



HAL
open science

Méthode de Galerkin Discontinue et intégrations explicites-implicites en temps basées sur un découplage des degrés de liberté. Applications au système des équations de Navier-Stokes.

Sophie Gérald

► **To cite this version:**

Sophie Gérald. Méthode de Galerkin Discontinue et intégrations explicites-implicites en temps basées sur un découplage des degrés de liberté. Applications au système des équations de Navier-Stokes.. Mécanique des fluides [physics.class-ph]. Université Pierre et Marie Curie - Paris VI, 2013. Français. NNT: . tel-00943621

HAL Id: tel-00943621

<https://theses.hal.science/tel-00943621v1>

Submitted on 8 Feb 2014

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.



**THESE DE DOCTORAT DE
L'UNIVERSITÉ PIERRE ET MARIE CURIE
(Université Paris VI)**

Spécialité : Mathématiques appliquées
(École Doctorale de Sciences Mathématiques de Paris-Centre)

Présentée par

Sophie GÉRALD

Pour l'obtention du grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

Méthode de Galerkin Discontinue et intégrations explicites-implicites
en temps basées sur un découplage des degrés de liberté.
Applications au système des équations de Navier-Stokes.

soutenue le 26 novembre 2013,

devant le jury composé de :

| | | |
|----------------------|--|--------------------|
| Rémi ABGRALL | Directeur de recherche, Inria Bordeaux | Président |
| Frédéric COQUEL | Directeur de recherche, CNRS | Directeur de thèse |
| Daniele A. DI PIETRO | Professeur, Université de Montpellier 2 | Examinateur |
| Edwige GODLEWSKI | Professeur, Université Pierre et Marie Curie | Examinatrice |
| Hervé GUILLARD | Directeur de recherche, Inria Sophia Antipolis | Rapporteur |
| Philippe HELLUY | Professeur, Université de Strasbourg | Rapporteur |
| Florent RENAC | Docteur-Ingénieur chercheur, Onera | Encadrant |

Université **P**ierre et **M**arie **C**urie (UPMC) - Paris VI
École **D**octorale de sciences mathématiques de Paris-Centre (ED 386)
Laboratoire Jacques-Louis Lions (UMR 7598)

Office **N**ational d'**E**tudes et de **R**echerches **A**érospatiales (Onera)
Département de **S**imulation **N**umérique des écoulements et **A**éroacoustique (DSNA)
Unité de méthodes **N**Umériques pour la **M**écanique des **F**luides (NUMF)

*à mon grand-oncle Jacques Reclus,
à mes parents Nadine et Jean,
à mon cher Maxime.*

Remerciements

Je veux ici exprimer toute ma reconnaissance à celles et ceux qui se sont intéressés à mon travail et grâce auxquels j'ai pu mener à bien cette thèse.

* * *

Je tiens tout d'abord à remercier les membres du jury pour le temps qu'ils ont consacré à mon travail, leurs questions et leurs remarques.

J'exprime toute ma gratitude à mon directeur de thèse Frédéric COQUEL. Il a dirigé cette thèse avec écoute, soutien, expertise et optimisme. J'espère que ma passion pour la simulation numérique fera de nouveau se croiser nos chemins.

J'assure de toute ma considération mon encadrant Florent RENAC. Sa disponibilité, ses connaissances, sa pédagogie et sa patience ont été un très beau cocktail pour faire vivre mon âme de chercheuse. J'espère que ce premier encadrement de thèse lui laissera également que de bons souvenirs. Je lui sais gré d'avoir donné la chance d'ajouter à ma formation de doctorante une expérience d'enseignement dans le supérieur.

Je suis particulièrement reconnaissante à Rémi ABGRALL de m'avoir fait l'honneur d'accepter de participer au jury et de le présider. Je n'oublie pas que c'est lui qui m'a donné la possibilité de regrouper les résultats du cas test du canal avec une bosse exposés lors du premier atelier sur les méthodes d'ordre élevé.

Je remercie vivement les rapporteurs Hervé GUILLARD et Philippe HELLUY. Leurs observations fort pertinentes m'ont permis d'approfondir certains points délicats de mon travail.

Je remercie sincèrement les examinateurs Edwige GODLEWSKI et Daniele A. DI PIETRO. Leurs questions m'ont permis d'étendre la vision de mon travail.

Même s'ils ont finalement dû décliner et n'ont pas pu être présents pour des raisons professionnelles le jour J, merci à Daniel BOUCHE et à Patrice HAURET d'avoir accepté le principe de mon invitation de participer au jury.

Enfin, je ne peux achever ces premiers remerciements sans évoquer la mémoire de Huy Duong BUI qui devait faire partie du jury mais dont la maladie en a décidé autrement. J'aurais tant aimé connaître la question qu'il m'aurait posée, je resterai donc dans le souvenir qu'il m'en aurait posé une.

* * *

Je n'oublie ni mes collègues de travail, ni les spécialistes en mécanique des fluides, en analyse numérique et en informatique grâce auxquels ont pris corps un certain nombre des idées qui m'ont amenée à mieux appréhender les objectifs de mon travail.

Je remercie les responsables des organismes qui m'ont accueillie de 2009 jusqu'aux derniers jours, en prenant soin que mes travaux se déroulent dans les meilleures conditions. Tout d'abord, je pense à Jean-Marie LE GOUEZ, directeur du DSNA de l'Onera et Vincent COUAILLIER, chef de l'unité NUMF de l'Onera, puis à Yvon MADAY et Benoît PERTHAME les directeurs successifs du laboratoire Jacques-Louis Lions de Paris VI, sans oublier Gilles GODEFROY et Didier SMETS les directeurs successifs de l'école doctorale ED386.

Je remercie chaleureusement les chercheurs qui m'ont tout naturellement fait profiter de leurs travaux pour une meilleure valorisation des miens. Merci à Francesco BASSI et Andrea CRIVELLINI pour l'échange des résultats sur le temps CPU de calcul de la matrice implicite. Merci à Herman J.H. CLERCX et Geert H. KEETELS pour la transmission de la solution spectrale de référence sur le cas test de l'interaction tourbillons-paroi. Merci à tous les groupes du premier atelier sur les méthodes d'ordre élevé pour la description de la méthode qu'ils ont utilisée. Merci à Juliette RYAN et Michel BORREL pour le partage de leur code Navier-Stokes. Merci à Marie-Claire LE PAPE pour la communication de son savoir-faire GMSH.

Un grand merci à Claude MARMIGNON pour avoir co-encadré la thèse aux côtés de Florent. Ses conseils sur les schémas de discrétisation temporelle ont été un point clé de ma recherche.

Je pense également à Didier BLAISE et Philippe CERONI pour leur aide sur les aspects informatiques, Laurent CAMBIER pour notre discussion sur la « quête du graal » en simulation aéronautique, Jean-Baptiste CHAPELIER pour l'aide à la mise en place du cas test de l'interaction tourbillons-paroi, Marta DE LA LLAVE PLATA, Philippe DELORME et Christophe PEYRET pour nos échanges sur la méthode de Galerkin discontinue, Thiên-Hiêp LÊ pour l'organisation de la journée des doctorants, Emeric MARTIN pour la lutte contre les fuites mémoires, et tous les autres collègues chercheurs, enseignants, docteurs, post-doctorants, doctorants et stagiaires pour les discussions et échanges qui ont fait avancer ma recherche.

De manière générale, je remercie tout le personnel qui a facilité mon travail, tant à l'Onera qu'à Paris VI. Un mot en particulier pour Ghislaine DENIS, secrétaire du DSNA, pour son sourire quotidien et sa joie de vivre. Une pensée également pour Annie GARCIA, autre secrétaire du DSNA, pour son aide à l'organisation des congrès et à la publication du mémoire final. Une dernière intention pour Jean-François VENUTI et Corentin LACOMBE, gestionnaires à l'école doctorale, le premier pour le lien qu'il m'a permis d'entretenir avec les activités de l'école et le second pour l'accompagnement dans les étapes du rendu du mémoire aux rapporteurs.

Je termine ici par tous mes professeurs qui ont su, par leur passion et leur pédagogie, orienter ma carrière dans ce monde formidable où se côtoient les mathématiques, la physique et la simulation numérique. M'exposer à en faire la liste serait trop risqué, je pourrais en oublier. Donc tout simplement : merci à tous.

* * *

Je me tourne maintenant vers mes amis, mes proches et ma famille auxquels je dois un soutien sans faille tant dans les moments de doute que dans ceux de certitude. Merci de m'avoir épaulée jusqu'à ce jour important dont j'ai longtemps pensé qu'il achèverait mes études, mais dont je réalise aujourd'hui qu'il est celui où tout commence.

Merci à vous amis, voisins, camarades du CMD, de l'UVSQ, de l'Ensta Paris-Tech, de l'Onera, de la LiOnera et d'EChO de votre présence à ma soutenance. Je pense plus particulièrement à Alexandra, Célia, Chloé, Françoise et Ky, Marie-Alix, Martine et Didier, Jean-Baptiste, Claude, Florian, Philippe-Henri et Raphaël.

Merci aux expéditeurs des textos Flavinois d'encouragement dès l'aube du jour J. J'ai été fière de compter parmi l'assistance, ma grand-mère et je suis heureuse qu'elle ait ainsi pu mettre un peu de concret sur mon travail de thèse. Je remercie tendrement les deux personnes courageuses qui ont osé relire mon mémoire, à la fois pour avoir corrigé quelques erreurs de plume mais aussi quelques coquilles dans les équations. Merci Maman et Papa. J'associe à cette pensée mes sœurs Catherine et Elisabeth et mon frère Armand, ainsi que leur famille. Je remercie Michèle, Daniel et Marie-Laure d'avoir assisté à ma soutenance. Cela m'a beaucoup touché.

Un dernier remerciement qui ne sera jamais à la mesure de ce que m'a donné celui qui a vécu la thèse au jour le jour à mes côtés, en acceptant bien des sacrifices. Ah ! Maxime, je crois qu'elle va être belle l'année 2014 !

* * *

Enfin merci à toi, lecteur, de t'intéresser à ce travail de thèse...

Sophie GÉRALD,
à l'Onera, le 26 novembre 2013.

Table des matières

| | |
|--|----------|
| Table des matières | ix |
| Nomenclature | xiii |
| Introduction générale | 1 |
| 0.1 Ecoulements instationnaires à convection dominante | 2 |
| 0.1.1 Phénoménologie des écoulements étudiés | 2 |
| 0.1.2 Modélisation et simulation des écoulements étudiés | 3 |
| 0.2 Etat de l'art en simulation numérique | 4 |
| 0.2.1 Importance de la simulation | 4 |
| 0.2.2 Codes industriels actuels | 5 |
| 0.2.3 Codes industriels de demain | 6 |
| 0.3 Méthode de Galerkin Discontinue | 7 |
| 0.3.1 Principe | 8 |
| 0.3.2 Avantages et inconvénients | 9 |
| 0.3.3 Historique | 10 |
| 0.4 Travaux de thèse | 14 |
| 0.4.1 Motivations et contributions | 15 |
| 0.4.2 Organisation du mémoire | 19 |
| 0.4.3 Communications | 21 |

I Ecoulements bidimensionnels et instationnaires de fluides visqueux et compressibles : de la modélisation au schéma numérique en espace et en temps **23**

| | |
|---|-----------|
| 1 Modélisation du problème physique considéré | 25 |
| 1.1 Avant-propos | 25 |
| 1.2 Système d'équations pour un fluide visqueux et compressible | 26 |
| 1.2.1 Système des équations de Navier-Stokes | 26 |
| 1.2.2 Nombres sans dimension | 28 |
| 1.2.3 Cas particulier du système des équations d'Euler | 29 |
| 1.3 Conditions aux limites | 30 |
| 1.3.1 Types de frontières | 30 |
| 1.3.2 Conditions aux limites considérées | 30 |
| 1.4 Conclusion partielle | 33 |

| | | |
|---|--|---------------|
| 2 | Discrétisation spatiale Galerkin Discontinue (GD) | 35 |
| 2.1 | Avant-propos | 35 |
| 2.2 | Formulation variationnelle | 36 |
| 2.3 | Méthode de Galerkin Discontinue | 36 |
| 2.3.1 | Discrétisation spatiale du domaine | 37 |
| 2.3.2 | Espaces fonctionnels discrets | 39 |
| 2.3.3 | Approximation interne | 41 |
| 2.3.4 | Choix des flux numériques | 43 |
| 2.4 | Fonctions de base et intégration numérique : éléments droits et courbes | 51 |
| 2.4.1 | Transformation de l'élément de référence vers l'élément physique présentant une géométrie courbe | 52 |
| 2.4.2 | Expression des fonctions de base | 55 |
| 2.4.3 | Intégration numérique | 58 |
| 2.5 | Conclusion partielle | 59 |
| 3 | Discrétisations temporelles explicites-implicites à coût réduit | 61 |
| 3.1 | Avant-propos | 61 |
| 3.2 | Prérequis sur les conditions de stabilité | 62 |
| 3.3 | Discrétisation explicite avec un schéma Runge-Kutta | 64 |
| 3.4 | Discrétisation explicite-implicite avec une décomposition d'opérateurs | 65 |
| 3.4.1 | Décomposition de Strang | 65 |
| 3.4.2 | Choix du schéma explicite et du schéma implicite | 67 |
| 3.4.3 | Coûts en temps de calcul de la résolution complète | 73 |
| 3.5 | Simplifications du schéma implicite en temps | 74 |
| 3.5.1 | Structure de la matrice implicite | 74 |
| 3.5.2 | Simplification existante : méthode sans jacobienne | 78 |
| 3.5.3 | Simplification innovante : découplage des degrés de liberté | 79 |
| 3.5.4 | Double simplification : méthode sans jacobienne et découplage | 85 |
| 3.6 | Comparaison des simplifications | 86 |
| 3.6.1 | Encombrement mémoire | 86 |
| 3.6.2 | Nombre d'opérations à virgule flottante | 87 |
| 3.7 | Conclusion partielle | 90 |
| II Schéma numérique : de la mise en œuvre à la validation et à l'évaluation des performances | | 91 |
| 4 | Mise en œuvre du schéma discret | 93 |
| 4.1 | Avant-Propos | 93 |
| 4.2 | Code de calcul dédié | 95 |
| 4.2.1 | Historique | 95 |
| 4.2.2 | Synthèse des principales étapes de programmation | 95 |
| 4.3 | Cas tests exploités | 96 |
| 4.3.1 | Synthèse des écoulements associés aux cas tests | 96 |
| 4.3.2 | Démarche logique | 99 |

| | | |
|----------|---|------------|
| 5 | Validation de la discrétisation spatiale Galerkin Discontinue | 101 |
| 5.1 | Écoulement dans un canal convergent-divergent | 101 |
| 5.1.1 | Avant-propos | 103 |
| 5.1.2 | Description du cas test | 103 |
| 5.1.3 | Description des résultats numériques | 107 |
| 5.2 | Interaction entre une paire de tourbillons et une paroi plane | 115 |
| 5.2.1 | Avant-propos | 116 |
| 5.2.2 | Description du cas test | 117 |
| 5.2.3 | Description des résultats numériques | 121 |
| 5.3 | Conclusion partielle | 125 |
| 6 | Validation des discrétisations temporelles explicites-implicites | 133 |
| 6.1 | Avant-propos | 133 |
| 6.2 | Prérequis pour la validation | 134 |
| 6.2.1 | Types de comparaisons effectuées pour la validation | 134 |
| 6.2.2 | Grandeurs caractéristiques des écoulements externes | 135 |
| 6.3 | Validation sur des cas académiques | 136 |
| 6.3.1 | Interaction entre une paire de tourbillons et une paroi plane | 137 |
| 6.3.2 | Écoulement autour du profil d'aile NACA0012 | 143 |
| 6.4 | Validation sur des cas d'applications | 150 |
| 6.4.1 | Écoulement autour du profil d'aile SD7003 | 150 |
| 6.4.2 | Écoulement autour d'un cylindre | 155 |
| 6.5 | Conclusion partielle | 161 |
| 7 | Évaluation des performances des discrétisations temporelles | 163 |
| 7.1 | Prérequis pour l'évaluation des performances | 163 |
| 7.1.1 | Temps Elapsed et temps CPU | 164 |
| 7.1.2 | Environnement informatique | 164 |
| 7.1.3 | Définition du <i>Gain</i> en temps CPU | 165 |
| 7.2 | Estimation des temps CPU intermédiaires | 165 |
| 7.3 | Estimation du temps CPU total | 167 |
| 7.4 | Estimation des gains intermédiaires | 171 |
| 7.4.1 | Gain sur la matrice implicite | 171 |
| 7.4.2 | Gain sur la matrice de préconditionnement | 173 |
| 7.4.3 | Gain sur l'inversion GMRes | 173 |
| 7.4.4 | Gain sur la reconstruction des modes d'ordre élevé | 175 |
| 7.5 | Estimation des gains sur la résolution totale | 177 |
| 7.5.1 | Analyse globale de rapidité | 177 |
| 7.5.2 | Analyse en ordre et en maillage | 179 |
| 7.5.3 | Aparté sur les simplifications utilisant le découplage | 181 |
| 7.6 | Conclusion partielle | 183 |
| | Perspectives | 185 |

| | |
|---|------------|
| III Annexes | 187 |
| A Adimensionnement | 189 |
| A.1 Principe pour le système des équations de Navier-Stokes | 189 |
| A.2 Mise en œuvre | 190 |
| B Consistance de la méthode Galerkin Discontinue pour l'équation de la chaleur monodimensionnelle scalaire | 193 |
| B.1 Méthode de Galerkin Discontinue | 193 |
| B.2 Problème de consistance lié à l'approche naïve | 195 |
| C Résultats numériques de précision pour la validation spatiale | 197 |
| D Résultats numériques de précision pour la validation temporelle | 199 |
| E Résultats expérimentaux de référence pour la validation temporelle | 201 |
| F Analyse des gains sur la résolution totale | 203 |
| F.1 Prérequis de cette analyse | 203 |
| F.2 Analyse pour la méthode sans jacobienne | 205 |
| F.3 Analyse pour la méthode avec découplage des degrés de liberté | 207 |
| F.4 Analyse pour la méthode sans jacobienne et avec découplage | 209 |
| G Papier & Article | 211 |
| G.1 Janvier 2012 - Papier AIAA | 211 |
| G.2 Juin 2013 - Article du Journal of Computational Physics | 222 |
| Bibliographie | 245 |
| Résumé/Abstract | 258 |

Nomenclature

Caractères latins

| | |
|--|--|
| \mathbf{A} | Matrice implicite |
| $\mathbf{A}_{\kappa+\kappa^+}$ | Bloc diagonal de la matrice \mathbf{A} |
| $\mathbf{A}_{\kappa+\kappa_i^-}, \mathbf{A}_{\kappa+\kappa^-}$ | Bloc extra-diagonal de la matrice \mathbf{A} avec $1 \leq i \leq n_f$ |
| \mathcal{B}_h^p | Base qui engendre l'espace \mathcal{V}_h^p |
| \mathcal{B}_κ^p | Base qui engendre l'espace \mathcal{P}^p |
| c | Vitesse du son dans le fluide |
| C_D | Coefficient aérodynamique global de traînée |
| $C_{D,p}, C_{D,v}$ | Composantes de C_D en pression et en volume |
| C_L | Coefficient aérodynamique global de portance |
| $C_{L,p}, C_{L,v}$ | Composantes de C_L en pression et en volume |
| C_p, C_v | Chaleurs spécifiques à pression constante et à volume constant |
| C_s | Constante de Sutherland |
| CFL | Valeur de la condition de type CFL |
| d | Dimension d'espace |
| dS, dV | Elements de surface et de volume |
| \mathcal{D} | Partie diagonale d'une matrice |
| e, e_c, E | Energies interne, cinétique et totale par unité de masse |
| e | Face du maillage |
| E_c | Energie cinétique globale par unité de masse |
| $\mathcal{E}_h, \mathcal{E}_i, \mathcal{E}_b$ | Ensemble des faces, des faces intérieures et des faces frontières |
| \mathbf{F} | Tenseur des flux physiques |
| $\hat{\mathbf{F}}$ | Vecteur des flux numériques |
| $\mathbf{F}_c, \mathbf{f}_{c,i}$ | Tenseur des flux physiques convectifs, composante suivant la direction x_i |
| $\hat{\mathbf{F}}_c$ | Vecteur des flux numériques convectifs |
| $\mathbf{F}_v, \mathbf{f}_{v,i}$ | Tenseur des flux physiques visqueux, composante suivant la direction x_i |
| $\hat{\mathbf{F}}_v$ | Vecteur des flux numériques visqueux |
| \mathbf{F}_κ | Transformation de I^d vers κ |
| $\mathbf{F}_{\kappa,1}, \mathbf{F}_{\kappa,2}$ | Transformations de I^2 vers T^2 et de T^2 vers κ |
| \mathbf{G} | Tenseur pour la réécriture de \mathbf{F}_v |
| $Gain$ | Gain en temps de calcul (ou CPU) avec la méthode COMP pour référence |
| h, H | Tailles caractéristiques minimum et maximum du maillage |
| h_κ | Diamètre de la plus grande boule incluse dans κ |

| | |
|----------------------------------|--|
| I | Matrice identité |
| I^d | Élément de référence du maillage (segment, carré) |
| $J_{\mathbf{F}_\kappa}$ | Jacobien de la transformation \mathbf{F}_κ |
| $\mathbb{J}_{\alpha,\beta}^p$ | Polynôme de Jacobi de degré p et d'ordres α, β |
| k | Conductivité thermique |
| k_2 | Paramètre de stabilisation du schéma LLF |
| \mathcal{L} | Partie triangulaire inférieure d'une matrice |
| L^1 | Espace des fonctions intégrables |
| L^2 | Espace des fonctions de carré intégrable |
| L^∞ | Espace des fonctions bornées |
| L_c | Longueur caractéristique d'un objet |
| \mathbf{L}_c | Vecteur de résidu discret des termes de convection |
| \mathcal{L}_c | Opérateur discret des termes de convection |
| \mathbf{L}_v | Vecteur de résidu discret des termes de diffusion |
| \mathcal{L}_v | Opérateur discret des termes de diffusion |
| \mathbb{L}^p | Polynôme de Legendre de degré p |
| m | Degré du polynôme d'approximation de la géométrie du maillage |
| M | Matrice de masse |
| M | Nombre de Mach local |
| M_∞ | Nombre de Mach de référence |
| n | Nombre d'itérations en temps |
| n_f | Nombre de faces d'un élément du maillage |
| n | Vecteur normal extérieur à une surface et unitaire |
| N | Nombre total d'éléments du maillage |
| N_{ddl} | Dimension de l'espace \mathcal{V}_h^p (ou nombre de degrés de liberté total) |
| N_{eq} | Nombre d'équations |
| N_g | Nombre de points de Gauss |
| N_p | Dimension de l'espace \mathcal{P}^p (ou nombre de degrés de liberté par élément) |
| N_{p_s} | Nombre de degrés de liberté de bas ordre par élément |
| \mathcal{N} | Nombre total de degrés de liberté par équation |
| p | Pression statique |
| p | Degré du polynôme d'approximation de la solution numérique |
| p_s | Degré de séparation entre les modes de bas ordre et ceux d'ordre élevé |
| P | Matrice de préconditionnement |
| Pr | Nombre de Prandtl |
| \mathcal{P}^p | Espace des polynômes de degré total au plus égal à p |
| q | Vecteur du flux de chaleur |
| r | Constante du gaz |
| \mathbf{r}_h^e | Opérateur de relèvement local associé à une face e |
| RA | Rapport d'aspect du maillage |
| Re_h | Nombre de Reynolds de maille |
| Re_∞ | Nombre de Reynolds de référence |
| \mathbf{R}_h | Opérateur de relèvement global |
| s | Entropie |
| St | Nombre de Strouhal |
| t | Temps physique |
| t | Vecteur tangent à une surface et unitaire |
| \mathcal{T} | Domaine d'étude temporel |
| T^2 | Triangle de référence du maillage |

| | |
|-----------------------|--|
| \mathbf{U} | Partie triangulaire supérieure d'une matrice |
| \mathbf{U}, u_i | Vecteur vitesse, composante suivant la direction x_i |
| v | Fonction test |
| v_h | Discrétisation de v |
| \mathcal{V}_h^p | Espace Galerkin discontinu |
| \mathbf{w} | Vecteur des variables conservatives |
| \mathbf{w}_h, w_h^i | Discrétisation de \mathbf{w}_h , composante telle que $1 \leq i \leq N_{eq}$ |
| \mathbf{W} | Vecteur total des degrés de liberté sur Ω_h |
| \mathbf{W}_c | Vecteur total des degrés de liberté des termes de convection |
| \mathbf{W}_v | Vecteur total des degrés de liberté des termes de diffusion |
| $W_\kappa^{i,l}$ | Degré de liberté $1 \leq l \leq N_p$ de l'équation $1 \leq i \leq N_{eq}$ sur κ |
| \mathbf{W}_κ | Vecteur des degrés de liberté sur κ |
| \mathbf{x}, x_i | Vecteur position, i -ème coordonnée dans le repère cartésien |
| Z | Enstrophie globale par unité de masse |

Caractères grecs

| | |
|------------------------------|--|
| α | Angle d'incidence d'un objet |
| γ | Indice polytropique |
| Γ_p | Paroi du domaine spatial |
| $\Gamma_{in}, \Gamma_{out}$ | Frontières d'entrée et de sortie du domaine spatial |
| δ | Epaisseur de couche limite |
| δ_{ij} | Symbole de Kronecker |
| δ_{SJ} | Paramètre de stabilisation de la méthode SJ |
| Δt | Pas de temps |
| ϵ_{SJ} | Constante permettant de déterminer δ_{SJ} |
| $\boldsymbol{\eta}, \eta_i$ | Vecteur position dans T^2 et composante suivant la direction x_i |
| η_r | Paramètre de stabilisation du schéma BR2 |
| θ | Température |
| θ_s | Température de référence pour la loi de Sutherland |
| κ | Élément physique du maillage (segment, triangle) |
| μ | Viscosité dynamique de surface |
| μ_s | Viscosité de référence à la température θ_s |
| ν | Viscosité cinématique |
| $\boldsymbol{\xi}, \xi_i$ | Vecteur position dans I^2 , composante suivant la direction x_i |
| ρ | Masse volumique |
| $\boldsymbol{\tau}$ | Tenseur des contraintes visqueuses |
| Υ | Variable auxiliaire du schéma BR2 |
| Υ_h | Discrétisation de Υ |
| ϕ^k | Fonction de la base \mathcal{B}_κ^p |
| $\boldsymbol{\varphi}^k$ | Fonction vectorielle de la base \mathcal{B}_κ^p |
| ω | Vorticité |
| $\Omega, \partial\Omega$ | Domaine d'étude physique et sa frontière |
| $\Omega_h, \partial\Omega_h$ | Discrétisation de Ω et de $\partial\Omega$ |
| Ω_{adm} | Espace des états admissibles |

Symboles en exposant

| | |
|-----------|--|
| (n) | Grandeur au temps $n\Delta t$ |
| (\star) | Prédiction d'une grandeur |
| $+, -$ | Traces d'une quantité sur une face du maillage |

Symboles en indice

| | |
|----------|--|
| $1, 2$ | Composantes horizontale et verticale |
| b | Grandeur prise sur la frontière $\partial\Omega$ |
| B | Grandeur prise au barycentre d'un élément |
| c | Grandeur liée aux phénomènes de convection |
| e | Grandeur prise sur une face du maillage |
| G | Grandeur prise en un point de Gauss |
| h | Grandeur liée à la discrétisation du domaine de calcul |
| v | Grandeur liée aux phénomènes de diffusion |
| ref | Grandeur caractéristique pour l'adimensionnement |
| ∞ | Grandeur prise à l'infini amont ou en champ lointain |

Symboles au-dessus

| | |
|--------------|---|
| $-$ | Indique un adimensionnement |
| \wedge | Indique un flux numérique |
| \sim | Grandeur associée aux degrés de liberté de bas ordre |
| $\hat{\sim}$ | Grandeur associée aux degrés de liberté d'ordre élevé |

Sigles et abréviations

| | |
|---------------|--|
| AIAA | American Institute of Aeronautics and Astronautics |
| AFOSR | Air Force Office of Scientific Research |
| Aghora | Algorithm generation for high order resolution in aerodynamics |
| BR2 | Second schéma (ou flux) de Bassi et Rebay |
| CFD | Computational Fluid Dynamics |
| CFL | Courant-Friedrichs-Lewy |
| COMP | Méthode de discrétisation en temps explicite-implicite sans simplification, associée à une méthode GD en espace |
| CPU | Central Processing Unit |
| DLR | Deutsches Zentrum für Luft- und Raumfahrt |
| DNS | Direct Numerical Simulation |
| Dune | Distributed and unified numerics environment |
| <i>elsA</i> | ensemble logiciel de simulation en Aérodynamique |
| EXPL | Méthode de discrétisation en temps explicite, associée à une méthode GD en espace |
| Fig. | Figure |
| GD | Galerkin Discontinu |
| GD(p) | Discrétisation Galerkin Discontinue de degré p |
| GMRes | Generalized Minimum Residual |
| Inria | Institut national de recherche en informatique et automatique |
| LLF | Local Lax-Friedrichs |
| MIT | Massachusetts Institute of Technology |
| NACA | National Advisory Committee for Aeronautics |
| Onera | Office national d'études et de recherches aérospatiales |
| RANS | Reynolds Average Navier-Stokes |
| RK | Runge-Kutta |
| SD | Selig-Donovan |
| SIMP p_s | Méthode de discrétisation en temps explicite-implicite avec simplification découplant les degrés de liberté, associée à une méthode GD en espace |
| SJ | Méthode de discrétisation en temps explicite-implicite avec simplification sans jacobienne, associée à une méthode GD en espace |
| SJ+SIMP p_s | Méthode en temps explicite-implicite avec double simplification sans jacobienne et découplant les degrés de liberté, associée à une méthode GD en espace |
| SRVT | Static-Rho-Velocity-Temperature |
| Tab. | Tableau |
| URANS | Unsteady Reynolds Average Navier-Stokes |
| VKI | von Karman Institute |

Introduction générale

Ce travail de thèse s'inscrit dans le domaine de la mécanique des fluides numérique (ou « CFD »¹) pour des applications au secteur de l'aéronautique. Il concerne principalement les écoulements bidimensionnels, instationnaires et à convection dominante de fluides visqueux compressibles. Ces écoulements sont modélisés par le système des équations de Navier-Stokes, lequel est d'abord discrétisé en espace à l'aide d'une méthode de Galerkin Discontinue (ou GD), puis discrétisé en temps au moyen d'une méthode explicite-implicite. Comme on le re-précisera, l'aspect novateur de la thèse se situe dans la mise en place du schéma temporel.

L'ensemble de ces travaux a été réalisé à, et a été financé par, l'Onera², centre français de la recherche aéronautique, spatiale et de défense. Au cœur des missions de l'Onera se trouvent « la réalisation et la mise en œuvre des moyens d'expérimentation et de simulation » [80]. Plus particulièrement, ce travail a été porté par l'unité NUMF³ rattachée à la branche « mécanique des fluides et énergétique » de l'Onera. Une des activités majeures de cette unité est l'étude à caractère amont ou finalisé pour le développement d'outils innovants de simulation numérique, voire de nouvelle génération.

Pour bien cerner ce travail, les quatre points suivants sont abordés au cours de cette introduction générale :

- le premier, paragraphe 0.1, explique la phénoménologie des écoulements instationnaires à convection dominante et les besoins de simulation numérique qu'ils impliquent ;
- le second point, paragraphe 0.2, fait un état de l'art des méthodes numériques actuelles en insistant sur les limitations des méthodes utilisées dans les codes à caractère industriel au vu des besoins futurs en CFD ;
- le troisième point, paragraphe 0.3, s'arrête sur la méthode d'ordre élevé GD pour en présenter, à la fois le principe en comparaison avec ses proches parentes que sont les deux méthodes des éléments finis et des volumes finis, et l'historique en soulignant les points en rapport avec les contributions de ce mémoire ;
- le dernier point, paragraphe 0.4, est réservé à l'ensemble des motivations et contributions majeures de ce travail, ainsi qu'à l'organisation du mémoire.

1. Acronyme anglais : « **C**omputational **F**luid **D**ynamics ».

2. Acronyme : « **O**ffice **n**ational d'**e**tudes et de **r**echerches **a**érospatiales ».

3. Acronyme : « méthodes **N**umériques pour la **M**écanique des **F**luides ».

0.1 Écoulements instationnaires à convection dominante

Le point de départ de ce travail est la compréhension des phénomènes physiques des écoulements étudiés (§ 0.1.1) et des besoins requis pour leur étude numérique (§ 0.1.2).

0.1.1 Phénoménologie des écoulements étudiés

Parmi les écoulements réels complexes, c'est-à-dire principalement instationnaires et turbulents, on s'intéresse aux écoulements à convection dominante caractérisés par un nombre de Reynolds Re_∞ grand ($Re_\infty \gg 1$) mais modéré (par exemple $100 \leq Re_\infty \leq 10^4$ dans ce travail). La définition de ce nombre est basée sur les échelles globales caractéristiques de l'écoulement :

$$Re_\infty = \frac{L_c U_\infty}{\nu},$$

où L_c est l'échelle spatiale du problème physique considéré (par exemple la taille d'un obstacle soumis à l'écoulement), U_∞ est la norme de la vitesse de l'écoulement en champ lointain et $\nu \geq 0$ est le coefficient de viscosité cinématique du fluide. Ce nombre quantifie globalement l'amplitude des phénomènes de convection (dus aux forces d'inertie) par rapport à celle des phénomènes de diffusion (dus aux forces visqueuses).

Dans les écoulements étudiés, on recense deux phénomènes physiques intervenant plus ou moins majoritairement en fonction des régions du domaine d'étude. Avant de décrire ces phénomènes, on définit le nombre de Reynolds de maille Re_h pour quantifier localement l'amplitude des phénomènes de convection par rapport à celle des phénomènes de diffusion. La définition de ce nombre est basée sur les échelles locales caractéristiques d'un élément κ du maillage du domaine d'étude :

$$Re_h = \frac{h_\kappa U_\kappa(h_\kappa, t_\kappa)}{\nu},$$

où h_κ est l'échelle spatiale locale (par exemple la taille de l'élément du maillage) et U_κ est l'échelle de vitesse locale de l'écoulement définie au moyen de h_κ et du temps *via* le temps t_κ d'évolution du phénomène considéré.

- Le premier phénomène est celui de couche limite et de son éventuel décollement. Il est localisé dans les régions proches des parois qui sont des zones fortement cisailées (il y a présence de forts gradients de la vitesse) où ont lieu des échanges importants de chaleur en raison du frottement du fluide. Il s'agit principalement des phénomènes de diffusion visqueuse et de diffusion thermique, évoluant à de petites échelles spatiales et à des vitesses plutôt lentes. Localement il vient $Re_h \ll 1$.
- Le second phénomène est celui de transport de matière. Il est localisé dans les régions éloignées des parois. Il est majoritairement question d'un phénomène de convection, évoluant à de grandes échelles spatiales et à des vitesses plutôt rapides. Localement il vient $Re_h \gg 1$.

Outre ces deux phénomènes, il apparaît parfois des instationnarités dues à la dynamique de l'écoulement, tout particulièrement pour les écoulements externes autour d'obstacles. Ces instationnarités sont générées à la paroi, entre autres par les contraintes de cisaillement, et se détachent dans le sillage sous l'apparence de structures tourbillonnaires se propageant sous l'influence des forces de convection. L'allée de Bénard-von Kàrmàn est un exemple connu de telles instationnarités. Une bonne capture de ces structures est essentielle pour le dimensionnement des aéronefs par exemple, puisqu'elles contribuent de façon importante aux coefficients aérodynamiques globaux (de portance, de traînée, *etc.*) et aux sources de bruits.

0.1.2 Modélisation et simulation des écoulements étudiés

Modèles d'équations. La modélisation des écoulements instationnaires et turbulents repose sur le système mixte hyperbolique-parabolique des équations de Navier-Stokes. C'est un système d'équations aux dérivées partielles non linéaires en espace, établi en considérant les principes fondamentaux de conservation de la masse, de la quantité de mouvement et de l'énergie. En présence d'un écoulement turbulent, plusieurs approches de modélisation existent en fonction de la nature et de la taille des échelles résolues et modélisées. Les trois principales sont les approches URANS⁴, LES⁵ et DNS⁶. Pour la lecture de la suite de ce mémoire, ces approches ne sont que mentionnées, par ordre croissant à la fois de leurs capacités à représenter la turbulence et de leurs exigences en ressources machine. Par exemple, l'approche URANS utilise le modèle le plus grossier pour représenter les écoulements turbulents, et également la moins coûteuse en ressources machine.

Simulation. Une difficulté majeure pour la simulation des écoulements étudiés est de capturer à la fois l'origine (en zone de proche paroi) et le transport (en zone de sillage) des structures tourbillonnaires sur de longues distances. En effet, dans ces deux zones, les échelles spatiales et temporelles des phénomènes présents sont très différentes. Deux points numériques importants sont alors à prendre en compte.

- Premièrement, la méthode numérique retenue ne doit induire ni trop de dissipation ni trop de dispersion. Il faut donc qu'elle soit suffisamment précise, sachant qu'elle est d'autant plus précise que son ordre est grand. On rappelle qu'une méthode numérique est d'ordre s lorsque l'erreur de troncature sur la solution est proportionnelle à h^s avec h une taille caractéristique du maillage. Autrement dit, s est la valeur du premier terme de la différence entre une solution analytique d'un problème et sa solution discrète approchée. On parle alors de méthode « de bas ordre » lorsque $s \leq 2$ et « d'ordre élevé » lorsque $s \geq 3$ [176].
- Deuxièmement, les pas d'espace et de temps doivent être adaptés localement en fonction des zones de l'écoulement, aux phénomènes que l'on souhaite capturer. D'un côté, il faut que les échelles d'espace et de temps soient comparables à

4. Acronyme anglais : « **U**nsteady **R**eynolds **A**verage **N**avier-**S**tokes » ou « **U**nsteady **R**ANS ».

5. Acronyme anglais : « **L**arge **E**ddy **S**imulation ».

6. Acronyme anglais : « **D**irect **N**umerical **S**imulation ».

celles des phénomènes de convection dans les zones de sillage. D'un autre côté, il faut que ces échelles soient comparables à celles des phénomènes de diffusion dans les zones de proche paroi.

C'est dans ce contexte numérique que s'inscrit ce travail de thèse. En particulier il met en œuvre une stratégie d'intégration temporelle efficace, associée à une discrétisation spatiale d'ordre élevé, pour les écoulements instationnaires à convection dominante. Avant de détailler les choix retenus pour le schéma numérique (*cf.* § 0.4), on revient sur l'état de l'art actuel en simulation numérique avec le paragraphe suivant.

0.2 Etat de l'art en simulation numérique

Ce paragraphe rappelle l'importance de la simulation numérique en CFD (§ 0.2.1), puis dresse un état des lieux des méthodes numériques existantes (§ 0.2.2) et futures (§ 0.2.3) dans les codes industriels pour la simulation des écoulements réels complexes. Les affirmations et les réflexions qui suivent s'inspirent principalement des différents articles suivants [158, 171, 43, 175, 96, 101, 176].

0.2.1 Importance de la simulation

Aujourd'hui c'est un fait, et Cédric Villani, médaille Fields 2010, le rappelait encore récemment : « dans tous les domaines scientifiques, la simulation numérique est véritablement devenue le troisième pilier de la recherche, aux côtés de la théorie et de l'expérience » [172].

Historiquement, l'étude de la mécanique des fluides était divisée en deux branches bien distinctes que sont la théorie et l'expérience. La première atteint ses limites lorsque les configurations géométriques se complexifient, que les modèles d'équations à résoudre ne sont plus simplifiables et que les solutions exactes restent inconnues. La seconde trouve ses limites dans la faisabilité technique ou temporelle d'une manipulation, et plus globalement lorsqu'il s'agit de parler budget. En réponse à ces limitations, la simulation numérique est apparue comme la troisième branche, et ce dès 1950, de pair avec l'augmentation des capacités informatiques de stockage mémoire et de temps de calcul (ou temps « CPU »⁷).

Cependant, la simulation ne constitue en aucun cas une substitution aux deux autres branches. En aéronautique, les essais en soufflerie sont toujours nécessaires tant que le « graal » de la simulation numérique n'est pas atteint, à savoir : la simulation numérique d'écoulements réalistes à physique complexe autour de configurations complètes d'engins (avions, drones, hélicoptères, *etc.*). Pour de telles applications industrielles, toute la difficulté du choix d'une méthode numérique repose sur le compromis entre la précision de la solution numérique, le temps de calcul nécessaire et la flexibilité (en lien avec le spectre des applications) du schéma complet.

7. Acronyme anglais : « **C**entral **P**rocessing **U**nit ».



Fig. 1 – Iso-valeurs de la pression autour d'un avion de transport civil complet en hypersustentation (fuselage, nacelles, ailes : becs, volets et spoilers braqués) pour la première fois et pour la prédiction des performances en configuration d'atterrissage. Calcul réalisé en 2011 avec le logiciel « *elsA* » [21] de l'Onera. Détails : calcul tridimensionnel en régime stationnaire et turbulent, utilisant une approche RANS et une méthode des volumes finis d'ordre deux, ainsi qu'un maillage structuré contenant plus de 33 millions de points. Illustration extraite de [21, figure 3], où les couleurs ont été converties en échelle de gris.

0.2.2 Codes industriels actuels

Quotidien des bureaux d'étude. L'état de l'art des codes CFD à caractère industriel couramment utilisés est essentiellement fondé sur les méthodes numériques d'ordre deux, dont les plus fréquentes sont les méthodes d'éléments finis et de volumes finis. Le monopole industriel de ces méthodes de bas ordre s'explique simplement par l'accent qui a été mis sur elles entre les années 70 et 90, leur conférant aujourd'hui un très bon niveau de maturité. En plus d'être simples à mettre en œuvre, elles sont alors précises, stables et efficaces sur de nombreuses applications, dans le sens où leurs résultats sont très satisfaisants en regard des marges d'erreurs acceptables en ingénierie (5% en général) [176].

Aujourd'hui en aéronautique, les cas d'écoulements les plus complexes régulièrement simulés dans les bureaux d'études sont les écoulements tridimensionnels, stationnaires et turbulents (approche RANS) sur l'enveloppe complète d'aéronefs [171]. A ce titre, les capacités du logiciel *elsA* développé par l'Onera [21] sont un bel exemple comme l'illustre la figure 1.

Vers d'autres besoins. En dépit des avancées significatives de ces vingt-trois dernières années, un grand nombre de problématiques de la CFD restent hors de portée des codes actuels, en particulier l'absence de prise en compte des instationnarités manque le plus à l'appel. En aéronautique, les problèmes les plus souvent mentionnés sont d'une part, la représentation et la compréhension des écoulements turbulents fortement décollés autour d'aéronefs au décollage et à l'atterrissage, et d'autre part, la prédiction et le contrôle des sources de bruits comme celle du train d'atterrissage. En plus de ces problèmes multi-échelles, on cite ceux multi-physiques tels que l'aéroacoustique, l'aéroélasticité, l'aérothermique, *etc.*

Pour les méthodes d'ordre deux, les barrières les plus grandes à la faisabilité de ces simulations sont la précision de la solution et le temps CPU de la résolution. D'un côté, ces méthodes induisent trop de dissipation numérique et de ce fait, elles

empêchent de prédire les phénomènes de transport sur de longues distances. Ceci est particulièrement vrai pour l'approche URANS. D'un autre côté, elles requièrent des maillages de grande résolution pour capturer les phénomènes à petites échelles et par conséquent, les temps de calcul sont rédhibitoires pour les supercalculateurs actuels. Ceci est notamment vrai pour les approches LES et DNS.

In fine, il faut aujourd'hui envisager des alternatives aux méthodes d'ordre deux.

0.2.3 Codes industriels de demain

Réponses aux enjeux scientifiques. Si dans la littérature, la loi empirique de Moore, ayant trait à l'évolution de la puissance des ordinateurs et à la complexité du matériel informatique, laisse parfois espérer la simulation du graal aéronautique pour les années 2080 [158, tableau 1], les scientifiques n'en restent pas moins convaincus que cela ne suffira pas à lever la barrière sur la précision des solutions. En effet, les efforts doivent porter conjointement sur le perfectionnement des modèles de turbulence, le développement d'outils numériques permettant de passer des ordres supérieurs de précision, l'approfondissement de schémas rigoureux de génération et d'adaptation de maillages, l'amélioration des techniques d'accélération de convergence, *etc.*

Pour les Etats-Unis et l'Europe, acteurs mondiaux majeurs en aéronautique, ces efforts sont cruciaux. Ils sont coordonnés par l'AFOSR⁸ aux Etats-Unis [73] et par l'Acare⁹ en Europe [71]. Ces deux instances fixent, entre autres, les stratégies de recherche et d'innovation pour l'aéronautique de demain. Parmi les objectifs communs, on retrouve le développement d'outils numériques permettant d'obtenir des ordres supérieurs de précision, dont les méthodes d'ordre élevé.

A ce titre, on cite les deux projets européens Adigma (2006-2009) et Idihom (2010-2013). Le premier concerne le développement de méthodes adaptatives d'ordre élevé pour les écoulements compressibles en aéronautique [96]. Le second concerne la maturité des méthodes d'ordre élevé pour des applications à caractère industriel [77]. On mentionne également les deux premiers ateliers internationaux sur ces méthodes, organisés en vue d'obtenir des cas tests de référence de difficulté croissante. Le premier, où certains travaux de cette thèse ont été présentés, était en janvier 2012 à Nashville (Etats-Unis) [83, 176], et le second était en mai 2013 à Cologne (Allemagne) [82].

Méthodes d'ordre élevé. Le principal argument en faveur des méthodes d'ordre élevé réside dans leurs propriétés de convergence qui permettent, au moins pour des solutions régulières, d'obtenir la même précision que les méthodes d'ordre deux, avec un nombre réduit de degrés de liberté. Il faut également souligner leurs faibles dissipation et dispersion numériques, et donc leur meilleure précision pour la résolution de certains phénomènes physiques comme les structures tourbillonnaires, la propagation d'ondes, les instabilités, *etc.* En aéronautique, ces méthodes ont déjà démontré leur compétitivité en matière de précisions et coûts par rapport aux méthodes de bas ordre pour les écoulements bidimensionnels, stationnaires ou instationnaires, de

8. Acronyme anglais : « Air Force Office of Scientific Research ».

9. Acronyme anglais : « Advisory council for aeronautics research in europe ».

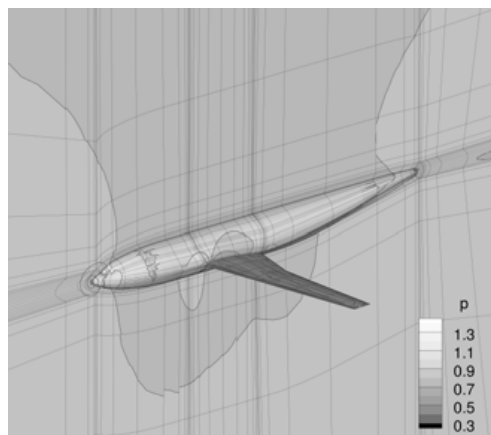


Fig. 2 – Iso-valeurs de la pression autour de la configuration « DLR-F6 »¹⁰ (fuselage et ailes) pour la prédiction des performances. Calcul réalisé en 2010 avec le code « Migale » des universités de Bergame et de Brescia [5] dans le cadre du projet Adigma. Détails : calcul tridimensionnel en régime stationnaire et turbulent, utilisant une approche RANS et une méthode d'ordre élevé de type de Galerkin discontinu d'ordre quatre, ainsi qu'un maillage d'hexaèdres contenant environ 1 million de degrés de liberté. Illustration extraite de [96, chapitre 3, figure 7].

fluides parfaits ou laminaires autour de configurations simples de profils d'aile [176]. En outre, la figure 2 illustre un avancement remarquable sur un calcul complexe.

Malgré leurs attraits et leurs démonstrations de compétences, les méthodes d'ordre élevé manquent de robustesse et restent coûteuses en temps de calcul. Il faut cependant rappeler qu'elles n'ont qu'une quinzaine d'années d'existence en CFD, soit vingt-cinq ans de moins que les méthodes de bas ordre.

In fine, ces nouvelles méthodes sont à percevoir comme une réponse aux limitations actuelles des codes CFD industriels, avec des améliorations scientifiques à mener pour les porter, dans un premier temps, au niveau de maturité des méthodes d'ordre deux.

0.3 Méthode de Galerkin Discontinue

Les méthodes d'ordre élevé sont nombreuses, leur spectre de précision allant jusqu'aux méthodes spectrales, qui sont les méthodes les plus précises avec une convergence exponentielle. Concernant le système des équations d'Euler ou de Navier-Stokes, le lecteur trouvera dans les articles de Ekaterineras [43] et de Wang [175] un panel de ces méthodes pour le cas des maillages respectivement, structurés et non structurés, et dans l'article de Lê *et al.* [101] un aperçu de ces méthodes à l'Onera.

Dans ce travail, la méthode utilisée est la méthode de Galerkin Discontinue (ou méthode GD). Dans ce paragraphe, son principe est tout d'abord présenté (§ 0.3.1), puis ses avantages et inconvénients (§ 0.3.2), et enfin son historique (§ 0.3.3).

10. Acronyme allemand : « Deutsches Zentrum für Luft- und Raumfahrt ».

0.3.1 Principe

La méthode GD fait le lien entre les méthodes de volumes finis et d'éléments finis. Tout d'abord, le domaine de calcul est partitionné en éléments formant un maillage. La solution numérique (ou solution discrète) est approchée par une combinaison linéaire de fonctions de forme (ou fonctions de base) en certains points du maillage. Ces fonctions sont bien souvent des polynômes par morceaux (en volumes finis, ce sont des fonctions constantes). Les inconnues du système (ou degrés de liberté) sont alors, soit les composantes de la solution numérique dans la base de polynômes et dans chaque élément (on parle de représentation modale), soit les valeurs de la solution en des points d'interpolation (on parle de représentation nodale). Leur détermination est réalisée comme suit.

A l'image de la méthode des éléments finis : le problème fort de départ est mis sous formulation faible (ou formulation variationnelle) en le multipliant par des fonctions tests engendrant l'espace des fonctions (ou espace d'approximation) dans lequel la solution est recherchée. Puis une intégration par parties est réalisée sur chaque élément du domaine (en éléments finis, cette intégration est réalisée sur tout le domaine). Les fonctions de forme et les fonctions tests sont choisies dans le même espace (méthode de Galerkin), cependant aucune continuité n'est imposée aux interfaces entre les éléments du domaine de calcul discrétisé (on parle de la méthode des éléments finis non conforme). L'expression de ces fonctions dépend de la géométrie des éléments du maillage.

A l'image de la méthode des volumes finis : les hypothèses de régularité sur la solution ne concernent que l'intérieur de chaque élément et par conséquent, cette dernière n'est pas définie de manière unique aux interfaces. Les flux physiques sur une interface sont alors approchés par des flux numériques faisant intervenir les valeurs moyennes et les sauts de la solution au niveau de l'interface, et permettant d'imposer faiblement les conditions de couplage de la solution discrète.

Choix des flux numériques : il n'est pas anodin puisqu'il détermine la consistance, la conservativité, la stabilité et la précision du schéma global et participe donc à sa convergence (théorème de Lax-Wendroff). Le principal obstacle à la détermination des flux numériques est la discontinuité globale de la solution numérique. En effet, cette dernière étant continue au sein d'un élément mais discontinue aux interfaces, la composante normale des flux à ces interfaces, qui apparaît dans l'expression discrète du schéma, n'est pas définie de manière unique et doit être approchée. De plus, pour le système des équations de Navier-Stokes, l'utilisation directe de flux numériques pour les termes de convection (associés à l'opérateur hyperbolique), à l'image de ceux utilisés dans les méthodes de volumes finis, ne pose pas de difficulté particulière. En revanche, ce n'est pas le cas pour les termes de diffusion (associés à l'opérateur parabolique), dont la présence des dérivées secondes en espace nécessite, comme on le verra, une attention particulière.

Finalement, on peut concevoir la méthode GD à la fois comme une méthode des éléments finis non conforme (ou discontinue) et comme une extension d'ordre

élevé naturelle de la méthode des volumes finis. On note que la formulation locale à l'élément de la méthode GD permet la mise en œuvre naturelle d'une formulation d'ordre élevé. Enfin, il existe une grande variété de méthodes GD selon le choix de la géométrie des éléments, des fonctions de forme et des fonctions tests, ou encore des flux numériques. Le lecteur peut consulter [28, 38] pour plus de détails sur les méthodes GD.

0.3.2 Avantages et inconvénients

Avantages. La méthode GD regroupe les avantages cumulés des méthodes de volumes finis et d'éléments finis. Ceux souvent mis en avant sont les suivants.

- Le découplage du système d'équations, c'est-à-dire que les matrices intervenant dans le système linéaire à résoudre sont creuses et structurées par blocs de taille égale au nombre de degrés de liberté d'un élément du maillage. Ce point permet d'envisager des techniques spécifiques de simplifications pour la résolution.
- La taille du stencil est indépendante de l'ordre de précision, c'est-à-dire que le calcul du résidu discret des équations ne dépend que de la solution dans l'élément et de ses voisins directs. On parle de méthode compacte. Ce point permet l'obtention d'un ordre arbitrairement élevé et l'adaptation aux architectures de code parallèle.
- L'application des conditions aux limites est faite au sens faible, c'est-à-dire sans l'imposer au travers de la définition de l'espace d'approximation, permettant ainsi une mise en œuvre simplifiée de la méthode GD.
- La sensibilité à la régularité des maillages est faible, grâce à la discontinuité de la solution entre les éléments. Ce point permet, non seulement l'adéquation avec des maillages non structurés et non conformes pour représenter les géométries industrielles, mais aussi le développement de stratégies de raffinement. Pour ces dernières, on note que l'idéal est de combiner un raffinement du maillage (adaptation h) dans les zones de moindres régularités de la solution avec une augmentation de l'ordre d'approximation (adaptation p) là où la solution est plus régulière. On parle de méthode d'adaptation hp .

Inconvénients. Certains désavantages de la méthode GD peuvent être cités.

- En présence de discontinuités (par exemple des chocs), on observe une perte de précision et de sens physique de la solution numérique, ainsi que l'apparition d'oscillations parasites. On parle de manque de robustesse de la méthode. Il faut développer des techniques de stabilisation comme un limiteur de pente ou ajouter de la viscosité artificielle au schéma numérique.
- Les résolutions temporelles sont encore inefficaces car jusqu'à il y a cinq ans environ (2008) l'accent des méthodes GD a porté sur le développement de discrétisations génériques, stables, consistantes et compactes. En particulier pour les stratégies explicites, il faut contourner les restrictions de stabilité contraignantes sur le pas de temps, et pour les stratégies implicites, il faut optimiser des méthodes de résolution de grands systèmes linéaires.

- Les coûts en encombrement mémoire et en temps CPU sont importants à cause des valeurs doubles de la solution aux interfaces augmentant le nombre de degrés de liberté. Il faut mettre l'accent sur des techniques de simplification des calculs et d'accélération de convergence.
- L'erreur géométrique peut devenir prépondérante devant l'erreur du schéma numérique en présence de frontières matérielles courbes. Il faut envisager des techniques de représentation iso-paramétrique de ces frontières et développer des générateurs de maillages adéquats.

Pour conclure sur cette présentation de la méthode GD, le tableau 1 compare les trois méthodes des éléments finis, des volumes finis et GD en donnant une idée de leurs qualités de réponse vis-à-vis des demandes actuelles en CFD.

| La méthode est-elle ... | VF | EF | GD |
|--|-----|----|-----|
| ... adaptée à la capture de chocs ? | ++ | + | ++ |
| ... adaptée à l'ordre élevé (supérieur à 2) ? | + | ++ | +++ |
| ... avec une faible dissipation ? | + | ++ | +++ |
| ... stable pour les lois de conservation ? | +++ | + | ++ |
| ... à stencil compact ? | + | ++ | +++ |
| ... sensible à la qualité du maillage ? | ++ | ++ | + |
| ... adaptée aux maillages non structurés ? | ++ | ++ | +++ |
| ... adaptée aux maillages non conformes ? | + | + | +++ |
| ... adaptée aux maillages mobiles / déformants ? | ++ | ++ | +++ |
| ... adaptée aux techniques d'adaptation ? | ++ | + | +++ |
| ... complexe à programmer ? | + | ++ | +++ |
| ... coûteuse en ressources mémoire ? | + | ++ | +++ |
| ... coûteuse en temps CPU ? | + | ++ | +++ |
| ... adaptée à une parallélisation ? | ++ | + | +++ |

Tab. 1 – Comparaison des méthodes de volumes finis (VF), d'éléments finis (EF) et de Galerkin discontinue (GD) s'inspirant de [98, tableau 1.2]. La signification des symboles est relative entre les méthodes et à considérer par ligne :

+ : peu / faible ; ++ : moyen ; +++ : beaucoup / fort.

0.3.3 Historique

L'historique de la méthode GD est maintenant présenté en insistant sur les axes importants au cœur de ce travail. Le lecteur peut consulter [26, partie I] et [62, chapitre 1] pour un historique plus détaillé.

0.3.3.1 Au commencement

La méthode GD a été introduite il y a exactement 40 ans par Reed et Hill (1973) [133] pour la résolution de l'équation stationnaire de transport des neutrons. Une analyse mathématique d'équations linéaires, scalaires et d'ordre un en espace

a été menée, entre autres, en premier par Lesaint et Raviart (1974) [105], puis par Johnson et Pitkaranta (1986) [85], Richter (1988) [142] et plus récemment par Houston *et al.* (2000 et 2002) [67, 68].

En particulier en supposant que la solution est suffisamment régulière, Johnson et Pitkaranta ont démontré une convergence en norme L^2 en loi de puissance de $\mathcal{O}(h^{p+1/2})$ pour une méthode GD(p) avec p le degré total maximum du polynôme d'approximation de la solution numérique et pour un partitionnement quelconque du maillage. Richter a établi un ordre de convergence optimal de $\mathcal{O}(h^{p+1})$ dans le cas de certains maillages structurés et non cartésiens, en deux dimensions d'espace. Enfin, Houston a prouvé une convergence en loi exponentielle de $\mathcal{O}(e^{-f(h)p})$ pour f une fonction de h , en supposant que la solution est analytique par morceaux.

0.3.3.2 Cas des opérateurs de premier ordre en espace

La méthode GD a d'abord été développée pour les équations aux dérivées partielles du premier ordre en espace. La plupart des développements ont été portés par Cockburn, Shu et leurs collègues. Ils sont synthétisés dans la revue très complète [28] et font suite à une série de papiers des mêmes auteurs. En particulier, leur travail souligne l'état de l'art de la résolution des systèmes d'équations hyperboliques de lois de conservation, non linéaires et dépendantes du temps en plusieurs dimensions d'espace. Il fournit des informations importantes sur la discrétisation, le choix des flux numériques, les limiteurs et le traitement des conditions aux limites aux frontières du domaine de calcul.

Dans le cas d'un opérateur hyperbolique, la plupart des flux numériques sont issus du développement des méthodes de volumes finis et respectent les propriétés de consistance et de conservativité du schéma. Ils sont solutions exactes ou approchées de problèmes de Riemann locaux, et décentrés puisque pour les problèmes hyperboliques les flux centrés sont inconditionnellement instables. Les flux numériques les plus souvent utilisés étant le flux exact de Godunov, ou ceux approchés de Engquist–Osher, Harten–Lax–van Leer (ou HLL et HLLC dans sa version modifiée), de Lax–Friedrichs (ou LxF et LLF dans sa version locale), Roe, de Vijayasundaram, *etc.* Le lecteur peut consulter le livre de Toro [162] pour plus de détails. Dans le cadre de la discrétisation du système des équations d'Euler (de type hyperbolique) avec un schéma GD en espace et explicite en temps, Cockburn et Shu [28] ainsi que Qiu *et al.* [131] effectuent une comparaison de plusieurs des flux mentionnés ci-dessus. La conclusion la plus remarquable est que selon l'expérience numérique, plus le degré p du polynôme augmente, moins le flux numérique choisi n'a d'importance sur la précision de la solution numérique.

0.3.3.3 Cas des opérateurs de second ordre en espace

L'extension de la méthode GD aux équations aux dérivées partielles du second ordre en espace a été étudiée séparément par deux communautés : celle des éléments finis discontinus, dès 1971 avec le travail de Nitsche, et celle GD, dès 1997 avec le travail de Bassi et Rebay. La première utilise une technique de pénalisation intérieure pour imposer faiblement aux interfaces les conditions aux limites de la solution et de ses dérivées, tandis que la seconde reconsidère l'opérateur du second ordre en espace

en un système d'équations du premier ordre (on parle de formulation mixte). Ces deux formulations ont été unifiées, analysées et comparées dans le célèbre article de Arnold *et al.* [4]. En particulier, cet article s'intéresse aux propriétés de consistance, de conservativité, de stabilité, de précision et à la taille du stencil des différentes formulations.

Dans le cas d'un opérateur elliptique ou parabolique, une approche directe (similaire aux termes de convection) des flux physiques visqueux par des flux numériques conduit à un schéma numérique inconsistant : c'est un phénomène dit de « crime variationnel » [28, 183]. Pour palier le problème d'inconsistance, deux techniques principales existent.

- La première est celle de la formulation mixte utilisée par Bassi et Rebay [9] en 1997. Elle introduit une équation supplémentaire pour le calcul des gradients qui prend en compte le saut de la solution aux interfaces. En fonction de la formulation, le schéma est compact ou non. Par ailleurs, le schéma a besoin d'être stabilisé, soit par des termes de pénalité intérieure, soit par des termes de viscosité artificielle avec, dans les deux cas, un paramètre à ajuster. Parmi les principales contributions où le schéma (donc le flux numérique) est à stencil compact, on peut citer : Bassi et Rebay avec leur second schéma dit BR2 [13, 7], Hartmann et Houston avec le schéma SIP [58], Peraire et Persson [124] avec leur schéma CDG. Pour la lecture de la suite de cette introduction, on mentionne également les deux schémas non compacts LDG de Cockburn et Shu [27] et BR1 de Bassi et Rebay [9] (*cf.* [4] pour une bibliographie plus complète).
- La seconde technique est celle de reconstruction ou de recouvrement local de la solution mise en œuvre par van Leer et Nomura [169, 170] en 2005. Elle consiste à fusionner localement des éléments partageant une interface, puis à calculer les flux sur ce nouvel élément en définissant localement une solution régulière de degré supérieur à celle que l'on aurait sans fusionnement. Cette technique permet alors de lisser les discontinuités aux interfaces et de s'abstenir localement des sauts de la solution et de son gradient. Bien qu'il ne soit plus question d'introduire un terme de pénalité ou de régler des paramètres, un obstacle reste : la construction d'une base locale pour le fusionnement dont la compacité du schéma est fortement dépendante. Parmi les principales contributions, on cite van Leer et Nomura avec la méthode RDG, Luo *et al.* avec la méthode rDG, Borrel et Ryan avec la méthode EDG, *etc.* (*cf.* [16] pour une bibliographie plus complète).

0.3.3.4 Association à une discrétisation temporelle

Depuis les années 2000, l'accent porte sur la recherche de discrétisations temporelles efficaces associées à la méthode GD. Parmi celles-ci, les méthodes explicites et implicites sont les plus courantes [43, 175].

D'un côté, les méthodes explicites sont relativement précises, faciles à mettre en œuvre, bien adaptées à la parallélisation, et ne requièrent pas de stockage mémoire conséquent. En revanche, ces méthodes sont conditionnellement stables imposant de fortes restrictions sur le pas de temps. Ce dernier point est particulièrement vrai pour les écoulements à convection dominante étudiés, où la résolution des différentes

échelles de l'écoulement demande d'utiliser des éléments de maillage de tailles très différentes. En GD, les méthodes explicites de type Runge-Kutta (ou RK) et Euler ont été parmi les premières étudiées, de pair avec des limiteurs de pentes ou des conditions TVD¹¹ pour stabiliser le schéma en présence des non-linéarités de l'écoulement [28].

D'un autre côté, les méthodes implicites sont inconditionnellement stables et permettent « en théorie » de lever les restrictions sur le pas de temps. Cependant, ces méthodes sont plus lourdes à mettre en œuvre en particulier à cause de la matrice du système linéaire à inverser. Cette dernière est d'autant plus grande que l'on souhaite un schéma précis. En particulier, la possibilité d'inverser ces larges systèmes est un point critique pour rendre efficace l'utilisation des méthodes GD. Ce point est au centre des préoccupations actuelles [28, 39].

En GD, la résolution du système linéaire implicite par une méthode directe n'est pas envisageable pour des applications réelles à cause des trop gros coûts en stockage mémoire et en temps de calcul. Les caractéristiques de la matrice implicite, dont sa structure par blocs, conduisent naturellement au choix de méthodes itératives. Le développement de ces dernières inclut la recherche de compatibilité avec une architecture parallèle, de techniques d'accélération de convergence (par exemple multi-grilles) et de simplifications des calculs du schéma, de schémas d'adaptation *etc.* A cette fin, plusieurs auteurs ont appliqué une méthode itérative de type Newton-Krylov ou de Runge-Kutta implicite avec succès sur le système des équations d'Euler ou de Navier-Stokes laminares ou turbulentes (approche RANS) pour les fluides compressibles [132, 12, 7, 57, 40].

De pair avec les méthodes itératives, différents préconditionnements ont été introduits, soit ils exploitent la structure de la matrice implicite et/ou ré-ordonnent les degrés de liberté [132, 46, 89, 127, 39, 15, 126], soit ils utilisent une méthode de condensation statique [59, 154]. Par ailleurs, concernant les techniques d'accélération de convergence, des méthodes multi-grilles en h ou p ont été envisagées dans [46, 114, 39]. Concernant, les simplifications possibles du schéma, on cite par exemple Yasue *et al.* [181] qui utilisent un schéma dit localement implicite car seuls les termes de l'élément courant sont implicités (« *pointwise relaxation algorithm* »), schéma qui est associé à une construction simplifiée de la matrice implicite pour les systèmes des équations de d'Euler et de Navier-Stokes en trois dimensions d'espace. Dolejší *et al.* [40] proposent, quant à eux, une méthode semi-implicite, où les termes linéaires du schéma discret sont traités implicitement et ceux non linéaires sont traités explicitement par extrapolation.

En outre, si la plupart des approches de la méthode GD consistent à discrétiser en espace, puis en temps, par une méthode des lignes, en considérant séparément les variables spatiales et la variable temporelle, une autre approche existe : la méthode GD espace-temps. Dans cette dernière, les variables sont couplées en espace et en temps, ce qui permet, entre autres, d'utiliser des maillages en mouvement et déformants. Elle a été introduite en 2002 par van der Vegt et van der Ven [167] et a été appliquée au système des équations de Navier-Stokes compressibles en 2006 par Klaij *et al.* [89].

11. Acronyme anglais : « **T**otal **V**ariation **D**iminishing ».

0.3.3.5 Applications à la CFD

La méthode GD connaît depuis plusieurs années un fort développement tant pour la résolution des problèmes hyperboliques que pour celle des problèmes elliptiques. Ce développement est bien illustré par la figure 3. En particulier, le premier atelier de travail international organisé à Newport (Etats-unis) en 1999 [26] sur ce sujet, ainsi que celui plus récent à Banff (Canada) en 2007 [29] donnent un large panorama des résultats théoriques, numériques et d'applications obtenus avec cette méthode. Un nombre important de résultats sur le cadre théorique est également disponible avec les récents ouvrages de Hesthaven et Warburton [62], Kanschäat [86], Rivière [143] et Di Pietro et Ern [38].

Dans ce travail, on s'est particulièrement intéressé aux applications de la méthode GD pour les systèmes des équations d'Euler [10, 109, 174, 128, 8] et de Navier-Stokes [9, 11, 7, 46, 58, 89, 90, 127, 39, 40, 16, 110, 50, 61, 15] (les citations mentionnées sont ici restreintes à celles utilisées dans la suite de ce mémoire).

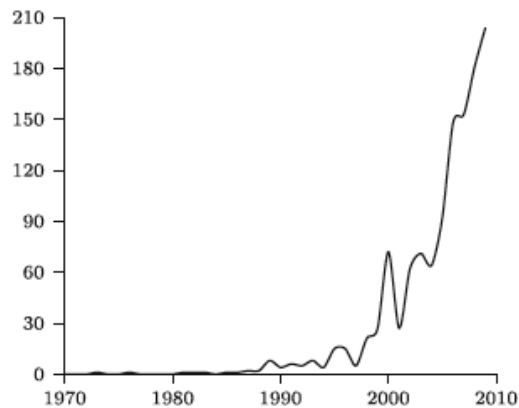


Fig. 3 – Nombre d'entrées de la base documentaire MathSciNet contenant l'expression anglaise « Discontinuos Galerkin » en fonction des années. Figure extraite de [38, figure 1].

0.4 Travaux de thèse

En reprenant les points forts des précédents paragraphes, il ressort le bilan suivant. Si la méthode d'ordre élevé de type GD peut répondre aux besoins de simulation numérique des écoulements instationnaires à convection dominante, son association avec une méthode en temps efficace, en matière d'encombrement mémoire et de temps de calcul, n'est en revanche pas au point.

C'est dans ce contexte que cette étude a été développée. Cette dernière consiste à **mettre en œuvre une stratégie d'intégration temporelle explicite-implicite efficace, associée à une discrétisation spatiale GD d'ordre élevé, pour les écoulements instationnaires à convection dominante de fluides visqueux compressibles modélisés par le système des équations de Navier-Stokes.**

Ce paragraphe est consacré à la présentations de cette étude. Il détaille d'abord l'ensemble de ses motivations et de ses contributions (§ 0.4.1), puis son plan par chapitre (§ 0.4.2), et enfin les communications qui en sont issues (§ 0.4.3).

0.4.1 Motivations et contributions

Lors de cette thèse, on a développé l'ensemble des outils nécessaires à la simulation numérique du système des équations de Navier-Stokes pour des solutions régulières. Il s'agit de la génération de maillages, du code de calcul dédié pour le schéma discret en espace et en temps et de l'analyse des résultats. On résume ici l'essentiel des contributions sans respecter leur ordre d'apparition dans le mémoire, en insistant sur les principales difficultés et en soulignant les résultats majeurs.

0.4.1.1 Approximation d'ordre élevé

Validation des ordres de convergence. La discrétisation spatiale GD proposée est d'ordre élevé. En particulier, on utilise une représentation polynomiale de la solution jusqu'à l'ordre sept. De plus, les calculs numériques sont effectués au moyen d'un code informatique dédié, lequel est issu d'un code interne Onera qui a été généralisé et complété pour les besoins de ce travail. Les résultats obtenus sont validés sur cinq cas tests d'écoulements réguliers internes ou externes de fluides parfaits ou visqueux, bien référencés dans la littérature :

1. l'écoulement interne, stationnaire et subsonique de fluide parfait dans un canal de type convergent-divergent [176, 83] (cas test 1) ;
2. l'écoulement interne, instationnaire et subsonique d'une paire de tourbillons visqueux et contra-rotatifs se dirigeant perpendiculairement à une paroi plane [87] (cas test 2) ;
3. l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour du profil d'aile « NACA0012 »¹² [96, 40] (cas test 3) ;
4. l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour du profil d'aile de Selig et Donovan (ou SD) SD7003 [166, 113] (cas test 4) ;
5. l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour et en aval d'un cylindre [79] (cas test 5).

La convergence en maillage en loi de puissance $p+1$ pour le degré p d'un polynôme donné [142] et celle en ordre en loi exponentielle p pour un maillage donné [68] sont vérifiées en détail sur les cas tests 1 et 2. La bonne valeur des ordres obtenus valide, non seulement la métrique du schéma numérique, mais également le choix des flux numériques.

Utilisation de flux numériques à stencil compact. Pour la méthode GD, le choix de flux numériques à stencil compact est une clé incontournable à une parallélisation efficace de la méthode. Comme on l'a déjà mentionné, ce choix est particulièrement délicat pour les termes de diffusion [28, 183]. En effet, sur l'exemple de l'équation de la chaleur monodimensionnelle scalaire, on démontre qu'un choix de flux similaire à celui des termes de convection conduit à un schéma numérique inconsistant.

Ici, pour les termes de convection, on utilise le flux LLF [156] dans une version modifiée [33] permettant de stabiliser le schéma par l'ajout de dissipation numérique.

12. Acronyme anglais : « National Advisory Committee for Aeronautics ».

Ce flux a été retenu pour sa simplicité de mise en œuvre sur un système d'équations non linéaires multi-dimensionnelles, pour son faible coût en nombre d'opérations à virgules flottantes par seconde et enfin pour ses performances d'approximation [28, 131, 96, 33].

Pour les termes de diffusion, on utilise la technique de formulation mixte appliquant le flux (ou schéma) BR2 [13, 7], lequel introduit des opérateurs de relèvement pour approcher le saut du gradient de la solution numérique aux interfaces. En particulier, ce choix fait suite à plusieurs études préliminaires sur des équations d'advection-diffusion scalaires linéaires et non linéaires en deux dimensions d'espace incluant les deux autres flux BR1 [9] et LDG [27]. Ces études ne sont pas présentées dans ce mémoire. Toutefois, on indique que le flux BR2 est celui qui montre les meilleures propriétés de compacité, de stabilité et de convergence réunies, lors d'une discrétisation explicite en temps. En outre, il garantit un ordre de convergence optimal de la norme L^2 en loi de puissance.

