sobolev et les espaces d'éléments finis pour les sous-domaines Ω_j et voir quelle formulation variationnelle nous pouvons définir sous ces sous-domaines. Nous introduisons d'abord simplement l'espace de Sobolev H^1 restreint aux domaines Ω_i par :

$$H^1(\Omega_i) := \{v_i = v_{|\,\Omega_i},\ v \in H^1(\Omega_{h,R})\}, \tag{6.3}$$

pour tout $i = 0, \cdots, J$. Soit $V^0_{h,R}$ l'espace d'éléments finis défini par

$$V^0_{h,R} = \{v_{h,R} \in V_{h,R},\ v_{h,R\,|\,\Gamma_0} = 0\}, \tag{6.4}$$

où $V_{h,R}$ est défini en (5.8). Nous introduisons maintenant les espaces d'éléments finis définis sur chaque sous-domaine par

$$V^{(i)} := V^0_{h,R} \cap H^1(\Omega_i), \tag{6.5}$$

et l'espace produit des ces espaces défini par

$$V := \prod_{i=0}^{J} V^{(i)}. \tag{6.6}$$

Nous définissons également les espaces des traces $\Lambda^{(i)}$ par

$$\Lambda^{(i)} := \{u^i_{h,R|\partial\Omega_i\cap\Gamma},\quad u^i_{h,R} \in V^{(i)}\}, \tag{6.7}$$

et l'espace produit Λ

$$\Lambda := \prod_{i=0}^{J} \Lambda^{(i)}. \tag{6.8}$$

Les notations avec les espaces produit signifient que nous dupliquons les degrés de libertés correspondant aux interfaces.

Nous voulons résoudre le problème variationnel suivant :

$$\begin{cases} \text{Trouver } u_{h,R} \in V^0_{h,R} \text{ tel que} \\ a_{h,R}(w_{h,R})(u_{h,R}, v_{h,R}) = \int_{\Omega_{h,R}} \varrho \, v_{h,R} \, dx \quad \forall v_{h,R} \in V^0_{h,R}, \end{cases} \tag{6.9}$$

où $a_{h,R}(w_{h,R})(u_{h,R}, v_{h,R}) = a^{(+)}_{h,R}(w_{h,R})(u_{h,R}, v_{h,R}) + a^{(-)}_{h,R}(w_{h,R})(u_{h,R}, v_{h,R})$ avec $a^{(+)}_{h,R}$ et $a^{(-)}_{h,R}$ définis en (5.12) et (5.13), et où $w_{h,R}$ est fixé dans l'espace d'éléments finis $\boldsymbol{V}_{h,R}$. Les notations sont exactement les mêmes que celles du chapitre 5. On note $\hat{a}^i_{h,R}$ la restriction de $a_{h,R}$ à chaque sous-domaine Ω_i, c'est à dire que pour $w^i_{h,R}$ fixé tel que $w^i_{h,R} = w_{h,R\,|\,\Omega_i}$,

$$\begin{aligned} \hat{a}^i_{h,R}(u^i_{h,R}, v^i_{h,R}) := & \int_{\Omega_i} (\nabla u^i_{h,R} \cdot \nabla v^i_{h,R}) \, dx \\ & + a_{1,h,R}\big(R_{1,h,R} S(w^i_{h,R}\chi_{\Omega_i}, u^i_{h,R}\chi_{\Omega_i}), R_{1,h,R} S(w^i_{h,R}\chi_{\Omega_i}, v^i_{h,R}\chi_{\Omega_i})\big) \\ & + \mathcal{R}e \int_{\Omega_i} D_1 u^i_{h,R} \cdot v^i_{h,R} \, dx + \mathcal{R}e \int_{\Omega_i} \big((w^i_{h,R} \cdot \nabla) u^i_{h,R} \cdot v^i_{h,R} + \frac{1}{2} (\operatorname{div} w^i_{h,R}) \, u^i_{h,R} \cdot v^i_{h,R} \big) \, dx \\ & + \delta_{i,J+1} \int_{\Gamma_{J+1}} (\frac{1}{R} + \frac{\mathcal{R}e}{2}) \, u^i_{h,R} v^i_{h,R} \, ds - \delta_{i,J+1} \int_{\Gamma_{J+1}} \frac{\mathcal{R}e}{2}) (w_{h,R} + e_1) \cdot n^{(h,R)} u^i_{h,R} v^i_{h,R} \, ds, \end{aligned} \tag{6.10}$$

$\forall u^i_{h,R}, v^i_{h,R} \in V^{(i)}$ pour tout $0 \leq i \leq J$ et où χ est la fonction indicatrice et δ le symbole de Kronecker.

La formulation variationnelle (6.9) correspond à la discrétisation du problème continu d'inconnue u et de second membre ϱ :

$$\begin{cases} -\triangle u + \mathcal{R}e \, ((w + e_1) \cdot \nabla) u + \dfrac{\mathcal{R}e}{2} (\nabla \cdot w) u = \varrho, \\ u_{\,|\,\Gamma_0} = 0, \ \dfrac{\partial u}{\partial n_{J+1}} = (\dfrac{1}{R} + \dfrac{\mathcal{R}e}{2}) u + \dfrac{\mathcal{R}e}{2} ((w + e_1) \cdot n) u. \end{cases} \tag{6.11}$$

La formulation variationnelle discrète de (6.11) sur chaque sous-domaine Ω_i est donnée par : Trouver $u^i_{h,R} \in V^{(i)}$ telle que,

$$a^i_{h,R}(u^i_{h,R}, v^i_{h,R}) = \int_{\Omega_i} \varrho \, v^i_{h,R} \, dx, \quad \forall v^i_{h,R} \in V^{(i)}, \tag{6.12}$$

avec

$$\forall u^i_{h,R}, v^i_{h,R} \in V^{(i)}, \quad a^i_{h,R}(u^i_{h,R}, v^i_{h,R}) := \hat{a}^i_{h,R}(u^i_{h,R}, v^i_{h,R}) + (1-\delta_{i,J+1}) \int_{\Gamma_{i+1}} \frac{\partial u^i_{h,R}}{\partial n_{i+1}} v^i_{h,R}\, dx - (1-\delta_{i,0}) \int_{\Gamma_i} \frac{\partial u^i_{h,R}}{\partial n_i} v^i_{h,R}\, dx. \quad (6.13)$$

On doit s'assurer que le problème est bien posé sur chaque interface, c'est à dire que la forme bilinéaire $a^i_{h,R}$ est coercive. On note qu'il n'y a pas lieu d'avoir une dérivée normale sur le bord intérieur du premier anneau car on a prescrit les conditions aux limites de Dirichlet.

6.2.2 Conditions aux limites à l'interface

On cherche des conditions aux limites à l'interface de Ω_i de telle manière que $a^i_{h,R}$ soit coercive et si possible qu'elles approchent les conditions aux limites artificielles exactes. On note B_i l'opérateur de conditions aux limites sur l'interface Γ_i pour $1 \leq i \leq J$. Pour l'opérateur de convection-diffusion, on rencontre deux types de conditions aux limites d'ordre 0 :

– les conditions aux limites de Robin [100] [3]

$$B_{i+1}u_i = \frac{\partial u_i}{\partial n_{i+1}} - \frac{1}{2}\mathcal{R}e\,(e_1 + w_i)\cdot n_{i+1} u_i, \quad (6.14)$$

et

$$B_i u_i = \frac{\partial u_i}{\partial n_i} + \frac{1}{2}\mathcal{R}e\,(e_1 + w_i)\cdot n_i u_i. \quad (6.15)$$

– Les conditions aux limite de Taylor d'ordre 0

$$B_{i+1}u_i = \frac{\partial u_i}{\partial n_{i+1}} - \frac{1}{2}\mathcal{R}e\ \min(0, (e_1 + w_i)\cdot n_{i+1}) u_i, \quad (6.16)$$

et

$$B_i u_i = \frac{\partial u_i}{\partial n_i} + \frac{1}{2}\mathcal{R}e\ \min(0, (e_1 + w_i)\cdot n_i) u_i. \quad (6.17)$$

On ne peut pas poser de conditions aux limites d'ordre 1, parce que la forme bilinéaire obtenue avec de telles limites n'est pas coercive. Des conditions aux limites d'ordre 2 auraient la forme suivante avec $\tau_{1,i}$ et $\tau_{2,i}$ deux vecteurs orthogonaux tangents à l'interface Γ_i et normalisés :

$$B_i u_i = \frac{\partial u_i}{\partial n_i} + \alpha_i u_i + \beta_i \frac{\partial u_i}{\partial \tau_{1,i}} + \gamma_i \frac{\partial u_i}{\partial \tau_{2,i}} + \delta_i \frac{\partial^2 u_i}{\partial \tau^2_{1,i}} + \varepsilon_i \frac{\partial^2 u_i}{\partial \tau^2_{2,i}} + \zeta_i \frac{\partial^2 u_i}{\partial \tau_{2,i} \partial \tau_{1,i}}, \quad (6.18)$$

et où $\alpha_i, \beta_i, \gamma_i, \delta_i, \varepsilon_i, \zeta_i$ sont des coefficients à déterminer. Prendre des conditions à l'interface de Taylor d'ordre 2 serait possible, mais dans ce cas, les coefficients β_i et γ_i auraient pour

dénominateur le terme $(e_1 + w_i) \cdot n_i$. Ce coefficient pouvant devenir très petit, les conditions à l'interface de Taylor d'ordre 2 donnent parfois des résultats de convergence bien pires qu'avec les conditions aux limites de Taylor d'ordre 0. Il peut même arriver que ces conditions aux limites fassent diverger l'algorithme de Schwarz sans recouvrement ([59, pp 175-176]). L'alternative proposée par Japhet [59] consiste à trouver un développement d'ordre 2 dont les coefficients minimisent le taux de convergence de l'algorithme de Schwarz : α_i est déterminé pour coïncider avec le développement de Taylor d'ordre 0. Pour les autre coefficients, il existe une formulation explicite des coefficients dans le cas d'une convection constante sur tout le domaine. Dans le cas de convection variable, il faut approcher ces coefficients par une méthode numérique. De plus, l'utilisation de conditions aux limites d'ordre 2 nécessite d'ajouter des conditions de raccord à l'intersection des sous-domaines. Adapter une telle technique en 3D n'est pas du tout aisée. On obtient déjà un problème de minimisation à 5 coefficients pour $\beta_i, \gamma_i, \delta_i, \varepsilon_i$ et ζ_i et il n'est du tout évident de trouver une méthodologie raisonnable pour minimiser le taux de convergence.

Pour ces raisons, nous n'envisageons pas d'autres conditions aux limites que des conditions aux limites d'ordre 0. Si on calcule la formulation variationnelle avec les conditions à l'interface (6.14)-(6.15), on obtiendra

$$\begin{aligned} a^i_{h,R}(u^i_{h,R}, v^i_{h,R}) := \hat{a}^i_{h,R}(u^i_{h,R}, v^i_{h,R}) - \\ \frac{\mathcal{R}e}{2}\left(\int_{\Gamma_{i+1}} u^i_{h,R} \cdot v^i_{h,R}(w + e_1) \cdot n_{i+1}\, ds - \int_{\Gamma_i} u^i_{h,R} \cdot v^i_{h,R}(w + e_1) \cdot n_i\, ds\right). \end{aligned} \tag{6.19}$$

Si on choisit les conditions aux limites (6.16)-(6.17), on aura

$$\begin{aligned} a^i_{h,R}(u^i_{h,R}, v^i_{h,R}) := \hat{a}^i_{h,R}(u^i_{h,R}, v^i_{h,R}) - \\ \frac{\mathcal{R}e}{2}\Bigg((1 - \delta_{i,J+1}) \int_{\Gamma_{i+1}} u^i_{h,R} \cdot v^i_{h,R} \max(0, (w + e_1) \cdot n_{i+1})\, ds \\ -(1 - \delta_{i,0}) \int_{\Gamma_i} u^i_{h,R} \cdot v^i_{h,R} \max(0, (w + e_1) \cdot n_i)\, ds\Bigg). \end{aligned} \tag{6.20}$$

Dans les deux cas, le problème (6.12) est bien posé dans Ω_i puisqu'on obtient la coercivité de $a^i_{h,R}$. De plus, l'algorithme de Schwarz est consistant puisque les intégrales de bord additionnelles s'annulent à la sommation. Cela revient à dire que

$$\sum_{i=0}^{J} a^i_{h,R}(u^i_{h,R}, v^i_{h,R}) = a_{h,R}(u_{h,R}, v_{h,R}). \tag{6.21}$$

Quelques tests préliminaires à l'intérieur d'une méthode FETI-1 semblent montrer qu'il y a peu de différence entre ces deux conditions aux limites. Nous utiliserons celle de Robin comme dans Toselli [100] et Achdou et al [2].

6.2.3 Méthode FETI

Soit $(\varphi_k)_{1\leq k\leq m}$ la base nodale de l'espace d'éléments finis de $V_{h,R}$ et $(\varphi_k^i)_{1\leq k\leq m_i}$ la base nodale de $V^{(i)}$. Le problème variationnel discret (6.9) est équivalent au problème algébrique suivant : Résoudre

$$\boldsymbol{A}u = l, \tag{6.22}$$

où $\boldsymbol{A}$ est la matrice d'éléments finis

$$\boldsymbol{A} := a_{h,R}(\varphi_j, \varphi_i)_{1\leq i,j\leq m},$$

et l est le vecteur défini par

$$l := \left(\int_{\Omega_{h,R}} \varrho\, \varphi_i\, dx\right)_{1\leq i\leq m}.$$

Nous appelons également Γ l'ensemble de tous les indices des noeuds de l'interface Γ. Il n'y aura pas d'ambiguïté quant à la signification de cette notation. Nous notons $I = \{I_0, \cdots, I_J\}$ les autres indices : cela signifie que les indices I_i correspondent aux noeuds internes de Ω_i pour $1 \leq i \leq J-1$, I_0 et I_J contiennent respectivement les indices des noeuds de $\overline{\Omega}_0 \setminus \Gamma_0$ et $\overline{\Omega}_J \setminus \Gamma_{J+1}$.

Considérons maintenant les problèmes variationnels (6.12) sur les sous-domaines Ω_i, pour $0 \leq i \leq J$ avec $a_{h,R}^i$ défini en (6.19). Notons $\boldsymbol{A}^{(i)}$ la matrice définie par

$$\boldsymbol{A}^{(i)} = a_{h,R}^i(\varphi_k^i, \varphi_j^i)_{1\leq j,k\leq m_i},$$

et le vecteur $l^{(i)}$ défini par

$$l^{(i)} = \left(\int_{\Omega_{h,R}} \varrho\varphi_k^i\, dx\right)_{1\leq k\leq m_i}.$$

Supposons que la base nodale $(\varphi_k^i)_{1\leq k\leq m_i}$ se décompose en $(\varphi_k^i)_{k\in I_i} \oplus (\varphi_k^i)_{k\in\{p_i,\cdots,m_i\}}$, où l'ensemble $\{p_i, \cdots, m_i\}$, qu'avec abus de notation nous continuerons à indicer avec Γ, correspond aux éléments associés au bord $\partial\Omega_i$ et où I_i correspond aux noeuds intérieurs dans leur numérotation globale. Chaque sous problème sur Ω_i peut s'écrire sous forme du système

suivant

$$\boldsymbol{A}^{(i)}u^{(i)} = \begin{pmatrix} \boldsymbol{A}_{I_i,I_i} & \boldsymbol{A}^{(i)}_{I_i,\Gamma} \\ \boldsymbol{A}^{(i)}_{\Gamma,I_i} & \boldsymbol{A}^{(i)}_{\Gamma,\Gamma} \end{pmatrix} \begin{pmatrix} u^{(i)}_{I_i} \\ u^{(i)}_{\Gamma} \end{pmatrix} = \begin{pmatrix} l^{(i)}_{I_i} \\ l^{(i)}_{\Gamma} \end{pmatrix}, \tag{6.23}$$

où

$$(\boldsymbol{A}^{(i)}_{I_i,\Gamma})_{j,k} = a^i_{h,R}(\varphi_k,\varphi_j), \quad (\boldsymbol{A}^{(i)}_{\Gamma,I_i})_{k,j} = a^i_{h,R}(\varphi_j,\varphi_k) \quad \forall k \in I_i, j \in \{p_i,\cdots,m_i\},$$

$$(\boldsymbol{A}^{(i)}_{\Gamma,\Gamma})_{j,k} = a^i_{h,R}(\varphi_k,\varphi_j), \quad \forall k,j \in \{p_i,\cdots,m_i\},$$

et

$$(\boldsymbol{A}^{(i)}_{I_i,I_i})_{j,k} = a^i_{h,R}(\varphi_k,\varphi_j), \quad \forall k,j \in I_i.$$

On note $\boldsymbol{S}_i$ le complément de Schur associé au système (6.23), c'est à dire

$$\forall 0 \leq i \leq J, \quad \boldsymbol{S}_i = \boldsymbol{A}^{(i)}_{\Gamma,\Gamma} - \boldsymbol{A}^{(i)}_{\Gamma,I_i}\boldsymbol{A}^{-1}_{I_i,I_i}\boldsymbol{A}^{(i)}_{I_i,\Gamma}. \tag{6.24}$$

On peut donc résoudre d'abord le problème sur l'interface de chaque Ω_i :

$$\boldsymbol{S}_i u^{(i)}_{\Gamma} = l^{(i)}_{\Gamma} - \boldsymbol{A}^{(i)}_{\Gamma,I_i}\boldsymbol{A}^{-1}_{I_i,I_i} l^{(i)}_{I_i}, \tag{6.25}$$

puis en déduire $u^{(i)}_{I_i}$ en résolvant l'équation relative aux noeuds internes de Ω_i :

$$\boldsymbol{A}_{I_i,I_i} u^{(i)}_{I_i} = l^{(i)}_{I_i} - \boldsymbol{A}^{(i)}_{I_i,\Gamma} u^{(i)}_{\Gamma}. \tag{6.26}$$

Posons

$$v^{(i)} = l^{(i)}_{\Gamma} - \boldsymbol{A}^{(i)}_{\Gamma,I_i}\boldsymbol{A}^{-1}_{I_i,I_i} l^{(i)}_{I_i}.$$

et $\tilde{v}_\Lambda \in \Lambda$ tel que

$$\tilde{v}_\Lambda = (v^{(0)},\cdots,v^{(J)})^T.$$

On pose $\tilde{\boldsymbol{S}} = \mathrm{diag}(\boldsymbol{S}_0,\cdots,\boldsymbol{S}_J)$. Pour la méthode FETI, on veut imposer la continuité de la solution à travers les interfaces, c'est à dire que l'on veut obtenir

$$u^{i+1}_{|\,\Gamma_{i+1}} = u^i_{|\,\Gamma_{i+1}}.$$

On utilise pour cela des multiplicateurs de Lagrange. Appelons n_Λ la taille de l'espace produit Λ et n_Γ le nombre de degrés de liberté de d'espace Γ formé de la réunion de toutes les interfaces. Pour imposer la continuité aux interfaces, nous introduisons la matrice $\boldsymbol{C}$ de taille $n_\Gamma \times n_\Lambda$ composée uniquement de $1, -1$ et 0 et correspondant à l'opération suivante :

$$\boldsymbol{C}w = \tilde{w} \quad \text{avec} \quad w = (w^{(0)},\cdots,w^{(J)})^T \in \Lambda \to \tilde{w}_{|\Gamma_i} = (w^{(i)} - w^{(i-1)})_{|\,\Gamma_i}.$$

La matrice $\boldsymbol{C}$ ainsi définie est de rang maximal. Imposer cette contrainte avec les multiplicateurs de Lagrange équivaut à résoudre le système suivant sur l'espace Λ

$$\begin{cases} \widetilde{\boldsymbol{S}} u_\Lambda + \boldsymbol{C}^T \lambda = \tilde{v}_\Lambda, \\ \boldsymbol{C} u_\Lambda = 0, \end{cases} \tag{6.27}$$

où λ est un multiplicateur de Lagrange et où $u_\Lambda = (u_\Gamma^{(0)}, \cdots, u_\Gamma^{(J)})^T \in \Lambda$ est un vecteur de taille n_Λ.