0.4.1.2 Représentation des frontières matérielles courbes

Représentation polynomiale des géométries courbes. En GD, il est nécessaire de prendre en compte les frontières matérielles courbes. Par exemple en 1997, Bassi et Rebay [10] ont simulé un écoulement bidimensionnel isentropique autour d'un cercle, approché par un polygone. Ils ont alors montré qu'une création importante d'entropie avait lieu à chaque sommet du polygone, donnant naissance à des oscillations non-physiques dans le sillage, et ce même avec un maillage raffiné.

Dans ce travail, pour les cas tests présentant des parois matérielles courbes, on utilise une représentation polynomiale de degré deux ou quatre, passant par trois ou cinq points pour décrire une frontière dans un élément physique du maillage [151].

Construction d'une base orthogonale de polynômes. La prise en compte des géométries courbes rend difficile le choix de la base orthonormale et hiérarchique dans laquelle on cherche la solution numérique. En effet, dans la pratique les polynômes qui engendrent la base sont évalués en des points de quadrature, eux-mêmes contenus dans un élément de référence. Il faut alors introduire une transformation de cet élément vers chaque élément physique courbe. Le calcul du Jacobien de cette transformation rend trop complexe la construction d'une base orthogonale dans l'élément de référence comme par exemple, celle de Dubiner [42] qui utilise le produit tensoriel de familles de polynômes définis sur un intervalle de l'axe des réels.

Pour pallier ce problème, on définit une base numérique directement dans l'élément physique à partir d'une base de monômes définis en coordonnées barycentriques (pour un meilleur conditionnement) et orthonormalisés par une procédure de Gram-Schmidt modifiée [134].

Avec le cas test 1, on teste l'efficacité de la représentation retenue pour les géométries courbes. En particulier, on montre que les convergences en maillage et en ordre sont garanties lors de l'utilisation d'une représentation iso-paramétrique des frontières.

0.4.1.3 Discrétisation temporelle adaptée aux écoulements étudiés

Implicitation des termes de flux visqueux. L'association d'une méthode GD à un schéma explicite en temps impose de grandes restrictions sur le pas de temps en raison des conditions de stabilité. En effet, des expériences numériques préliminaires sur l'équation linéaire d'advection pure en une dimension d'espace montrent que pour une discrétisation GD de degré p , la restriction est de type **Courant-Friedrichs-Lewy** (ou CFL) et décroît linéairement en p [28], tandis que pour l'équation linéaire de diffusion pure la restriction est de type Von Neumann et décroît quadratiquement en p [108, 51]. En général, lors de la simulation des écoulements à convection dominante, la capture correcte des phénomènes dans la couche limite en zone de proche paroi impose de raffiner le maillage à cet endroit et conduit localement à $Re_h \ll 1$. La valeur du pas de temps est alors fixée par celle du pas de temps visqueux et devient très petite, voire réductrice pour des applications réelles.

Dans ce travail, on choisit alors de discrétiser implicitement les termes de diffusion afin de s'affranchir de la condition de stabilité très contraignante de type Von Neumann, tout en gardant une discrétisation explicite pour les termes de convection afin de restituer correctement les mécanismes de transport.

Résolution séparée des termes de convection et de diffusion. Dans [139], on s'est intéressé à la résolution de l'équation d'advection-diffusion linéaire scalaire avec une discrétisation spatiale GD et une discrétisation temporelle explicite-implicite (comme expliqué ci-dessus) utilisant un schéma d'Euler sur les termes de convection et d'Euler rétrograde sur les termes de diffusion. Par une analyse de Von Neumann, on montre que les conditions de stabilité du schéma numérique sont :

$$\Delta t \leq \frac{2\nu}{c^2}, \quad Re_h \leq 6, \quad (1)$$

avec c la vitesse de convection. La première condition correspond à une limitation du pas de temps et la seconde correspond à une restriction de la méthode aux écoulements à diffusion dominante. En particulier, cette dernière est inappropriée pour les écoulements à convection dominante de ce travail.

C'est pourquoi, on remédie à cette limitation en utilisant une décomposition d'opérateurs de type Strang [160]. Cette décomposition permet de résoudre séparément les intégrations des termes de convection et de diffusion, et de proposer ainsi une méthode numérique dédiée pour chacun des opérateurs associés aux différents termes. En particulier, le schéma explicite retenu est celui de Heun [63, 52], et celui implicite est celui de Crank-Nicolson [34, 36]. Dans ce dernier cas, pour inverser le système linéaire, on a recours à une méthode itérative GMRes¹³ à redémarrages et préconditionnée avec l'algorithme de Jacobi par blocs.

Avec l'ensemble des cas tests 2 à 5, on montre à la fois que la stratégie temporelle explicite-implicite retenue lève la restriction de type Von Neumann par rapport à une stratégie explicite, et qu'elle est très bien adaptée aux écoulements instationnaires à convection dominante. En particulier les cas tests 4 et 5 sont de très bons exemples d'applications de ces écoulements, en raison de leur couche limite fortement décollée.

13. Acronyme anglais : « **G**eneralized **M**inimum **R**esidual ».

0.4.1.4 Simplifications des discrétisations temporelles

Les schémas implicites en temps conduisent à l'inversion de systèmes linéaires d'autant plus grands que le maillage est raffiné ou que la solution recherchée est précise, et donc à des temps de calcul rédhitoires. Par exemple, lors de la résolution du système des équations RANS avec une méthode GD utilisant le flux BR2 pour les termes de diffusion, Crivellini et Bassi [35] montrent que le temps de construction de la matrice implicite est croissant avec le degré p du polynôme et atteint 80% du temps total de résolution pour $p = 6$.

Dans ce travail, on met en place deux stratégies de simplification du schéma numérique implicite, que l'on combine ensuite.

- La première simplification est celle sans jacobienne car elle approche le produit de la matrice jacobienne de dérivation du résidu discret par un vecteur au moyen d'une différence finie d'ordre un. Elle permet une réduction des coûts (encombrement mémoire et temps CPU) lors de la construction de la matrice implicite et du produit de cette matrice avec un vecteur dans l'inversion GMRes. On précise que cette simplification a été développée très récemment (2008) pour les écoulements de fluides visqueux laminaires [123, 121, 91, 15] et turbulents [35] dans le cadre de la méthode GD.
- La seconde simplification exploite le caractère modale et celui hiérarchique de la base de polynômes. Elle repose sur un découplage des degrés de liberté de la solution numérique à l'intérieur d'un élément du maillage avec ceux des éléments voisins. Elle permet une réduction des coûts lors des constructions de la matrice implicite et de la matrice de préconditionnement, et lors du produit de ces deux matrices avec un vecteur dans l'inversion GMRes. On précise que cette simplification a été initiée en 2010 sur l'équation de la chaleur scalaire en une dimension d'espace [37], puis a été mise en œuvre avec succès sur des équations d'advection-diffusion scalaires, linéaires et non linéaires, en deux dimensions d'espace pour les écoulements à diffusion [139] ou convection dominante [53, 138]. Très récemment (2013), le découplage des degrés de liberté, tel qu'il est appliqué pour cette simplification, a été réutilisé dans [15] mais uniquement sur la matrice de préconditionnement, alors qu'ici il s'applique aux deux matrices implicite et de préconditionnement.

Avec l'ensemble des cas tests 2 à 5, on montre, non seulement que les trois stratégies temporelles simplifiées capturent les phénomènes physiques à l'identique de la méthode non simplifiée, mais également qu'elles préservent les propriétés de stabilité de cette dernière. Par ailleurs, les gains théoriques et obtenus avec les méthodes simplifiées, en comparaison avec la méthode sans simplification, sont en bon accord. Notamment, la méthode combinant les deux simplifications est jusqu'à 3, 25 fois plus rapide sur la résolution totale que la méthode sans simplification.

0.4.1.5 Génération de maillages d'ordre élevé

D'un côté, un maillage adéquat pour simuler des écoulements instationnaires à convection dominante est un maillage où la taille caractéristique des éléments est adaptée aux échelles locales des phénomènes à capturer. D'un autre côté, un

maillage judicieux en GD est aussi un maillage grossier pour exploiter le potentiel de résolution en h/p de la méthode, et tenant compte des géométries courbes pour garantir un ordre élevé.

Dans ce travail, on a construit l'ensemble des maillages des cas tests 2 à 5 simulant des écoulements de fluides visqueux à l'aide du logiciel libre Gmsh permettant de générer des maillages curvilignes [54, 55]. Ces maillages sont à la fois non structurés triangulaires pour s'adapter aux géométries complexes, et fortement anisotropes pour capturer les différents phénomènes physiques. Ils présentent, dans les zones de proche paroi, des éléments étirés dans la direction du gradient de vitesse décroissant, et dans les zones de sillage, des éléments asymétriques resserrés sur la ligne de fuite. A ce titre, le rapport d'aspect global des maillages, rapport du plus grand diamètre inscrit d'un élément sur le plus petit, se situe entre 1 (maillage isotrope) et 2 125 (maillage anisotrope), et le rapport d'aspect global des éléments, rapport du plus grand côté d'un élément sur le plus petit, se situe entre 1 (élément isotrope) et 640 (élément anisotrope). Ainsi, pour l'ensemble des applications numériques, on observe toujours la très bonne capture des phénomènes physiques en comparaison aux références de la littérature.

Toutefois, on soulève ici une difficulté non négligeable pour la construction de maillages à géométries courbes. Il n'existe pas à ce jour, comme le rappelait encore le compte-rendu du premier atelier sur les méthodes d'ordre élevé [176], de générateurs de maillages curvilignes efficaces capables de générer des maillages d'ordre élevé adaptés aux zones de couche limite. Le principal problème est dû au chevauchement (ou croisement) possible des éléments proches des géométries courbes, conduisant à des valeurs négatives pour le Jacobien de la transformation des éléments courbes vers l'élément de référence. Ce souci s'est répercuté dans ce travail, par des maillages contenant, non seulement des éléments avec des rapports d'aspect moins grands que souhaités, mais également avec un nombre relativement important d'éléments.

0.4.2 Organisation du mémoire

Ce travail de thèse s'articule en la présente introduction et deux parties structurées en sept chapitres au total, lesquels sont complétés par sept annexes.

Comme l'ensemble des chapitres (hormis le quatrième) se terminent par une conclusion partielle recensant les résultats importants établis, il n'y a pas de conclusion générale dans ce mémoire. Toutefois, une dernière partie s'intéresse aux perspectives que l'on peut envisager pour la suite. Par ailleurs, l'annexe G reproduit certains papiers et articles issus de ce travail, lesquels sont détaillés au paragraphe 0.4.3.

L'idée des deux parties de ce mémoire est de séparer les aspects théoriques et pratiques :

- la partie I présente le problème physique à modéliser, ainsi que le schéma numérique discret en espace et en temps ;
- la partie II est réservée à la mise en œuvre du schéma discret dans un code de calcul, à la validation et à l'évaluation des performances de ce même schéma.

0.4.2.1 Contenu de la partie I

Le chapitre 1 présente le système des équations de Navier-Stokes (ou d'Euler) modélisant un écoulement bidimensionnel et instationnaire de fluides visqueux (ou parfaits) et compressibles, ainsi que les conditions aux limites venant les compléter.

Pour ce chapitre, l'annexe A s'intéresse à l'adimensionnement des systèmes d'équations. Elle développe le principe permettant de déterminer les nombres sans dimension caractéristiques du système, ainsi que la mise en œuvre permettant d'exprimer les variables de l'écoulement sous forme adimensionnée.

Le chapitre 2 est consacré à la discrétisation spatiale du système des équations de Navier-Stokes avec une méthode GD. Il introduit le cadre mathématique de la méthode en détaillant les étapes d'écriture de la formulation faible, de discrétisation du domaine de calcul, de définition des espaces fonctionnels et de choix des flux numériques. En particulier, le flux LLF relatif aux termes de convection et le flux BR2 relatif aux termes de diffusion sont exprimés. La dernière partie du chapitre est réservée à la prise en compte des frontières matérielles courbes du domaine de calcul, à la construction de la base de fonctions et à la formule de quadrature numérique pour approcher les intégrales du schéma.

Pour ce chapitre, l'annexe B revient sur le choix d'un flux numérique pour les termes de diffusion. Il est redémontré, sur l'exemple de l'équation de la chaleur monodimensionnelle scalaire, qu'un choix de flux similaire à celui des termes de convection conduit à un schéma numérique inconsistant (phénomène de « crime variationnel »).

Le chapitre 3 est dédié à la discrétisation temporelle du système des équations de Navier-Stokes. D'une part, il aborde la discrétisation explicite utilisant une méthode de Runge-Kutta servant à certaines validations de la partie II. D'autre part, il met en œuvre la discrétisation explicite-implicite utilisant une méthode de décomposition d'opérateurs de Strang. En particulier, la résolution explicite des termes de convection avec le schéma de Heun et la résolution implicite des termes de diffusion avec le schéma de Crank-Nicolson sont exprimées. Ensuite, les trois simplifications de la résolution implicite, reposant sur une approximation de la matrice implicite par une méthode sans jacobienne et/ou sur un découplage des degrés de liberté, sont expliquées séparément. La dernière partie du chapitre dresse un bilan théorique comparatif des intégrations temporelles explicites-implicites sur les aspects d'encombrements mémoire et de nombre d'opérations à virgule flottante.

0.4.2.2 Contenu de la partie II

Le chapitre 4 constitue une sorte de prologue aux trois chapitres suivants. Il contient la description du code de calcul dédié développé d'une part, et celle des cas tests envisagés servant à valider le code d'autre part.

Le chapitre 5 se concentre sur la validation de la discrétisation spatiale GD associée à une discrétisation temporelle explicite. Les cas tests 1 et 2 sont étudiés pour valider la discrétisation des termes de flux respectivement, convectifs et visqueux. En particulier, le cas test 1 démontre l'importance de prendre en compte les parois

courbes d'un domaine de calcul, et permet de juger la compétitivité des méthodes GD vis-à-vis d'autres méthodes d'ordre élevé grâce aux résultats du premier atelier sur les méthodes d'ordre élevé.

Pour ce chapitre, l'annexe C présente des résultats numériques complémentaires du cas test 1 sur les erreurs globales en norme L^1 , L^2 et L^∞ sur l'entropie, ainsi que les ordres de convergence associés à ces erreurs.

Le chapitre 6 permet la validation des méthodes explicites-implicites en temps associées à la discrétisation GD en espace précédemment validée. Les quatre cas tests 2, 3, 4 et 5 sont alors considérés séparément.

Ce chapitre comprend deux annexes. L'annexe D présente des résultats numériques complémentaires du cas test 2 sur les erreurs globales en norme L^1 et L^∞ sur la quantité de mouvement projetée selon une direction d'espace, ainsi que les ordres de convergence associés à ces erreurs. L'annexe E reproduit les courbes expérimentales utilisées comme les références du cas test 5.

Le chapitre 7 termine sur l'évaluation des performances des méthodes explicites-implicites en temps précédemment validées. Il étudie les gains en temps CPU sur les étapes de calculs intermédiaires et sur la résolution globale pour les trois méthodes en temps simplifiées en comparaison avec la méthode complète. Cette étude repose sur une analyse paramétrique en fonction principalement du degré du polynôme, du nombre d'éléments et du rapport d'aspect global des maillages. En particulier, les gains théoriques établis au chapitre 3 sont mis en regard avec les gains réellement mesurés.

Pour ce chapitre, l'annexe F détaille l'analyse paramétrique des gains obtenus sur la résolution totale des trois intégrations temporelles explicites-implicites simplifiées.

0.4.3 Communications

Les études menées au cours de cette thèse ont donné lieu aux publications et présentations orales suivantes.

- 06/2013 - **Article** *J. Comput. Phys.*, 251, p. 272–291 : F. Renac, S. Gérald, C. Marmignon, and F. Coquel, *Fast-time implicit-explicit discontinuous Galerkin method for the compressible Navier-Stokes equations*. Cet article [136] est reporté en annexe G, il reprend certains éléments de ce mémoire.
- 01/2012 - **Conférence** « 50th American Institute of Aeronautics and Astronautics (AIAA) Aerospace Sciences Meeting » (Etats-Unis), Papier AIAA2012-0731 : S. Gérald, F. Renac, C. Marmignon, J. Ryan, and F. Coquel, *Fast-time implicit-explicit discontinuous Galerkin method for convection-diffusion flow problems*. Ce papier [53] est reporté en annexe G.
- 01/2012 - **Atelier** « 1st international Workshop on high-order methods » [176, 83] (Etats-Unis), *Présentation des cas tests C1.1 et C1.4*.
- 09/2011 - **Séminaire des doctorants du Cermics** (Ecole des Ponts Paris-Tech, France), *Présentation des travaux de thèse*.
- 01/2010 - 01/2011 - 01/2012 - **Journées des doctorants Onera** (France), *Présentation des travaux de thèse*.

Première partie

**Écoulements bidimensionnels et
instationnaires de fluides visqueux
et compressibles : de la
modélisation au schéma numérique
en espace et en temps**

Chapitre 1

Modélisation du problème physique considéré

Dans une première partie, ce chapitre présente le système des équations de Navier-Stokes pour un fluide visqueux et compressible au sein d'un écoulement bidimensionnel et instationnaire. Dans la limite d'un nombre de Reynolds infini, ce système se ramène au système des équations d'Euler.

Dans une seconde partie, ce chapitre décrit les conditions aux limites venant compléter ces deux systèmes.

1.1 Avant-propos

Les équations de Navier-Stokes, dont l'origine remonte au XVIII^e siècle, décrivent le mouvement d'un fluide assimilé à un milieu continu. En 1755, Leonhard Euler applique les lois de Newton à des volumes infinitésimaux de fluide sans tenir compte des phénomènes de viscosité donnant vie aux équations qui portent son nom. Puis en 1822, Claude-Louis Navier complète ces équations avec un modèle de viscosité à l'échelle atomique qui est redéfini directement pour l'échelle macroscopique par Adhémar Barré de Saint-Venant en 1843 et par George Gabriel Stokes en 1845 : les équations de Navier-Stokes sont alors établies. A ces dernières s'ajoutent les deux équations de conservation de la masse et de l'énergie, formant le système des équations de Navier-Stokes.

Le système des équations de Navier-Stokes n'a jamais été aussi important que depuis l'an 2000, date à laquelle l'institut Clay de mathématiques l'a proposé dans la liste des sept « problèmes du millénaire ». En effet, pour ce système, aucun théorème d'existence global de solutions régulières assorties de conditions initiales réalistes ne peut être énoncé aujourd'hui [104, 161]. Plusieurs pistes de recherche sont toujours à l'étude, et portent sur la démonstration d'un tel théorème pour tous les temps, ou le cas échéant sur la démonstration de l'apparition de singularités en temps fini [157].

La première partie de ce chapitre, paragraphe 1.2, décrit le système des équations de Navier-Stokes pour un fluide visqueux et compressible au sein d'un écoulement bidimensionnel et instationnaire. En particulier, le système des équations d'Euler est donné dans la limite d'un grand nombre de Reynolds. La seconde partie du chapitre, paragraphe 1.3, présente les conditions aux limites venant compléter les deux systèmes d'équations.

1.2 Système d'équations pour un fluide visqueux et compressible

Dans ce paragraphe, le système des équations de Navier-Stokes est établi en tenant compte des propriétés du fluide considéré et en considérant une loi d'état sur la pression dans l'approximation des gaz parfaits (§ 1.2.1). Ensuite, des précisions sont apportées sur les nombres sans dimension caractéristiques de l'écoulement considéré, et donc du système d'équations établi (§ 1.2.2). En particulier, le système des équations d'Euler est donné dans la limite d'un nombre de Reynolds infini (§ 1.2.3).

1.2.1 Système des équations de Navier-Stokes

Le système des équations de Navier-Stokes s'obtient en considérant les principes fondamentaux de conservation de la masse, de la quantité de mouvement et de l'énergie.

Soit Ω un ouvert de \mathbb{R}^d avec d la dimension d'espace, et soit $N_{eq} = d+2$ le nombre d'équations du système. Dans ce travail $d = 2$ et donc $N_{eq} = 4$, sauf précision contraire. Soit $\mathbf{x} = (x_1, x_2)^\top$ la variable d'espace dans le repère cartésien, avec l'exposant \top pour désigner l'opérateur de transposée, et soit t la variable temporelle appartenant à $\mathcal{T} = (0, +\infty)$.

Sous forme conservative, le système des équations de Navier-Stokes s'écrit :

$$\partial_t \mathbf{w} + \nabla \cdot \mathbf{F}_c(\mathbf{w}) - \nabla \cdot \mathbf{F}_v(\mathbf{w}, \nabla \mathbf{w}) = 0, \quad \forall (\mathbf{x}, t) \in \Omega \times \mathcal{T}, \quad (1.1a)$$

$$\mathbf{w}(\mathbf{x}, 0) = \mathbf{w}^{(0)}(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega, \quad (1.1b)$$

auquel s'ajoutent des conditions aux limites appropriées (*cf.* § 1.3 pour plus de détails sur ces conditions).

Dans le système (1.1), \mathbf{w} appartient à un ouvert convexe de \mathbb{R}^{d+2} et désigne le vecteur des variables conservatives :

$$\mathbf{w} = \begin{pmatrix} \rho \\ \rho \mathbf{U} \\ \rho E \end{pmatrix},$$

où ρ est la masse volumique, \mathbf{U} appartient à \mathbb{R}^d et est le vecteur vitesse tel que $\mathbf{U} = (u_1, u_2)^\top$, et E est l'énergie totale par unité de masse telle que $E = e + e_c$ avec e l'énergie interne par unité de masse, qui doit encore être définie, et $e_c = \mathbf{U}^2/2$ l'énergie cinétique par unité de masse. L'ouvert convexe où est recherché \mathbf{w} s'appelle l'espace des états admissibles et est défini comme suit

$$\Omega_{adm} = \left\{ \mathbf{w} \in \mathbb{R}^{d+2} : \rho > 0, \mathbf{U} \in \mathbb{R}^d, e > 0 \right\}. \quad (1.2)$$

Il correspond aux domaines de définition du tenseur des flux convectifs \mathbf{F}_c et du tenseur des flux visqueux \mathbf{F}_v , lesquels appartiennent à $\mathbb{R}^{(d+2) \times d}$ et prennent la forme générale suivante :

$$\mathbf{F}_c = \begin{pmatrix} \rho \mathbf{U}^\top \\ \rho \mathbf{U} \otimes \mathbf{U} + p \mathbf{I} \\ (\rho E + p) \mathbf{U}^\top \end{pmatrix}, \quad \mathbf{F}_v = \begin{pmatrix} 0 \\ \boldsymbol{\tau} \\ \mathbf{U}^\top \cdot \boldsymbol{\tau} - \mathbf{q}^\top \end{pmatrix}, \quad (1.3)$$

où $\mathbf{U} \otimes \mathbf{U} = \mathbf{U}\mathbf{U}^\top$, p est la pression statique, \mathbf{I} est la matrice identité de \mathbb{R}^d , $\boldsymbol{\tau}$ appartient à $\mathbb{R}^{d \times d}$ et est le tenseur des contraintes visqueuses, enfin $\mathbf{q} = (q_1, q_2)^\top$ est le vecteur du flux de chaleur. La définition de ces quantités est donnée par la nature du fluide considéré. Il s'agit ici d'un fluide newtonien assimilé à un gaz parfait décrit par une loi d'état et deux modèles de loi de comportement. Les définitions de ces lois sont données ci-après.

Tout d'abord, une loi d'état du type $e = e(\rho, p)$ permet de fermer le système (1.1) de quatre équations à cinq inconnues (ρ, u_1, u_2, p, E) . Elle porte sur la pression et relie les variables thermodynamiques (p, ρ, e) par :

$$p = (\gamma - 1)\rho e, \quad (1.4)$$

où le gaz est considéré comme calorifiquement et thermiquement parfait, et de fait $e = C_v \theta$ avec C_v la chaleur massique à pression constante, et $\gamma = C_p/C_v$ est l'indice polytropique¹ avec C_p la chaleur massique à pression constante. Il vient les relations $C_p = \gamma r/(\gamma - 1)$ et $C_v = r/(\gamma - 1)$, avec $r = C_p - C_v$ la constante du gaz¹.

Ensuite, pour définir le tenseur des contraintes visqueuses $\boldsymbol{\tau}$, le premier modèle de loi de comportement de fluide newtonien est considéré. Il traduit la linéarité entre le gradient de vitesse du fluide et le tenseur des contraintes *via* les coefficients de proportionnalité de viscosité dynamique de surface μ et de volume λ . Pour un gaz polyatomique comme l'air, ces coefficients ne dépendent que de la température θ et vérifient l'hypothèse de Stokes $2\mu(\theta) + 3\lambda(\theta) = 0$. Le tenseur $\boldsymbol{\tau}$ est alors symétrique et s'exprime ainsi :

$$\boldsymbol{\tau} = \mu(\theta) \left(-\frac{2}{3} (\nabla \cdot \mathbf{U}) \mathbf{I} + \nabla \mathbf{U} + \nabla \mathbf{U}^\top \right), \quad (1.5)$$

où le coefficient de viscosité dynamique μ suit la loi de Sutherland :

$$\mu(\theta) = \mu_s \sqrt{\frac{\theta}{\theta_s} \frac{1 + C_s/\theta_s}{1 + C_s/\theta}}, \quad (1.6)$$

avec C_s la constante de Sutherland¹ et $\mu_s = \mu(\theta_s)$ la viscosité de référence à une température de référence θ_s connue².

Enfin, pour définir le flux de chaleur \mathbf{q} , le second modèle de loi de comportement de Fourier est considéré. Il traduit la linéarité entre le flux de chaleur échangé par conduction et le gradient de température *via* le coefficient de proportionnalité de conductivité thermique $k(\theta)$. Le flux \mathbf{q} s'exprime alors ainsi :

$$\mathbf{q} = -k(\theta) \nabla \theta, \quad \text{avec } k(\theta) = \mu(\theta) \frac{C_p}{Pr}, \quad (1.7)$$

où Pr est le nombre de Prandtl¹.

1. Pour les applications numériques, les caractéristiques de l'air sec dans les conditions normales de température et de pression sont retenues : $\gamma = 1,4$, $r = 287,053 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$, $C_s = 110,4 \text{ K}$ et $Pr = 0,725$.

2. Les variables de référence sont définies à partir des conditions de l'écoulement infini amont du problème, lesquelles sont fixées par le choix de la condition initiale (1.1b).

Ainsi, l'écoulement bidimensionnel et instationnaire d'un fluide visqueux est modélisé par un système non linéaire de quatre équations aux dérivées partielles d'ordre un en temps et d'ordre deux en espace, mixte hyperbolique-parabolique sous les hypothèses (1.2) à (1.7).

Remarque 1.1. *En considérant la convention de sommation sur l'indice répété $i = 1$ et $i = 2$, l'équation (1.1a) est équivalente à :*

$$\frac{\partial \mathbf{w}}{\partial t} + \frac{\partial \mathbf{f}_{c,i}(\mathbf{w})}{\partial x_i} - \frac{\partial \mathbf{f}_{v,i}(\mathbf{w}, \nabla \mathbf{w})}{\partial x_i} = 0, \quad \forall (\mathbf{x}, t) \in \Omega \times \mathcal{T},$$

où $\mathbf{f}_{c,i}$ et $\mathbf{f}_{v,i}$ appartiennent à \mathbb{R}^{d+2} et représentent les composantes vectorielles des tenseurs de flux, soit $\mathbf{F}_c = (\mathbf{f}_{c,1}, \mathbf{f}_{c,2})^\top$ et $\mathbf{F}_v = (\mathbf{f}_{v,1}, \mathbf{f}_{v,2})^\top$.

Remarque 1.2. (Adimensionnement) *Le système (1.1) est formulé dans un jeu de variables adimensionnées afin de réaliser des calculs entre des nombres du même ordre de grandeur et donc minimisant les erreurs numériques. L'adimensionnement retenu pour le système (1.1) est appelé adimensionnement « SRVT »¹ [30]. Il est détaillé en annexe A, en particulier le tableau A.2 page 192 donne l'expression de chaque variable adimensionnée. Pour la compréhension de la suite de ce travail, les points suivants sont mentionnés.*

- *L'adimensionnement choisi utilise les quatre échelles de grandeurs caractéristiques L_{ref} , ρ_{ref} , U_{ref} et θ_{ref} pour décrire les quatre unités fondamentales de toutes les variables de l'écoulement. Ces unités sont celles de longueur, de masse, de température et de temps.*
- *Dans la pratique, les grandeurs caractéristiques retenues sont définies à partir de la longueur caractéristique de l'obstacle soumis à l'écoulement L_c et, selon les applications, des conditions infinies amont ou de champ lointain de l'écoulement ρ_∞ , U_∞ et θ_∞ .*

1.2.2 Nombres sans dimension

A partir du système des équations de Navier-Stokes il est possible de déterminer des nombres sans dimension caractéristiques de l'écoulement considéré, et donc du système lui-même. Ainsi, en appliquant le théorème de Vaschy-Buckingham et en utilisant l'adimensionnement SRVT (cf. annexe A pour plus de détails sur cet adimensionnement), les quatre nombres, sans dimension et caractéristiques de l'écoulement, suivants apparaissent : l'angle d'incidence de l'obstacle soumis à l'écoulement, puis les nombres de Mach M_∞ , de Reynolds Re_∞ et de Prandtl Pr . Ils sont définis

1. Acronyme anglais : « **S**tatic-**R**ho-**V**elocity-**T**emperature ».

par :

$$\alpha = \arctan\left(\frac{u_{2\infty}}{u_{1\infty}}\right), \quad (1.8a)$$

$$M_\infty = \frac{U_\infty}{c_\infty}, \quad (1.8b)$$

$$Re_\infty = \frac{\rho_\infty U_\infty L_c}{\mu(\theta_\infty)}, \quad (1.8c)$$

$$Pr = \frac{\mu(\theta_\infty) C_p}{k(\theta_\infty)}, \quad (1.8d)$$

avec $\mathbf{U}_\infty = (u_{1\infty}, u_{2\infty})^\top$ la vitesse infinie amont, $U_\infty = \|\mathbf{U}_\infty\|$, et avec $c_\infty = \sqrt{\gamma r \theta_\infty}$ la vitesse c du son dans le fluide évaluée à la température θ_∞ .

Le nombre de Mach est caractéristique des écoulements compressibles. Son carré est proportionnel au rapport de l'énergie cinétique par l'énergie interne (énergies par unité de masse). Dans ce travail, seuls des écoulements subsoniques sont considérés.

Le nombre de Reynolds est caractéristique des écoulements visqueux. Il compare les forces d'inertie aux forces visqueuses. Dans ce travail, son ordre de grandeur n'excède pas 10^4 , ce qui correspond à des écoulements développant une couche limite laminaire et un sillage turbulent (aucun écoulement pleinement turbulent n'est donc réalisé).

Le nombre de Prandtl est caractéristique du fluide. Il compare les transferts de quantité de mouvement associés aux forces visqueuses aux transferts de chaleur par conduction. Pour les gaz, ces deux phénomènes sont du même ordre de grandeur et il en résulte un nombre de Prandtl de l'ordre de l'unité.

Enfin, un dernier nombre sans dimension, caractéristique des écoulements instationnaires, est introduit ici. Sa valeur est intrinsèque à l'écoulement et ne constitue donc pas une donnée d'entrée du problème, contrairement aux trois nombres présentés précédemment. C'est le nombre de Strouhal défini par :

$$St = \frac{L_c}{U_\infty t_{ref}}, \quad (1.9)$$

où t_{ref} représente l'échelle temporelle caractéristique du phénomène étudié. Ce nombre est, par exemple, introduit dans l'étude du détachement tourbillonnaire observé dans le sillage des obstacles soumis à un écoulement instationnaire.

1.2.3 Cas particulier du système des équations d'Euler

Dans la limite d'un nombre de Reynolds infini, le système des équations de Navier-Stokes se réduit au système des équations d'Euler. En effet, le nombre de Reynolds peut être ré-interprété comme le rapport entre le temps caractéristique des phénomènes de diffusion, $t_{ref,v} = \rho_\infty L_c^2 / \mu(\theta_\infty)$, et le temps caractéristique des phénomènes de convection, $t_{ref,c} = L_c / U_\infty$, soit $Re_\infty = t_{ref,v} / t_{ref,c}$.

Aussi, lorsque $Re_\infty \gg 1$ l'écoulement est dominé par les forces d'inertie liées aux phénomènes de convection, tandis que les phénomènes de diffusion, beaucoup plus lents, sont négligés aux échelles L_c et $t_{ref,c}$ considérées. Le système d'équations

qui en résulte est le système des équations d'Euler et l'écoulement est dit de fluide parfait. Sous forme conservative, le système des équations d'Euler s'écrit :

$$\partial_t \mathbf{w} + \nabla \cdot \mathbf{F}_c(\mathbf{w}) = 0, \quad \forall (\mathbf{x}, t) \in \Omega \times \mathcal{T}, \quad (1.10a)$$

$$\mathbf{w}(\mathbf{x}, 0) = \mathbf{w}^{(0)}(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega, \quad (1.10b)$$

auquel s'ajoutent des conditions aux limites appropriées (*cf.* § 1.3 pour plus de détails sur ces conditions). Ainsi, l'écoulement bidimensionnel et instationnaire d'un fluide parfait est modélisé par un système non linéaire de quatre équations aux dérivées partielles d'ordre un en temps et en espace, hyperbolique sous les hypothèses (1.2) et (1.4).

A défaut, lorsque le nombre de Reynolds n'est pas grand, les forces visqueuses liées aux phénomènes de diffusion ne sont plus négligeables aux échelles considérées et le modèle de fluide parfait n'est plus valable. Il faut alors revenir au système des équations de Navier-Stokes (1.1). C'est le cas des zones de couche limite et de sillage.

1.3 Conditions aux limites

L'objectif de ce paragraphe est de décrire les deux types de frontières du domaine Ω (§ 1.3.1) et de donner les conditions aux limites considérées pour déterminer complètement le problème physique associé au système des équations de Navier-Stokes (1.1) ou d'Euler (1.10) (§ 1.3.2).

1.3.1 Types de frontières

Pour les problèmes physiques considérés dans ce travail, la frontière $\partial\Omega$ du domaine de calcul est composée de deux types de frontières. D'un côté, la frontière Γ_p de l'obstacle soumis à l'écoulement, qui est une frontière matérielle et imperméable (ou paroi). D'un autre côté, la frontière $\Gamma_{in} \cup \Gamma_{out}$ externe au domaine de calcul, qui est une frontière artificielle, perméable et fixée par l'utilisateur. Dans les notations précédentes, les indices *p*, *in* et *out* font référence à une paroi respectivement, à l'entrée du domaine de calcul et à sa sortie. En conclusion, il vient $\partial\Omega = \Gamma_p \cup (\Gamma_{in} \cup \Gamma_{out})$.

Les figures 1.1 illustrent les deux cas possibles de décomposition de la frontière $\partial\Omega$. Ces cas sont en fait une généralisation des configurations rencontrées dans les applications numériques de la partie II : l'écoulement interne confiné entre des parois (figure 1.1a) et l'écoulement externe autour d'un obstacle (figure 1.1b).

Pour la suite, l'état frontière est défini par $\mathbf{w}_b = (\rho_b, \rho_b \mathbf{U}_b, \rho_b E_b)^\top$ en tout point de $\partial\Omega$. En fonction des conditions aux limites appliquées, cet état dépend de l'état intérieur $\mathbf{w}^+ = (\rho^+, \rho^+ \mathbf{U}^+, \rho^+ E^+)^\top$ et de l'état physique infini amont ou bien de champ lointain $\mathbf{w}_\infty = (\rho_\infty, \rho_\infty \mathbf{U}_\infty, \rho_\infty E_\infty)^\top$. De plus, \mathbf{n} est la normale unitaire sortante du domaine Ω par convention.

1.3.2 Conditions aux limites considérées

D'un point de vue général, le nombre de conditions aux limites à imposer au système des équations de Navier-Stokes (1.1) repose sur un juste accord entre la nature

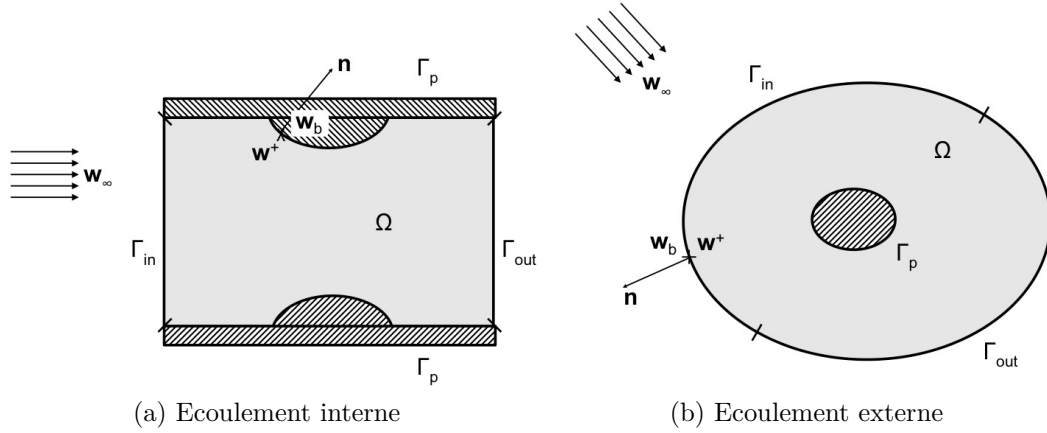


Fig. 1.1 – Exemples de configurations des frontières $\Gamma_p \cup (\Gamma_{in} \cup \Gamma_{out})$ du domaine Ω . Les notations \mathbf{w}^+ , \mathbf{w}_b et \mathbf{w}_∞ désignent les états respectivement : intérieur, frontière et infini amont (cas d'un écoulement interne) ou de champ lointain (cas d'un écoulement externe). Par convention, \mathbf{n} est la normale unitaire sortante du domaine.

mathématique de ses équations (mixte hyperbolique-parabolique) et les propriétés physiques de l'écoulement (cinématique, température).

Dans le cas d'équations hyperboliques (système des équations d'Euler), le traitement des conditions aux limites sur la frontière du domaine de calcul est fondé sur la méthode des caractéristiques. L'extension de cette méthode au système mixte complet (système des équations de Navier-Stokes) est naturelle pour les zones éloignées de la paroi pouvant être assimilées à un écoulement de fluide parfait, c'est-à-dire sur Γ_{in} et Γ_{out} . En revanche, ce n'est plus le cas en zone de proche paroi où la couche limite se développe, c'est-à-dire sur Γ_p . Il faut alors tenir compte des phénomènes physiques mis en jeu.

1.3.2.1 Conditions sur les frontières perméables $\Gamma_{in} \cup \Gamma_{out}$

Ces conditions sont celles de non-réflexion. Leur détermination repose sur l'application de la méthode des caractéristiques aux systèmes des équations d'Euler et de Navier-Stokes. On rappelle que pour ce dernier système, les frontières $\Gamma_{in} \cup \Gamma_{out}$ doivent se situer « suffisamment loin » de la paroi, ce qui en général fait suite à une étude approfondie sur la valeur de certains coefficients (par exemple : les coefficients aérodynamiques globaux).

Afin de décrire comment les conditions de non-réflexion sont imposées, le paragraphe suivant fait un bref rappel sur la méthode des caractéristiques et plus particulièrement sur les relations caractéristiques qu'elle permet de définir pour le système (1.10). Pour une présentation complète de la méthode, le lecteur est renvoyé à l'ouvrage [56, chapitre IV] pour le cadre mathématique théorique, ou [65, chapitre 11] pour une approche plus physique, et [118] pour une application concrète aux systèmes des équations d'Euler et de Navier-Stokes.

Les quatre relations caractéristiques (ou relations de compatibilité) discrétisées, notées $(RCi)_{1 \leq i \leq 4}$, sont déduites de la matrice jacobienne de dérivation des flux convectifs suivant la normale \mathbf{n} (soit $n_1 \partial \mathbf{f}_{c,1} / \partial \mathbf{w} + n_2 \partial \mathbf{f}_{c,2} / \partial \mathbf{w}$). Elles sont associées

aux quatre valeurs propres λ_i et s'expriment comme suit :

$$\lambda_1 = \mathbf{U} \cdot \mathbf{n} - c ; \quad p_b - (\rho c)^+ \mathbf{U}_b \cdot \mathbf{n} = p^+ - (\rho c)^+ \mathbf{U}^+ \cdot \mathbf{n} \quad (RC1),$$

$$\lambda_2 = \mathbf{U} \cdot \mathbf{n} ; \quad \mathbf{U}_b \cdot \mathbf{t} = \mathbf{U}^+ \cdot \mathbf{t} \quad (RC2),$$

$$\lambda_3 = \mathbf{U} \cdot \mathbf{n} ; \quad \left(\frac{p}{\rho^\gamma}\right)_b = \left(\frac{p}{\rho^\gamma}\right)^+ \quad (RC3),$$

$$\lambda_4 = \mathbf{U} \cdot \mathbf{n} + c ; \quad p_b + (\rho c)^+ \mathbf{U}_b \cdot \mathbf{n} = p^+ + (\rho c)^+ \mathbf{U}^+ \cdot \mathbf{n} \quad (RC4),$$

où \mathbf{t} est le vecteur tangent à la frontière $\partial\Omega$. Les valeurs propres $(\lambda_i)_{1 \leq i \leq 4}$ donnent la vitesse de propagation de l'information au sein du domaine. En particulier, si $\lambda_i > 0$ (respectivement $\lambda_i < 0$), l'information se propage de l'intérieur vers l'extérieur (respectivement de l'extérieur vers l'intérieur) du domaine Ω et la caractéristique est dite sortante (respectivement entrante).

L'imposition des conditions aux limites de non-réflexion se fait alors comme suit. Lorsque $\lambda_i > 0$, la relation caractéristique sortante correspondante est imposée. Tandis que lorsque $\lambda_i < 0$, la caractéristique entrante correspondante est remplacée par une condition au limite dite physique. Concrètement, cela signifie que les variables intérieures sont remplacées par les variables infinies amont dans les relations caractéristiques, exceptée la quantité $(\rho c)^+$ qui reste calculée avec l'état intérieur par cohérence avec l'évaluation des valeurs propres au cours des itérations de calcul. L'utilisation des conditions infinies amont est en accord avec le fait que la relation caractéristique est entrante et donc que l'information vient de l'extérieur du domaine.

Pour Γ_{in} , frontière d'entrée subsonique, il vient $\lambda_{1,2,3} < 0$ et $\lambda_4 > 0$. L'état frontière \mathbf{w}_b est donné par la résolution du problème de quatre équations à quatre inconnues : $(RC1)$, $(RC2)$, $(RC3)$ modifiées et $(RC4)$ conservée.

Pour Γ_{out} , frontière de sortie subsonique, il vient (en général) $\lambda_1 < 0$ et $\lambda_{2,3,4} > 0$. L'état frontière \mathbf{w}_b est donné par la résolution du problème de quatre équations à quatre inconnues : $(RC1)$ modifiée et $(RC2)$, $(RC3)$, $(RC4)$ conservées.

1.3.2.2 Conditions sur les frontières imperméables Γ_p

Ces conditions dépendent de la nature du fluide et prennent donc en compte les effets visqueux portant sur la vitesse du fluide à la paroi (condition cinématique) et les échanges thermiques (condition thermique). On précise ici, que pour les applications numériques de la partie II, la paroi est considérée immobile et adiabatique.

Pour le système des équations d'Euler, seule la condition cinématique est nécessaire. Il s'agit de la condition de glissement (condition de Dirichlet) :

$$\mathbf{U}_b \cdot \mathbf{n} = 0, \quad \forall(\mathbf{x}, t) \in \Gamma_p \times \mathcal{T}.$$

Pour le système des équations de Navier-Stokes, la condition cinématique est celle de non-glissement ou d'adhérence à la paroi (condition de Dirichlet) en accord avec les phénomènes physiques qui apparaissent dans la couche limite :

$$\mathbf{U}_b = 0, \quad \forall(\mathbf{x}, t) \in \Gamma_p \times \mathcal{T}.$$

A cette condition s'ajoute la condition physique sur les échanges thermiques (condition de Neumann) :

$$\nabla\theta|_b \cdot \mathbf{n} = 0, \quad \forall(\mathbf{x}, t) \in \Gamma_p \times \mathcal{T}.$$

1.3.2.3 Cas particulier de la condition de symétrie

Cette condition permet de réduire la taille d'un domaine de calcul et donc de réduire le coût en temps de calcul et en capacité de mémoire. La frontière où est appliquée cette condition agit comme un miroir et repose sur le postulat suivant : des deux côtés de la frontière le même phénomène physique se déroule, imposant donc une valeur identique aux variables situées à une même distance de part et d'autre.

Dans la pratique, la condition de symétrie s'écrit, pour une quantité scalaire z ou vectorielle \mathbf{z} quelconque :

$$\begin{aligned}z_b &= z^+, \\ \mathbf{z}_b \cdot \mathbf{n} &= 0, \\ \mathbf{z}_b \cdot \mathbf{t} &= \mathbf{z}^+ \cdot \mathbf{t}.\end{aligned}$$

1.4 Conclusion partielle

Les points suivants constituent le bilan de ce chapitre.

- Les modèles mathématiques du problème physique de l'écoulement bidimensionnel et instationnaire d'un fluide compressible ont été rappelés. A savoir :
 - le système hyperbolique des équations d'Euler pour un fluide parfait ;
 - le système hyperbolique-parabolique des équations de Navier-Stokes pour un fluide visqueux.
- Ces systèmes sont complétés en tenant compte des propriétés du fluide considéré, lesquelles se traduisent par :
 - la loi de comportement pour les fluides newtonien modélisant les contraintes visqueuses ;
 - la loi de comportement modélisant les transferts thermiques.
- Les systèmes sont fermés grâce à une loi d'état sur la pression dans l'approximation des gaz parfaits.
- Trois conditions aux limites viennent compléter ces deux systèmes d'équations, les conditions de :
 - paroi adiabatique (glissement ou adhérence) ;
 - non-réflexion (entrée ou sortie subsonique) ;
 - symétrie.
- Les équations sont écrites sous une forme adimensionnelle en utilisant l'adimensionnement SRVT. Quatre nombres sans dimension caractéristiques de ces deux systèmes d'équations apparaissent alors (voir également annexe A) :
 - l'angle d'incidence de l'obstacle soumis à l'écoulement α ;
 - le nombre de Mach M_∞ ;
 - le nombre de Reynolds Re_∞ ;
 - le nombre de Prandtl Pr .

A présent, il s'agit d'établir le schéma numérique discret de ces systèmes d'équations. Pour ce faire, le chapitre 2 est consacré à la discrétisation en espace et le chapitre 3 à la discrétisation en temps.

Chapitre 2

Discrétisation spatiale Galerkin Discontinue (GD)

Ce chapitre décrit la discrétisation spatiale du système des équations de Navier-Stokes établi au chapitre 1. Une méthode de Galerkin Discontinue (GD) est utilisée : écriture de la formulation faible, discrétisation du domaine de calcul, définition des espaces fonctionnels, choix des flux numériques. Il s'agit du flux local de Lax-Friedrichs (LLF) [156] et de celui du second schéma de Bassi et Rebay (BR2) [13, 7] pour les discrétisations respectivement, des termes de convection et de ceux de diffusion. La dernière partie du chapitre concerne la mise en œuvre de la méthode GD, en particulier pour prendre en compte les frontières matérielles courbes du domaine.

2.1 Avant-propos

Les méthodes de type GD font le lien entre la méthode des éléments finis et la méthode des volumes finis. En effet, le problème de départ (1.1) (problème fort) est mis sous forme variationnelle (problème faible) comme pour la méthode des éléments finis. Cependant aucune continuité n'est imposée à la solution faible ni aux fonctions tests à l'interface entre les éléments du domaine de calcul discrétisé. En raison de cette discontinuité, les flux physiques sont approchés par des flux numériques aux interfaces comme pour la méthode des volumes finis. En revanche, la solution faible est représentée sur chaque élément par des fonctions de base polynomiales et non par des fonctions constantes. Pour la mise en place de la méthode GD considérée dans ce travail, les principales références sont [28, 38].

La première partie de ce chapitre, paragraphe 2.2, rappelle la formulation faible du système des équations de Navier-Stokes (1.1). La seconde partie, paragraphe 2.3, s'intéresse à la discrétisation (ou approximation interne) de cette formulation par une méthode GD. L'objectif étant d'obtenir un schéma numérique compact respectant les propriétés de consistance et de conservativité, des flux numériques adéquats sont définis aux interfaces du domaine approché. La dernière partie, paragraphe 2.4, est réservée à certains aspects de mise en œuvre de la méthode GD comme la manière de prendre en compte les frontières matérielles courbes pour minimiser l'erreur géométrique pouvant devenir prépondérante devant l'erreur d'approximation du schéma numérique dans une simulation d'ordre élevé [3, 10].

2.2 Formulation variationnelle

Ce paragraphe établit la formulation variationnelle (ou faible) du problème (1.1), dit fort. On considère les définitions suivantes pour les espaces fonctionnels $L^2(\Omega)$ et $H^1(\Omega)$:

Définition 2.1. (Espace $L^2(\Omega)$) Soit $L^2(\Omega)$ l'espace des fonctions de carré intégrable

$$L^2(\Omega) = \left\{ f : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R} : \int_{\Omega} f^2(\mathbf{x}) dV < \infty \right\},$$

muni du produit scalaire et de la norme définis respectivement par :

$$(f, g)_{L^2(\Omega)} = \int_{\Omega} f(\mathbf{x})g(\mathbf{x})dV, \quad \|f\|_{L^2(\Omega)} = \sqrt{(f, f)_{L^2(\Omega)}}.$$

Définition 2.2. (Espace $H^1(\Omega)$) Soit $H^1(\Omega)$ l'espace des fonctions de $L^2(\Omega)$ dont la dérivée première appartient à $L^2(\Omega)$

$$H^1(\Omega) = \left\{ f \in L^2(\Omega) : \nabla f \in [L^2(\Omega)]^d \right\}.$$

Une solution faible $\mathbf{w} : \Omega \times \mathcal{T} \rightarrow \Omega_{adm}$ du problème (1.1) est recherchée dans l'espace fonctionnel $\mathcal{V} \equiv \mathcal{C}^0(\mathcal{T}; [H^1(\Omega)]^{d+2})$. La formulation variationnelle du problème s'obtient en multipliant chacune des équations du système (1.1) par une fonction test v appartenant à $H^1(\Omega)$, puis en intégrant sur Ω . Après intégration par parties, il vient : Trouver \mathbf{w} appartenant à \mathcal{V} , telle que pour tout v appartenant à $H^1(\Omega)$ et à chaque instant t dans \mathcal{T} , \mathbf{w} soit solution de

$$\int_{\Omega} v \partial_t \mathbf{w} dV - \int_{\Omega} (\mathbf{F}_c - \mathbf{F}_v) \nabla v dV + \oint_{\partial\Omega} v (\mathbf{F}_c - \mathbf{F}_v) \mathbf{n} dS = 0. \quad (2.1)$$

Remarque 2.1. L'intégration par parties est réalisée sur tout le domaine Ω et permet de réduire la régularité requise sur \mathbf{w} et d'imposer faiblement les conditions aux limites par le biais du troisième terme de l'équation (2.1).

Remarque 2.2. Dans la littérature [58], la fonction test est parfois prise sous forme vectorielle donnant (2.1) sous forme scalaire et facilitant l'étude de ses propriétés. Cette écriture est strictement équivalente à celle considérée dans ce travail qui présente le système d'équations tel qu'il est mis en œuvre dans le code de calcul utilisé pour les applications numériques de la partie II.

2.3 Méthode de Galerkin Discontinue

Ce paragraphe exprime l'approximation interne (ou forme discrète) de la formulation variationnelle (2.1) précédente en utilisant une méthode GD compacte. Comme mentionné dans l'introduction de ce chapitre, l'idée est de considérer une solution faible \mathbf{w} polynomiale par morceaux, c'est-à-dire polynomiale sur chaque élément du domaine de calcul préalablement discrétisé (ou du maillage) mais non continue aux interfaces entre les éléments.

Pour cela, un partitionnement Ω_h du domaine de calcul Ω composé d'éléments κ est introduit (§ 2.3.1). Ensuite, l'espace fonctionnel particulier utilisé en raison de l'absence de continuité aux interfaces entre les éléments est défini, ainsi que les caractéristiques de la base qui l'engendre (§ 2.3.2). Ces deux premiers points permettent d'écrire la formulation variationnelle (2.1) sur chacun des éléments de la discrétisation, puis sur le domaine tout entier : c'est l'approximation interne (§ 2.3.3). Cette dernière fait intervenir des flux numériques pour respecter les propriétés de consistance et de conservativité du schéma numérique obtenu. Finalement, le choix des flux numériques retenus, pour obtenir un schéma numérique à stencil compact, est expliqué (§ 2.3.4). Il s'agit du flux local de Lax-Friedrichs (LLF) [156] pour la discrétisation des termes de convection, et de celui du second schéma de Bassi et Rebay (BR2) [13, 7] pour la discrétisation des termes de diffusion. Ces flux sont fonction de la solution numérique supposée discontinue aux interfaces entre les éléments.

2.3.1 Discrétisation spatiale du domaine

Le domaine Ω est partitionné en un ensemble fini Ω_h de N éléments κ , simpliciaux (par exemple, des segments en une dimension d'espace et des triangles en deux dimensions d'espace), droits ou courbes, et disjoints, soit :

$$\Omega_h = \bigcup_{\kappa \in \Omega_h} \kappa.$$

De plus, $\partial\Omega_h$ désigne la partition de $\partial\Omega$. Ainsi, les maillages considérés sont non structurés et tels que :

- l'intérieur d'un élément κ est non vide,
- l'intersection entre deux éléments ne peut être qu'un point en une dimension d'espace et un segment en deux dimensions d'espace,
- cette intersection est soit commune avec un autre élément, soit une partie de la frontière $\partial\Omega_h$.

Le paramètre h représente la taille caractéristique du maillage et détermine sa finesse. Le paramètre H est également introduit. Ces paramètres sont définis par :

$$h = \min \{h_\kappa, \kappa \in \Omega_h\}, \quad (2.2a)$$

$$H = \max \{h_\kappa, \kappa \in \Omega_h\}, \quad (2.2b)$$

avec h_κ le diamètre de la plus grande boule contenue dans l'élément κ . Une mesure du caractère anisotrope du maillage est alors donnée par le rapport d'aspect global du maillage, noté RA , défini par

$$RA = \frac{H}{h}. \quad (2.3)$$

Enfin, l'ensemble des faces de Ω_h est noté \mathcal{E}_h . Il regroupe l'ensemble des faces intérieures \mathcal{E}_i et l'ensemble des faces frontières \mathcal{E}_b (qui correspond à $\partial\Omega_h$), soit :

$$\mathcal{E}_h = \mathcal{E}_i \cup \mathcal{E}_b \quad \text{avec} \quad \begin{cases} \mathcal{E}_i = \{e \in \mathcal{E}_h : e \cap \partial\Omega_h = \emptyset\}, \\ \mathcal{E}_b = \{e \in \mathcal{E}_h : e \in \partial\Omega_h\}. \end{cases} \quad (2.4)$$

Par ailleurs, soit $n_f = d+1$ le nombre d'éléments voisins d'un élément, qui correspond donc au nombre de faces de cet élément. La figure 2.1 illustre ces notations.

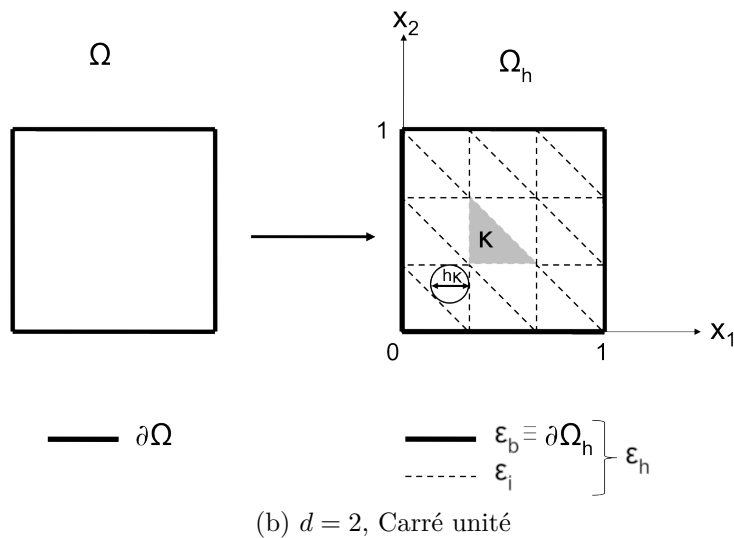
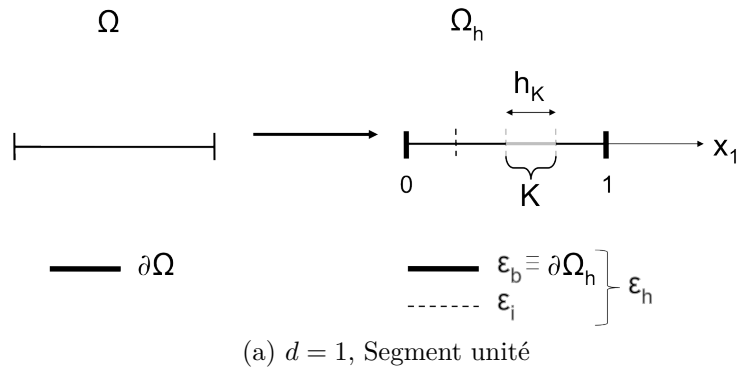


Fig. 2.1 – Illustration de la discrétisation de Ω et de ses notations pour $d = 1$ (a) et pour $d = 2$ (b). Cas de maillages à éléments simpliciaux droits.

| | | | | | | | |
|-------------------|---|---|---|----|----|----|----|
| p | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| N_p ($d = 1$) | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| N_p ($d = 2$) | 1 | 3 | 6 | 10 | 15 | 21 | 28 |

Tab. 2.1 – Dimension de l'espace $\mathcal{P}^p(\kappa)$ en fonction de p pour les deux cas $d = 1$ et $d = 2$.

2.3.2 Espaces fonctionnels discrets

Dans l'approche GD, les fonctions tests engendrant l'espace où est recherchée la solution numérique sont supposées polynomiales (ou exponentielles ou trigonométriques) mais sans condition de continuité aux interfaces. De ce fait, la solution numérique est naturellement discontinue aux interfaces entre les éléments. Ce paragraphe donne la forme de l'espace d'approximation (§ 2.3.2.1) ainsi que le choix des fonctions de base qui l'engendrent (§ 2.3.2.2).

2.3.2.1 Espace pour la méthode de Galerkin discontinue

On considère la définition suivante pour l'espace fonctionnel $H^1(\Omega_h)$:

Définition 2.3. (*Espace $H^1(\Omega_h)$*) Soit $H^1(\Omega_h)$ l'espace des fonctions appartenant à $L^2(\Omega_h)$ dont la restriction à l'élément κ appartient à $H^1(\kappa)$:

$$H^1(\Omega_h) = \left\{ f \in L^2(\Omega_h) : f|_{\kappa} \in H^1(\kappa), \forall \kappa \in \Omega_h \right\}.$$

La formulation (2.1) étant considérée pour chaque élément κ , les composantes du vecteur solution numérique $\mathbf{w}_{h|\kappa}$ sont cherchées dans l'espace \mathcal{V}_h^p inclus dans $H^1(\Omega_h)$, parfois appelé « espace GD », qui est l'espace des fonctions polynomiales par morceaux sur Ω_h , défini par

$$\mathcal{V}_h^p = \left\{ f \in L^2(\Omega_h) : f|_{\kappa} \in \mathcal{P}^p(\kappa), \forall \kappa \in \Omega_h \right\}, \quad (2.5)$$

avec $\mathcal{P}^p(\kappa)$ l'espace engendré par des polynômes définis sur κ et de degré total au plus égal à p , avec $p \geq 0$:

$$\mathcal{P}^p(\kappa) = \left\{ f \in L^2(\kappa) : f(x_1, x_2) = \sum_{0 \leq k+l \leq p} \alpha_{k+l} x_1^k x_2^l, \alpha_{k+l} \in \mathbb{R} \right\}, \quad (2.6)$$

où par simplicité, on s'est placé en deux dimensions d'espace. La restriction à la dimension un est obtenue en fixant l à 0. Dans la pratique, ces polynômes sont évalués en des points de quadrature, eux-mêmes contenus dans un domaine de référence. Il convient alors d'introduire une transformation de ce domaine vers l'élément physique κ . Cette transformation est exposée au paragraphe 2.4 pour plus de clarté. Enfin, la dimension des deux espaces, décrits en (2.6) et (2.5), est respectivement :

$$\begin{aligned} \dim(\mathcal{P}^p) &= \prod_{i=1}^d \frac{(p+i)}{i} = N_p, \\ \dim(\mathcal{V}_h^p) &= \text{card}(\Omega_h) \times \dim(\mathcal{P}^p) = N \times N_p. \end{aligned}$$

A titre d'exemple, le tableau 2.1 donne les valeurs de N_p pour $d = 1$ et $d = 2$. En outre, on indique que la notation $\text{GD}(p)$, en lien avec la définition (2.5) de l'espace GD, sera parfois utilisée dans la suite de ce mémoire.

Remarque 2.3. Dans la pratique, le choix de l'espace d'approximation où est recherchée la solution n'est pas unique pour les méthodes GD. En particulier, Yuan et

Shu [182] ont montré qu'il est possible d'obtenir des conditions pour lesquelles des espaces de fonctions non polynomiales (par exemple : exponentielles ou trigonométriques) donnent la même estimation d'erreur et la même stabilité que des espaces de fonctions polynomiales.

Remarque 2.4. Comme aucune continuité n'est imposée aux interfaces des éléments κ , la méthode GD constitue une méthode des éléments finis non conforme : $[\mathcal{V}_h^p]^{d+2}$ n'est pas inclus dans \mathcal{V} .

Remarque 2.5. En prenant \mathcal{V}_h^p égal à \mathcal{V}_h^0 , c'est-à-dire en prenant les fonctions de base et la solution numérique constantes par élément, et avec le choix des flux numériques détaillé au paragraphe 2.3.4, la méthode GD correspond à la méthode des volumes finis au premier ordre en espace.

2.3.2.2 Choix de la base de polynômes

Soit $\mathcal{B}_\kappa^p = (\phi^k)_{1 \leq k \leq N_p}$ une base qui engendre l'espace $\mathcal{P}^p(\kappa)$. Afin d'établir l'approximation interne, il est commode de choisir les fonctions tests v_h comme étant les fonctions de cette base, soit :

$$v_h \in \left\{ \phi^k \in \mathcal{P}^p(\kappa) : 1 \leq k \leq N_p \right\}, \quad (2.7)$$

conduisant à N_p équations par équation du système de départ (1.1). Dans le cadre de ce travail, la base \mathcal{B}_κ^p retenue présente les trois caractéristiques suivantes : elle est modale, orthonormale et hiérarchique (cf. remarque 2.6).

Premièrement, la base est modale, c'est-à-dire que les inconnues du problème sont les composantes de \mathbf{w}_h dans la base \mathcal{B}_κ^p . Ainsi, dans chaque élément κ , la solution numérique est reconstruite grâce à une combinaison linéaire des fonctions de base, soit :

$$\mathbf{w}_h(\mathbf{x}, t) = \sum_{l=1}^{N_p} \mathbf{W}_\kappa^l(t) \phi^l(\mathbf{x}), \quad \forall \mathbf{x} \in \kappa, \quad \kappa \in \Omega_h, \quad \forall t \in \mathcal{T}, \quad (2.8)$$

où

$$\mathbf{W}_\kappa^l = \left(W_\kappa^{1,l}, \dots, W_\kappa^{N_{eq},l} \right)^\top.$$

L'écriture de \mathbf{w}_h sous la forme (2.8) est appelée décomposition modale [65, chapitre 9]. Par commodité, le vecteur contenant les degrés de liberté sur l'élément κ est introduit :

$$\mathbf{W}_\kappa = \left(\left(W_\kappa^{1,l} \right)_{1 \leq l \leq N_p}, \dots, \left(W_\kappa^{i,l} \right)_{1 \leq l \leq N_p}, \dots, \left(W_\kappa^{N_{eq},l} \right)_{1 \leq l \leq N_p} \right)^\top. \quad (2.9)$$

Il constitue les $N_{eq}N_p$ inconnues du problème sur κ (on rappelle que $N_{eq} = d + 2$). De même le vecteur complet des degrés de liberté sur Ω_h est introduit :

$$\mathbf{W} = \left(\mathbf{W}_\kappa \right)_{\kappa \in \Omega_h}^\top = \left(W_\kappa^{i,l} \right)_{\kappa \in \Omega_h, 1 \leq i \leq N_{eq}, 1 \leq l \leq N_p}^\top. \quad (2.10)$$

Il constitue les $NN_{eq}N_p$ inconnues du problème sur Ω_h et est rangé par degré de liberté l pour chaque composante i du vecteur des variables conservatives \mathbf{w}_h , et ce pour chaque élément κ de la discrétisation.

Deuxièmement, la base est orthonormale, c'est-à-dire que les éléments de la base sont orthonormaux deux à deux vis-à-vis du produit intérieur (ou produit scalaire) défini sur κ , soit :

$$\int_{\kappa} \phi^k(\mathbf{x})\phi^l(\mathbf{x})dV = |\kappa|\delta_{kl}, \quad \forall \kappa \in \Omega_h, \quad 1 \leq k, l \leq N_p, \quad (2.11)$$

où $|\kappa|$ représente la mesure de κ (par exemple : sa longueur en une dimension d'espace, son aire en deux dimensions d'espace) et δ_{kl} est le symbole de Kronecker.

Troisièmement, la base est hiérarchique, c'est-à-dire, en considérant la définition 2.4, que sur chaque élément κ , l'extension d'une base de polynômes de degré p à une autre de degré $p + 1$ se fait naturellement en ajoutant $(N_{p+1} - N_p)$ polynômes supplémentaires.

Définition 2.4. (Base hiérarchique) Soit k un entier strictement positif. On dit que la famille de polynômes (ϕ^1, \dots, ϕ^k) forme une base hiérarchique de $\mathcal{P}^k(\kappa)$ si pour tout $1 \leq l \leq k$, la sous-famille (ϕ^1, \dots, ϕ^l) forme une base de $\mathcal{P}^l(\kappa)$ [chapitre 4][44].

Enfin, le choix de la base dépend du type d'élément physique κ et pour chaque type il existe de nombreuses bases. Cependant, on note qu'à p fixé, la précision de la méthode GD est indépendante du choix des fonctions de base qui engendrent le même espace d'approximation. Dans ce travail :

- deux bases ont été retenues pour les applications numériques sur les maillages d'éléments simpliciaux droits : celle des polynômes de Legendre [2, chapitres 8 et 22] en une dimension d'espace et celle de Dubiner [42], [153, chapitre 3] en deux dimensions d'espace ;
- une base numérique [134], [38, annexe A] a été retenue pour les applications numériques sur les maillages d'éléments simpliciaux courbes.

Ces trois bases sont exposées au paragraphe 2.4 pour plus de clarté.

Remarque 2.6. En anticipant sur les développements à venir, les trois caractéristiques retenues pour la base \mathcal{B}_{κ}^p sont motivées par la discrétisation temporelle explicite-implicite simplifiée du paragraphe 3.5.3 page 79.

D'un côté pour la partie explicite en temps, l'orthonormalité de la base conduit à une matrice de masse bien conditionnée (même lorsque p augmente) et diagonale, de fait, facilement inversible à moindre coût pour chaque pas de temps.

D'un autre côté pour la partie implicite en temps, la base modale et hiérarchique permet la simplification de mise en œuvre. En particulier, son caractère hiérarchique justifie qu'il soit utilisé une sous-famille pour engendrer un espace d'approximation de dimension inférieure. Cet espace joue un rôle important pour les performances de la simplification envisagée.

2.3.3 Approximation interne

La discrétisation du domaine étant réalisée, puis la forme des fonctions de base et celle de l'inconnue discrète étant données respectivement en (2.7) et (2.8), la discrétisation de la formulation variationnelle (2.1) sur chaque élément κ s'écrit :

Trouver \mathbf{w}_h appartenant à $[\mathcal{V}_h^p]^{d+2}$, telle que pour tout κ appartenant à Ω_h et ϕ^k élément de \mathcal{B}_κ^p et à chaque instant t dans \mathcal{T} , \mathbf{w}_h soit solution de

$$\begin{aligned} \int_{\kappa} \phi^k \partial_t \mathbf{w}_h dV - \int_{\kappa} \left(\mathbf{F}_c(\mathbf{w}_h) - \mathbf{F}_v(\mathbf{w}_h, \boldsymbol{\Upsilon}_h) \right) \nabla \phi^k dV \\ + \oint_{\partial\kappa} \phi^k \left(\hat{\mathbf{F}}_c - \hat{\mathbf{F}}_v \right) dS = 0, \end{aligned} \quad (2.12)$$

où $\boldsymbol{\Upsilon}_h$ est une variable auxiliaire représentant au niveau discret $\nabla \mathbf{w}_h$, et où $\hat{\mathbf{F}}_c$ et $\hat{\mathbf{F}}_v$ sont des vecteurs de flux numériques, qui seront tous les trois définis au paragraphe 2.3.4. Cependant, on souligne dès à présent que sur les intégrales aux frontières d'un élément, les flux numériques sont des approximations des composantes normales des flux physiques $\mathbf{F}_c \mathbf{n}$ et $\mathbf{F}_v \mathbf{n}$, qui permettent d'assurer les propriétés de consistance et de conservativité du schéma. En effet, les seules hypothèses de régularité formulées sur les fonctions appartenant à \mathcal{V}_h^p concernent l'intérieur de chaque élément κ . De ce fait, \mathbf{w}_h n'est pas définie de manière continue aux interfaces entre les éléments. Comme pour la méthode des volumes finis, les flux numériques pallient ce problème.

Remarque 2.7. *L'intégration par parties est réalisée ici sur chaque élément κ et permet de coupler les degrés de liberté de la solution numérique entre les éléments via les flux numériques.*

Avant de présenter les trois quantités $\boldsymbol{\Upsilon}_h$, $\hat{\mathbf{F}}_c$ et $\hat{\mathbf{F}}_v$, l'approximation interne (2.12) est reconsidérée en tenant compte de la sommation sur tous les éléments κ appartenant à Ω_h . Pour ce faire, les fonctions $(\phi^k)_{1 \leq k \leq N_p}$ de la base \mathcal{B}_κ^p sont étendues à tout le domaine Ω_h en posant $\phi^k = 0$ sur $\Omega_h \setminus \kappa$. Soit $\mathcal{B}_h^p = (\mathcal{B}_\kappa^p)_{\kappa \in \Omega_h}$ une base de l'espace \mathcal{V}_h^p , où par construction chaque fonction de la base \mathcal{B}_h^p est bien à support dans un seul élément κ appartenant à Ω_h . Il vient donc pour la relation (2.12) : Trouver \mathbf{w}_h appartenant à $[\mathcal{V}_h^p]^{d+2}$, telle que pour tout ϕ^k élément de \mathcal{B}_h^p et à chaque instant t dans \mathcal{T} , \mathbf{w}_h soit solution de

$$\begin{aligned} \sum_{\kappa \in \Omega_h} \int_{\kappa} \phi^k \partial_t \mathbf{w}_h dV - \sum_{\kappa \in \Omega_h} \int_{\kappa} \left(\mathbf{F}_c(\mathbf{w}_h) - \mathbf{F}_v(\mathbf{w}_h, \boldsymbol{\Upsilon}_h) \right) \nabla \phi^k dV \\ + \sum_{\kappa \in \Omega_h} \oint_{\partial\kappa} \phi^k \left(\hat{\mathbf{F}}_c - \hat{\mathbf{F}}_v \right) dS = 0, \end{aligned} \quad (2.13)$$

qui suppose un traitement spécifique aux frontières du domaine qui sera détaillé au paragraphe 2.3.4 suivant.

Enfin, pour faciliter la lecture de la suite, cette dernière équation est reconsidérée sous la forme générique suivante : Trouver \mathbf{w}_h appartenant à $[\mathcal{V}_h^p]^{d+2}$, telle que pour tout ϕ^k élément de \mathcal{B}_h^p et à chaque instant t dans \mathcal{T} , \mathbf{w}_h soit solution de

$$\int_{\Omega_h} \phi^k \partial_t \mathbf{w}_h dV + \mathcal{L}_c(\mathbf{w}_h, \phi^k) + \mathcal{L}_v(\mathbf{w}_h, \phi^k) = 0, \quad (2.14)$$

où la notation suivante a été adoptée pour le premier terme :

$$\int_{\Omega_h} f dV = \sum_{\kappa \in \Omega_h} \int_{\kappa} f dV, \quad \forall f \in L^2(\Omega_h), \quad (2.15)$$

et où \mathcal{L}_c et \mathcal{L}_v sont les opérateurs de discrétisation des termes respectivement, de convection et de diffusion (encore appelés résidus discrets) qu'il reste à définir en fonction de la variable auxiliaire Υ_h , des flux physiques et des flux numériques retenus. Enfin la condition initiale est discrétisée selon :

$$\int_{\Omega_h} \phi^k \mathbf{w}_h(\mathbf{x}, 0) dV = \int_{\Omega_h} \phi^k \mathbf{w}^{(0)}(\mathbf{x}) dV, \quad \forall \phi^k \in \mathcal{B}_h^p. \quad (2.16)$$

Les équations (2.14) et (2.16) constituent alors le schéma semi-discret en espace.

Remarque 2.8. *Pour assurer une bonne précision de la solution numérique, toutes les intégrales présentes dans le schéma semi-discret sont évaluées avec une méthode d'intégration numérique de type quadrature Gauss. Cette dernière est exposée au paragraphe 2.4 pour plus de clarté.*

2.3.4 Choix des flux numériques

Le choix des flux numériques convectifs et visqueux permettant de définir les opérateurs \mathcal{L}_c et \mathcal{L}_v constitue la dernière étape pour compléter la formulation GD.

A priori, un choix de flux numériques issus du développement des méthodes de volumes finis et respectant les propriétés de consistance et de conservativité du schéma peut convenir pour les méthodes GD, à condition que le résidu discret dans un élément de discrétisation présente un stencil compact. En effet, cette propriété de compacité est un avantage important de la méthode GD. Toutefois, on indique dès à présent que si l'utilisation directe d'un flux numérique pour les termes de convection ne pose pas de difficulté particulière, ce n'est en revanche plus le cas pour les termes de diffusion qui nécessitent une attention particulière.

Avant de présenter séparément les flux numériques retenus dans ce travail pour les termes de convection (§ 2.3.4.2) et pour les termes de diffusion (§ 2.3.4.3), les notations utilisées à l'interface entre deux éléments du domaine de calcul discrétisé sont d'abord introduites (§ 2.3.4.1).

2.3.4.1 Traces et opérateurs

Soit une interface e appartenant à \mathcal{E}_i et κ un élément de Ω_h ayant e pour une de ses interfaces. Soit \mathbf{n} la normale unitaire à e et extérieure à κ . L'élément κ est alors noté κ^+ (élément dit courant), et l'unique autre élément partageant e est noté κ^- (élément dit voisin). De plus, z^+ (respectivement z^-) est définie comme la trace de la fonction scalaire z définie pour tout point \mathbf{x}_e de l'interface e et appartenant à κ^+ (respectivement κ^-) par :

$$\forall \mathbf{x}_e \in e, \quad z^\pm(\mathbf{x}_e) = \lim_{\mathbf{x} \rightarrow \mathbf{x}_e, \mathbf{x} \in \kappa^\pm} z(\mathbf{x}).$$

La figure 2.2 illustre ces notations en deux dimensions d'espace. Par ailleurs, on définit les opérateurs de moyenne $\{z\}$ et de saut $[[z]]$ d'une fonction scalaire z respectivement comme suit :

$$\{z\} = \begin{cases} \frac{(z^+ + z^-)}{2} & \text{si } e \in \mathcal{E}_i, \\ z_b & \text{si } e \in \mathcal{E}_b, \end{cases}$$

$$[[z]] = \begin{cases} z^+ - z^- & \text{si } e \in \mathcal{E}_i, \\ z^+ - z_b & \text{si } e \in \mathcal{E}_b, \end{cases}$$

où z_b représente soit la trace imposée par une condition aux limites de type Dirichlet, soit la trace de z à savoir $z_b = z^+$. La définition de ces opérateurs pour une quantité vectorielle s'étend naturellement. La figure 2.3 illustre les opérateurs de moyenne et de saut en une dimension d'espace.

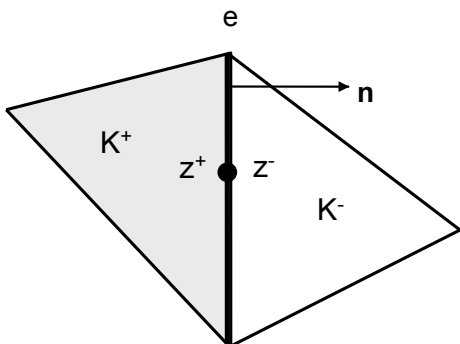


Fig. 2.2 – Exemple pour $d = 2$ des notations à l'interface e . Élément courant κ^+ , élément voisin κ^- , et définitions des traces z^\pm sur l'interface, ainsi que de la normale \mathbf{n} unitaire à e et extérieure à κ^+ .

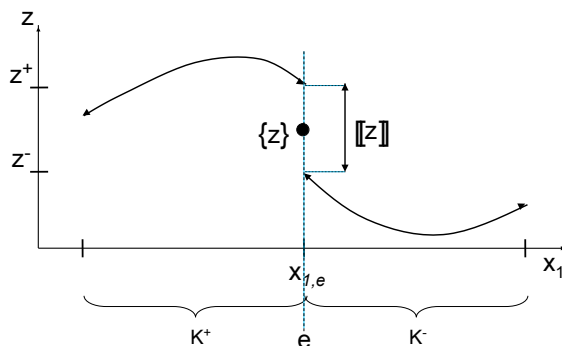


Fig. 2.3 – D'après [38, figure 1.3]. Exemple pour $d = 1$ des opérateurs de moyenne et de saut. L'interface e est réduite à un point séparant deux intervalles adjacents.

2.3.4.2 Discrétisation des termes de convection

Ce paragraphe a pour but de donner les expressions du vecteur des flux numériques convectifs $\hat{\mathbf{F}}_c$, ainsi que de l'opérateur discret \mathcal{L}_c . A cette fin, les propriétés de consistance et de conservativité d'un flux numérique sont définies, puis l'écriture (2.12) de l'approximation interne pour les termes de convection est reconsidérée afin de démontrer que le schéma numérique semi-discret en espace est consistant et conservatif.

Propriétés du flux. Tout d'abord, le flux numérique ne doit dépendre que des états \mathbf{w}_h^+ et \mathbf{w}_h^- afin d'assurer un stencil compact. Ceci explique pourquoi l'approximation de la composante normale du tenseur des flux physiques $\mathbf{F}_c \mathbf{n}$ par le vecteur des flux numériques $\hat{\mathbf{F}}_c$ dans (2.12) a été faite de la manière suivante :

$$\oint_{\partial\kappa} \phi^k \mathbf{F}_c(\mathbf{w}_h) \mathbf{n} dS \approx \oint_{\partial\kappa} \phi^k \hat{\mathbf{F}}_c(\mathbf{w}_h^+, \mathbf{w}_h^-, \mathbf{n}) dS, \quad \forall \kappa \in \Omega_h, \quad 1 \leq k \leq N_p.$$

De plus, le flux numérique doit être consistant et conservatif pour assurer la convergence du schéma en cas de stabilité (théorème de Lax-Wendroff). La définition de la consistance est :

Définition 2.5. (Flux consistant) Un flux numérique $\hat{\mathbf{F}}$ est dit consistant avec le flux physique \mathbf{F} , s'il vérifie la propriété suivante :

$$\hat{\mathbf{F}}(\mathbf{w}, \mathbf{w}, \mathbf{n}) = \mathbf{F}(\mathbf{w}) \mathbf{n}, \quad \forall \mathbf{w} \in \mathcal{V}.$$

Autrement dit, le flux numérique est consistant si on retrouve le flux physique en remplaçant \mathbf{w}_h par \mathbf{w} dans la relation ci-dessus. Enfin, ce flux numérique est recherché conservatif au sens de la définition suivante :

Définition 2.6. (Flux conservatif) Un flux numérique $\hat{\mathbf{F}}$ associé au flux physique \mathbf{F} est dit conservatif s'il vérifie la propriété suivante à travers chaque interface :

$$\hat{\mathbf{F}}(\mathbf{w}_h^+, \mathbf{w}_h^-, \mathbf{n}) = -\hat{\mathbf{F}}(\mathbf{w}_h^-, \mathbf{w}_h^+, -\mathbf{n}), \quad \forall \mathbf{w}_h \in [\mathcal{V}_h^p]^{d+2}.$$

Retour sur l'approximation interne. A présent, les propriétés 2.5 et 2.6 sur les flux numériques sont justifiées afin d'assurer les propriétés de consistance et de conservativité du schéma global (2.13) des termes de convection. C'est l'objet des deux définitions et des deux lemmes suivants.

Pour ce faire, une loi de conservation du type $\partial_t \mathbf{w} + \nabla \cdot \mathbf{F}(\mathbf{w}) = 0$ est temporairement considérée dans Ω . Son approximation interne associée, utilisant un flux numérique $\hat{\mathbf{F}}$, est alors :

$$\sum_{\kappa \in \Omega_h} \int_{\kappa} \phi^k \partial_t \mathbf{w}_h dV - \sum_{\kappa \in \Omega_h} \int_{\kappa} \mathbf{F}(\mathbf{w}_h) \nabla \phi^k dV + \sum_{\kappa \in \Omega_h} \oint_{\partial \kappa} \phi^k \hat{\mathbf{F}}(\mathbf{w}_h^+, \mathbf{w}_h^-, \mathbf{n}) dS = 0. \quad (2.17)$$

Définition 2.7. (Schéma consistant) Le schéma numérique décrit par (2.17) est consistant si la solution \mathbf{w} du problème fort est également solution du problème discret (2.17), c'est-à-dire si :

$$\sum_{\kappa \in \Omega_h} \int_{\kappa} \phi^k \partial_t \mathbf{w} dV - \sum_{\kappa \in \Omega_h} \int_{\kappa} \mathbf{F}(\mathbf{w}) \nabla \phi^k dV + \sum_{\kappa \in \Omega_h} \oint_{\partial \kappa} \phi^k \hat{\mathbf{F}}(\mathbf{w}^+, \mathbf{w}^-, \mathbf{n}) dS = 0.$$

Lemme 2.1. Le schéma numérique décrit par (2.17) est consistant si et seulement si le flux numérique associé est consistant.

Démonstration. On pose $\mathbf{w}_h = \mathbf{w}$ dans l'approximation interne (2.17) :

$$\sum_{\kappa \in \Omega_h} \int_{\kappa} \phi^k \partial_t \mathbf{w} dV - \sum_{\kappa \in \Omega_h} \int_{\kappa} \mathbf{F}(\mathbf{w}) \nabla \phi^k dV + \sum_{\kappa \in \Omega_h} \oint_{\partial \kappa} \phi^k \hat{\mathbf{F}}(\mathbf{w}, \mathbf{w}, \mathbf{n}) dS = 0,$$

où $\mathbf{w}^+ = \mathbf{w}^- = \mathbf{w}$ par continuité de \mathbf{w} , solution régulière du problème fort. Cette propriété de régularité permet également d'intégrer le second terme par parties, il vient :

$$\sum_{\kappa \in \Omega_h} \int_{\kappa} \phi^k (\partial_t \mathbf{w} + \nabla \cdot \mathbf{F}(\mathbf{w})) dV - \sum_{\kappa \in \Omega_h} \oint_{\partial \kappa} \phi^k (\mathbf{F}(\mathbf{w}) \mathbf{n} - \hat{\mathbf{F}}(\mathbf{w}, \mathbf{w}, \mathbf{n})) dS = 0,$$

or le premier terme est nul puisque \mathbf{w} vérifie la formulation forte du problème. Il reste

$$\sum_{\kappa \in \Omega_h} \oint_{\partial \kappa} \phi^k (\mathbf{F}(\mathbf{w}) \mathbf{n} - \hat{\mathbf{F}}(\mathbf{w}, \mathbf{w}, \mathbf{n})) dS = 0.$$

Cette dernière relation est vérifiée si et seulement si le flux numérique $\hat{\mathbf{F}}$ est consistant au sens de la définition 2.5. □

Définition 2.8. (Schéma conservatif) Le schéma numérique (2.17) est conservatif si

$$\sum_{\kappa \in \Omega_h} \int_{\kappa} \partial_t \mathbf{w}_h dV + \sum_{\kappa \in \Omega_h} \oint_{\partial \kappa} \mathbf{F}(\mathbf{w}_b(\mathbf{w}_h^+, \mathbf{n})) \mathbf{n} dS = 0,$$

où \mathbf{w}_b est un opérateur permettant d'imposer les conditions aux limites physiques sur \mathcal{E}_b (cf. § 1.3 page 30 pour plus de détails sur ces conditions). Autrement dit, la variation de la grandeur \mathbf{w}_h à l'intérieur du domaine ne résulte que de la différence des flux entrant et sortant aux frontières du domaine en l'absence de termes sources.

Lemme 2.2. Le schéma numérique décrit par (2.17) est conservatif si et seulement si le flux numérique associé est conservatif.

Démonstration. Soit κ appartenant à Ω_h , on pose $\phi^k = 1_{\kappa}$ la fonction indicatrice, restreinte à κ et nulle ailleurs, dans l'approximation interne (2.17) :

$$\sum_{\kappa \in \Omega_h} \int_{\kappa} 1_{\kappa} \partial_t \mathbf{w}_h dV - \sum_{\kappa \in \Omega_h} \int_{\kappa} \mathbf{F}(\mathbf{w}_h) \nabla \phi^k dV + \sum_{\kappa \in \Omega_h} \oint_{\partial \kappa} 1_{\kappa} \hat{\mathbf{F}}(\mathbf{w}_h^+, \mathbf{w}_h^-, \mathbf{n}) dS = 0,$$

où les intégrales de bord peuvent être réécrites de la manière suivante (cf. remarque 2.9) :

$$\begin{aligned} \sum_{\kappa \in \Omega_h} \oint_{\partial \kappa} \hat{\mathbf{F}}(\mathbf{w}_h^+, \mathbf{w}_h^-, \mathbf{n}) dS &= \sum_{e \in \mathcal{E}_i} \int_e \hat{\mathbf{F}}(\mathbf{w}_h^+, \mathbf{w}_h^-, \mathbf{n}) + \hat{\mathbf{F}}(\mathbf{w}_h^-, \mathbf{w}_h^+, -\mathbf{n}) dS \\ &+ \sum_{e \in \mathcal{E}_b} \int_e \mathbf{F}(\mathbf{w}_b(\mathbf{w}_h^+, \mathbf{n})) \mathbf{n} dS. \end{aligned} \quad (2.18)$$

En regroupant les deux relations précédentes, il vient :

$$\begin{aligned} \sum_{\kappa \in \Omega_h} \int_{\kappa} \partial_t \mathbf{w}_h dV + \sum_{e \in \mathcal{E}_i} \int_e \hat{\mathbf{F}}(\mathbf{w}_h^+, \mathbf{w}_h^-, \mathbf{n}) + \hat{\mathbf{F}}(\mathbf{w}_h^-, \mathbf{w}_h^+, -\mathbf{n}) dS \\ + \sum_{e \in \mathcal{E}_b} \int_e \mathbf{F}(\mathbf{w}_b(\mathbf{w}_h^+, \mathbf{n})) \mathbf{n} dS = 0. \end{aligned}$$

Le schéma est conservatif si et seulement si le second terme s'annule, c'est-à-dire si et seulement si le flux numérique est lui-même conservatif au sens de la définition 2.6. \square

Remarque 2.9. Dans la réécriture (2.18), une vision par face (membre de droite) a été adoptée, par opposition à une vision par élément (membre de gauche). Autrement dit, les frontières de la partition Ω_h sont regardées les unes après les autres (par opposition aux frontières de chaque élément κ les uns après les autres). En s'intéressant à chaque face séparément, les faces de bord interviennent une seule fois dans le calcul alors que les faces internes interviennent deux fois : une première fois vis-à-vis de l'élément κ^+ et une seconde fois vis-à-vis de l'élément κ^- . Enfin, il faut noter que sur les faces de bord, le flux considéré est bien le flux physique.

Expression du flux. L'opérateur \mathcal{L}_c est associé à la partie hyperbolique du système des équations de Navier-Stokes. Il est obtenu en utilisant un flux numérique qui respecte les propriétés de consistance et de conservativité et qui est solution exacte ou approchée de problèmes de Riemann locaux à chaque interface. Un panorama des différents flux numériques issus des méthodes de volumes finis a déjà été donné dans l'introduction générale de ce mémoire (cf. § 0.3.3.2 page 11). On rappelle [28, 96] que selon l'expérience numérique, plus le degré p du polynôme augmente, moins le flux numérique choisi n'a d'importance sur la précision de la solution numérique.

Dans ce travail, le flux de Lax-Friedrichs local (LLF) [156] a été retenu pour sa simplicité de mise en œuvre sur un système d'équations non linéaires multidimensionnelles, pour son faible coût en nombre d'opérations à virgules flottantes par seconde et enfin pour ses performances d'approximation sur lesquelles on revient au paragraphe 5.1.3 page 107 [28, 131].

Le flux numérique LLF est un flux décentré amont dont la version légèrement modifiée suivante est utilisée ici [33] :

$$\hat{\mathbf{F}}_c(\mathbf{w}_h^+, \mathbf{w}_h^+, \mathbf{n}) = \{\mathbf{F}_c(\mathbf{w}_h)\} \mathbf{n} + k_2 \frac{a}{2} \llbracket \mathbf{w}_h \rrbracket, \quad (2.19)$$

avec

$$a = \max \left\{ \left| \frac{\partial \mathbf{F}_c \mathbf{n}}{\partial \mathbf{w}} \right| : \mathbf{w} = \mathbf{w}_h^\pm \right\},$$

et où $|\partial \mathbf{F}_c \mathbf{n} / \partial \mathbf{w}|$ représente le rayon spectral de la matrice jacobienne des flux convectifs suivant la direction de la normale \mathbf{n} , et k_2 est un paramètre de stabilisation appartenant à $[0, 1]$. Ce dernier permet de passer continûment d'un flux centré ($k_2 = 0$) au flux LLF original ($k_2 = 1$). Par conséquent, plus k_2 est grand, plus le schéma numérique est stabilisé par l'ajout de dissipation numérique.

La discrétisation des termes de convection s'écrit :

$$\begin{aligned} \mathcal{L}_c(\mathbf{w}_h, \phi) &= - \int_{\Omega_h} \mathbf{F}_c(\mathbf{w}_h) \nabla_h \phi dV \\ &+ \int_{\mathcal{E}_i} \llbracket \phi \rrbracket \hat{\mathbf{F}}_c(\mathbf{w}_h^+, \mathbf{w}_h^-, \mathbf{n}) dS \\ &+ \int_{\mathcal{E}_b} \phi^+ \mathbf{F}_c(\mathbf{w}_b(\mathbf{w}_h^+, \mathbf{n})) \mathbf{n} dS, \end{aligned} \quad (2.20)$$

où l'opérateur de gradient discret ∇_h est défini sur chaque élément κ appartenant à Ω_h comme :

$$(\nabla_h f)|_\kappa = \nabla(f|_\kappa), \quad \forall f \in H^1(\Omega_h), \quad (2.21)$$

où les notations suivantes ont été adoptées pour les intégrales de surface :

$$\int_{\mathcal{E}_i} f dS = \sum_{e \in \mathcal{E}_i} \int_e f dS, \quad (2.22a)$$

$$\int_{\mathcal{E}_b} f dS = \sum_{e \in \mathcal{E}_b} \int_e f dS, \quad \forall f \in H^1(\Omega_h). \quad (2.22b)$$

Remarque 2.10. L'intégrale de volume dans (2.20) n'existe pas pour la méthode des volumes finis. Il est précisé ici, que son évaluation induit un surplus de calcul non négligeable de la méthode GD par rapport à la méthode des volumes finis. Il en est de même de l'utilisation d'intégrations numériques pour évaluer chaque intégrale.

2.3.4.3 Discrétisation des termes de diffusion

L'opérateur \mathcal{L}_v est associé à la partie parabolique du système des équations de Navier-Stokes. La présence des termes de dérivées secondes en espace oblige à considérer aux interfaces, non plus uniquement des sauts sur la solution, mais également des sauts sur le gradient de la solution. Dans le cadre des méthodes GD, une approche directe (similaire aux termes de convection) des flux physiques visqueux par des flux numériques conduit à un schéma numérique inconsistant : c'est un phénomène dit de « crime variationnel » [28, 183]. Le lecteur trouvera en annexe B un exemple de ce phénomène dans le cas simplifié de l'équation de la chaleur monodimensionnelle scalaire.

Pour pallier le problème d'inconsistance, deux techniques principales existent, elles ont déjà été abordées dans l'introduction générale de ce mémoire (*cf.* § 0.3.3.3 page 11) : celle de la formulation mixte [4] et celle de reconstruction ou de recouvrement local de la solution [169, 170]. Parmi les techniques de formulation mixte, celle appliquant le second schéma (ou flux) de Bassi et Rebay (BR2) [13, 7] a été retenue dans ce travail. Le schéma BR2 introduit des opérateurs de relèvement pour approcher le saut du gradient de la solution numérique aux interfaces. Il est intéressant puisqu'il garantit, outre un schéma consistant et conservatif, un ordre de convergence optimal de la norme L^2 de l'erreur entre la solution analytique et la solution numérique sur des équations scalaires linéaires ou non en deux dimensions d'espace. De plus, son stencil est compact [4].

À présent, la technique de formulation mixte utilisant le schéma BR2 est présentée pour le système des équations de Navier-Stokes. Tout d'abord, pour contourner le problème du crime variationnel, les méthodes de formulation mixte reconsidèrent la partie parabolique du système des équations de Navier-Stokes en un système couplé du premier ordre, avec $\mathbf{\Upsilon}$ appartenant à $\mathbb{R}^{(d+2) \times d}$ la variable auxiliaire, qui s'écrit :

$$\mathbf{\Upsilon} = \nabla \mathbf{w}, \quad (2.23a)$$

$$\partial_t \mathbf{w} - \nabla \cdot (\mathbf{G}(\mathbf{w}) \mathbf{\Upsilon}) = 0, \quad (2.23b)$$

où la première équation est appelée équation des gradients et où on a tenu compte de la remarque 2.11 suivante.

Remarque 2.11. *Le flux visqueux est linéaire par rapport au gradient du vecteur des variables conservatives. En considérant la convention de sommation sur les indices répétés $1 \leq i, j \leq 2$, il vient*

$$\begin{aligned} \nabla \cdot \mathbf{F}_v(\mathbf{w}, \nabla \mathbf{w}) &= \nabla \cdot (\mathbf{G}(\mathbf{w}) \nabla \mathbf{w}), \\ &= \frac{\partial}{\partial x_i} \left(G_{ij}(\mathbf{w}) \frac{\partial \mathbf{w}}{\partial x_j} \right), \end{aligned}$$

où les matrices $(G_{ij})_{i,j=1,2}$ appartiennent à $\mathbb{R}^{(d+2) \times (d+2)}$ et sont telles que :

$$G_{ij} = \frac{\partial \mathbf{f}_{v,i}}{\partial \frac{\partial \mathbf{w}}{\partial x_j}}.$$

L'introduction du tenseur \mathbf{G} , de composantes matricielles $(G_{ij})_{i,j=1,2}$, est uniquement utilisée pour formuler théoriquement le schéma BR2. Pour plus de détails sur l'écriture de \mathbf{G} , voir [58] par exemple.

En suivant les étapes pour établir l'approximation interne expliquées au paragraphe 2.3.3 et en utilisant les relations (2.15) et (2.22) sur les opérateurs intégraux, l'approximation interne du système (2.23) s'écrit directement sous la forme générique : *Trouver le couple $(\Upsilon_h, \mathbf{w}_h)$ appartenant à $[\mathcal{V}_h^p]^{(d+2) \times d} \times [\mathcal{V}_h^p]^{d+2}$, tel que pour tout couple (φ^k, ϕ^k) éléments de $[\mathcal{B}_h^p]^d \times \mathcal{B}_h^p$, il vienne à chaque instant t dans \mathcal{T}*

$$\int_{\Omega_h} \Upsilon_h \varphi^k dV + \int_{\Omega_h} (\nabla \cdot \varphi^k) \mathbf{w}_h dV - \int_{\mathcal{E}_i} (\{\mathbf{w}_h\} \otimes \mathbf{n}) [\varphi^k] dS - \int_{\mathcal{E}_b} (\mathbf{w}_b \otimes \mathbf{n}) \varphi^{k,+} dS = 0, \quad (2.24a)$$

$$\int_{\Omega_h} \phi^k \partial_t \mathbf{w}_h dV + \int_{\Omega_h} (\mathbf{G}\Upsilon_h) \nabla \phi^k dV - \int_{\mathcal{E}_i} [\phi^k] \{\mathbf{G}\Upsilon_h\} \mathbf{n} dS - \int_{\mathcal{E}_b} \phi^{k,+} \mathbf{F}_v(\mathbf{w}_b, \nabla \mathbf{w}_b) \mathbf{n} dS = 0, \quad (2.24b)$$

où $\mathbf{w}_b = \mathbf{w}_b(\mathbf{w}_h^+, \mathbf{n})$ et $\nabla \mathbf{w}_b = \nabla \mathbf{w}_b(\mathbf{w}_h^+, \mathbf{n})$ sont des opérateurs permettant d'imposer les conditions aux limites physiques sur \mathcal{E}_b , et où en première approche, on a considéré les flux numériques centrés $\hat{\mathbf{w}}_h = \{\mathbf{w}_h\} \otimes \mathbf{n}$ et $\mathbf{G}\Upsilon_h = \{\mathbf{G}\Upsilon_h\} \mathbf{n}$ sur \mathcal{E}_i , et ce en accord avec le phénomène de diffusion qui ne privilégie aucune direction spatiale dans l'écoulement.

A ce stade, le système (2.24) constitue le schéma BR1 [9] qui consiste, à chaque pas de temps, à résoudre (2.24a) pour obtenir Υ_h , puis à résoudre (2.24b) pour obtenir \mathbf{w}_h . Cependant, l'utilisation de l'équation des gradients (2.24a) pour déterminer Υ_h conduit à un schéma numérique non compact et d'ordre de convergence non optimal pour les problèmes purement elliptiques [11, 4]. La non-compactité s'observe facilement puisque le calcul de la solution numérique au sein d'un élément κ appartenant à Ω_h demande de connaître les valeurs de Υ_h dans les éléments voisins *via* le terme $\{\mathbf{G}\Upsilon_h\}$ dans (2.24b), qui font elles-mêmes intervenir les valeurs de \mathbf{w}_h dans les éléments voisins des voisins *via* le terme $\{\mathbf{w}_h\}$ dans (2.24a).

Afin, entre autres, de rendre le stencil du schéma plus compact, deux opérateurs de relèvement sont introduits. Ils se substituent à l'équation (2.24a) et permettent d'approcher directement Υ_h dans l'équation (2.24b). Il s'agit du schéma BR2 [13, 7].

Relèvement global pour l'équation des gradients. Il est obtenu au moyen d'une réécriture de l'équation (2.24a). On considère tout d'abord, la proposition suivante qui utilise la relation de Leibniz :

Proposition 2.1. *Pour tout couple de vecteurs (\mathbf{f}, \mathbf{g}) appartenant à $[\mathcal{H}^1(\Omega_h)]^{d+2} \times [\mathcal{H}^1(\Omega_h)]^d$, la propriété algébrique suivante est vérifiée :*

$$\int_{\Omega_h} (\nabla_h \cdot \mathbf{g}) \mathbf{f} dV = - \int_{\Omega_h} (\nabla_h \mathbf{f}) \mathbf{g} dV + \int_{\mathcal{E}_i} \left((\{\mathbf{f}\} \otimes \mathbf{n}) [\mathbf{g}] + ([\mathbf{f}] \otimes \mathbf{n}) \{\mathbf{g}\} \right) dS + \int_{\mathcal{E}_b} (\mathbf{f}^+ \otimes \mathbf{n}) \mathbf{g}^+ dS,$$

où l'opérateur ∇_h a été défini à la relation (2.21) et où l'opérateur de divergence discret $\nabla_h \cdot$ suit une définition similaire.

Démonstration. Il suffit d'intégrer par parties le membre de gauche dans chaque élément κ , puis de sommer les différentes contributions. On utilise à cette fin, l'identité suivante :

$$(\{\mathbf{f}\} \otimes \mathbf{n}) \llbracket \mathbf{g} \rrbracket + (\llbracket \mathbf{f} \rrbracket \otimes \mathbf{n}) \{\mathbf{g}\} = (\mathbf{f}^+ \otimes \mathbf{n}) \mathbf{g}^+ - (\mathbf{f}^- \otimes \mathbf{n}) \mathbf{g}^-. \quad (2.25)$$

□

Ensuite, en posant $\mathbf{f} \equiv \mathbf{w}_h$ et $\mathbf{g} \equiv \boldsymbol{\varphi}^k$ dans la relation (2.25) ci-avant, l'application à (2.24a) de la proposition 2.1 conduit à :

$$\int_{\Omega_h} (\boldsymbol{\Upsilon}_h - \nabla_h \mathbf{w}_h) \boldsymbol{\varphi}^k dV + \int_{\mathcal{E}_i} (\llbracket \mathbf{w}_h \rrbracket \otimes \mathbf{n}) \{\boldsymbol{\varphi}^k\} dS + \int_{\mathcal{E}_b} (\llbracket \mathbf{w}_h \rrbracket \otimes \mathbf{n}) \boldsymbol{\varphi}^{k,+} dS = 0. \quad (2.26)$$

On définit alors l'opérateur de relèvement global comme suit :

Définition 2.9. (Relèvement global) Soit $\mathbf{R}_h(\llbracket \mathbf{w}_h \rrbracket)$ appartenant à $\mathbb{R}^{(d+2) \times d}$, l'opérateur de relèvement global, tel que

$$\int_{\Omega_h} \phi^k \mathbf{R}_h(\llbracket \mathbf{w}_h \rrbracket) dV = - \int_{\mathcal{E}_i} \{\phi^k\} \llbracket \mathbf{w}_h \rrbracket \otimes \mathbf{n} dS - \int_{\mathcal{E}_b} \phi^{k,+} \llbracket \mathbf{w}_h \rrbracket \otimes \mathbf{n} dS, \quad \forall \phi^k \in \mathcal{B}_h^p.$$

Au sens des distributions, (2.25) s'écrit

$$\boldsymbol{\Upsilon}_h \triangleq \nabla \mathbf{w}_h + \mathbf{R}_h(\llbracket \mathbf{w}_h \rrbracket). \quad (2.27)$$

Remarque 2.12. Le relèvement global est une correction du gradient de la solution numérique pour tenir compte de ses discontinuités aux interfaces. En particulier, si la solution \mathbf{w}_h est continue aux interfaces il vient $\llbracket \mathbf{w}_h \rrbracket = 0$, et donc $\mathbf{R}_h \equiv 0$.

Relèvement local pour l'équation d'évolution. La résolution de l'équation d'évolution (2.24b) nécessite de calculer $\{\mathbf{G}\boldsymbol{\Upsilon}_h\}$. Avec (2.27), une première approche serait de choisir $\{\mathbf{G}\boldsymbol{\Upsilon}_h\} = \{\mathbf{G}\nabla \mathbf{w}_h\} + \{\mathbf{G}\mathbf{R}_h\}$. Avec un tel choix, le schéma constitué de (2.26) et (2.24b) reste non compact puisque l'introduction de l'opérateur de relèvement global ne constitue qu'une réécriture du schéma BR1. On définit alors l'opérateur de relèvement local, pour se substituer à l'opérateur de relèvement global aux interfaces, comme suit :

Définition 2.10. (Relèvement local) Soit $\mathbf{r}_h^e(\llbracket \mathbf{w}_h \rrbracket)$ appartenant à $\mathbb{R}^{(d+2) \times d}$, l'opérateur de relèvement local à une face e , tel que

$$\int_{\kappa^+ \cup \kappa^-} \phi^k \mathbf{r}_h^e(\llbracket \mathbf{w}_h \rrbracket) dV = \begin{cases} - \int_e \{\phi^k\} \llbracket \mathbf{w}_h \rrbracket \otimes \mathbf{n} dS & \text{si } e \in \mathcal{E}_i, \\ - \int_e \phi^{k,+} \llbracket \mathbf{w}_h \rrbracket \otimes \mathbf{n} dS & \text{si } e \in \mathcal{E}_b, \end{cases} \quad \forall \phi^k \in \mathcal{B}_h^p.$$

Remarque 2.13. L'opérateur de relèvement local est restreint aux deux éléments κ^+ et κ^- partageant l'interface e et par conséquent, il est identiquement nul en dehors du support $\kappa^+ \cup \kappa^-$. De plus, comme \mathbf{r}_h^e n'est pas continu à la traversée de e , il convient de distinguer sa trace dans l'élément courant de celle dans l'élément voisin, soit respectivement $\mathbf{r}_h^{e,+}$ et $\mathbf{r}_h^{e,-}$.

Au sens des distributions, les deux opérateurs sont reliés comme suit :

$$\mathbf{R}_h \triangleq \sum_{e \in \mathcal{E}_h} \mathbf{r}_h^e. \quad (2.28)$$

Démonstration. Il s'agit de partir de la définition 2.9 de l'opérateur de relèvement global en utilisant les relations (2.22) sur les opérateurs intégraux. On regroupe les sommes sur les ensembles de faces à l'aide de (2.4), puis on utilise la définition 2.10 de l'opérateur de relèvement local. Il vient :

$$\begin{aligned} \int_{\Omega_h} \phi^k \mathbf{R}_h(\llbracket \mathbf{w}_h \rrbracket) dS &= - \sum_{e \in \mathcal{E}_h} \int_e \{ \phi^k \} \llbracket \mathbf{w}_h \rrbracket \otimes \mathbf{n} dS, \quad \forall \phi^k \in \mathcal{B}_h^p, \\ &= \sum_{e \in \mathcal{E}_h} \int_{\kappa^+ \cup \kappa^-} \phi^k \mathbf{r}_h^e(\llbracket \mathbf{w}_h \rrbracket) dV, \quad \forall \phi^k \in \mathcal{B}_h^p, \end{aligned}$$

or l'égalité précédente est vraie pour tout ϕ^k appartenant à \mathcal{B}_h^p . Donc l'égalité (2.28) est vraie au sens des distributions. \square

Le schéma BR2 s'obtient alors en remplaçant $\{\mathbf{G}\Upsilon_h\}$ par $\{\mathbf{G}\nabla \mathbf{w}_h\} + \{\mathbf{G}\mathbf{r}_h^e\}$ pour l'évaluation du flux numérique. La

$$\begin{aligned} \mathcal{L}_v(\mathbf{w}_h, \phi) &= - \int_{\Omega_h} \mathbf{F}_v(\mathbf{w}_h, \nabla_h \mathbf{w}_h + \mathbf{R}_h) \nabla \phi dV \\ &\quad + \int_{\mathcal{E}_i} \llbracket \phi \rrbracket \hat{\mathbf{F}}_v(\mathbf{w}_h^+, \nabla \mathbf{w}_h^+ + \mathbf{r}_h^{e,+}, \mathbf{w}_h^-, \nabla \mathbf{w}_h^- + \mathbf{r}_h^{e,-}, \mathbf{n}) dS \\ &\quad + \int_{\mathcal{E}_b} \phi^+ \mathbf{F}_v(\mathbf{w}_b, \nabla \mathbf{w}_b + \eta_r \mathbf{r}_h^e(\llbracket \mathbf{w}_h \rrbracket)) \mathbf{n} dS, \end{aligned} \quad (2.29)$$

avec

$$\begin{aligned} \mathbf{F}_v(\mathbf{w}_h, \nabla_h \mathbf{w}_h + \mathbf{R}_h) &= \mathbf{G}(\mathbf{w}_h) (\nabla_h \mathbf{w}_h + \mathbf{R}_h(\llbracket \mathbf{w}_h \rrbracket)), \\ \hat{\mathbf{F}}_v(\mathbf{w}_h^+, \nabla \mathbf{w}_h^+ + \mathbf{r}_h^{e,+}, \mathbf{w}_h^-, \nabla \mathbf{w}_h^- + \mathbf{r}_h^{e,-}, \mathbf{n}) &= \left\{ \mathbf{G}(\mathbf{w}_h) (\nabla \mathbf{w}_h + \eta_r \mathbf{r}_h^e(\llbracket \mathbf{w}_h \rrbracket)) \right\}, \end{aligned}$$

où η_r est un paramètre de stabilisation (*cf.* remarque 2.14).

Remarque 2.14. Dans la discrétisation BR2 originale de Bassi et Rebay [13] le paramètre η_r n'apparaît pas et, de fait, $\eta_r = 1$. Brezzi et al. [17] prouvent que $\eta_r \geq 3$ est une condition suffisante à la stabilité du schéma BR2 dans le cas de la discrétisation de l'équation de Poisson bidimensionnelle sur des maillages triangulaires. Dans le cas du système des équations de Navier-Stokes, on a pris $\eta_r = 1$ pour les applications numériques de la partie II sans rencontrer de problème de stabilisation particulier.

2.4 Fonctions de base et intégration numérique : éléments droits et courbes

A ce stade, pour déterminer complètement le problème semi-discret en espace, il reste à définir la base modale, orthonormale et hiérarchique \mathcal{B}_κ^p de l'espace $\mathcal{P}^p(\kappa)$, et

à calculer les intégrales présentes dans le système (2.14). Ce qui suit concerne plus particulièrement l'aspect pratique de la méthode GD, et est relativement classique dans le cas de maillages à éléments droits, mais plus délicat dans le cas de maillages à éléments courbes. Ces derniers apparaissent, par exemple, au voisinage des parois courbes. En GD, il est nécessaire de les prendre en compte pour minimiser l'erreur géométrique qui peut devenir prépondérante devant l'erreur d'approximation du schéma numérique dans une simulation d'ordre élevé [3, 10]. L'application numérique du paragraphe 5.1 page 101, illustre cette propriété.

Concernant le choix de la base \mathcal{B}_κ^p pour les éléments droits, il suit les travaux de [42] qui introduisent une transformation d'un élément de référence $I^d = [-1, 1]^d$ vers l'élément physique κ . Quelle que soit la dimension d'espace, cet élément de référence permet d'utiliser le produit tensoriel de familles de polynômes définis sur un intervalle de l'axe des réels, afin d'établir une base orthogonale vis-à-vis du produit intérieur (2.11) dans I^d . Pour les éléments courbes, il n'est pas connu de base orthogonale vis-à-vis du produit intérieur dans l'élément κ . En revanche, une telle base peut être construite à partir de l'orthonormalisation de monômes définis dans κ .

Concernant l'intégration numérique, toutes les intégrales de l'approximation interne sont approchées par une formule de quadrature de type Gauss dont les points d'intégration sont définis dans I^d .

Ainsi, dans la pratique, qu'il s'agisse de la construction des fonctions de base (§ 2.4.2) ou de l'écriture de la quadrature numérique (§ 2.4.3), il est nécessaire de définir la transformation de l'élément de référence I^d vers le triangle physique κ , notée \mathbf{F}_κ . C'est l'objet du paragraphe suivant.

2.4.1 Transformation de l'élément de référence vers l'élément physique présentant une géométrie courbe

Cette transformation fait le lien entre les coordonnées des points d'intégration uniquement définis à l'intérieur de l'élément de référence I^d (cf. § 2.4.3) et les coordonnées connues de l'élément physique κ . Elle est directement donnée en deux dimensions d'espace et valable dans le cas général de maillages à éléments droits ou courbes. On considère les trois systèmes de coordonnées suivants, illustrés respectivement par les figures 2.4a, 2.4b et 2.4c.

- Le premier système de coordonnées est $(0, x_1, x_2)$, les points sont notés $\mathbf{x} = (x_1, x_2)^\top$ et le triangle physique κ appartenant à Ω_h est défini grâce à ses sommets \mathbf{x}_{S_1} , \mathbf{x}_{S_2} et \mathbf{x}_{S_3} . Chacune des faces $(e_{ij})_{i,j=1,2,3;i \neq j}$ du triangle, a pour sommets \mathbf{x}_{S_i} et \mathbf{x}_{S_j} , et est décrite par $m + 1$ points (cf. définition 2.11). Les coordonnées de ces points sont généralement fournies par un logiciel de construction de maillages. A titre d'exemple, la figure 2.5 ci-contre illustre la différence dans l'orientation des normales à la paroi dans le cas d'éléments droits ($m = 1$) ou courbes ($m > 1$).
- Le second système de coordonnées est $(0, \eta_1, \eta_2)$, les points sont notés $\boldsymbol{\eta} = (\eta_1, \eta_2)^\top$ et le triangle de référence est défini tel que :

$$T^2 = \{(\eta_1, \eta_2) : 0 < \eta_2 \leq \eta_1 < 1\}.$$

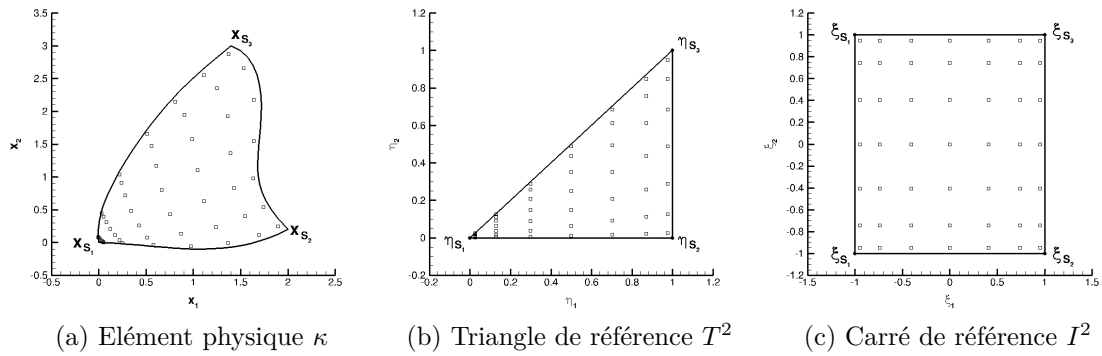


Fig. 2.4 – Systèmes de coordonnées et images des points de quadrature de Gauss dans chaque élément pour $m = 6$.

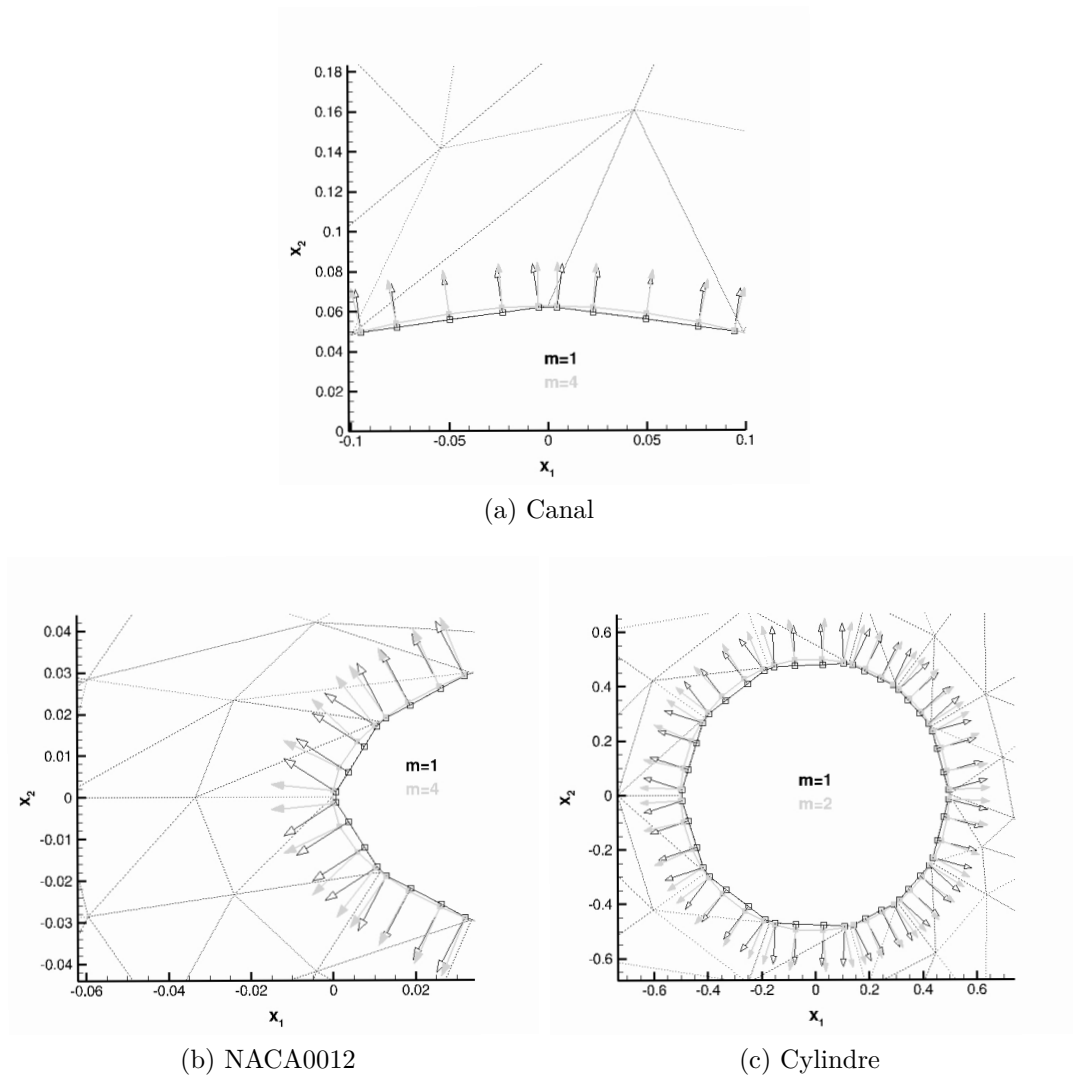


Fig. 2.5 – Comparaison des normales à la paroi aux points d'intégration numérique pour trois configurations rencontrées dans les applications numériques de la partie II sur des maillages de type linéaire ($m = 1$) en noir ou de type curviligne ($m = 2$ ou $m = 4$) en gris.

- Le troisième système de coordonnées est (O, ξ_1, ξ_2) , les points sont notés $\boldsymbol{\xi} = (\xi_1, \xi_2)^\top$ et le carré de référence est défini tel que :

$$I^2 = \{(\xi_1, \xi_2) : -1 < \xi_1, \xi_2 < 1\}.$$

Décrire la transformation \mathbf{F}_κ du carré de référence I^2 vers le triangle physique κ , revient à décrire les points physiques \mathbf{x} à l'intérieur de l'élément κ , comme les images des points $\boldsymbol{\xi}$ issus de l'élément de référence I^2 . Deux étapes sont alors nécessaires : définir la transformation du carré de référence vers le triangle de référence, notée $\mathbf{F}_{\kappa,1}$, puis celle du triangle de référence vers le triangle physique, notée $\mathbf{F}_{\kappa,2}$. Il vient donc $\mathbf{F}_\kappa = \mathbf{F}_{\kappa,2} \circ \mathbf{F}_{\kappa,1}$.

La première transformation $\mathbf{F}_{\kappa,1}$ est directe, elle s'écrit :

$$\begin{aligned} \mathbf{F}_{\kappa,1} : I^2 &\rightarrow T^2 \\ \boldsymbol{\xi} &\mapsto \mathbf{F}_{\kappa,1}(\boldsymbol{\xi}) = \boldsymbol{\eta}, \end{aligned} \quad (2.30)$$

avec

$$\eta_1 = \frac{1 + \xi_1}{2}, \quad \eta_2 = \frac{(1 + \xi_1)(1 + \xi_2)}{4}.$$

La seconde transformation $\mathbf{F}_{\kappa,2}$ est plus complexe car elle doit prendre en compte les courbures éventuelles du triangle physique κ . Soit \mathbf{C}_{ij} l'unique interpolation polynomiale de degré m passant par les $m + 1$ points décrivant la face $(e_{ij})_{i,j=1,2,3;i \neq j}$:

$$\begin{aligned} \mathbf{C}_{ij} : [0, 1] &\rightarrow e_{ij} \subset \mathbb{R}^2 \\ \eta_1 &\mapsto \mathbf{C}_{ij}(\eta_1) = \mathbf{x}, \end{aligned} \quad (2.31)$$

telle que $\mathbf{C}_{ij}(0) = \mathbf{x}_{S_i}$ et $\mathbf{C}_{ij}(1) = \mathbf{x}_{S_j}$ représentent les sommets de la face.

Définition 2.11. *Un élément κ possédant chaque arête décrite par (2.31) est dit d'ordre $m + 1$.*

La face e_{12} est d'abord considérée, alors une transformation possible de T^2 vers κ est [151] :

$$\mathbf{x} = \frac{\eta_1 - \eta_2}{\eta_1} \mathbf{C}_{12}(\eta_1) + \frac{\eta_2(1 - \eta_1)}{\eta_1} \mathbf{x}_{S_1} + \eta_2 \mathbf{x}_{S_3}, \quad \forall \mathbf{x} \in e_{12}. \quad (2.32)$$

En utilisant une approche similaire pour les autres faces, la transformation $\mathbf{F}_{\kappa,2}$ s'écrit :

$$\begin{aligned} \mathbf{F}_{\kappa,2} : T^2 &\rightarrow \kappa \subset \mathbb{R}^2 \\ \boldsymbol{\eta} &\mapsto \mathbf{F}_{\kappa,2}(\boldsymbol{\eta}) = \mathbf{x}, \end{aligned} \quad (2.33)$$

avec

$$\begin{aligned} \mathbf{x} = & \frac{\eta_1 - \eta_2}{\eta_1} \mathbf{C}_{12}(\eta_1) + \frac{\eta_1 - \eta_2}{1 - \eta_2} \mathbf{C}_{23}(\eta_2) + \frac{1 - \eta_1}{1 - \eta_2} \mathbf{C}_{13}(\eta_2) \\ & - \frac{(\eta_1 - \eta_2)(1 - \eta_1)}{\eta_1} \mathbf{x}_{S_1} - (\eta_1 - \eta_2) \mathbf{x}_{S_2}. \end{aligned}$$

Remarque 2.15. Si la face e_{12} est considérée pour des éléments droits, alors il vient $\mathcal{C}_{12} = (1 - \eta_1)\mathbf{x}_{S_1} + \eta_1\mathbf{x}_{S_2}$. De fait, à partir de (2.32), la transformation bilinéaire classique est bien retrouvée pour $\mathbf{F}_{\kappa,2}$:

$$\mathbf{x} = (1 - \eta_1)\mathbf{x}_{S_1} + (\eta_1 - \eta_2)\mathbf{x}_{S_2} + \eta_2\mathbf{x}_{S_3}.$$

Remarque 2.16. Dans la littérature, d'autres options possibles sont présentées pour (2.32) et elles ne semblent pas affecter les résultats [151]. Par exemple, Kopriva propose [93, chapitre 6] une interpolation linéaire pour représenter les points à l'intérieur des éléments et une interpolation transfinie pour représenter les faces. L'utilisation des polynômes de Lagrange pour représenter l'élément complet est aussi proposée.

2.4.2 Expression des fonctions de base

Dans ce travail, trois bases \mathcal{B}_κ^p sont utilisées suivant la dimension d'espace concernée et le type de maillages considéré. D'une part, pour l'application en une dimension d'espace de l'annexe B, il s'agit de la base des polynômes de Legendre [2, chapitres 8 et 22] (§ 2.4.2.1). D'autre part, pour les applications numériques en deux dimensions d'espace de la partie II, il s'agit de la base dite de Dubiner [42], [153, chapitre 3] sur les maillages à éléments droits (§ 2.4.2.2) et d'une base dite numérique [134], [38, annexe A] issue de l'orthonormalisation de monômes sur les maillages à éléments courbes (§ 2.4.2.3). Ces trois bases sont succinctement présentées en donnant l'expression des fonctions qui les composent.

2.4.2.1 Base des polynômes de Legendre en une dimension d'espace

Les polynômes de Legendre [2, chapitres 8 et 22] sont un exemple très classique d'une famille de polynômes, orthogonaux et hiérarchiques, issus du procédé d'orthonormalisation de Gram-Schmidt à partir des polynômes de la base canonique pour le produit scalaire $(f, g)_{L^2([0,1])}$. Soit \mathbb{L}^p le polynôme de Legendre de degré p défini sur le segment de référence $I = [-1, 1]$, il vient :

$$\phi^k(x_1) = \mathbb{L}^{k-1}(\xi_1), \quad \forall x_1 \in \kappa, \quad \kappa \in \Omega_h, \quad 1 \leq k \leq N_p,$$

où la relation entre x_1 et ξ_1 est donnée par la transformation $F_\kappa(\xi_1) = x_1$ avec $x_1 = h_\kappa \xi_1 / 2 + x_{1,\kappa}$. De plus, l'orthogonalité de la base \mathcal{B}_κ^p est donnée dans l'élément de référence I vis-à-vis du produit intérieur suivant :

$$\int_I \mathbb{L}^{k-1}(\xi_1) \mathbb{L}^{k'-1}(\xi_1) d\xi_1 = \frac{2}{2k-1} \delta_{kk'},$$

où $\delta_{kk'}$ est le symbole de Kronecker.

2.4.2.2 Base de Dubiner en deux dimensions d'espace : éléments droits

L'idée clé de la construction d'une base orthogonale dans le cas de maillages à éléments droits en deux dimensions d'espace, repose sur l'utilisation du produit tensoriel de polynômes connus et définis sur un intervalle de l'axe des réels dans

chaque direction d'espace. Si cette idée est facilement envisageable pour des éléments de type rectangle dont la surface peut être décrite indépendamment des deux directions d'espace, elle l'est moins pour des éléments de type triangle.

Néanmoins, une base orthogonale et hiérarchique sur des triangles a été développée par plusieurs auteurs comme Dubiner [42] dans le cadre des méthodes spectrales (*cf.* [153, chapitre 3] pour un historique). Elle est construite par le produit tensoriel des polynômes de Jacobi définis sur le segment de référence I .

Soit $\mathbb{J}_{\alpha,\beta}^p$ le polynôme de Jacobi unidimensionnel, de degré p et d'ordres α, β , et on rappelle que $\mathbb{J}_{0,0}^p = \mathbb{L}^p$ [2, chapitre 22]. Il vient :

$$\phi^{k(i,j)}(\mathbf{x}) = \mathbb{J}_{0,0}^j(\xi_2)(1 + \xi_1)^j \mathbb{J}_{0,2j+1}^i(\xi_1), \quad \forall \mathbf{x} \in \kappa, \quad \kappa \in \Omega_h, \quad 1 \leq k(i,j) \leq N_p,$$

où la relation entre \mathbf{x} et $\boldsymbol{\xi}$ est donnée par la transformation \mathbf{F}_κ définie en (2.30) et (2.33), et, avec le choix d'ordonnancement des degrés de liberté donné par :

$$k(i,j) = 1 + (i+1)j + \frac{i+1}{2}i + \frac{j+1}{2}j, \quad 0 \leq i+j \leq p. \quad (2.34)$$

La figure 2.6 illustre, pour $p = 4$, le rangement des fonctions de base $\phi^{k(i,j)}$ par valeurs de k croissantes en fonction des indices i et j associés aux directions d'espace respectivement, ξ_1 et ξ_2 . En reconsidérant l'écriture précédente comme :

$$\phi^{k(i,j)}(\mathbf{x}) = \mathbb{P}^j(\xi_2)\mathbb{P}^{j,i}(\xi_1), \quad \forall \mathbf{x} \in \kappa, \quad \kappa \in \Omega_h, \quad 1 \leq k(i,j) \leq N_p,$$

avec $\mathbb{P}^j(\xi_2) = \mathbb{J}_{0,0}^j(\xi_2)$ et $\mathbb{P}^{j,i}(\xi_1) = (1 + \xi_1)^j \mathbb{J}_{0,2j+1}^i(\xi_1)$, on remarque d'une part, que le polynôme $\phi^{k(i,j)}$ est de degré j en ξ_2 et i en ξ_1 grâce au facteur $(1 + \xi_1)^j$, et, d'autre part, qu'il constitue un produit tensoriel généralisé dans I^2 . De plus, l'orthogonalité de la base \mathcal{B}_κ^p est donnée dans l'élément de référence I^2 vis-à-vis du produit intérieur suivant :

$$\int_{I^2} \mathbb{J}_{0,0}^j(\xi_2)\mathbb{J}_{0,0}^{j'}(\xi_2)(1 + \xi_1)^{j+j'} \mathbb{J}_{0,2j+1}^i(\xi_1)\mathbb{J}_{0,2j'+1}^{i'}(\xi_1)d\xi_1d\xi_2 = \frac{4^j}{(2j+1)(i+j+1)} \delta_{ii'}\delta_{jj'},$$

où $\delta_{ii'}$ (ou $\delta_{jj'}$) est le symbole de Kronecker.

2.4.2.3 Base numérique en deux dimensions d'espace : éléments courbes

Dans le cas de maillages à éléments courbes, l'approche précédente par produit tensoriel ne permet pas d'obtenir une base orthogonale dans l'élément de référence. En particulier, le calcul du jacobien de la transformation \mathbf{F}_κ rend difficile cette orthogonalisation.

Néanmoins, une base orthonormale et hiérarchique pour éléments courbes est proposée par Remacle *et al.* [134] dans le cadre de la méthode GD, elle est détaillée dans [38, annexe A]. Récemment, Bassi *et al.* [6] l'ont utilisée avec le schéma BR2 en proposant une quadrature simplifiée. Cette base est construite directement dans le triangle physique à partir d'une base de monômes définis en coordonnées barycentriques (pour un meilleur conditionnement) et orthonormalisés par une procédure de Gram-Schmidt modifiée.

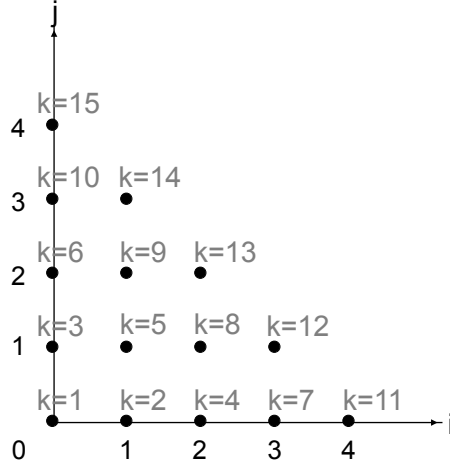


Fig. 2.6 – Illustration du rangement des fonctions de base $\phi^{k(i,j)}$ pour $p = 4$ et $d = 2$.

Soit $\mathbf{x}_B = (x_{1,B}, x_{2,B})^\top$ le barycentre d'un élément κ et soit $m^{i,j}$ le monôme de degré $i + j$ défini par :

$$m^{ij}(\mathbf{x}) = (x_1 - x_{1,B})^i (x_2 - x_{2,B})^j, \quad \forall \mathbf{x} \in \kappa, \quad \kappa \in \Omega_h, \quad i, j \geq 0.$$

Ces monômes servent à l'initialisation d'une base de départ, notée $\check{\mathcal{B}}_\kappa^p = (\check{\phi}^k)_{1 \leq k \leq N_p}$, pour la procédure d'orthonormalisation. Un choix possible est [6] :

$$\check{\phi}^{k(i,j)} = \frac{m_{ij}(\mathbf{x})}{\|m_{ij}(\mathbf{x})\|_{L^2(\kappa)}}, \quad \forall \mathbf{x} \in \kappa, \quad \kappa \in \Omega_h, \quad 1 \leq k(i,j) \leq N_p,$$

avec $k(i,j)$ donné par la relation (2.34). Les fonctions de la base finale \mathcal{B}_κ^k sont telles que :

$$\phi^k(\mathbf{x}) = \frac{1}{\alpha_{kk}} \check{\phi}^k(\mathbf{x}) - \sum_{j=1}^{k-1} \frac{\alpha_{kj}}{\alpha_{kk}} \phi^j(\mathbf{x}), \quad \forall \mathbf{x} \in \kappa, \quad \kappa \in \Omega_h, \quad 1 \leq k \leq N_p,$$

où α_{kk} et α_{kj} sont les coefficients issus du procédé de Gram-Schmidt modifié et tels que :

$$\alpha_{kk} = \sqrt{(\check{\phi}^k, \check{\phi}^k)_{L^2(\kappa)} - \sum_{j=1}^{k-1} (\check{\phi}^k, \phi^j)_{L^2(\kappa)}^2},$$

$$\alpha_{kj} = (\check{\phi}^k, \phi^j)_{L^2(\kappa)}.$$

De plus, l'orthogonalité de la base \mathcal{B}_κ^p est donnée dans l'élément physique κ vis-à-vis du produit intérieur suivant :

$$\int_{\kappa} \phi^k(\mathbf{x}) \phi^{k'}(\mathbf{x}) dx_1 dx_2 = |\kappa| \delta_{kk'},$$

où $\delta_{kk'}$ est le symbole de Kronecker.

Remarque 2.17. *Le produit intérieur précédent étant au mieux égal à la précision sur l'approximation de l'intégrale, la base est dite numérique, et ce, contrairement à celle de Legendre ou de Dubiner dont les intégrales sont évaluées analytiquement.*

2.4.3 Intégration numérique

Toutes les intégrales de volume (respectivement de surface) intervenant dans le schéma numérique décrit par (2.14) sont approchées au moyen d'une méthode d'intégration numérique (ou quadrature) dans l'élément de référence I^d (respectivement I^{d-1}). Par exemple, pour l'intégrale de volume d'une fonction f sur κ , il vient :

$$\int_{\kappa} f(\mathbf{x}) d\mathbf{x} = \int_{I^d} (f \circ \mathbf{F}_{\kappa})(\boldsymbol{\xi}) J_{\mathbf{F}_{\kappa}}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \quad (2.35)$$

où $J_{\mathbf{F}_{\kappa}}(\boldsymbol{\xi}) = |\partial\mathbf{x}/\partial\boldsymbol{\xi}|$ est le jacobien de la transformation \mathbf{F}_{κ} définie en (2.30) et (2.33).

Parmi les quadratures existantes, celle de Gauss a été choisie, entre autres car elle est exacte pour un polynôme f de degré $2p - 1$ au plus, avec p points pris dans chaque direction d'espace sur le domaine d'intégration [153, chapitre 2]. Ainsi, en approchant l'intégrale (2.35) par la quadrature de Gauss, il vient :

$$\int_{\kappa} f(\mathbf{x}) d\mathbf{x} \approx \sum_{i=1}^{N_g} \omega_{G,i} (f \circ \mathbf{F}_{\kappa})(\boldsymbol{\xi}_{G,i}) J_{\mathbf{F}_{\kappa}}(\boldsymbol{\xi}_{G,i}) d\boldsymbol{\xi},$$

où $\omega_{G,i}$ et $\boldsymbol{\xi}_{G,i}$ représentent respectivement les poids et les points de la quadrature (cf. figure 2.4). En raison de l'intervalle d'intégration $[-1, 1]$ dans chaque direction d'espace, l'ensemble des points $(\boldsymbol{\xi}_{G,i})_{1 \leq i \leq N_g}$ correspond à tous les couples possibles du produit tensoriel des racines du polynôme de Legendre $(\mathbb{L}^i(\xi_1))_{0 \leq i \leq p}$. Ces points sont donc tous contenus à l'intérieur de l'élément de référence I^d . Le lecteur peut consulter [130, chapitre 4] pour leur détermination. Dans la pratique, pour une approximation GD de degré de polynôme p , $p + 1$ points sont pris dans chaque direction d'espace pour intégrer exactement des polynômes de degré $2p + 1$. A titre d'exemple, le tableau 2.2 donne les valeurs de N_g pour $d = 1$ et $d = 2$.

En supposant un élément droit, une quadrature d'ordre $2p$ (respectivement $2p + 1$) suffit pour calculer les intégrales de volume (respectivement de surface) dans l'approximation interne (2.14) [25]. En revanche, pour un élément courbe, une telle quadrature ne suffit plus car la transformation \mathbf{F}_{κ} n'est plus linéaire. Cependant, les expériences numériques du chapitre 5 (en particulier § 5.1 page 101), et d'autres travaux [6] montrent que les quadratures utilisées suffisent pour obtenir les ordres de convergence attendus de la méthode GD.

| | | | | | | | | |
|-------|-----------|---|---|---|----|----|----|----|
| p | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| N_g | $(d = 1)$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | $(d = 2)$ | 1 | 4 | 9 | 16 | 25 | 36 | 49 |

Tab. 2.2 – Nombre de points d'intégration utilisés pour les intégrales de surface $d = 1$ ou de volume $d = 2$ en fonction de p .

2.5 Conclusion partielle

Les points suivants constituent le bilan de ce chapitre.

- Le schéma numérique semi-discret en espace du système des équations d'Euler ou de Navier-Stokes a été établi au moyen d'une discrétisation en espace de type GD compacte.
- Les points clés de cette discrétisation GD, qui considère une solution numérique discontinue aux interfaces entre les éléments du maillage, ont été mis en avant :
 - un espace variationnel GD pour chaque élément du maillage ;
 - des flux numériques consistants et conservatifs pour remplacer les flux physiques qui ne sont pas définis aux interfaces.
- Le choix des flux numériques utilisés pour obtenir un schéma numérique compact a été expliqué, en insistant sur la difficulté de traiter les termes de diffusion (voir également annexe B). A savoir :
 - le flux décentré amont LLF pour les termes de convection [156] ;
 - le flux (ou schéma) BR2 pour les termes de diffusion [13, 7], il fait intervenir deux opérateurs de relèvement, lesquels sont une correction du gradient de la solution numérique pour tenir compte de ses discontinuités aux interfaces.
- La discrétisation repose sur les maillages simpliciaux droits ou courbes.
- La base de fonctions retenue présente trois caractéristiques motivées par la discrétisation temporelle explicite-implicite, elle est :
 - modale ;
 - orthonormale ;
 - hiérarchique.
- Certains aspects pratiques ont été évoqués pour la mise en œuvre de la discrétisation en espace, soit :
 - la transformation d'un élément physique vers un élément de référence avec prise en compte des éléments à parois courbes ;
 - les fonctions de base qui engendrent l'espace d'approximation GD dans le cas d'éléments droits (bases de Legendre [2, chapitres 8 et 22], de Dubiner [42]) ou courbes (base numérique [134], [38, annexe A]) ;
 - la méthode d'intégration numérique de Gauss pour approcher toutes les intégrales du schéma semi-discret en espace.

A présent, il s'agit d'établir le schéma numérique discret complet du système des équations d'Euler ou de Navier-Stokes en tenant compte de la dérivée en temps du schéma semi-discret en espace. C'est l'objet du prochain chapitre.

Chapitre 3

Discrétisations temporelles explicites-implicites à coût réduit

Ce chapitre décrit la discrétisation temporelle du système des équations de Navier-Stokes, déjà discrétisé en espace au chapitre 2. Une méthode de décomposition d'opérateurs est utilisée, où les termes de convection sont résolus explicitement et ceux de diffusion implicitement. Trois simplifications du schéma implicite sont proposées et reposent sur une approximation de la matrice implicite et/ou sur un découplage des degrés de liberté. La dernière partie du chapitre dresse un bilan comparatif des intégrations temporelles.

3.1 Avant-propos

L'ensemble des choix concernant l'intégration temporelle explicite-implicite simplifiée proposée fait suite à plusieurs études sur des équations d'advection-diffusion scalaires linéaires et non linéaires en deux dimensions d'espace [139, 53, 138], comme expliqué dans l'introduction générale de ce mémoire (*cf.* § 0.4.1 page 15). En particulier ces choix sont en lien avec les trois motivations principales suivantes :

1. comme la restriction la plus importante sur le pas de temps provient des termes de diffusion, la résolution de ces derniers est réalisée implicitement en temps tandis que celle des termes de convection est réalisée explicitement afin d'éviter trop de dissipation sur les phénomènes de transport ;
2. pour utiliser une méthode adaptée aux problèmes de convection dominante, les résolutions des termes de convection et de diffusion, donc des opérateurs hyperbolique et parabolique, sont réalisées séparément ;
3. comme la résolution d'un schéma implicite requiert l'inversion d'un système linéaire d'autant plus grand que le maillage est raffiné ou que la solution recherchée est précise, des simplifications visant à réduire la taille du système sont proposées.

La première partie de ce chapitre, paragraphe 3.2, revient sur les conditions de stabilité issues de l'analyse de Von Neumann. La seconde partie, paragraphe 3.3, est réservée à la discrétisation temporelle explicite du système semi-discret (2.14) utilisée pour certaines validations. Les trois dernières parties, paragraphes 3.4, 3.5 et 3.6, sont consacrées respectivement, à la mise en place de la discrétisation explicite-implicite, aux simplifications envisagées du schéma implicite et à l'analyse de ces dernières sur les aspects d'encombrement mémoire et de nombre d'opérations à virgule flottante.

3.2 Prérequis sur les conditions de stabilité

Ce paragraphe revient sur les deux conditions de stabilité de type Courant-Friedrichs-Lewy (ou CFL) et de type Von Neumann imposant une forte restriction sur le pas de temps lors de l'utilisation d'une discrétisation temporelle explicite associée à une discrétisation spatiale GD. L'objectif est de mesurer le poids de ces restrictions à l'origine des intégrations temporelles explicites-implicites envisagées dans la suite de ce travail.

La détermination des conditions de stabilité repose sur l'analyse de stabilité de Von Neumann [65, chapitre 7]. L'application rigoureuse de cette analyse pour le système des équations de Navier-Stokes est complexe, elle est rarement faite, il lui est substitué celle sur l'équation d'advection-diffusion monodimensionnelle scalaire et instationnaire (*cf.* par exemple [65, chapitre 8] ou [89] et leurs références) :

$$\frac{\partial w}{\partial t} + c \frac{\partial w}{\partial x_1} - \nu \frac{\partial^2 w}{\partial x_1^2} = 0, \quad \forall (x_1, t) \in \Omega \times \mathcal{T}, \quad (3.1a)$$

$$w(x_1, 0) = w^{(0)}(x_1), \quad \forall x_1 \in \Omega, \quad (3.1b)$$

complétée par des conditions de périodicité aux bords. Ici, $\Omega = [0, 1]$ est le segment unité discrétisé uniformément en N segments réguliers de longueur $h = 1/N$. En outre, l'équation (3.1a) représente le transport de la quantité scalaire w dans un écoulement à la vitesse uniforme $c \geq 0$ (ou coefficient de convection) et de diffusion $\nu \geq 0$ (ou coefficient de viscosité), et l'équation (3.1b) sa condition initiale.

Lors de l'utilisation d'un schéma GD en espace associé à un schéma explicite en temps, les conditions de stabilité à respecter pour l'équation d'advection pure, soit $\nu = 0$, (ou condition de CFL) et pour l'équation de diffusion pure, soit $c = 0$ (ou condition de Von Neumann), sont respectivement :

$$\frac{c\Delta t_c}{h} \lesssim \frac{1}{2p+1}, \quad (3.2a)$$

$$\frac{\nu\Delta t_v}{h^2} \lesssim \frac{1}{(2p+1)^2}, \quad (3.2b)$$

où $\Delta t_c > 0$ et $\Delta t_v > 0$ représentent les pas de temps des schémas numériques discrets de l'équation respectivement, d'advection pure et de diffusion pure. En particulier, la relation (3.2a) a été établie par Cockburn et Shu [28] dans le cas d'un schéma en temps de type Runge-Kutta d'ordre s tel que $p = s - 1$ (*cf.* § 3.3 pour plus de détails sur ces schémas). Elle est exacte pour $p = 0$ et $p = 1$, et empirique pour $p \geq 2$ avec un écart inférieur à 5% observé dans les expériences numériques [28]. Par ailleurs, la relation (3.2b) est établie par Lörcher *et al.* [108, 51] à partir d'expérimentations numériques.

In fine, le pas de temps $\Delta t > 0$ pour l'équation complète d'advection-diffusion (3.1a) est soumis à la condition de stabilité suivante :

$$\begin{aligned} \Delta t &\leq \min(\Delta t_c, \Delta t_v), \\ &\leq \frac{h}{c(2p+1)} \min\left(1, \frac{Re_h}{2p+1}\right), \end{aligned} \quad (3.3)$$

où le nombre de Reynolds de maille Re_h défini par :

$$Re_h = \frac{ch}{\nu}$$

est introduit pour faciliter la distinction des régimes à convection ou diffusion dominante. Ce nombre quantifie l'amplitude des phénomènes de convection par rapport à celle des phénomènes de diffusion à l'échelle des longueurs basée sur la taille caractéristique du maillage h . La condition de stabilité (3.3) s'interprète comme suit :

- lorsque $Re_h/(2p + 1) \gg 1$, c'est-à-dire lorsque Re_h (ou h) est très grand ou p est petit, le pas de temps est limité par l'échelle temporelle d'évolution des phénomènes de convection et la condition de stabilité de type CFL se comporte en loi de décroissance linéaire en p (cf. relation (3.2a)) ;
- *a contrario*, lorsque $Re_h/(2p + 1) \ll 1$, c'est-à-dire lorsque Re_h (ou h) est très petit ou p est grand, le pas de temps est limité par l'échelle temporelle d'évolution des phénomènes de diffusion et la condition de stabilité de type Von Neumann se comporte en loi de décroissance quadratique en p (cf. relation (3.2b)).

En général, lors de la simulation des écoulements à convection dominante comme ceux étudiés dans ce travail, la capture correcte du décollement de la couche limite en zone de proche paroi impose de raffiner le maillage à cet endroit et conduit localement à $Re_h \ll 1$. La valeur du pas de temps est alors fixée par celle du pas de temps visqueux ($\Delta t \approx \Delta t_v$), elle devient très petite, voire rédhibitoire pour des applications réelles. C'est pourquoi, on envisage une discrétisation implicite des termes de diffusion pour s'affranchir de la condition très contraignante de type Von Neumann, tout en gardant une discrétisation explicite des termes de convection pour restituer correctement les mécanismes de transport. Cette approche est relativement « classique » dans la littérature (cf. par exemple [152] dans le cadre des méthodes GD).

En revenant au problème initial (1.1) de la résolution du système des équations de Navier-Stokes, le pas de temps est défini par les relations suivantes en fonction de l'intégration temporelle utilisée :

- dans le cas d'une discrétisation temporelle explicite (cf. § 3.3 pour plus de détails sur cette discrétisation), il vient de façon similaire à (3.3) :

$$\Delta t = CFL \times \min \left\{ \left(\frac{h_\kappa}{\|\mathbf{U}\| + c}, \frac{\rho h_\kappa^2}{2\gamma \frac{\mu}{Pr}} \right); \kappa \in \Omega_h \right\}; \quad (3.4)$$

- et dans le cas d'une discrétisation temporelle explicite-implicite (cf. § 3.4 pour plus de détails sur cette discrétisation), il vient :

$$\Delta t = CFL \times \min \left\{ \frac{h_\kappa}{\|\mathbf{U}\| + c}; \kappa \in \Omega_h \right\}; \quad (3.5)$$

où CFL désigne la valeur de la condition de stabilité de type CFL (ou condition de CFL dans la suite) du schéma numérique.