Le problème revient alors à résoudre le système

$$\boldsymbol{C}\widetilde{\boldsymbol{S}}^{-1}\boldsymbol{C}^T \lambda = \boldsymbol{C}\widetilde{\boldsymbol{S}}^{-1}\tilde{v}_\Lambda. \tag{6.28}$$

Posons pour la suite $\boldsymbol{F} = \boldsymbol{C}\tilde{\boldsymbol{S}}^{-1}\boldsymbol{C}^T$.

6.3 Choix de l'algorithme FETI

L'équation (6.28) est non symétrique et sera résolue par un GMRES préconditionné. Le travail de préconditionnement est le point crucial de la méthode FETI. Dans le cas symétrique, plusieurs types de préconditionneurs ont été utilisés : préconditionneur de type Dirichlet [47], préconditionneur de type Neumann-Neumann [70]. Toselli [100] a ensuite généralisé les résultats de Klawonn obtenus pour les problèmes symétriques au problème de convection-diffusion à l'aide des résultats de Achdou et al dans [2]. Il propose deux préconditionneurs basés sur les conditions aux limites de Robin, et nomme les deux méthodes qui en découlent FETI-1 pour le FETI classique et FETI-2 pour la méthode FETI avec ajout d'un espace grossier.

6.3.1 Préconditionneur de Toselli pour le FETI-1

Soit $(\mu_i)_{1\leq i\leq J}$ une suite d'applications linéaires définies de la façon suivante : Soit x un sommet de Γ_i, alors x appartient aux deux sous-domaines $\overline{\Omega}_{i-1}$ et $\overline{\Omega}_i$. Il existe donc deux indices j et k tels que $x = x_j \in \overline{\Omega}_{i-1}$ et $x = x_k \in \overline{\Omega}_i$. Nous définissons alors la famille des μ_i par

$$\mu_i(x) = \frac{\boldsymbol{A}_{j,j}^{(i-1)}}{\boldsymbol{A}_{j,j}^{(i-1)} + \boldsymbol{A}_{k,k}^{(i)}}. \tag{6.29}$$

On appelle $\boldsymbol{D}_i$ la matrice associée à la fonction μ_i et on pose $\boldsymbol{D} = \mathrm{diag}(\mathrm{D}_1, \cdots, \mathrm{D}_\mathrm{J})$. Le préconditionneur proposé par Toselli dans le cadre du FETI-1 est

$$\boldsymbol{M}^{-1} = (\boldsymbol{C}\boldsymbol{D}^{-1}\boldsymbol{C}^T)^{-1}\boldsymbol{C}\boldsymbol{D}^{-1}\tilde{\boldsymbol{S}}\boldsymbol{D}^{-1}\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{D}^{-1}\boldsymbol{C}^T)^{-1}. \tag{6.30}$$

Ce préconditionneur est appliqué à gauche à l'intérieur d'un GMRES sans redémarrage, ce qui signifie que l'on résout par un GMRES l'équation

$$\boldsymbol{M}^{-1}\boldsymbol{F}\lambda = \boldsymbol{M}^{-1}\boldsymbol{C}\tilde{\boldsymbol{S}}^{-1}\tilde{v}_\Gamma. \tag{6.31}$$

Multiplier par $\boldsymbol{S}_i$ nécessite d'inverser $\boldsymbol{A}_{I_i,I_i}$, ce qui correspond à résoudre un problème de convection-diffusion sur Ω_i avec des conditions aux limites de Dirichlet. On observe également que

$$\begin{pmatrix} \boldsymbol{A}_{I_i,I_i} & \boldsymbol{A}^{(i)}_{I_i,\Gamma} \\ \boldsymbol{A}^{(i)}_{\Gamma,I_i} & \boldsymbol{A}^{(i)}_{\Gamma,\Gamma} \end{pmatrix} \begin{pmatrix} u^{(i)}_{I_i} \\ u^{(i)}_\Gamma \end{pmatrix} = \begin{pmatrix} 0 \\ l^{(i)}_\Gamma \end{pmatrix} \Leftrightarrow \begin{cases} \boldsymbol{S}_i u^{(i)}_\Gamma = l^{(i)}_\Gamma \\ \boldsymbol{A}_{I_i,I_i} u^{(i)}_{I_i} = -\boldsymbol{A}^{(i)}_{I_i,\Gamma} u^{(i)}_\Gamma \end{cases}.$$

On en déduit que pour tout $v^{(i)} \in \Lambda^{(i)}$, $\boldsymbol{S}_i^{-1} v^{(i)}$ revient à projeter sur l'interface la solution du problème de convection-diffusion suivant défini avec des conditions aux limites de Robin :

$$\begin{pmatrix} \boldsymbol{A}_{I_i,I_i} & \boldsymbol{A}^{(i)}_{I_i,\Gamma} \\ \boldsymbol{A}^{(i)}_{\Gamma,I_i} & \boldsymbol{A}^{(i)}_{\Gamma,\Gamma} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ v^{(i)} \end{pmatrix}.$$

Chaque multiplication par $\boldsymbol{M}^{-1}$ nécessite la résolution d'un problème de convection-diffusion avec des conditions aux limites du Dirichlet, tandis que le produit par $\boldsymbol{F}$ implique la résolution d'un problème de convection-diffusion avec des conditions aux limites de Robin.

La matrice $\boldsymbol{D}$ permet de tenir compte du nombre de domaines qui contribue à chaque sous-interface et est une généralisation du préconditionneur introduit par Widlund et Klawonn pour des problèmes symétriques dans [70].

6.3.2 Préconditionneur simplifié pour le FETI-1

Nous avons un domaine très particulier, et la mise à l'échelle, via la matrice $\boldsymbol{D}$, n'apporte pas beaucoup d'informations pour notre problème. Si on se rapporte aux travaux de L.M. Hernandez [56], un bon préconditionneur pour le complément de Schur $\boldsymbol{C}\widetilde{\boldsymbol{S}}^{-1}\boldsymbol{C}^T$ est simplement ${\boldsymbol{C}^\dagger}^T\widetilde{\boldsymbol{S}}\boldsymbol{C}^\dagger$ où $\boldsymbol{C}^\dagger$ est le pseudo-inverse de $\boldsymbol{C}$. Ce pseudo-inverse est défini par

$$\boldsymbol{C}^\dagger = (\boldsymbol{C}\boldsymbol{C}^T)^{-1}\boldsymbol{C}^T,$$

mais comme par construction de notre domaine, nous avons $\boldsymbol{C}\boldsymbol{C}^T = 2\boldsymbol{I}$, il suffit de prendre comme préconditionneur

$$\boldsymbol{M}^{-1} = \frac{1}{4}\boldsymbol{C}\widetilde{\boldsymbol{S}}\boldsymbol{C}^T.$$

Quelques tests préliminaires ont montré que ce préconditionneur est un peu plus efficace que le préconditionneur de Toselli.

6.3.3 Préconditionnement pour le FETI-2

L'action du FETI-1 devient moins efficace lorsque le nombre de sous-domaines devient grand. On introduit donc une variante, appelée FETI-2, qui signifie que l'on doit définir un espace dit "grossier" et que nous devons résoudre le problème projeté parallèlement à cet espace grossier. D'après Achdou et al, le mauvais conditionnement de $\widetilde{\boldsymbol{S}}^{-1}$ vient du fait $a^i_{h,R}(u,u)$ devient très petit si u est constante. Pour éviter cette situation, il suffit de résoudre le problème de Robin à une constante près. Cela revient à définir cet espace grossier via un opérateur de projection $\boldsymbol{P}$ sur cet espace grossier de manière à ce que le problème de convection-diffusion projeté $\boldsymbol{P}^T\boldsymbol{F}\boldsymbol{P}$ revient à un problème de convection-diffusion résolu à une constante près. On définit un tel opérateur de la façon suivante (voir [100]) : Sur chaque sous-domaine Ω_i, nous résolvons le problème adjoint de convection-diffusion avec conditions de Robin sur le bord du domaine et avec pour second membre la fonction constante égale à 1 sur tout le domaine. Cela revient en terme d'éléments finis à résoudre

$$\left(\boldsymbol{A}^{(i)}\right)^T w^{(i)} = L^{(i)}, \tag{6.32}$$

avec

$$L^{(i)}_k = \int_{\Omega_i} \varphi^{(i)}_k \, dx, \quad \forall k = 1, \cdots, m_i.$$

$\tilde{\boldsymbol{Z}}_i$ est le vecteur défini sur l'espace $\Lambda^{(i)}$ et dont la restriction aux noeuds de $\partial\Omega_i$ correspond à la solution $w^{(i)}$ de (6.32) et $\boldsymbol{Z}_i$ le vecteur de taille $\dim(\Lambda)$ obtenu par extension de $\tilde{\boldsymbol{Z}}_i$ par O en dehors de $\partial\Omega_i$. Nous définissons maintenant la matrice

$$\boldsymbol{G} = \boldsymbol{C}\boldsymbol{Z} \text{ avec } \boldsymbol{Z} = (\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_J).$$

Nous définissions la projection $\boldsymbol{P}$ par

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{F}\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{F}.$$

On notera que

$$\boldsymbol{P}^T\boldsymbol{F}\boldsymbol{P} = \boldsymbol{F}\boldsymbol{P},$$

et que

$$\forall x \in \mathrm{Ker}(\boldsymbol{P}), \quad \boldsymbol{P}^T(x - \boldsymbol{F}x) = x - \boldsymbol{F}x.$$

Ces deux relations sont importantes pour l'implantation du code, car elles justifient qu'il n'est jamais nécessaire de calculer $\boldsymbol{P}^T$.

Le FETI-2 est en fait un GMRES projeté, qui peut être vu comme un GMRES appliqué au problème préconditionné suivant

$$\boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^T\boldsymbol{F}\lambda = \boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^T b \quad \text{avec } b := \boldsymbol{C}\widetilde{\boldsymbol{S}}^{-1}\tilde{v}_\Gamma, \tag{6.33}$$

avec pour donnée initiale $\lambda_0 = \boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{F}\boldsymbol{G})^{-1}\boldsymbol{G}^T b$. Le choix de λ_0 implique que

$$\boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^T(b - \boldsymbol{F}\lambda_0) = \boldsymbol{P}\boldsymbol{M}^{-1}(b - \boldsymbol{F}\lambda_0).$$

Un tel choix permet d'éviter la calcul de $\boldsymbol{P}^T$. De plus, le préconditionneur envoie le vecteur initial de l'espace de Krylov associé au GMRES dans $\mathrm{Im}(\boldsymbol{P})$, ce qui signifie que la multiplication par $\boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^T\boldsymbol{F}$ se fait dans l'espace $\mathrm{Im}(\boldsymbol{P})$ à chaque itération. Or, si $x \in \mathrm{Im}(\boldsymbol{P}), \boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^T\boldsymbol{F}x = \boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^T\boldsymbol{F}\boldsymbol{P}x = \boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{F}x$. Cela signifie que, dans une itération, nous calculons un produit par $\boldsymbol{P}$, un produit par $\boldsymbol{M}^{-1}$ et un produit par $\boldsymbol{F}$. La matrice carrée $\boldsymbol{G}^T\boldsymbol{F}\boldsymbol{G}$, de taille $J \times J$, est assemblée une fois pour toute en début de résolution, puis on calcule sa factorisation incomplète. A chaque étape du FETI-2, nous devons donc résoudre un problème de Dirichlet sur chaque sous-structure (préconditionneur $\boldsymbol{M}^{-1}$), deux problèmes de Robin (pour chaque produit par $\boldsymbol{F}$, un dans le complément de Schur et un dans $\boldsymbol{P}$) et un problème grossier (inversion de $\boldsymbol{G}^T\boldsymbol{F}\boldsymbol{G}$).

6.3.4 Performances connues pour la méthode FETI

On trouve dans la littérature plusieurs tests concernant les méthodes de décomposition de domaines pour le problème d'advection-diffusion. Beaucoup d'entre eux consistent à comparer plusieurs conditions aux limites [5, 101, 2, 1]. Nous souhaitons étudier la dépendance du nombre d'itérations en fonction de la taille de la maille h, du diamètre des sous-domaines H et du nombre de Reynolds $\mathcal{R}e$. Achdou et al [2] nous donnent déjà des indications sur les performances du préconditionneur de type Robin-Robin avec éventuellement ajout d'une

projection sur un espace grossier. Sur une batterie de test en 2D sans projection sur l'espace grossier, on observe que, pour une décomposition de domaine en bandes verticales, le nombre d'itérations du GMRES croît linéairement avec le nombre de sous-domaines. Par contre, le taux de convergence de la méthode est indépendant de la taille de la maille. A h fixé, on observe une croissance linéaire en $1/H$ dans le cas des fluides tournants. L'ajout d'un espace grossier rend cette croissance sous linéaire. Par contre, l'espace grossier proposé par les auteurs n'améliore pas les résultats dans les autres cas de figures. Des test préliminaires en 3D montrent que le taux de convergence est toujours indépendant de la taille de la maille, et que le nombre d'itérations reste faible pour un petit nombre de sous-domaines.

Pour la méthode FETI, les tests n'ont été fait qu'en 2D pour les problèmes non-symétriques avec des maillages conformes non structurés (des triangulations de Delaunay). Les performances dépendent des lignes de champ de la solution.

Dans le cas où les lignes de champs du vecteur d'advection ne sont pas fermées, le nombre d'itérations des méthodes FETI-1 et FETI-2 reste borné quand $\mathcal{R}e$ devient grand ou quand h devient petit : on notera cependant que le nombre d'itérations croît avec $\mathcal{R}e$ avant d'atteindre sa borne supérieure dans le cas du FETI-2 et que quand $\mathcal{R}e$ est petit, le nombre d'itérations des deux méthodes croissent légèrement lorsqu'on raffine. Par contre, le nombre d'itérations augmente linéairement quand on augmente le nombre de sous-domaines pour le FETI-1, et est stable pour le FETI-2 à condition que $\mathcal{R}e$ soit petit (de l'ordre de la dizaine en 2D !).

Lorsque le fluide est tournant (présente des lignes de champs fermées), le nombre d'itérations n'est indépendant de la taille de la maille que lorsque $\mathcal{R}e$ est petit. Par contre, pour une taille de maille fixée, le nombre d'itérations croit avec $\mathcal{R}e$, mais tend à stagner quand $\mathcal{R}e$ devient vraiment très grand. La dépendance en H est la même que dans le cas précédent.

6.4 Résultats numériques dans un domaine borné

6.4.1 Description de l'algorithme

Pour tous les résultats numériques, les matrices $\boldsymbol{A}_{I_i,I_i}$ sont inversées par un GMRES préconditionné par une factorisation incomplète. On arrête les itérations lorsque le résidu relatif est inférieur à 10^{-14}. $\tilde{\boldsymbol{S}}$ est inversé par un GMRES préconditionné par la matrice diagonale formée par les éléments diagonaux de $\boldsymbol{A}_{\Gamma,\Gamma} - \boldsymbol{A}_{\Gamma,I}\mathrm{diag}(\boldsymbol{A}_{I,I})^{-1}\boldsymbol{A}_{I,\Gamma}$. Le critère d'arrêt est d'avoir un résidu relatif inférieur à 10^{-12}. Les résolutions sont faites de manière

très précises, car les matrices à inverser sont très mal conditionnées. En effet, choisir une trop grande tolérance dans les critères d'arrêt provoque la stagnation de la méthode. La matrice $\boldsymbol{G}^T\boldsymbol{F}\boldsymbol{G}$ est assemblée explicitement en faisant le produit par chacun des vecteurs de la base canonique, puis sa factorisation LU est calculée. On arrête les itérations générales de la méthode FETI lorsque la norme euclidienne du résidu est 10^{-6} fois inférieure à la norme euclidienne du résidu initial. On notera que le critère d'arrêt ne peut pas porter sur le résidu relatif à cause du mauvais conditionnement du second membre quand le nombre d'itérations augmente.

6.4.2 Résultats dans un domaine fixe

Étant donné que notre décomposition de domaine a une géométrie particulière, nous allons essayer de faire un analogue des résultats de Toselli par rapport aux notions de lignes de champ. Nous nous plaçons d'abord dans le domaine fixe situé entre les deux sphères de centre O et de rayons $R = 1$ et $R = 2$. On le décomposera en deux, quatre puis 8 sous-domaines annulaires. La figure 6.1 montre la projection du maillage en 2D pour 4 sous-domaines. Nous appelons h la taille de la maille dans le premier anneau. Elle se déduit du

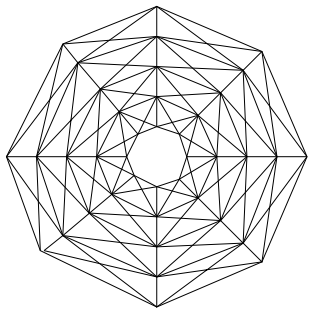


FIG. 6.1 – Maillage dans le domaine fixe

nombre de raffinements N (voir le tableau 6.1). Nous considérons que des conditions de

N	2 anneaux	4 anneaux	8 anneaux	16 anneaux
0	1/4	1/8	1/16	1/32
1	1/8	1/16	1/32	1/64
2	1/16	1/32	1/64	1/128

TAB. 6.1 – Taille de la maille

Dirichlet homogènes sont associées aux polyèdres intérieurs et extérieurs. La particularité de notre décomposition de domaine en anneaux fausse les repères par rapport aux notions de lignes de courant fermées et ouvertes. Nous allons considérer trois cas : le cas ou la convection est dans une direction fixe, la cas ou la convection est tournante (c'est à dire qu'elle est tangente aux bord des sous-domaines), et le cas ou la convection est radiale. Au niveau des conditions aux limites de Robin, cela donne des résultats très différents. Quand la convection est radiale, le terme de Robin $\mathcal{R}e\ w \cdot n$ est important et croît avec $\mathcal{R}e$. Dans le cas où elle est tournante, ce terme est nul, c'est à dire que les conditions aux limites de Robin reviennent en fait à des conditions aux limites de Neumann. Dans le cas d'une convection fixe, l'aspect Robin sera dominant aux pôles et l'aspect Neumann sera dominant à l'équateur.

– **Convection fixe**

Nous choisissons comme second membre

$$f(x) = (\|x\| - 1)(\|x\| - 2).$$

Nous choisissons $w = 0$, ce qui correspond à une convection fixe égale à e_1.

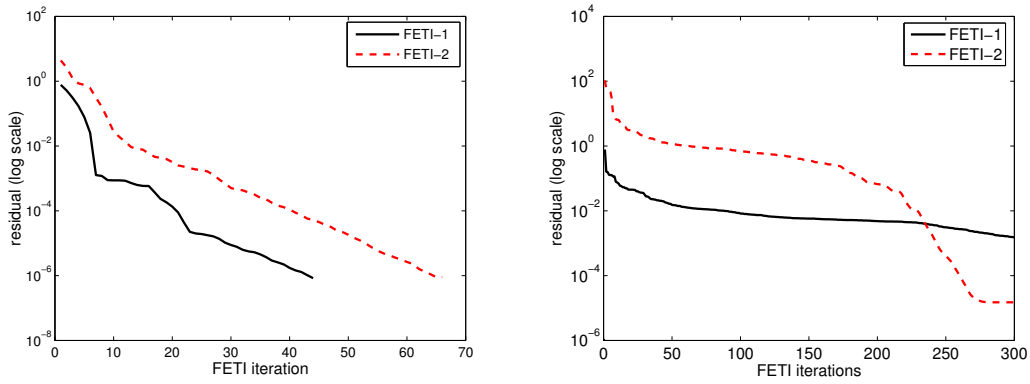
		2 anneaux			4 anneaux			8 anneaux			16 anneaux		
		$\mathcal{R}e =$			$\mathcal{R}e =$			$\mathcal{R}e =$			$\mathcal{R}e =$		
N	méthode	0.1	*1*	**10**	0.1	*1*	**10**	0.1	*1*	**10**	0.1	*1*	**10**
0	FETI-1	5	*6*	**12**	8	*16*	**38**	16	*35*	**92**	36	*75*	**197**
0	FETI-2	5	*6*	**10**	15	*18*	**30**	33	*41*	**78**	65	*77*	**181**
1	FETI-1	4	*6*	**34**	8	*17*	**79**	21	*39*	**184**	43	*86*	**>300**
1	FETI-2	4	*6*	**15**	15	*17*	**35**	33	*36*	**103**	73	*57*	**>300**
2	FETI-1	4	*6*	**69**	7	*16*	**119**	20	*38*	**182**	44	*85*	**>300**
2	FETI-2	4	*6*	**20**	13	*15*	**45**	19	*32*	**97**	65	*60*	**>300**

TAB. 6.2 – Nombre d'itération du GMRES, Convection fixe, pas de stabilisation en vitesse.

– **Convection tournante**

Cette fois-ci, nous choisissons comme vecteur d'advection w le vecteur défini par :

$$w(x) = \Lambda(x) \begin{pmatrix} y - z \\ z - x \\ x - y \end{pmatrix}$$



(a) $\mathcal{R}e = 0.1$, 16 anneaux,N=2

(b) N=1, $\mathcal{R}e = 10$, 16 anneaux

FIG. 6.2 – résidu pour une convection fixe

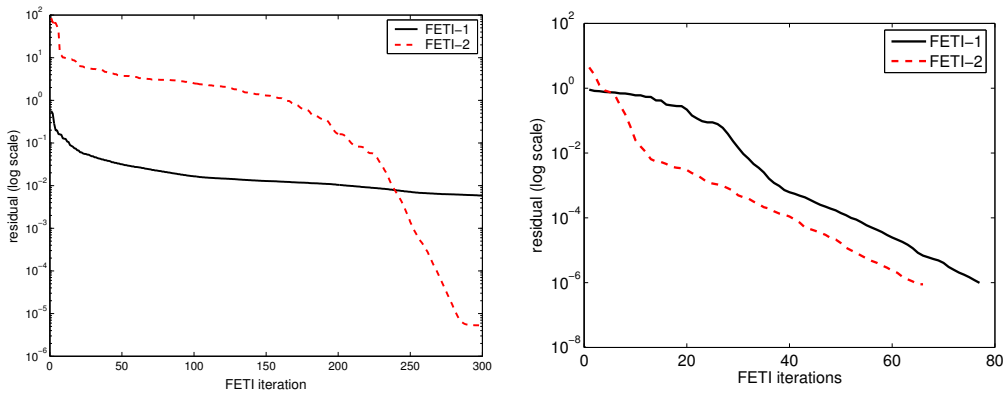
avec

$$\Lambda(x) = 40|x|^{-3}(1 - 3|x|^{-1} + 3|x|^{-2} - |x|^{-3})$$

		2 anneaux			4 anneaux			8 anneaux			16 anneaux		
		$\mathcal{R}e =$			$\mathcal{R}e =$			$\mathcal{R}e =$			$\mathcal{R}e =$		
N	méthode	0.1	*1*	**10**	0.1	*1*	**10**	0.1	*1*	**10**	0.1	*1*	**10**
0	FETI-1	5	*6*	**9**	15	*16*	**28**	36	*30*	**87**	68	*106*	**190**
0	FETI-2	5	*5*	**9**	17	*18*	**30**	37	*44*	**79**	65	*77*	**181**
1	FETI-1	4	*5*	**26**	16	*22*	**62**	39	*87*	**240**	89	*249*	**112**
1	FETI-2	4	*6*	**15**	15	*15*	**35**	32	*35*	**106**	69	*57*	**102**
2	FETI-1	4	*6*	**35**	15	*22*	**83**	35	*53*	**294**	77	*194*	**>300**
2	FETI-2	4	*5*	**21**	13	*18*	**45**	28	*32*	**102**	65	*60*	**>300**

TAB. 6.3 – Nombre d'itération du GMRES, convection tournante, pas de stabilisation en vitesse.

– **Convection radiale**



(a) $\mathcal{R}e = 10$, 16 anneaux, N=2

(b) $\mathcal{R}e = 0.1$, 16 anneaux, N=2

FIG. 6.3 – Résidu dans le cas d'une convection tournante

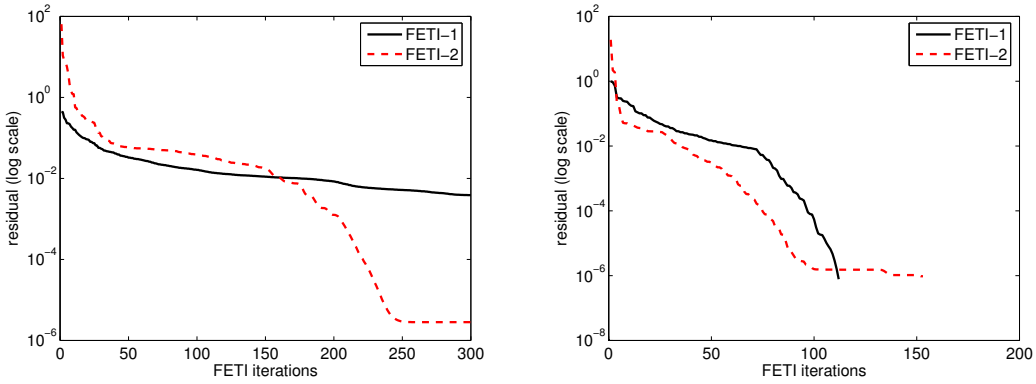
On choisit comme terme convectif

$$w(x) = (r-1)(r-2)\begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

et comme solution $u = \dfrac{1}{\mathcal{R}e} w$. On choisit le second membre f défini par $f = -\triangle u + \mathcal{R}e\, w \cdot \nabla u$.

		2 anneaux			4 anneaux			8 anneaux			16 anneaux		
		$\mathcal{R}e =$			$\mathcal{R}e =$			$\mathcal{R}e =$			$\mathcal{R}e =$		
N	méthode	0.1	*1*	**10**	0.1	*1*	**10**	0.1	*1*	**10**	0.1	*1*	**10**
0	FETI-1	5	*5*	**6**	13	*15*	**19**	29	*33*	**67**	65	*84*	**197**
0	FETI-2	5	*6*	**9**	23	*17*	**29**	31	*38*	**78**	67	*72*	**181**
1	FETI-1	4	*4*	**6**	15	*17*	**34**	35	*43*	**149**	75	*118*	**>300**
1	FETI-2	4	*6*	**15**	15	*17*	**35**	33	*35*	**101**	72	*52*	**>300**
2	FETI-1	4	*4*	**7**	15	*16*	**26**	34	*38*	**112**	80	*88*	**>300**
2	FETI-2	4	*5*	**20**	13	*15*	**45**	30	*31*	**152**	73	*59*	**>300**

TAB. 6.4 – Nombre d'itération du GMRES, convection radiale, pas de stabilisation en vitesse.



(a) $\mathcal{R}e = 10$, 16 sous-domaines, N=1

(b) $\mathcal{R}e = 10$, 8 sous-domaines, N=2

FIG. 6.4 – Convection radiale

6.4.3 Bilan des résultats numériques

Pour l'indépendance vis à vis de la taille de la maille, on obtient des résultats tout à fait semblables à ceux de la littérature dans les trois cas, c'est à dire qu'on observe une faible croissance du nombre d'itérations pour des petits nombres de Reynolds et que ce nombre d'itération est borné pour des Reynolds plus grands.

Par contre, nous n'observons pas de borne supérieure du nombre d'itérations du FETI-2 quand le nombre de sous-domaines augmente. On peut donner deux suppositions pour cela : notre nombre de sous-domaines reste petit par rapport à celui utilisé par Toselli. La seconde hypothèse est que, dans Toselli, on n'observe une stagnation du nombre de sous-domaines que pour Reynolds très petit. Cette situation n'a peut-être pas d'équivalent en 3D.

Dans aucun de nos cas, nous n'observons d'indépendance du nombre d'itérations en fonction du nombre de Reynolds. Notre géométrie et le fait que nous travaillons en 3D équivaut probablement au cas le moins favorable de Toselli.

On constate que le FETI-2 est moins sensible au cas-test considéré, parce que le FETI-2 corrige la perte de coercivité due à l'annulation de la composante normale du vecteur d'advection intervenant dans la condition aux limites de Robin. On constate d'ailleurs que l'apport du FETI-2 est moins important dans le cas d'un convection radiale. Cependant, le FETI-2 nécessite plus d'opérations à chaque étape, et est souvent plus lent que le FETI-1. On peut en tirer la conclusion suivante : Lorsque la condition à l'interface ne se réduit pas

à des conditions aux limites proches de celles de Neumann (cas d'une convection radiale, par exemple), les problèmes locaux sont bien posés, et la méthode FETI-2 n'apporte pas d'amélioration. Elle peut même donner des résultats pires (par exemple dans le tableau 6.4 pour 8 anneaux et $\mathcal{R}e = 10$ dans le cas radial). Cela peut se comprendre par la fait que nous résolvons des problèmes supplémentaires mal conditionnés et qui n'apportent rien. Par contre, quand le terme normal de la condition aux limites à l'interface s'annule (cas d'une convection tournante, par exemple), on voit que le nombre d'itérations du FETI-1 a tendance à exploser (tableau 6.4 avec $\mathcal{R}e = 1$ pour 16 anneaux ou $\mathcal{R}e = 10$ pour 8 anneaux). Dans ces cas, le FETI-2 apporte une amélioration significative. On peut également noter que dans tous les cas, pour $\mathcal{R}e = 10$ et 16 anneaux, les deux méthodes explosent complètement. On a une stagnation forte de la courbe du résidu dans ces deux cas.

On peut également noter que Toselli teste le problème d'advection diffusion avec un terme d'ordre zéro, ce qui n'est pas notre cas. Plus précisément, il travaille sur la discrétisation d'un problème du type $-\triangle u + w \cdot \nabla u + cu = f$, avec les conditions aux limites appropriées. Or ce terme a pour effet d'améliorer la coercivité. En effet, si nous ajoutons la matrice de masse à notre matrice de convection-diffusion, nous obtenons des résultats légèrement meilleurs : ils sont présentés dans le tableau 6.5 pour 1 raffinement dans le cas-test de la convection radiale. On pourra aussi remarquer que Toselli choisit toujours un second membre nul, ce qui permet

	2 anneaux			4 anneaux			8 anneaux			16 anneaux		
$\mathcal{R}e$	0.1	1	10	0.1	1	10	0.1	1	10	0.1	1	10
FETI-1	4	4	6	13	14	31	27	30	144	57	89	>300
FETI-2	4	6	16	13	15	31	29	33	86	62	70	>300

TAB. 6.5 – Avec un terme d'ordre 0

d'éviter les problèmes de mauvais conditionnement venant du second membre.

L'observation des graphes 6.2, 6.3 et 6.4 montre que FETI-2 stagne souvent vers 10^{-6} à cause du mauvais conditionnement des sous-problèmes. Or, nous avons dû déjà prendre des critères d'arrêt très stricts pour faire converger cette méthode dans la plupart des cas. Nous remarquons cependant que dans les cas ou le FETI-1 stagne fortement (voir les graphes 6.2b), 6.3a) et 6.4a)), le FETI-2 a une décroissance nettement meilleure et arriverait à convergence en moins de 300 itérations à condition d'affaiblir le critère d'arrêt.

6.5 Résultats numériques pour le problème extérieur

Pour préciser les ordres de grandeurs, le tableau récapitulatif 6.6 donne, pour le premier anneau, la taille de la maille h, le nombre de noeuds et le nombre de tétraèdres en fonction du nombre de raffinements N.