Enfin pour la suite, les notations suivantes sont considérées. Soit $n \geq 0$ un entier représentant le nombre d'itérations en temps du schéma numérique et soit $\Delta t > 0$ le pas de temps entre les instants $t^{(n)}$ et $t^{(n+1)}$, avec $t^{(n)} = n\Delta t$. La solution numérique semi-discrète au temps $t^{(n)}$ du schéma (2.14) est $\mathbf{w}_h^{(n)}(\cdot) = \mathbf{w}_h(\cdot, t^{(n)})$.

3.3 Discrétisation explicite avec un schéma Runge-Kutta

Dans ce travail, une discrétisation temporelle explicite utilisant un schéma de Runge-Kutta (ou RK) sert à valider la discrétisation spatiale GD décrite au chapitre 2. Ce paragraphe donne l'expression de ce schéma.

De façon générale, les schémas RK constituent une famille importante d'intégration temporelle explicite en temps. De fait, leurs propriétés ont été largement étudiées. Par exemple, le lecteur peut consulter [65, chapitre 9] et ses références, ainsi que [28] et [38, chapitre 3] pour des applications dans le cadre des méthodes GD. Il ressort de ces ouvrages un intérêt d'utilisation de ces schémas pour leur précision, leur stabilité et leur simplicité.

Le principe de ces derniers est le suivant. Soit $s > 0$ appartenant à \mathbb{N} , les schémas RK consistent à évaluer \mathcal{L}_c et \mathcal{L}_v en s différentes valeurs de $\mathbf{w}_h^{(n)}$ notées $(\mathbf{w}_h^{(n,i)})_{1 \leq i \leq s}$ et comprises dans l'intervalle de temps générique $[t^{(n)}, t^{(n+1)}]$ pour obtenir une approximation d'ordre élevé de $\mathbf{w}_h^{(n+1)}$. Soit une écriture du schéma RK s , dit schéma « à s pas », suivante pour le schéma numérique semi-discret (2.14) :

$$\mathbf{w}_h^{(n,0)} = \mathbf{w}_h^{(n)}, \quad (3.6a)$$

$$\mathbf{w}_h^{(n,i)} = \sum_{j=0}^{i-1} \left(a_{ij} \mathbf{w}_h^{(n,j)} - b_{ij} \Delta t \left(\mathcal{L}_c(\mathbf{w}_h^{(n,j)}, \phi^k) + \mathcal{L}_v(\mathbf{w}_h^{(n,j)}, \phi^k) \right) \right) \quad (3.6b)$$

$$\mathbf{w}_h^{(n+1)} = \mathbf{w}_h^{(n,s)}, \quad (3.6c)$$

où $1 \leq i \leq s$, et où $(a_{ij})_{1 \leq i \leq s, 0 \leq j \leq i-1}$ et $(b_{ij})_{1 \leq i \leq s, 0 \leq j \leq i-1}$ sont les matrices triangulaires inférieures permettant de déterminer entièrement le schéma telles que :

$$a_{ij} \geq 0 \text{ avec } \sum_{j=0}^{i-1} a_{ij} = 1 \text{ et } \forall 1 \leq i \leq s,$$

$$b_{ij} = 0 \text{ pour } j < i - 1 \text{ et } \forall 1 \leq i \leq s.$$

L'écriture d'un schéma RK sous cette forme, en particulier le choix des coefficients de la matrice \mathbf{b} , permet de réduire les coûts mémoire en ne stockant que les informations sur les opérateurs \mathcal{L}_c et \mathcal{L}_v relatives au sous-pas $i - 1$ [65].

Les coefficients des matrices \mathbf{a} et \mathbf{b} sont choisis pour obtenir le meilleur ordre possible et la restriction de stabilité la moins contraignante [28]. Pour les applications numériques de la partie II, les deux schémas RK s d'ordre s suivants ont été retenus :

- le schéma RK3 de Shu-Osher[155], il est utilisé pour la discrétisation explicite portant sur les termes hyperboliques seuls (opérateur \mathcal{L}_c) lors de la résolution du système des équations d'Euler, c'est-à-dire uniquement au paragraphe 5.1 page 101 :

$$\mathbf{a} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}; \quad (3.7)$$

- le schéma RK2 de Heun [63, 52], il est utilisé, soit pour la discrétisation explicite et portant sur les termes hyperboliques et paraboliques (opérateurs \mathcal{L}_c et

\mathcal{L}_v) lors de la résolution du système des équations de Navier-Stokes, c'est-à-dire principalement au paragraphe 5.2 page 115, soit pour le schéma explicite de la discrétisation explicite-implicite comme expliqué au paragraphe 3.4 suivant :

$$\mathbf{a} = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \quad (3.8)$$

Définition 3.1. *Dans la suite de ce travail, la résolution du système des équations de Navier-Stokes avec une discrétisation spatiale GD (relations (2.14), (2.20) et (2.29)) et une discrétisation temporelle explicite RK2 (relations (3.6) et (3.8)) ou RK3 (relations (3.6) et (3.7)) est désignée par l'appellation « EXPL ».*

3.4 Discrétisation explicite-implicite avec une décomposition d'opérateurs

Souvent considérée comme une stratégie de « diviser pour mieux régner », la méthode de décomposition d'opérateurs (dite à pas fractionnaires, ou de projection) est utilisée ici dans le cadre du système des équations de Navier-Stokes pour séparer la résolution des phénomènes physiques de convection (opérateur \mathcal{L}_c) et de diffusion (opérateur \mathcal{L}_v). Cette séparation est particulièrement bien adaptée aux problèmes de convection dominante ($Re_\infty \gg 1$) considérés dans ce travail, elle permet d'utiliser des méthodes numériques dédiées pour chacun des opérateurs (par exemple explicite pour l'un et implicite pour l'autre). Le lecteur peut se référer à [66] et [111, chapitre 3] (et leurs références) pour une présentation théorique et numérique récente de la méthode dans le cas respectivement, des équations de type hyperbolique-parabolique non linéaires et du système des équations de Navier-Stokes.

Ainsi, la décomposition d'opérateurs retenue pour le schéma semi-discret est d'abord mise en place (§ 3.4.1), puis les méthodes numériques associées à chaque opérateur sont détaillées (§ 3.4.2), et enfin les coûts en temps CPU du schéma discret obtenu sont étudiés pour mieux cerner les points d'améliorations envisagés (§ 3.4.3).

3.4.1 Décomposition de Strang

Le schéma numérique semi-discret (2.14) est discrétisé en temps *via* la méthode de décomposition d'opérateurs de Strang [160] afin de résoudre séparément les intégrations des termes de convection et de diffusion. Cette méthode est précise à l'ordre deux grâce à sa formulation qui utilise symétriquement les opérateurs \mathcal{L}_c et \mathcal{L}_v sur des sous-intervalles réguliers du pas de temps. En GD, elle est par exemple utilisée dans [184] sur un système d'équations d'advection-réaction ou dans [61] sur un système des équations de Navier-Stokes incompressibles.

La figure 3.1 est une schématisation de la décomposition d'opérateurs de Strang. La solution numérique semi-discrete $\mathbf{w}_h^{(n)}$ est avancée en temps en trois étapes sur trois sous-intervalles de l'intervalle $[t^{(n)}, t^{(n+1)}]$ comme suit.

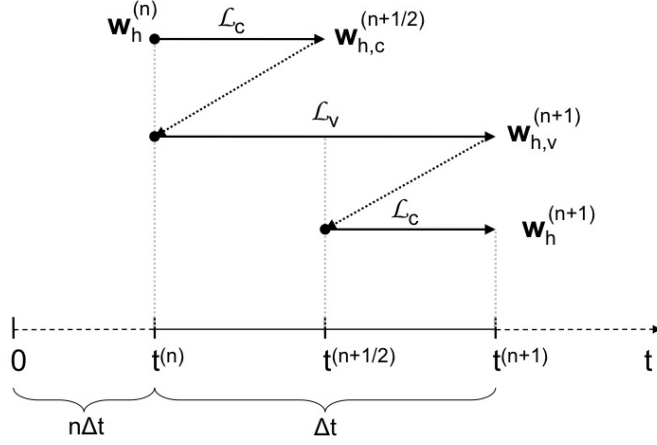


Fig. 3.1 – Schématisation de la décomposition d’opérateurs de Strang [160] sur l’intervalle de temps $[t^{(n)}, t^{(n+1)}]$. Résolution du problème hyperbolique sur $[t^{(n)}, t^{(n+1/2)}]$, puis du problème parabolique sur $[t^{(n)}, t^{(n+1)}]$, et à nouveau du problème hyperbolique sur $[t^{(n+1/2)}, t^{(n+1)}]$.

1. Le problème hyperbolique est résolu pour $\mathbf{w}_{h,c}$ sur l’intervalle $[t^{(n)}, t^{(n+1/2)}]$:

$$\int_{\Omega_h} \phi^k \partial_t \mathbf{w}_{h,c} dV + \mathcal{L}_c(\mathbf{w}_{h,c}, \phi^k) = 0, \quad \forall \phi^k \in \mathcal{B}_h^p, \quad \forall t \in (t^{(n)}, t^{(n+\frac{1}{2})}], \quad (3.9a)$$

$$\mathbf{w}_{h,c}(\cdot, t^{(n)}) = \mathbf{w}_h(\cdot, t^{(n)}). \quad (3.9b)$$

2. Le problème parabolique est résolu pour $\mathbf{w}_{h,v}$ sur l’intervalle $[t^{(n)}, t^{(n+1)}]$:

$$\int_{\Omega_h} \phi^k \partial_t \mathbf{w}_{h,v} dV + \mathcal{L}_v(\mathbf{w}_{h,v}, \phi^k) = 0, \quad \forall \phi^k \in \mathcal{B}_h^p, \quad \forall t \in (t^{(n)}, t^{(n+1)}], \quad (3.10a)$$

$$\mathbf{w}_{h,v}(\cdot, t^{(n)}) = \mathbf{w}_{h,c}(\cdot, t^{(n+\frac{1}{2})}). \quad (3.10b)$$

3. Le problème hyperbolique est résolu pour $\mathbf{w}_{h,c}$ sur l’intervalle $[t^{(n+1/2)}, t^{(n+1)}]$:

$$\int_{\Omega_h} \phi^k \partial_t \mathbf{w}_{h,c} dV + \mathcal{L}_c(\mathbf{w}_{h,c}, \phi^k) = 0, \quad \forall \phi^k \in \mathcal{B}_h^p, \quad \forall t \in (t^{(n+\frac{1}{2})}, t^{(n+1)}], \quad (3.11a)$$

$$\mathbf{w}_{h,c}(\cdot, t^{(n+\frac{1}{2})}) = \mathbf{w}_{h,v}(\cdot, t^{(n+1)}). \quad (3.11b)$$

La solution au temps $t^{(n+1)}$ est donc $\mathbf{w}_h^{(n+1)}(\cdot) = \mathbf{w}_{h,c}(\cdot, t^{(n+1)})$.

La solution numérique peut s’écrire de façon formelle comme la composition des opérateurs \mathcal{L}_c et \mathcal{L}_v (non commutatifs) suivante sur un pas de temps :

$$\mathbf{w}_h^{(n+1)} \approx e^{\frac{\Delta t}{2} \mathcal{L}_c} e^{\Delta t \mathcal{L}_v} e^{\frac{\Delta t}{2} \mathcal{L}_c} \mathbf{w}_h^{(n)}.$$

Il est facile de s’assurer que l’erreur de la méthode de Strang est bien globalement d’ordre deux en temps dans des cas simples, comme celui où les deux opérateurs sont des matrices. En effet, il suffit d’utiliser le développement limité de la fonction exponentielle à l’ordre deux en Δt autour de zéro dans l’équation précédente. La démonstration est faite, par exemple, dans [160] ou [56, chapitre IV]. Pour les cas plus complexe d’opérateurs non bornés ou non linéaires, le lecteur peut se référer à [66, chapitre 6] et [111, chapitre 3].

Remarque 3.1. Du point de vue du formalisme, il est bien entendu possible d'inverser la séquence des opérateurs \mathcal{L}_c et \mathcal{L}_v : $\mathbf{w}_h^{(n+1)} \approx e^{\frac{\Delta t}{2}\mathcal{L}_v} e^{\Delta t\mathcal{L}_c} e^{\frac{\Delta t}{2}\mathcal{L}_v} \mathbf{w}_h^{(n)}$, mais d'après [159] et [66, chapitre 4] et des tests numériques, la décomposition se terminant par l'opérateur hyperbolique \mathcal{L}_c reste la plus efficace. Elle maintient en particulier la stabilité du schéma. En outre, la résolution de l'opérateur \mathcal{L}_c est moins coûteuse que celle de l'opérateur \mathcal{L}_v .

3.4.2 Choix du schéma explicite et du schéma implicite

A présent il reste à choisir les schémas numériques d'intégration temporelle pour obtenir le schéma discret final. Les méthodes possibles sont nombreuses et fonctionnent dans des classes d'écoulements. A titre d'exemple, Turek [165] en compare plusieurs, théoriquement et numériquement, pour le système des équations de Navier-Stokes.

Dans ce travail, les choix sont guidés par deux idées essentielles, en plus de celle d'explicitement les termes de convection et d'impliciter ceux de diffusion comme mentionné dans l'introduction de ce chapitre. Premièrement, il faut que l'ordre de la méthode de Strang soit maintenu, ce qui est le cas si chaque résolution des systèmes (3.9), (3.10) et (3.11) se fait également avec une méthode numérique d'ordre deux minimum. Deuxièmement, il faut que le schéma global reste stable, ce qui est *a priori* le cas si ces mêmes méthodes numériques le sont aussi.

Les deux paragraphes suivants présentent les schémas numériques retenus, à savoir le schéma explicite de Heun [63, 52] pour les termes de convection (§ 3.4.2.1) et celui de Crank-Nicolson [34, 36] pour les termes de diffusion (§ 3.4.2.2). Cette combinaison des schémas temporels a été utilisée avec succès avec une méthode GD sur le système des équations de Navier-Stokes dans [61].

Définition 3.2. Pour faciliter la lecture de la suite de ce travail, le nombre entier N_{ddl} d'inconnues du problème est introduit, soit $N_{ddl} = NN_{eq}N_p$.

3.4.2.1 Schéma explicite retenu pour le problème hyperbolique

Les problèmes hyperboliques (3.9) et (3.11) sont discrétisés avec le schéma explicite de Heun qui est précis au second ordre en temps et optimal en restriction de type CFL [63, 52]. L'écriture du schéma de Heun est celle d'un schéma RK à deux pas et a déjà été donnée par les relations (3.6) et (3.8) pour trouver \mathbf{w}_h à l'instant $t^{(n+1)}$. Ici, elle est appliquée pour trouver $\mathbf{w}_{h,c}$ aux instants $t^{(n+1/2)}$ et $t^{(n+1)}$.

Pour le problème (3.9) par exemple, l'approximation de la dérivée temporelle s'écrit : Trouver $\mathbf{w}_{h,c}^{(n+1/2)}$ appartenant à $[\mathcal{V}_h^p]^{d+2}$, telle que pour tout ϕ^k élément de \mathcal{B}_h^p , il vienne à chaque instant $t^{(n)}$

$$\int_{\Omega_h} \phi^k \frac{\mathbf{w}_{h,c}^{(*)} - \mathbf{w}_h^{(n)}}{\Delta t/2} dV = -\mathcal{L}_c(\mathbf{w}_h^{(n)}, \phi^k), \quad (3.12a)$$

$$\int_{\Omega_h} \phi^k \frac{\mathbf{w}_{h,c}^{(n+1/2)} - \mathbf{w}_h^{(n)}}{\Delta t/2} dV = -\frac{1}{2} \left(\mathcal{L}_c(\mathbf{w}_h^{(n)}, \phi^k) + \mathcal{L}_c(\mathbf{w}_{h,c}^{(*)}, \phi^k) \right), \quad (3.12b)$$

où $\mathbf{w}_{h,c}^{(*)}$ est une prédiction de la valeur de la solution $\mathbf{w}_{h,c}$ au temps $t^{(n+1/2)}$ (cf. remarque 3.2).

Remarque 3.2. La méthode de Heun peut être vue comme une méthode de type prédicteur-correcteur. L'équation (3.12a) représente le schéma prédicteur qui approche la dérivée temporelle par la méthode d'Euler tandis que l'équation (3.12b) représente le schéma correcteur qui approche la dérivée temporelle par la méthode des trapèzes. C'est pourquoi, la méthode de Heun est parfois dite d'Euler modifiée (ou méthode des trapèzes).

Le problème discret (3.9) prend alors la forme du système linéaire suivant :

$$\mathbf{W}_c^{(n+\frac{1}{2})} = \frac{1}{2} \left(\mathbf{W}_c^{(*)} + \mathbf{W}^{(n)} - \frac{\Delta t}{2} \mathbf{M}^{-1} \mathbf{L}_c(\mathbf{W}_c^{(*)}) \right), \quad (3.13)$$

avec :

$$\mathbf{W}_c^{(*)} = \mathbf{W}^{(n)} - \frac{\Delta t}{2} \mathbf{M}^{-1} \mathbf{L}_c(\mathbf{W}^{(n)}),$$

et de façon identique, le second problème discret (3.11) prend la forme :

$$\mathbf{W}_c^{(n+1)} = \frac{1}{2} \left(\mathbf{W}_c^{(*)} + \mathbf{W}_v^{(n+1)} - \frac{\Delta t}{2} \mathbf{M}^{-1} \mathbf{L}_c(\mathbf{W}_c^{(*)}) \right), \quad (3.14)$$

avec

$$\mathbf{W}_c^{(*)} = \mathbf{W}_v^{(n+1)} - \frac{\Delta t}{2} \mathbf{M}^{-1} \mathbf{L}_c(\mathbf{W}_v^{(n+1)}),$$

où les notations ci-après sont utilisées.

- $\mathbf{W}^{(n)}$ appartient à $\mathbb{R}^{N_{adi}}$, il représente le vecteur complet des degrés de liberté sur Ω_h au temps $t^{(n)}$ défini en chaque instant par (2.10).
- \mathbf{M} appartient à $\mathbb{R}^{N_{adi} \times N_{adi}}$, elle est la matrice de masse définie par le produit intérieur des fonctions de la base \mathcal{B}_h^p défini par la relation (2.11). Cette matrice est diagonale grâce à l'orthogonalité de cette base. Le détail de sa structure est approfondi au paragraphe 3.5.1, en particulier à la définition 3.4.
- $\mathbf{L}_c(\mathbf{W}^{(n)})$ appartient à $\mathbb{R}^{N_{adi}}$, il est le vecteur de résidu des termes de convection dont chaque composante est définie par la projection de l'opérateur de discrétisation \mathcal{L}_c donné par la relation (2.20) et évalué en $\mathbf{w}_h^{(n)}$ au temps $t^{(n)}$ sur l'espace d'approximation $[\mathcal{V}_h^p]^{d+2}$.

3.4.2.2 Schéma implicite retenu pour le problème parabolique

Le problème parabolique (3.10) est discrétisé avec le schéma implicite de Crank-Nicolson qui est précis au second ordre en temps et qui présente la plus petite erreur de troncature parmi les méthodes inconditionnellement stables de second ordre [34, 36].

Pour le problème (3.10), l'approximation de la dérivée temporelle s'écrit : Trouver $\mathbf{w}_{h,v}^{(n+1)}$ appartenant à $[\mathcal{V}_h^p]^{d+2}$, telle que pour tout ϕ^k élément de \mathcal{B}_h^p , il vienne à chaque instant $t^{(n)}$

$$\int_{\Omega_h} \frac{\mathbf{w}_{h,v}^{(n+1)} - \mathbf{w}_{h,c}^{(n+\frac{1}{2})}}{\Delta t} \phi^k dV = -\frac{1}{2} \left(\mathcal{L}_v(\mathbf{w}_{h,v}^{(n+1)}, \phi^k) + \mathcal{L}_v(\mathbf{w}_{h,c}^{(n+\frac{1}{2})}, \phi^k) \right).$$

Cependant, l'opérateur \mathcal{L}_v étant non linéaire par rapport à $\mathbf{w}_{h,v}$, il est linéarisé au moyen d'un développement limité à l'ordre un en $t^{(n+1)}$ autour de $\mathbf{w}_{h,v}^{(n)}$. Il vient :

$$\mathcal{L}_v(\mathbf{w}_{h,v}^{(n+1)}, \phi^k) = \mathcal{L}_v(\mathbf{w}_{h,v}^{(n)}, \phi^k) + \left. \frac{\partial \mathcal{L}_v}{\partial \mathbf{w}_h} \right|_{\mathbf{w}_h = \mathbf{w}_{h,v}^{(n)}} (\mathbf{w}_{h,v}^{(n+1)} - \mathbf{w}_{h,v}^{(n)}),$$

où $\partial \mathcal{L}_v / \partial \mathbf{w}_h$ appartient à $\mathbb{R}^{N_{dat} \times N_{dat}}$ et constitue la matrice jacobienne de dérivation du résidu discret, par rapport au vecteur \mathbf{w}_h , estimée à l'instant $t^{(n)}$. Finalement, le problème discret (3.10) s'écrit sous la forme du système linéaire suivant :

$$\mathbf{A} \Delta \mathbf{W}^{(n+1)} = -\mathbf{L}_v(\mathbf{W}_c^{(n+\frac{1}{2})}), \quad (3.15)$$

avec

$$\mathbf{A} = \frac{1}{\Delta t} \mathbf{M} + \left. \frac{1}{2} \frac{\partial \mathbf{L}_v}{\partial \mathbf{W}} \right|_{\mathbf{w} = \mathbf{w}_v^{(n)}}, \quad (3.16)$$

et

$$\Delta \mathbf{W}^{(n+1)} = \mathbf{W}_v^{(n+1)} - \mathbf{W}_c^{(n+\frac{1}{2})},$$

où les notations suivantes ci-après sont utilisées.

- \mathbf{A} appartient à $\mathbb{R}^{N_{dat} \times N_{dat}}$, elle est la matrice implicite à inverser. Cette matrice est non symétrique, creuse et structurée en N lignes de $(n_f + 1)$ blocs (matrices) de dimension $N_{eq} N_p \times N_{eq} N_p$, avec n_f le nombre de voisins d'un élément κ (cf. § 2.3.1 page 37). Le détail de cette structure est approfondi au paragraphe 3.5.1. Lorsqu'aucune simplification n'est envisagée, cette matrice est calculée explicitement, c'est-à-dire termes à termes.
- $\mathbf{L}_v(\mathbf{W}_c^{(n+\frac{1}{2})})$ appartient à $\mathbb{R}^{N_{dat}}$, il est le vecteur de résidu des termes de diffusion dont chaque composante est définie par la projection de l'opérateur de discrétisation \mathcal{L}_c donné par la relation (2.20) et évalué en $\mathbf{w}_{h,c}^{(n+\frac{1}{2})}$ au temps $t^{(n+\frac{1}{2})}$ sur l'espace d'approximation $[\mathcal{V}_h^p]^{d+2}$.

Définition 3.3. Dans la suite de ce travail, la résolution du système des équations de Navier-Stokes avec une discrétisation spatiale GD (relations (2.14), (2.20) et (2.29)) et une discrétisation temporelle explicite-implicite complète (relations (3.13), (3.15) et (3.14)) est désignée par l'appellation « COMP ».

A présent, il reste à résoudre le système linéaire (3.15). La grande taille de ce système pour des applications réelles, ainsi que les caractéristiques de la matrice \mathbf{A} rendant sa factorisation coûteuse, amènent naturellement au choix d'une méthode de résolution itérative plutôt qu'une méthode directe.

Les méthodes itératives construisent une suite d'approximations de la solution numérique, à partir d'une estimation initiale de la solution jusqu'à ce qu'un certain critère de convergence, fixé par l'utilisateur, soit atteint. Par ailleurs, la rapidité de convergence de ces méthodes est fortement améliorée en utilisant un préconditionneur. Quelques exemples de méthodes itératives utilisées de pair avec une méthode GD en espace ont déjà été donnés dans l'introduction générale de ce mémoire (cf. § 0.3.3.4 page 12).

Dans ce travail, le choix de la méthode itérative retenue est motivé par la structure creuse de la matrice \mathbf{A} . Ainsi, une méthode de type Newton-Krylov reposant

sur le principe généralisant la minimisation du résidu (ou GMRes) et optimisant le coût de stockage mémoire est appliquée au système (3.15) préalablement préconditionné. C'est la méthode GMRes à redémarrages, notée GMRes(s) où $s > 0$ est un entier. Ces deux points, préconditionneur et méthode GMRes(s), sont maintenant expliqués pour le système (3.15). Pour plus de détails, le lecteur peut consulter [147, chapitre 9] et [44, chapitre 11].

Préconditionnement du système. Soit \mathbf{P} appartenant à $\mathbb{R}^{N_{dat} \times N_{dat}}$ une matrice de préconditionnement (ou préconditionneur) inversible et associée au système préconditionné à gauche suivant :

$$\mathbf{P}^{-1} \mathbf{A} \Delta \mathbf{W}^{(n+1)} = -\mathbf{P}^{-1} \mathbf{L}_v(\mathbf{W}_c^{(n+\frac{1}{2})}),$$

lequel est équivalent à (3.15) au sens où il admet la même solution. La matrice \mathbf{P} est choisie de telle sorte que la matrice $\mathbf{P}^{-1} \mathbf{A}$ ait de meilleures propriétés que la matrice \mathbf{A} en ce qui concerne, par exemple, le conditionnement. En résumé, cette matrice se présente comme un compromis entre la réduction du nombre d'itérations nécessaires pour la convergence de la méthode itérative utilisée et les coûts de calculs supplémentaires qu'elle implique par itération.

Le préconditionneur utilisé est un algorithme de Jacobi par blocs, à savoir bloc diagonal. On l'a retenu pour son efficacité sur des systèmes bien conditionnés, pour sa simplicité de mise en œuvre, pour le peu d'espace mémoire qu'il requiert et pour ses coûts de calculs raisonnables [127, 39, 126]. Il consiste à appliquer une décomposition LU aux N blocs de dimension $N_{eq} N_p \times N_{eq} N_p$ de la diagonale de la matrice implicite \mathbf{A} (cf. § 3.5.1 pour plus de détails sur la structure de \mathbf{A}). Le lecteur peut consulter [130, chapitre 2] pour cette décomposition.

Remarque 3.3. *Dans la pratique, les blocs diagonaux de la matrice implicite sont directement remplacés par la décomposition LU, laquelle permet à la fois le produit matrice-vecteur et l'inversion requise par l'étape de préconditionnement. Ce remplacement évite le stockage supplémentaire de la matrice \mathbf{P} .*

Dans la suite de ce travail, on fait le choix de ne plus indiquer la matrice de préconditionnement dans les équations. Toutefois, l'étape de décomposition devant être réalisée, on y fera référence comme l'étape de construction de la matrice \mathbf{P} .

Remarque 3.4. *Dans la littérature [132, 127, 39] le préconditionneur LU incomplet par blocs est souvent préféré lors de l'utilisation d'une méthode GD car il permet de tenir compte de la matrice \mathbf{A} complète avec des coûts d'encombrement mémoire et de temps de calcul raisonnables sur une large gamme de problèmes. Cependant, il n'est pas retenu dans ce travail où des problèmes instationnaires sont considérés. En effet, ces derniers requièrent l'utilisation de petits pas de temps conduisant à une matrice implicite à diagonale dominante bien conditionnée. De fait, un préconditionneur LU incomplet n'apporterait pas une meilleure efficacité de la résolution du système linéaire.*

Méthode GMRes(s). La méthode GMRes en général [148] est une méthode de résolution de système linéaire particulièrement bien adaptée aux matrices de grandes

tailles, creuses et non symétriques. Son avantage, outre d'être très stable, est qu'elle ne requiert que des multiplications de la matrice \mathbf{A} par un vecteur, des inversions de systèmes pour la matrice \mathbf{P} , des produits scalaires et des additions de vecteurs.

Soit $\Delta\mathbf{W}^{(n+1,0)}$ une valeur de départ approchée de la solution du système (3.15), soit $\mathbf{r}^{(0)}$ le vecteur résidu associé défini par :

$$\mathbf{r}^{(0)} = -\mathbf{L}_v(\mathbf{W}_c^{(n+\frac{1}{2})}) - \mathbf{A}\Delta\mathbf{W}^{(n+1,0)},$$

et soit \mathcal{K}^i l'espace de Krylov d'ordre i avec $1 \leq i \leq s$ un entier, l'espace vectoriel engendré par $\mathbf{r}^{(0)}$ et ses $i - 1$ produits itérés par \mathbf{A} , défini par :

$$\mathcal{K}^i = \text{vect} \left\{ \mathbf{r}^{(0)}, \mathbf{A}\mathbf{r}^{(0)}, \dots, \mathbf{A}^{i-1}\mathbf{r}^{(0)} \right\}.$$

On précise que l'entier s correspond à la dimension maximale de l'espace de Krylov pour un $\mathbf{r}^{(0)}$ donné (cf. remarque 3.5). La méthode GMRes(s) appliquée au système (3.15), consiste alors à construire, à l'étape i , l'itérée $\Delta\mathbf{W}^{(n+1,i)}$ dans l'espace $\Delta\mathbf{W}^{(n+1,0)} + \mathcal{K}^i$ de telle sorte que la suite $(\Delta\mathbf{W}^{(n+1,i)})_{1 \leq i \leq s}$ converge vers la solution du système lorsque s tend vers l'infini, et que la norme L^2 du résidu préconditionné $\mathbf{r}^{(i)}$ soit minimisée sur le même espace $\Delta\mathbf{W}^{(n+1,0)} + \mathcal{K}^i$, avec :

$$\mathbf{r}^{(i)} = -\mathbf{L}_v(\mathbf{W}_c^{(n+\frac{1}{2})}) - \mathbf{A}\Delta\mathbf{W}^{(n+1,i)}.$$

Autrement dit, il s'agit d'un problème classique d'optimisation en dimension i de type moindres carrés. Il se décompose alors en deux principales étapes : la construction d'une base de l'espace de Krylov d'une part et la minimisation du résidu d'autre part. La première étape implique le calcul d'une base orthonormale de dimension i de l'espace \mathcal{K}^i par l'intermédiaire de l'algorithme dit d'Arnoldi. La seconde étape de l'algorithme se réduit à un problème de moindres carrés et repose sur une suite de rotations de Givens. La mise en œuvre de ces deux étapes n'est pas détaillée ici. Le lecteur peut consulter [32] pour l'algorithme utilisé.

Remarque 3.5. *A chaque itération i de la méthode GMRes(s), les i vecteurs de taille N_{dat} qui engendrent l'espace \mathcal{K}^i doivent être connus et donc stockés. Le coût de ce stockage mémoire par itération augmente alors linéairement avec i et devient rédhibitoire pour des applications réelles. C'est pourquoi chaque espace \mathcal{K}^i est limité à une taille maximale de s vecteurs [148].*

Par conséquent, si la convergence requise sur le résidu $\mathbf{r}^{(i)}$ n'a pas été obtenue au bout de s itérations, la solution du problème de minimisation est fixée à $\Delta\mathbf{W}^{(n+1,s)}$ et l'algorithme est redémarré avec la nouvelle initialisation : $\Delta\mathbf{W}^{(n+1,0)} = \Delta\mathbf{W}^{(n+1,s)}$. Cette variante de la méthode GMRes originale, notée GMRes(s), est donc dite à redémarrages.

Dans la suite de ce mémoire, on parlera de la méthode de résolution GMRes (ou de l'inversion GMRes), la référence à la variante GMRes(s) ne sera faite que si nécessaire.

Remarque 3.6. *Pour les applications numériques de la partie II, on a pris l'initialisation telle que $\Delta\mathbf{W}^{(n+1,0)} = 0$, $s = 10$ vecteurs de Krylov maximum et une précision relative en norme L^2 sur le résidu $(\mathbf{r}^i)_{1 \leq i \leq s}$ de 10^{-5} pour stopper la convergence.*

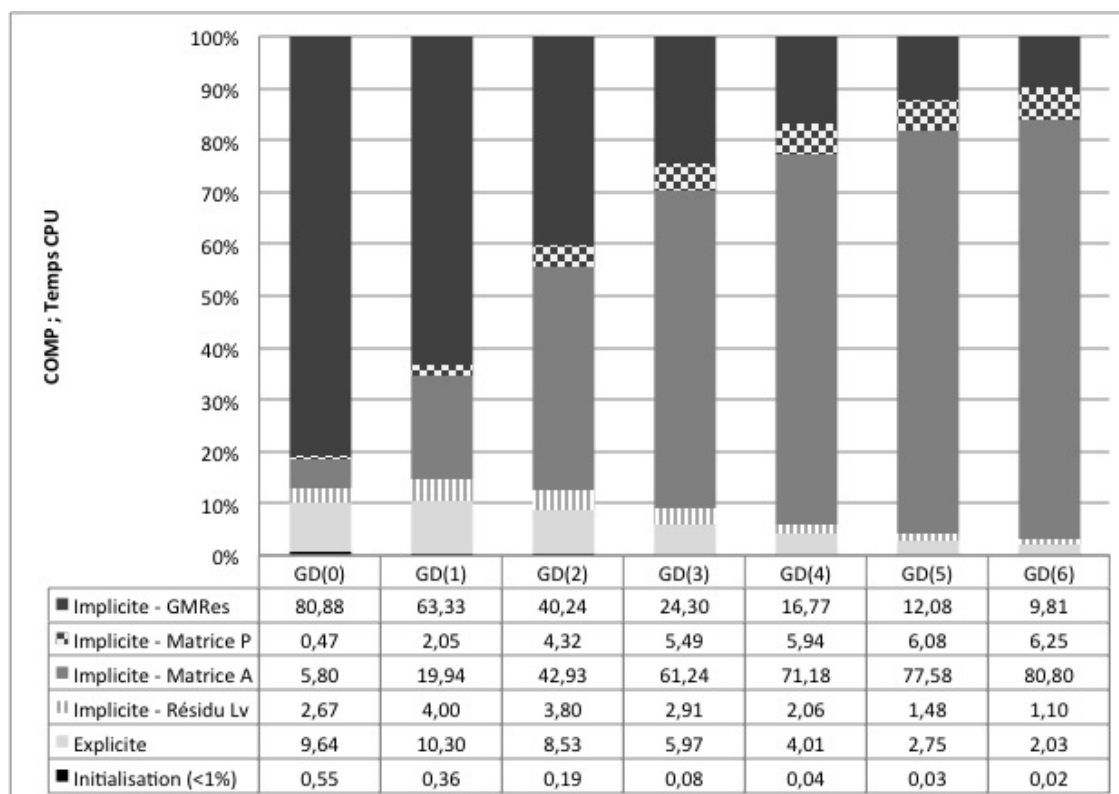


Fig. 3.2 – Illustration des contributions relatives (en pourcentage) aux coûts en temps CPU pour les étapes de la résolution explicite-implicite du schéma COMP. Calculs pour le cas test 2, sur le maillage uniforme $N = 2\,888$ et pour $0 \leq p \leq 6$. La table de correspondance étapes-couleurs est indiquée dans la marge gauche du tableau.

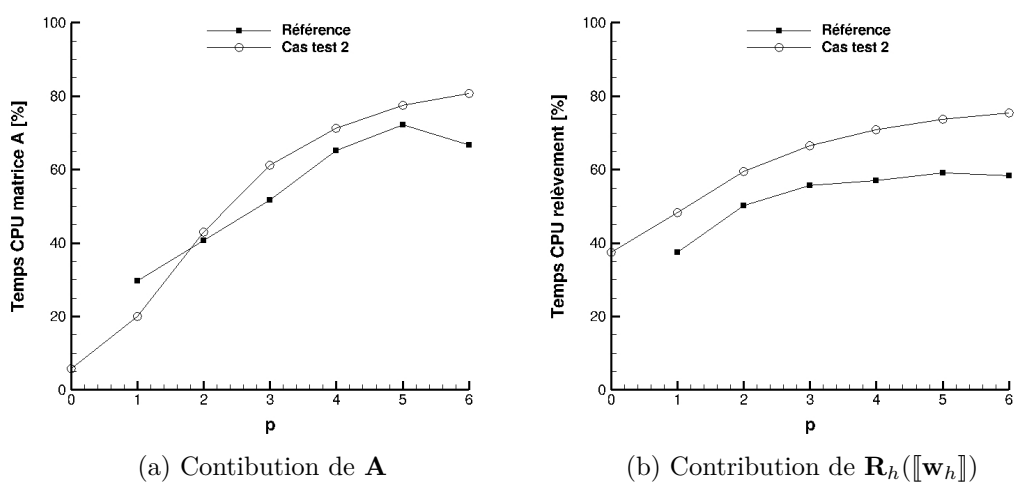


Fig. 3.3 – En (a) : temps de construction (en pourcentage) en fonction du degré de polynôme de la matrice implicite, ramené au temps CPU de la résolution totale. En (b) : temps de construction (en pourcentage) en fonction du degré de polynôme de la matrice jacobienne de dérivation du relèvement global $\mathbf{R}_h(\llbracket \mathbf{w}_h \rrbracket)$ par rapport au vecteur \mathbf{W}_v dans les intégrales de volume, ramené au temps CPU de la résolution totale de la matrice implicite. Comparaison entre la méthode GD utilisée dans [figure 1][35] et celle COMP utilisée avec le cas test 2 de ce mémoire.

3.4.3 Coûts en temps de calcul de la résolution complète

Le schéma numérique discret du système des équations de Navier-Stokes (1.1) se résume donc à la résolution des trois systèmes d'équations (3.13), (3.15) et (3.14) à chaque pas de temps. L'objet de ce paragraphe est d'évaluer les coûts en temps CPU de cette résolution complète sur un exemple numérique concret afin de justifier la recherche d'un schéma explicite-implicite simplifié.

Pour cela, l'exemple considéré est un des cas tests de la partie II simulant l'écoulement d'une paire de tourbillons visqueux et contra-rotatifs se dirigeant perpendiculairement à une paroi plane, soit le cas test 2 (*cf.* § 5.2 page 115 pour plus de détails sur ce cas test). Les coûts en temps CPU sont calculés uniquement pour les étapes significatives de résolution des schémas explicite et implicite. Pour une meilleure vision, cette dernière étape de résolution implicite est décomposée suivant l'étape de construction du vecteur résidu des termes de diffusion \mathbf{L}_v , puis celles de construction de la matrice implicite \mathbf{A} et de construction de la matrice de préconditionnement \mathbf{P} , et enfin celle de l'inversion GMRes.

La figure 3.2 indique la répartition de ces coûts moyens, évalués sur $n = 500$ itérations, du temps CPU en pourcentage sur un maillage uniforme tel que $N = 2\,888$ et pour $0 \leq p \leq 6$. Son analyse amène à plusieurs observations.

- Tout d'abord, elle atteste que le temps de résolution du schéma implicite est très important par rapport à celui des deux résolutions explicites. En effet, pour $p \leq 1$, ces deux dernières ne représentent que 10% du temps total de résolution, et pour p plus grand, cette proportion ne fait que décroître.
- Ensuite, elle confirme que l'étape de construction de la matrice implicite peut s'avérer très coûteuse voire réhibitoire pour des applications réelles, en particulier pour des simulations numériques d'ordre élevé (comme cela est visible sur la figure) et dans le cas d'un grand nombre d'éléments (non visible sur la figure). En effet le temps de cette construction atteint 80% du temps total de résolution pour $p = 6$.
- Par ailleurs, pour cette étape de construction de la matrice implicite, il est intéressant de noter, avec la figure 3.3a, que Crivellini et Bassi [35] observent un comportement similaire pour le coût de construction de \mathbf{A} lors de la résolution du système des équations RANS avec une méthode GD utilisant le flux BR2 pour les termes de diffusion. En particulier, la contribution de la matrice jacobienne de dérivation du relèvement global $\mathbf{R}_h(\llbracket \mathbf{w}_h \rrbracket)$ par rapport au vecteur solution des termes visqueux \mathbf{W}_v dans les intégrales de volume est la plus importante. En effet, elle représente jusqu'à 76% du temps de construction de la matrice pour $p = 6$ comme l'illustre la figure 3.3b. L'utilisation du schéma BR2 joue donc un rôle significatif dans les coûts en temps CPU de la construction de \mathbf{A} .
- Concernant le coût de construction de la matrice de préconditionnement, il est peu significatif dans la résolution du système linéaire implicite lorsque p est petit ($p \leq 4$). Lorsque p devient plus grand, il est comparable à celui de l'inversion GMRes et atteint, par exemple, 6% lorsque $p = 6$.
- Enfin, cette analyse révèle que les deux étapes de construction de la matrice implicite et de l'inversion GMRes sont les plus significatives du temps de réso-

lution total, avec un poids quasiment identique de 40% lorsque $p = 2$ et qui se compense sinon. Globalement le coût de l'inversion GMRes diminue lorsque p augmente : il est de 81% pour $p = 0$ et huit fois plus petit pour $p = 6$, tandis que le coût de construction de la matrice \mathbf{A} augmente lorsque p augmente : il est de 6% pour $p = 0$ et quatorze fois plus grand pour $p = 6$.

Au vu de ces résultats, et sachant que certaines applications réelles peuvent prendre plusieurs mois de résolution lors d'écoulements pleinement turbulents par exemple, il convient d'adopter des stratégies de simplification du schéma numérique implicite. Dans la communauté scientifique, deux stratégies se démarquent principalement : exploiter la structure de la matrice implicite [132, 46, 89, 127, 39, 15, 126] et accélérer la convergence du schéma implicite [46, 114, 109, 90, 39, 181, 40, 35]. Dans ce travail, c'est principalement la première stratégie qui est mise en œuvre et qui fait l'objet du prochain paragraphe.

3.5 Simplifications du schéma implicite en temps

Différentes techniques permettent d'envisager une réduction des coûts en temps CPU de la résolution du schéma numérique comme cela a déjà été mentionné dans l'introduction générale de ce mémoire (*cf.* § 0.3.3.4 page 12). Dans ce travail, les techniques proposées se focalisent sur une simplification de la construction de la matrice implicite et il en résulte des gains, non seulement sur cette construction, mais aussi sur l'inversion du système préconditionné avec la méthode GMRes.

Ces techniques sont au nombre de trois et reposent : soit sur une approximation de la matrice implicite (§ 3.5.2), soit sur un découplage partiel des degrés de liberté (§ 3.5.3), soit sur les deux à la fois (§ 3.5.4). Pour une meilleure compréhension des simplifications envisagées, la structure de la matrice implicite \mathbf{A} est d'abord détaillée dans le cas du système des équations de Navier-Stokes (§ 3.5.1).

3.5.1 Structure de la matrice implicite

La matrice implicite \mathbf{A} est définie en (3.16). Sa dimension étant $N_{ddl} \times N_{ddl}$, trois niveaux de construction correspondant aux trois nombres N , N_{eq} et N_p peuvent être considérés (§ 3.5.1.1, § 3.5.1.2 et § 3.5.1.3). Ces trois niveaux sont illustrés par la figure 3.4 pour le système des équations de Navier-Stokes (1.1) (soit $N_{eq} = 4$) avec $N = 4$ triangles et $p = 2$ (soit $N_p = 6$). La structure complète de la matrice \mathbf{A} est alors donnée sur la figure 3.5a.

3.5.1.1 Premier niveau de construction

Le premier niveau de construction concerne les N éléments de la discrétisation Ω_h du domaine (*cf.* figure 3.4b). La matrice est structurée en N lignes de $n_f + 1$ blocs (matrices) de dimension $N_{eq}N_p \times N_{eq}N_p$. La forme compacte suivante est adoptée pour décrire la structure creuse de ce niveau :

$$\mathbf{A} = \left(\mathbf{A}_{\kappa^+\kappa^+} | \mathbf{A}_{\kappa^+\kappa_1^-} | \dots | \mathbf{A}_{\kappa^+\kappa_{n_f}^-} \right)_{\kappa^+ \in \Omega_h}, \quad (3.17)$$

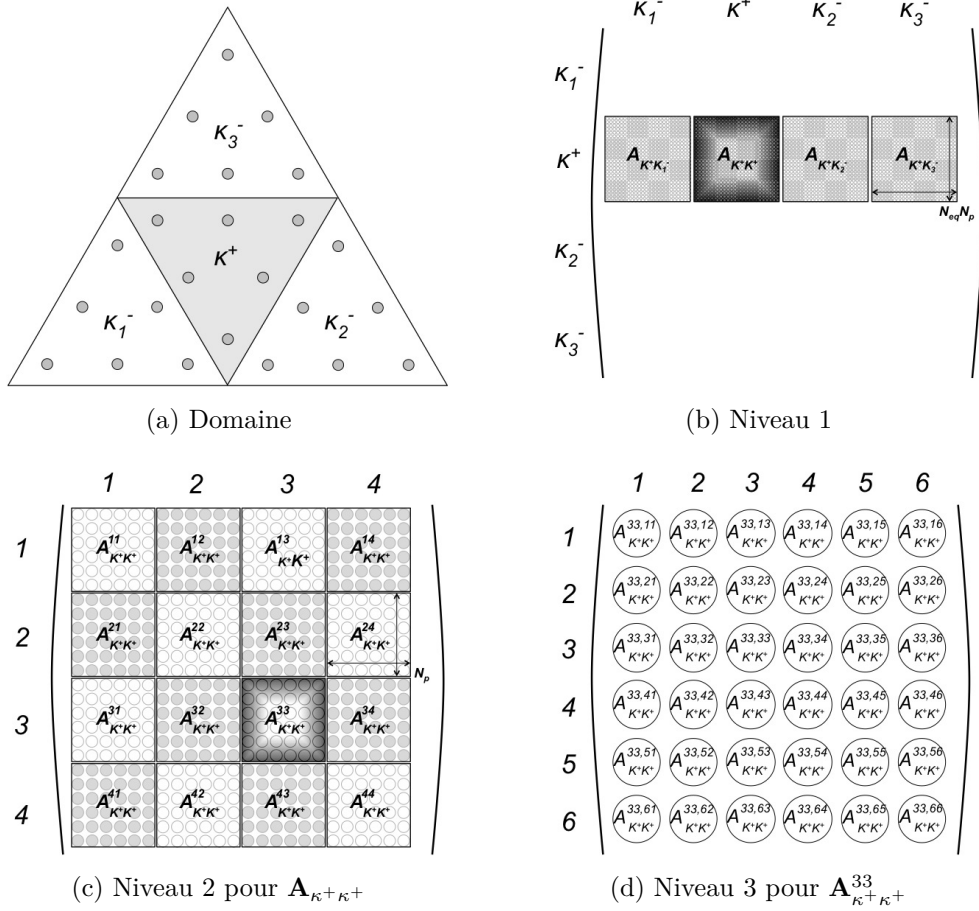
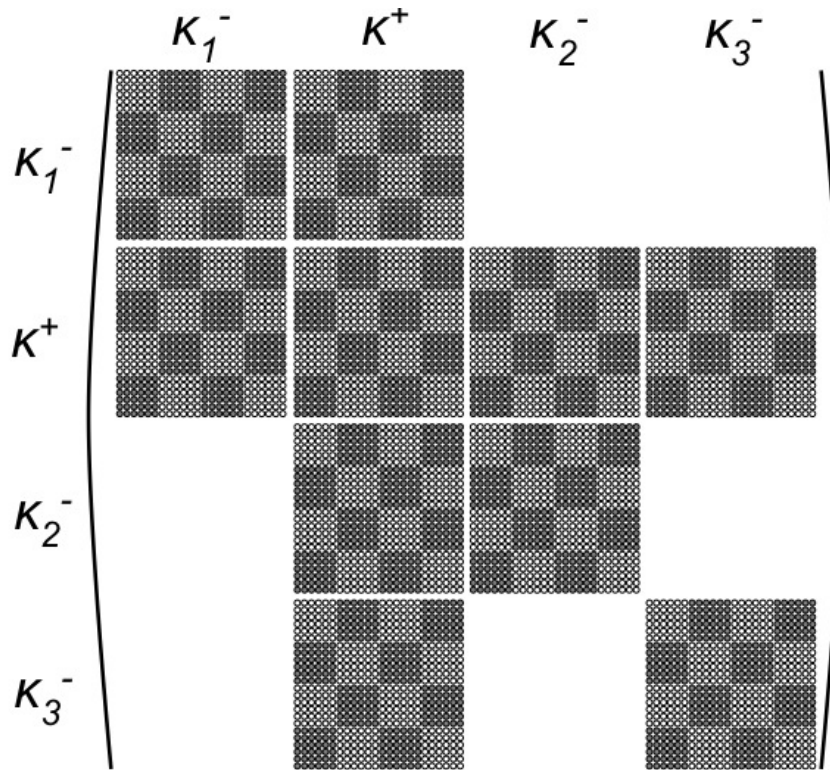


Fig. 3.4 – Illustration des trois niveaux de construction de la matrice implicite. Exemple pour le système des équations de Navier-Stokes (1.1) (soit $N_{eq} = 4$) avec $N = 4$ triangles et $p = 2$ (soit $N_p = 6$). Les degrés de liberté par élément sont représentés par des cercles.

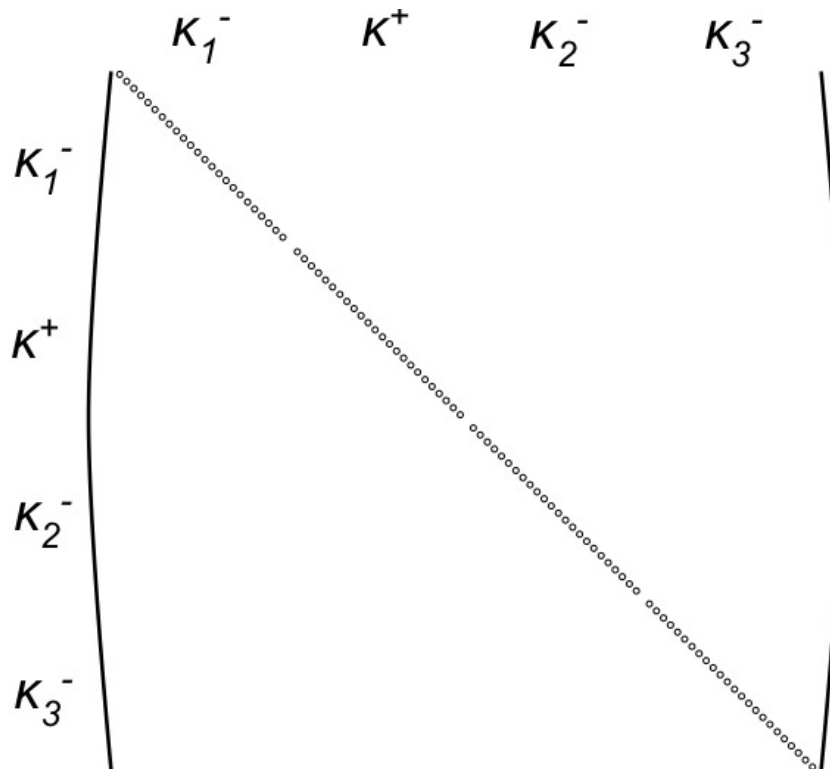
où $\mathbf{A}_{\kappa^+ \kappa^+}$ désigne le bloc diagonal lié à l'élément courant κ^+ , et $\mathbf{A}_{\kappa^+ \kappa_i^-}$ désigne le bloc extra-diagonal i lié aux voisins de l'élément κ^+ pour $1 \leq i \leq n_f$. Pour alléger les notations, les blocs extra-diagonaux sont notés $\mathbf{A}_{\kappa^+ \kappa^-}$ par la suite.

Ainsi, chacune des N lignes de blocs correspond à la relation (3.16) où le vecteur des degrés de liberté sur un élément κ^+ est considéré, soit \mathbf{W}_{κ^+} défini par la relation (2.9). Chaque bloc diagonal $\mathbf{A}_{\kappa^+ \kappa^+}$ correspond aux dérivées possibles de ce vecteur par rapport au même vecteur évalué sur l'élément κ^+ . Il s'agit du couplage, interne à un élément, des degrés de liberté $W_{\kappa^+}^{i,l}$, pour $1 \leq i \leq N_{eq}$ et $1 \leq l \leq N_p$. De même, chaque bloc extra-diagonal $\mathbf{A}_{\kappa^+ \kappa^-}$ correspond aux dérivées possibles de ce vecteur par rapport au vecteur \mathbf{W}_{κ^-} évalué sur un des élément voisins κ^- . Autrement dit, il s'agit du couplage, entre les éléments, des degrés de liberté des vecteurs \mathbf{W}_{κ^+} et \mathbf{W}_{κ^-} .

Par ailleurs, le bloc $\mathbf{A}_{\kappa^+ \kappa^+}$ contient les termes de la matrice de masse (cf. définition 3.4).



(a) Matrice A



(b) Matrice M

Fig. 3.5 – Structure complète de la matrice implicite (a) et de la matrice de masse (b). Exemple pour le système des équations de Navier-Stokes déjà considéré à la figure 3.4a.

3.5.1.2 Second niveau de construction

Le second niveau de construction concerne les N_{eq} composantes du vecteur des variables conservatives \mathbf{w}_h (cf. figure 3.4c). Chaque bloc du niveau de construction précédent est structuré en N_{eq} lignes de N_{eq} blocs (matrices) de dimension $N_p \times N_p$. La forme suivante est adoptée pour décrire la structure de ce niveau :

$$\mathbf{A}_{\kappa^+\kappa^\pm} = \left(\mathbf{A}_{\kappa^+\kappa^\pm}^{ij} \right)_{1 \leq i, j \leq N_{eq}}, \quad \forall \kappa^+ \in \Omega_h.$$

Ainsi, chacune des N_{eq} lignes de blocs correspond à la relation (3.16) où une composante i du vecteur des degrés de liberté sur un élément κ^+ est considérée. Chacune des N_{eq} colonnes de blocs correspond aux couplages de cette composante avec la même composante i (blocs diagonaux) ou avec une autre composante j (blocs extra-diagonaux), de ce même vecteur évalué sur l'élément κ^+ ou du vecteur évalué sur l'élément κ^- , pour $1 \leq i, j \leq N_{eq}$.

3.5.1.3 Troisième niveau de construction

Le dernier niveau de construction concerne les N_p degrés de liberté par élément κ (cf. figure 3.4d). Chaque bloc du niveau de construction précédent est structuré en une matrice de dimension $N_p \times N_p$. La forme suivante est adoptée pour décrire la structure de ce niveau :

$$\mathbf{A}_{\kappa^+\kappa^\pm}^{ij} = \left(A_{\kappa^+\kappa^\pm}^{ij,kl} \right)_{1 \leq k, l \leq N_p}, \quad \forall \kappa^+ \in \Omega_h, \quad 1 \leq i, j \leq N_{eq}. \quad (3.18)$$

Ainsi, chacune des N_p lignes correspond à la relation (3.16) où un degré de liberté k d'une composante i du vecteur solution sur un élément κ^+ est considéré, soit $W_{\kappa^+}^{i,k}$ avec $1 \leq i \leq N_{eq}$ et $1 \leq k \leq N_p$. Chacune des N_p colonnes correspond aux couplages de ce degré avec le même degré k (termes diagonaux) ou avec un autre degré l (termes extra-diagonaux), de la même composante i ou d'une autre composante j , de ce même vecteur évalué sur l'élément κ^+ ou du vecteur évalué sur l'élément κ^- , pour $1 \leq i, j \leq N_{eq}$ et $1 \leq k, l \leq N_p$.

Remarque 3.7. *La différence entre une solution d'ordre élevé ou non se situe dans la dimension des matrices du niveau 3. Par exemple, pour la méthode des volumes finis, cette dimension est 1×1 . La matrice \mathbf{A} est alors constituée de N lignes de $n_f + 1$ matrices de dimension $N_p \times N_p$, dans le cas d'un schéma compact pour le système des équations de Navier-Stokes considéré dans ce travail.*

Définition 3.4. (Matrice de masse) *La matrice \mathbf{M} est de dimension $N_{ddl} \times N_{ddl}$, c'est pourquoi les trois niveaux de construction de la matrice implicite sont également considérés. En reprenant les notations précédentes de la matrice \mathbf{A} , la matrice de masse est définie par :*

$$\mathbf{M} = \text{diag} \left(\left(\mathbf{M}_{\kappa^+} \right)_{\kappa^+ \in \Omega_h} \right),$$

avec

$$\mathbf{M}_{\kappa^+} = \text{diag} \left(\left(\mathbf{M}_{\kappa^+}^{ii} \right)_{1 \leq i \leq N_{eq}} \right), \quad \forall \kappa^+ \in \Omega_h,$$

et avec

$$\mathbf{M}_{\kappa^+}^{ii} = \text{diag} \left(M_{\kappa^+}^{ii,kk} \right)_{1 \leq k \leq N_p}, \quad \forall \kappa^+ \in \Omega_h, \quad 1 \leq i \leq N_{eq},$$

où chaque composante $(M_{\kappa^+}^{ii,kk})_{1 \leq k \leq N_p}$ est définie par la relation (2.11). Au final, la matrice de masse est contenue dans la stricte diagonale de \mathbf{A} . Ceci est illustré par la figure 3.5b.

3.5.2 Simplification existante : méthode sans jacobienne

La méthode de résolution complète requiert la construction de la matrice \mathbf{A} pour l'inversion GMRes de la résolution du problème implicite. En raison du temps considérable nécessaire à cette construction, il est préférable de stocker \mathbf{A} plutôt que de recourir à un produit matrice-vecteur à la volée. En effet, dans cette dernière option, il est rappelé que la méthode GMRes nécessite plusieurs produits matrice-vecteur pour la construction de l'espace de Krylov, dont le coût d'un seul est comparable à la construction de la matrice \mathbf{A} , et ce du fait de la taille importante des blocs $\mathbf{A}_{\kappa^+\kappa^\pm}$ lorsque p est grand [35].

La simplification proposée est dite sans jacobienne (ou SJ) car elle approche le produit de la matrice jacobienne de dérivation du résidu discret par un vecteur au moyen d'une différence finie d'ordre un. Par conséquent, elle permet une réduction des coûts (encombrement mémoire et temps CPU) lors de la construction de la matrice \mathbf{A} et du produit de \mathbf{A} avec un vecteur dans l'inversion GMRes (*cf.* § 3.6 pour plus de détails sur ces coûts).

Cette simplification est désormais une alternative classique des méthodes implicites de Newton-Krylov lorsqu'il s'agit de réduire l'encombrement mémoire. Le lecteur peut consulter l'article de Knoll et Keyes [92] pour un historique d'utilisation des méthodes sans jacobienne. En GD, si ces méthodes sont apparues avec succès il y a une dizaine d'années (2001) pour les écoulements de fluides parfaits [132, 64], elles ne sont en revanche devenues à la mode que très récemment (2008) pour les écoulements de fluides visqueux laminaires [123, 121, 91, 15] et turbulents [35].

La méthode de simplification sans jacobienne consiste à approcher le produit de la matrice $\partial \mathbf{L}_v / \partial \mathbf{W}$ avec le vecteur $\Delta \mathbf{W}^{(n+1)}$ dans la relation (3.15) par une différence finie au premier ordre. Le système linéaire implicite de Crank-Nicolson (3.15) est alors résolu avec la méthode GMRes pour laquelle le produit de la matrice \mathbf{A} par un vecteur \mathbf{z} quelconque est remplacé par :

$$(\mathbf{A}\mathbf{z})^\# = \frac{1}{\Delta t} \mathbf{M}\mathbf{z} + \frac{1}{2\delta_{SJ}} \left(\mathbf{L}_v(\mathbf{W}_v^{(n)} + \delta_{SJ}\mathbf{z}) - \mathbf{L}_v(\mathbf{W}_v^{(n)}) \right). \quad (3.19)$$

où δ_{SJ} est un paramètre de stabilisation défini par [125, 35] :

$$\delta_{SJ} \approx \epsilon_{SJ} \frac{\sqrt{1 + \|\mathbf{W}_v^{(n)}\|_{L^2(\Omega_h)}}}{\|\mathbf{z}\|_{L^2(\Omega_h)}},$$

avec ϵ_{SJ} une constante dont le carré représente l'erreur d'arrondi de la machine de calcul (*cf.* remarque 3.8). De plus, pour les besoins du préconditionnement, les blocs diagonaux $\mathbf{A}_{\kappa^+\kappa^+}$ de la matrice implicite sont entièrement construits. Enfin,

la relation (3.19) fait apparaître le produit matrice-vecteur \mathbf{Mz} qui est peu coûteux puisque \mathbf{M} est diagonale. Elle indique également la nécessité de construire et de stocker un second résidu discret évalué en $\mathbf{W}_v^{(n)} + \delta_{SJ}\mathbf{z}$.

Remarque 3.8. *Dans la littérature, plusieurs choix existent pour définir le paramètre δ_{SJ} [92, 122]. La détermination de ce dernier peut s'avérer critique pour l'efficacité et la stabilité d'un schéma numérique. Le choix retenu ici se base sur les travaux de Crivellini et Bassi [35] car ces derniers résolvent le système des équations de Navier-Stokes avec un schéma discret très proche de celui de ce travail : méthode GD en espace, flux BR2 pour déterminer les flux numériques visqueux, méthode de Newton-Krylov, simplification sans jacobienne.*

Pour les applications numériques de la partie II, on a pris $\epsilon_{SJ} = 10^{-8}$ comme suggéré par [35], s'inspirant eux-mêmes de [125].

Définition 3.5. *Dans la suite de ce travail, la résolution du système des équations de Navier-Stokes avec une discrétisation spatiale GD (relations (2.14), (2.20) et (2.29)) et une discrétisation temporelle explicite-implicite simplifiée (relations (3.13), (3.15) en utilisant (3.19) et (3.14)) est désignée par l'appellation « SJ ».*

3.5.3 Simplification innovante : découplage des degrés de liberté

Dans la simplification SJ, les gains en temps CPU ne sont pas aussi importants qu'espérés en raison de la construction du préconditionneur. Ceci est d'autant plus vrai avec, par exemple, l'utilisation d'un préconditionneur LU incomplet qui requiert la construction complète de \mathbf{A} [35].

La nouvelle méthode de simplification proposée exploite les deux caractères modale et hiérarchique de la base de polynômes \mathcal{B}_κ^p . Elle repose sur un découplage des degrés de liberté de la solution numérique à l'intérieur d'un élément κ du maillage avec ceux des éléments voisins. Par conséquent, elle permet une réduction des coûts (encombrement mémoire et temps CPU) lors des constructions des matrices \mathbf{A} et \mathbf{P} , et lors des produits de \mathbf{A} et \mathbf{P}^{-1} avec un vecteur dans l'inversion GMRes (*cf.* § 3.6 pour plus de détails sur ces coûts).

Cette simplification a été initiée en 2010 sur l'équation de la chaleur scalaire en une dimension d'espace [37], puis a été mise en œuvre avec succès sur des équations d'advection-diffusion scalaires linéaires et non linéaires en deux dimensions d'espace pour les écoulements à diffusion [139] ou convection dominante [53, 138]. Très récemment, le découplage des degrés de liberté, tel qu'il est appliqué pour cette simplification, a été repris dans [15] mais uniquement sur la matrice de préconditionnement \mathbf{P} , alors qu'ici il est appliqué aux deux matrices \mathbf{A} et \mathbf{P} .

Pour décrire cette simplification, la notion de mode est introduite (§ 3.5.3.1), puis le découplage est illustré sur un exemple simple d'un système d'équations scalaires en une dimension d'espace avant de le généraliser aux dimensions d'espace supérieures (§ 3.5.3.2), enfin la résolution du système implicite simplifié est expliquée (§ 3.5.3.3).

3.5.3.1 Notion de mode

La décomposition modale de \mathbf{w}_h au sein d'un élément κ a été donnée par la relation (2.8). On appelle « mode d'ordre l », la composante de \mathbf{w}_h suivant la l -ième fonction ϕ^l de la base \mathcal{B}_κ^p avec $1 \leq l \leq N_p$, c'est-à-dire la quantité $\mathbf{W}_\kappa^l(t)\phi^l(\mathbf{x})$.

Par ailleurs, soit p_s l'entier tel que $0 \leq p_s \leq p$ permettant de séparer les modes de bas ordre tels que $0 \leq q \leq p_s$ des modes d'ordre élevé tels que $p_s < q \leq p$. Ainsi, la relation (2.8) est reconsidérée sous la forme suivante :

$$\mathbf{w}_h(\mathbf{x}, t) = \underbrace{\sum_{l=1}^{N_{p_s}} \tilde{\mathbf{W}}_\kappa^l(t)\phi^l(\mathbf{x})}_{\text{modes de bas ordre}} + \underbrace{\sum_{l=N_{p_s}+1}^{N_p} \hat{\mathbf{W}}_\kappa^l(t)\phi^l(\mathbf{x})}_{\text{modes d'ordre élevé}}, \quad \forall \mathbf{x} \in \kappa, \kappa \in \Omega_h, \forall t \in \mathcal{T},$$

avec $N_{p_s} = (p_s + 1)(p_s + 2)/2$, et où les symboles $\tilde{}$ et $\hat{}$ portent respectivement, sur les $N_{eq}N_{p_s}$ degrés de liberté de bas ordre et sur les $N_{eq}(N_p - N_{p_s})$ degrés de liberté d'ordre élevé dans κ .

3.5.3.2 Principe de la simplification

Pour plus de clarté, la méthode de simplification est d'abord introduite dans le cas du système couplé des N_{eq} équations de Navier-Stokes en une dimension d'espace (la généralisation au cas des dimensions d'espace supérieures étant directe), puis elle est écrite avec les notations matricielles.

Simplification en une dimension d'espace. Soit Ω_h une discrétisation du domaine Ω qui consiste en N segments réguliers κ et soit w_h^i une composante donnée de la solution discrète, autrement dit une solution de l'équation i du système, pour $1 \leq i \leq N_{eq}$. En prenant l'exemple où le couple $(p, p_s) = (2, 0)$, la composante w_h^i est complètement déterminée par la somme des trois modes suivante sur chaque segment κ :

$$w_h^i(x_1, t) = \tilde{W}_\kappa^{i,1}(t)\phi^1(x_1) + \hat{W}_\kappa^{i,2}(t)\phi^2(x_1) + \hat{W}_\kappa^{i,3}(t)\phi^3(x_1), \quad \forall x_1 \in \kappa, \kappa \in \Omega_h, \forall t \in \mathcal{T}, \quad (3.20)$$

où

- $\tilde{W}_\kappa^{i,1}(t)\phi^1(x_1)$ représente le mode de bas ordre correspondant au mode constant avec $\tilde{W}_\kappa^{i,1}$ la valeur moyenne de la solution numérique sur le segment κ ;
- $\hat{W}_\kappa^{i,2}(t)\phi^2(x_1)$ et $\hat{W}_\kappa^{i,3}(t)\phi^3(x_1)$ représentent les modes d'ordre élevé correspondant aux modes respectivement, linéaire et quadratique.

D'une manière générale, la somme (3.20) permet de reconstruire la solution w_h^i au sein d'un élément. Elle peut donc être vue comme le couplage des degrés de liberté au sein de cet élément. Ce couplage interne à l'élément correspond aux blocs diagonaux $\mathbf{A}_{\kappa^+\kappa^+}$ dans la notation (3.17). Par ailleurs, la détermination de la solution w_h^i sur l'ensemble du domaine Ω_h est assurée à la fois par les flux numériques et les opérateurs de relèvement qui permettent le couplage des modes entre deux éléments κ^+ et κ^- . Ce couplage inter-éléments correspond aux blocs extra-diagonaux

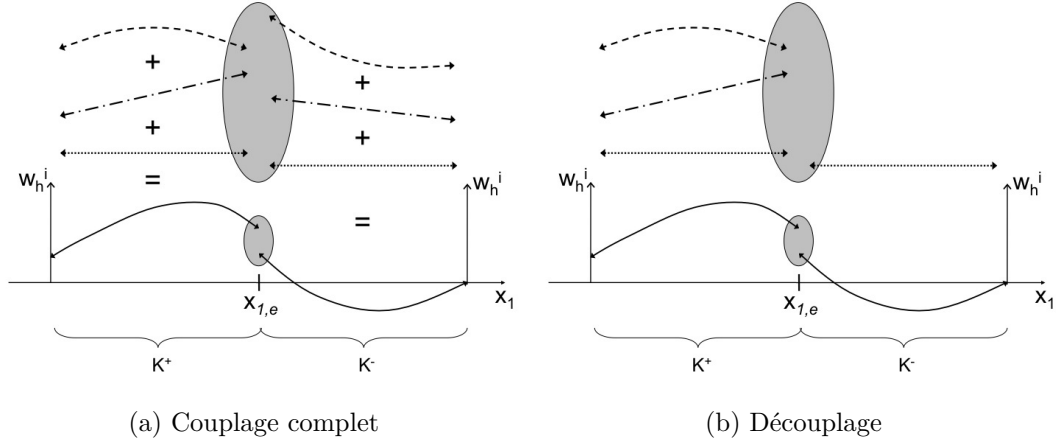


Fig. 3.6 – Exemple de la simplification pour $d = 1$, $p = 2$ et pour la composante i de \mathbf{w}_h avec $1 \leq i \leq N_{eq}$. En (a) : pas de simplification, où sur chaque segment κ , la solution (trait plein) est la somme des trois modes : constant (trait point), linéaire (trait point-pointillé) et quadratique (trait pointillé). La zone grisée indique le couplage des modes au point frontière $x_{1,e}$ entre les segments. En (b) : simplification pour $p_s = 0$.

$\mathbf{A}_{\kappa^+\kappa^-}$ dans la notation (3.17). La figure 3.6a illustre le couplage des trois modes de l'exemple (3.20) sans simplification.

Dans le cas d'une seule équation scalaire la simplification repose sur l'hypothèse 3.1 suivante :

Hypothèse 3.1. (Découplage) *Le couplage de tous les modes de la solution sur le segment κ^+ avec les modes de bas ordre de la solution sur le segment κ^- est dominant. Autrement dit, le couplage de tous les modes de la solution sur κ^+ avec les modes d'ordre élevé de la solution sur κ^- peut être négligé.*

En considérant w_h^i la solution scalaire de l'équation définie par la relation (3.20) avec $1 \leq i \leq N_{eq}$, l'application de cette hypothèse permet de négliger les deux couplages suivants :

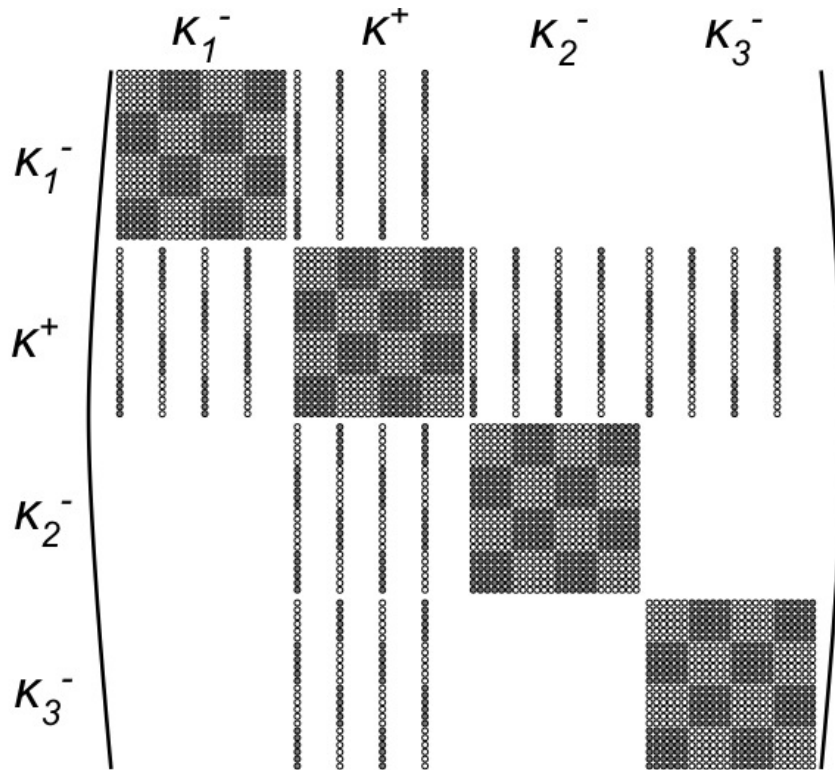
- celui entre les trois modes de la solution w_h^i sur κ^+ et le mode linéaire sur κ^- ;
- celui entre les trois modes de la solution w_h^i sur κ^+ et le mode quadratique sur κ^- .

La figure 3.6b illustre cette procédure de simplification.