N	0	1	2	3	4
h	1/2	1/4	1/8	1/16	1/32
Nombre de noeuds	46	294	2122	16146	125986
Nombre de tétraèdres	144	1152	9216	73728	589824

TAB. 6.6 – Ordre de grandeur

a) Test avec un second membre polynomial : On choisit comme second membre la fonction définie par

$$f(x) = \left(\begin{smallmatrix}10\\2\\2\end{smallmatrix}\right)\frac{\|x\|^2-1}{\|x\|^5}.$$

La matrice $\boldsymbol{A}$ correspond à la discrétisation de l'opérateur continu $-\triangle + \mathcal{R}e\ D_1$ résolu avec des conditions de Dirichlet homogènes sur la sphère. Le critère d'arrêt pour FETI-1 et FETI-2 sont les mêmes que pour le cas borné. Les résultats pour le domaine extérieur sont très mauvais. Cette fois-ci, le FETI-2 n'améliore plus du tout les résultats, mais fait souvent bien pire. Cela signifie que ce n'est pas un problème de perte de coercivité, mais de conditionnement qui explose avec la constante de Poincaré. Le FETI-2 ajoute à chaque itération un problème avec des conditions aux limites de Robin et un problème grossier dont la construction a nécessité la résolution de plusieurs problèmes de convection-diffusion avec des conditions aux limites de Robin. Pour le domaine extérieur, le nombre de sous-domaines n'est plus le seul problème pour le mauvais conditionnement de nos matrices. On pouvait s'attendre à de tels résultats au vue des résultats précédents. En effet, on ajoute deux difficultés majeures aux problèmes traités dans le paragraphe 6.4 : la constante de Poincaré qui est proportionnelle au rayon du domaine et le rapport d'homothétie qui est très important pour notre maillage exponentiel (égal à 2). Il pourrait être intéressant par la suite de changer d'échelle pour réduire les problèmes de conditionnement dus à l'homothétie [6, 7].

b) Test avec la dernière itération non-linéaire : Dans la dernière itération non-linéaire du problème de Navier-Stokes présenté au chapitre précédent, nous devons résoudre un pro-

N	$\mathcal{R}e$	2 anneaux		3 anneaux		4 anneaux		5 anneaux	
FETI		1	2	1	2	1	2	1	2
0	1	12	10	17	17	24	22	33	34
	5	14	12	25	22	36	33	47	44
1	1	11	12	19	25	32	45	54	63
	5	26	31	54	56	83	86	114	113
2	1	10	12	18	28	38	59	82	287
	5	38	51	86	100	176	>300	246	>300

TAB. 6.7 – Méthode FETI pour le problème extérieur

blème de point selle avec pour second membre le dernier résidu non-linéaire de notre méthode de quasi-Newton. Dans le préconditionneur que nous avons introduit dans les premiers chapitres, nous remplaçons le solveur du bloc convection-diffusion, c'est à dire que nous remplaçons la méthode du GMRES préconditionnée par ILU(0) par la méthode FETI-1. Le critère d'arrêt du FGMRES(100) externe est d'obtenir un résidu relatif inférieur à 10^{-1} et nous itérons le FETI-1 jusqu'à réduire le résidu de 10^{-4}. Les problèmes locaux définis sur chaque sous-domaines sont résolus par un GMRES préconditionné par une factorisation incomplète jusqu'à un résidu relatif de 10^{-5}. Les résultats sont présentés dans le tableau 6.8. La première valeur donne le nombre d'itérations du FGMRES, et le seconde (en caractère gras) donne le nombre moyen d'itérations de la méthode FETI-1 nécessaires à la convergence. On note que le nombre d'itérations

	R	4	8	16	32
$\mathcal{R}e = 1$	N=0	16 - **15.1**	19 - **15**	12 - **17.3**	8 - **21.1**
	N=1	23 - **15.0**	29 - **15**	43 - **17**	27 - **21.0**
	N=2	38 - **18.7**	100 - **15.1**	>300 - **19.7**	>300 - **24.2**
$\mathcal{R}e = 5$	N=0	27 - **15.0**	60 - **19.1**	61 - **26.5**	31 - **25.6**
	N=1	35 - **15.1**	45 - **19.1**	38 - **16.7**	43 - **26.2**
	N=2	>300 - **16.7**	>300 - **20.6**	>300 - **31.7**	>300 - **30.5**

TAB. 6.8 – Nombre d'itérations du FETI-1/FGMRES lors de la dernière itération non-linéaire

du FETI-1 n'explose plus comme dans la cas de la convection-diffusion. Cependant, le

nombre d'itérations du FGMRES est plus important que dans le cas du préconditionneur avec ILU-GMRES sur le bloc convection-diffusion. C'est en partie dû au fait que le critère d'arrêt ne porte que sur l'erreur du problème à l'interface. Les matrices $\boldsymbol{A}_{I_i,I_i}$ ont un conditionnement qui croît avec $1/h$ (inégalités inverses habituelles), avec i (via la constante de Poincaré) et avec $\mathcal{R}e$ (pour le terme non symétrique). De ce fait, l'erreur faite sur $u_\Gamma^{(i)}$ est amplifiée lorsqu'on calcule u_{I_i}.

6.6 Conclusion

Nous avons vu que l'adaptation de la méthode FETI au problèmes de convection-diffusion tridimensionnels reste un problème très difficile. C'est la raison pour laquelle peu de tests en trois dimensions ont été réalisés pour des problèmes non symétriques. Les travaux de Jin Li [75] sur les systèmes de Stokes et de Navier-Stokes montrent également que le passage à des problèmes non symètriques est loin d'être évident. Il y a cependant quelques espoirs au vue des résultats pour les équations de Navier-Stokes à l'intérieur de la méthode de quasi-Newton et l'étude de préconditionneurs de type décomposition de domaine mérite d'être poursuivie.

Perspectives

L'objectif du travail présenté ici était d'approcher la solution des équations de Navier-Stokes stationnaires dans un voisinage de l'obstacle. Pour cela, on a maillé presque uniformément une couronne autour de l'obstacle, puis déduit le reste du maillage par une homothétie de rapport 2. Les matrices associées à des formes bilinéaires n'ont donc pas eu besoin d'être assemblées sur tout le domaine, puisqu'elles se déduisent linéairement de la matrice assemblée sur le premier anneau. Considérer un schéma implicite pour les équations stationnaires nous a empêché de réduire le stockage du bloc convection-diffusion. Une perspective d'utilisation pour de l'instationnaire aurait été de considérer les équations évolutives associées à un schéma où la non-linéarité est traitée en explicite. Cela aurait permis de plus de travailler avec des nombres de Reynolds plus grands pour lesquels la solution n'est pas nécessairement stationnaire. Ces travaux appellent à poursuivre dans deux directions distinctes : premièrement, on pourra travailler à la réduction du stockage en modifiant le maillage et en jouant sur les itérations non-linéaires ; deuxièmement, il pourra être très appréciable de réduire le temps de calcul en parallélisant le code.

Réduction du stockage grâce au maillage exponentiel

Ainsi, on peut améliorer le maillage en supprimant tout raffinement et en ne travaillant que sur le nombre de couches. Cela permet d'avoir très peu de degrés de liberté stockés au niveau du mailleur et des matrices. Dans ce cas, le rapport d'homothétie reste à définir. Dans [7], les auteurs ont montré que ce paramètre dépendait du nombre de couches et il serait intéressant d'en faire de même pour notre problème. Pour poursuivre cette idée de réduction du stockage, on peut également envisager de faire des itérations en temps et de ne considérer le problème stationnaire que comme situation d'équilibre du problème d'évolution. Si nous

considérons un schéma d'ordre 1 en temps et en espace, nous pouvons nous ramener à un schéma du type :

$$\begin{cases} \dfrac{u^{n+1}-u^n}{\Delta t} - \triangle u^{n+1} + \mathcal{R}e\ D_1 u^{n+1} + \nabla p^{n+1} = f^n - \mathcal{R}e\ u^n \cdot \nabla u^n, \\ \nabla \cdot u^{n+1} = 0, \end{cases} \tag{6.34}$$

avec les conditions aux limites appropriées sur le bord du domaine, $(u^n, p^n)^T$ et f^n étant respectivement la solution des équations de Navier-Stokes non stationnaires et son second membre au temps $n\Delta t$. On peut bien sûr considérer des schémas d'ordres supérieurs, tant que la non-linéarité reste au second membre. Dans ce cas, le système d'Oseen à résoudre à chaque itération reste le même, et le bloc convection-diffusion est associé à l'opérateur

$$\frac{1}{\Delta t} - \triangle + \mathcal{R}eD_1.$$

Le gain de stockage suffirait à compenser les contraintes liées à la condition CFL. Si nous travaillons avec des nombres de Reynolds supérieurs à ceux considérés dans cette thèse, il semble judicieux de choisir comme préconditionneur du complément de Schur celui de Silvester et al [41], c'est à dire prendre $\boldsymbol{L}_p\boldsymbol{A}^{-1}\boldsymbol{Q}$ où $\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{L}_p$ sont respectivement les matrices de convection-diffusion avec les conditions aux limites du problème, de masse et du Laplacien avec conditions aux limites de Neumann. Il nous faudra dans ce cas trouver un préconditionneur pour les problèmes de Poisson et de convection-diffusion. Les matrices n'étant pas assemblées, il n'est pas possible de calculer des factorisations incomplètes des problèmes globaux. On pourra éventuellement considérer des préconditionneurs diagonaux, de type SOR ou de type décomposition de domaine.

Optimisation en terme de parallélisation

Par ailleurs, il est possible de considérer une méthode de décomposition de domaine où chaque sous-domaine correspond à un secteur angulaire. Cela permet d'avoir un nombre de couches de notre maillage indépendant du nombre de processeurs. Dans ce manuscrit, nous n'avons considéré la méthode de décomposition de domaine que pour le bloc convection-diffusion. Or nous pouvons nous demander s'il est plus efficace de faire de la décomposition de domaine pour le problème complet. Dans [75], Li propose une méthode FETI pour les équations de Navier-Stokes. Les tests donnent des résultats satisfaisants dans la cavité entraînée

en deux dimensions jusqu'à des nombres de Reynolds allant jusqu'à 1000 lorsque la discrétisation est munie d'une stabilisation en vitesse. L'adaptation en trois dimensions nécessiterait probablement des aménagements.

Bibliographie

[1] Y. Achdou, C. Japhet, P. Le Tallec, F. Nataf, F. Rogier et M. Vidrascu, *Domain decomposition methods for non-symmetric problems*, *in* Eleventh International Conference on Domain Decomposition Methods (London, 1998), DDM.org, Augsburg, 1999, p. 3–17 (electronic).

[2] Y. Achdou, P. Le Tallec, F. Nataf et M. Vidrascu, *A domain decomposition preconditioner for an advection-diffusion problem*, Comput. Methods Appl. Mech. Engrg., 184 (2000), p. 145–170. Vistas in domain decomposition and parallel processing in computational mechanics.

[3] Y. Achdou et F. Nataf, *A Robin-Robin preconditioner for an advection-diffusion problem*, C. R. Acad. Sci. Paris Sér. I Math., 325 (1997), p. 1211–1216.

[4] M. Adams, *Algebraic multigrid methods for constrainted linear systems with applications to contact problems in solid mechanics*, Numer. Linear Algebra Appl., 0 (2004), p. 1–66.

[5] A. Alonso, R. L. Trotta et A. Valli, *Coercive domain decomposition algorithms for advection-diffusion equations and systems*, J. Comput. Appl. Math., 96 (1998), p. 51–76.

[6] F. Alouges, *Computation of the demagnetizing potential in micromagnetics using a coupled finite and infinite elements method*, ESAIM Control Optim. Calc. Var., 6 (2001), p. 629–647 (electronic).

[7] F. Alouges, J. Laminie et S. M. Mefire, *Exponential meshes and three-dimensional computation of a magnetic field*, Numer. Methods Partial Differential Equations, 19 (2003), p. 595–637.

[8] P. Arbens et R. Geus, *A comparison of solvers for large eigenvalue problems occuring in the design of resonant cavities*, Numer. Linear Algebra Appl., 6 (1999), p. 3–16.

[9] R. E. Bank et R. K. Smith, *An algebraic multilevel multigraph algorithm*, SIAM J. Sci. Comput., 23 (2002), p. 1572–1592.

[10] H. Bao, J. Bielak, O. Ghattas, L. Kallivokas, D. O'Hallaron et J. X. J. Shewchuk, *Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers*, Comput. Methods Appl. Mech. Eng.

[11] M. Bercovier et O. Pironneau, *Error estimates for finite element method solution of the stokes problem in primitive variables*, Numer. Math., 33 (1997), p. 211–224.

[12] J. Bey et A. Reusken, *On the convergence of basic iterative methods for convection-diffusion equations*, Numer. Linear Algebra Appl., 6 (1999), p. 329–352.

[13] A. Borzì et G. Borzì, *An algebraic multigrid method for a class of elliptic differential systems*, SIAM J. Sci. Comput., 25 (2003), p. 302–323.

[14] D. Braess et R. Sarazin, *An efficient smoother for the Stokes problem*, Appl. Numer. Math., 23 (1997), p. 3–19. Multilevel methods (Oberwolfach, 1995).

[15] J. H. Bramble, D. Y. Kwak et J. E. Pasciak, *Uniform convergence of multigrid V-cycle iterations for indefinite and nonsymmetric problems*, SIAM J. Numer. Anal., 31 (1994), p. 1746–1763.

[16] J. H. Bramble, J. E. Pasciak et A. T. Vassilev, *Analysis of non-overlapping domain decomposition algorithms with inexact solves*, Math. Comp., 67 (1998), p. 1–19.

[17] S. C. Brenner et L. Scott, *The mathematical theory of finite element methods*, Springer-Verlag, New York, 1994.

[18] F. Brezzi et J. Pitkaranta, *On the stabilization of finite element approximations of the stokes equations, efficient solutions of elliptic systems*, Notes Numer. Fluid Mech., 10 (1984), p. 11–19.

[19] S. F. Brezzi et M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York, 1991.

[20] A. N. Brooks et T. J. R. Hughes, *Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg., 32 (1982), p. 199–259. FENOMECH '81, Part I (Stuttgart, 1981).

[21] J. Cahouet et J. Chabard, *Some fast 3d finite element solvers for the generalized stokes problem*, Int. J. Numer. Meth. Fluids., 8 (1988), p. 869–895.

[22] C. Calgaro, P. Deuring et D. Jennequin, *Computation of 3d exterior stationary incompressible navier-stokes flowswith non-zero velocity at infinity*, *in* PAMM, vol. 3, 2003, p. 533 – 534.

[23] ——, *A preconditioner for generalized saddle point problems : Application to 3d navier stokes equations*, Numer. Meth. in P.D.E., (à paraître).

[24] S. Campbell, I. Ipsen, C. Kelley et C. Meyer, *Gmres and the minimal polynomial*, BIT, 36 (1996), p. 664–675.

[25] Z.-H. Cao, *Fast Uzawa algorithm for generalized saddle point problems*, Appl. Numer. Math., 46 (2003), p. 157–171.

[26] ——, *Fast iterative solution of stabilized Navier-Stokes systems*, Appl. Math. Comput., 157 (2004), p. 219–241.

[27] C. Carlenzoli et A. Quarteroni, *Adaptive domain decomposition methods for advection-diffusion problems*, *in* Modeling, mesh generation, and adaptive numerical methods for partial differential equations (Minneapolis, MN, 1993), vol. 75 de IMA Vol. Math. Appl., Springer, New York, 1995, p. 165–186.

[28] L. M. Carvalho, L. Giraud et G. Meurant, *Local preconditioners for two-level non-overlapping domain decomposition methods*, Numer. Linear Algebra Appl., 8 (2001), p. 207–227.

[29] T. Chacón-Rebollo et E. Chacón Vera, *Study of a non-overlapping domain decomposition method : Poisson and Stokes problems*, Appl. Numer. Math., 48 (2004), p. 169–194.

[30] J. Chehab, A. Cohen, D. Jennequin, J. Nieto, C. Roland et J. Roche, *An adaptive particle-in-cell method using multi-resolution analysis*, *in* Numerical Methodes for Hyperbolic and Kinetic Problems, I. L. in Mathematics et T. Physics, éds., vol. 7, EMS, 2005, p. 29–42.

[31] S. Chen, D. Donoho et M. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev., 43 (2001), p. 129–159.

[32] M. Crouzeix, *Étude d'une méthode de linéarisation. Résolution numérique des équations de Stokes stationnaires. Application aux équations de Navier-Stokes stationnaires*,

in Approximation et méthodes itératives de résolution d'inéquations variationnelles et de problèmes non linéaires. IRIA Cahier, No. 12, Inst. Recherche Informat. Automat., Rocquencourt, 1974, p no paging available.

[33] P. Deuring, *Finite element methods for the Stokes system in three-dimensional exterior domains*, Math. Methods Appl. Sci., 20 (1997), p. 245–269.

[34] P. Deuring, *A finite element method for computing 3d exterior stationary navier-stokes flows. i. stability and numerical test*, Rap. tech. 244, LMPA, 2005.

[35] P. Deuring, *A finite element method for computing 3d exterior stationary navier-stokes flows. ii. error estimates.*, Rap. tech. 245, LMPA, 2005.

[36] P. Deuring et S. Kračmar, *Exterior stationary Navier-Stokes flows in 3D with non-zero velocity at infinity : approximation by flows in bounded domains*, Math. Nachr., 269/270 (2004), p. 86–115.

[37] V. Dolean, F. Nataf et S. Lanteri, *Construction of interface conditions for solving the compressible Euler equations by non-overlapping domain decomposition methods*, Internat. J. Numer. Methods Fluids, 40 (2002), p. 1485–1492. LMS Workshop on Domain Decomposition Methods in Fluid Mechanics (London, 2001).

[38] H. Elman, *Preconditioning for the steady-state navier-stokes equations with low viscosity*, SIAM J. Sci. Comput., 20 (1999), p. 1299–1316.

[39] ——, *Preconditioners for saddle point problems arising in computational fluid dynamics*, Appl. Numer. Math., 43 (2002), p. 75–89.

[40] H. Elman et M. H. Schultz, *Preconditioning by fast direct methods for nonself-adjoint nonseparable elliptic equations*, SIAM J. Numer. Anal., 23 (1986), p. 44–57.

[41] H. C. Elman et D. Silvester, *Fast nonsymmetric iterations and preconditioning for navier-stokes equations*, SIAM J. Sci. Comput., 17 (1996), p. 33–46.

[42] H. C. Elman, D. J. Silvester et A. J. Wathen, *Performance and analysis of saddle point preconditioners for the discrete steady-state navier-stokes equations*, Numer. Math., 90 (2002), p. 665–688.

[43] C. Farhat, L. Crivelli et F.-X. Roux, *A transient FETI methodology for large-scale parallel implicit computations in structural mechanics*, Internat. J. Numer. Methods Engrg., 37 (1994), p. 1945–1975.

[44] C. Farhat, M. Lesoinne, P. LeTallec, K. Pierson et D. Rixen, *FETI-DP : a dual-primal unified FETI method. I. A faster alternative to the two-level FETI method*, Internat. J. Numer. Methods Engrg., 50 (2001), p. 1523–1544.

[45] C. Farhat, M. Lesoinne et K. Pierson, *A scalable dual-primal domain decomposition method*, Numer. Linear Algebra Appl., 7 (2000), p. 687–714. Preconditioning techniques for large sparse matrix problems in industrial applications (Minneapolis, MN, 1999).

[46] C. Farhat et J. Mandel, *The two-level FETI method for static and dynamic plate problems. I. An optimal iterative solver for biharmonic systems*, Comput. Methods Appl. Mech. Engrg., 155 (1998), p. 129–151.

[47] C. Farhat et F.-X. Roux, *Implicit parallel processing in structural mechanics*, Comput. Mech. Adv., 2 (1994), p 124.

[48] L. P. Franca et S. L. Frey, *Stabilized finite element methods ii. the incompressible navier-stokes equations*, Comput. Methods Appl. Mech. Eng., 99 (1992), p. 209–233.

[49] V. Girault et P. Raviart, *Finite element methods for Navier-Stokes equations*, vol. 5 de Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1986. Theory and algorithms.

[50] C. I. Goldstein, *The finite element method with nonuniform mesh sizes for unbounded domains*, Math. Comp., 36 (1981), p. 387–404.

[51] C. I. Goldstein, *Multigrid methods for elliptic problems in unbounded domains*, SIAM J. Numer. Anal., 30 (1993), p. 159–183.

[52] G. H. Golub et C. F. Van Loan, *Matrix computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, third éd., 1996.

[53] I. Graham, A. Spence et E.Vainikko, *Parallel iterative methods for navier-stokes equations and application to eigenvalue computation*, Concurrency Comput. Pract. Exp.

[54] W. Hackbush et T. Probst, *Downing gauss-seidel smoothing for convection diffusion problems*, Numer. lineear Algebra Appl., 4 (1997), p. 85–102.

[55] L. Halpern et M. Schatzman, *Artificial boundary conditions for incompressible viscous flows*, SIAM J. Math. Anal., 20 (1989), p. 308–353.

[56] L. Hernandez Ramos, *Projection obliques alternées pour les systèmes linéaires couplés*, Rap. tech. 02024, Laboratoire Jacques-Louis Lions, 2002.

[57] E. Hienard, *Méthodes d'éléments finis et moindre carrés pour la résolution des equations de Navier Stokes*, Thèse doctorat, Université d'Orsay (Paris XI), 2003.

[58] Q. Hu et J. Zou, *Two new variants of nonlinear inexact Uzawa algorithms for saddle-point problems*, Numer. Math., 93 (2002), p. 333–359.

[59] C. Japhet, *Méthode de décomposition de domaine et conditions aux limites artificielles en mécanique des fluides : méthode optimisée d'ordre 2.*, Thèse doctorat, Université de Paris XIII, 1998.

[60] C. Japhet, F. Nataf et F. Rogier, *The optimized order 2 method. application to convection-diffusion problems*, Future Generation Computer Systems FUTURE, 18 (2001), p. 17–30.

[61] C. Japhet, F. Nataf et F.-X. Roux, *Extension of a coarse grid preconditioner to non-symmetric problems*, *in* Domain decomposition methods, 10 (Boulder, CO, 1997), vol. 218 de Contemp. Math., Amer. Math. Soc., Providence, RI, 1998, p. 279–286.

[62] B. Jiang, T. Lin et L. A. Povinelli, *Large scale computation of incompressible viscous flow by least square finite element method*, Comput. Methods Appl. Mech. Eng., 114 (1994), p. 213–231.

[63] V. John et L. Tobiska, *Numerical performance of smoothers in coupled multigrid methods for the parallel solution of the incompressoble navier-stokes equations*, Int. J. Numer. Meth. Fluids., 133 (2000), p. 453–473.

[64] D. Kay, D. Loghin et A. Wathen, *A preconditioner for the steady-state navier-stokes equation*, SIAM J. Sci. Comput., 24 (2002), p. 237–256.

[65] C. Kelley, *Iterative Methods for linear and nonlinear equations*, vol. 16, SIAM Frontiers in applied mathematics, 1995.

[66] H. Kim, J. Xu et L. Zikatanov, *A multigrid method based on graph matching for convection-diffusion equations*, Numer. Lin. Algebra Appl., 10 (2003), p. 181–195.

[67] A. Klawonn, *Block-triangular preconditioners for saddle point problems with penalty term*, SIAM J. Sci. Comput., (1998), p. "172–184".

[68] ——, *An optimal preconditioner for a class of saddle point problems with a penalty term*, SIAM J. Sci. Comput., 19 (1998), p. "549–552".

[69] A. Klawonn et G. Starke, *Block triangular preconditioners for nonsymmetric saddle point problems : field-of-values analysis*, Numer. Math., 81 (1999), p. "577–594".

[70] A. Klawonn et O. B. Widlund, *FETI and Neumann-Neumann iterative substructuring methods : connections and new results*, Comm. Pure Appl. Math., 54 (2001), p. 57–90.

[71] A. Klawonn, O. B. Widlund et M. Dryja, *Dual-primal FETI methods for three-dimensional elliptic problems with heterogeneous coefficients*, SIAM J. Numer. Anal., 40 (2002), p. 159–179 (electronic).

[72] J. Koseff et R. Street, *The lid-driven cavity flow : a synthesis of qualitative and quantitative observations*, J. Fluids Engrg., 106 (1984), p. "390–398".

[73] P. Krzyzanowski, *On block preconditioners for nonsymmetric saddle point problems*, SIAM J. Sci. Comput., 23 (2001), p. "157–169".

[74] R. Lehoucq et D. Sorensen, *Deflation techniques for an implicitly restarted arnoldi iteration*, SIAM J. Matrix. Anal. Appl, 17 (1996), p. "789–821".

[75] J. Li, *A dual-primal FETI method for solving Stokes/Navier-Stokes equations*, *in* Domain decomposition methods in science and engineering, Natl. Auton. Univ. Mex., México, 2003, p. 225–231 (electronic).

[76] D. Loghin, *Analysis of preconditioned picard iterations for the navier-stokes equations*, Rap. tech. 01/10, Oxford University, Numerical Ananlysis Group, 2001.

[77] D. Loghin et A. J. Wathen, *Schur complement preconditioners for the navier-stokes equations*, Int. J. Numer. Meth. Fluids., 40 (2002), p. "403–412".

[78] G. Lube, L. Müller et H. Müller, *A new non-overlapping domain decomposition method for stabilized finite element methods applied to the non-stationary Navier-Stokes equations*, Numer. Linear Algebra Appl., 7 (2000), p. 449–472. Numerical linear algebra methods for computational fluid flow problems.

[79] G. Lube, L. Müller et F. C. Otto, *A non-overlapping domain decomposition method for the advection-diffusion problem*, Computing, 64 (2000), p. 49–68.

[80] G. H. G. M. Benzi et J. Liesen, *Numerical solution of saddle point problems.*

[81] F. Nataf et F. Rogier, *Factorization of the convection-diffusion operator and the Schwarz algorithm*, Math. Models Methods Appl. Sci., 5 (1995), p. 67–93.

[82] N.Kechkar et D. Silvester, *Analysis of locally stabilized mixed finite element methods for the stokes problem.*

[83] S. Norburn et D. Silvester, *Stabilised vs stable mixed methods for incompressible flow*, Comput. Methods Appl. Mech. Eng., 166 (1998), p. "131–141".

[84] ——, *Fourier analysis of stabilised q1-q1 mixed finite element approximation*, SIAM J. Numer. Anal., 39 (2001), p. "817–833".

[85] J. T. Oden, *Finite elements. Vol. IV*, The Texas Finite Element Series, IV, Prentice Hall Inc., Englewood Cliffs, NJ, 1983. Mathematical aspects, In collaboration with Graham F. Carey.

[86] M. Olshanskii, *A low order finite element method for the navier-stokes equations of steady incompressible flow : a stabilization issue and iterative methods*, Comput. Methods Appl. Mech. Eng., 191 (2002), p. "5515–5536".

[87] C. E. Powell et D. Silvester, *Optimal preconditioning for raviart-thomas mixed formulation of second-order elliptic problems*, Rap. tech. 399, NAREP, 2002.

[88] A. Quarteroni et A. Valli, *Numerical approximation of partial differential equations*, vol. 23 de Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1994.

[89] ——, *Domain decomposition methods for partial differential equations*, Numerical Mathematics and Scientific Computation, The Clarendon Press Oxford University Press, New York, 1999. Oxford Science Publications.

[90] T. C. Rebollo, *A term by term stabilization algorithm for finite element solution of incompressible fluid flow problems.*, Numer. Math., 79 (1998), p. "282–319".

[91] Y. Saad, *A flexible inner-outer preconditioned gmres algorithm*, SIAM J. Sci. Comput., 14 (1993), p. "461–469".

[92] J. Schöberl et W. Zulehner, *On schwarz-type smoothers for saddle point problems*, Numer. Math., 95 (2003), p. "377–399".

[93] C. Shu, L. Wang et Y. Chew, *Numerical computation of three-dimensional incompressible navier-stokes equations in primitive variable form by dq method*, Int. J. Numer. Meth. Fluids., 43 (2002), p. "345–368".

[94] D. Silvester, H. Elman, D. Kay et A. Wathen, *Efficient preconditioning of the linearized navier-stokes equations for incompressible flow*, J. Comput. Appl. Math., 128 (2001), p. "261–279".

[95] D. Silvester et A. Wathen, *Fast iterative solution of stabilised stokes systems. ii : Using general block preconditioners*, SIAM J. Numer. Anal., 31 (1994), p. "1352–1367".

[96] J. Stoer et R. Burlirsch, *Introduction to Numerical Analysis*, vol. 12 de Texts in Applied Mathematics, Springer verlag, 1993.

[97] K. Stuben, *A review of algebraic multigrid*, J. Comput. Appl. Math., 128 (2001), p. 281–309.

[98] R. Temam, *Navier-Stokes equations*, AMS Chelsea Publishing, Providence, RI, 2001. Theory and numerical analysis, Reprint of the 1984 edition.

[99] F. Thomasset, *Implementation of finite element methods for Navier-Stokes equations*, Springer Series in Computational Physics, Springer-Verlag, New York, 1981.

[100] A. Toselli, *FETI domain decomposition methods for scalar advection-diffusion problems*, Comput. Methods Appl. Mech. Engrg., 190 (2001), p. 5759–5776.

[101] R. L. Trotta, *Multidomain finite elements for advection-diffusion equations*, Appl. Numer. Math., 21 (1996), p. 91–118.

[102] S. V. Tsynkov, *Numerical solution of problems on unbounded domains. A review*, Appl. Numer. Math., 27 (1998), p. 465–532. Absorbing boundary conditions.

[103] S. Turek, *Efficient solvers for incompressible flow problems*, vol. 6 de Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin, 1999.

[104] F. Wang et J. Xu, *A crosswind block iterative method for convection-dominated problems*, SIAM J. Sci. Comput., 21 (1999), p. 620–645.

[105] A. Wathen, D.Loghin, D.Kay, H. Elman et D. Silvester, *A preconditioner for the 3d oseen equation*, 2002.

[106] A. Wathen et D. Silvester, *Fast iterative solution of stabilised stokes systems. part i : Using simple diagonal preconditioners*, SIAM J. Numer. Anal., 30 (1993), p. "630–649".

Deuxième partie

Une méthode adaptative pour le système de Vlasov-Poisson

Chapitre 7

CEMRACS 2003 : An Adaptive Particle-In-Cell method using multi-resolution analysis

Cette partie correpond l'article écrit en collaboration avec J.-P. Chehab, A. Cohen, J.J. Nieto, Ch. Roland et J.R. Roche, et publié dans Numerical Methods for Hyperbolic and Kinetic Problems .

Abstract : In this paper, we introduce a new PIC method based on an adaptive multi-resolution scheme for solving one dimensional Vlasov-Poisson equation. Our approach is based on a description of the solution by particles of unit weight and on a reconstruction of the density at each time step of the numerical scheme by an adaptive wavelet technique : the density is firstly estimated in a proper wavelet basis as a distribution function from the current empirical data and then "de-noised" by a thresholding procedure. The so-called Landau damping problem is considered for validating our method. The numerical results agree with those obtained by the classical PIC scheme, suggesting that this multi-resolution procedure could be extended with success to plasma dynamics in higher dimensions.

7.1 Introduction

The kinetic motion of a physic plasma of charged particles in which the collisions between particles are neglected is usually modeled by the Vlasov equation [A.4],

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - (E \cdot \nabla_v) f = 0, \qquad (x, v) \in \mathbb{R}^d \times \mathbb{R}^d, \tag{7.1}$$

$f(x, v, t)$ being the distribution function and E the electrostatic field.. The self-consistent field produced by the charge of the particles is

$$E_{self}(x,t) = -\nabla_x \phi_{self}(x,t) \quad \text{where} \quad -\Delta_x \phi_{self} = \rho(x,t), \quad \rho(x,t) = \int f(x,v,t)\, dv. \tag{7.2}$$

The system is closed with an initial data $f(x, v, 0) = f_0(x, v)$ and some decay conditions for the Poisson equation.

If $E = E_{self,}$ this model is clearly dispersive due to the repulsive forces and then, in order to confine the particles in a bounded domain, as usual, we consider an additional given external potential $\phi_{ext}(x)$ and rewrite E as :

$$E(x,t) = E_{self} + E_{ext} := -(\nabla_x \phi_{self} + \nabla_x \phi_{ext}). \tag{7.3}$$

In this project we are interested in the numerical resolution of the repulsive VP system (7.1)-(7.2)-(7.3) endowed with an appropriate initial data by means of the *particle in cell* (PIC) method. In the classical PIC method (see [A.6, A.12]) the initial data is approximated by set of particles, and the method aims to follow the trajectories (the characteristic curves) of these particles. In order to reflect the distribution function f_0, the initial set of particles can either be uniformly distributed and weighted with the value of f_0 at the corresponding point, or distributed randomly according to the distribution function f_0 and identically weighted. In this paper we follow the second approach.

The main difficulty in the PIC method lies in the construction of the characteristic curves

$$\begin{cases} \dfrac{dX(t)}{dt} = V(t), \\[2ex] \dfrac{dV(t)}{dt} = E(X(t), t), \end{cases}$$

because of the nonlinearity due to the self-consistent potential. This requires to rebuild the charge density ρ at each time step in order to solve the Poisson equation and to obtain the

electric field. Generally, this method gives good results with a relatively small number of particles but produces some numerical noise which prevent from describing precisely the tail of the distribution function. To overcome this drawback, it has been proposed to solve the problem by Eulerian methods [A.1] or semi Lagrangian methods [A.2]. Here, we propose to combine the PIC method with density estimation techniques based on wavelet thresholding [A.8] in order to reduce the noise level : the density $\rho(x,t)$ is estimated at time $t > 0$ by an expansion of the type

$$\sum_k \hat{c}_{J_1,k}\ \varphi_{J_1,k}(x) + \sum_{j=J_1}^{J_0} \Big(\sum_{k=0}^{2^j-1} T_\eta(\hat{d}_{j,k})\ \psi_{j,k}(x) \Big). \tag{7.4}$$

Here $\varphi_{J_1,k}$ and $\psi_{j,k}$ are the scaling functions and the wavelets, respectively. The scaling and detail coefficients $\hat{c}_{J_1,k}$ and $\hat{d}_{j,k}$ are estimated from the particle distribution at time t, and T_η is a thresholding operator at level η. Both the threshold level η and the finest resolution level J_0 are chosen depending on the number of particles.

The paper is organized as follows : in section 2 we describe how the density is estimated by wavelets, in particular we present several thresholding strategies. Then, in section 3 we present the new numerical scheme, making a comparison with the classical PIC method. In section 4, we give a numerical illustration with the simulation of the so-called Landau damping.

7.2 Density estimation by wavelet thresholding

Wavelet decompositions have been widely studied since the last two decades both from the theoretical and practical point of view. In a nutshell, these decompositions are based on a hierarchy of nested approximation spaces $(V_j)_{j\geq 0}$ which should be thought as finite element spaces of mesh size $h \sim 2^{-j}$, endowed with a nodal basis of the form $\varphi_{j,k} := 2^{j/2}\varphi(2^j \cdot -k)$. The functions $\varphi_{j,k}$ are often referred to as primal scaling functions. A projector onto V_j is of the form

$$P_j f := \sum_k c_{j,k}\varphi_{j,k} \ \text{ with } \ c_{j,k} := \langle f, \tilde{\varphi}_{j,k}\rangle, \tag{7.5}$$

where $\tilde{\varphi}_{j,k}$ are dual scaling functions. The primal and dual wavelets $\psi_{j,k}$ and $\tilde{\psi}_{j,k}$ characterize the update between two successive level of approximation in the sense that

$$P_{j+1}f - P_j f := \sum_k d_{j,k}\psi_{j,k} \ \text{ with } \ d_{j,k} := \langle f, \tilde{\psi}_{j,k}\rangle, \tag{7.6}$$

We refer to [A.7] for a classical introduction on wavelets, [A.5] for more information on their application to numerical simulation of PDE's.

In the particular context of PIC methods, we are interested in the reconstruction of the density $\rho(x,t)$ from the locations $(x_i)_{i=1,\cdots,N}$ of the particles at time t. As explained in the introduction, this reconstruction has the form 7.4 where $\hat{c}_{J_1,k}$ and $\hat{d}_{j,k}$ are estimators of the exact coefficients $c_{J_1,k} := \int \rho(x,t)\tilde{\varphi}_{J_1,k}(x)dx$ and $d_{J_1,k} := \int \rho(x,t)\tilde{\psi}_{j,k}(x)dx$ from the empirical distribution according to

$$\hat{c}_{J_1,k} := \frac{1}{N}\sum_{i=1}^{N} \varphi_{J_1,k}(x_i), \tag{7.7}$$

and

$$\hat{d}_{j,k} := \frac{1}{N}\sum_{i=1}^{N} \psi_{j,k}(x_i). \tag{7.8}$$

The interest of the thresholding procedure, in contrast to a simple projection or regularization at a fixed scale j which would compute $\sum_k \hat{c}_{j,k}\varphi_{j,k}$ is twofold : (i) the regularization level j is allowed to vary locally in the sense that the procedure might retain coefficients $d_{j,k}$ at scale j only for some k which typically corresponds to the regions where the density has sharp transitions and requires more resolution, and (ii) the local regularization level automatically adapts to the unknown amount of smoothness of the density through the thresholding procedure which only depends on the number N of samples.