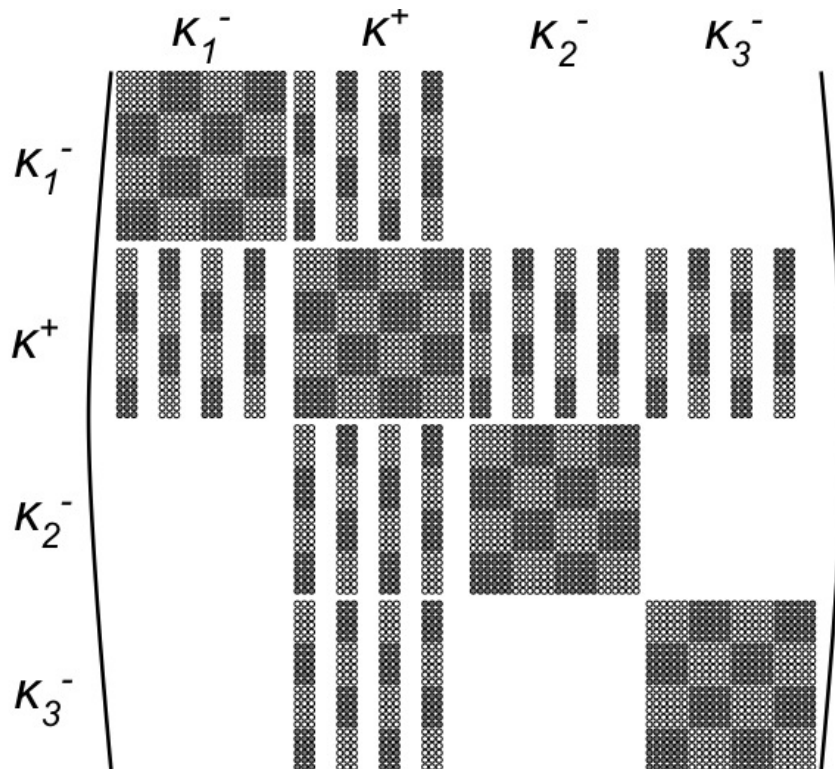
Dans le cas d'un système d'équations, l'hypothèse 3.1 se généralise aisément. En considérant w_h^i et w_h^j deux composantes données de la solution discrète définies toutes deux par la relation (3.20) avec $1 \leq i, j \leq N_{eq}$ et $i \neq j$, l'application de cette hypothèse permet de négliger les couplages suivants :

- les deux vus précédemment pour la composante w_h^i et pour la composante w_h^j ;
- ceux croisés entre les trois modes de la composante w_h^i (respectivement w_h^j) sur κ^+ et les modes linéaire et quadratique de la composante w_h^j (respectivement w_h^i) sur κ^- pour $j \neq i$.

Enfin, dans le cas de dimensions d'espace supérieures, la généralisation est également directe. En effet, la séparation modale dans (3.20) est appliquée selon le degré total de chacun des modes.



(a) Cas $(p, p_s) = (2, 0)$



(b) Cas $(p, p_s) = (2, 1)$

Fig. 3.7 – Structure simplifiée de la matrice implicite. Exemple pour le système des équations de Navier-Stokes déjà considéré à la figure 3.4a.

Ecriture sous forme matricielle. Chacune des matrices du niveau 3 de construction, donnée par la relation (3.18), est reconsidérée en séparant les degrés de liberté de bas ordre de ceux d'ordre élevé. Il vient la forme suivante :

$$\mathbf{A}_{\kappa^+\kappa^\pm}^{ij} = \begin{pmatrix} \tilde{\mathbf{A}}_{\kappa^+\kappa^\pm}^{ij} & \mathbf{B}_{\kappa^+\kappa^\pm}^{ij} \\ \mathbf{C}_{\kappa^+\kappa^\pm}^{ij} & \hat{\mathbf{A}}_{\kappa^+\kappa^\pm}^{ij} \end{pmatrix}, \quad \forall \kappa^+ \in \Omega_h, \quad 1 \leq i, j \leq N_{eq},$$

où $\tilde{\mathbf{A}}_{\kappa^+\kappa^\pm}^{ij}$ est une matrice carrée de dimension $N_{p_s} \times N_{p_s}$, et $\hat{\mathbf{A}}_{\kappa^+\kappa^\pm}^{ij}$ est également une matrice carrée de dimension $(N_p - N_{p_s}) \times (N_p - N_{p_s})$.

La simplification, qui repose sur l'hypothèse 3.1, revient à remplacer chaque sous-bloc extra-diagonal $(\mathbf{A}_{\kappa^+\kappa^-}^{ij})_{1 \leq i, j \leq N_{eq}}$ par le nouveau sous-bloc extra-diagonal simplifié $(\mathbf{A}_{\kappa^+\kappa^-}^{*ij})_{1 \leq i, j \leq N_{eq}}$ tel que :

$$\mathbf{A}_{\kappa^+\kappa^-}^{*ij} = \begin{pmatrix} \tilde{\mathbf{A}}_{\kappa^+\kappa^-}^{ij} & 0 \\ \mathbf{C}_{\kappa^+\kappa^-}^{ij} & 0 \end{pmatrix}, \quad \forall \kappa^+ \in \Omega_h, \quad 1 \leq i, j \leq N_{eq},$$

les sous-blocs diagonaux $(\mathbf{A}_{\kappa^+\kappa^+}^{ij})_{1 \leq i, j \leq N_{eq}}$ étant, quant à eux, conservés.

A l'échelle de la matrice implicite complète \mathbf{A} , la simplification revient alors à considérer la nouvelle matrice simplifiée \mathbf{A}^* sous la forme compacte suivante (cf. figure 3.7) :

$$\begin{aligned} \mathbf{A}^* &= \left(\mathbf{A}_{\kappa^+\kappa^+} \mid \mathbf{A}_{\kappa^+\kappa_1^-}^* \mid \dots \mid \mathbf{A}_{\kappa^+\kappa_{n_f}^-}^* \right)_{\kappa^+ \in \Omega_h} \\ &= \left(\left(\begin{pmatrix} \tilde{\mathbf{A}}_{\kappa^+\kappa^+}^{ij} & \mathbf{B}_{\kappa^+\kappa^+}^{ij} \\ \mathbf{C}_{\kappa^+\kappa^+}^{ij} & \hat{\mathbf{A}}_{\kappa^+\kappa^+}^{ij} \end{pmatrix} \right)_{1 \leq i, j \leq N_{eq}} \mid \left(\begin{pmatrix} \tilde{\mathbf{A}}_{\kappa^+\kappa_1^-}^{ij} & 0 \\ \mathbf{C}_{\kappa^+\kappa_1^-}^{ij} & 0 \end{pmatrix} \right)_{1 \leq i, j \leq N_{eq}} \mid \right. \\ &\quad \left. \dots \mid \left(\begin{pmatrix} \tilde{\mathbf{A}}_{\kappa^+\kappa_{n_f}^-}^{ij} & 0 \\ \mathbf{C}_{\kappa^+\kappa_{n_f}^-}^{ij} & 0 \end{pmatrix} \right)_{1 \leq i, j \leq N_{eq}} \right)_{\kappa^+ \in \Omega_h}. \end{aligned} \quad (3.21)$$

Finalement, la simplification affecte uniquement la structure des blocs extra-diagonaux $\mathbf{A}_{\kappa^+\kappa^-}$.

3.5.3.3 Résolution du problème implicite simplifié

La résolution du problème linéaire implicite (3.15), où la matrice \mathbf{A} est remplacée par la matrice simplifiée \mathbf{A}^* donnée en (3.21), se fait en deux étapes pour chaque pas de temps. L'idée est de séparer le calcul des degrés de liberté de bas ordre, au nombre de $NN_{eq}N_{p_s}$, du calcul des degrés de liberté d'ordre élevé, au nombre de $NN_{eq}(N_p - N_{p_s})$. Les premiers sont déterminés en résolvant un système linéaire implicite et les seconds au moyen d'une étape de reconstruction explicite.

Avant de présenter les deux étapes, les notations suivantes sont adoptées et permettent de réécrire la matrice \mathbf{A}^* avec quatre sous-matrices. Soit les matrices carrées $\tilde{\mathbf{A}}$ et $\hat{\mathbf{A}}$ appartenant respectivement à $\mathbb{R}^{NN_{eq}N_{p_s} \times NN_{eq}N_{p_s}}$ et $\mathbb{R}^{NN_{eq}(N_p - N_{p_s}) \times NN_{eq}(N_p - N_{p_s})}$, et soit les matrices \mathbf{B} et \mathbf{C} appartenant respectivement à $\mathbb{R}^{NN_{eq}N_{p_s} \times NN_{eq}(N_p - N_{p_s})}$ et

$\mathbb{R}^{NN_{eq}(N_p - N_{p_s}) \times NN_{eq}N_{p_s}}$, telles que :

$$\tilde{\mathbf{A}} = \left(\left(\tilde{\mathbf{A}}_{\kappa^+ \kappa^+}^{ij} \right)_{1 \leq i, j \leq N_{eq}} \mid \left(\tilde{\mathbf{A}}_{\kappa^+ \kappa_1^-}^{ij} \right)_{1 \leq i, j \leq N_{eq}} \mid \dots \mid \left(\tilde{\mathbf{A}}_{\kappa^+ \kappa_{n_f}^-}^{ij} \right)_{1 \leq i, j \leq N_{eq}} \right)_{\kappa^+ \in \Omega_h}, \quad (3.22a)$$

$$\hat{\mathbf{A}} = \left(\left(\hat{\mathbf{A}}_{\kappa^+ \kappa^+}^{ij} \right)_{1 \leq i, j \leq N_{eq}} \mid \left(\hat{\mathbf{A}}_{\kappa^+ \kappa_1^-}^{ij} \right)_{1 \leq i, j \leq N_{eq}} \mid \dots \mid \left(\hat{\mathbf{A}}_{\kappa^+ \kappa_{n_f}^-}^{ij} \right)_{1 \leq i, j \leq N_{eq}} \right)_{\kappa^+ \in \Omega_h}, \quad (3.22b)$$

$$\mathbf{B} = \left(\left(\mathbf{B}_{\kappa^+ \kappa^+}^{ij} \right)_{1 \leq i, j \leq N_{eq}} \mid 0 \mid \dots \mid 0 \right)_{\kappa^+ \in \Omega_h}, \quad (3.22c)$$

$$\mathbf{C} = \left(\left(\mathbf{C}_{\kappa^+ \kappa^+}^{ij} \right)_{1 \leq i, j \leq N_{eq}} \mid 0 \mid \dots \mid 0 \right)_{\kappa^+ \in \Omega_h}. \quad (3.22d)$$

Première étape. Cette étape ne considère que l'avancement en temps des $NN_{eq}N_{p_s}$ degrés de liberté de bas ordre au temps $t^{(n+1)}$. Ces derniers sont obtenus en résolvant le schéma linéaire implicite de Crank-Nicolson suivant avec la méthode GMRES :

$$\tilde{\mathbf{A}} \Delta \tilde{\mathbf{W}}^{(n+1)} = -\tilde{\mathbf{L}}_v(\mathbf{W}_c^{(n+\frac{1}{2})}) - \mathbf{B} \Delta \hat{\mathbf{W}}^{(n)}, \quad (3.23)$$

où :

- $\Delta \tilde{\mathbf{W}}^{(n+1)} = \tilde{\mathbf{W}}_v^{(n+1)} - \tilde{\mathbf{W}}_c^{(n+1/2)}$ appartient à $\mathbb{R}^{NN_{eq}N_{p_s}}$, il correspond à l'incrément en temps sur les degrés de liberté de bas ordre $(W_{\kappa}^{i,l,(n+1)})_{\kappa \in \Omega_h, 1 \leq i \leq N_{eq}, 1 \leq l \leq N_{p_s}}^{\top}$;
- $\Delta \hat{\mathbf{W}}^{(n)} = \hat{\mathbf{W}}_v^{(n)} - \hat{\mathbf{W}}_c^{(n-1/2)}$ appartient à $\mathbb{R}^{NN_{eq}(N_p - N_{p_s})}$, il correspond à l'incrément sur les degrés de liberté d'ordre élevé $(W_{\kappa}^{i,l,(n)})_{\kappa \in \Omega_h, 1 \leq i \leq N_{eq}, N_{p_s} + 1 \leq l \leq N_p}$ connus au temps $t^{(n)}$;
- $\tilde{\mathbf{L}}_v(\mathbf{W}_c^{(n+1/2)})$ appartient à $\mathbb{R}^{NN_{eq}N_{p_s}}$, il désigne le vecteur de résidu des termes de diffusion associé aux degrés de liberté de bas ordre.

Seconde étape. Cette étape permet de déterminer les $NN_{eq}(N_p - N_{p_s})$ degrés de liberté d'ordre élevé au temps $t^{(n+1)}$. Ces derniers sont obtenus par une reconstruction explicite à partir des degrés de liberté de bas ordre déterminés à l'étape 1. En considérant la décomposition suivante pour la matrice carrée $\hat{\mathbf{A}}$ définie en (3.22b) : $\hat{\mathbf{A}} = \hat{\mathbf{L}} + \hat{\mathbf{D}} + \hat{\mathbf{U}}$, avec $\hat{\mathbf{L}}$ sa partie strictement inférieure, $\hat{\mathbf{D}}$ sa diagonale et $\hat{\mathbf{U}}$ sa partie strictement supérieure, la reconstruction se réduit à la résolution du système triangulaire suivant :

$$(\hat{\mathbf{L}} + \hat{\mathbf{D}}) \Delta \hat{\mathbf{W}}^{(n+1)} = -\hat{\mathbf{L}}_v(\mathbf{W}_c^{(n+\frac{1}{2})}) - \mathbf{C} \Delta \tilde{\mathbf{W}}^{(n+1)} - \hat{\mathbf{U}} \Delta \hat{\mathbf{W}}^{(n)}, \quad (3.24)$$

où :

- $\Delta \hat{\mathbf{W}}^{(n+1)} = \hat{\mathbf{W}}_v^{(n+1)} - \hat{\mathbf{W}}_c^{(n+1/2)}$ appartient à $\mathbb{R}^{NN_{eq}(N_p - N_{p_s})}$, il correspond aux degrés de liberté d'ordre élevé $(W_{\kappa}^{i,l,(n+1)})_{\kappa \in \Omega_h, 1 \leq i \leq N_{eq}, N_{p_s} + 1 \leq l \leq N_p}$ à déterminer ;
- $\Delta \hat{\mathbf{W}}^{(n)} = \hat{\mathbf{W}}_v^{(n)} - \hat{\mathbf{W}}_c^{(n-1/2)}$ appartient à $\mathbb{R}^{NN_{eq}(N_p - N_{p_s})}$, il correspond aux degrés de liberté d'ordre élevé $(W_{\kappa}^{i,k,(n+1)})_{\kappa \in \Omega_h, 1 \leq i \leq N_{eq}, l \leq k \leq N_p}$ qui ne sont pas encore connus au temps $t^{(n+1)}$ et qui sont alors explicités ;
- $\hat{\mathbf{L}}_v(\mathbf{W}_c^{(n+1/2)})$ appartient à $\mathbb{R}^{NN_{eq}(N_p - N_{p_s})}$, il désigne le vecteur de résidu des termes de diffusion associé aux degrés de liberté d'ordre élevé.

Conclusion. Les deux étapes (3.23) et (3.24) peuvent se résumer à la résolution du système linéaire global suivant :

$$\mathcal{A}\Delta\mathbf{W}^{(n+1)} = -\mathcal{L}_v(\mathbf{W}_c^{(n+\frac{1}{2})}) - \mathcal{B}\Delta\mathbf{W}^{(n)}, \quad (3.25)$$

où les matrices \mathcal{A} et \mathcal{B} appartiennent à $\mathbb{R}^{N_{ddt} \times N_{ddt}}$, elles sont telles que :

$$\mathcal{A} = \begin{pmatrix} \tilde{\mathbf{A}} & 0 \\ \mathbf{C} & \hat{\mathcal{L}} + \hat{\mathcal{D}} \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} 0 & \mathbf{B} \\ 0 & \hat{\mathcal{U}} \end{pmatrix},$$

et où les vecteurs $\Delta\mathbf{W}$ et \mathcal{L}_v appartiennent à $\mathbb{R}^{N_{ddt}}$, ils sont tels que :

$$\Delta\mathbf{W} = \begin{pmatrix} \Delta\tilde{\mathbf{W}} \\ \Delta\hat{\mathbf{W}} \end{pmatrix}, \quad \mathcal{L}_v = \begin{pmatrix} \tilde{\mathbf{L}}_v \\ \hat{\mathbf{L}}_v \end{pmatrix}.$$

On précise que la matrice de préconditionnement n'est nécessaire qu'à l'étape 1 pour obtenir les modes de bas ordre. Par ailleurs, le problème implicite simplifié (3.25) est bien équivalent au problème implicite complet de départ (3.15) à résoudre puisque $\mathcal{A} + \mathcal{B} = \mathbf{A}$. Ceci assure la consistance en temps de la méthode simplifiée utilisée.

Remarque 3.9. Dans la pratique, la valeur de l'entier p_s est choisie de façon à autoriser le gain en temps CPU le plus important. Pour les applications numériques de la partie II, il s'agit de la valeur $p_s = 0$. Toutefois, une analyse de sensibilité du gain est réalisée au paragraphe 7.5.3 page 181.

Définition 3.6. Dans la suite de ce travail, la résolution du système des équations de Navier-Stokes avec une discrétisation spatiale GD (relations (2.14), (2.20) et (2.29)) et une discrétisation temporelle explicite-implicite simplifiée (relations (3.13), (3.25) et (3.14)) est désignée par l'appellation « SIMP p_s ».

3.5.4 Double simplification : méthode sans jacobienne et découplage

L'idée est de regrouper les deux simplifications SJ et SIMP p_s précédentes dans un seul et même schéma.

Afin de garder les avantages des deux méthodes, certains points clés doivent être respectés. Concernant la méthode SJ, il est question de la construction des blocs diagonaux (et uniquement diagonaux) $\mathbf{A}_{\kappa+\kappa+}$ de la matrice implicite pour permettre le calcul de la matrice de préconditionnement. Concernant la méthode SIMP p_s , il est question du découplage des degrés de liberté et de la résolution du système linéaire en deux étapes.

En fait, il survient une difficulté dans la mise en place de la double simplification dès lors qu'apparaît l'évaluation d'un produit matrice-vecteur faisant intervenir les blocs extra-diagonaux de la matrice implicite. Cela concerne d'une part, l'étape 1 avec le produit de la matrice $\tilde{\mathbf{A}}$ avec un vecteur appelé depuis l'algorithme GMRes du système (3.23), et d'autre part, l'étape 2 avec le produit $\mathbf{C}\Delta\tilde{\mathbf{W}}^{(n+1)}$ lors de la résolution du système (3.24). Pour pallier cette difficulté, ces deux produits sont approchés par une différence finie au premier ordre, en suivant le principe de l'approche sans jacobienne.

Première étape. Le système linéaire implicite de Crank-Nicolson (3.23) est résolu avec la méthode GMRes pour laquelle le produit $\tilde{\mathbf{A}}\tilde{\mathbf{z}}$ est remplacé par le produit approché au premier ordre suivant :

$$\left(\tilde{\mathbf{A}}\tilde{\mathbf{z}}\right)^\# = \frac{1}{\Delta t}\tilde{\mathbf{M}}\tilde{\mathbf{z}} + \frac{1}{2\delta_{SJ}}\left(\tilde{\mathbf{L}}_v(\mathbf{W}_v^{(n)} + \delta_{SJ}\tilde{\mathbf{z}}) - \tilde{\mathbf{L}}_v(\mathbf{W}_v^{(n)})\right), \quad (3.26)$$

où $\tilde{\mathbf{M}}$ est la matrice de masse diagonale associée aux modes de bas ordre et de dimension $NN_{eq}N_{p_s} \times NN_{eq}N_{p_s}$.

Seconde étape. La résolution du système triangulaire (3.24) est inversé en remplaçant le produit $\mathbf{C}\Delta\tilde{\mathbf{W}}^{(n+1)}$ par le produit approché au premier ordre suivant :

$$\left(\mathbf{C}\Delta\tilde{\mathbf{W}}^{(n+1)}\right)^\# = \frac{1}{2\delta_{SJ}}\left(\hat{\mathbf{L}}_v(\mathbf{W}_v^{(n)} + \delta_{SJ}\Delta\tilde{\mathbf{W}}^{(n+1)}) - \hat{\mathbf{L}}_v(\mathbf{W}_v^{(n)})\right). \quad (3.27)$$

Définition 3.7. Dans la suite de ce travail, la résolution du système des équations de Navier-Stokes avec une discrétisation spatiale GD (relations (2.14), (2.20) et (2.29)) et une discrétisation temporelle explicite-implicite simplifiée (relations (3.13), (3.23) en utilisant (3.26), (3.24) en utilisant (3.27) et (3.14)) est désignée par l'appellation « SJ+SIMP p_s ».

3.6 Comparaison des simplifications

Ce paragraphe propose une comparaison des quatre discrétisations explicites-implicites en temps (COMP, SJ, SIMP p_s , SJ+SIMP p_s) en matière d'encombrement mémoire (§ 3.6.1) et de coût asymptotique en nombre d'opérations à virgule flottante (§ 3.6.2), deux aspects déterminants dans l'appréciation d'une méthode pour des applications réelles.

Cette comparaison suppose un grand nombre d'éléments et un degré de polynôme élevé. De plus, elle se concentre sur la résolution implicite et elle concerne les étapes de calculs les plus significatives. Enfin, les différentes observations sont principalement axées sur les trois méthodes simplifiées par rapport à la méthode COMP.

3.6.1 Encombrement mémoire

La matrice de préconditionnement ne nécessitant pas de stockage supplémentaire (cf. remarque 3.3), l'encombrement mémoire est surtout évalué sur la matrice implicite \mathbf{A} puisqu'elle demande la ressource mémoire la plus importante. Cet encombrement est récapitulé dans le tableau 3.1 par élément (N au total) pour chacune des méthodes explicites-implicites en temps. L'analyse du tableau amène les observations suivantes.

- Pour la méthode COMP, la matrice implicite complète \mathbf{A} est stockée, ce qui correspond à $NN_{eq}^2N_p^2$ termes non nuls pour les blocs diagonaux et à $Nn_fN_{eq}^2N_p^2$ termes non nuls pour les blocs extra-diagonaux.

- Pour la méthode SJ, seuls les blocs diagonaux de la matrice \mathbf{A} sont entièrement stockés comme la méthode COMP. Ceci constitue une importante économie de stockage par rapport à la méthode COMP. En outre, le second résidu visqueux $\mathbf{L}_v(\mathbf{W}_v^{(n)} + \delta_{SJ}\Delta\mathbf{W}^{(n+1)})$ de la relation (3.19) ne requiert qu'un stockage de $NN_{eq}N_p$ termes supplémentaires.
- Pour la méthode SIMP p_s , les blocs diagonaux de la matrice implicite sont entièrement stockés comme la méthode COMP. Quant aux blocs extra-diagonaux, le découplage des degrés de liberté conduit à l'économie des sous-matrices $\hat{\mathbf{A}}$ et \mathbf{B} , ce qui correspond à $Nn_fN_{eq}^2(N_p - N_{p_s})N_p$ termes ni calculés ni stockés. En revanche, l'étape supplémentaire de reconstruction demande le stockage du vecteur $\Delta\hat{\mathbf{W}}^{(n)}$, ce qui correspond à un surplus de stockage de $NN_{eq}(N_p - N_{p_s})$ termes non nuls.
- Pour la méthode SJ+SIMP p_s , seuls les blocs diagonaux de la matrice implicite et le second résidu visqueux $\tilde{\mathbf{L}}_v(\mathbf{W}_v^{(n)} + \delta_{SJ}\tilde{\mathbf{z}}^{(n+1)})$ de la relation (3.26) sont entièrement stockés comme pour la méthode SJ. Cependant ce dernier est réduit de $NN_{eq}(N_p - N_{p_s})$ termes par rapport à la méthode SJ. Par ailleurs, l'étape de reconstruction demande le stockage du vecteur $\Delta\hat{\mathbf{W}}^{(n)}$ comme la méthode SIMP p_s , mais également le stockage supplémentaire du troisième résidu $\hat{\mathbf{L}}_v(\mathbf{W}_v^{(n)} + \delta_{SJ}\Delta\tilde{\mathbf{W}}^{(n+1)})$ de la relation (3.27).

Finalement, la méthode SJ présente l'économie de stockage la plus importante, suivie par les méthodes SJ+SIMP p_s puis SIMP p_s qui permettent également des économies importantes.

| | COMP | SJ | SIMP p_s | SJ+SIMP p_s |
|---|---------------------------|------------------|---|-------------------------|
| Matrice \mathbf{A} | $(n_f + 1)N_{eq}^2 N_p^2$ | $N_{eq}^2 N_p^2$ | $N_{eq}^2 N_p^2 + n_f N_{eq}^2 N_p N_{p_s}$ | $N_{eq}^2 N_p^2$ |
| 2^{nd} résidu \mathbf{L}_v cf. (3.19) & (3.26) | – | $N_{eq}N_p$ | – | $N_{eq}N_{p_s}$ |
| Vecteur $\Delta\hat{\mathbf{W}}^{(n)}$ | – | – | $N_{eq}(N_p - N_{p_s})$ | $N_{eq}(N_p - N_{p_s})$ |
| 3^e résidu \mathbf{L}_v cf. (3.27) | – | – | – | $N_{eq}(N_p - N_{p_s})$ |

Tab. 3.1 – Evaluation asymptotique de l'encombrement mémoire par élément pour chaque méthode explicite-implicite.

3.6.2 Nombre d'opérations à virgule flottante

Le coût asymptotique, pour N et p grands, en nombre d'opérations à virgule flottante est uniquement évalué pour les étapes de calculs les plus significatives : construction du vecteur résidu des termes de diffusion \mathbf{L}_v , constructions de la matrice implicite \mathbf{A} et de la matrice de préconditionnement \mathbf{P} , inversion GMRes, et enfin reconstruction des modes d'ordre élevé.

| | COMP | SJ | SIMP p_s | SJ+SIMP p_s |
|--|-------------------------------|-------------------------------|---|-------------------------------------|
| Résidu \mathbf{L}_v | $\alpha_1 N_g N_{eq} N_p$ | $\alpha_1 N_g N_{eq} N_p$ | $\alpha_1 N_g N_{eq} N_p$ | $\alpha_1 N_g N_{eq} N_p$ |
| Matrice \mathbf{A} | | | | |
| Bloc diagonal | $\alpha_2 N_g N_{eq}^2 N_p^2$ | $\alpha_2 N_g N_{eq}^2 N_p^2$ | $\alpha_2 N_g N_{eq}^2 N_p^2$ | $\alpha_2 N_g N_{eq}^2 N_p^2$ |
| Bloc extra-diagonal | $\alpha_3 N_g N_{eq}^2 N_p^2$ | 0 | $\alpha_3 N_g N_{eq}^2 N_p N_{p_s}$ | 0 |
| Matrice \mathbf{P} | $\frac{2}{3} N_{eq}^3 N_p^3$ | $\frac{2}{3} N_{eq}^3 N_p^3$ | $\frac{2}{3} N_{eq}^3 N_{p_s}^3$ | $\frac{2}{3} N_{eq}^3 N_{p_s}^3$ |
| Inversion GMRes | | | | |
| $\mathbf{A}\mathbf{z}$ | $2(n_f + 1) N_{eq}^2 N_p^2$ | $\alpha_1 N_g N_{eq} N_p$ | $2(n_f + 1) N_{eq}^2 N_{p_s}^2$ | $\alpha_1 N_g N_{eq} N_p$ |
| $\mathbf{P}^{-1}\mathbf{z}$ | $2N_{eq}^2 N_p^2$ | $2N_{eq}^2 N_p^2$ | $2N_{eq}^2 N_{p_s}^2$ | $2N_{eq}^2 N_{p_s}^2$ |
| Reconstruction | | | | |
| $\mathbf{B}\Delta\hat{\mathbf{W}}^{(n)}$ | 0 | 0 | $2N_{eq}^2 (N_p - N_{p_s}) N_{p_s}$ | $2N_{eq}^2 (N_p - N_{p_s}) N_{p_s}$ |
| $\mathbf{C}\Delta\tilde{\mathbf{W}}^{(n+1)}$ | 0 | 0 | $2(n_f + 1) N_{eq}^2 (N_p - N_{p_s}) N_{p_s}$ | $\alpha_1 N_g N_{eq} N_p$ |
| $\hat{\mathbf{U}}\Delta\hat{\mathbf{W}}^{(n)}$ | 0 | 0 | $N_{eq}^2 (N_p - N_{p_s})^2$ | $N_{eq}^2 (N_p - N_{p_s})^2$ |
| $(\hat{\mathbf{L}} + \hat{\mathbf{D}})^{-1}\Delta\hat{\mathbf{W}}^{(n+1)}$ | 0 | 0 | $N_{eq}^2 (N_p - N_{p_s})^2$ | $N_{eq}^2 (N_p - N_{p_s})^2$ |

Tab. 3.2 – Evaluation asymptotique du nombre d’opérations à virgule flottante par élément pour chaque méthode explicite-implicite.

| | SJ | SIMP p_s | SJ+SIMP p_s |
|-----------------------|---|---|---|
| Résidu \mathbf{L}_v | 1 | 1 | 1 |
| Matrice \mathbf{A} | $1 + \frac{\alpha_3}{\alpha_2}$ | $\frac{1 + \alpha_3/\alpha_2}{1 + \alpha_3 N_{p_s}/\alpha_2 N_p}$ | $1 + \frac{\alpha_3}{\alpha_2}$ |
| Matrice \mathbf{P} | 1 | $\left(\frac{N_p}{N_{p_s}}\right)^3$ | $\left(\frac{N_p}{N_{p_s}}\right)^3$ |
| Inversion GMRes | $\frac{2(n_f+2)N_{eq}N_p}{\alpha_1 N_g + 2N_{eq}N_p}$ | $\left(\frac{N_p}{N_{p_s}}\right)^2$ | $\frac{2(n_f+2)N_{eq}N_p}{\alpha_1 N_g + 2N_{eq}N_{p_s}^2/N_p}$ |

Tab. 3.3 – Evaluation asymptotique du gain théorique pour chaque méthode explicite-implicite simplifiée à partir du nombre d’opérations à virgule flottante (*cf.* tableau 3.2).

Le tableau 3.2 récapitule les coûts asymptotiques par élément (N au total) pour chacune des méthodes en temps et le tableau 3.3 indique le gain théorique asymptotique qui en découle. Ce gain théorique est simplement calculé en effectuant le rapport du nombre d'opérations à virgule flottante de la méthode COMP sur celui d'une autre méthode. Les nombres α_1 , α_2 et α_3 introduits sont tels que $\alpha_{1 \leq i \leq 3} = \mathcal{O}(1)$. L'analyse conjointe de ces deux tableaux amène les observations suivantes.

- Pour le calcul des résidus explicites \mathbf{L}_v , le nombre d'opérations est identique pour toutes les méthodes.
- La construction de la matrice implicite représente le nombre d'opérations le plus conséquent pour toutes les méthodes lorsque p devient grand [35], comme on l'a déjà vu au paragraphe 3.4.3. Les deux méthodes SJ et SJ+SIMP p_s présentent un gain très important pour la construction de \mathbf{A} dû à la construction des seuls blocs-diagonaux. Pour ces mêmes blocs, la méthode SIMP p_s présente un gain intermédiaire, évalué à $(N_p/N_{p_s})^2$, entre la méthode COMP et les méthodes SJ et SJ+SIMP p_s .
- Pour la construction de la matrice de préconditionnement, le nombre d'opérations est réduit pour les deux méthodes SIMP p_s et SJ+SIMP p_s , qui sont donc moins sensibles à cette étape. En effet, le préconditionnement est uniquement nécessaire pour la résolution du système linéaire (3.23), laquelle concerne les modes de bas ordre. La taille de ce système est alors de $NN_{eq}N_{p_s}$ contre $NN_{eq}N_p$ pour les méthodes COMP et SJ, et le gain est évalué à $(N_p/N_{p_s})^3$.
- Concernant l'inversion GMRes, les deux opérations matrice-vecteur $\mathbf{A}\mathbf{z}$ et $\mathbf{P}^{-1}\mathbf{z}$ pour un vecteur \mathbf{z} quelconque représentent le nombre d'opérations le plus important par itération. Ce nombre est réduit pour les trois méthodes SJ, SIMP p_s et SJ+SIMP p_s . D'une part pour la méthode SIMP p_s , le gain est de $(N_p/N_{p_s})^2$ pour les mêmes raisons que précédemment : le système linéaire est de taille réduite $NN_{eq}N_{p_s}$. D'autre part, pour les méthodes SJ et SJ+SIMP p_s , le gain est positif puisque le produit $\mathbf{A}\mathbf{z}$ est approché par celui simplifié (3.19) ou (3.26) et correspond concrètement à la construction de la matrice de masse et d'un second résidu visqueux. On précise que le coût du produit $\mathbf{A}\mathbf{z}$ de la méthode SJ+SIMP p_s est moindre que celui de la méthode SJ puisqu'il bénéficie du système linéaire réduit de la méthode SIMP p_s .
- La reconstruction concerne les méthodes SIMP p_s et SJ+SIMP p_s et représente un coût modéré. Notamment les opérations les plus conséquentes sont le produit matrice-vecteur $\hat{\mathbf{U}}\Delta\hat{\mathbf{W}}^{(n)}$ et la résolution du système triangulaire $(\hat{\mathbf{L}} + \hat{\mathbf{D}})^{-1}\Delta\hat{\mathbf{W}}^{(n+1)}$. Pour la méthode SJ+SIMP p_s , le calcul approché du produit $\mathbf{C}\Delta\tilde{\mathbf{W}}^{(n+1)}$ est également à prendre en considération car il correspond à la construction d'un troisième résidu.

Finalement, chacune des deux méthodes SJ et SIMP p_s représentent des gains théoriques importants à la fois sur la construction de la matrice implicite et sur la résolution du système linéaire par une méthode itérative préconditionnée. De ce fait, des gains prometteurs sont attendus de la méthode SJ+SIMP p_s regroupant les deux stratégies sans jacobienne et du découplage des degrés de liberté.

3.7 Conclusion partielle

Les points suivants constituent le bilan de ce chapitre.

- Le schéma numérique discret du système des équations d'Euler ou de Navier-Stokes a été établi au moyen :
 - soit d'une intégration temporelle totalement explicite (EXPL) ;
 - soit d'une intégration temporelle explicite-implicite, sans simplification (COMP) ou avec simplification (SJ, SIMP_{p_s} ou $\text{SJ}+\text{SIMP}_{p_s}$).
- Dans le cas d'une intégration temporelle explicite, le schéma numérique retenu est celui RK2 de Heun [63, 52] d'ordre deux ou celui RK3 de Shu-Osher [155] d'ordre trois.
- Dans le cas d'une intégration temporelle explicite-implicite, le schéma numérique utilise la décomposition d'opérateurs de Strang [160] d'ordre deux, pour résoudre séparément la discrétisation des termes de convection et celle des termes de diffusion. Les schémas numériques associés à ces deux discrétisations sont :
 - un schéma explicite d'ordre deux de Heun [63, 52] sur les termes de convection, afin de ne pas induire trop de dissipation des mécanismes de transport ;
 - un schéma implicite d'ordre deux de Crank-Nicolson [34, 36] sur les termes de diffusion, afin de s'affranchir de la condition de stabilité de type Von Neumann. Ce schéma est résolu avec une méthode itérative préconditionnée de GMRes à redémarrages.
- Une étude des coûts en temps CPU a démontré la nécessité de simplifier la discrétisation temporelle implicite en vue d'applications réelles, ainsi :
 - pour $p \leq 2$ le coût le plus important concerne l'étape d'inversion GMRes préconditionnée, plus globalement ce coût diminue avec p ;
 - pour $p \geq 3$ le coût le plus important concerne l'étape de construction de la matrice implicite, il est dû au terme de relèvement global dans la mise en place du schéma BR2, plus globalement ce coût augmente avec p .
- Trois simplifications du schéma implicite COMP ont été mises en œuvre :
 - celle sans jacobienne (SJ) approchant le calcul complet de la matrice jacobienne de dérivation du résidu discret par rapport au vecteur solution des termes visqueux par une différence finie d'ordre un ;
 - celle découplant les degrés de liberté entre les éléments du maillage (SIMP_{p_s}) ;
 - celle regroupant les deux simplifications précédentes ($\text{SJ}+\text{SIMP}_{p_s}$).
- Une comparaison théorique des méthodes sur l'encombrement mémoire et sur le nombre d'opérations à virgule flottante a été faite. Elle laisse présager des gains importants sur ces deux aspects pour l'ensemble des simplifications.

A présent, il s'agit de valider le schéma discret. C'est l'objet de la partie II suivante.

Deuxième partie

Schéma numérique : de la mise en
œuvre à la validation et à
l'évaluation des performances

Chapitre 4

Mise en œuvre du schéma discret

Ce chapitre est consacré à la mise en œuvre du schéma discret du système des équations de Navier-Stokes développé dans la partie I précédente. Il est divisé en deux avec la description du code de calcul dédié développé d'une part, et celle des cas tests envisagés servant à valider le code d'autre part. Ces descriptions sont accompagnées de schémas pour une meilleure vue d'ensemble et un meilleur suivi des prochains chapitres.

4.1 Avant-Propos

Dans le cadre, par exemple, des programmes américains de l'AFOSR [73] et des projets européens Adigma [96] et Idihom [77], nombre d'institutions ont développé (ou amélioré) leur propre code de calcul pour la CFD, en y incluant des méthodes d'ordre élevé dont celle de GD. Parmi ces codes, il vient par exemple¹ :

- « AeroSol » de l'Inria [72],
- « COOLFluiD » du VKI [74],
- « Dune » qui est libre [76],
- « Migale » des universités de Bergame et de Brescia [5],
- « Padge » du DLR [57],
- « ProjectX » du MIT [106],
- « XFlow » de l'université du Michigan [81].

Cette liste n'est pas exhaustive mais elle reflète l'intérêt de la communauté scientifique pour développer une alternative aux codes industriels actuels d'ordre deux.

Concernant plus particulièrement l'Onera, différents codes de calcul existent. Dans le domaine de l'aérodynamique appliquée il vient, en parallèle ou suite aux travaux de thèse de Berde [14], Drozo [41], Burbeau [19] et Léger [102], les codes de calcul « Spot » [16] et « Aghora » [140] pour les problèmes d'aérodynamique externe et interne, ou le code « Space » [103] pour les problèmes d'aéroacoustique.

Le code développé dans ce travail est un code de développement dédié, principalement issu d'une généralisation du code Spot. La première partie de ce chapitre, paragraphe 4.2, présente une vue d'ensemble du code et de ses principales étapes de programmation. La seconde partie, paragraphe 4.3, décrit les cas tests exploités et leur rôle dans la validation du code, et par conséquent du schéma numérique discret associé.

1. Pour la signification des acronymes, voir la nomenclature page xvii.

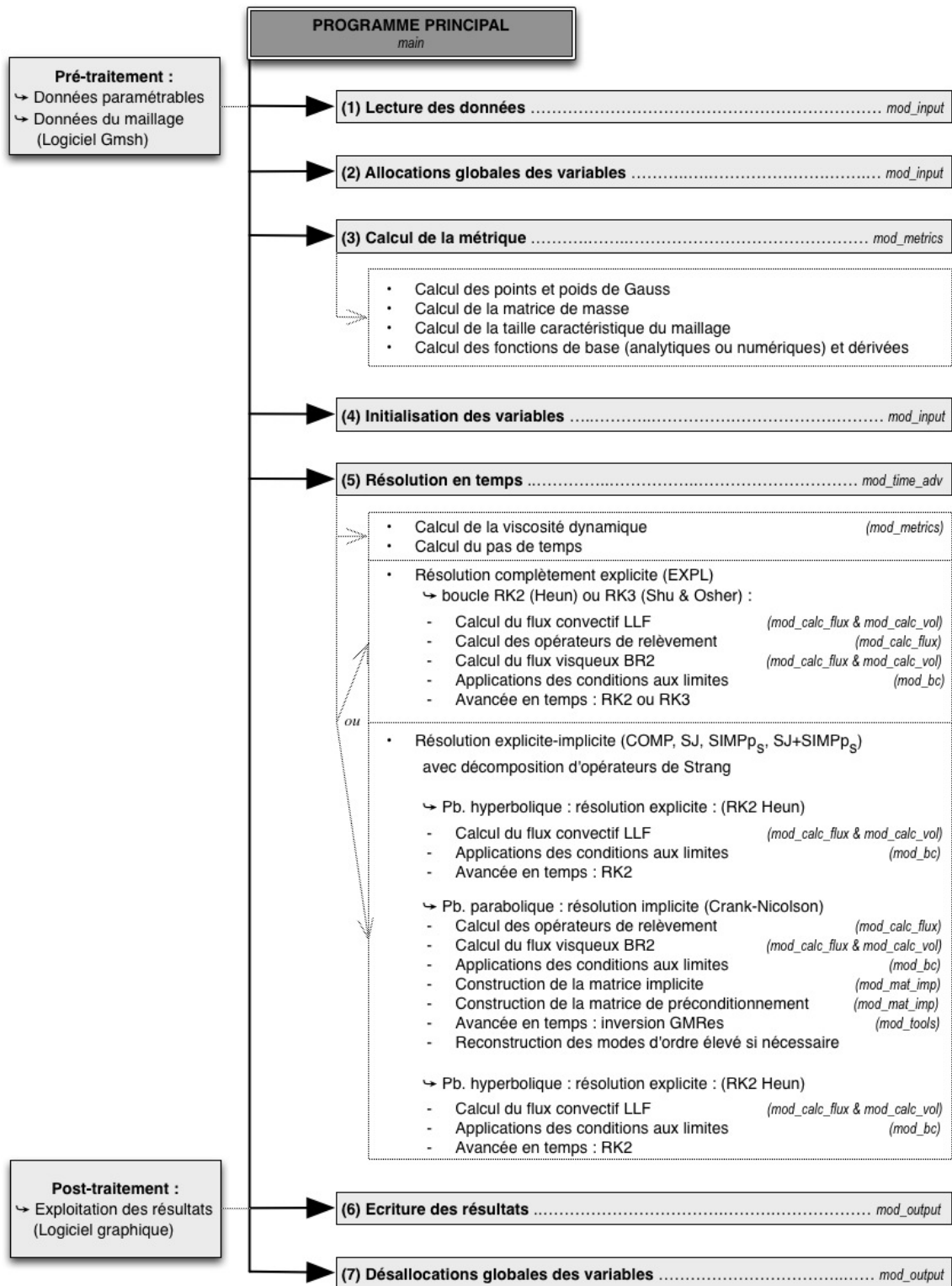


Fig. 4.1 – Schéma résumant les étapes de programmation du code de calcul développé, ainsi que le nom des modules concernés. Les sept étapes clés sont dans les cadres gris clairs et les sous-étapes sont dans les cadres blancs. Abréviation : problème (Pb.).

4.2 Code de calcul dédié

Les calculs numériques sont effectués au moyen d'un code informatique dédié. Les paragraphes suivants permettent d'obtenir une vue d'ensemble de ce code avec un rapide historique (§ 4.2.1) et une synthèse des étapes de programmation (§ 4.2.2).

4.2.1 Historique

Le présent code de développement étant l'aboutissement d'une généralisation du code Spot, ce dernier est succinctement décrit ci-dessous.

Le code Spot [16] est un code Onera de développement, écrit en Fortran 90 par Michel Borrel et Juliette Ryan dans les années 1990. Il est destiné à la résolution du système des équations de Navier-Stokes bidimensionnelles pour un fluide compressible. Il utilise une discrétisation GD en espace et une discrétisation RK explicite en temps. Les flux numériques présents sont le flux convectif LLF et celui visqueux « elastoplast » [16] proche des méthodes de recouvrement de van Leer [169, 170].

Le code développé dans ce travail est né en 2010 à partir du code Spot. Il est écrit en Fortran 90, sous forme de modules (*cf.* § 4.2.2 pour cet aspect) et est séquentiel. Concernant la forme, le squelette des grandes étapes de programmation de Spot est conservé : lectures/écritures des paramètres d'entrées/sorties, enchaînement des boucles de calculs des intégrales (volumes, surfaces dont les frontières). Concernant le fond, le flux LLF est conservé, tandis que les fonctions de base, le flux BR2, la généralisation à un degré de polynôme quelconque, la prise en compte de géométries courbes ou la discrétisation temporelle explicite-implicite ont été ajoutés.

4.2.2 Synthèse des principales étapes de programmation

Le présent code de calcul considère le système des équations d'Euler ou de Navier-Stokes pour les écoulements bidimensionnels de fluide compressible, en utilisant la discrétisation décrite dans la partie I. Il comporte plusieurs étapes principales de programmation (§ 4.2.2.1) précédées d'une phase de pré-traitement où sont générés des fichiers de paramètres, et suivies d'une phase de post-traitement où sont analysés les fichiers de résultats (§ 4.2.2.2).

4.2.2.1 Sept étapes principales de programmation

Les étapes sont détaillées par ordre d'enchaînement avec le schéma de la figure 4.1. Ce schéma permet de bien comprendre comment sont combinées les méthodes numériques en espace et en temps utilisées.

Ainsi, à partir des fichiers de paramètres en entrée du code, un scénario de calcul se déroule en sept étapes clés contenues dans le fichier principal nommé *main* :

1. la lecture des données,
2. les allocations globales,
3. le calcul de la métrique,
4. l'initialisation des variables,
5. la résolution en temps,
6. l'écriture des résultats,
7. les désallocations globales.

Ces étapes, ainsi que leurs sous-étapes, sont réalisées grâce à un certain nombre de fonctions regroupées en modules de préfixe *mod* comme l'illustre le schéma de la figure 4.1. Cette organisation modulaire du code présente de nombreux avantages de programmation dont une meilleure lisibilité ou encore une diminution du risque d'erreur de programmation.

4.2.2.2 Etapes de lecture et d'écriture des données

Ce paragraphe détaille les étapes de lecture et d'écriture des données afin d'aborder certaines généralités sur la présentation des résultats numériques des prochains chapitres.

La lecture des paramètres d'entrée est associée à la phase de pré-traitement. Elle est fonction de l'écoulement simulé et concerne à la fois des données paramétrables et du maillage. D'un côté, les données paramétrables sont de natures physique ou numérique. Elles concernent : le cas test, la méthode numérique en espace et en temps et l'écriture des fichiers de sortie. D'un autre côté, les données du maillage sont générées avec le logiciel libre Gmsh [54, 55] présentant une interface graphique. Elles concernent : les nœuds, les faces (arêtes), les triangles et les conditions aux limites à imposer. En outre, Gmsh permet de générer des maillages curvilignes.

L'écriture des paramètres de sortie est associée à la phase de post-traitement. Elle est également fonction de l'écoulement simulé. Les données concernent : le vecteur des variables conservatives à un temps donné, la valeur de certaines quantités locales (par exemple aux parois) et l'historique temporel de certaines quantités. Elles sont exploitées avec un outil de visualisation graphique.

Remarque 4.1. *Toutes les illustrations, en deux dimensions d'espace, de la solution numérique sur son domaine de calcul utilisent une visualisation améliorée, où chaque triangle du domaine est divisé en neuf sous-triangles égaux et la solution numérique est reconstruite en chacun des neufs barycentres associés.*

4.3 Cas tests exploités

Ce paragraphe donne une vue d'ensemble des cinq cas tests utilisés avec le code de calcul (§ 4.3.1) et une vue d'ensemble de la démarche logique suivie pour valider ce dernier de pair avec le schéma numérique (§ 4.3.2).

4.3.1 Synthèse des écoulements associés aux cas tests

Au total, cinq cas tests sont utilisés pour valider le code. La désignation de chacun d'entre eux, l'écoulement associé, les maillages utilisés et l'utilité dans la démarche logique de validation et d'estimation des performances sont résumés dans le tableau 4.1. Ces cas tests concernent :

1. l'écoulement interne, stationnaire et subsonique d'un fluide parfait dans un canal de type convergent-divergent comme dans [176, 83];

2. l'écoulement interne, instationnaire et subsonique d'une paire de tourbillons visqueux et contra-rotatifs se dirigeant perpendiculairement à une paroi plane comme dans [87] ;
3. l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour du profil d'aile NACA0012 [78, 84] comme dans [96, 40] ;
4. l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour du profil d'aile SD7003 [150] comme dans [166, 113] ;
5. l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour et en aval d'un cylindre comme dans [79].

Parmi ces cinq cas tests, un seul, le cas test 1, est décrit par le système des équations d'Euler, tandis que les quatre suivants sont décrits par le système des équations de Navier-Stokes. Ces derniers sont classiques des problématiques en aérodynamique, en particulier à cause des phénomènes physiques mis en jeu : formation de nappes de vorticit  (cas test 2), d collement de couche limite laminaire (cas tests 2, 3, 4 et 5), ou encore d tachements tourbillonnaires (cas tests 3, 4 et 5).

| | | | | | | |
|---------------|-------------------------------|----------------------------------|-----------------------------|---|----------------------|-------------------------------|
| D signation | Cas test | 1 | 2 | 3 | 4 | 5 |
| | Nom | Canal | Int raction tourb.-paroi | NACA0012 | SD7003 | Cylindre |
| Description | R f rence (★) | [176, 83] | [87] | [96, 40] | [166, 113] | [79] |
| | Description | Chap. 5,   5.1 | Chap. 5,   5.2 | Chap. 6 ,   6.3.2 | Chap. 6,   6.4.1 | Chap. 6,   6.4.2 |
| Ecoulement 2D | Type | Interne | Interne | Externe | Externe | Externe |
| | Equation | E | NS | NS | NS | NS |
| | Cond. lim. | Paroi Non-r flex. Sym trie | Paroi | Paroi Non-r flex. | Paroi Non-r flex. | Paroi Non-r flex. |
| | Solution | Statio. | Insta. | Insta. | Insta. | Insta. |
| | t | ∞ | 1 | 60 | 25 | 150 |
| | M_∞ | 0,5 | 0,01 | 0,5 | 0,2 | 0,2 |
| Maillage | Origine | R cup r  | Cr e | Cr e | Cr e | Cr e |
| | Nombre | 12 | 8 | 3 | 1 | 1 |
| | m | 1 & 4 | 1 | 4 | 4 | 2 |
| | RA | – | 1, 4, 20, 143 & 600 | 444, 1 685 & 2 125 | 11 676 | 630 |
| D marche | Utilit  | Validation spatiale | | Validation et performance temporelle | | |
| | Comparaison en plus de (★) | [65, 46, 114] [107, 178] | [24, 97] | [135] | [48] | [146, 163, 179] [120, 115] |

Tab. 4.1 – R capitulation des cinq cas tests, avec t le temps physique de simulation, m le degr  d'approximation de la g om trie (*cf.* d finition 2.11 page 54) et RA le rapport d'aspect global (*cf.* relation (2.3) page 37). Abr viations : Tourbillons (tourb.), Chapitre (Chap.), Bidimensionnel (2D), Euler (E), Navier-Stokes (NS), Conditions aux limites (Cond. lim.), Non-reflexion (Non-r flex.), Stationnaire (Statio.), Instationnaire (Insta.).

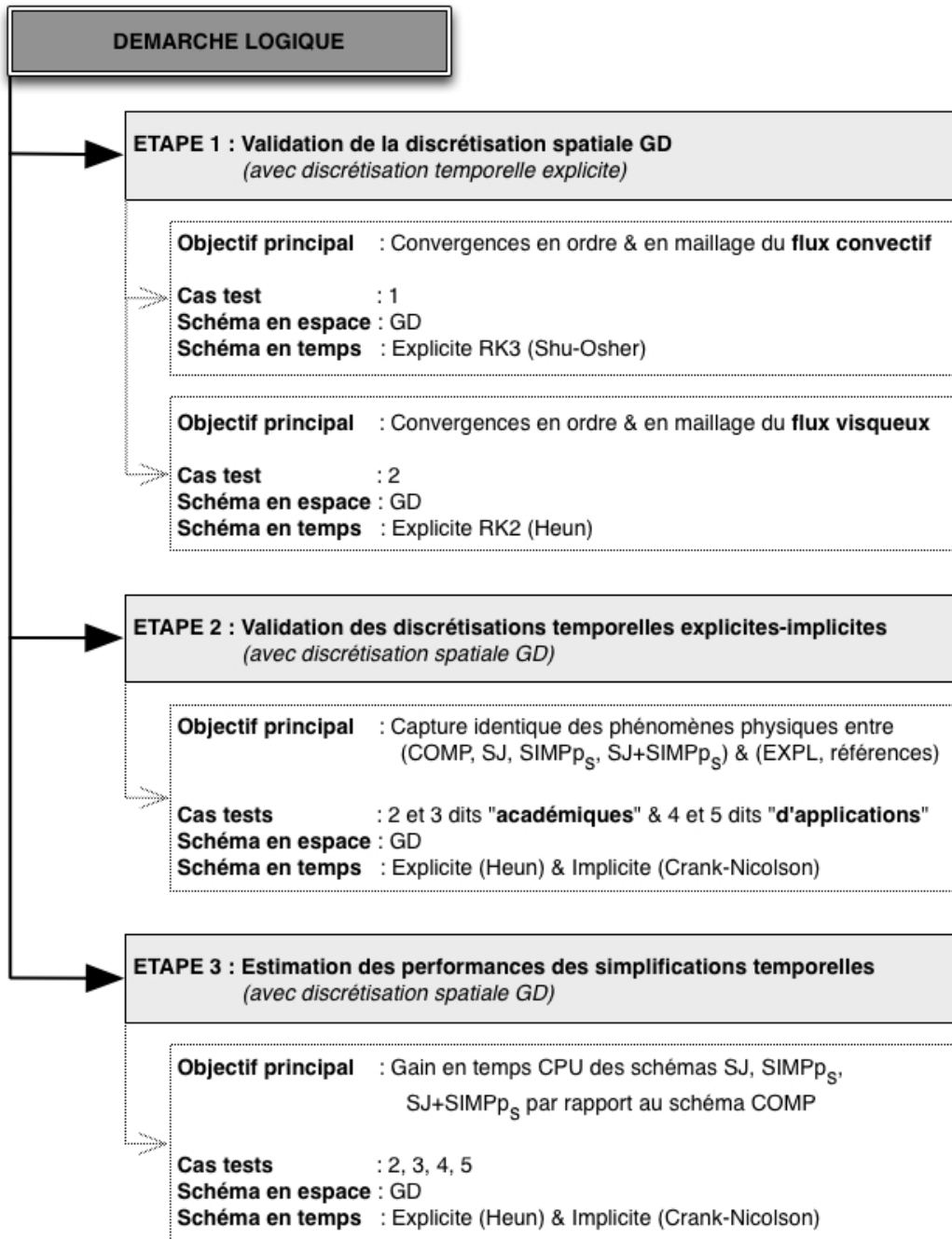


Fig. 4.2 – Schéma expliquant la démarche logique de validation du code de calcul développé.

Sachant que dans ce travail on s'intéresse à des écoulements à convection dominante, les cas tests 2 et 3 sont considérés comme académiques, surtout en raison de la géométrie du problème, et les cas tests 4 et 5 sont considérés comme plus applicatifs, surtout en raison de la couche limite qui est très fortement décollée. De plus, pour le cas test 4, le nombre de Reynolds est important, et pour le cas test 5, les effets visqueux jouent un rôle prédominant dans la physique du problème.

4.3.2 Démarche logique

La simulation des différents cas tests envisagés respecte une démarche logique servant à valider et évaluer le schéma numérique développé dans ce travail. Cette démarche comporte trois étapes détaillées par le schéma de la figure 4.2, dont certaines concernent plusieurs cas tests comme cela est décrit dans le bas du tableau 4.1.

4.3.2.1 Etape 1

La première étape est la validation de la discrétisation spatiale de type GD associée à une discrétisation temporelle explicite (EXPL) de type RK, et ce *via* les cas tests 1 et 2. En particulier, la validation de la discrétisation des termes de flux convectifs, relation (2.20), est réalisée *via* le cas test 1, et celle des termes de flux visqueux, relation (2.29), est réalisée *via* le cas test 2. Ces validations consistent en une comparaison des résultats obtenus sur certaines grandeurs caractéristiques des écoulements simulés avec ceux issus de la littérature.

4.3.2.2 Etape 2

La seconde étape est la validation des quatre méthodes explicites-implicites en temps, COMP, SJ, $\text{SIMP}p_s$ et $\text{SJ}+\text{SIMP}p_s$, associées à la discrétisation GD en espace de l'étape précédente, et ce à la fois *via* les cas tests académiques 2 et 3 et *via* les cas tests d'applications 4 et 5. Pour ce faire, l'idée est de comparer les captures des phénomènes physiques de ces quatre méthodes avec celles des références. En particulier, les résultats du schéma EXPL de l'étape 1 sont utilisés pour le cas test 2, et ceux issus de la littérature sont utilisés pour les cas tests 3, 4 et 5.

4.3.2.3 Etape 3

La troisième étape est l'estimation des performances des méthodes explicites-implicites simplifiées en temps, SJ, $\text{SIMP}p_s$ et $\text{SJ}+\text{SIMP}p_s$, et ce *via* les cas tests 2, 3, 4 et 5. L'estimation consiste à comparer les résultats en temps CPU entre le schéma COMP et les schémas SJ, $\text{SIMP}p_s$ et $\text{SJ}+\text{SIMP}p_s$, en particulier concernant le gain sur les étapes de calculs intermédiaires et le gain sur la résolution globale.

A présent, il s'agit de mettre en œuvre les étapes 1, 2 et 3 de validation du code interne de développement. C'est l'objet des trois prochains chapitres 5, 6 et 7.

Chapitre 5

Validation de la discrétisation spatiale Galerkin Discontinue

Ce chapitre se concentre sur l'étape 1 de validation de la discrétisation spatiale GD associée à une discrétisation temporelle explicite (EXPL). Deux cas tests simulant des écoulements internes et subsoniques sont alors considérés.

D'une part (§ 5.1), le cas test 1, d'un écoulement de fluide parfait dans un canal de type convergent-divergent permet la validation de la discrétisation des termes de flux convectifs.

D'autre part (§ 5.2), le cas test 2, d'un écoulement d'une paire de tourbillons visqueux et contra-rotatifs impactant perpendiculairement une paroi plane permet la validation de la discrétisation des termes de flux visqueux.

5.1 Écoulement dans un canal convergent-divergent

Pour valider la discrétisation spatiale des termes de flux convectifs (relation (2.20)), on simule l'écoulement interne, stationnaire et subsonique d'un fluide parfait dans un canal de type convergent-divergent (cas test 1) comme proposé par le premier atelier sur les méthodes d'ordre élevé qui s'est tenu en janvier 2012 à Nashville aux Etats-Unis [176, 83]. C'est un problème décrit par le système des équations d'Euler donné en (1.10).

A travers ce cas test, deux points sont particulièrement mis en avant. D'un côté, on montre l'importance de prendre en compte les parois courbes d'un domaine de calcul. D'un autre côté, on utilise les résultats de l'atelier pour juger la compétitivité de différentes méthodes GD vis-à-vis d'autres méthodes d'ordre élevé comme celles de différences finies, d'éléments finis, de volumes finis, *etc.*

La présentation du cas test est articulée comme suit. Tout d'abord, le paragraphe 5.1.1 revient sur la pertinence de ce cas test pour valider une méthode numérique. Puis, le paragraphe 5.1.2 présente la géométrie du domaine de calcul, les maillages et les paramètres physiques et numériques utilisés. Enfin, le paragraphe 5.1.3 synthétise les résultats numériques permettant de valider la discrétisation spatiale retenue pour les termes de convection.

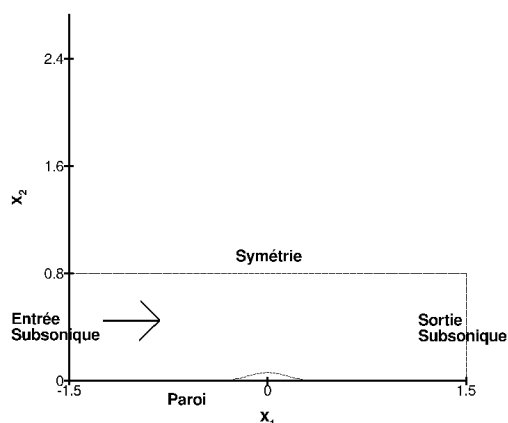


Fig. 5.1 – (*Canal*). Géométrie, conditions aux limites du domaine de calcul et condition initiale.

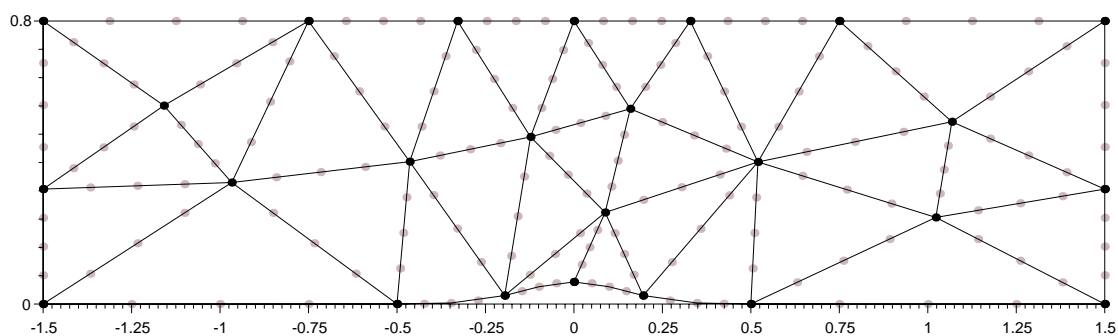


Fig. 5.2 – (*Canal*). Vue globale du maillage grossier tel que $N = 32$: linéaire ($m = 1$) décrit par les points noirs uniquement, ou curviligne ($m = 4$) décrit par les points noirs et gris.

| N | Nb. points total | | Nb. points sur Γ_p | | h | RA |
|--------|------------------|-------------|---------------------------|-------------|----------------------|------|
| | ($m = 1$) | ($m = 4$) | ($m = 1$) | ($m = 4$) | | |
| 32 | 25 | 193 | 7 | 25 | $h_1 \approx 0,1$ | 4 |
| 128 | 81 | 705 | 13 | 49 | $h_2 \approx h_1/2$ | 4 |
| 512 | 289 | 2 689 | 25 | 97 | $h_3 \approx h_1/4$ | 4 |
| 2 048 | 1 089 | 10 497 | 49 | 193 | $h_4 \approx h_1/8$ | 4 |
| 8 192 | 4 225 | 41 473 | 97 | 385 | $h_5 \approx h_1/16$ | 4 |
| 32 768 | 16 641 | 164 865 | 193 | 769 | $h_6 \approx h_1/32$ | 4 |

Tab. 5.1 – (*Canal*). Description des maillages utilisés. Abréviation : Nombre (Nb.).

5.1.1 Avant-propos

D'une façon générale, l'écoulement bidimensionnel et stationnaire de fluide parfait dans un canal convergent-divergent est représentatif de nombreux écoulements internes comme ceux dans les tuyères. Lors d'une première approche, ces écoulements dans les conduits sont souvent considérés isentropiques, c'est-à-dire adiabatiques (sans échange de chaleur) et réversibles (sans frottement si le conduit est suffisamment court). Ils sont décrits par le système des équations d'Euler et leur simulation s'impose de plus en plus comme un exemple académique de vérifications d'une méthode numérique [153, chapitre 10] ou [65, chapitre 11]. En effet, cet exemple bénéficie d'une géométrie simple et symétrique (suivant les axes horizontal et vertical), et une solution précise peut être obtenue sur un maillage raffiné. De plus, l'évaluation de l'erreur globale en norme L^2 sur l'entropie (*cf.* relation (5.1)) est un excellent indicateur des effets de la dissipation numérique révélant la précision de la méthode utilisée.

On souhaite indiquer que l'écoulement dans un canal convergent-divergent proposé par l'atelier « méthodes d'ordre élevé » en 2012 [176, 83] est une variante de celui plus répandu autour d'une bosse dans un canal droit initié par l'atelier « GAMM »¹ sur les écoulements transsoniques en 1979 [144]. La différence repose, en partie, sur la condition aux limites appliquée à la paroi supérieure : symétrie ou glissement.

Concernant la littérature des méthodes GD, on a relevé que si plusieurs choix existent pour la fonction décrivant la courbure de la bosse : gaussienne dite de « Ni » [109, 45, 128, 69], autre gaussienne [178, 46], sinusöide [167, 8], ou demi-cercle [107, 132, 45]; ce cas test est souvent utilisé pour : évaluer les performances d'une méthode pour les fluides compressibles et incompressibles [45, 128], tester les capacités de la méthode pour la capture de chocs [69], juger l'efficacité de l'intégration temporelle [132, 109, 8], ou vérifier la convergence spatiale en maillage et celle en ordre grâce à l'évaluation de l'erreur sur l'entropie [107, 178, 167, 46, 114].

Dans ce travail, on s'intéresse particulièrement à ces deux convergences en maillage et en ordre afin de valider la discrétisation spatiale GD retenue pour les termes de flux convectifs.

5.1.2 Description du cas test

5.1.2.1 Géométrie du domaine de calcul et maillages

La figure 5.1 décrit la géométrie du domaine de calcul, les conditions aux limites considérées et la condition initiale. Le domaine est tel que $\Omega \subset [-1, 5; 1, 5] \times [0; 0, 8]$. Le canal est rétréci au col ($x_1 = 0$). Sa partie inférieure est décrite par la fonction gaussienne $x_2 = 0,0625 \exp(-25x_1^2)$ dont le maximum est en $(x_1; x_2) = (0; 0,0625)$. Les conditions aux limites appliquées sont une condition de glissement (paroi immobile) en bas du domaine, une condition de symétrie en haut, une condition d'entrée subsonique à gauche et une condition de sortie subsonique à droite.

On rappelle que m représente le degré d'approximation de la géométrie d'un maillage (*cf.* définition 2.11 page 54). Au total douze maillages de deux types différents sont utilisés : linéaire avec $m = 1$ ou curviligne avec $m = 4$ (*cf.* figure 5.2).

1. Acronyme allemand : « Gesellschaft für Angewandte Mathematik und Mechanik. »

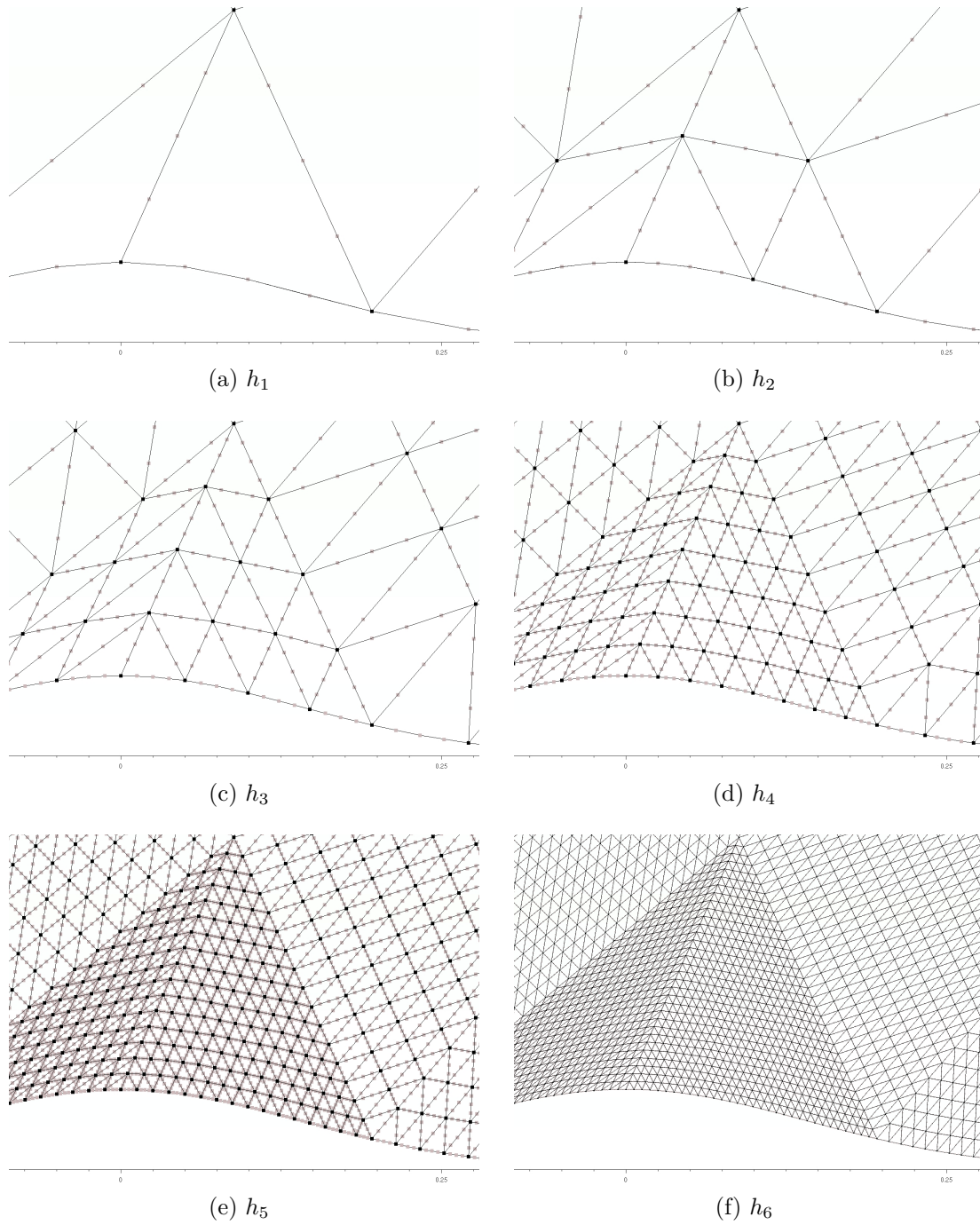


Fig. 5.3 – (*Canal*). Vues agrandies des maillages curvilignes ($m = 4$) au niveau du col en $x_1 = 0$. Les points noirs seuls décrivent les maillages linéaires ($m = 1$). Les points gris permettent d’obtenir des éléments d’ordre élevé.

Une comparaison de l'orientation des normales à la paroi aux points d'intégration numérique a été donnée sur la figure 2.5a page 53. De plus, les maillages curvilignes sont fournis par l'atelier [176, 83] et les maillages linéaires en sont une adaptation. Chacun des deux types de maillages est constitué d'une série de six maillages de plus en plus raffinés à partir du maillage le plus grossier tel que $h = h_1$ où $h_1 \approx 0,1$ (cf. remarque 5.1 et tableau 5.1). Le raffinement consiste en un découpage de chaque triangle du maillage précédent en quatre sous-triangles identiques. Enfin, la figure 5.3 donne une vue agrandie au col pour chaque maillage.

Remarque 5.1. *Pour ce cas test, la taille caractéristique h d'un maillage (cf. relation (2.2a)) est indiquée par l'entier i avec $1 \leq i \leq 6$ afin de différencier les différents maillages. Il vient alors $h = h_1$ pour le maillage le plus grossier.*

5.1.2.2 Paramètres physiques et numériques

Pour tous les calculs sur les cinq premiers maillages, le champ initial $\mathbf{w}^{(0)}(\mathbf{x})$ est uniforme. Pour le sixième maillage, il est uniforme pour $p = 0$ puis repris d'un champ convergé $p - 1$ pour $p \geq 1$ en raison du temps des calculs. De plus, le nombre de Mach amont est fixé à $M_\infty = 0,5$.

On s'intéresse à une solution stationnaire. On utilise alors une méthode de marche en temps pour résoudre le problème et le pas de temps utilisé est local afin d'accélérer la convergence vers l'état stationnaire. La discrétisation spatiale est celle décrite au chapitre 2. La discrétisation temporelle concerne uniquement des termes de convection. Elle est complètement explicite avec un schéma RK3 de Shu-Osher [155] (cf. relations (3.6) et (3.7)). Pour le présent schéma, les valeurs de la condition de CFL sont alors imposées par le pas de temps convectif seul. Elles sont semblables à [28] et valent :

| p | 0 | 1 | 2 | 3 | 4 | 5 |
|--------------------------|------|------|------|------|------|------|
| CFL pour h_1, h_2, h_3 | 1,40 | 0,40 | 0,25 | 0,15 | 0,10 | 0,07 |
| pour h_4, h_5, h_6 | 1,20 | 0,40 | 0,20 | 0,15 | 0,10 | 0,07 |

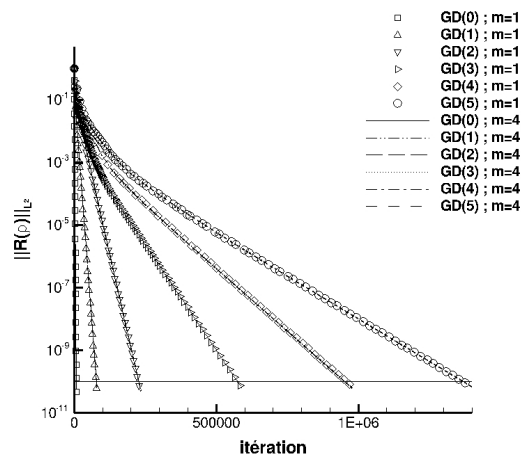
En particulier, on observe que quel que soit le maillage, lorsque p augmente, les valeurs de la condition de CFL décroissent linéairement en p suivant le facteur $2p+1$. Cette observation est conforme à ce qui a été rapporté au paragraphe 3.2 page 62 dans le cas d'un problème purement hyperbolique discrétisé explicitement en temps.

L'évaluation des résultats suit les consignes de l'atelier [176, 83]. A chaque pas de temps on évalue la norme L^2 du résidu discret $\mathbf{L}_c(\mathbf{W}^{(n)})$ pour l'équation de continuité. L'état stationnaire est alors considéré comme atteint lorsque cette erreur a diminué de dix ordres de grandeur par rapport à l'erreur initiale (cf. figure 5.4). Une fois cet état atteint, la précision de la solution numérique est calculée avec l'erreur globale en norme L^2 sur l'entropie $s = p/\rho^\gamma$. Cette erreur est définie par :

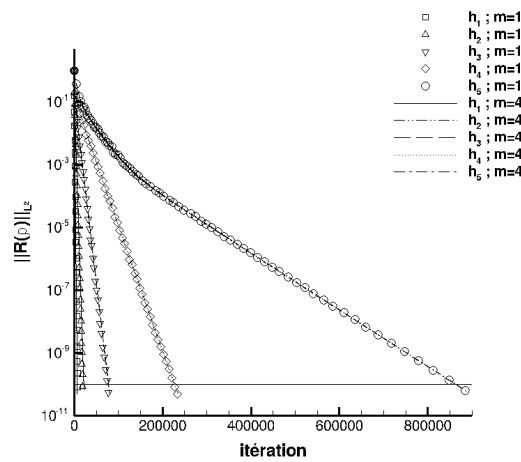
$$\|s - s_\infty\|_{L^2(\Omega_h)} = \sqrt{\frac{\int_{\Omega_h} \left(\frac{p/\rho^\gamma}{p_\infty/\rho_\infty^\gamma} - 1\right)^2 dv}{\int_{\Omega_h} dv}}, \quad (5.1)$$

où les quantités p_∞ et ρ_∞ sont données dans le tableau A.2 page 192 (cf. remarque 5.2).

Remarque 5.2. On rappelle que pour un écoulement isentropique l'équation d'état sur l'entropie s s'écrit : $\partial_t s + \mathbf{U} \cdot \nabla s = 0$ et signifie que s se conserve le long des trajectoires. De plus, si l'écoulement est stationnaire alors $s = s_\infty$ le long des lignes de courant. Comme ici s_∞ est imposée sur la frontière d'entrée ($x_1 = -1, 5$), alors $s = s_\infty$ en tout point. Donc la quantité donnée en (5.1) doit être nulle. Son évaluation numérique donne une indication sur la dissipation numérique du schéma.



(a) h_4



(b) $p = 2$

Fig. 5.4 – (Canal). Norme L^2 du résidu discret de la masse volumique en fonction du nombre d'itérations (échelle logarithmique en ordonnée). Calculs sur le maillage h_4 pour $0 \leq p \leq 5$ (a). Calculs pour $p = 2$ sur les maillages h_1, h_2, h_3, h_4 et h_5 (b). Superposition des résultats pour les maillages linéaires (symboles) et curvilignes (traits).

5.1.3 Description des résultats numériques

La description des résultats numériques s'intéresse à différentes validations. La première est le tracé des iso-valeurs du nombre de Mach local pour rendre compte de la cohérence des phénomènes physiques dans le canal (§ 5.1.3.1). La seconde est l'évaluation de la création d'entropie à la paroi pour observer le comportement de la dissipation numérique sur les deux types de maillages (§ 5.1.3.2). La dernière étape est l'étude de l'erreur sur l'entropie en matière de précision et d'efficacité (§ 5.1.3.3).

5.1.3.1 Iso-valeurs du nombre de Mach local

Soit $M = \|\mathbf{U}\|/c$ le nombre de Mach local, avec $c = \sqrt{\gamma p/\rho}$ la vitesse du son dans le fluide. La figure 5.5 présente les iso-valeurs de M pour le maillage linéaire et curviligne h_2 et pour un degré de polynôme $p = 2$.

On retrouve bien la physique d'une tuyère convergente-divergente subsonique, puisque l'écoulement est accéléré dans le convergent et décéléré dans le divergent. Au col, on relève la valeur $M \approx 0,78$.

Par ailleurs, si on observe la symétrie attendue [65] par rapport à la droite d'équation $x_1 = 0$ des iso-valeurs de M pour le maillage curviligne, ce n'est en revanche pas le cas pour le maillage linéaire.

Plus généralement, à comparaisons équivalentes (maillage et degré p fixés), cette symétrie est plus facile à obtenir sur les maillages curvilignes, laissant pressentir une meilleure précision de la solution pour ces derniers. En particulier, on montre avec la validation suivante sur la création d'entropie que l'asymétrie lorsque $m = 1$ est due à une plus grande dissipation numérique du schéma que lorsque $m = 4$.

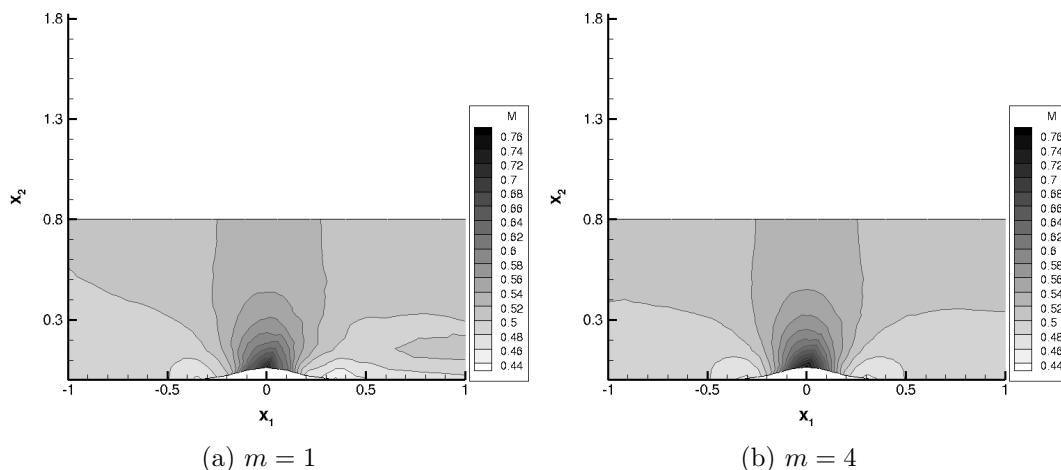


Fig. 5.5 – (Canal). Iso-valeurs du nombre de Mach local. Comparaison entre le maillage h_2 linéaire (a) et le maillage h_2 curviligne (b) pour $p = 2$.

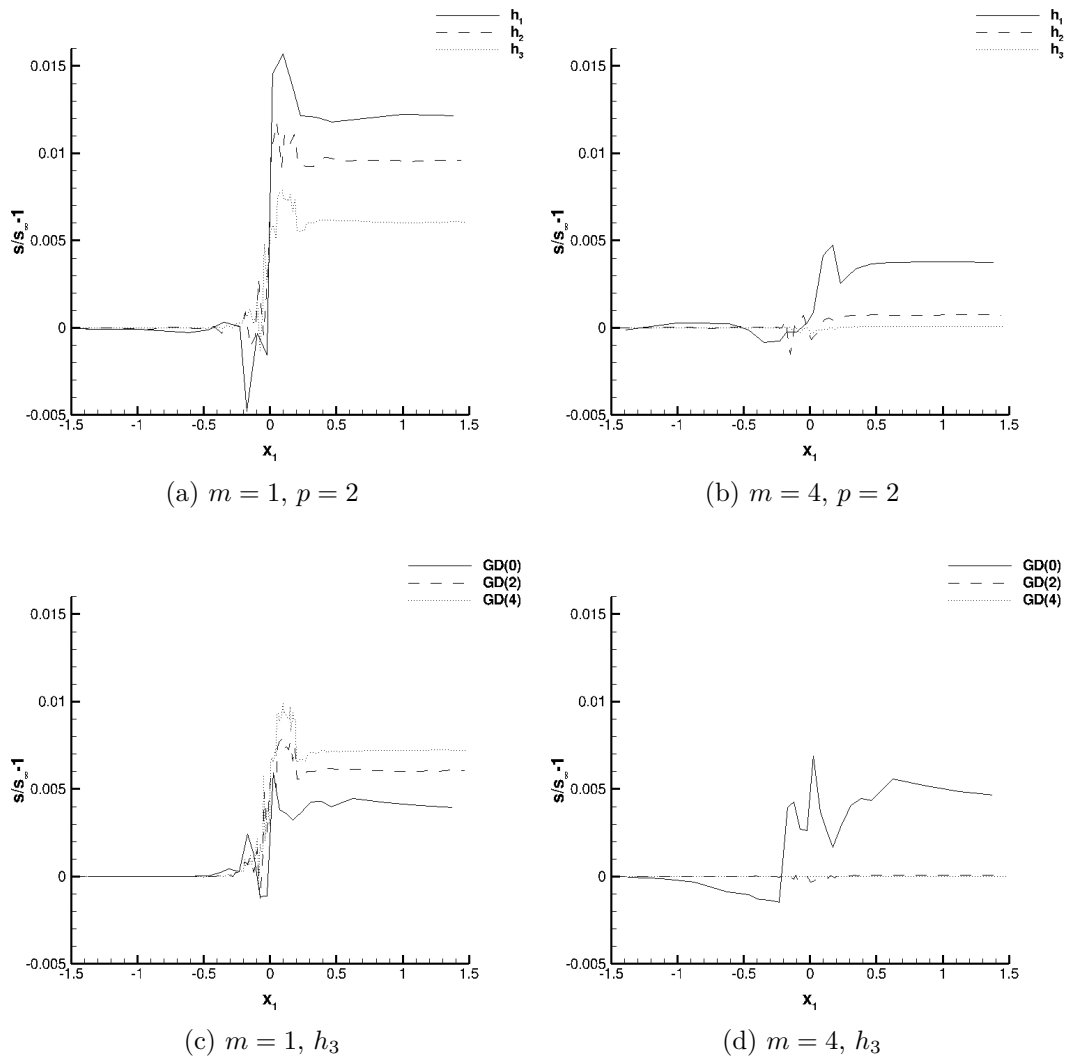


Fig. 5.6 – (Canal). Création d'entropie à la paroi. Comparaison entre les maillages linéaires (gauche) et ceux curvilignes (droite) pour le cas où $p = 2$ sur les maillages h_1, h_2 et h_3 (haut), puis pour le cas du maillage h_3 avec $p = 0, p = 2$ et $p = 4$ (bas).

5.1.3.2 Création d'entropie à la paroi

La seconde validation vérifie dans quelle mesure la relation $s = s_\infty$ est respectée (*cf.* remarque 5.2). La quantité $s/s_\infty - 1$ est alors évaluée à la paroi inférieure, car à cet endroit les phénomènes physiques sont les plus importants et les gradients de la solution sont les plus forts.

Les figures 5.6a et 5.6b donnent la valeur de $s/s_\infty - 1$ pour $p = 2$ fixé et pour les maillages h_1 , h_2 et h_3 respectivement, linéaires ou curvilignes. Pour les deux types de maillages, on observe que la quantité $s/s_\infty - 1$ tend vers zéro quand h diminue. Cependant, pour le même maillage (h_1 par exemple), la création d'entropie est trois fois moindre dans le cas curviligne (plateau situé vers 0,004 pour $x_1 \approx 1,0$) que dans le cas linéaire (plateau situé vers 0,012 pour $x_1 \approx 1,0$). Ce point indique que la convergence en maillage est respectée avec ou sans la prise en compte des géométries courbes.

Les figures 5.6c et 5.6d donnent la valeur de $s/s_\infty - 1$ pour le maillage h_3 fixé respectivement, linéaire ou curviligne, et pour p pair tel que $0 \leq p \leq 4$. Si pour le maillage curviligne la quantité $s/s_\infty - 1$ tend vers zéro quand p augmente, ce n'est en revanche pas le cas pour le maillage linéaire lorsque $x_1 > 0$. Ce point indique que la convergence en ordre est respectée seulement avec la prise en compte des géométries courbes. En effet, c'est ce qui est confirmé par la validation suivante.

5.1.3.3 Erreur globale en norme L^2 sur l'entropie

La troisième validation concerne l'étude de l'erreur globale en norme L^2 sur l'entropie, non seulement en matière de précision pour évaluer les convergences en maillage et en ordre du schéma numérique GD, mais aussi en matière d'efficacité pour évaluer les convergences en degrés de liberté et en temps.

Cette étude est réalisée, d'abord avec les résultats de la méthode GD retenue dans ce travail, puis avec ceux des méthodes utilisées lors de l'atelier [176, 83].

Résultats de la méthode GD retenue.

Convergence en maillage. On rappelle que la précision optimale attendue, pour la méthode GD utilisée, est en loi de puissance telle que

$$\|s - s_\infty\|_{L^2(\Omega_h)} \approx h^{p+1},$$

comme prouvé théoriquement dans [142, théorème 3.1] pour un problème hyperbolique, linéaire, de solution suffisamment régulière en maillages structurés et non cartésiens en deux dimensions d'espace, et comme vérifié numériquement dans [46, 114] sur un cas test similaire à celui considéré ici.

Ainsi, pour apprécier la précision obtenue, on trace l'erreur globale en norme L^2 sur l'entropie (5.1) en fonction de $1/\sqrt{\mathcal{N}}$, avec \mathcal{N} le nombre total de degrés de liberté par équation (*cf.* remarque 5.3). Le choix de la grandeur d'abscisse $1/\sqrt{\mathcal{N}}$ est proposée par l'atelier [176, 83] et permet une comparaison de toutes les méthodes présentées. Ici, on a $\mathcal{N} = NN_p$.

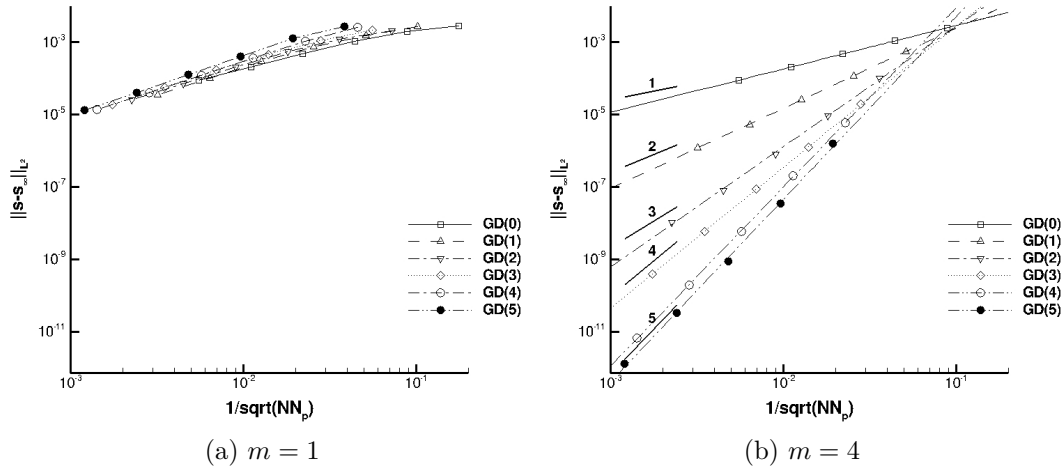


Fig. 5.7 – (*Canal*). Erreur globale en norme L^2 sur l'entropie en fonction de $1/\sqrt{NN_p}$ (échelles logarithmiques). Comparaison entre les maillages linéaires (a) et ceux curvilignes (b) pour $0 \leq p \leq 5$. Les chiffres donnent la valeur des pentes des segments associés.

| p | | 0 | 1 | 2 | 3 | 4 | 5 |
|-----|-------------|-------|-------|-------|-------|-------|-------|
| a | ($m = 1$) | 1, 12 | 1, 37 | 1, 41 | 1, 47 | 1, 58 | 1, 65 |
| a | ($m = 4$) | 1, 20 | 2, 22 | 3, 33 | 3, 90 | 4, 95 | 5, 05 |

Tab. 5.2 – (*Canal*). Ordres de précision a en espace pour les maillages linéaires ($m = 1$) et curvilignes ($m = 4$) tels que l'erreur globale en norme L^2 sur l'entropie soit de l'ordre de $(1/\sqrt{NN_p})^a$ pour chaque degré p de polynôme. La moyenne est réalisée sur les cinq maillages les plus raffinés.

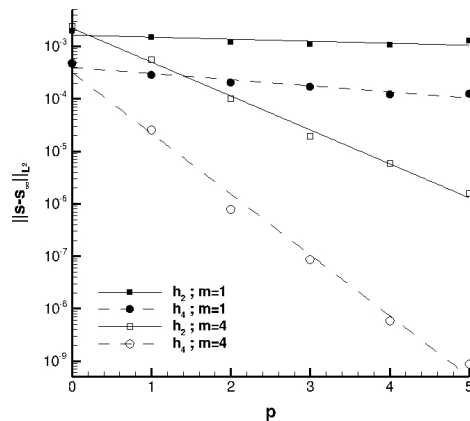


Fig. 5.8 – (*Canal*). Erreur globale en norme L^2 sur l'entropie en fonction du degré de polynôme (échelle logarithmique en ordonnée). Comparaison entre les maillages linéaires (symboles pleins) et ceux curvilignes (symboles creux) pour les deux maillages h_2 (trait plein) et h_4 (trait pointillé).

Ce premier résultat de précision est donné sur la figure 5.7 pour les deux types de maillages et pour $0 \leq p \leq 5$. De plus, le tableau 5.2 récapitule les ordres de convergence a tels que l'erreur regardée $\|s - s_\infty\|_{L^2(\Omega_h)}$ soit de l'ordre de $(1/\sqrt{NN_p})^a$ pour chaque valeur de p . On renvoie à l'annexe C pour les valeurs des erreurs mesurées et le calcul des ordres intermédiaires pour chaque p .

Remarque 5.3. Soit $\|s - s_\infty\|_{L^2(\Omega_h)} \approx h^{p+1}$ alors pour $h \approx 1/\sqrt{N}$ cette précision est équivalente à $\|s - s_\infty\|_{L^2(\Omega_h)} \approx (1/\sqrt{N})^{p+1}$, ou encore à $\|s - s_\infty\|_{L^2(\Omega_h)} \approx (1/\sqrt{N})^{p+1}$ à la constante N_p près.

La comparaison des figures 5.7a pour les maillages linéaires et 5.7b pour ceux curvilignes exhibe une convergence en maillage différente. En effet, à p fixé, plus le maillage est raffiné (soit plus $1/\sqrt{NN_p}$ est petit) plus l'erreur sur l'entropie diminue. Cependant, pour $p = 3$ par exemple, l'erreur sur l'entropie diminue de quatre ordres de grandeur quand le raffinement du maillage augmente pour $m = 4$, tandis qu'elle reste au même ordre de grandeur (10^{-4}) pour $m = 1$.

Ce résultat est confirmé au tableau 5.2. En effet, seules les valeurs pour les maillages curvilignes donnent une convergence en loi de puissance comme attendu, c'est-à-dire que l'ordre de précision de la méthode numérique est en $p + 1$ pour $0 \leq p \leq 4$.

Toutefois pour $p = 5$, on note que l'ordre obtenu est cinq contrairement à la valeur six attendue. Effectivement, la représentation est ici sous-paramétrique ($m < p$) et l'ordre de la méthode GD est limité par l'ordre d'approximation des éléments d'ordre élevé, à savoir cinq pour $m = 4$.

Convergence en ordre. On rappelle que la précision attendue pour la méthode GD utilisée est en loi exponentielle, soit telle que :

$$\|s - s_\infty\|_{L^2(\Omega_h)} \approx e^{-f(h)p},$$

avec f une fonction de h comme prouvé théoriquement dans [68, théorème 5.2] pour un problème linéaire, scalaire et de solution suffisamment régulière, et comme montré numériquement dans [107, 178] sur un cas test similaire à celui considéré ici.

Ainsi, pour apprécier la précision obtenue, on trace l'erreur globale en norme L^2 sur l'entropie (5.1) en fonction de p . Ce second résultat de précision est donné sur la figure 5.8 pour les deux types de maillages h_2 et h_4 et pour $0 \leq p \leq 5$.

La comparaison entre les maillages linéaires et curvilignes exhibe une convergence en ordre différente, comme constaté précédemment. En effet, à maillage fixé, plus p augmente plus l'erreur sur l'entropie diminue de façon exponentielle comme attendu. Cependant, pour le maillage h_4 par exemple, l'erreur sur l'entropie diminue de six ordres de grandeur quand p augmente pour $m = 4$, tandis qu'elle reste au même ordre de grandeur (10^{-4}) pour $m = 1$.

Par ailleurs, on note que pour les maillages curvilignes, le taux de convergence augmente significativement lorsque le raffinement augmente.

Efficacité. Ce résultat est donné sur les figures 5.9a et 5.9b montrant l'erreur globale en norme L^2 sur l'entropie (5.1), en fonction respectivement, du nombre total de degrés de liberté N_{ddl} et du temps CPU, pour les maillages curvilignes et pour $0 \leq p \leq 5$.

On constate que pour atteindre un niveau de précision requis, par exemple 10^{-6} , plus le degré du polynôme est élevé moins le nombre total de degrés de liberté est élevé. Il s'ensuit que le temps CPU requis pour atteindre ce niveau de précision est d'autant plus faible. Ces résultats illustrent bien un intérêt majeur de l'utilisation des méthodes d'ordre élevé, en comparaison par exemple, aux méthodes usuelles de volumes finis d'ordre deux.

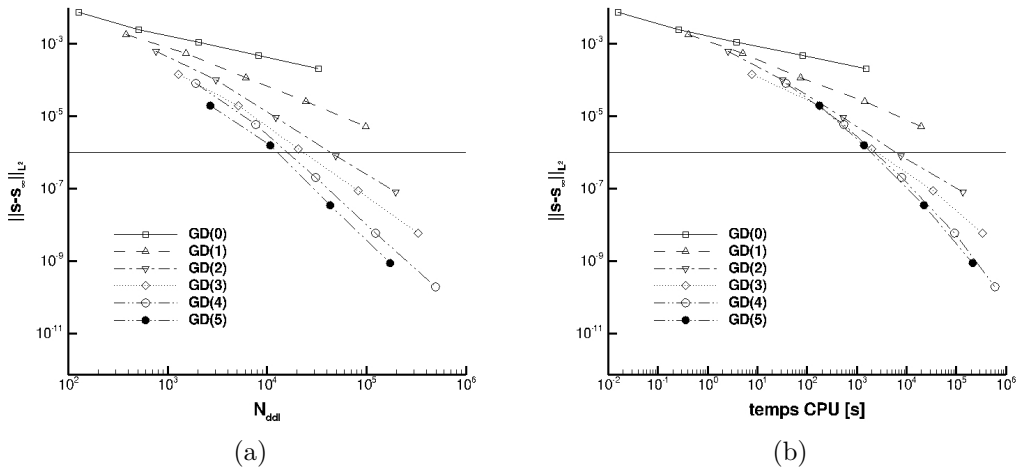


Fig. 5.9 – (Canal). Erreur globale en norme L^2 sur l'entropie en fonction du nombre total de degrés de liberté (a) ou du temps CPU (b) (échelles logarithmiques). Calculs sur les maillages curvilignes pour $0 \leq p \leq 5$.

Résultats des méthodes utilisées lors de l'atelier.

Le cas test du canal de type convergent-divergent est référencé comme le cas test C1.1 de l'atelier « méthode d'ordre élevé » [176, 83] et onze groupes y ont participé sous la direction de Rémi Abgrall.

Le tableau 5.3 récapitule les points importants des méthodes numériques de chaque groupe. A savoir : la méthode en espace avec le type de flux lorsqu'il existe, la méthode en temps, le type de maillage et la valeur de m lorsque c'est possible, et enfin, un article de référence de la méthode en question (en complément de [83]). On note que sept groupes ont proposé une méthode GD lors de l'atelier, ce qui reflète bien l'intérêt grandissant de la communauté scientifique pour cette méthode.

La figure 5.10 donne, quant à elle, l'erreur sur l'entropie (5.1) en fonction de $1/\sqrt{N}$ pour chaque degré p et pour tous les groupes. Une méthode classique de volumes finis d'ordre deux implicite en temps [99, 100], issue du code « Tau » [75] et utilisant les maillages de l'atelier [176, 83], sert de référence. Par ailleurs, on précise que le groupe « Y » utilise des maillages spécifiques raffinés en fonction de N_{ddl} , ce

| Symbole du groupe | Auteurs |
|-------------------|---|
| « A » | R. Abgrall, M. Ricchiuto, D. de Santis |
| « F » | K. Fidkowski, S. Khosravi |
| « G » | M.C. Galbraith, P.D. Orkwis, J.A. Benek |
| « □ » | S. Gérald, F. Renac |
| « H » | H.T. Huynh |
| « L » | Y. Li, Z.J. Wang |
| « O » | C. Ollivier-Gooch |
| « T » | P. Trontin |
| « V » | E. van der Weide, M. Svård |
| « W » | L. Wang, W.K. Anderson, J.T. Erwin |
| « Y » | M. Yano, D. Darmofal |

(a) Noms des participants.

| Groupe | Méthode en espace | Type de flux | Méthode en temps | Type de maillage | m | Référence de la méthode |
|--------|----------------------|--------------|------------------|------------------|-----|-------------------------|
| « A » | DR | LxF | I | Tri. (★) | 4 | [1] |
| « F » | GD | Roe | I | Tri. | 5 | [47] |
| « G » | GD | Roe | I | Quad. | 4 | [49] |
| « □ » | GD | LLF | E | Tri. (★) | 4 | [53] |
| « H » | GD | Roe | E | Tri. (★) | 4 | [70, 50] |
| « L » | GD | Roe | I | Tri. (★) | 4 | [70, 177] |
| « O » | VF | Roe | I | Tri. (★) | 3 | [117, 112] |
| « T » | EF (iso-géométrique) | - | I | Quad. | - | [164] |
| « V » | DF | - | I | Quad. | - | [116] |
| « W » | EF (Petrov-Galerkin) | - | I | Tri. | p | [174] |
| | GD | HLLC | I | Tri. | p | |
| « Y » | GD | Roe | I | Tri. | 5 | [180] |

(b) Méthodes utilisées. Abréviations : distribution du résidu (DR), Galerkin discontinue (GD), volumes finis (VF), éléments finis (EF), différences finies (DF); Lax Friedrichs (LxF), local Lax Friedrichs (LLF), Harten-Lax-van Leer modifié (HLLC); implicite (I), explicite (E); triangle (Tri.), quadrangles (Quad.). Le symbole (★) indique que les maillages utilisés sont ceux de l'atelier.

Tab. 5.3 – (*Canal*). Description succincte des groupes et des méthodes utilisées lors de l'atelier « méthode d'ordre élevé » [176, 83].

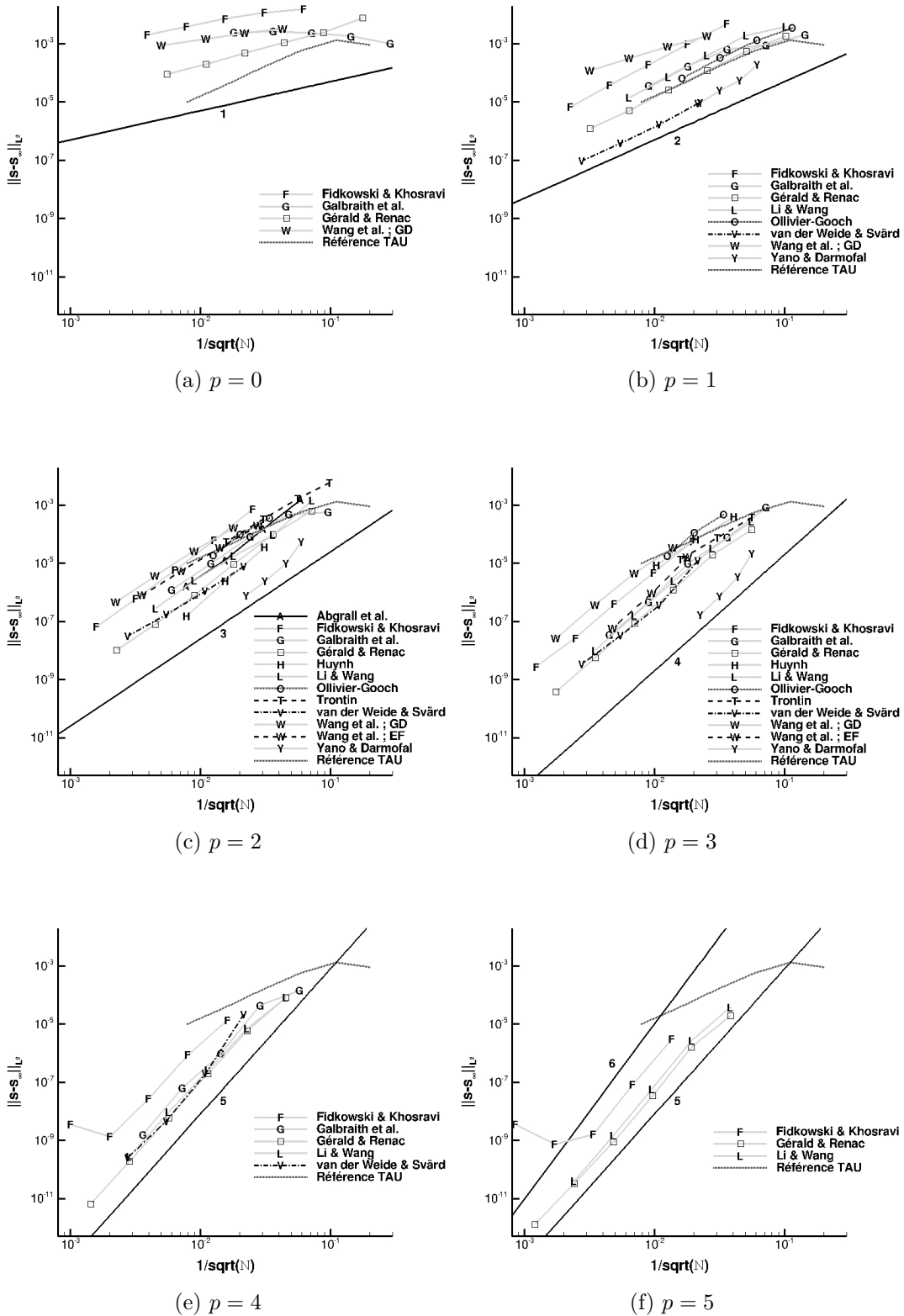


Fig. 5.10 – (Canal). Erreur globale en norme L^2 sur l'entropie en fonction de $1/\sqrt{N}$ (échelles logarithmiques). Comparaison entre les calculs des participants à l'atelier « méthode d'ordre élevé » [176, 83] pour $0 \leq p \leq 5$ et le calcul de référence du code TAU [75] (trait pointillé). Les chiffres donnent la valeur des pentes des droites associées.

qui explique leurs très bons résultats comparés à tous les autres [176]. L'étude de la figure est divisée en deux temps.

- Dans un premier temps, on compare toutes les méthodes d'ordre élevé avec celle de référence. Pour $p = 1$, on retrouve que la précision des méthodes d'ordre élevé est du même ordre de grandeur que celle de la méthode de référence (exception pour les groupes « F, W »). Pour $p > 1$ cette précision est, pour tous les groupes, au moins identique si ce n'est plus grande, et ce plus p augmente. Notamment, pour un degré p , elle est asymptotique d'ordre $p + 1$, sauf pour $p = 5$ où le groupe « L » est limité par l'ordre d'approximation de la géométrie du maillage comme dans ce travail.
- Dans un second temps, on met en avant le comportement général de certaines méthodes. Tout d'abord, on observe que les résultats des deux méthodes d'éléments finis des groupes « T, W » sont similaires, et ceux de la méthode différences finies du groupe « V » sont particulièrement bons. Ensuite, concernant les flux numériques des méthodes GD, on note que celui utilisé dans ce travail (LLF) est très satisfaisant. Pour $p = 1$ par exemple, il donne de meilleurs résultats que le flux de Roe (groupes « F, G, L, O ») ou de HLLC (groupe « W »). Ceci vient conforter le choix fait du flux LLF modifié (2.19) avec $k_2 = 0,05$, pour ses qualités d'approximation où la dissipation numérique peut être paramétrée (et donc minimisée) pour les faibles valeurs de p ($p \leq 1$) lors de la résolution de problèmes où la solution est régulière [33].

Finalement, le nombre important de groupes ayant présenté une méthode GD et l'unanimité des bons résultats de convergence asymptotique dès $p > 1$ démontrent bien que les méthodes GD sont compétitives au sein des méthodes d'ordre élevé et qu'elles représentent une alternative majeure aux méthodes classiques.

Remarque 5.4. *Lors de l'atelier, les résultats d'efficacité ont aussi été mis en commun, mais ils ne sont pas présentés ici. En effet, comme suggéré par Abgrall [176, 83] leur comparaison est peu pertinente puisque les disparités sur le type de maillages, l'optimisation, le parallélisme ou encore la structure des données sont trop importantes.*

5.2 Interaction entre une paire de tourbillons et une paroi plane

Pour valider la discrétisation spatiale des termes de flux visqueux (relation (2.29)), on simule l'écoulement interne, instationnaire et subsonique d'une paire de tourbillons visqueux et contra-rotatifs se dirigeant perpendiculairement à une paroi plane (cas test 2) comme utilisé dans [87]. C'est un problème décrit par le système des équations de Navier-Stokes donné en (1.1). En outre, ce cas test étant décrit pour les écoulements de fluides incompressibles, il est adapté en considérant un faible nombre de Mach.

La présentation du cas test est articulée comme suit. Tout d'abord, le paragraphe 5.2.1 revient sur la difficulté de capturer les phénomènes physiques intervenant lorsqu'une paire de tourbillons heurte une paroi. Puis, le paragraphe 5.2.2

présente la géométrie du domaine de calcul, les maillages et les paramètres physiques et numériques utilisés. Enfin, le paragraphe 5.2.3 synthétise les résultats numériques permettant de valider la discrétisation spatiale retenue pour les termes de diffusion.

5.2.1 Avant-propos

D'une façon générale, l'interaction entre une paire de tourbillons et une paroi est un cas intéressant en mécanique des fluides à cause des phénomènes physiques complexes à petites échelles qui interviennent. En effet, lorsque les tourbillons atteignent la paroi, cette interaction conduit à la formation de nappes (ou structures) de vorticité de forte amplitude dans la zone de couche limite de paroi. Ces structures s'enroulent et donnent naissance à un tourbillon secondaire de signe opposé à celui de départ (*cf.* figure 5.11). Il s'ensuit des interactions plus complexes entre les différentes structures dont l'intensité décroît au cours du temps.

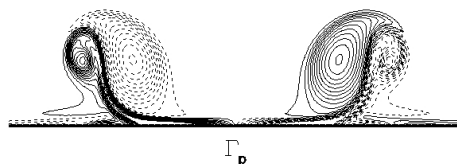


Fig. 5.11 – (*Interaction tourbillons-paroi*). Illustration de l'enroulement des nappes de vorticité (trait plein de gauche ou trait pointillé de droite) se formant dans la couche limite lorsqu'un tourbillon (trait pointillé de gauche ou trait plein de droite) heurte une paroi Γ_p avec condition de non-glissement. Ici, la paroi d'interaction se situe en bas contrairement à l'application numérique présentée où elle se situe à droite.

Ces structures tourbillonnaires secondaires ont été mises en évidence par Harvey et Perry en 1971 [60]. Leur étude numérique en deux dimensions d'espace pour le cas test d'une paire de tourbillons contra-rotatifs venant heurter une paroi plane est réalisée par Orlandi en 1990 [119]. Ce dernier constate avec succès l'accord qualitatif de la formation de structures secondaires, et même tertiaires, avec l'expérience de Van Heijst et Flor de 1989 [168] qui consiste à injecter du fluide en continu dans un volume fermé de faible épaisseur occupé par un fluide stratifié. Par la suite, de nombreux auteurs utilisent le même cas test d'Orlandi pour valider qualitativement la fiabilité de leur méthode [95, 31, 120]. Plus récemment, en 2006, Clercx et Bruneau [24] ont repris ce cas test dans une étude quantitative. En plus des visualisations des iso-valeurs de la vorticité, ils analysent l'évolution en temps de valeurs globales telles que l'énergie cinétique ou l'enstrophie. En particulier, les auteurs établissent la précision spatiale nécessaire pour capturer la physique à la paroi par rapport à l'épaisseur de couche limite.

Ainsi, l'exigence de précision requise pour la simulation de la paire de tourbillons constitue un bon moyen de validation d'une méthode numérique. On cite ici les méthodes spectrales pour les simulations numériques directes (ou DNS). Par exemple, Keetels *et al.* [87] comparent les performances d'une méthode spectrale de Fourier et d'une méthode en base d'ondelettes, toutes deux combinées avec une méthode de pénalisation, avec une méthode spectrale de Fourier-Chebyshev de référence où

$Re_\infty = 1\,000$. Kramer *et al.* [94] s'intéressent à de plus grands nombres Reynolds tels que $Re_\infty \gtrsim \mathcal{O}(10^4)$ et se comparent à la méthode spectrale de référence précédente. Enfin, Laizet et Lamballais [97] comparent eux aussi, les performances d'une méthode mixant résolution dans l'espace physique et résolution dans l'espace spectral avec la méthode spectrale de référence précédente.

Dans ce travail, on utilise également les résultats de Keetels *et al.* [87] afin de valider la discrétisation spatiale GD retenue pour les termes de flux visqueux.

5.2.2 Description du cas test

5.2.2.1 Géométrie du domaine de calcul et maillages

La figure 5.12 décrit la géométrie du domaine de calcul, les conditions aux limites considérées et la condition initiale. Le domaine est tel que $\Omega = [-1; 1] \times [-1; 1]$. Les conditions aux limites appliquées sont uniquement des conditions de non-glissement (paroi immobile et adiabatique).

On rappelle que RA représente le rapport d'aspect global d'un maillage (*cf.* relation (2.3)). Au total six maillages de deux types différents sont utilisés : uniforme où $RA = 1$ (*cf.* figure 5.13a) ou avec rapport d'aspect variable où $RA > 1$ (*cf.* figure 5.13b). Pour contrôler leur précision, ils sont construits avec le même nombre de rectangles dans chaque direction d'espace et coupés en deux. La découpe est réalisée en chevrons pour respecter la symétrie de l'écoulement, laissant apparaître deux zones distinctes : $x_2 > 0$ et $x_2 < 0$. La figure 5.14 donne une vue agrandie de tous les maillages à la paroi.

Le choix des maillages uniformes a été déterminé par les observations de Clercx et Bruneau [24]. Ces derniers ont montré qu'il faut une taille caractéristique du maillage h cinq fois plus petite que l'épaisseur de couche limite δ , estimée *a priori* par la relation $\delta \approx 1/4\sqrt{Re_\infty}$, pour obtenir une précision correcte des phénomènes physiques à la paroi. C'est pourquoi, les trois maillages uniformes retenus sont de plus en plus raffinés jusqu'à obtenir $h \approx \delta$ (*cf.* tableau 5.4). L'idée est de capter les phénomènes physiques en utilisant des polynômes de degré élevé, plutôt que d'avoir un maillage très raffiné tel que $h \approx \delta/5$.

La construction d'un maillage avec rapport d'aspect consiste en un resserrement des mailles vers la zone d'interaction tourbillons-paroi en suivant une progression géométrique selon x_1 , notée PG_{x_1} et définie par

$$PG_{x_1} = \exp\left(\frac{\log(RA_{x_1})}{N-1}\right),$$

où $RA_{x_1} = \Delta x_{1,g}/\Delta x_{1,d}$ correspond au rapport des longueurs horizontales entre le triangle le plus à gauche et celui le plus à droite. Dans ce travail, on applique les trois rapports d'aspects $RA_{x_1} = 10$, $RA_{x_1} = 100$ et $RA_{x_1} = 1\,000$ sur le maillage le plus grossier (*cf.* tableau 5.4). L'objectif n'est pas d'obtenir une plus grande précision de la solution, mais de tester la méthode GD retenue sur des maillages fortement anisotropes, similaires à ceux utilisés en CFD pour les écoulements cisailés.

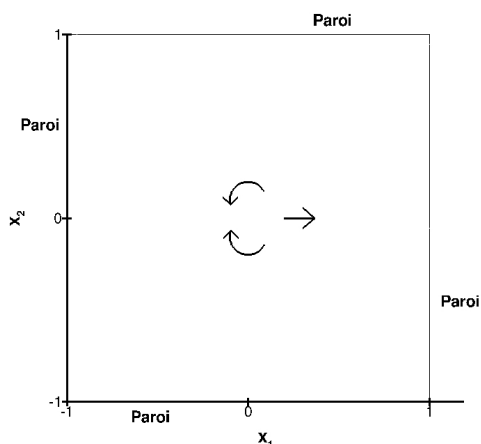


Fig. 5.12 – (*Interaction tourbillons-paroi*). Géométrie, conditions aux limites du domaine de calcul et condition initiale.

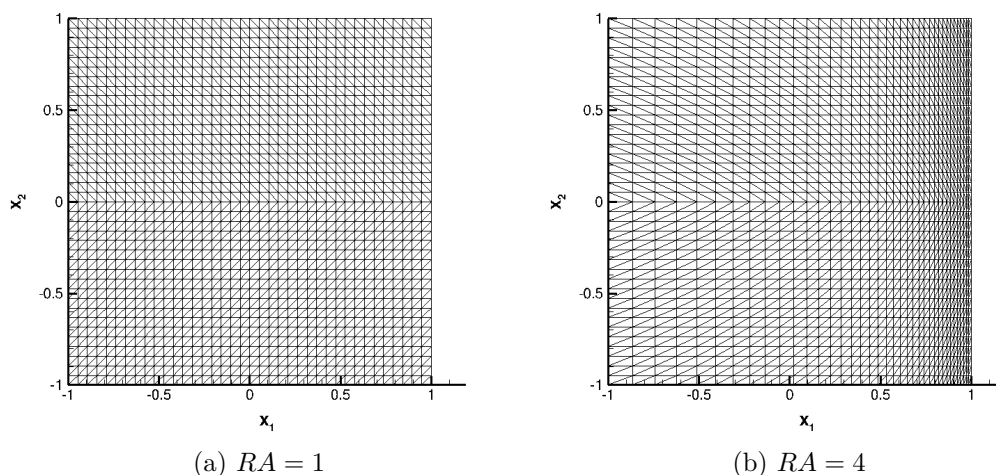
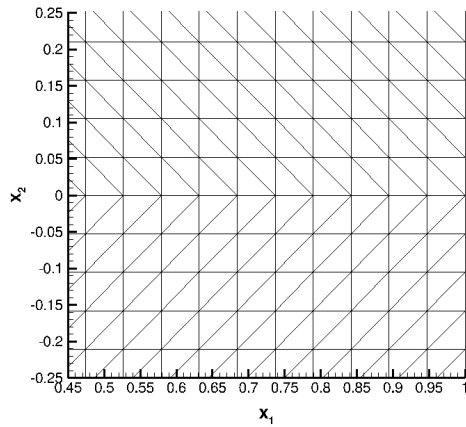


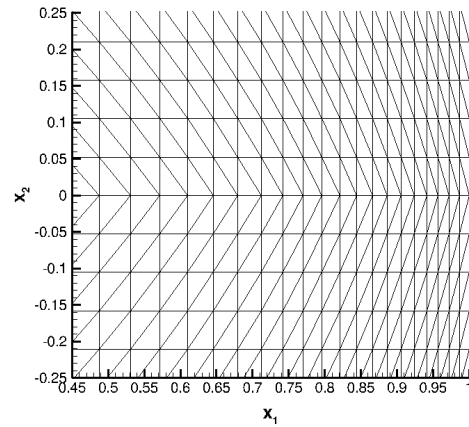
Fig. 5.13 – (*Interaction tourbillons-paroi*). Vue globale du maillage grossier $N = 2\,888$ dans le cas uniforme (a) et dans le cas avec rapport d'aspect (b).

| N | Nb. points sur Γ_p | h | RA | RA_{x_1} | |
|----------------------|---------------------------|--|------|------------|---|
| $2(10)^2 = 200$ | 40 | $1,12 \times 10^{-1} \approx 15\delta$ | 1 | 1 | |
| $2(18)^2 = 648$ | 72 | $6,51 \times 10^{-2} \approx 8\delta$ | 1 | 1 | |
| $2(38)^2 = 2\,888$ | 152 | $3,08 \times 10^{-2} \approx 4\delta$ | 1 | 1 | ★ |
| $2(74)^2 = 10\,952$ | 296 | $1,58 \times 10^{-2} \approx 2\delta$ | 1 | 1 | ★ |
| $2(148)^2 = 43\,808$ | 592 | $7,92 \times 10^{-3} \approx \delta$ | 1 | 1 | ★ |
| $2(38)^2 = 2\,888$ | 152 | $1,17 \times 10^{-2}$ | 4 | 10 | ★ |
| $2(38)^2 = 2\,888$ | 152 | $2,31 \times 10^{-3}$ | 20 | 100 | ★ |
| $2(38)^2 = 2\,888$ | 152 | $3,40 \times 10^{-4}$ | 143 | 1 000 | ★ |
| $2(38)^2 = 2\,888$ | 152 | $8,22 \times 10^{-5}$ | 600 | 5 000 | |

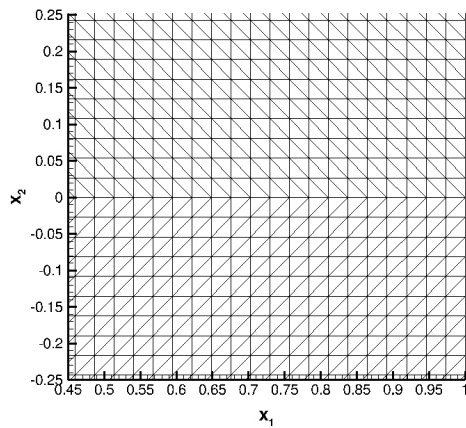
Tab. 5.4 – (*Interaction tourbillons-paroi*). Description des maillages utilisés. Le nombre entre parenthèses correspond au nombre de subdivisions dans chaque direction d'espace. Le symbole ★ correspond aux maillages considérés dans ce chapitre, les autres sont utilisés dans la suite de ce mémoire. Abréviation : Nombre (Nb.).



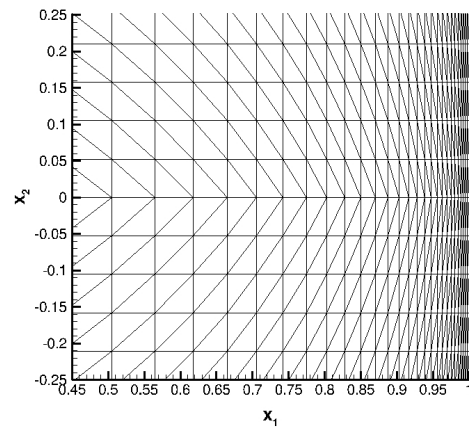
(a) $N = 2\ 888$, $RA = 1$



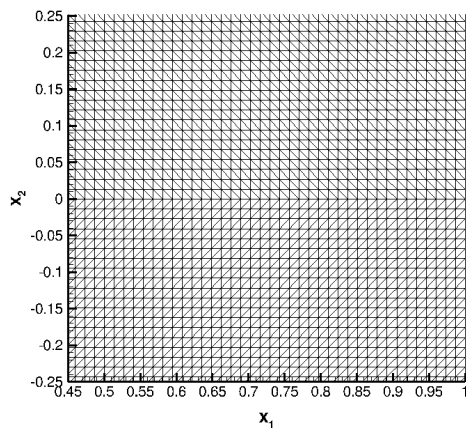
(b) $N = 2\ 888$, $RA = 4$



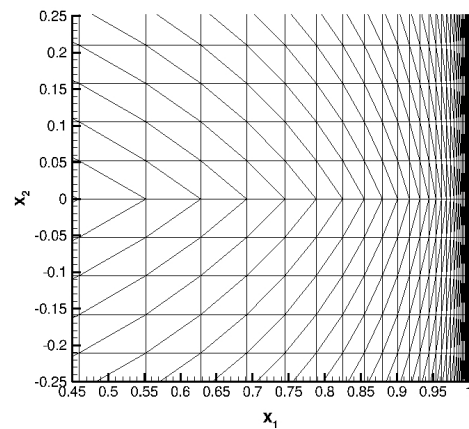
(c) $N = 10\ 952$, $RA = 1$



(d) $N = 2\ 888$, $RA = 20$



(e) $N = 43\ 808$, $RA = 1$



(f) $N = 2\ 888$, $RA = 143$

Fig. 5.14 – (Interaction tourbillons-paroi). Vues agrandies à la paroi en $x_1 = 1$ des maillages utilisés : soit uniformes (gauche), soit avec rapport d'aspect variable (droite).

5.2.2.2 Paramètres physiques et numériques

Soit $\omega = \partial u_2 / \partial x_1 - \partial u_1 / \partial x_2$ la vorticité. L'initialisation de la paire de tourbillons est la superposition de deux tourbillons de Lamb isolés. Les centres de ces tourbillons sont situés en $\mathbf{x}_{C_i} = (x_{1,C_i}; x_{2,C_i})^\top = (0; \pm 0, 1)^\top$ avec $i = A$ et $i = B$, et leur vorticité initiale respective est telle que $\omega_i^{(0)} : r_i \in [0; +\infty) \mapsto \omega_i(r_i) \in \mathbb{R}$ avec

$$\omega_i^{(0)}(r_i) = \pm \omega_s \left(1 - \left(\frac{r_i}{r_s} \right)^2 \right) e^{-\left(\frac{r_i}{r_s} \right)^2},$$

où $r_i = \sqrt{(x_1 - x_{1,C_i})^2 + (x_2 - x_{2,C_i})^2}$ représente la distance au centre \mathbf{x}_{C_i} d'un tourbillon, $r_s = 0, 1$ le rayon de chaque tourbillon au-delà duquel la vorticité change de signe, et $\omega_s = 299, 528385375226$ la vorticité maximale en valeur absolue aux centres \mathbf{x}_{C_i} . Avec cette initialisation, la paire de tourbillons se dirige vers la paroi de droite située en $x_1 = 1$. Le champ de vitesse initiale $\mathbf{U}^{(0)}$ est donné en tout point par :

$$\begin{aligned} u_1^{(0)}(\mathbf{x}) &= -\frac{1}{2} |\omega_s| (x_2 - x_{2,C_A}) e^{-\left(\frac{r_A}{r_s} \right)^2} + \frac{1}{2} |\omega_s| (x_2 - x_{2,C_B}) e^{-\left(\frac{r_B}{r_s} \right)^2}, \\ u_2^{(0)}(\mathbf{x}) &= \frac{1}{2} |\omega_s| (x_1 - x_{1,C_A}) e^{-\left(\frac{r_A}{r_s} \right)^2} - \frac{1}{2} |\omega_s| (x_1 - x_{1,C_B}) e^{-\left(\frac{r_B}{r_s} \right)^2}. \end{aligned}$$

De plus, $\rho^{(0)} = 1$, $\theta^{(0)} = 1$. Aussi, le nombre de Reynolds amont est basé sur la demi-largeur de Ω et la vitesse moyenne, il est fixé à $Re_\infty = 1\,000$ pour comparer les résultats obtenus avec [87]. Le nombre de Mach amont est fixé à $M_\infty = 0,01$ pour limiter les effets de compressibilité comme suggéré dans [23]. Le lecteur peut se référer à [24] pour plus de détails sur l'initialisation de ce cas test.

La précision de la solution numérique est évaluée avec la distribution de vorticité et les évolutions temporelles des deux quantités physiques que sont l'énergie cinétique globale E_c et l'enstrophie globale Z . Ces dernières sont définies par :

$$\begin{aligned} E_c(t) &= \frac{1}{2} \int_{\Omega} \mathbf{U}^2(\mathbf{x}, t) dv, \\ Z(t) &= \frac{1}{2} \int_{\Omega} \omega^2(\mathbf{x}, t) dv, \end{aligned}$$

avec en particulier $E_c(0) = 2$ et $Z(0) = 800$.

On s'intéresse à une simulation pour un temps physique $t = 1$. La discrétisation spatiale est celle décrite au chapitre 2. La discrétisation temporelle est complètement explicite avec un schéma RK2 de Heun [63, 52] (*cf.* relations (3.6) et (3.8)). Pour le présent schéma, les valeurs de la condition de CFL sont imposées par le minimum entre le pas de temps convectif et celui visqueux (*cf.* relation (3.4)). Elles valent :

| p | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|--------------------|-------|-------|--------|--------|--------|--------|--------|
| $CFL \quad RA = 1$ | 0, 50 | 0, 25 | 0, 125 | 0, 075 | 0, 05 | 0, 025 | 0, 015 |
| $RA > 1$ | 0, 50 | 0, 20 | 0, 10 | 0, 05 | 0, 025 | — | — |

En particulier, on fait les observations suivantes.

- Quel que soit le maillage uniforme ($RA = 1$), lorsque p augmente, les valeurs de la condition de CFL décroissent plutôt linéairement en p . Cette observation indique que le pas de temps est limité par les phénomènes de convection.
- Pour les maillages avec rapport d'aspect variable ($RA > 1$), où la taille caractéristique de la maille la plus grande est comparable à celle des maillages uniformes, on s'aperçoit que lorsque p augmente, les valeurs de la condition de CFL décroissent plus fortement avec une évolution comprise entre les lois de comportement prédites par les termes de convection et de diffusion. Cette observation indique que le pas de temps est influencé par les phénomènes de diffusion, lesquels imposent une restriction de stabilité plus stricte à l'origine des méthodes explicites-implicites proposées dans ce travail.

Ces deux observations sont en accord avec ce qui a été rapporté au paragraphe 3.2 page 62 dans le cas d'un problème mixte hyperbolique-parabolique discrétisé explicitement en temps.

5.2.3 Description des résultats numériques

La description des résultats numériques s'intéresse à différentes validations. La première est le tracé des iso-valeurs de la vortacité pour montrer la bonne capture des phénomènes physiques à la paroi (§ 5.2.3.1). La seconde est l'étude des courbes d'évolution temporelle des grandeurs globales d'énergie cinétique et d'enstrophie en matière de précision (§ 5.1.3.2).

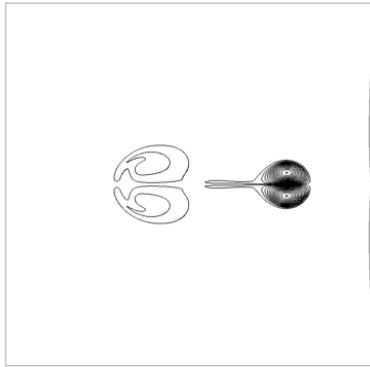
L'ensemble de ces validations repose sur une comparaison des résultats obtenus avec ceux de la méthode spectrale de référence de Keetels *et al.* [87]. Leur méthode utilise une résolution spatiale très précise avec 2 048 modes de Fourier dans la direction de l'écoulement et 1 024 modes de Chebyshev dans la direction transverse.

5.2.3.1 Iso-valeurs de la vortacité

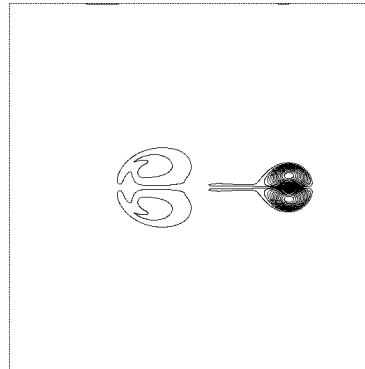
La figure 5.9 présente les iso-valeurs de ω en six instants différents pour le maillage uniforme le plus raffiné $N = 43\,808$ et pour un degré de polynôme $p = 3$. On analyse d'abord cette figure en reprenant les phénomènes physiques apparaissant à la paroi. On renvoie le lecteur à [119, 24] pour plus de détails sur ces phénomènes.

On observe les phénomènes physiques suivants pour les différents instants.

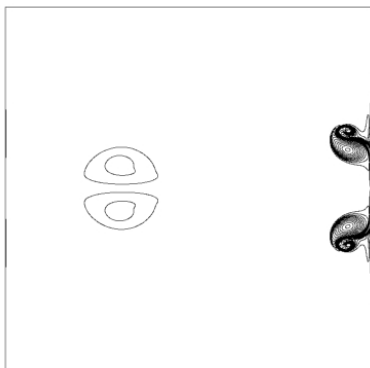
- En $t = 0,2$ la paire de tourbillons contra-rotatifs est symétrique et se dirige vers la paroi de droite où une couche limite constituée de nappes de vortacité se forme (*cf.* remarque 5.5). Comme l'écoulement est symétrique, on s'intéresse uniquement au domaine $x_2 \geq 0$. Dans ce domaine, la vortacité du tourbillon est positive et celle des nappes de vortacité est négative. L'impact entre le tourbillon et la paroi a lieu vers $t \approx 0,3$.
- En $t = 0,4$ l'inertie du tourbillon de départ (ou primaire) entraîne les nappes de vortacité de la couche limite qui s'enroulent pour former un tourbillon secondaire. Une nouvelle paire antisymétrique de tourbillons est alors créée.
- Pour $0,5 \leq t \leq 0,6$, la paire de tourbillons est entraînée et se décolle de la paroi jusqu'à rupture des nappes de vortacité dans la couche limite. On parle de



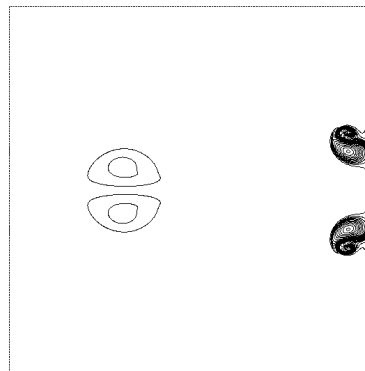
(a) $t = 0, 2$, Référence



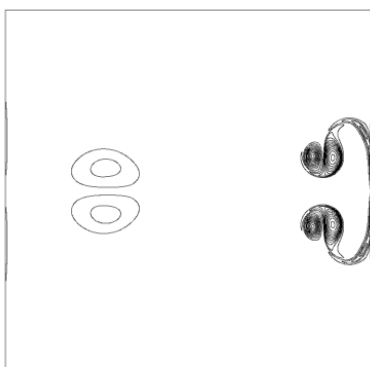
(b) $t = 0, 2$, GD



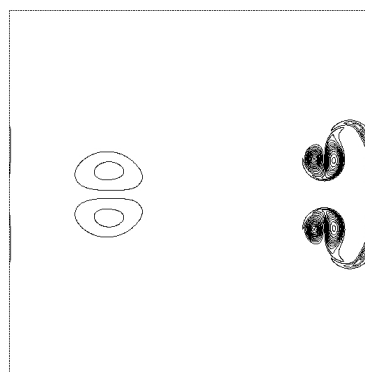
(c) $t = 0, 4$, Référence



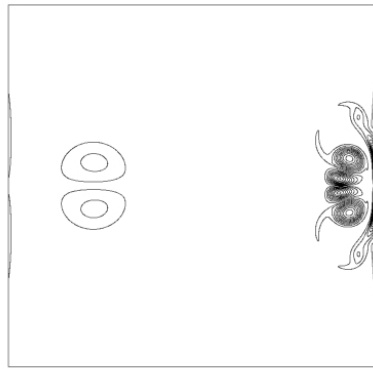
(d) $t = 0, 4$, GD



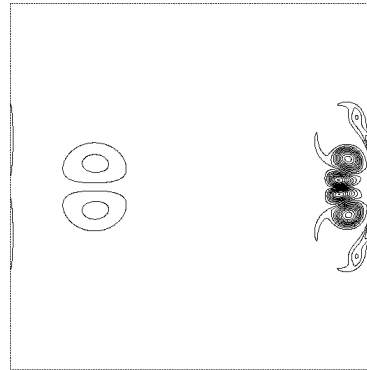
(e) $t = 0, 5$, Référence



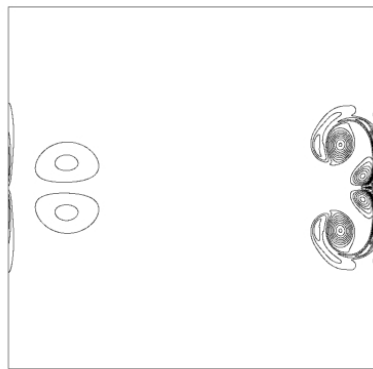
(f) $t = 0, 5$, GD



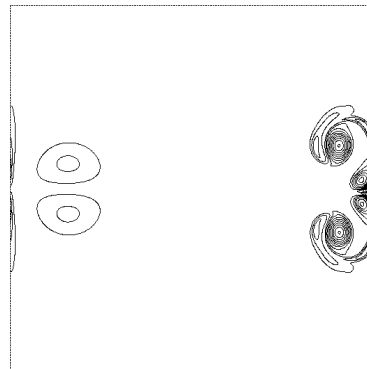
(g) $t = 0,6$, Référence



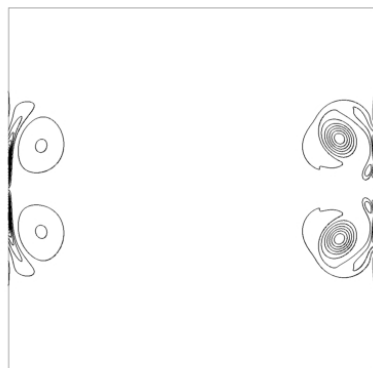
(h) $t = 0,6$, GD



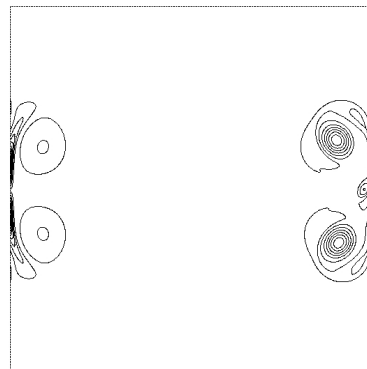
(i) $t = 0,8$, Référence



(j) $t = 0,8$, GD



(k) $t = 1,2$, Référence



(l) $t = 1,2$, GD

Fig. 5.9 – (*Interaction tourbillons-paroi*). Iso-valeurs de la vorticité à différents instants. Comparaison entre le calcul spectral de référence [87] (gauche) et le calcul GD sur le maillage uniforme $N = 43\,808$ pour $p = 3$ (droite). Les niveaux représentés sont $-270, \dots, -10, 10, \dots, 270$.

« rebond ». La vorticit  de tourbillon primaire  tant plus importante au sein de la paire, cette derni re suit une trajectoire circulaire concave et se redirige vers la paroi. Le second impact entre cette paire et la paroi a lieu vers $t \approx 0,6$.

- En $t = 0,8$ la paire de tourbillons se s pare. D’un c t , le tourbillon primaire pr sente une assez forte amplitude de vorticit  pour entra ner,   nouveau, les nappes de vorticit  de la couche limite. Ces derni res s’enroulent pour former un tourbillon tertiaire et il en r sulte une nouvelle paire de tourbillons. D’un autre c t , le tourbillon secondaire ne pr sente plus assez de vorticit , il est dissip  dans la couche limite par les effets visqueux.
- En $t = 1,2$ la dissipation de vorticit  est importante. La paire nouvellement form e ne parvient pas   rebondir et continue   s’enrouler tout en se dissipant dans la couche limite.

Ainsi, sur la figure 5.9 on voit que la m thode GD reproduit aussi bien les ph nom nes physiques   la paroi que la m thode de r f rence, et ce aux m mes instants. De plus, on constate la tr s nette sym trie de l’ coulement par rapport   la droite d’ quation $x_2 = 0$. Cette sym trie se d grade l g rement   partir de $t = 1,2$   cause du manque de pr cision que requi rent normalement les plus petites structures [24]. C’est pourquoi par la suite on choisit une simulation sur un temps physique $t = 1$.

Remarque 5.5. *A $t = 0,2$ on observe en fait deux paires de tourbillons. Celle se dirigeant vers la gauche correspond   l’initialisation de la vorticit  pour le rayon $r_i > r_s$ avec $i = A$ ou $i = B$. Elle est de moindre intensit  que la paire principale se dirigeant vers la droite et se dissipe rapidement sous l’action des effets visqueux.*

5.2.3.2 Evolutions temporelles de l’ nergie cin tique et de l’ensrophie

La seconde validation concerne l’ tude des courbes d’ volution temporelle des grandeurs globales d’ nergie cin tique E_c et d’ensrophie Z en mati re de pr cision, et ce pour  valuer la convergence en maillage et celle en ordre de la m thode GD.

Convergence en maillage. Elle est d crite sur les figures 5.10 et 5.11 o  sont trac es les  volutions temporelles de l’ nergie cin tique E_c et de l’ensrophie Z , pour les deux types de maillages et pour les degr s $p = 2$ et $p = 4$.

Pour les deux degr s p repr sent s sur les maillages uniformes, on constate que plus le maillage est raffin , plus les  volutions en temps des grandeurs globales E_c et Z se rapprochent de la solution de r f rence spectrale. Il y a donc convergence en maillage de la m thode GD pour les maillages uniformes.

En revanche, on fait l’observation contraire sur les maillages avec rapport d’aspect variable. Ceci est directement li  au choix du rapport d’aspect.

- Premièrement, il se trouve que la vorticit  est regroup e   90% aux parois de gauche et de droite [87], or plus le rapport d’aspect est grand, plus le resserrement des mailles   la paroi de droite est, certes, important, mais plus le d serrerment   la paroi oppos e de gauche est  galement important. En outre, on rel ve que lorsque certains auteurs raffinent, ils le font vers les deux parois [97].

- A cela s'ajoute le fait que le rapport d'aspect privilégie la direction de propagation de la paire de tourbillons au détriment de la direction transverse, or les phénomènes physiques étudiés sont relativement isotropes. On peut remarquer que pour le calcul très précis de la méthode spectrale de référence [87] il y a plus de degrés de liberté dans la direction principale de l'écoulement que dans la direction transverse.

Ces deux observations confortent les résultats obtenus : il existe un rapport d'aspect optimal, ici $RA = 4$, pour lequel la capture des phénomènes physiques est la meilleure.

Convergence en ordre. Elle est décrite sur les figures 5.12 et 5.13 où sont tracées les évolutions temporelles de l'énergie cinétique E_c et de l'ensrophie Z , pour tous les maillages et pour tous les degrés. On indique que les courbes pour $p = 0$ ne sont pas pertinentes, elles ne sont donc pas représentées ici. En outre, pour le maillage uniforme le plus grossier, on s'est également intéressé aux degrés $p = 5$ et $p = 6$.

Pour chaque maillage, on observe que plus le degré p est élevé, plus les évolutions en temps des grandeurs globales E_c et Z se rapprochent de leur solution de référence spectrale respective. Il y a donc convergence en ordre de la méthode GD. On note également la convergence en ordre très rapide de la méthode GD d'ordre élevé $p \geq 3$ sur le maillage uniforme le plus grossier $N = 2\,888$, et ce en comparaison aux méthodes classiques d'ordre deux qui ne sont pas capables de restituer les phénomènes physiques complexes de cet écoulement [97].

Remarque 5.6. *On remarque que pour chaque maillage il existe un degré p (et vice versa) pour lequel l'ensemble des phénomènes physiques à la paroi est capturé. En effet, d'un côté la première collision entre la paire de tourbillons initiale et la paroi à $t \approx 0,3$ est bien accompagnée d'une forte diminution d'énergie cinétique totale et d'un pic de création d'ensrophie. D'un autre côté, la seconde collision à $t \approx 0,6$ se traduit par un second pic d'ensrophie. On rappelle que les deux quantités globales E_c et Z sont reliées par la relation $dE_c/dt = -2Z/Re_\infty$ [24].*

5.3 Conclusion partielle

La figure 5.14 et les points suivants constituent le bilan de ce chapitre.

- La validation de la discrétisation spatiale GD associée à une discrétisation explicite en temps (EXPL) a été établie au moyen des validations :
 - de la discrétisation des termes de convection (relation (2.20) associée à une discrétisation explicite RK3 de Shu-Osher (cas test 1) ;
 - de la discrétisation des termes de diffusion (relation (2.29) associée à une discrétisation explicite RK2 de Heun (cas test 2) ;
 - des ordres de convergence en loi de puissance $p + 1$ pour un degré p de polynôme donné (convergence en maillage), et en loi exponentielle en p pour un maillage donné (convergence en ordre) (cas test 1) ;
 - de la métrique du schéma numérique (cas tests 1 et 2).

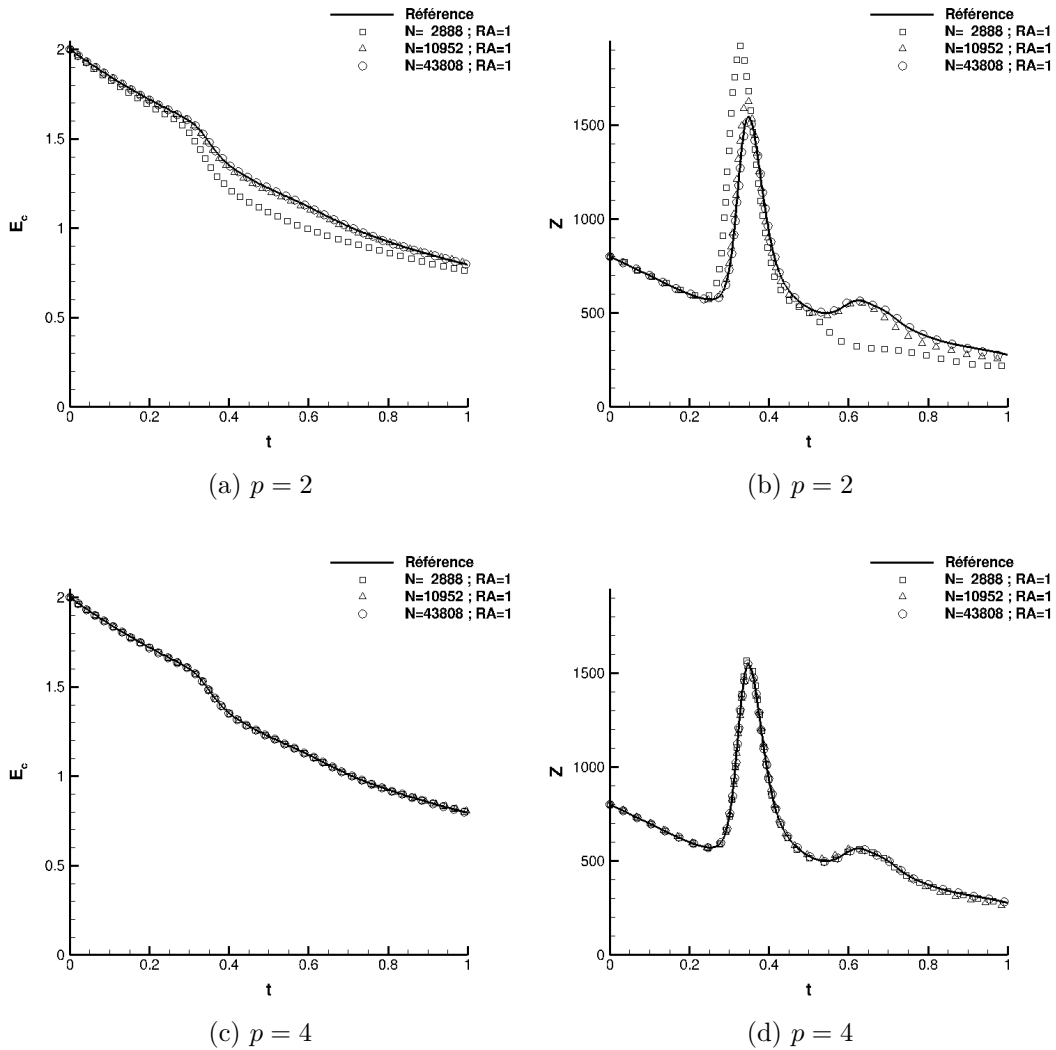


Fig. 5.10 – (*Interaction tourbillons-paroi*). Evolution de l'énergie cinétique (gauche) et évolution de l'énstrophie (droite) en fonction du temps. Comparaison entre le calcul spectral de référence [87] (trait) et le calcul GD (symboles) sur les maillages uniformes pour $p = 2$ (haut) et $p = 4$ (bas).

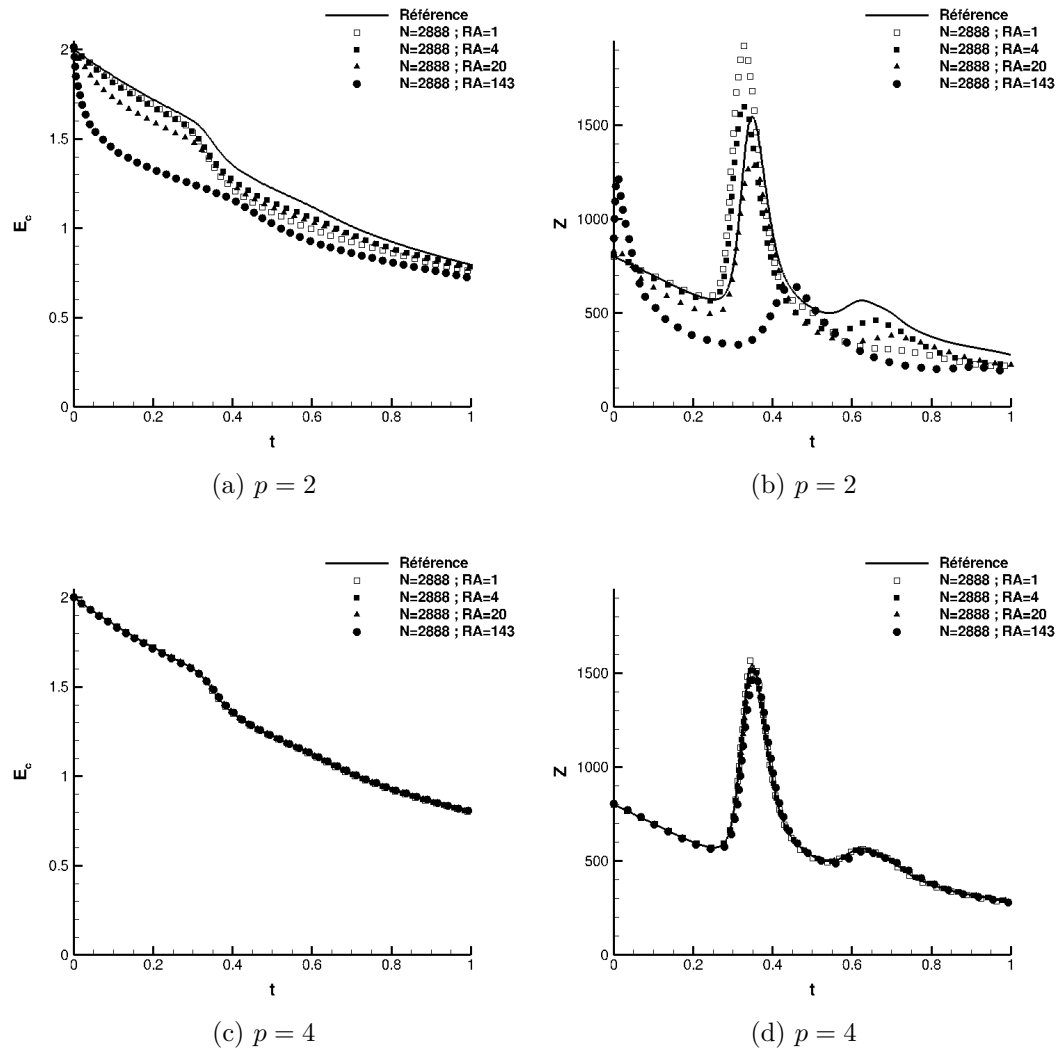


Fig. 5.11 – (*Interaction tourbillons-paroi*). Evolution de l'énergie cinétique (gauche) et évolution de l'enstrophie (droite) en fonction du temps. Comparaison entre le calcul spectral de référence [87] (trait) et le calcul GD (symboles) sur les maillages avec rapport d'aspect variable pour $p = 2$ (haut) et $p = 4$ (bas).