Following Donoho *et al* [A.8], the maximal scale level J_0 and the threshold η depend on the number of samples according to

$$2^{J_0} \sim N^{1/2} \tag{7.9}$$

and

$$\eta \sim \sqrt{\log(N)/N}. \tag{7.10}$$

Another choice proposed in [A.8] is a threshold parameter which also depends on the scale level j according to $\eta = \eta_j = K\sqrt{j/N}$. Two techniques are generally used to threshold the details : "hard" thresholding defined by $T_\eta(y) = y\chi_{\{|y|\geq\eta\}}$ and "soft" thresholding defined by $T_\eta(y) = \mathrm{Sign}(y)\max\{0, |y| - \eta\}$. We shall precise thresholding strategies that we choose for our applications in section 4.

The density reconstruction method varies with the choice of the wavelet basis. This choice is dictated by two constraints :

1. Numerical simplicity : according to 7.7 and 7.8, the coefficients are estimated through the evaluation of dual scaling functions $\tilde{\varphi}_{J_1,k}$ and dual wavelets $\tilde{\psi}_{j,k}$ at the points x_i. It is therefore useful that these functions have a simple analytical form. In particular, high order compactly supported orthonormal wavelets cannot be used since they do not have an explicit analytical expression.
2. High order accuracy and smoothness : the primal wavelet system should have high order accuracy and smoothness in order to ensure the quality of the approximation of $\rho(x,t)$ by the expansion 7.4.

The choice of the Haar system is good with respect to the first constraint, since in this case the scaling function $\tilde{\varphi} = \varphi$ is simply the box function $\chi_{[0,1]}$, so that the estimation of a scaling coefficient $c_{j,k} = \langle \rho, \tilde{\varphi}_{j,k} \rangle$ simply amounts in counting the points falling in the interval $I_{j,k} = [2^{-j}k, 2^{-j}(k+1)[$:

$$\hat{c}_{j,k} := 2^{j/2} \frac{1}{N} \#\{i \; ; \; x_i \in I_{j,k}\}. \tag{7.11}$$

In particular, we can compute the $\hat{c}_{J_0,k}$ at the finest scale level and use the Haar transform algorithm to compute the $\hat{d}_{j,k}$ according to the classical relations :

$$\hat{c}_{j,k} = \frac{\hat{c}_{j+1,2k} + \hat{c}_{j+1,2k+1}}{\sqrt{2}} \quad \text{and} \quad \hat{d}_{j,k} = \frac{\hat{c}_{j+1,2k} - \hat{c}_{j+1,2k+1}}{\sqrt{2}}. \tag{7.12}$$

However, this choice is not good with respect to the second constraint since piecewise constant functions are low order accurate. In order to fix this defect, while preserving the numerical simplicity of the method, we propose to use a higher order (third order) reconstruction still based on the box function $\chi_{[0,1]}$ as $\tilde{\varphi}$, as proposed by Ami Harten in [A.11]. This means that the coefficients $\hat{c}_{j,k}$ are still defined by 7.11, but the relation between the approximation and detail coefficients $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ is modified according to

$$\hat{d}_{j,k} = \frac{1}{\sqrt{2}} \hat{c}_{j+1,2k} - \hat{c}_{j,k} - \frac{1}{8}(\hat{c}_{j,k-1} - \hat{c}_{j,k+1}). \tag{7.13}$$

This is an instance of the so-called *lifting scheme* introduced in [A.14]. Using this relation, we estimate all the coefficients $c_{j,k}$ and $d_{j,k}$ for $J = J_1, \cdot, J_0 - 1$ and we apply the thresholding operator T_η to the estimated coefficients $\hat{d}_{j,k}$.

It should be remarked that the primal scaling functions $\varphi_{J_1,k}$ and wavelets $\psi_{j,k}$ do not have an explicit analytical expression, in contrast to the dual scaling functions and wavelets. However, we can reconstruct the estimator 7.4 at arbitrarily fine resolution by applying the

reconstruction formulae

$$\hat{c}_{j+1,2k} = \sqrt{2}[\hat{c}_{j,k} - \frac{1}{8}(\hat{c}_{j,k-1} - \hat{c}_{j,k+1}) + T_\eta(\hat{d}_{j,k})], \tag{7.14}$$

and

$$\hat{c}_{j+1,2k+1} = \sqrt{2}\hat{c}_{j,k} - \hat{c}_{j+1,2k}. \tag{7.15}$$

It is also possible to construct wavelet-like multiscale decompositions where both the dual and primal functions have a simple analytical expression, based on the quasi-interpolation operator

$$P_j f := \sum_k c_{j,k}\varphi_{j,k}, \quad c_{j,k} = \langle f, \varphi_{j,k}\rangle, \tag{7.16}$$

where $\varphi = (1 - |x|)_+$ is the classical hat function. We therefore estimate the coefficients by

$$\hat{c}_{j,k} := \frac{1}{N}\sum_{i=1}^{N} \varphi_{j,k}(X_i), \tag{7.17}$$

at the finest scale $j = J_0$ and derive them recursively at coarser levels by the formula

$$\hat{c}_{j,k} = \frac{1}{\sqrt{2}}\hat{c}_{j+1,2k} + \frac{1}{2\sqrt{2}}(\hat{c}_{j+1,2k+1} + \hat{c}_{j+1,2k-1}).$$

In this case, the detail components at level j reads

$$(P_{j+1} - P_j)f = \sum_k [\hat{d}^0_{j,k}\varphi_{j+1,2k} + \hat{d}^+_{j,k}\varphi_{j+1,2k+1} + \hat{d}^-_{j,k}\varphi_{j+1,2k-1}], \tag{7.18}$$

with

$$\begin{aligned}
\hat{d}^0_{j,k} &:= \tfrac{\sqrt{2}-1}{\sqrt{2}}\hat{c}_{j+1,2k} - \tfrac{1}{2\sqrt{2}}\hat{c}_{j+1,2k+1} - \tfrac{1}{2\sqrt{2}}\hat{c}_{j+1,2k-1},\\
\hat{d}^+_{j,k} &:= \tfrac{2\sqrt{2}-1}{4\sqrt{2}}\hat{c}_{j+1,2k+1} - \tfrac{1}{2\sqrt{2}}\hat{c}_{j+1,2k} - \tfrac{1}{4\sqrt{2}}\hat{c}_{j+1,2k-1},\\
\hat{d}^-_{j,k} &:= \tfrac{2\sqrt{2}-1}{4\sqrt{2}}\hat{c}_{j+1,2k-1} - \tfrac{1}{2\sqrt{2}}\hat{c}_{j+1,2k} - \tfrac{1}{4\sqrt{2}}\hat{c}_{j+1,2k}.
\end{aligned}$$

The triplet $(d^0_{j,k}, d^+_{j,k}, d^-_{j,k})$ plays the role of the wavelet coefficient and it is jointly thresholded in order to preserve the density mass.

In the sequel of the paper, we shall denote W0 for the first algorithm based on the lifting scheme we have described and W1 for the second algorithm based on the quasi-interpolation operator. In the numerical scheme, we apply these algorithms and reconstruct the denoised density at the finest level J_0 on which we apply the Poisson solver to derive the electric field.

7.3 Numerical schemes

We present here the new scheme we introduce in the paper (PICONU[1]) as a modification of the classical PIC method which will be used to compare the numerical results. Of course, the considered Vlasov-Poisson equation is one dimensional in space and in velocity, so we can write a formal expression using the fundamental solution of the Poisson equation. Indeed, we have

$$-\triangle\Phi_{self} = \rho \Leftrightarrow \Phi_{self} = \frac{1}{2}|x| * \rho \Leftrightarrow E_{self} = \frac{1}{2}\frac{x}{|x|} * \rho.$$

On the other hand, if we denote $X_i(t)$ the position of the ith particle at time t for $i = 1 \dots N$, the density is

$$\rho(x,t) = \frac{1}{N}\Big(\sum_{i=1}^{N} \delta_{X_i(t)}\Big),$$

where δ_ξ stands for the Dirac measure at point ξ. Hence, the self-consistent field E_{self} can be computed by the following formula (written in general dimension d)

$$E_{self}(x,t) = \frac{1}{2N}\Big(\sum_{i=1}^{N} \frac{x - X_i}{|x - X_i|^d}\Big).$$

However, we underline that, in the practical point of view, we can not proceed in such a way in higher dimensions ($d > 1$) due to the singularity of the Green kernel, and that the adaptive method proposed in this paper is aimed at being extended, e.g., to 2-D Vlasov-Poisson problem.

7.3.1 The PIC method

The PIC method consists in following the track to particles with position X_i and velocity V_i along the characteristic curves

$$\begin{aligned}
&\frac{dX_i(t)}{dt} = V_i(t),\\
&\frac{dV_i(t)}{dt} = E(X_i(t),t),\\
&X_i(0) = x_i,\ V_i(0) = v_i.
\end{aligned}$$

Let f_0 be the initial distribution, the distribution at time $t = T$ is computed as follows :

[1] which stands in French for PIC Ondelettes NUmérique

- Initialization : build (x_i, v_i), N pair of random variables drawn of the initial distribution f_0.
- Time marching scheme : the (nonlinear) characteristic equation is split and integrated as follows : set $\delta t = \dfrac{T}{N_{max}}$ where N_{max} is the number of time steps, then, for $n = 0, \cdots$ c

$$V^{n+1/2} = V^n + \frac{\delta t}{2} E^n(X^n) \tag{7.19a}$$

$$X^{n+1} = X^n + \delta t\, V^{n+1/2} \tag{7.19b}$$

$$\texttt{Build}\ \rho^{n+1} \tag{7.19c}$$

$$\texttt{Solve}\ -\Delta_h \phi^{n+1} = \rho^{n+1} \tag{7.19d}$$

$$E^{n+1}(X^{n+1}) = \nabla_h \phi^{n+1}(X^{n+1}) \tag{7.19e}$$

$$V^{n+1} = V^{n+1/2} + \frac{\delta t}{2}\, E^{n+1}(X^{n+1}) \tag{7.19f}$$

- Build f^{Nmax} by interpolation.

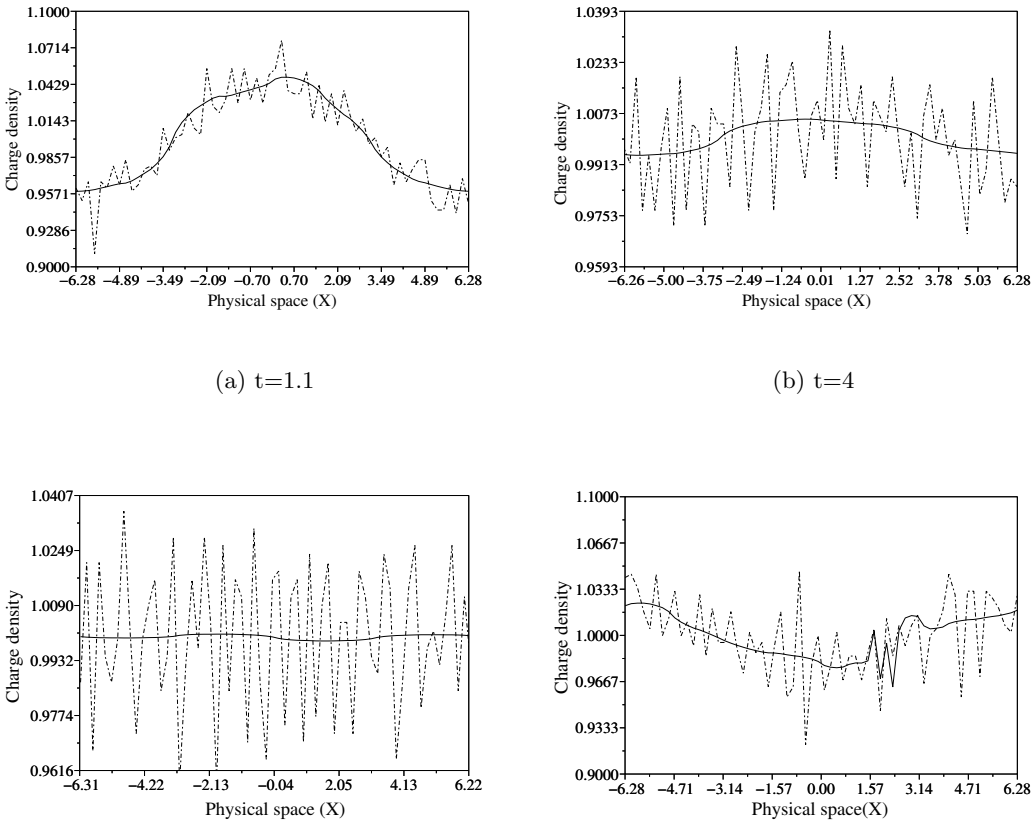
In step 7.19c, we must compute the charge density ρ on the discrete grid points. Two classical methods are Nearest Grid Point (NGP) and Cloud In Cell (CIC) : they consist on $P0$ and $P1$ interpolations respectively. According to Birdsall and Langdon in [A.3, p.19-23], CIC reduces the noise relative to the NGP. Higher order techniques could be used too, some of them consist on quadratic and cubic spline interpolations. In our numerical results, we will compare our schemes to a PIC method with CIC and NGP density reconstruction.

7.3.2 The adaptive scheme (PICONU)

The PICONU scheme differs from the classical PIC method in the step (7.19c). The density is computed by Donoho's technique described in section 7.2. For this method, we have to select the finest and the coarsest resolution level (J_0 and J_1), the threshold and the mesh size for the Poisson equation (7.19d). In the classical density estimation, the noise appears when the mesh size is locally too small. We expect to find the "good threshold" which refines the density mesh only in the region where there are a lot of particles.

7.4 Numerical results

The numerical results presented hereafter were obtained with SCILAB, the (free) numerical software of the INRIA [A.13]. As a validation of our scheme, we consider the simulation of



(a) t=1.1

(b) t=4

(c) t=6.1

(d) t=7.4

FIG. 7.1 – Landau damping with $\alpha = 0.1$. Charge density computed with 26000 particles, in solid line : the PICONU method (W0), in dashed line : the classical PIC (NGP).

the so-called Landau damping. This is indeed a significant numerical test, due to its difficulty in simulating, and it has been considered by several authors for validating a code, see, e.g., [A.3, A.9] and references therein. This test consists in the observation of the decay rate of the electrostatic energy obtained when the initial distribution is the perturbed Maxwellian distribution defined by

$$\forall (x,v) \in [-\frac{L}{2}, \frac{L}{2}] \times \mathbb{R}, \quad f_0(x,v) = \frac{1}{\sqrt{2\pi}} \exp\left(-v^2/2{v_{th}}^2\right)(1 + \alpha \cos(kx))$$

where v_{th} is the thermal mean velocity, $L = \dfrac{2\pi}{k}$ and k, α are positive constants with $\alpha \ll 1$. Let us recall that if E denotes the electric field, the electrostatic energy is defined by $\displaystyle\int_{-L/2}^{L/2} E^2\, dx$.

We consider that the tail of the distribution does not contribute to the problem for $|v| > v_{max}$ for some v_{max} large enough. We will choose $v_{th} = 1$ and $v_{max} = 6$.
More precisely, one must observe numerically that

– The decay rate of electrostatic energy defines a line of director coefficient

$$\gamma_L = \sqrt{\frac{\pi}{8}} k^{-3} \exp(-\frac{1}{2k^2} - \frac{3}{2}),$$

– Oscillations frequency of the electrostatic energy must be

$$\omega^2 = 1 + 3k^2.$$

The Landau damping is very sensitive to the initial distribution and we follow [A.3] using for that purpose the so-called "quiet start" initialization for α small enough. For all the tests, we do vary only α taking as fixed value $k = 0.5$.

7.4.1 Comparison between NGP and W0.

As a reference for validating our new scheme, we shall compare the results obtained with the classical PIC method where the charge density is computed using the NGP technique described in [A.3, pp 21,22] to the wavelets build from Haar system. In the finest resolution level J_0 is equal to the coarsest one J_1, these two methods are equivalent. In the figures 7.1,7.2 and 7.3 we plotted the charge density, the electrostatic field and the discrete electrostatic energy of the plasma. The graph of the electrostatic energy is in a log-scale and the line corresponds to the theoretical decay of director coefficient γ_L.

We used the following parameters : the time step δt is chosen equal to 0.1. The threshold is the one given in (7.10) and more precisely, we choose $\eta = 0.5 \times \sqrt{j/N}$. The finest and coarsest resolution levels are 6 and 2, this corresponds to about 800 particles per cell. Then we use NGP on a grid of 2^6 intervals.

The NGP method requires a high number of particles per cell. The only way to reduce it is to increase the accuracy of the interpolation. The wavelets, based on the same interpolation, inherit the same problem. However, we observe that wavelets allow to reduce efficiently the noise of the method. This is obvious in figure 7.3 and 7.4 looking at the minima of electrostatic energy : the local minima are obtained theoretically when the electrostatic field vanishes and numerically, these minima have the same magnitude than the discrete L^2-norm of the noise. Furthermore, we observe that the charge density and the electrostatic field are smoother than

in the simple $P0$ interpolation case. Whatever noisy is the charge density, the electrostatic field is nearly smooth. It is due to the particular case of the one dimensional integration which has a smoothing effect. In higher dimension, the smoothness will take a greater importance and the use of wavelets should be more pertinent.