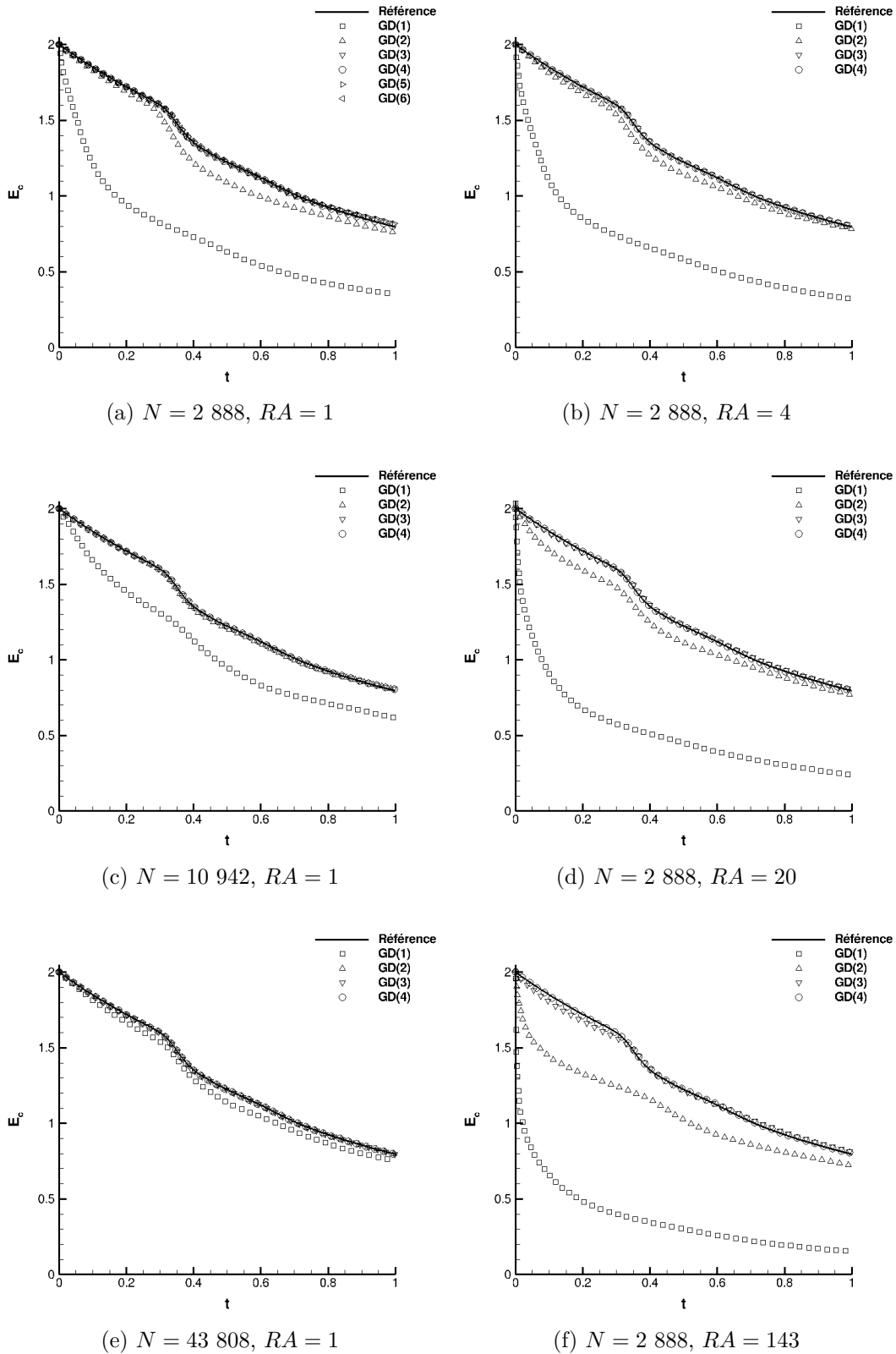


Fig. 5.12 – (*Interaction tourbillons-paroi*). Evolution de l'énergie cinétique en fonction du temps. Comparaison entre le calcul spectral de référence [87] (trait) et le calcul GD (symboles) sur les maillages uniformes (gauche) ou avec rapport d'aspect variable (droite) et pour $1 \leq p \leq 6$ (maillage uniforme grossier) ou $1 \leq p \leq 4$ (autres maillages).

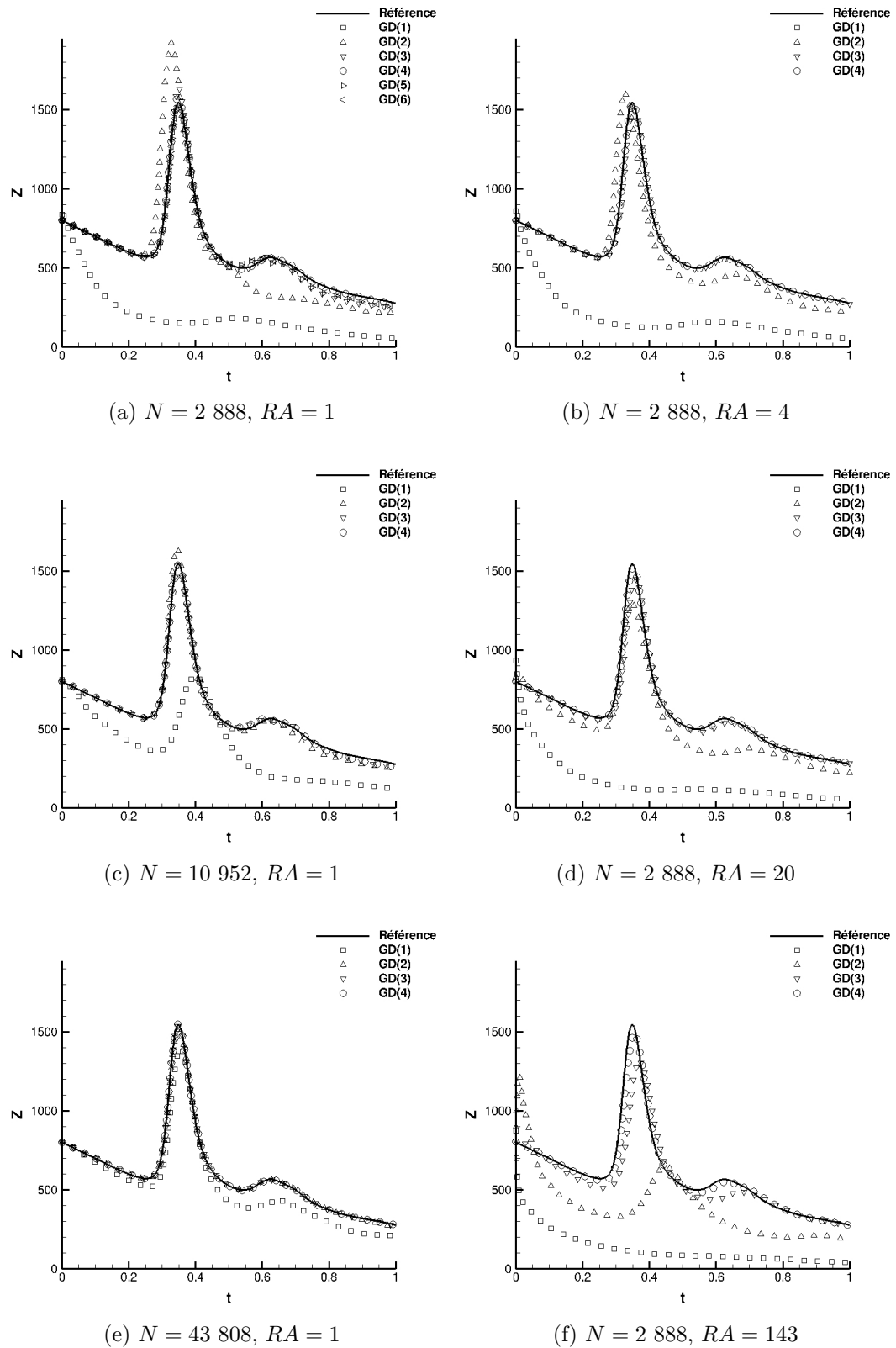


Fig. 5.13 – (*Interaction tourbillons-paroi*). Evolution de l'entrophie en fonction du temps. Comparaison entre le calcul spectral de référence [87] (trait) et le calcul GD (symboles) sur les maillages uniformes (gauche) ou avec rapport d'aspect variable (droite) pour $1 \leq p \leq 6$ (maillage uniforme grossier) ou $1 \leq p \leq 4$ (autres maillages).

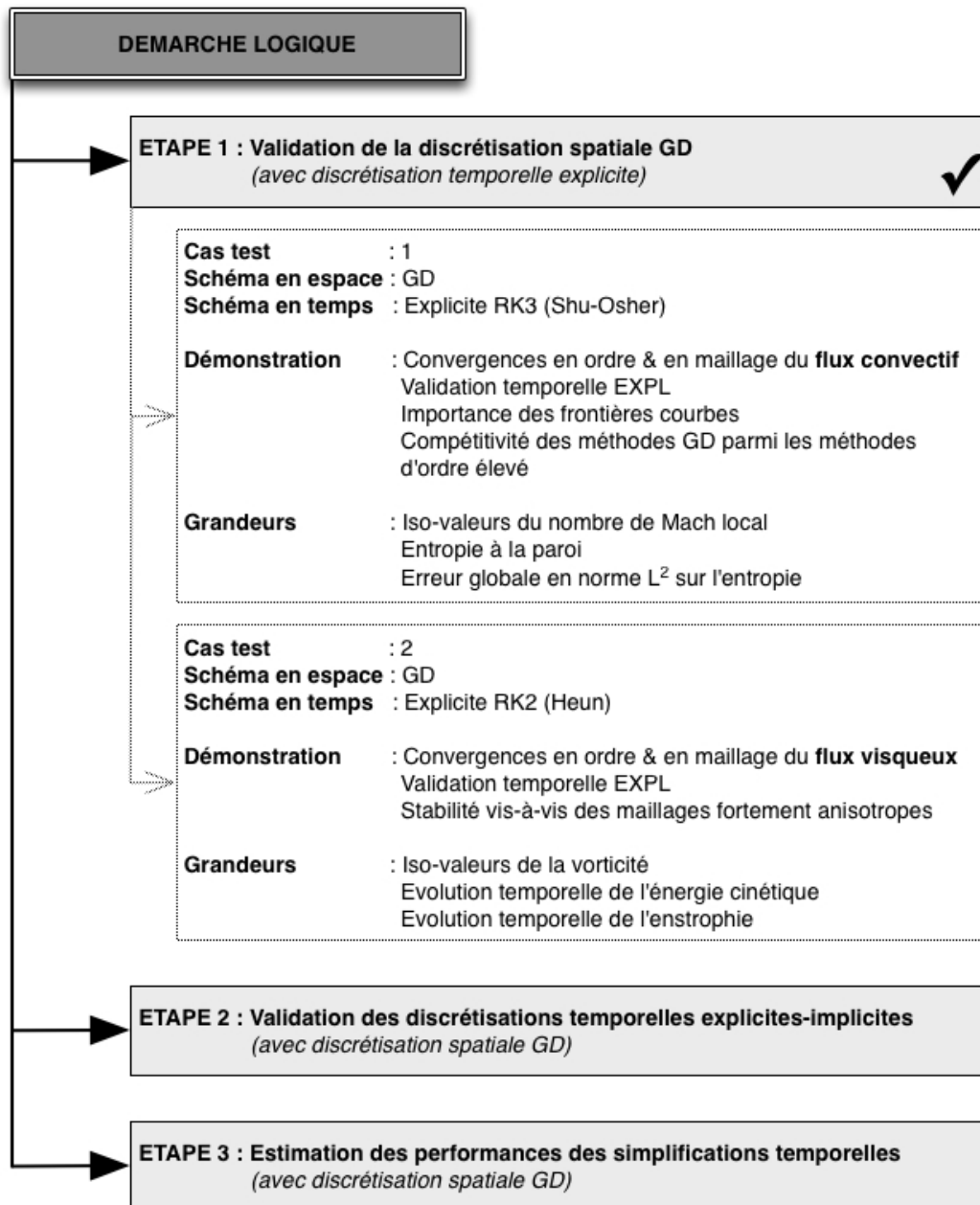


Fig. 5.14 – Schéma récapitulant l'étape 1 de validation du code de calcul développé.

- Certains points clés de la méthode GD ont été mis en avant. A savoir :
 - l'importance de prendre en compte les parois courbes pour obtenir les ordres de convergence théoriques en maillage et en ordre (cas test 1) ;
 - son avantage vis-à-vis des méthodes classiques d'ordre deux à la fois concernant la précision puisque l'on peut faire varier l'ordre de précision de la méthode, et à la fois concernant l'efficacité puisque pour une précision voulue, le temps CPU est d'autant plus faible que p est grand (cas test 1) ;
 - la précision très correcte des phénomènes physiques liés aux petites structures apparaissant au voisinage de la paroi, et ce pour un nombre de degrés de liberté moindre que celui requis par la méthode spectrale de référence [87] (cas test 2) ;
 - la non-altération globale de la précision spatiale sur une large gamme de degrés de polynôme et pour différents types de maillages plus ou moins fortement anisotropes (cas tests 1 et 2) ;
 - sa compétitivité au sein des méthodes d'ordre élevé (cas tests 1 et 2).

A présent, il s'agit de mettre en œuvre les étapes 2 et 3 de validation du code de calcul. C'est l'objet des deux prochains chapitres 6 et 7.

Chapitre 6

Validation des discrétisations temporelles explicites-implicites

Ce chapitre se concentre sur l'étape 2 de validation des méthodes explicites-implicites en temps, COMP, SJ, SIMP p_s et SJ+SIMP p_s , associées à la discrétisation GD en espace déjà validée au chapitre 5.

Quatre cas tests d'écoulements internes et externes, instationnaires et subsoniques de fluides visqueux sont alors simulés : le cas test 2 vu précédemment, les cas tests 3 et 4 d'écoulements autour des profils NACA0012 et SD7003, et le cas test 5 d'écoulement autour d'un cylindre. Les résultats obtenus sont alors comparés à une ou plusieurs références.

6.1 Avant-propos

Dans la suite de ce travail, on impose $p_s = 0$, sauf mention contraire (*cf.* remarque 3.9 page 85 pour plus de détails sur l'entier p_s). On s'intéresse donc à la validation des quatre méthodes COMP, SJ, SIMP0 et SJ+SIMP0.

De plus, les quatre cas tests sont regroupés en deux catégories ayant un rapport avec le degré de complexité de capture des phénomènes physiques apparaissant lors de l'écoulement. D'un côté, les cas tests 2 et 3 appartiennent à la catégorie des cas académiques permettant de valider les quatre stratégies temporelles en profondeur, et d'un autre côté, les cas tests 4 et 5 appartiennent à la catégorie des cas tests d'applications permettant de tester la robustesse des mêmes méthodes.

Ce chapitre est organisé comme suit. Le paragraphe 6.2.1 détaille certains prérequis sur les paramètres physiques et numériques utiles à la validation temporelle, et les deux paragraphes suivants 6.3 et 6.4 sont associés respectivement aux deux catégories de cas tests servant à cette validation.

En outre, on précise que chacun des nouveaux cas tests 3, 4 et 5 est présenté suivant le même schéma avec tout d'abord, des rappels expliquant le choix du cas test en question, puis une description du cas test (géométrie, maillages, paramètres physiques et numériques), et enfin une description des résultats numériques.

6.2 Prérequis pour la validation

Ce paragraphe détaille les quantités physiques ou numériques évaluées pour la validation des discrétisations temporelles explicites-implicites. Il est question de prérequis, lesquels concernent les types de comparaisons envisagées entre les quantités (§ 6.2.1), et les écoulements externes (§ 6.2.2).

6.2.1 Types de comparaisons effectuées pour la validation

Pour atteindre l'objectif de validation de ce chapitre, l'idée est systématiquement de vérifier la bonne capture des phénomènes physiques et leur précision. A cette fin, on considère les trois comparaisons suivantes :

- celle des discrétisations COMP, SJ, SIMP0 et SJ+SIMP0 avec celle EXPL ;
- celle des discrétisations COMP, SJ, SIMP0 et SJ+SIMP0 avec d'autres simulations issues de la littérature ;
- celle des discrétisations COMP, SJ, SIMP0 et SJ+SIMP0 entre elles.

Toutes les comparaisons d'une discrétisation EXPL avec une discrétisation COMP, SJ, SIMP0 ou SJ+SIMP0 s'entendent, sauf mention contraire, à critères équivalents, c'est-à-dire que :

- la discrétisation explicite est de type RK2 de Heun d'ordre deux par cohérence avec les discrétisations explicites-implicites qui utilisent la décomposition d'opérateurs de type Strang également d'ordre deux (afin d'alléger les écritures de ce chapitre, on ne précisera plus systématiquement que la discrétisation EXPL est de type RK2 de Heun) ;
- les calculs sont réalisés avec les valeurs suivantes de la condition de CFL :

| p | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---------------------|------|------|-------|-------|------|-------|-------|
| CFL explicite | 0,50 | 0,25 | 0,125 | 0,075 | 0,05 | 0,025 | 0,015 |
| explicite-implicite | 1,00 | 0,50 | 0,25 | 0,15 | 0,10 | 0,05 | 0,03 |

où, dans le cas d'une discrétisation explicite, ces valeurs sont imposées par le minimum entre le pas de temps convectif et celui visqueux (*cf.* relation (3.4)), tandis que dans le cas d'une discrétisation explicite-implicite, ces valeurs sont imposées par le pas de temps convectif (*cf.* relation (3.5)).

Concernant les valeurs de la condition de CFL, on apporte les précisions suivantes.

- Dans le cas de la discrétisation explicite, on a déjà observé que ces valeurs de la condition de CFL décroissent plutôt linéairement en p , indiquant que le pas de temps est dominé par les phénomènes de convection (*cf.* § 5.2.2.2 page 120).
- Dans le cas de la discrétisation explicite-implicite, on observe que quel que soit le maillage, lorsque p augmente, les valeurs de la condition de CFL décroissent linéairement en p . La restriction de stabilité imposée par les termes de diffusion est bien levée par rapport à la discrétisation explicite (*cf.* § 5.2.2.2 page 120). Ces observations sont en accord avec ce qui a été rapporté au paragraphe 3.2 page 62 dans le cas d'un problème mixte hyperbolique-parabolique discrétisé explicitement-implicitement en temps.

- Le facteur deux observé en faveur des stratégies explicites-implicites correspond au facteur minimum attendu en utilisant la décomposition d'opérateurs de Strang, où le problème hyperbolique est résolu deux fois sur un demi pas de temps (cf. relations (3.13) et (3.14)), et sachant que, tant pour la discrétisation explicite que pour celles explicites-implicites, le pas de temps est limité par les phénomènes de convection.
- Un facteur supérieur à deux pourrait être atteint si dans le cas d'une discrétisation explicite, le pas de temps était systématiquement limité par les phénomènes de diffusion. En particulier, ce serait le cas pour des maillages contenant des éléments avec des grands rapports d'aspect¹, car la condition sur Re_h serait alors plus contraignante. A ce titre, le cas test 5 est un bon exemple d'application puisque, comme on le verra, le schéma EXPL est instable pour tous les degrés de polynôme considérés ($0 \leq p \leq 6$) et toutes les valeurs de la condition de CFL non rédhibitoires en temps de calcul (ici, on a choisi $CFL \geq 10^{-2}$).
- On touche donc ici à une limitation de ce travail, tout en rappelant que l'objectif premier n'est pas de multiplier la valeur de CFL au maximum entre la discrétisation explicite et celles explicites-implicites, mais il est de valider les méthodes proposées afin qu'elles soient stables dans la limite $Re_h/(2p+1) \ll 1$.

Enfin, pour les trois types de comparaisons, on définit les deux erreurs suivantes.

Définition 6.1. (Erreur sur la solution numérique) L'erreur en norme L^2 sur la solution numérique est définie par

$$\|\mathbf{w}_{REF} - \mathbf{w}_{h,AUTRE}\|_{L^2(\Omega_h)} = \sqrt{\frac{\int_{\Omega_h} (\mathbf{w}_{REF} - \mathbf{w}_{h,AUTRE})^2 dv}{\int_{\Omega_h} dv}}, \quad (6.1)$$

où \mathbf{w}_{REF} et $\mathbf{w}_{h,AUTRE}$ sont les solutions numériques issues d'un schéma différent (EXPL, COMP, SJ, SIMP0 ou SJ+SIMP0) re-précisé en fonction des cas tests.

Définition 6.2. (Erreur en pourcentage) L'erreur en pourcentage sur une quantité quelconque z est définie par

$$\frac{z_{REF} - z_{AUTRE}}{z_{REF}} * 100, \quad (6.2)$$

où z_{REF} et z_{AUTRE} sont issues d'un schéma différent (EXPL, COMP, SJ, SIMP0 ou SJ+SIMP0) re-précisé en fonction des cas tests.

6.2.2 Grandeurs caractéristiques des écoulements externes

Parmi les quatre cas tests simulés lors de cette seconde étape de validation, un seul est représentatif des écoulements internes (cas test 2) et trois sont représentatifs des écoulements externes (cas tests 3, 4 et 5). Ces derniers n'ont pas encore été abordés dans ce travail et on apporte alors des précisions sur les phénomènes physiques apparaissant et sur les grandeurs évaluées.

1. Le rapport d'aspect d'un élément est le rapport de son plus grand côté sur son plus petit, il est à différencier du rapport d'aspect RA d'un maillage donné par la relation (2.3).

Les écoulements autour de corps solides sont ici étudiés à des nombres de Reynolds amont faibles ou modérés ($150 \leq Re_\infty \leq 10^4$) et à de faibles incidences ($0^\circ \leq \alpha \leq 4^\circ$). De fait, ils présentent un comportement instationnaire marqué par deux phénomènes.

- Le premier phénomène est le décollement laminaire de la couche limite dû à la présence d'un gradient de pression adverse. Ce gradient est généré par la forme de l'objet et par l'incidence de l'écoulement amont qui induisent un contournement du front d'attaque et donc une accélération en proche paroi au niveau de l'extrados. Les valeurs des coefficients aérodynamiques globaux caractérisent ce décollement.
- Le second phénomène concerne l'ensemble des instabilités de sillage dues au détachement de tourbillons périodiques alternés en aval de l'objet. Le nombre de Strouhal, déjà donné relation (1.9), permet la caractérisation de ces instabilités.

Concernant les coefficients aérodynamiques, on souhaite ici apporter les précisions suivantes pour une meilleure compréhension de la suite. De manière générale, pour les écoulements externes présentant une structure soumise au vent, l'évaluation des forces exercées par le fluide sur la structure rend compte des performances de l'ensemble. En deux dimensions d'espace, ces forces sont celles de traînée et de portance quantifiées par les coefficients aérodynamiques globaux du même nom [65, chapitre 11].

- Plus précisément, les coefficients aérodynamiques globaux de traînée C_D et de portance C_L sont calculés dans le repère aérodynamique lié au vent. Ce repère est obtenu par la rotation d'angle α avec le repère physique $(0, x_1, x_2)$ lié à la structure, et où α est donné par la relation (1.8a). Dans le repère aérodynamique, C_D représente la composante dans la direction de l'écoulement et C_L celle qui lui est perpendiculaire.
- Chaque coefficient est composé de la somme d'un effort de pression (indiqué par p) et d'un effort de frottement tangentiel (indiqué par v), ce dernier est dû à la viscosité du fluide. Soit $C_i = C_{i,p} + C_{i,v}$ avec $i = D$ ou $i = L$.
- Soit $C_\infty = \rho_\infty \mathbf{U}_\infty^2 L_c^2 / 2$, l'expression adimensionnée de ces coefficients est :

$$C_D = C_{D,p} + C_{D,v} = \frac{1}{C_\infty} \int_{\Gamma_p} (p\mathbf{n} - \boldsymbol{\tau} \cdot \mathbf{n}) \cdot (\cos \alpha, \sin \alpha)^\top dS,$$

$$C_L = C_{L,p} + C_{L,v} = \frac{1}{C_\infty} \int_{\Gamma_p} (p\mathbf{n} - \boldsymbol{\tau} \cdot \mathbf{n}) \cdot (-\sin \alpha, \cos \alpha)^\top dS.$$

- Dans la pratique, ces coefficients sont calculés au cours du temps. On évalue donc leur moyenne temporelle sur un certain nombre de périodes, une fois le régime instationnaire établi.

6.3 Validation sur des cas académiques

Les cas tests académiques 2 et 3 sont ici considérés pour la validation des quatre méthodes COMP, SJ, SIMP0 et SJ+SIMP0. Les résultats sont respectivement présentés aux paragraphes 6.3.1 et 6.3.2.

6.3.1 Interaction entre une paire de tourbillons et une paroi plane

6.3.1.1 Rappel du cas test

La simulation de ce cas test a été entièrement décrite au paragraphe 5.2 page 115, où elle a servi à valider la discrétisation GD en espace des termes de diffusion. Pour la suite on rappelle que $M_\infty = 0,01$, $Re_\infty = 1\,000$ et $t \in [0, 1]$.

Par ailleurs, pour la présente validation, seuls les maillages uniformes ($RA = 1$) sont utilisés. Les valeurs de la condition de CFL utilisées ont été données au paragraphe 6.2.1.

6.3.1.2 Description des résultats numériques

La description des résultats numériques s'intéresse aux deux aspects suivants. D'un côté, on vérifie la bonne capture des phénomènes physiques de chacune des simulations explicites-implicites en comparant les iso-valeurs de la vorticit   d'une part, et les courbes d'  volution temporelle de l'  nergie cin  tique et de l'enstrophie d'autre part, avec la solution de r  f  rence. D'un autre c  t  , on   value la pr  cision en maillage de chacune des solutions num  riques explicites-implicites en calculant l'erreur globale en norme L^2 sur la quantit   de mouvement projet  e selon x_1 par rapport    la solution de r  f  rence.

En outre, pour ce cas test, la solution de r  f  rence est celle issue du sch  ma EXPL valid   au chapitre 5.

Iso-valeurs de la vorticit  . La figure 6.1 pr  sente les iso-valeurs de ω sur le maillage uniforme ($RA = 1$) le plus grossier $N = 2\,888$, pour $p = 4$ et $t = 0, 8$. On rappelle que la convergence en ordre est tr  s rapide sur ce maillage et tend vers le calcul spectral de r  f  rence [87] d  s $p = 3$ (*cf.* figure 5.13a page 129).

Sur les figures 6.1b    6.1e, on observe la tr  s bonne capture des ph  nom  nes physiques pour chacun des quatre calculs explicites-implicites en comparaison avec celui de r  f  rence. Sur la vue agrandie figure 6.1f, on note   galement une superposition forte des iso-valeurs de l'ensemble des calculs, laissant supposer peu d'  carts de pr  cision entre eux. En particulier, on insiste sur la pr  cision du tourbillon tertiaire, form   par l'enroulement des nappes de vorticit      la paroi (*cf.* § 5.2.3 page 121 pour l'explication des ph  nom  nes physiques). Ce dernier est captur   avec une tr  s nette sym  trie par rapport    la droite d'  quation $x_2 = 0$, et ce malgr   le maillage grossier.

Evolutions temporelles de l'  nergie cin  tique et de l'enstrophie. Les figures 6.2 et 6.3 pr  sentent les courbes d'  volution temporelle, respectivement $E_c(t)$ et $Z(t)$, toujours sur le maillage uniforme ($RA = 1$) le plus grossier $N = 2\,888$, pour $1 \leq p \leq 6$ et $t = 1$.

Pour chacune des sous-figures, repr  sentant une simulation    un degr   p diff  rent, les discr  tisations temporelles explicites-implicites sont unanimement confondues avec celle de r  f  rence. L'ensemble des ph  nom  nes physiques est donc tr  s bien reproduit. Ce point confirme l'  quivalence des m  thodes COMP, SJ, SIMP0 et SJ+SIMP0 d  j   observ  e avec les iso-valeurs de ω .

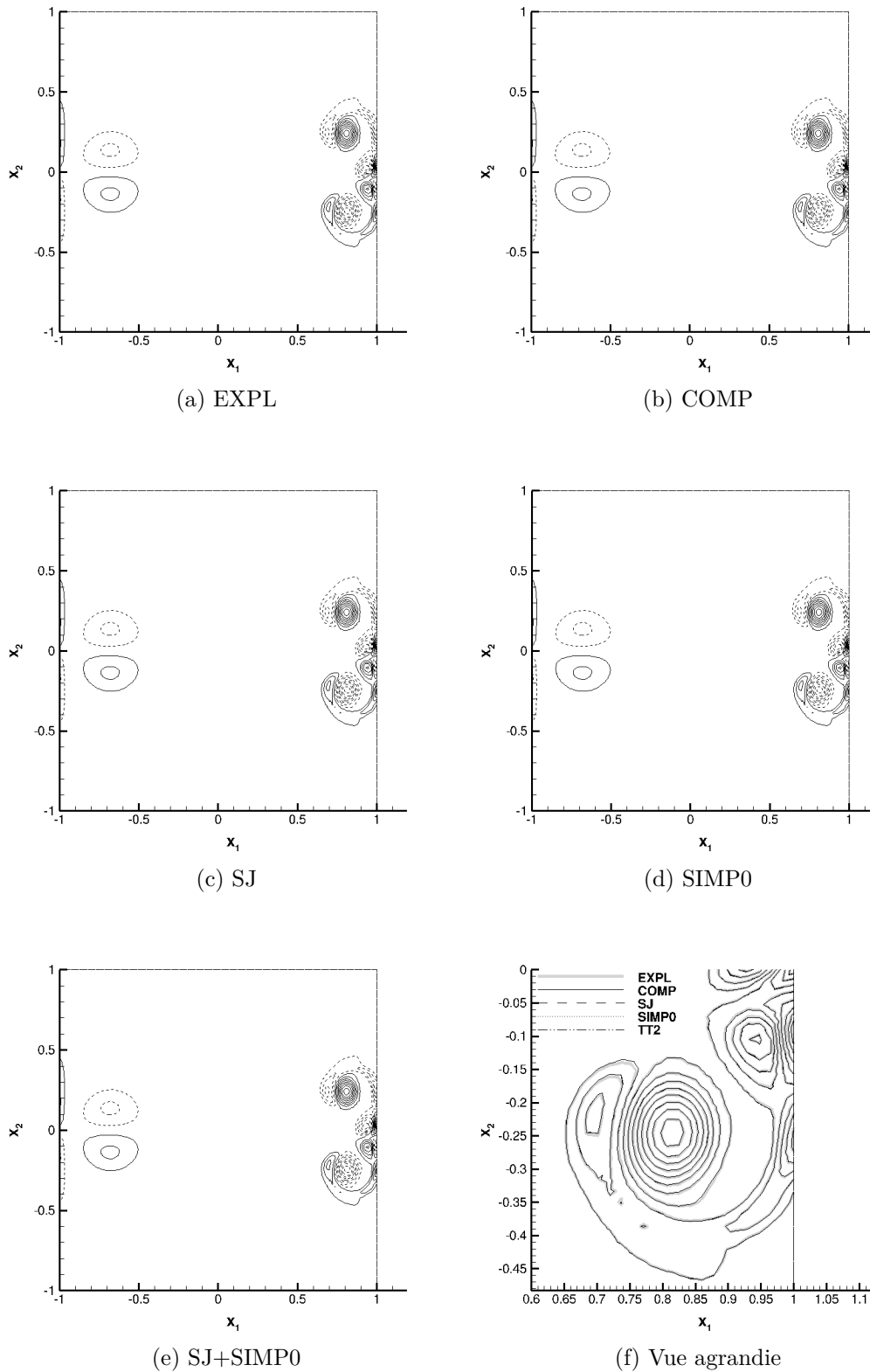


Fig. 6.1 – (*Interaction tourbillons-paroi*). Iso-valeurs de la vorticité. Comparaison entre la simulation de référence (a) et les discrétisations explicites-implicites (b), (c), (d) et (e) sur le maillage uniforme $N = 2\,888$, pour $p = 3$ et $t = 0,8$. La figure (f) montre une vue agrandie. Les niveaux représentés sont $-270, \dots, -10, 10, \dots, 270$, en trait pointillé pour les valeurs négatives ou en trait plein pour les valeurs positives.

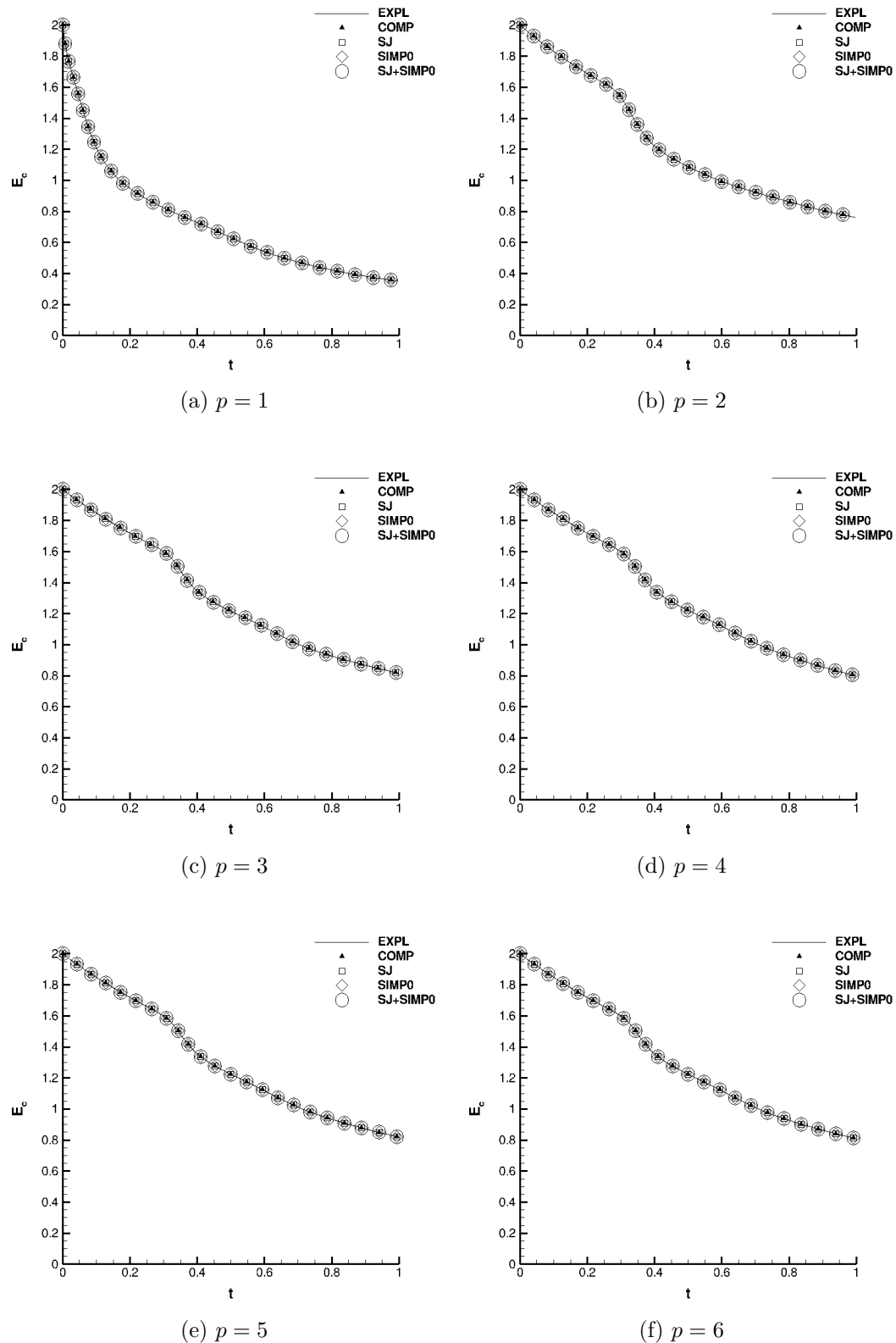


Fig. 6.2 – (*Interaction tourbillons-paroi*). Evolution de l'énergie cinétique en fonction du temps. Comparaison entre différentes discrétisations temporelles : explicite (trait) et explicites-implicites (symboles) sur le maillage uniforme $N = 2\ 888$ et pour $1 \leq p \leq 6$.

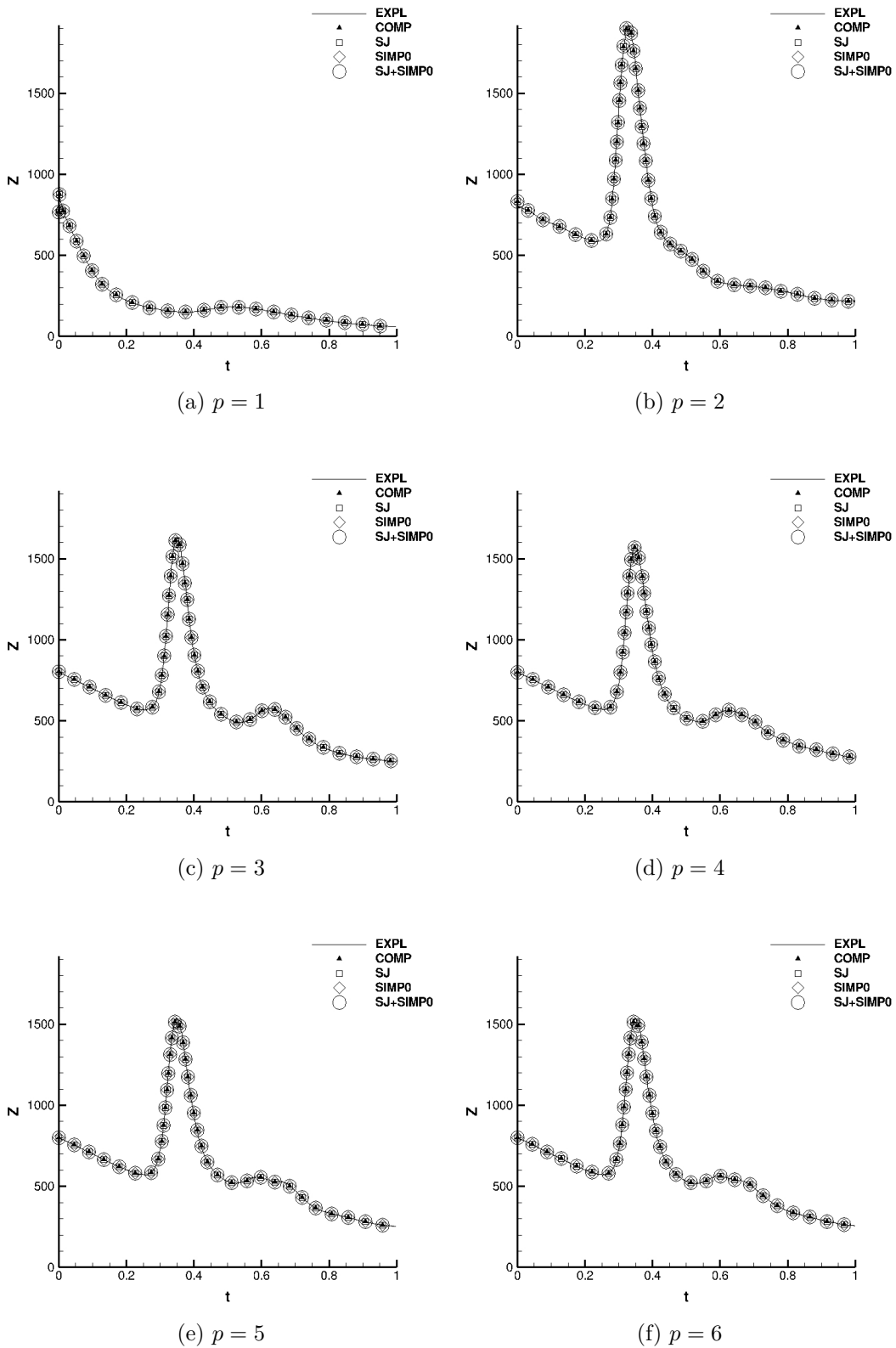


Fig. 6.3 – (*Interaction tourbillons-paroi*). Evolution de l'enstrophie en fonction du temps. Comparaison entre différentes discrétisations temporelles : explicite (trait) et explicites-implicites (symboles) sur le maillage uniforme $N = 2\,888$ et pour $1 \leq p \leq 6$.

Erreur sur la solution numérique. La seconde validation est obtenue avec l'erreur globale en norme L^2 sur la quantité de mouvement projetée selon x_1 , soit l'erreur sur $(\rho u_1)_h$. On trace cette erreur en fonction de $1/\sqrt{NN_p}$ pour apprécier la convergence en maillage. Une telle étude ayant déjà été menée en détail lors de l'étape 1 (cf. § 5.1.3.3 page 109), on se concentre ici sur les résultats principaux.

On rappelle que l'erreur sur $(\rho u_1)_h$ est définie par la relation (6.1), où l'on évalue la précision d'un calcul explicite-implicite en utilisant comme référence un calcul explicite où $p = 4$ et $CFL = 10^{-2}$ afin d'isoler au mieux l'erreur temporelle.

Le résultat est donné sur la figure 6.4 pour cinq maillages uniformes, pour $1 \leq p \leq 4$ et $t = 0, 1$. De plus, le tableau 6.1 récapitule les valeurs des erreurs mesurées et des ordres intermédiaires de convergence a tels que l'erreur regardée $\|(\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE}\|_{L^2(\Omega_h)}$ soit de l'ordre de $(1/\sqrt{NN_p})^a$ pour chaque p . On renvoie à l'annexe D pour ces mêmes calculs en norme L^1 et L^∞ .

Tout d'abord, on remarque que le comportement de toutes les méthodes explicites-implicites est identique, comme le révèle la parfaite superposition des résultats sur la figure 6.4 et les erreurs aboutissant aux mêmes ordres de précision dans le tableau 6.1. Ensuite à $p \leq 3$ fixé, on observe la convergence en maillage attendue en loi de puissance en $(1/\sqrt{NN_p})^{p+1}$ uniquement pour les raffinements les plus grossiers des maillages. Par exemple pour $p = 1$, l'ordre deux est bien atteint lorsque N augmente, mais ce n'est pas le cas pour $p = 3$ où l'ordre quatre est atteint pour $N = 2\,888$ puis décroît pour N plus grand. Par ailleurs, on note que pour le degré $p = 4$ l'ordre ne dépasse pas deux quel que soit N . Cependant, pour un maillage fixé, une solution d'ordre plus élevée permet une erreur significativement plus petite.

Tous ces aspects confondus permettent de comprendre qu'une fois un certain degré de polynôme et un certain raffinement de maillage atteints, la précision spatiale en h^{p+1} de la solution numérique explicite-implicite est rattrapée par la précision temporelle d'ordre deux du schéma numérique.

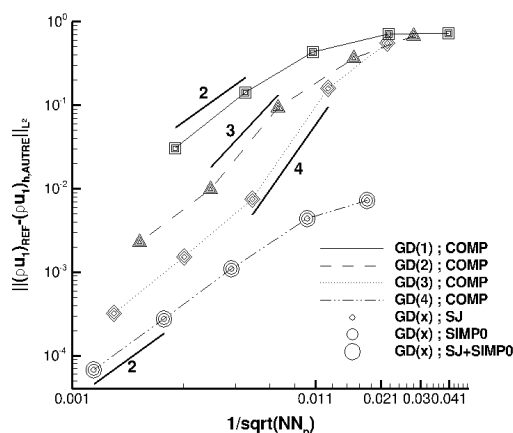


Fig. 6.4 – (*Interaction tourbillons-paroi*). Erreur en norme L^2 sur la quantité de mouvement projetée selon x_1 en fonction de $1/\sqrt{NN_p}$ (échelles logarithmiques). Comparaison des calculs COMP (trait plein), SJ (symboles petits), SIMP0 (symboles moyens) et SJ+SIMP0 (symboles grands) pour les maillages uniformes et pour différents degrés de polynôme. Les chiffres donnent la valeur des pentes des segments associés.

| | $1/\sqrt{NN_p}$ | Erreur : $\ (\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE}\ _{L^2(\Omega_h)}$ | | | | a |
|---------|-----------------------|--|------------------|------------------|------------------|------|
| | | COMP | SJ | SIMP0 | SJ + SIMP0 | |
| $p = 1$ | $1/\sqrt{200 * 3}$ | $7,215387E - 01$ | $7,215387E - 01$ | $7,215389E - 01$ | $7,215389E - 01$ | — |
| | $1/\sqrt{648 * 3}$ | $7,161010E - 01$ | $7,161010E - 01$ | $7,161011E - 01$ | $7,161011E - 01$ | 0,01 |
| | $1/\sqrt{2888 * 3}$ | $4,311248E - 01$ | $4,311248E - 01$ | $4,311264E - 01$ | $4,311264E - 01$ | 0,68 |
| | $1/\sqrt{10952 * 3}$ | $1,413693E - 01$ | $1,413693E - 01$ | $1,413739E - 01$ | $1,413739E - 01$ | 1,67 |
| | $1/\sqrt{43808 * 3}$ | $3,076470E - 02$ | $3,076469E - 02$ | $3,076846E - 02$ | $3,076844E - 02$ | 2,20 |
| $p = 2$ | $1/\sqrt{200 * 6}$ | $6,728332E - 01$ | $6,728332E - 01$ | $6,728381E - 01$ | $6,728381E - 01$ | — |
| | $1/\sqrt{648 * 6}$ | $3,635218E - 01$ | $3,635218E - 01$ | $3,635304E - 01$ | $3,635304E - 01$ | 1,05 |
| | $1/\sqrt{2888 * 6}$ | $9,260377E - 02$ | $9,260375E - 02$ | $9,262818E - 02$ | $9,262817E - 02$ | 1,83 |
| | $1/\sqrt{10952 * 6}$ | $9,884677E - 03$ | $9,884675E - 03$ | $9,888696E - 03$ | $9,888693E - 03$ | 3,36 |
| | $1/\sqrt{43808 * 6}$ | $2,300850E - 03$ | $2,300847E - 03$ | $2,304866E - 03$ | $2,304863E - 03$ | 2,10 |
| $p = 3$ | $1/\sqrt{200 * 10}$ | $5,570918E - 01$ | $5,570918E - 01$ | $5,570988E - 01$ | $5,570988E - 01$ | — |
| | $1/\sqrt{648 * 10}$ | $1,593688E - 01$ | $1,593688E - 01$ | $1,594031E - 01$ | $1,594031E - 01$ | 2,13 |
| | $1/\sqrt{2888 * 10}$ | $7,487701E - 03$ | $7,487691E - 03$ | $7,493355E - 03$ | $7,493351E - 03$ | 4,09 |
| | $1/\sqrt{10952 * 10}$ | $1,527034E - 03$ | $1,527079E - 03$ | $1,520074E - 03$ | $1,520076E - 03$ | 2,39 |
| | $1/\sqrt{43808 * 10}$ | $3,229160E - 04$ | $3,229205E - 04$ | $3,228140E - 04$ | $3,228142E - 04$ | 2,24 |
| $p = 4$ | $1/\sqrt{200 * 15}$ | $7,169216E - 03$ | $7,169215E - 03$ | $7,168894E - 03$ | $7,168892E - 03$ | — |
| | $1/\sqrt{648 * 15}$ | $4,403603E - 03$ | $4,403602E - 03$ | $4,403762E - 03$ | $4,403762E - 03$ | 0,83 |
| | $1/\sqrt{2888 * 15}$ | $1,101699E - 03$ | $1,101697E - 03$ | $1,101911E - 03$ | $1,101911E - 03$ | 1,85 |
| | $1/\sqrt{10952 * 15}$ | $2,772841E - 04$ | $2,773006E - 04$ | $2,773006E - 04$ | $2,779816E - 04$ | 2,07 |
| | $1/\sqrt{43808 * 15}$ | $6,724045E - 05$ | $6,724162E - 05$ | $6,746280E - 05$ | $6,746278E - 05$ | 2,04 |

Tab. 6.1 – (*Interaction tourbillons-paroi*). Ordres de précision a en espace tels que l'erreur en norme L^2 sur la quantité de mouvement projetée selon x_1 soit de l'ordre de $(1/\sqrt{NN_p})^a$ pour $1 \leq p \leq 4$.

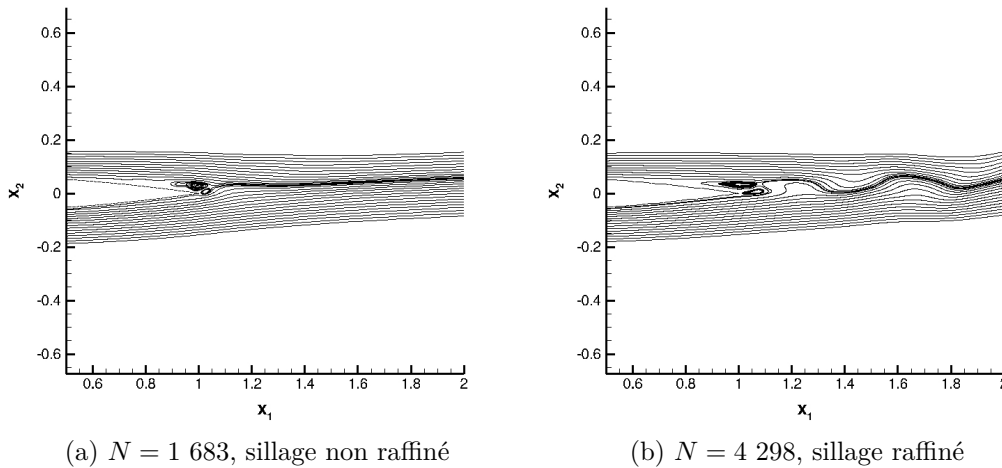


Fig. 6.5 – (*NACA0012*). Illustration de la bulle de décollement à l'extrados du profil d'aile. Comparaison entre les calculs COMP sur un maillage grossier sans instabilités de sillage (a) et sur un maillage raffiné avec instabilités de sillage (b) pour $p = 1$ et $t = 30$.

6.3.2 Écoulement autour du profil d'aile NACA0012

Ce cas test (cas test 3) concerne l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour d'un profil d'aile du comité consultatif national pour l'aéronautique (ou NACA) : le NACA0012 [78, 84], comme utilisé dans [96, 40].

6.3.2.1 Avant-propos

Les ailes NACA sont nées en 1929 dans les « souffleries de densité variable » du centre de recherche de Langley. L'objectif était d'obtenir des caractéristiques de référence d'une quantité importante d'ailes pour des nombres de Reynolds proches de ceux des vols réels des appareils moyens ($Re_\infty = \mathcal{O}(10^6)$) [78].

L'ensemble de ces ailes est regroupé en familles suivant la géométrie du profil et une série de chiffres leur est alors associée (ici 0012). En particulier, la famille des ailes NACA à quatre chiffres est la famille « mère » car elle prend en compte les deux paramètres fondamentaux pour définir une géométrie de surface portante : la ligne de cambrure (définie à partir des deux premiers chiffres) et la répartition d'épaisseur le long de cette ligne (définie à partir des deux derniers chiffres). On note que les profils NACA00xx sont symétriques.

On rappelle que si l'épaisseur est importante d'un point de vue structurel, la cambrure l'est d'un point de vue aérodynamique. C'est pourquoi les études à Langley ont débuté par des profils symétriques. On retient d'elles, que le meilleur rapport du maximum de portance sur le minimum de traînée, lequel doit être le plus grand possible pour un profil réussi, est obtenu pour une épaisseur d'environ 12% [84, figure 8]. Ceci explique l'utilisation du profil NACA0012 comme profil étalon dans les souffleries, et par suite logique comme profil de validation numérique dans les groupes de travail (par exemple [18, 96, 176]).

Ici, on a retenu la configuration « obligatoire » du projet Adigma [96], référencée comme le cas test « MTC3 »². C'est un cas test particulièrement intéressant qui exhibe une couche limite laminaire décollée avec présence d'une bulle de décollement à l'extrados du bord de fuite, et qui peut conduire à un comportement stationnaire ou instationnaire du sillage en fonction du raffinement du maillage choisi et de la discrétisation en temps utilisée (*cf.* figure 6.5 et § 6.3.2.2). On se sert donc les résultats du projet Adigma [135, 96], en partie repris par Dolejší *et al.* [40], pour valider les quatre discrétisations temporelles explicites-implicites.

6.3.2.2 A propos du cas test MTC3 du projet Adigma

On souhaite apporter des précisions sur le cas test MTC3 du projet Adigma servant de référence [135, 96]. En effet, ce cas test simule *a priori* un écoulement stationnaire présentant une couche limite laminaire avec bulle de décollement en bord de fuite. Cependant, on remarque que si l'ensemble des partenaires capture cette bulle, une minorité fait état de la présence des instabilités de sillage [96, chapitre 33]. Par conséquent, une analyse de convergence des résultats semble difficile.

On rappelle que ces instabilités correspondent à un transport de matière et sont donc principalement dues aux termes de convection. Leur capture est fonction des

2. Acronyme anglais : « **M**andatory **T**est **C**ase number **3** ».

résolutions en espace et en temps de ces mêmes termes. Dans le cadre du projet, la majorité des partenaires utilisait une méthode GD précise en espace combinée à une méthode implicite en temps connue pour dissiper les phénomènes de convection. Ceci par opposition à la minorité qui utilise la même méthode en espace combinée à une méthode explicite en temps plus précise favorisant la capture des instabilités. C'est le cas dans ce travail puisque les termes de convection sont explicités.

Ici, on fait le choix de comparer les valeurs des coefficients aérodynamiques obtenues avec celles du projet Adigma [135, 96] parce qu'on a observé des valeurs proches. Toutefois, par soucis de rigueur, on précise à chaque fois le type de discrétisation temporelle et si les instabilités de sillage ont été observées (*cf.* tableau 6.3).

6.3.2.3 Description du cas test

Le profil d'aile est tel que son bord d'attaque est situé en $(x_1; x_2) = (0; 0)^\top$. Celui retenu fait partie de la famille des NACA à quatre chiffres : le premier chiffre indique la cambrure du profil en pourcentage de longueur de corde L_c (ici 0 donc le profil est symétrique), le second chiffre indique la position de la cambrure en dixième de cette longueur, et les deux derniers chiffres indiquent l'épaisseur maximale du profil en pourcentage toujours de cette même longueur (ici 12 donc une épaisseur de $\delta_p = 0,12L_c$). De plus, le profil est entièrement déterminé à partir de son équation mathématique donnant l'épaisseur au-dessus et en-dessous de la ligne de cambrure :

$$x_2 = \pm \frac{\delta_p}{0,20} \left(0,29690\sqrt{x_1} - 0,12600x_1 - 0,35160x_1^2 + 0,28430x_1^3 - 0,10150x_1^4 \right).$$

Le détail complet de la géométrie du profil peut être consulté dans [84]. La condition aux limites appliquée est celle de non-glissement (paroi immobile et adiabatique). La frontière externe au domaine est située à environ $20L_c$ et les conditions aux limites appliquées sont celles d'entrée et de sortie subsoniques.

Au total trois maillages sont utilisés, lesquels sont décrits dans le tableau 6.2 et des vues globales et agrandies sont données sur les figures respectivement 6.6 et 6.7. Ils sont composés de triangles d'ordre élevé tels que $m = 4$. Concernant le maillage le plus grossier, il présente un unique raffinement des mailles vers la paroi. C'est pourquoi, il capte bien la bulle de décollement mais pas les instabilités de sillage. Concernant les deux autres maillages, ils présentent deux zones de raffinement afin de pouvoir contrôler au mieux l'épaisseur des mailles dans les directions des forts gradients de l'écoulement. D'une part, il y a la zone de proche paroi où se développe la couche limite et qui correspond aux mailles rectangulaires coupées en deux suivant une diagonale. D'autre part, il y a la zone de sillage où se développent les instabilités et qui correspond aux mailles resserrées le long de la ligne de fuite. Ainsi, ces deux maillages captent les deux phénomènes physiques de bulle de décollement et d'échappement tourbillonnaire.

Concernant l'initialisation, le vecteur des variables conservatives $\mathbf{w}^{(0)}(\mathbf{x})$ est donné pour $\rho^{(0)} = 1$, $u_1^{(0)} = \cos(\alpha)$, $u_2^{(0)} = \sin(\alpha)$ avec l'incidence $\alpha = 2^\circ$ et $\theta^{(0)} = 1$, et il est uniforme pour tous les calculs. Le nombre de Reynolds amont est basé sur la longueur de corde du profil et la vitesse amont, il est fixé à $Re_\infty = 5\,000$ et le nombre de Mach amont est fixé à $M_\infty = 0,5$. De plus, on s'intéresse à une

simulation pour un temps physique $t = 60$. Le lecteur peut se référer à [96, 40] pour plus de détails sur l'initialisation de ce cas test. Enfin, les valeurs de la condition de CFL utilisées ont été données au paragraphe 6.2.1.

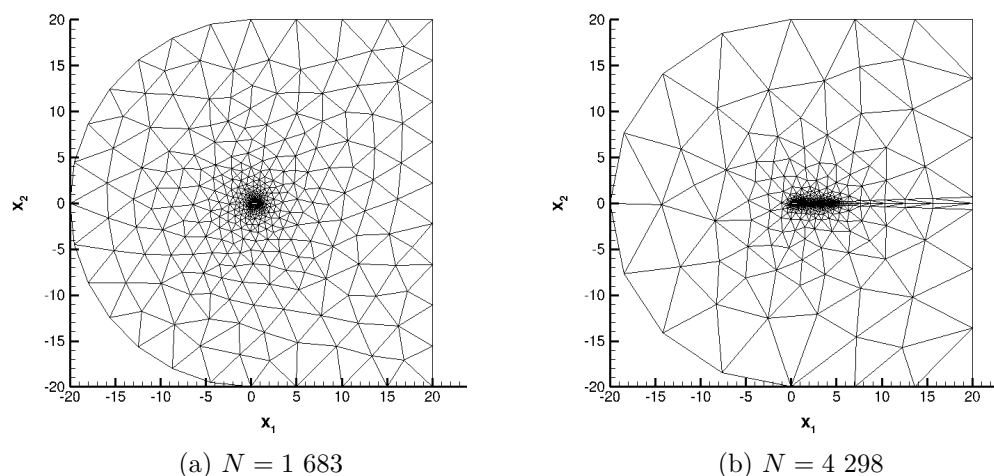


Fig. 6.6 – (NACA0012). Vue globale des maillages en C utilisés dans le cas, soit d'un raffinement unique à la paroi (a), soit d'un raffinement à la paroi et dans le sillage (b).

| N | Nb. points sur Γ_p ($m = 4$) | RA |
|-------|--|-------|
| 1 683 | 537 | 444 |
| 4 298 | 365 | 2 125 |
| 7 368 | 481 | 1 685 |

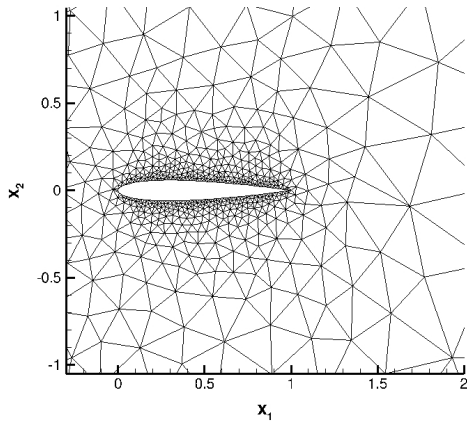
Tab. 6.2 – (NACA0012). Description des maillages utilisés. Abréviations : Nombre (Nb.).

6.3.2.4 Description des résultats numériques

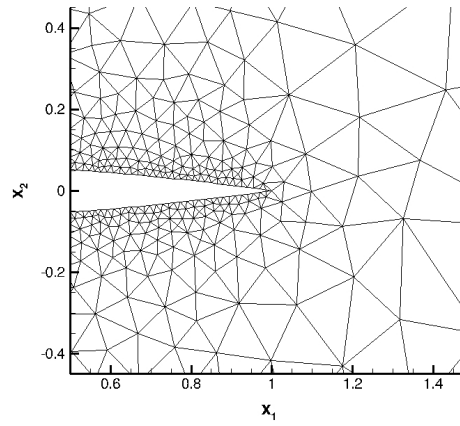
La description des résultats numériques suit la même logique que pour le cas test précédent : vérification de la capture des phénomènes physiques et précision des quatre discrétisations temporelles COMP, SJ, SIMP0 et SJ+SIMP0.

Iso-valeurs du nombre de Mach local. La figure 6.8 présente les iso-valeurs de M . Plus précisément, la figure 6.8a reproduit la simulation GD de discrétisation temporelle explicite de référence de Taube *et al.* [96, chapitre 5, figure 1], réalisée telle que $\mathcal{N} = 33\ 600$ ($p = 5$). Quant à elles, les figures 6.8b à 6.8f illustrent les simulations GD de discrétisations temporelles COMP, SJ, SIMP0 et SJ+SIMP0, réalisées sur le maillage intermédiaire et telles que $\mathcal{N} = 42\ 980$ ($N = 4\ 298$, $p = 3$) et $t = 60$.

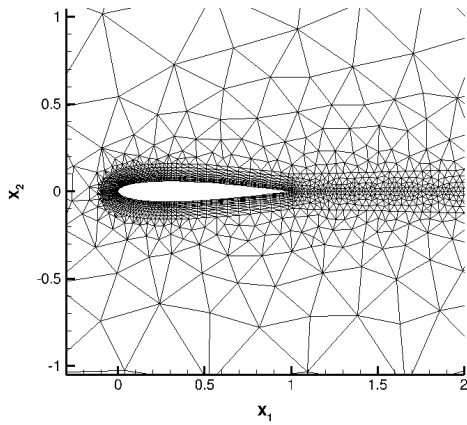
Sur les figures 6.8b à 6.8e, on observe la très bonne capture des instabilités de sillage pour chacun des quatre calculs explicites-implicites en comparaison avec la



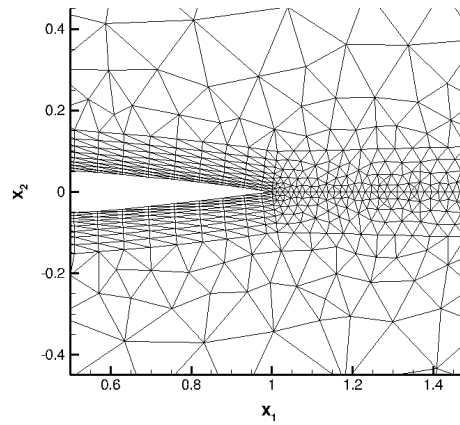
(a) $N = 1\ 683$



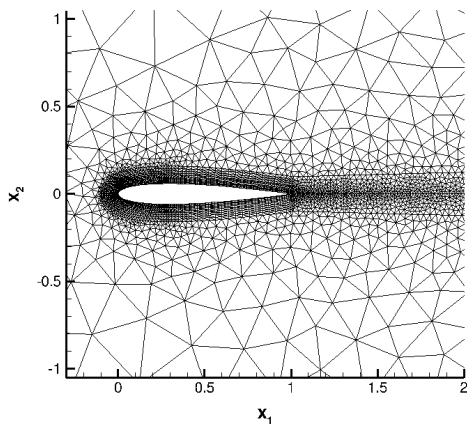
(b) $N = 1\ 683$



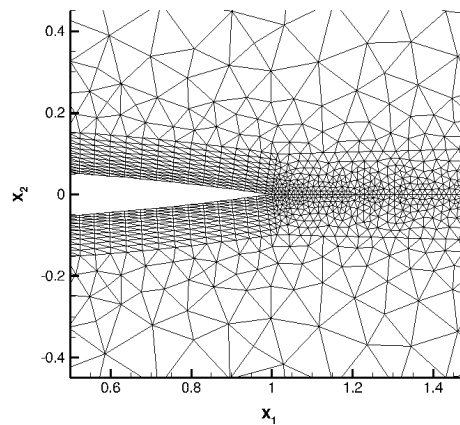
(c) $N = 4\ 298$



(d) $N = 4\ 298$



(e) $N = 7\ 368$



(f) $N = 7\ 368$

Fig. 6.7 – (NACA0012). Vues agrandies des maillages utilisés : profil complet (gauche) et bord de fuite où se situe la bulle de décollement (droite).

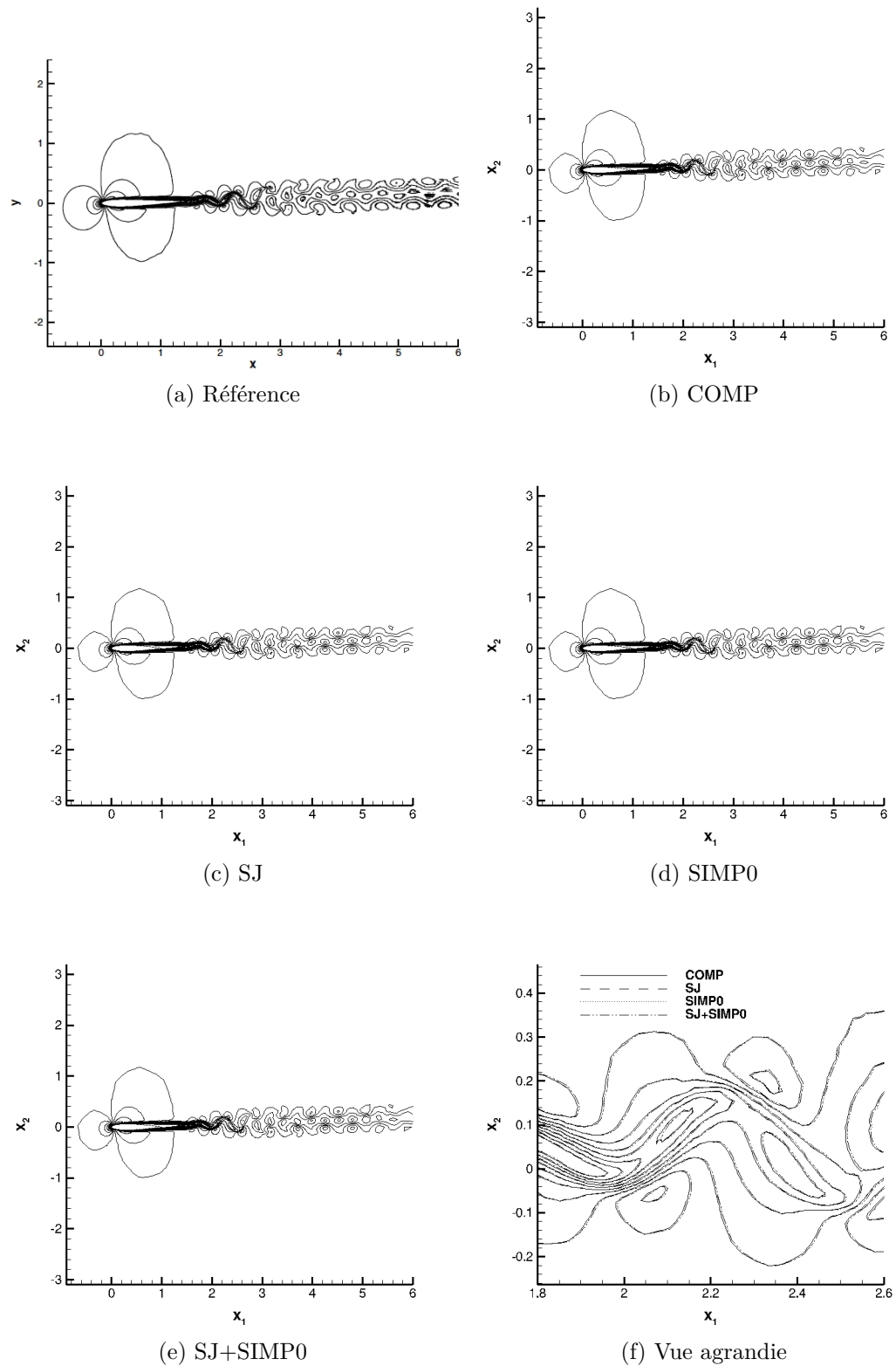


Fig. 6.8 – (*NACA0012*). Iso-valeurs du nombre de Mach local. Comparaison entre la simulation GD explicite de référence telle que $\mathcal{N} = 33\,600$ à un instant t non précisé [96, chapitre 5, figure 1] (a) et les simulations GD explicites-implicites telles que $\mathcal{N} = 42\,980$ à $t = 60$ (b), (c), (d) et (e). La figure (f) montre une vue agrandie. Il y a 20 niveaux représentés : 0,05; ...; 0,60.

référence [96, chapitre 5], et ce pour une résolution comparable du nombre total de degrés de liberté par équation. Sur la vue agrandie figure 6.8f, on note également une bonne superposition des iso-valeurs de l'ensemble des calculs, laissant supposer peu d'écart de précision entre eux.

Valeurs des coefficients de traînée et de portance. L'étude des valeurs des coefficients aérodynamiques C_D et C_L est organisée en deux temps. Dans un premier temps, on compare ces valeurs entre la discrétisation explicite-implicite COMP et celles du projet Adigma servant de référence. Dans un second temps, on compare ces valeurs entre les quatre discrétisations COMP, SJ, SIMP0 et SJ+SIMP0 et la discrétisation EXPL équivalente.

Remarque 6.1. *Concernant les calculs de référence, ils sont principalement issus du projet Adigma. Concernant les calculs COMP, la valeur des coefficients correspond à une moyenne temporelle : soit pour t appartenant à $[50, 55]$ pour le maillage le plus grossier ne capturant pas les instabilités de sillage, soit sur 10 périodes à partir de $t = 50$ pour les deux autres maillages capturant ces instabilités (cf. § 6.2).*

Comparaison : résultats COMP et Adigma. D'une part, le tableau 6.3 indique les valeurs de C_D et C_L de plusieurs simulations GD explicites ou implicites extraites de la littérature [135, 96, 40] et des simulations GD de discrétisation temporelle COMP pour tous les maillages et pour $1 \leq p \leq 4$. On précise que \mathcal{N} , le nombre total de degrés de liberté par équation, permet une comparaison globale de toutes les simulations, et le symbole \star permet d'identifier les simulations capturant des instabilités de sillage (cf. § 6.3.2.2 pour plus de détails sur cette capture).

On observe que l'ensemble des valeurs des coefficients C_D et C_L obtenues avec les calculs COMP se rapproche (par valeurs supérieures) des valeurs de référence les plus hautes, et ce plus le maillage est raffiné et plus le degré p est élevé. Ce point démontre une certaine convergence globale en maillage et également une en ordre. Par ailleurs, on voit que pour $\mathcal{N} = 110\,520$ les deux coefficients aérodynamiques sont compris dans la gamme des valeurs de référence.

D'autre part, la figure 6.9 donne les valeurs des coefficients C_D et C_L des simulations GD de discrétisation temporelle COMP en fonction de \mathcal{N} pour tous les maillages et pour $1 \leq p \leq 4$. Pour chacun des coefficients on indique un intervalle de confiance calculé par une extrapolation de Richardson [141, 145, 96, chapitre 33] sur les deux maillages les plus raffinés et pour $p = 4$. Il vient :

$$\begin{aligned} C_{D, Richardson} &= (5,703 \pm 0,234) \times 10^{-2}, \\ C_{L, Richardson} &= (4,127 \pm 0,723) \times 10^{-2}. \end{aligned}$$

Une première observation est la différence de taille de l'intervalle de confiance entre les deux coefficients, à savoir un intervalle presque trois fois plus important pour le coefficient C_L . Ceci montre une convergence plus lente à s'établir pour ce dernier, comme relevé par les partenaires d'Adigma [135, 96, 40]. La seconde observation est le nombre total de degrés de liberté par équation minimal pour converger, à savoir $\mathcal{N} = 12\,894$ pour le coefficient C_D et $\mathcal{N} = 25\,788$ pour le coefficient C_L . Ces deux valeurs sont très représentatives de celles de référence indiquées dans le tableau 6.3.

| Simulation | | \mathcal{N} | Détail | | $C_D (10^{-2})$ | $C_L (10^{-2})$ |
|-----------------------|------------------|-------------------|--------|---|-----------------|-----------------|
| DLR | [135] | 6 400 | I | | 5,61 | 3,747 |
| Cenaero | [135] | $\approx 40\,000$ | I | * | 5,725 | 3,860 |
| U. de Bergame | [135] | 9 408 | I | | 5,65 | 3,853 |
| U. de Nottingham | [135] | $\approx 20\,000$ | I | | 5,608 | 3,721 |
| U. de Stuttgart | [135] | 14 300 | E | * | 5,70 | 4,35 |
| U. de Nanjing | [96, chapitre 6] | 14 412 | E | | 5,55 | 4,25 |
| Onera | [96, chapitre 8] | 19 200 | E | * | 5,604 | 3,640 |
| Dolejší <i>et al.</i> | [40] | 88 494 | I | | 5,60683 | 3,7171 |
| $N = 1\,683$ | $p = 1$ | 5 049 | COMP | | 6,470 | 1,504 |
| 1 683 | 2 | 10 098 | COMP | | 5,894 | 5,240 |
| 1 683 | 3 | 16 830 | COMP | | 5,830 | 5,601 |
| 1 683 | 4 | 25 245 | COMP | | 5,794 | 4,643 |
| 4 298 | 1 | 12 894 | COMP | * | 5,888 | 5,538 |
| 4 298 | 2 | 25 788 | COMP | * | 5,752 | 4,341 |
| 4 298 | 3 | 42 980 | COMP | * | 5,758 | 4,427 |
| 4 298 | 4 | 64 470 | COMP | * | 5,754 | 4,242 |
| 7 368 | 1 | 22 104 | COMP | * | 5,766 | 4,885 |
| 7 368 | 2 | 44 208 | COMP | * | 5,753 | 4,366 |
| 7 368 | 3 | 73 680 | COMP | * | 5,774 | 4,528 |
| 7 368 | 4 | 110 520 | COMP | * | 5,716 | 4,157 |

Tab. 6.3 – (*NACA0012*). Valeur des coefficients de traînée et de portance. Comparaison entre les simulations GD explicites ou implicites de référence (haut) et la simulation GD explicite-implicite COMP sur tous les maillages pour $1 \leq p \leq 4$ (bas). Abréviations : Université (U.), Implicite (I), Explicite (E). Le symbole \star indique que le sillage présente des instabilités (*cf.* § 6.3.2.2).