7.4.2 Comparison between CIC and W1.

The simulation parameters are chosen as follows : the highest level of resolution equals 7 and the number of particles equals 10 000. It implies that there is about 80 particles per cell. The time step δt equals 0.1. For the density estimation using wavelets, the coarsest resolution level is equal to 3 and the threshold is this one prescribed by Donoho, that is $K \times \sqrt{j/N}$. The coefficient R_T gives the rate of thresholded coefficients that is the ratio of the mean value of thresholded coefficients at each time step. The figure 7.5 gives the electrostatic energy computed with the classical PIC (CIC). We observe that the classical PIC fails when α becomes small. On the contrary, the adaptive method gives some better results (see figure 7.6). A finer analysis of the threshold shows that all the coefficients are thresholded on the two finest grid. This is a natural consequence of the landau test : the distribution of particles corresponds to a small perturbation of the uniform distribution. Since we use only a first order reconstruction, the charge density is less smooth than in the case of $W0$.

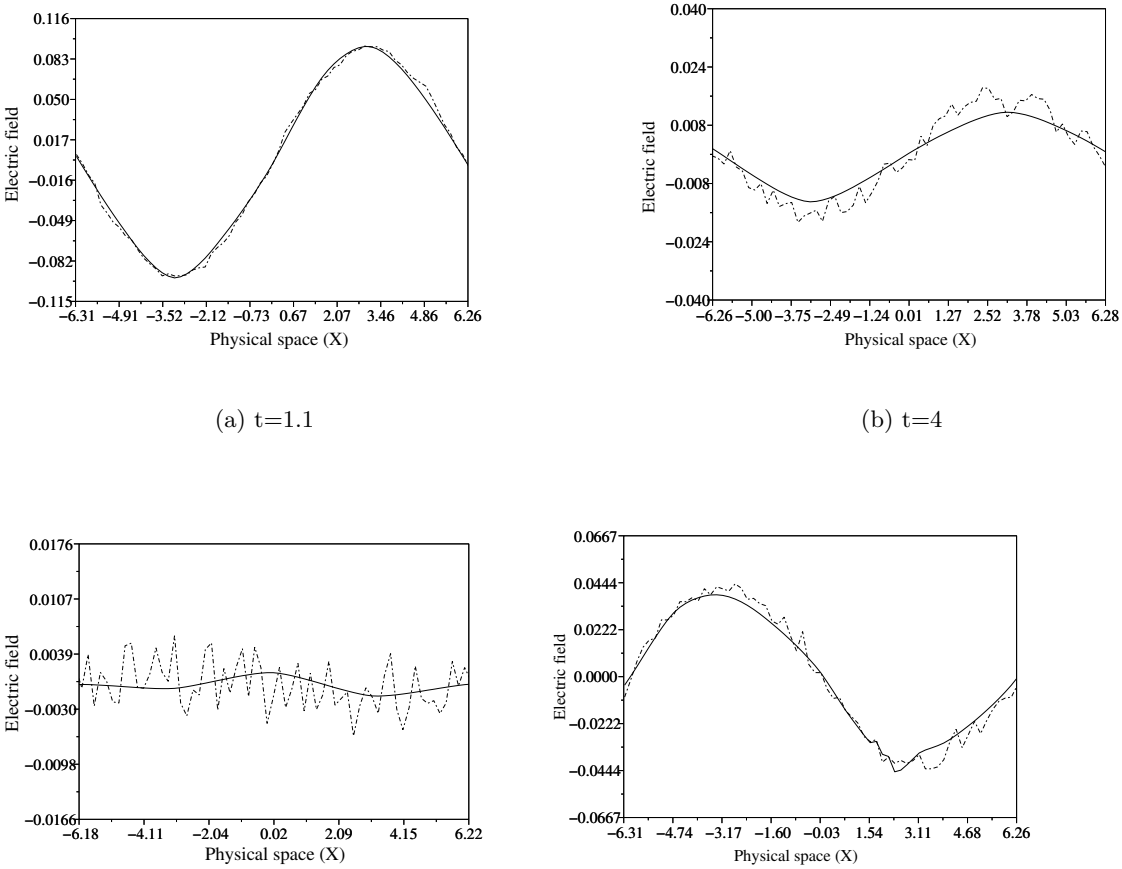
7.5 Concluding remarks and perspectives

The results presented in this paper show that the adaptive wavelet reconstruction of the density for the Vlasov-Poisson equation is a promising approach to solve such plasma dynamics in a Lagrangian framework even though the choice of appropriated wavelets must still be discussed. The W0 wavelets are not completely satisfying because they require a high number of particles. Moreover, these methods are unable to verify the landau damping test for small perturbation magnitude α. We are interested in the numerical simulation where there are less than 100 particles per cell. The W1 wavelets satisfy this condition but do not smooth the density. The results proved that threshold helps to find the appropriate (adaptive) mesh and it should become crucial in tests where particles are very dispersed. Moreover, highest order reconstruction is particularly efficient to reduce the noise. A compromise has to be found between the reconstruction and accuracy order which minimize the computational time.

We have considered here one dimensional Vlasov-Poisson but our approach will be extended in a near future to higher dimensional problems for which the Eulerian framework becomes more costly in terms of CPU time since large numbers of grid points must be used in that case.

Acknowledgements

The authors are very greatful to Prof. Eric Sonnendrücker for the numerous discussions we had with him and for the precious advise he gave us.



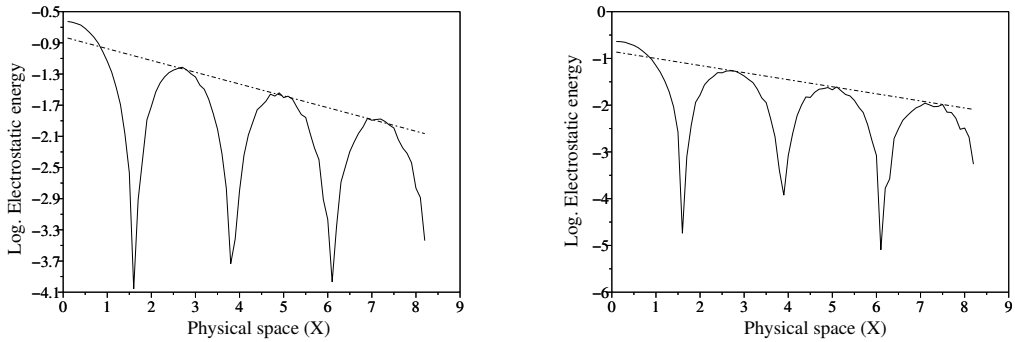
(a) t=1.1

(b) t=4

(c) t=6.1

(d) t=7.4

FIG. 7.2 – Landau damping with $\alpha = 0.1$. Electrostatic field computed with 26000 particles, in solid line : the PICONU method(W0), in dashed line :the classical PIC (NGP).



(a) Classical PIC (NGP)

(b) PICONU method (W0)

FIG. 7.3 – Landau damping with $\alpha = 0.1, k = 0.5$. Electrostatic energy

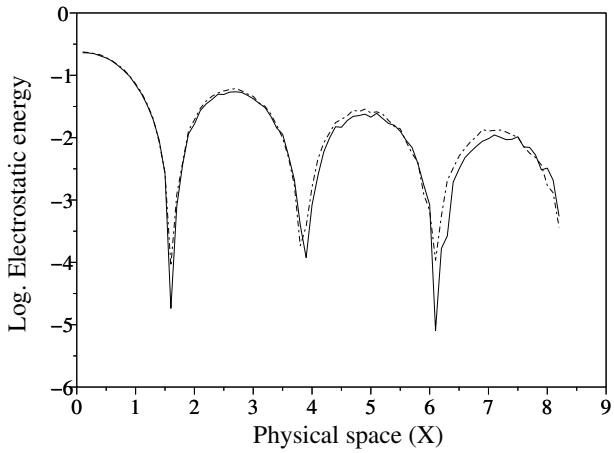
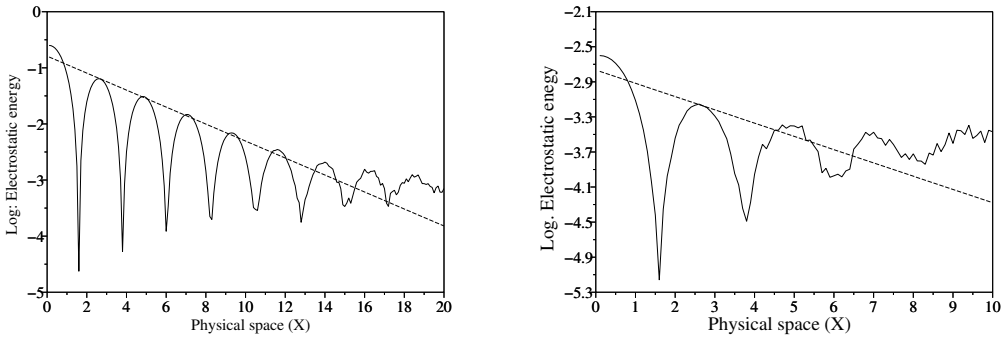
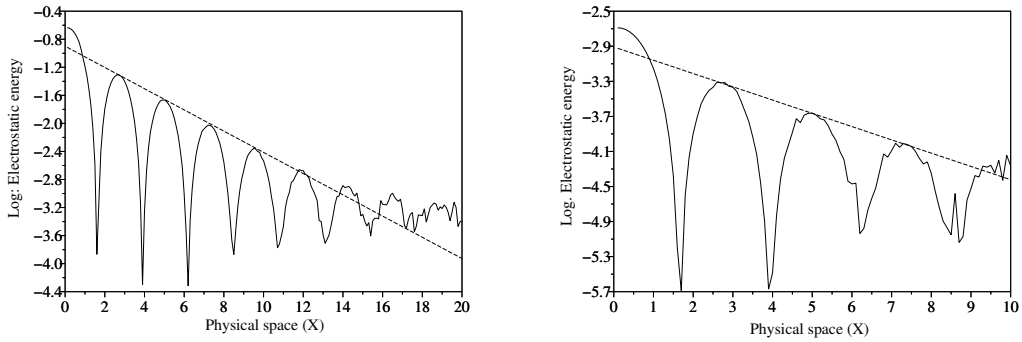


FIG. 7.4 – Superposition of electrostatic energy with 26000 particles. In solid line : damping obtain with PICONU method (W0), In dashed line : classical PIC method (NGP).



(a) $\alpha = 0.1$ (b) $\alpha = 0.01$

FIG. 7.5 – linear Landau damping with classical PIC (CIC). 10000 particles, 128 cells.



(a) $\alpha = 0.1, R_T = 13\%, K = 0.5$ (b) $\alpha = 0.01, R_T = 96\%, K = 1.5$

FIG. 7.6 – Linear landau damping, PICONU (W1), 10000 particles

Bibliographie

[A.1] T.D. Arber, R.G.L. Vann, A critical comparison of Eulerian-Grid-Based Vlasov solvers, in J. Compt. Phys. 180, pp 339-357, 2002.

[A.2] Besse, N., Sonnendrücker, E. Semi-Lagrangian schemes for the Vlasov equation on an unstructured mesh of phase space. J. Comput. Phys. 191, No.2, 341-376, 2003.

[A.3] Birdsall C.K., Langdon, A.B., *Plasma Physics via Computer Simulation*, Institute of Physics Publishing Bristol and Philadelphia, 1991.

[A.4] Chapman S., Cowling T.G., *The mathematical theory of non-uniform gases*, Cambridge University Press, 1970.

[A.5] Cohen A., Numerical Analysis of Wavelet Methods, Elsevier, North-Holland, 2003.

[A.6] Cottet G.-H., Raviart P.-A., Particle methods for the one-dimensional Vlasov-Poisson equations, Siam J. Numer. Anal., vol 21, no. 1, 1984.

[A.7] Daubechies I., Ten Lectures on Wavelets, SIAM, 1992.

[A.8] Donoho D.L., Johnstone I.M., Kerkyacharian G., Picard, D., *Density estimation by wavelet thresholding.* Ann. Statist. **24**, no. 2, 508–539, 1996.

[A.9] Filbet F., *Contribution à l'analyse et la simulation numérique de l'équation de Vlasov*, Thèse Université Henri Poincaré, Nancy, France, 2001.

[A.10] Filbet F., Sonnendrücker E., Bertrand P., Conservative numerical schemes for the Vlasov equation, J. Comput. Phys. 172, pp 166-187, 2001

[A.11] Harten A., Discrete multiresolution and generalized wavelets, J. Appl. Num. Math. 12, 153-193, 1993

[A.12] Raviart P.A., *An analysis of particle method* , Lect. Notes Math. 1127, 243–324, 1985.

[A.13] Scilab, `http ://scilabsoft.inria.fr`

[A.14] Sweldens W., The lifting scheme : a custom design construction of biorthogonal wavelets, Appl. Comp. Harm. Anal. 3, 186-200, 1996.

Annexe A

Définitions et rappels

A.1 Résidu au sens des éléments finis

Soit Ω un ouvert borné de $\mathbb{R}^3$. Soit V_h un espace d'éléments finis de dimension n inclus dans $L^2(\Omega)$. On note $(\varphi_i)_{1\leq i\leq n}$ les fonctions de base et $(s_i)_{1\leq i\leq n}$ les noeuds du maillage tels que $\varphi_i(sj) = \delta_{ij}$. On veut résoudre l'équation variationnelle suivante :

$$a_h(u_h, v_h) = (f, v_h)_{L^2}, \quad \forall v_h \in \boldsymbol{V}_h.$$

Soit $A_h : V_h \to V_h$ l'opérateur discret défini par

$$a_h(u_h, v_h) = (A_h u_h, v_h)_L^2(\Omega).$$

Le résidu pour les éléments finis est $res_h = A_h u_h - f$ et on aimerait estimer la norme L^2 de ce résidu. On note RES le vecteur défini par $RES(i) = res_h(s_i)$. Soit $\boldsymbol{A}$ la matrice $n \times n$ obtenue en discrétisant par éléments finis, i.e. $\boldsymbol{A}_{i,j} = a_h(\varphi_j, \varphi_i)$. De même, soit $F \in \mathbb{R}^n$ tel que pour $i = 1, \cdots, n, \quad F_i = (f, \varphi_i)$ Numériquement, on résout le système $\boldsymbol{A}U = F$ d'inconnue U et le résidu numérique naturel est $D = \boldsymbol{A}U - F$. On note $(\ ,\ )_E$ la norme euclidienne. Soit v_h une fonction-test et soit V tel que $\forall i = 1, \cdots, n, \quad V_i = v_h(s_i)$. On a

$$(D, V)_E = (\boldsymbol{A}U - F, V)_E = a(u_h, v_h) - (f, v_h)_{L^2} = (res_h, v_h)_{L^2} = (\boldsymbol{Q}RES, V)_E,$$

où $\boldsymbol{Q}$ est la matrice de masse. Donc $D = \boldsymbol{Q}RES$. Pour estimer une approximation de $\|res_h\|_2^2 = (QRES, RES) = (D, \boldsymbol{Q}^{-1}D)$.

De la même façon, on a une estimé du la norme L^2 d'une fonction de V_h :

$$\|v_h\|_2^2 = (\boldsymbol{Q}V, V)_E \sim \sum_{i=1}^{n} \boldsymbol{Q}(i,i)V_i^2.$$

Sur un maillage régulier, on a $\boldsymbol{Q}(i,i) \sim Ch^3 \sim CN^{-1}$ où N est le nombre de degrés de liberté. Alors on peut approcher la norme L^2 par

$$\|v_h\|_2 \sim N^{-1/2}\|V\|_E.$$

A.2 M-matrice

A est une M-matrice si A s'écrit sous la forme $A = sI - B$ où $B \geq 0$ et $s \geq \varrho(B)$, avec $\varrho(B)$ le rayon spectral de B.

Annexe B

Bibliothèques

B.1 Fonctions LAPACK utilisées

B.1.1 Fonction `dhseqr_`

Calcule les valeurs propres de la matrice d'Hessenberg réelle H et éventuellement les matrices T et Z de la décomposition de Schur

$$H = ZTZ^T$$

avec Z orthogonale.

Si on spécifie une matrice Q en entrée, dhseqr_ calcule la décomposition d'une matrice quelconque A pouvant être réduite à la forme d'Hessenberg par $A = QHQ^T$.

B.1.2 Fonction `dgeev_`

dgeev_ calcule les valeurs propres ainsi que les vecteurs propres à droite ou à gauche pour une matrice non symétrique réelle. Utilise la factorisation QR de dgeqrf_.

B.1.3 Fonction `dgeqrf_`

dgeqrf_ calcule la factorisation QR d'une matrice réelle quelconque. La matrice R correspond au bloc triangulaire supérieur de la matrice de sortie et la matrice Q est représentée par le produit matriciel suivant :

$$Q = H(1)H(2)...H(k), \text{ avec } k = min(m, n),$$

et où les

$$H(i) = I - \tau_i v_i v_i^T.$$

Les τ_i sont stockés dans un vecteur et les vecteurs v_i se déduisent de la partie triangulaire inférieure matrice de sortie (voir [52, pp 208-213] pour plus de détails).

B.1.4 Fonction `dorgqr_`

dorgqr_ assemble la matrice Q définie dans dgeqrf_ comme produit des $H(i)$.

B.2 Description de SparseLib++

SparseLib++ est une bibliothèque écrite en C++ qui contient les stockages creux CSR, CSC, permet les produits matrice-vecteurs creux en utilisant Sparse BLAS. Elle connaît également le format Harwell Boing et est capable de convertir tous les formats de stockage. SparseLib++ contient aussi plusieurs classes de préconditionneurs : préconditionneur diagonal, Cholesky incomplet et ILU(0). Elle est construite à partir de LAPACK, MV++ qui est une bibliothèque C++ qui définit les classes de vecteurs. IML++ contient, sous forme de template, les méthodes itératives classiques (BiCG, BiCGSTAB, CG, CGS,GMRES, QMR).

Nous allons décrire rapidement quelques classes de Sparselib++ utilisées dans le code :

Classe VECTOR_double

Cette classe contient un pointeur en double précision et la dimension du vecteur. Elle permet l'addition de vecteurs. On peut calculer directement la norme euclidienne, la norme sup d'un vecteur et le produit scalaire de deux vecteurs de même taille.

MV_Vec_Index

Contient deux entiers. Permet d'extraire un sous vecteur d'indices consécutifs compris entre ces deux entiers et de faire directement des sommes et affectations de sous-vecteurs de même taille.

Classe CompRow_Mat_double

Cette classe contient les trois tableaux en double précision nécessaires à la définition d'une matrice creuse en stockage CSR. Elle permet le produit matrice vecteur ainsi que le produit par sa transposée. Elle permet d'accéder directement aux valeurs de la matrice par leur numéro de lignes et colonnes et permet la modification des valeurs appartenant à la structure creuse

de la matrice.

Classe CompRow_LU_double

Calcule la factorisation de type ILU(0) d'une matrice creuse donnée et inverse, via blas, les blocs triangulaires.

B.3 Description des classes du code pour la décomposition de domaine

Classe projddm

Cette classe gère les correspondances entre la numérotation globale et les numérotations par sous domaines. Elle permet de transformer un vecteur u défini sur $\Omega_{h,R}$ en un pointeur de vecteurs $u[i]$ définis sur Ω_i pour $0 \leq i \leq J$ et réciproquement. Elle permet aussi les restrictions, plongement, et toutes le opérations relatives à le gestion des deux numérotations. La numérotation locale est la suivante : d'abord les noeuds intérieurs ordonnés suivant l'algorithme de Cuthill Mac Kee inverse, de manière la réduite le profil pour pouvoir utiliser des factorisations incomplètes avec seuillage, ensuite, les noeuds du bord intérieur, et enfin ceux du bord extérieur. Cette classe permet également calculer le produit d'un vecteur défini sur l'espace produit Λ par la matrice des contraintes de continuité $\boldsymbol{C}$. Elle permet aussi le produit d'un vecteur défini sur Γ par $\boldsymbol{C}^T$.

Classe ddmSolver

Cette classe permet de décomposer les matrices de rigidités $\boldsymbol{A}^{(i)}$ définies sur chaque sous domaines $\Omega_i, \quad 0 \leq i \leq J$ en 4 sous matrices : la matrice $\boldsymbol{A}_{I_i,I_i}$ définie uniquement sur les noeuds intérieurs et dont nous calculons la factorisation incomplète, puis les matrices $\boldsymbol{A}^{(i)}_{I_i,\Gamma}$, $\boldsymbol{A}^{(i)}_{\Gamma,I_i}$, et $\boldsymbol{A}^{(i)}_{\Gamma,\Gamma}$. A partir de ces données, cette classe est capable de calculer les produits matriciels suivants et de résoudre les équations suivantes :

- produit par $\boldsymbol{S}_i$ défini en 6.24
- résolution du problème à l'interface 6.25 sur un sous domaine Ω_i.
- résolution de l'équation relative aux noeuds intérieurs 6.26
- produit par $\boldsymbol{C}\widetilde{\boldsymbol{S}}\boldsymbol{C}^T$.

Classe ddm

Cette classe utilise les propriétés du maillage exponentiel en permettant de faire le produit matrice vecteur sur tout le domaine à partir d'une matrice assemblée seulement sur le premier

anneau. Valable pour $\boldsymbol{B}, \boldsymbol{E}$ et $\boldsymbol{Q}$. Permet aussi de faire le produit par la transposée.

Classe Matddm

Cette classe est utilisée pour la matrice de convection-diffusion. Cette classe contient les matrices $\boldsymbol{A}^{(i)}$ définies sur chaque sous-domaine. Elle permet de calculer directement le produit matrice vecteur $\boldsymbol{A}u$ où u est un vecteur défini sur le domaine complet.

Classe PrecFETI

Classe donnant le préconditionneur de la méthode FETI. Elle calcule par sa fonction solve le produit par $\boldsymbol{M}^{-1} = \boldsymbol{C}\boldsymbol{S}\boldsymbol{C}^T$ dans le cas du FETI-1 et par $\boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^T$ dans le cas du FETI-2. Dans ce cas du FETI-2, nous construisons la matrice $\boldsymbol{G}$ lors de la création de la classe.

Classe feti

Cette classe permet, à partir des classes précédentes de résoudre le problème global à l'interface en appliquant un GMRES sur $\boldsymbol{M}^{-1}\boldsymbol{C}\widetilde{\boldsymbol{S}}\boldsymbol{C}^T$ ou $\boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^T\boldsymbol{C}\widetilde{\boldsymbol{S}}\boldsymbol{C}^T$. Elle résout ensuite le problème complet en calculant les solutions sur les noeuds intérieurs de chaque sous-domaine et en regroupant tous les résultats en un vecteur défini sur la numérotation globale.

Classe bloc_mult

Permet de définir par la simple opération * le produit d'un tableau de vecteur par la matrice de point selle par blocs $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{E} \end{pmatrix}$. Elle permet aussi de définir le produit par sa transposée. Cette classe gère, lors du produit, les conditions de Dirichlet liées au bloc $\boldsymbol{A}$.

Classe ml_triang_preconditioner

Cette classe définit le préconditionneur triangulaire pour le problème de point-selle. Elle permet, par sa fonction solve, de calculer le produit par $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{Q} \end{pmatrix}^{-1}$ pour le préconditionneur à gauche et le produit par $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix}^{-1}$ pour le préconditionneur à droite. Plus exactement, $\boldsymbol{Q}$ est le "mass lumping" de la matrice de masse et $\boldsymbol{A}$ est inversé par la méthode choisi (FETI ou GMRES préconditionné par ILU(0)).

Annexe C

Mise En oeuvre

C.1 Raffinement du maillage

Nous raffinons le maillage de la façon suivante : Nous démarrons avec 6 hexaèdres dont les sommets appartiennent soit à la sphère de rayon 1, soit à celle de rayon 2. Les deux sphères sont donc initialement approchées par deux polyèdres dont les faces sont des carrés. Nous raffinons chacun des hexaèdres en 8 hexaèdres autant que fois que nous le souhaitons puis à la dernière étape, décomposons les hexaèdres en 24 tétraèdres (voir figure C.1). Pour avoir un maillage adapté à notre problème, nous allons modifier le raffinement. Chaque centre d'une face ou centre de l'hexaèdre introduit lors du raffinement est déplacé de la façon suivante : chaque centre est le milieu d'un segment dont les extrémités sont à égale distance de origine O ; nous déplaçons ce centre par une homothétie de centre O de manière à ce que ce point et les deux extrémités du segment appartiennent au même cercle de centre O. Cette étape peut-être améliorée dans le cadre du maillage exponentiel : dans ce cas, le mieux est de ne jamais raffiner ; il suffit en effet de construire une seule couche d'hexaèdres dont la taille des côtés est de l'ordre de la taille de la maille souhaitée, de décomposer en tétraèdres, puis de construire le reste du maillage par homothétie.

C.2 Assemblage du problème de Navier-Stokes

On génère le maillage uniquement sur le premier anneau. Les matrices $\boldsymbol{E}, \boldsymbol{B}$ et $\boldsymbol{Q}$ sont linéaires. On peut se contenter de les assembler sur le premier anneau avec des conditions aux limites de Neumann homogènes puis on en déduit leur valeur sur tout le domaine. En

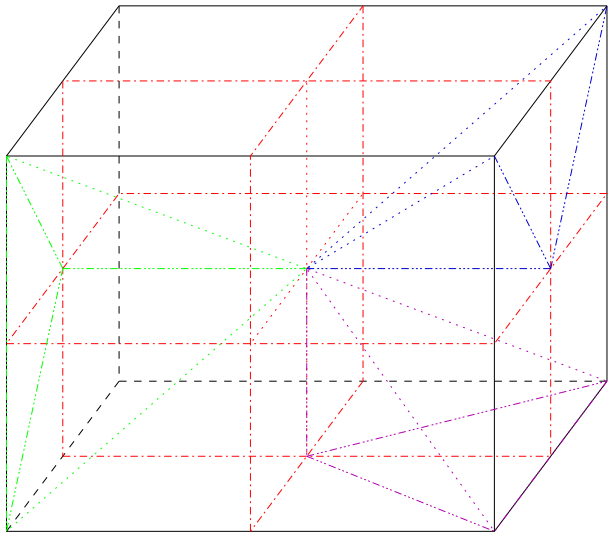


FIG. C.1 – Raffinement d'un hexaèdre en 24 tétraèdres

effet, comme notre maillage est homothétique, les fonctions de base se déduisent d'un anneau à l'autre par

$$\varphi_j^i(x) = \varphi_j^0(2^{-i}x) = \varphi_j^0(y), \quad \forall x \in \Omega_j, \forall 1 \leq j \leq m_0, \quad \text{avec} y \in \Omega_0. \tag{C.1}$$

On appelle $\boldsymbol{Q}^i, \boldsymbol{B}^i$ et $\boldsymbol{E}^i$ les matrices définies de la façon suivante :

$$\boldsymbol{Q}_{j,k}^i := \int_{\Omega_i} \varphi_k^i(x) \cdot \varphi_j^i(x)\, dx = 2^{3i} \int_{\Omega_0} \varphi_k^0(y) \cdot \varphi_j^0(y)\, dy, \tag{C.2}$$

$$\boldsymbol{B}_{j,k}^i := \int_{\Omega_i} (\nabla \cdot \varphi_k^i(x)) \cdot \varphi_j^i(x)\, dx = 2^{2i} \int_{\Omega_0} (\nabla \cdot \varphi_k^0(y)) \cdot \varphi_j^0(y)\, dy, \tag{C.3}$$

et

$$\begin{aligned} \boldsymbol{E}_{j,k}^i := \sum_{K \in \mathcal{T}_{h,R} \cap \Omega_i} \tau_{2,K} \int_{\Omega_i} \nabla \varphi_k^i(x) \cdot \nabla \varphi_j^i(x)\, dx &= 2^i \sum_{K' \in \mathcal{T}_{h,R} \cap \Omega_0} (2^{2i} \tau_{2,K'}) \int_{\Omega_0} \nabla \varphi_k^0(y) \cdot \nabla \varphi_j^O(y)\, dy \\ &= 2^{3i} \sum_{K' \in \mathcal{T}_{h,R} \cap \Omega_0} \tau_{2,K'} \int_{\Omega_0} \nabla \varphi_k^0(y) \cdot \nabla \varphi_j^O(y)\, dy \end{aligned} \tag{C.4}$$

On appelle n_i le nombre de noeuds de $\overline{\Omega}_i$ et n le nombre de noeuds de $\overline{\Omega_{h,R}}$. Pour $1 \leq i \leq J$, on définit $\boldsymbol{R}_i$ la matrice de taille $n \times n_i$ correspondant à la restriction d'un vecteur aux seuls noeuds de Ω_i et $\boldsymbol{R}_i^T$ la matrice de taille $n_i \times n$ correspondant à la prolongation par 0 d'un vecteur définis sur $\overline{\Omega}_i$.

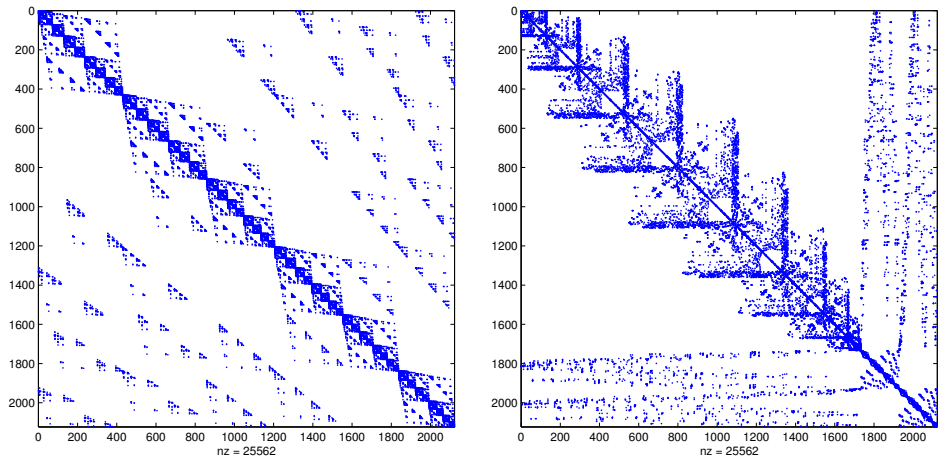
On a alors les relations suivantes :

$$\boldsymbol{Q} = \sum_{i=0}^{J} \boldsymbol{R}_i^T \boldsymbol{Q}^i \boldsymbol{R}_i, \quad \boldsymbol{E} = \sum_{i=0}^{J} \boldsymbol{R}_i^T \boldsymbol{E}^i \boldsymbol{R}_i \quad \text{et} \quad \boldsymbol{B} = \sum_{i=0}^{J} \boldsymbol{R}_i^T \boldsymbol{B}^i \boldsymbol{R}_i. \tag{C.5}$$

Cela signifie qu'on peut simplement déduire les matrice $\boldsymbol{Q}, \boldsymbol{B}$ et $\boldsymbol{E}$ à partir de l'assemblage sur Ω_0. La matrice de convection diffusion ne pourra pas être assemblée sur un seul anneau à cause de la non linéarité, mais on assemblera $J+1$ matrices $\boldsymbol{A}^i, 0 \leq i \leq J+1$ avec les conditions aux limites de Robin (6.14) et (6.15).

C.3 Réordonnancement

Nous numérotons d'abord les noeuds intérieurs, puis ensuite, les noeuds des interfaces. Sur les noeuds intérieurs, nous appliquons l'algorithme de Cuthill Mac Kee pour réduire le profil de la matrice, dans le but d'appliquer une factorisation incomplète de type ILUT.



(a) Matrice d'origine

(b) Matrice après permutation inverse de Cuthill Mac Kee sur l'ensemble des noeuds intérieurs

FIG. C.2 – Réordonnancement

La figure C.2(a) montre la structure de la matrice sur un anneau telle qu'elle a été générée par le mailleur. Nous voyons qu'une factorisation appliquée sur la sous matrice extraite correspondant aux noeuds intérieurs serait catastrophique, car cela donnerait une matrice pleine,

car une factorisation a pour effet de remplir le profil. Nous avons appliqué l'algorithme de Reverse Cuthill Mac Kee pour réduire le profil de la matrice. Nous voyons dans les figures C.2(b) la structure en flèche particulière à cette permutation. Les flèches sont toujours orientées vers la droite pour la permutation de Cuthill Mac Kee inverse.

Titre : Calculs d'écoulements extérieurs incompressibles

Résumé : Le but de cette thèse est d'approcher numériquement la solution des équations de Navier-Stokes stationnaires incompressibles dans un domaine extérieur tridimensionnel. Pour cela, nous imposons des conditions aux limites bien choisies sur le bord libre de notre domaine de calcul. Nous discrétisons ensuite par des éléments finis de même ordre avec stabilisation, ce qui implique que la linéarisation de notre problème est un problème de point selle généralisé. Nous choisissons de résoudre le système complet par une méthode de Krylov. La difficulté réside dans deux problèmes de préconditionnement : celui du complément de Schur et celui du bloc convection-diffusion.

Dans un premier temps, nous montrons que la matrice de masse est un équivalent spectral du complément de Schur, ce qui implique que le nombre d'itérations de notre méthode est indépendant de la taille de l'espace de discrétisation. Nous étudions théoriquement le comportement des valeurs propres du problème préconditionné en fonction du nombre de Reynolds dans le cas de la cavité entraînée. Nous ajoutons ensuite l'influence du rayon de troncature pour le problème extérieur. Les résultats numériques tridimensionnels viennent confirmer la théorie et montrent la robustesse de la méthode.

Ensuite, nous proposons une méthode de décomposition de domaines sans recouvrement pour le problème de convection-diffusion dans laquelle nous imposons la continuité de la solution par des multiplicateurs de Lagrange. Nous étudions les performances d'un préconditionneur pour le problème à l'interface et étendons ainsi à la dimension trois les résultats numériques bidimensionnels de la littérature.

La dernière partie du manuscrit est indépendante du sujet de thèse : elle relate un travail portant sur la physique des plasmas effectué à l'occasion du CEMRACS 2003.

Mots clés : Préconditionnement, problèmes de point selle, grand systèmes linéaires creux, méthode de décomposition de domaine, équations de Navier-Stokes, méthode d'éléments finis.

Title : Computations of exterior Navier-Stokes flows

Abstract : We aim to approach the solution of stationary incompressible Navier-Stokes systems in a three-dimensional exterior domain. So, we impose some appropriate conditions to the free boundary of the computational domain. We discretize by equal order finite elements with stabilization, thus the linearisation leads to a generalized saddle point problem. We choose to solve the complete system by a Krylov method. The remaining difficulties are the preconditioning of two matrices : the Schur complement matrix and the convection-diffusion matrix.

At first, we prove that the mass matrix is spectrally equivalent to the Schur complement, which means that our iteration count is independent of the size of the discretization space. We study theoretically the spectrum of the preconditioned problem with respect to the Reynolds number when the test-case is the driven cavity. Next, we additionally analyze the dependence on the truncation radius for the exterior problem. The numerical three-dimensional results well confirm the theory and show the method robustness.

Afterwards, we propose a non-overlapping domain decomposition method for the convection-diffusion problem where the continuity of the solution is imposed by Lagrange multipliers. We study the performance of a preconditioner for the interface problem, so that we extend to the three-dimensional case some two-dimensional numerical results of the literature.

An other part of this document, independent of the thesis subject, is devoted to a work concerning plasma physics realized during CEMRACS 2003.

Key words : Preconditioning, saddle point problem, large sparse linear systems, domain decomposition method, Navier-Stokes equations, finite element method.