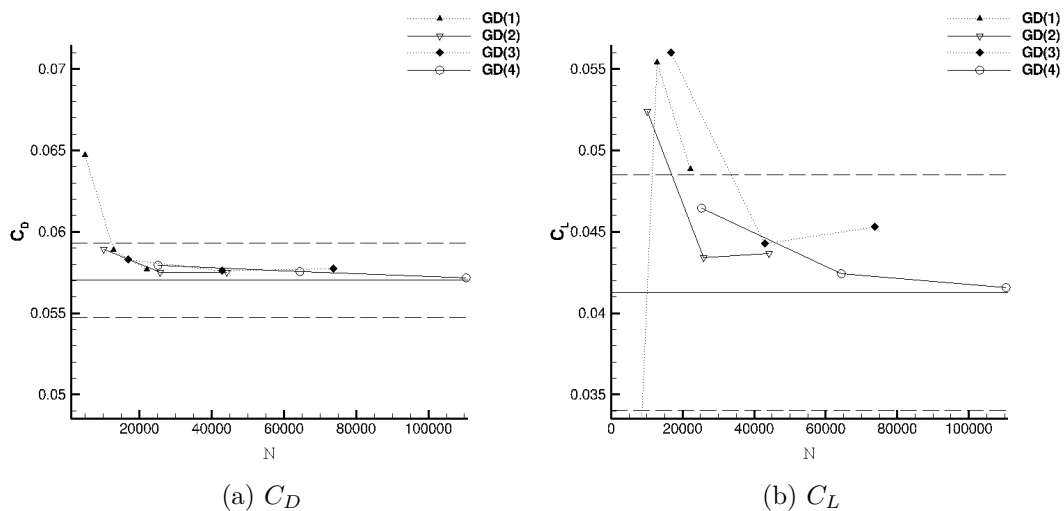


Fig. 6.9 – (*NACA0012*). Convergence des coefficients de traînée (a) et de portance (b) en fonction du nombre total de degrés de liberté par équation. Calculs pour la discrétisation COMP sur tous les maillages pour $1 \leq p \leq 4$. Les traits horizontaux correspondent à un intervalle de confiance calculé par une extrapolation de Richardson.

Comparaison : résultats COMP, SJ, SIMP0, SJ+SIMP0 et EXPL. Elle est reportée au tableau 6.4 donnant l'erreur en pourcentage (6.2) faite sur les coefficients C_D et C_L entre chaque discrétisation explicite-implicite et son équivalent explicite. Ici, on s'intéresse à un seul calcul où $N = 4\,298$, $p = 3$ et $t = 60$.

Tout d'abord, pour chacun des coefficients, les valeurs du tableau indiquent une erreur très petite et similaire pour toutes les discrétisations : environ 0,4% pour C_D et 2% pour C_L . Ensuite, on observe que l'erreur sur le coefficient C_L est presque cinq fois plus importante que celle sur le coefficient C_D , traduisant à nouveau une convergence plus difficile à établir pour le coefficient de portance. Enfin, on remarque, principalement avec le coefficient C_L , que les calculs COMP et SJ d'une part, et les calculs SIMP0 et SJ+SIMP0 d'autre part, présentent une erreur quasi-identique.

| | COMP | SJ | SIMP0 | SJ+SIMP0 |
|------------------|------|------|-------|----------|
| Erreur C_D (%) | 0,36 | 0,37 | 0,37 | 0,37 |
| Erreur C_L (%) | 1,87 | 1,86 | 2,14 | 2,13 |

Tab. 6.4 – (NACA0012). Erreur en pourcentage sur les coefficients de traînée et de portance pour les méthodes COMP, SJ, SIMP0 et SJ+SIMP0. Calculs sur le maillage $N = 4\,298$ pour $p = 3$ et $t = 60$.

6.4 Validation sur des cas d'applications

Deux cas d'applications pratiques sont ici considérés pour la validation des quatre méthodes COMP, SJ, SIMP0 et SJ+SIMP0. Il s'agit des cas tests 4 et 5 dont les résultats sont respectivement présentés aux paragraphes 6.4.1 et 6.4.2.

6.4.1 Écoulement autour du profil d'aile SD7003

A l'image du cas test 3, celui-ci (cas test 4) concerne également l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour d'un profil d'aile de Selig et Donovan : le SD7003 [150], comme utilisé dans [166, 113].

6.4.1.1 Avant-propos

Les ailes SD sont nées en 1989 dans les souffleries de l'université de Princeton suite à des tests sur des maquettes de planeurs radio-controlées. L'objectif était d'obtenir des caractéristiques de référence d'une quantité importante d'ailes pour des nombres de Reynolds proches de ceux des vols réels des appareils de petites tailles et de faibles vitesses comme les drones ou les micro-véhicules aériens ($10^3 \leq Re_\infty \leq 10^5$) [150].

Le profil SD7003 appartient à la famille des profils d'aile pour les maquettes de planeurs. Outre ses performances aérodynamiques reconnues puisqu'il a permis à Joe Wurts et Daryl Perkins d'obtenir respectivement la première et seconde place au championnat du monde F3B de maquette en 1991, ce profil est souvent étudié numériquement pour la large bulle de décollement qui se forme à l'extrados du profil, et ce même à faible incidence (*cf.* figure 6.10 et [48, figure 5.23]).

On rappelle que cette bulle est connue pour perturber les performances aérodynamiques des appareils (augmentation de la traînée, décrochage) et son contrôle est un vaste sujet de recherches. Par exemple, le lecteur peut consulter les études de Galbraith [48] et de Visbal *et al.* [173] sur le profil SD7003 pour une large gamme de nombre de Reynolds ($10^4 \leq Re_\infty \leq 9 \times 10^4$) et d'angle d'attaque ($2^\circ \leq \alpha \leq 14^\circ$). En particulier, on retient que dans le cas considéré ici, soit $Re_\infty = 10^4$ et $\alpha = 4^\circ$, l'écoulement est bien instationnaire, bidimensionnel et laminaire avec de faibles variations dans la direction de l'envergure, la couche limite se décolle avec formation d'une bulle mais sans recollement, et il s'ensuit un détachement de tourbillons périodiques.

Parmi les études GD sur le profil SD7003, on cite [166, 113, 110, 176]. Dans ce travail, on se sert de récentes simulations utilisant, soit une méthode des différences finies au sixième ordre [48, 173], soit une méthode GD [166, 113] pour valider les quatre discrétisations temporelles explicites-implicites.

Finalement, ce cas test est dit d'application car d'une part, le profil d'aile cambré et le nombre de Reynolds important sont à l'origine de l'accélération à l'extrados induisant un fort décollement de la couche limite, et d'autre part, les instationnarités de sillage sont très marquées.

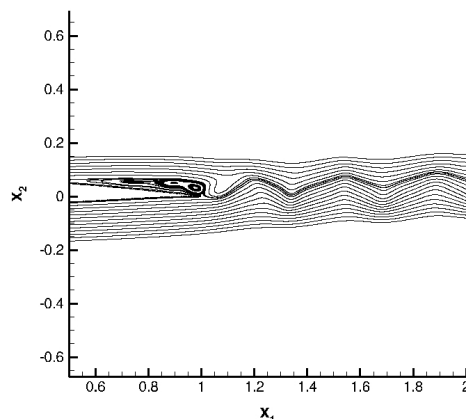


Fig. 6.10 – (SD7003). Illustration de la bulle de décollement à l'extrados du profil d'aile. Calcul COMP sur le maillage $N = 11\,676$ pour $p = 2$ et $t = 25$.

6.4.1.2 Description du cas test

Le profil d'aile est tel que son bord d'attaque est situé en $(x_1; x_2) = (0; 0)^\top$. Ce profil est non symétrique. Soit L_c sa longueur de corde, alors son épaisseur est de $0,085L_c$, sa cambrure de $0,0145L_c$ en $x_1 = 0,35L_c$ et son bord de fuite est arrondi avec un rayon de $r = 10^{-4}L_c$. Le détail complet de la géométrie du profil peut être consulté dans [150, 176]. La condition aux limites appliquée est celle de non-glissement (paroi immobile et adiabatique). La frontière externe au domaine est située à environ $20L_c$ et les conditions aux limites appliquées sont celles d'entrée et de sortie subsoniques.

Un seul maillage est utilisé, lequel est décrit dans le tableau 6.5 et deux vues, une globale et une agrandie, sont données sur les figures respectivement, 6.11a et 6.11b. Il est composé de triangles d'ordre élevé tels que $m = 4$. Comme pour le profil NACA0012, il y a deux zones de raffinement.

Concernant l'initialisation, le vecteur des variables conservatives $\mathbf{w}^{(0)}(\mathbf{x})$ est donné pour $\rho^{(0)} = 1$, $u_1^{(0)} = \cos(\alpha)$, $u_2^{(0)} = \sin(\alpha)$ avec l'incidence $\alpha = 4^\circ$ et $\theta^{(0)} = 1$, et il est pris uniforme pour tous les calculs. Le nombre de Reynolds amont est basé sur la longueur de corde du profil et la vitesse amont, il est fixé à $Re_\infty = 10\,000$ et le nombre de Mach amont est $M_\infty = 0,2$. De plus, on s'intéresse à une simulation pour un temps physique $t = 25$. Le lecteur peut se référer à [166, 113] pour plus de détails sur l'initialisation de ce cas test. Enfin, les valeurs de la condition de CFL utilisées ont été données au paragraphe 6.2.1.

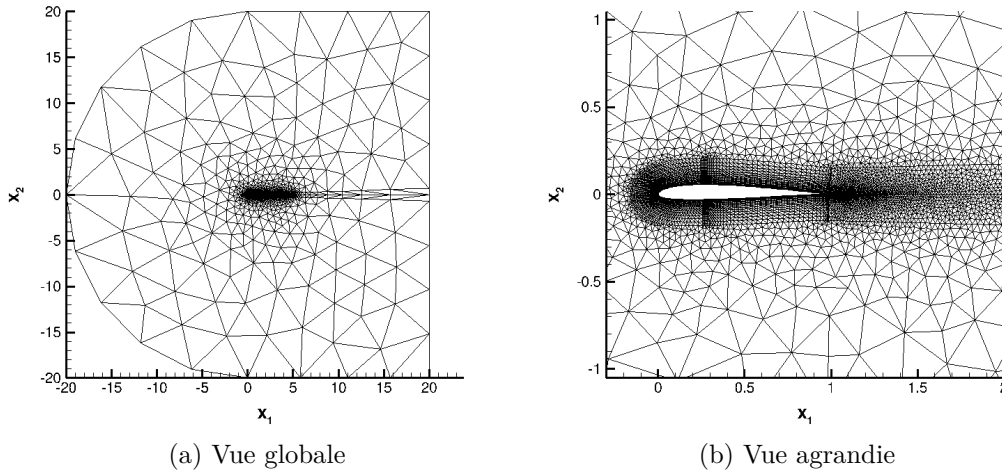


Fig. 6.11 – (*SD7003*). Maillage utilisé.

| N | Nb. points sur Γ_p ($m = 4$) | RA |
|--------|--|-------|
| 11 676 | 633 | 5 191 |

Tab. 6.5 – (*SD7003*). Description du maillage utilisé. Abréviation : Nombre (Nb.).

6.4.1.3 Description des résultats numériques

La description des résultats numériques est similaire à celle des cas tests 2 et 3.

Iso-valeurs de la vorticit . La figure 6.12 pr sente les iso-valeurs de ω . Plus pr cis ment, la figure 6.12a reproduit la simulation GD de r f rence de Uranga *et al.* [166, figure³ 7], r alis e en trois dimensions d'espace telle que $p = 3$ et $t = 25$. Quant   elles, les figures 6.12b   6.12f illustrent les simulations GD de discr tisations temporelles COMP, SJ, SIMP0 et SJ+SIMP0, r alis es telles que $p = 2$ et $t = 25$.

Sur ces figures, on observe la tr s bonne capture des instabilit s de sillage pour chacun des calculs explicites-implicites en comparaison avec la r f rence [166] (ou [113, figure 2.b]). Cette observation est confirm e sur la vue agrandie figure 6.12f, en particulier avec la capture des tourbillons dans le sillage sans trop de dissipation.

3. Les couleurs de l'image ont  t  converties en  chelle de gris.

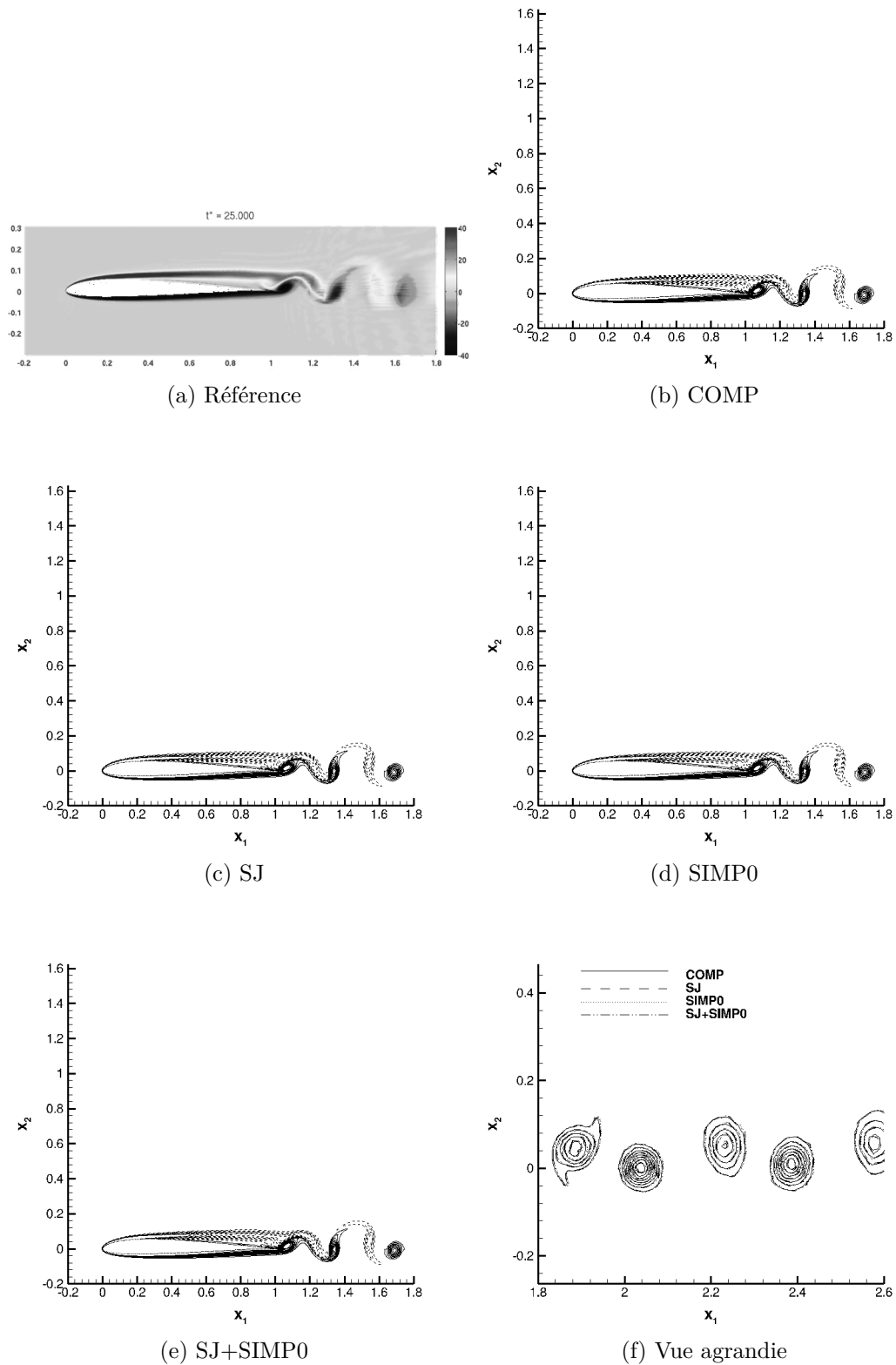


Fig. 6.12 – (*SD7003*). Iso-valeurs de la vorticité. Comparaison à $t = 25$ entre la simulation GD de référence telle que $p = 3$ [166, figure 7] (a) et les simulations GD explicites-implicites telles que $p = 2$ (b), (c), (d), et (e). La figure (f) montre une vue agrandie. Il y a 20 niveaux représentés sur l'intervalle $[-40, 40]$, en trait pointillé pour les valeurs négatives ou en trait plein pour les valeurs positives.

Valeurs des coefficients de traînée et de portance. L'étude des valeurs des coefficients aérodynamiques C_D et C_L est organisée en deux temps à l'image du cas test 2 sur le profil NACA0012.

Comparaison : résultats COMP et références. Elle est reportée au tableau 6.6, lequel indique les valeurs de C_D et C_L de plusieurs simulations extraites de la littérature [48, 166] et des simulations COMP pour $1 \leq p \leq 4$.

Remarque 6.2. Concernant les calculs COMP, la valeur indiquée des coefficients correspond à une moyenne temporelle sur 10 périodes à partir de $t = 20$.

D'une façon générale, l'ensemble des valeurs des coefficients C_D et C_L obtenues avec les calculs COMP est en accord avec celles de référence, notamment dès $p = 2$. De plus, comme le profil NACA0012, on observe des valeurs du coefficient C_D plus rapprochées que celles du coefficient C_L .

| Simulation | | Détail | | $C_D (10^{-2})$ | $C_L (10^{-1})$ |
|----------------------|---------|-----------------|--|-----------------|-----------------|
| Galbraith | [48] | 3D, DF, Ordre 6 | | 4,7 | 3,6 |
| Uranga <i>et al.</i> | [166] | 3D, GD, $p = 3$ | | 4,978 | 3,755 |
| $N = 11\ 676$ | $p = 1$ | 2D, GD, COMP | | 5,015 | 4,042 |
| $11\ 676$ | 2 | 2D, GD, COMP | | 4,934 | 3,692 |
| $11\ 676$ | 3 | 2D, GD, COMP | | 4,928 | 3,697 |
| $11\ 676$ | 4 | 2D, GD, COMP | | 4,956 | 3,701 |

Tab. 6.6 – (SD7003). Valeur des coefficients de traînée et de portance. Comparaison entre les simulations de référence (haut) et la simulation COMP pour $1 \leq p \leq 4$ (bas). Abréviations : deux ou trois dimensions d'espace (2D ou 3D), différences finies (DF), Galerkin discontinue (GD).

Comparaison : résultats COMP, SJ, SIMP0, SJ+SIMP0 et EXPL. Elle est reportée au tableau 6.7 donnant l'erreur en pourcentage (6.2) faite sur les coefficients C_D et C_L entre chaque discrétisation explicite-implicite et son équivalent explicite. Ici, on s'intéresse à un seul calcul où $p = 2$ et $t = 25$.

Pour chacun des coefficients, les valeurs du tableau indiquent une erreur très petite et similaire pour toutes les discrétisations explicites-implicites, soit inférieure à 0,1%. De plus, comme le profil NACA0012, on observe, cette fois-ci avec les deux coefficients, que les calculs COMP et SJ d'une part, et les calculs SIMP0 et SJ+SIMP0 d'autre part, présentent une erreur quasi-identique.

| | COMP | SJ | SIMP0 | SJ+SIMP0 |
|------------------|------|------|-------|----------|
| Erreur C_D (%) | 0,04 | 0,05 | 0,02 | 0,02 |
| Erreur C_L (%) | 0,02 | 0,04 | -0,09 | -0,09 |

Tab. 6.7 – (SD7003). Erreur en pourcentage sur les coefficients de traînée et de portance pour les méthodes COMP, SJ, SIMP0 et SJ+SIMP0. Calculs pour $p = 2$ et $t = 25$.

6.4.2 Écoulement autour d'un cylindre

Le dernier cas test (cas test 5) est l'écoulement externe, instationnaire et subsonique d'un fluide visqueux autour et en aval d'un cylindre comme utilisé dans le programme national pour les applications axées sur la recherche en CFD (ou « NPARC »⁴) [79].

6.4.2.1 Avant-propos

D'une façon générale, ce cas test est très répandu en mécanique des fluides tant pour les études expérimentales (*cf.* [149, chapitres II et IX]) que numériques (*cf.* [65, chapitres 2 et 11]). En effet, en science on peut relever deux priorités fondamentales : celle de la reproductibilité, permettant de démontrer l'existence de lois, et celle de l'étalonnage des mesures, permettant les comparaisons des mesures. A ce titre, la configuration simple du cylindre est un avantage certain.

La littérature abondante sur ce corps de référence a permis de mettre en évidence les mécanismes physiques successifs apparaissant dans l'écoulement en fonction du nombre de Reynolds. Par exemple, le lecteur peut consulter l'article de Williamson [179] et ses références pour l'examen de ces mécanismes et un historique des résultats expérimentaux jusqu'en 1996. En particulier, on retient que dans le cas considéré ici, soit $Re_\infty = 150$, l'écoulement est bien instationnaire, bidimensionnel et laminaire, la couche limite se décolle en aval du cylindre sous le poids des forces d'inertie, et il s'ensuit deux zones de recirculation qui se détachent alternativement et sont convectées dans le sillage formant ainsi une allée de Bénard-von Kàrmàn. On note que la symétrie de l'écoulement en moyenne temporelle conduit à un coefficient de portance nul en moyenne.

Parmi les études GD récentes sur le cylindre, on cite [107, 20, 110]. Dans ce travail, on utilise à la fois des résultats issus d'expérimentations [146, 163, 179] et de simulations utilisant, soit une méthode particules-grilles [120], soit une méthode des différences finies au troisième ordre [115] pour valider les quatre discrétisations temporelles explicites-implicites.

Finalement, ce cas test est dit d'application car les effets visqueux jouent un rôle prédominant dans la physique de l'écoulement à la paroi et sont à l'origine d'un manque de stabilité lors d'une discrétisation temporelle explicite. En effet, le schéma EXPL est instable pour tous les degrés de polynôme considérés ($0 \leq p \leq 6$) et toutes les valeurs de la condition de CFL non réhibitoires en temps de calcul (ici, on a choisi $CFL \geq 10^{-2}$).

6.4.2.2 Description du cas test

Le cylindre est de diamètre $L_c = 1,0$ et est centré en $(x_1; x_2) = (0; 0)^\top$. La condition aux limites appliquée est celle de non-glissement (paroi immobile et adiabatique). La frontière externe au domaine est circulaire de diamètre $30L_c$ et les conditions aux limites appliquées sont celles d'entrée et de sortie subsoniques.

Un seul maillage est utilisé, lequel est décrit dans le tableau 6.8 et des vues, globale et agrandie, sont données sur les figures respectivement, 6.13a et 6.13b.

4. Acronyme anglais : « **N**ational **P**rogram for **A**pplications-oriented **R**esearch in **C**FD ».

En outre, on a indiqué l'emplacement du point \mathbf{x}_{St} servant à calculer le nombre de Strouhal (cf. remarque 6.3). Il est composé de triangles d'ordre élevé tels que $m = 2$. À l'identique des deux précédents cas tests d'écoulement externe, ce maillage présente deux zones de raffinement. On note également qu'en amont du cylindre, les rectangles sont coupés en chevrons par rapport à l'axe des abscisses pour respecter au mieux la symétrie de l'écoulement.

Concernant l'initialisation, le vecteur des variables conservatives $\mathbf{w}^{(0)}(\mathbf{x})$ est donné pour $\rho^{(0)} = 1$, $u_1^{(0)} = 1$, $u_2^{(0)} = 0$ et $\theta^{(0)} = 1$, et il est pris uniforme pour tous les calculs. Le nombre de Reynolds amont est basé sur le diamètre du cylindre et la vitesse amont, il est fixé à $Re_\infty = 150$ et le nombre de Mach amont est $M_\infty = 0,2$. De plus, on s'intéresse à une simulation pour un temps physique $t = 150$. Le lecteur peut se référer à [79] pour plus de détails sur l'initialisation de ce cas test. Enfin, les valeurs de la condition de CFL utilisées ont été données au paragraphe 6.2.1.

Remarque 6.3. (Nombre de Strouhal) Ce nombre caractérise l'instabilité de Bénard-von Kármán grâce à la fréquence de détachement des tourbillons f_{ref} (cf. relation (1.9)). Dans la pratique, cette fréquence correspond à une valeur moyenne temporelle de la période de la pression (par exemple) en un point \mathbf{x}_{St} situé derrière le corps solide et éloigné du centre de la ligne de fuite. En effet, si cette dernière condition n'est pas respectée, alors la valeur de f_{ref} est doublée car la ligne de fuite est influencée par les tourbillons inférieurs et supérieurs [146]. Ici, on a pris $\mathbf{x}_{St} = (1,070; 0,313)^\top$ comme suggéré dans [79].

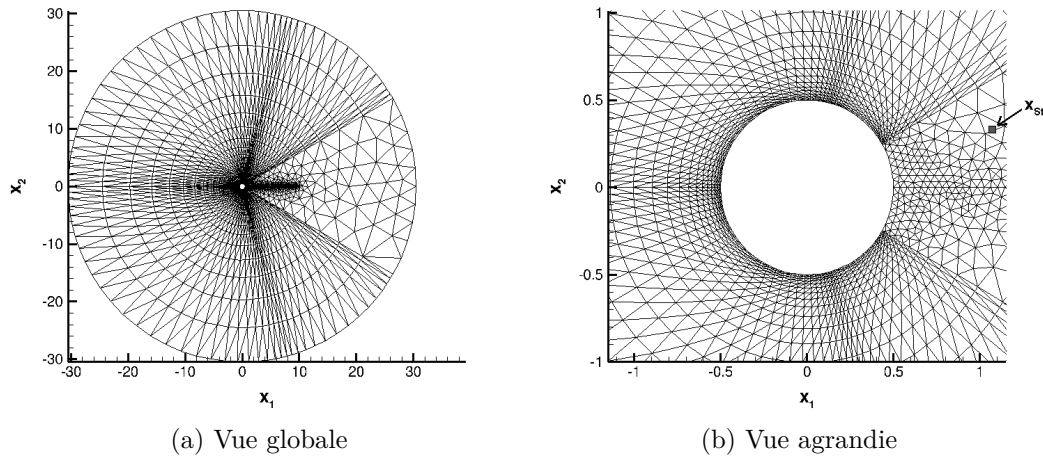


Fig. 6.13 – (Cylindre). Maillage utilisé et point \mathbf{x}_{St} servant à calculer le nombre St .

| N | Nb. points sur Γ_p ($m = 2$) | RA |
|-------|--|-----|
| 7 384 | 185 | 630 |

Tab. 6.8 – (Cylindre). Description du maillage utilisé. Abréviation : Nombre (Nb.).

6.4.2.3 Description des résultats numériques

La description des résultats numériques est similaire à celle des précédents cas tests.

Iso-valeurs du nombre de Mach local. La figure 6.14 présente les iso-valeurs de M . Plus précisément, la figure 6.14a reproduit la simulation avec la méthode des différences finies d'ordre trois de référence de Nichols et Heikkinen [115, figure⁵ 13]. Quant à elles, les figures 6.14b à 6.14f illustrent les simulations GD de discrétisations temporelles COMP, SJ, SIMP0 et SJ+SIMP0, réalisées telles que $p = 2$ et $t = 150$.

Sur ces figures, on observe la très bonne capture de l'allée de Bénard-von Kàrmàn pour chacun des quatre calculs explicites-implicites en comparaison avec la référence [115, chapitre 13]. Sur la vue agrandie figure 6.14f, on note également une superposition forte des iso-valeurs pour les discrétisations COMP et SJ d'une part, et les discrétisations SIMP0 et SJ+SIMP0 d'autre part, laissant supposer peu d'écarts de précision entre elles. On remarque que ce dernier point vient appuyer les résultats de précision sur l'erreur en pourcentage observés avec les deux cas tests 3 et 4.

Valeurs du coefficient de traînée et du nombre de Strouhal. L'étude des valeurs du coefficient aérodynamique C_D et du nombre de Strouhal est organisée en deux temps à l'image des précédents cas tests sur les profils d'aile.

Comparaison : résultats COMP et références. Elle est reportée au tableau 6.9, lequel indique les valeurs de C_D et St d'études expérimentales [146, 163, 179] et de simulations numériques [120, 115] extraites de la littérature et des simulations COMP pour $1 \leq p \leq 3$.

D'une façon générale, l'ensemble des valeurs obtenues avec les calculs COMP est en accord avec celles de référence, et plus particulièrement dès $p = 2$.

Remarque 6.4. *Concernant les expérimentations de référence, on se restreint aux auteurs les plus mentionnés [146], Tritton [163] et Williamson [179]. On précise que l'on a reproduit en annexe E, les copies des courbes où ont été lues les différentes valeurs reportées dans le tableau 6.9. Concernant les calculs COMP, la valeur indiquée des deux grandeurs correspond à une moyenne temporelle sur les 20 dernières périodes avant $t = 150$.*

Comparaison : résultats SJ, SIMP0, SJ+SIMP0 et COMP. Elle est reportée au tableau 6.10 donnant l'erreur en pourcentage (6.2) faite sur les grandeurs C_D et St entre chaque discrétisation explicite-implicite simplifiée et son équivalent explicite-implicite COMP, et ce faute de pouvoir obtenir une discrétisation EXPL équivalente comme indiquée dans l'introduction de ce cas test (*cf.* § 6.4.2.1). Ici, on s'intéresse à un seul calcul où $p = 2$ et $t = 150$.

On observe que pour chacune des grandeurs, les valeurs du tableau indiquent une erreur très petite et similaire pour toutes les discrétisations, soit environ $10^{-2}\%$ pour C_D et $10^{-3}\%$ pour St .

5. Les couleurs de l'image ont été converties en échelle de gris.

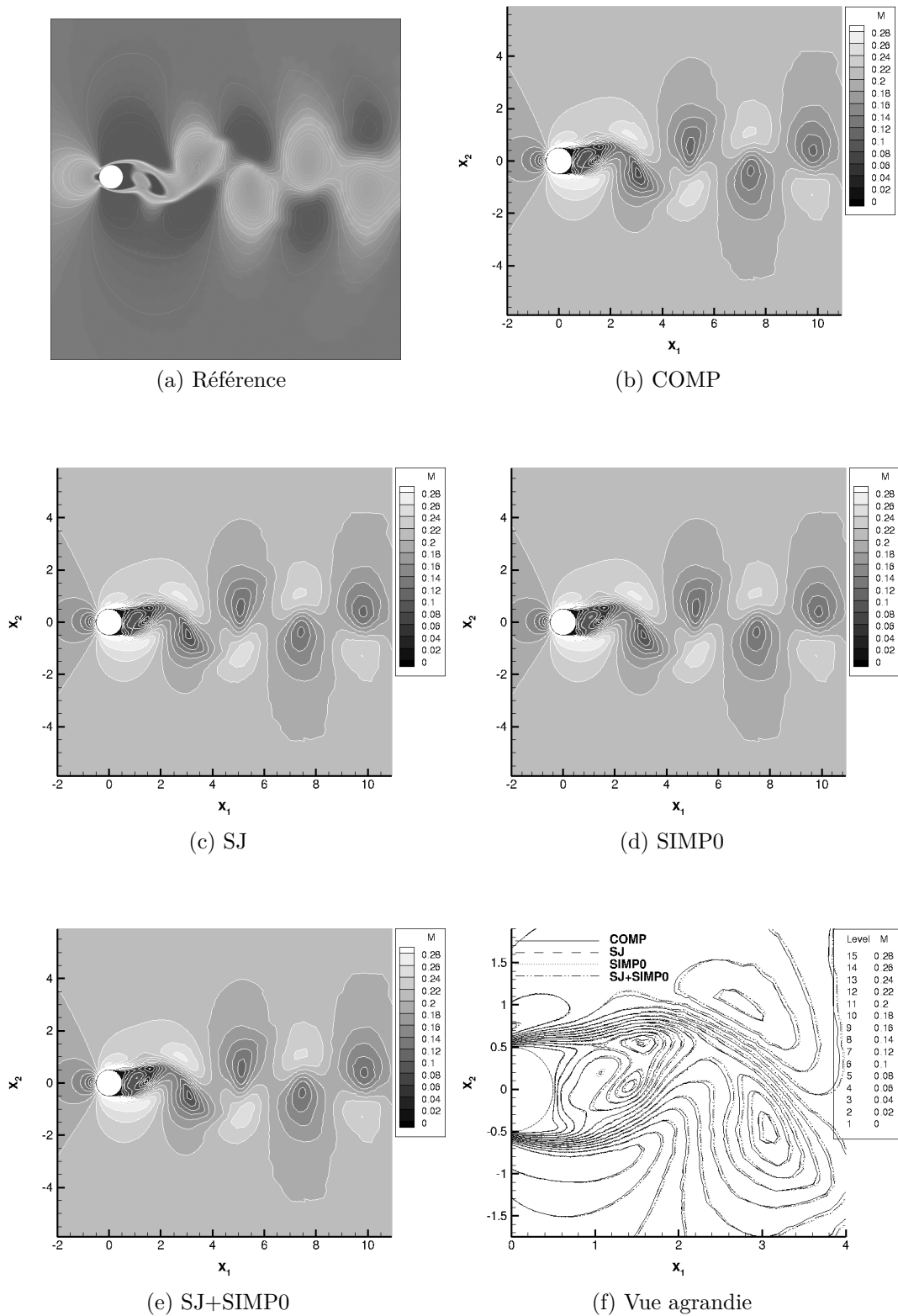


Fig. 6.14 – (*Cylindre*). Iso-valeurs du nombre de Mach local. Comparaison entre la simulation de référence à un instant t non précisé [115, figure 13] (a) et les simulations explicites-implicites telles que $p = 2$ à $t = 150$ (b), (c), (d) et (e). La figure (f) montre une vue agrandie.

| Expérience ou Simulation | | Détail | C_D | St |
|---------------------------|---------|-------------|---------------|-----------------|
| Roshko | [146] | | – | 0,182 |
| Tritton | [163] | | $\approx 1,3$ | – |
| Williamson | [179] | | – | $\approx 0,184$ |
| Ould Salihi <i>et al.</i> | [120] | PG | – | 0,183 |
| Nichols et Heikkinen | [115] | DF, Ordre 3 | 1,34 | [0,179; 0,182] |
| $N = 7\ 384$ | $p = 1$ | GD, COMP | 1,308 | 0,171 |
| $7\ 384$ | 2 | GD, COMP | 1,351 | 0,184 |
| $7\ 384$ | 3 | GD, COMP | 1,349 | 0,184 |

Tab. 6.9 – (*Cylindre*). Valeur du coefficient de traînée et du nombre de Strouhal. Comparaison entre les expériences ou simulations de référence (haut) et la simulation COMP pour $1 \leq p \leq 3$ (bas). Les valeurs données pour les expériences ont été lues graphiquement (*cf.* annexe E). Abréviations : particules-grilles (PG), différences finies (DF), Galerkin discontinue (GD).

| | SJ | SIMP0 | SJ+SIMP0 |
|------------------|-------------------|-------------------|-------------------|
| Erreur C_D (%) | 0,01 | 0,01 | 0,02 |
| Erreur St (%) | $\approx 10^{-3}$ | $\approx 10^{-3}$ | $\approx 10^{-3}$ |

Tab. 6.10 – (*Cylindre*). Erreur en pourcentage sur le coefficient de traînée et le nombre de Strouhal pour les méthodes SJ, SIMP0 et SJ+SIMP0. Calculs pour $p = 2$ et $t = 150$.

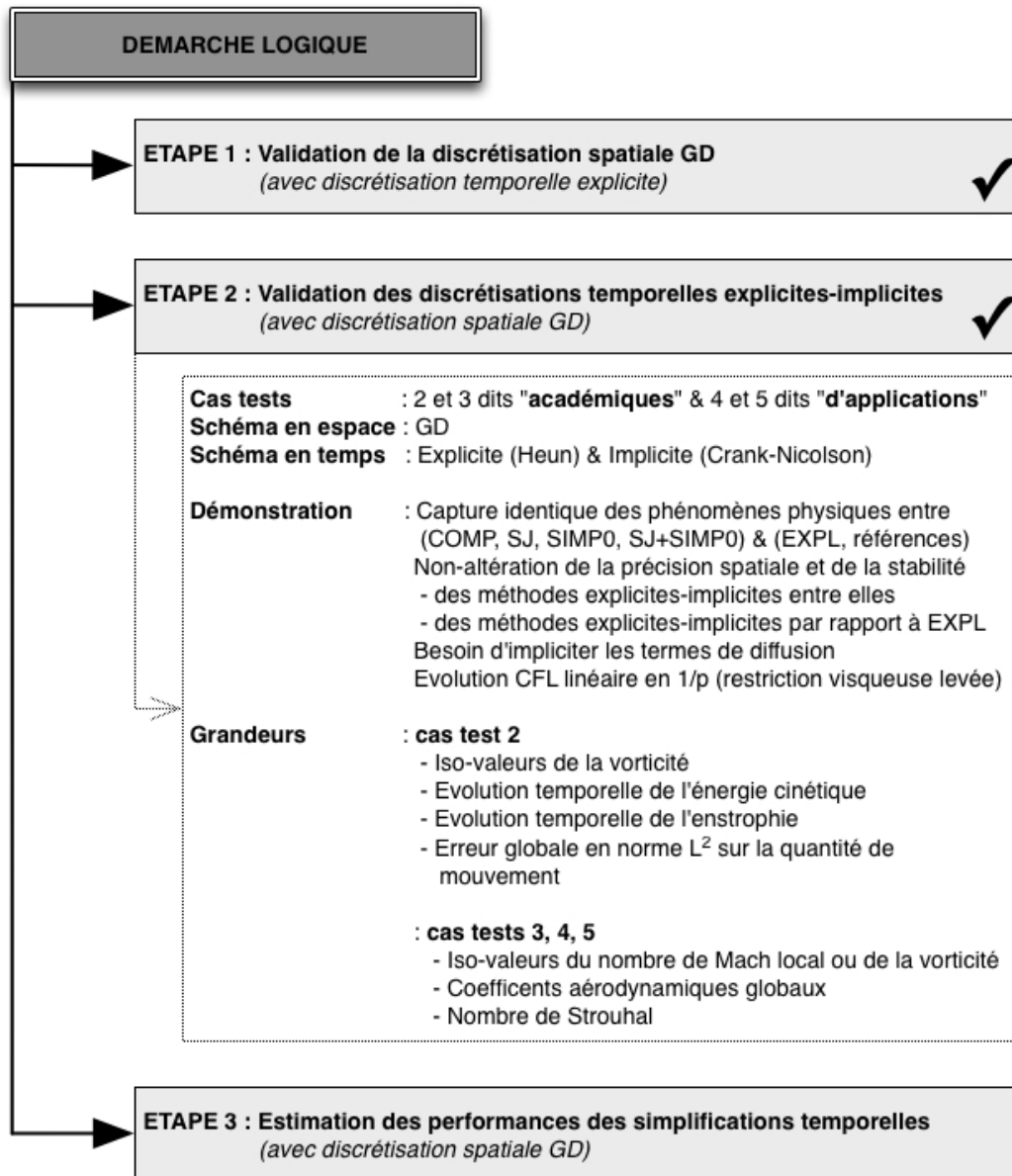


Fig. 6.15 – Schéma récapitulant l'étape 2 de validation du code de calcul développé.

6.5 Conclusion partielle

La figure 6.15 et les points suivants constituent le bilan de ce chapitre.

- La validation des quatre discrétisations temporelles explicites-implicites (COMP, SJ, SIMP0 et SJ+SIMP0) associées à la discrétisation spatiale GD a été établie au moyen des validations :
 - de la bonne capture des phénomènes physiques en comparaison à des références (cas tests 2, 3, 4 et 5), dont la création des petites structures tourbillonnaires à la paroi (cas test 2) ou l'apparition de la bulle de décollement et des instabilités de sillage (cas tests 3, 4 et 5);
 - des ordres de convergence en loi de puissance $p + 1$ pour un degré p de polynôme (cas test 2), autrement dit la précision spatiale n'est altérée par aucune des discrétisations temporelles explicites-implicites;
 - de l'exactitude des valeurs des coefficients aérodynamiques (cas tests 3, 4 et 5) ou du nombre de Strouhal (cas test 5);
 - sur un grand nombre de cas tests représentatifs des écoulements instationnaires à convection dominante, dont des cas d'applications réalistes où la couche limite est fortement décollée (cas tests 4 et 5).
- Certains points clés des méthodes explicites-implicites ont été mis en avant. A savoir :
 - l'évolution des valeurs de la condition de CFL suit la loi imposée par les termes de convection en $1/(2p + 1)$ démontrant que la restriction de stabilité imposée par les termes de diffusion est bien levée par la stratégie explicite-implicite par rapport à celle explicite;
 - la robustesse de l'ensemble des méthodes explicites-implicites simplifiées qui préservent les propriétés de stabilité de la méthode complète;
 - la nécessité d'utiliser une intégration temporelle implicite sur les termes de diffusion pour les écoulements où les effets visqueux sont prédominants aux parois (cas test 5).

A présent, il s'agit de mettre en œuvre la dernière étape de validation du code de calcul. C'est l'objet du prochain chapitre.

Chapitre 7

Evaluation des performances des discrétisations temporelles

Ce chapitre se concentre sur l'étape 3 d'évaluation des performances des méthodes explicites-implicites en temps, COMP, SJ, SIMP0 et SJ+SIMP0, précédemment validées au chapitre 6.

Les gains en temps CPU sur les étapes de calculs intermédiaires et sur la résolution globale sont alors étudiés pour les trois méthodes en temps simplifiées SJ, SIMP0 et SJ+SIMP0 en comparaison avec la méthode complète COMP. Cette étude porte sur une analyse paramétrique en fonction du degré du polynôme p , du nombre d'éléments N , du rapport d'aspect des maillages RA et de l'entier de couplage p_s entre les modes lors de l'étape de reconstruction. L'étude concerne les cas tests 2, 3, 4 et 5.

Ce chapitre est organisé comme suit. Le paragraphe 7.1 détaille certains prérequis sur la notion de mesure d'un temps d'exécution servant à l'évaluation des performances. Puis les deux paragraphes 7.2 et 7.3 donnent les premières tendances du comportement des quatre méthodes : COMP, SJ, SIMP0 et SJ+SIMP0, en fonction du temps CPU sur les étapes de calculs intermédiaires et sur la résolution globale. Enfin les deux paragraphes suivants 7.4 et 7.5 sont associés aux évaluations des gains en temps CPU respectivement sur les étapes de calculs intermédiaires et sur la résolution globale en fonction de p , N , RA et p_s pour les trois méthodes explicites-implicites simplifiées : SJ, SIMP0 et SJ+SIMP0.

7.1 Prérequis pour l'évaluation des performances

Ce paragraphe revient tout d'abord sur la notion de mesure d'un temps d'exécution comme le temps CPU (§ 7.1.1), puis il donne les caractéristiques du ordinateur utilisé pour réaliser les mesures (§ 7.1.2), et enfin il définit la quantité *Gain* servant à estimer les performances des discrétisations explicites-implicites simplifiées (§ 7.1.3).

7.1.1 Temps Elapsed et temps CPU

Il existe principalement deux types de mesures en temps : le temps Elapsed et le temps CPU. D'un côté le temps Elapsed est le temps réel d'exécution d'un programme pour l'utilisateur, autrement dit le temps de la montre. Il est fonction de la charge de calcul de la machine et donc il fluctue. Sa mesure se fait avec la commande « SYSTEM_CLOCK » en Fortran.

D'un autre côté, le temps CPU regroupe le temps de calcul qu'un processeur passe à l'exécution d'un programme. Dans ce travail, comme on utilise un processeur dédié, le temps ne dépend pas de la charge de travail de la machine et est donc une constante. Sa mesure se fait avec la commande « CPU_TIME » en Fortran.

Ainsi le temps Elapsed est composé de la somme du temps CPU et d'un temps système d'exécution (des entrées et des sorties par exemple). Dans ce travail, c'est le temps CPU qui est retenu comme mesure des performances des différentes méthodes.

7.1.2 Environnement informatique

Tous les calculs numériques de mesure du temps CPU sont réalisés sur le calculateur scalaire hautes performances « Baal ». C'est un calculateur destiné à des tâches de développement. Sa description est donnée dans le tableau 7.1 à titre indicatif pour toute personne voulant comparer ses mesures de calcul en temps CPU à celles communiquées dans ce travail.

En particulier, le cluster¹ Baal est issu de la gamme Altix ICE de SGI. Son système d'exploitation est Suse Linux Enterprise Server en version 10 SP2. Il est composé de quatre nœuds de calcul à base de processeurs Intel Nehalem et de deux nœuds de calcul à base de processeurs Intel Westmere [140]. Le nœud d'accueil *baal00* est dédié à la compilation et aux opérations de post-traitements, tandis que les autres sont destinés aux calculs de simulations.

Dans ce travail, on a retenu le nœud dédié *n002* pour l'ensemble des calculs en temps CPU.

| Nœud | Fonction | Mémoire | Processeur Intel | | | |
|--------------------|----------|---------|------------------|-----------|--------|-----------|
| | | | Gamme | Nb. Cœurs | Modèle | Fréquence |
| <i>baal00</i> | accueil | 24 Go | Nehalem | 4 | E5540 | 2,53 GHz |
| <i>n00</i> [1 – 4] | calcul | 24 Go | Nehalem | 4 | X5560 | 2,80 GHz |
| <i>n00</i> [5 – 6] | calcul | 48 Go | Westmere | 6 | E5650 | 2,67 GHz |

Tab. 7.1 – Composition du calculateur Baal [140].

1. Terme anglais signifiant « grappe de serveurs ».

7.1.3 Définition du *Gain* en temps CPU

L'évaluation des performances des trois discrétisations explicites-implicites simplifiées en temps repose sur la quantité *Gain* donnée par la définition 7.1 suivante.

Définition 7.1. (*Gain en temps CPU*) *Le Gain en temps CPU est défini comme le rapport du temps CPU nécessaire pour atteindre un nombre fixé de $n = 500$ itérations entre la simulation COMP servant de référence et une autre simulation AUTRE=SJ, SIMP0 ou SJ+SIMP0, soit :*

$$\text{Gain AUTRE} = \frac{\text{temps CPU (COMP)}}{\text{temps CPU (AUTRE)}}$$

Pour la compréhension de la suite, les deux points suivants sont mentionnés.

- Dans le cas des écoulements externes, on repart d'un calcul où les gradients de la solution sont les plus forts, et ce pour tester la robustesse des différentes stratégies d'intégration. Ainsi quel que soit p , le calcul repart des instants $t = 60$ pour le cas test 3, $t = 25$ pour le cas test 4 et $t = 150$ pour le cas test 5.
- Les valeurs de la condition CFL utilisées ont été données au paragraphe 6.2.1 page 134.

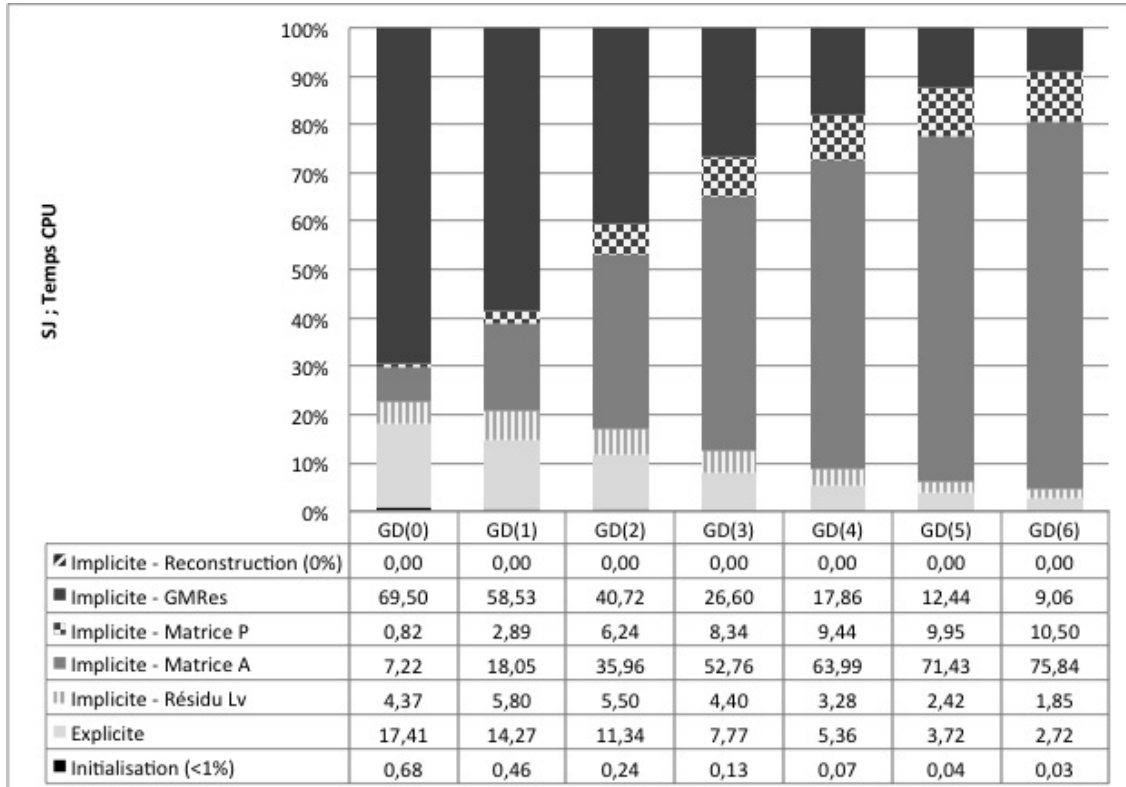
7.2 Estimation des temps CPU intermédiaires

Ce paragraphe reprend l'idée du paragraphe 3.4.3 page 73 qui était d'observer les coûts en temps CPU sur les étapes de calculs intermédiaires avec l'exemple numérique du cas test 2. Ces coûts sont ici considérés pour les trois méthodes simplifiées SJ, SIMP0 et SJ+SIMP0.

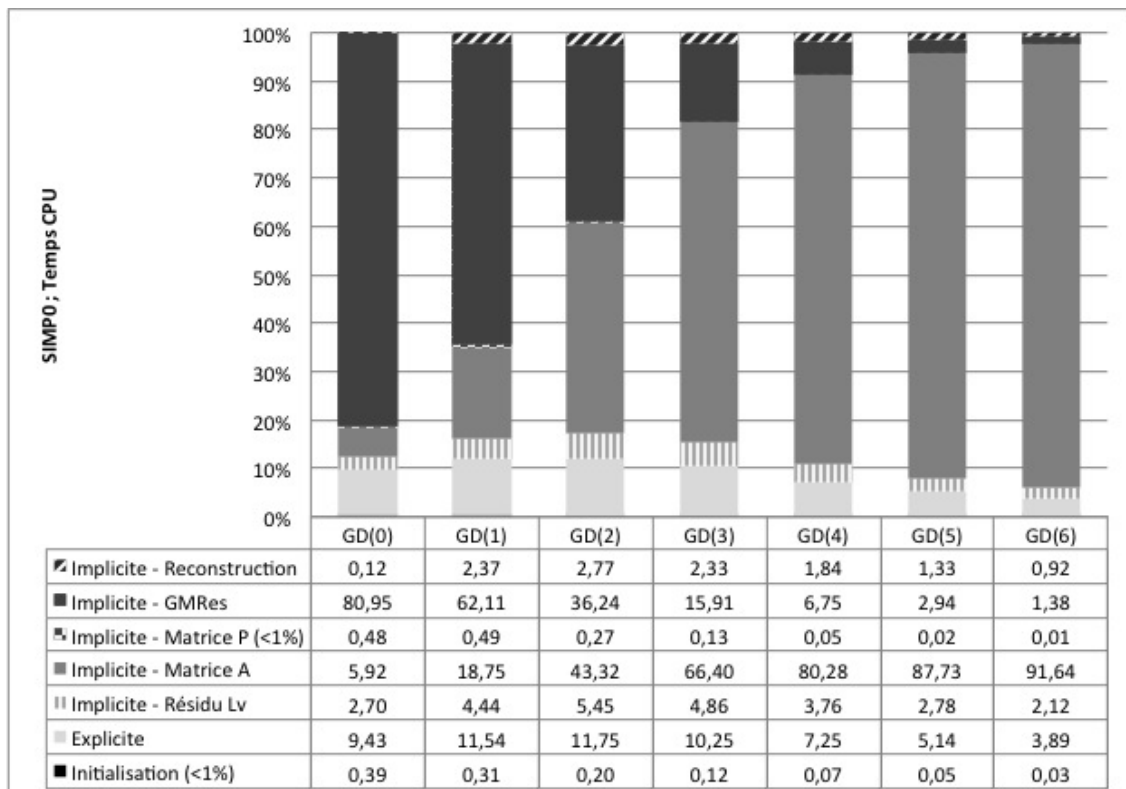
La figure 7.1 indique la répartition de ces coûts moyens en pourcentage sur le maillage uniforme tel que $N = 2\,888$ et pour $0 \leq p \leq 6$. Ces coûts étant en pourcentage, ils ne permettent pas une comparaison des méthodes entre elles. Toutefois, on relève les points suivants avec les figures 7.1 et 3.2 page 72.

- Premièrement, comme pour la méthode COMP, les deux coûts des étapes de construction de la matrice implicite et de l'inversion GMRes sont les plus significatifs du temps de résolution total pour l'ensemble des trois méthodes simplifiées, avec un poids qui se compense lorsque p augmente.
- Deuxièmement, le coût de construction de la matrice de préconditionnement, qui augmentait avec p pour la méthode COMP, devient insignifiant pour les deux méthodes SIMP0 et SJ+SIMP0 quel que soit p . Ces deux dernières sont bien moins sensibles au choix du préconditionneur. Toutefois, pour la méthode SJ, le coût devient plus conséquent que l'inversion GMRes lorsque $p \geq 6$.
- Troisièmement, le coût de la reconstruction est plus important pour la méthode SJ+SIMP0 que pour la méthode SIMP0, à cause de la construction du troisième résidu lors du calcul du produit matrice-vecteur (3.27).

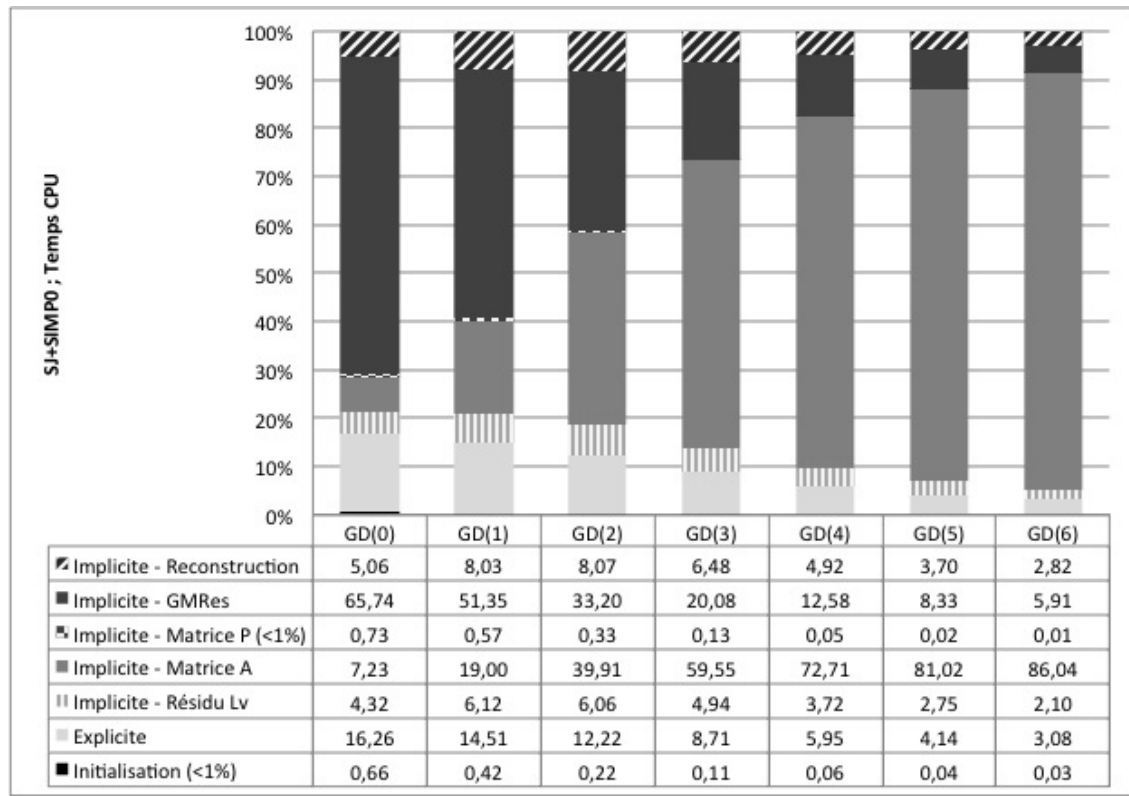
Pour compléter cette analyse des coûts, le paragraphe suivant s'intéresse à l'estimation du temps CPU sur la résolution complète.



(a) méthode SJ



(b) méthode SIMP0



(c) méthode SJ+SIMP0

Fig. 7.1 – Illustration des contributions relatives (en pourcentage) aux coûts en temps CPU pour les étapes de la résolution explicite-implicite des schémas simplifiés. Calculs pour le cas test 2, sur le maillage uniforme $N = 2\,888$ et pour $0 \leq p \leq 6$. La table de correspondance étapes-couleurs est indiquée dans la marge gauche du tableau.

7.3 Estimation du temps CPU total

Dans ce paragraphe, on regarde graphiquement le temps CPU nécessaire à la résolution numérique du problème discret pour les méthodes COMP, SJ, SIMP0 et SJ+SIMP0. L'objectif est d'observer les comportements des méthodes explicites-implicites simplifiées par rapport à la méthode complète.

Ces comparaisons sont illustrées avec les quatre groupes de figures : 7.2, 7.3a & 7.3b, 7.3c & 7.3d et 7.3e & 7.3f qui correspondent aux quatre cas tests 2, 3, 4 et 5 considérés. Sur chacune de ces figures on trace l'évolution de certaines quantités globales caractéristiques des écoulements simulés, d'abord en fonction du temps physique t , puis en fonction de l'équivalent en temps CPU, et ce pour les quatre stratégies temporelles explicites-implicites. Pour plus de clarté, le tableau 7.2 reprend synthétiquement les détails des calculs et indique les gains en temps CPU.

- La première comparaison concerne les courbes en fonction du temps physique. Pour chaque cas test, l'ensemble des courbes représentatif de l'ensemble des méthodes est complètement superposé. Cette observation est en accord avec les précédentes conclusions sur les validations spatiales et temporelles.

| Figure | | Détails du calcul | | | | Gains en temps CPU | | |
|-------------|----------|-------------------|-------|-----|-----|--------------------|-------|----------|
| n^o | Cas test | N | RA | p | t | SJ | SIMP0 | SJ+SIMP0 |
| 7.2 | 2 | 2 888 | 1 | 4 | 0,8 | 1,53 | 1,70 | 1,69 |
| 7.3a & 7.3b | 3 | 4 298 | 2 125 | 3 | 10 | 1,57 | 1,79 | 1,98 |
| 7.3c & 7.3d | 4 | 11 676 | 5 191 | 2 | 4 | 2,03 | 1,40 | 2,69 |
| 7.3e & 7.3f | 5 | 7 384 | 630 | 2 | 80 | 1,91 | 1,55 | 2,62 |

Tab. 7.2 – Récapitulation des calculs correspondant aux figures 7.2 et 7.3, et évaluation des gains en temps CPU avec la méthode COMP comme référence.

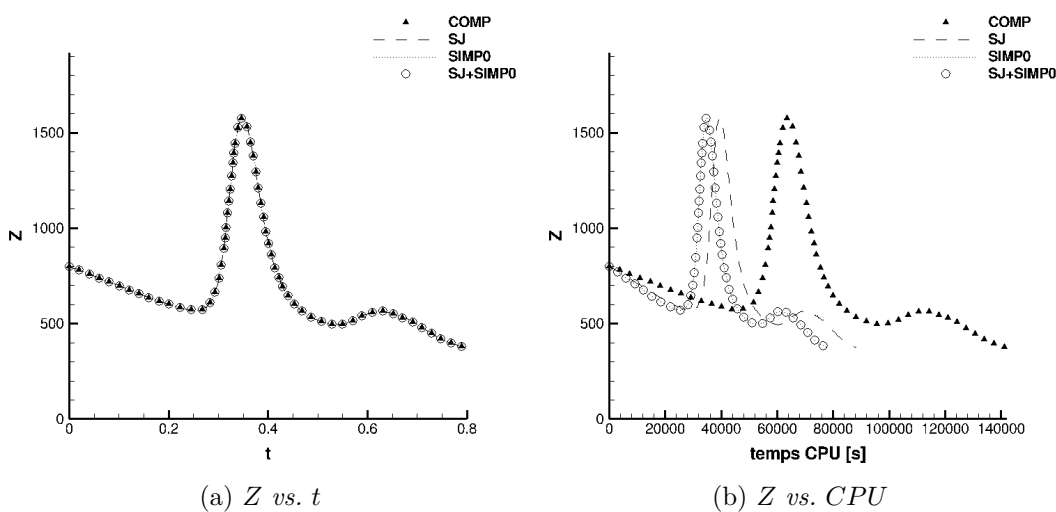


Fig. 7.2 – (*Interaction tourbillons-paroi*). Evolution de l'enstrophie en fonction du temps physique (a) ou du temps CPU correspondant (b). Comparaison entre les différentes discrétisations temporelles explicites-implicites sur le maillage uniforme ($RA = 1$) tel que $N = 2\,888$, pour $p = 4$ et $t = 0,8$.

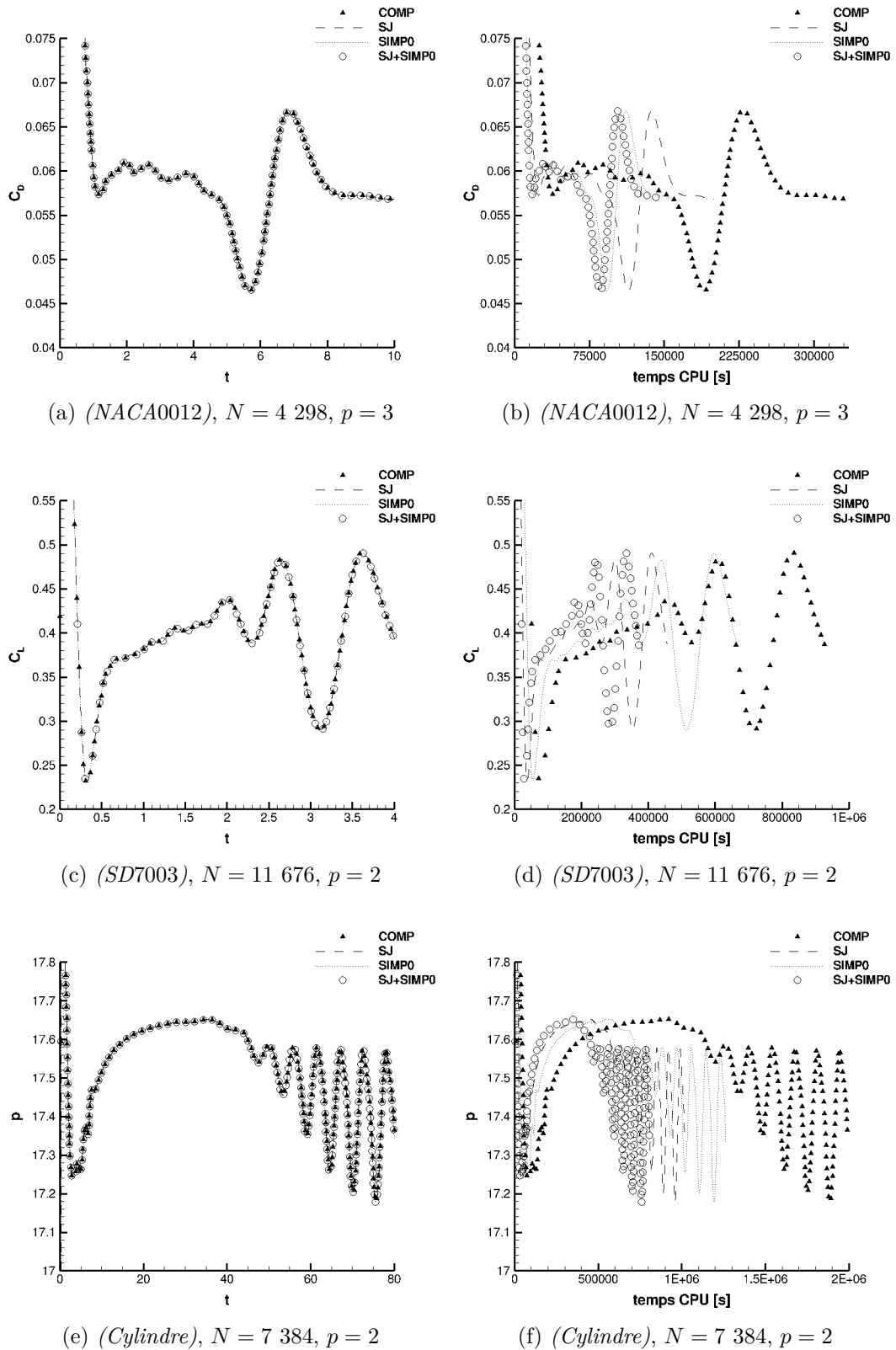


Fig. 7.3 – Evolution pour différents cas tests du coefficient de traînée (haut), du coefficient de portance (milieu) ou de la pression en x_{St} (bas) en fonction du temps physique (gauche) ou du temps CPU équivalent (droite). Comparaison entre les différentes discrétisations temporelles explicites-implicites.

| Détail des maillages du cas test 2 | | | | | | | |
|------------------------------------|------|-------|--------|-------|-------|-------|---------|
| N | 648 | 2 888 | 10 952 | 2 888 | 2 888 | 2 888 | |
| RA | 1 | 1 | 1 | 20 | 143 | 600 | |
| <i>Gain SJ</i> | | | | | | | Théorie |
| $p = 1$ | 1,65 | 1,55 | 1,64 | 1,59 | 1,64 | 1,55 | 1,77 |
| 2 | 1,73 | 1,68 | 1,74 | 1,69 | 1,74 | 1,66 | 1,77 |
| 3 | 1,75 | 1,76 | 1,77 | 1,71 | 1,77 | 1,77 | 1,77 |
| 4 | 1,76 | 1,72 | 1,79 | 1,72 | 1,78 | 1,79 | 1,77 |
| 5 | 1,77 | 1,71 | 1,80 | 1,73 | 1,79 | 1,79 | 1,77 |
| 6 | 1,76 | 1,77 | 1,81 | 1,74 | 1,79 | 1,80 | 1,77 |
| <i>Gain SIMP0</i> | | | | | | | Théorie |
| $p = 1$ | 1,17 | 1,19 | 1,21 | 1,13 | 1,19 | 1,12 | 1,41 |
| 2 | 1,40 | 1,42 | 1,44 | 1,35 | 1,43 | 1,35 | 1,57 |
| 3 | 1,54 | 1,54 | 1,56 | 1,47 | 1,55 | 1,54 | 1,64 |
| 4 | 1,60 | 1,62 | 1,65 | 1,54 | 1,62 | 1,61 | 1,68 |
| 5 | 1,65 | 1,66 | 1,71 | 1,59 | 1,66 | 1,66 | 1,71 |
| 6 | 1,66 | 1,73 | 1,76 | 1,62 | 1,74 | 1,69 | 1,72 |
| <i>Gain SJ + SIMP0</i> | | | | | | | Théorie |
| $p = 1$ | 1,95 | 1,58 | 1,64 | 1,59 | 1,65 | 1,67 | 1,77 |
| 2 | 1,81 | 1,67 | 1,74 | 1,69 | 1,74 | 1,76 | 1,77 |
| 3 | 1,84 | 1,70 | 1,77 | 1,72 | 1,77 | 1,77 | 1,77 |
| 4 | 1,84 | 1,71 | 1,79 | 1,73 | 1,78 | 1,78 | 1,77 |
| 5 | 1,84 | 1,71 | 1,80 | 1,73 | 1,78 | 1,78 | 1,77 |
| 6 | 1,82 | 1,76 | 1,81 | 1,75 | 1,80 | 1,79 | 1,77 |

Tab. 7.3 – (*Matrice implicite*). Gain en temps CPU, mesuré et théorique, des méthodes SJ, SIMP0 et SJ+SIMPO. Calculs sur le cas test 2, pour six maillages représentatifs où N et RA varient et pour $1 \leq p \leq 6$. Le gain théorique est défini au tableau 3.3 page 88 et est évalué sur les trois plus grandes valeurs de p .

- La seconde comparaison concerne les courbes en fonction du temps CPU. Pour chacun des cas tests, les calculs COMP sont les plus lents, ce qui conforte le réel intérêt des stratégies explicites-implicites développées. Concernant les stratégies simplifiées, si la méthode SJ+SIMP0 est globalement la plus rapide avec un gain atteignant 2,69, on note que les méthodes SJ et SIMP0 sont également compétitives.

Afin de mieux dégager les tendances en fonction des paramètres d'un calcul comme p , N ou RA , les deux paragraphes suivants proposent une étude plus approfondie des gains sur les étapes de calculs intermédiaires et sur la résolution totale.

7.4 Estimation des gains intermédiaires

Pour les trois méthodes simplifiées SJ, SIMP0 et SJ+SIMP0, ce paragraphe compare les gains théoriques établis au tableau 3.3 page 88 et les gains réellement obtenus, et ce pour certaines étapes principales de calculs intermédiaires. L'objectif est d'apprécier le respect des différentes tendances et d'évaluer les constantes $(\alpha_i)_{1 \leq i \leq 3}$. On souligne que l'évaluation exacte de ces dernières est difficile puisque les gains théoriques concernent les opérations les plus significatives et les gains mesurés concernent toutes les opérations dans leur intégralité.

L'analyse porte sur le cas test 2 offrant une large gamme de maillages où N et RA varient, et pour $1 \leq p \leq 6$. Elle concerne les étapes de calculs de construction : de la matrice implicite (§ 7.4.1) et de celle de préconditionnement (§ 7.4.2), de l'inversion GMRes (§ 7.4.3) et de reconstruction des modes d'ordre élevé (§ 7.4.4).

7.4.1 Gain sur la matrice implicite

Le tableau 7.3 donne les gains en temps CPU mesurés et théoriques pour la construction de la matrice \mathbf{A} . Son analyse permet une estimation du rapport α_3/α_2 .

Evaluations de *Gain SJ* et de *Gain SIMP0 + SJ*.

- Pour chaque maillage, plus p est grand plus les deux gains tendent vers une constante comme attendu. En effet, la dépendance en p est très faible.
- Pour chaque degré, et quel que soit le maillage, les deux gains restent du même ordre de grandeur, en particulier si p est grand. Ainsi, les deux gains sont plutôt indépendants du maillage comme attendu.
- L'estimation asymptotique retenue de la constante α_3/α_2 est 0,77.

Evaluation de *Gain SIMP0*.

- Pour chaque maillage, plus p augmente plus *Gain SIMP0* croît. Ainsi, le gain dépend faiblement de p comme attendu.
- Pour chaque degré, et quel que soit le maillage, *Gain SIMP0* reste du même ordre de grandeur. Ainsi, le gain est indépendant du maillage comme attendu.

Le classement des méthodes par rapidité décroissante est alors : SJ+SIMP0 et SJ à égalité grâce à la construction des seuls blocs diagonaux de \mathbf{A} , puis SIMP0.

| Détail des maillages du cas test 2 | | | | | | | |
|------------------------------------|------|-------|--------|-------|-------|-------|---------|
| N | 648 | 2 888 | 10 952 | 2 888 | 2 888 | 2 888 | |
| RA | 1 | 1 | 1 | 20 | 143 | 600 | |
| <i>Gain SJ</i> | | | | | | | Théorie |
| $p = 1$ | 1,05 | 0,99 | 1,00 | 0,99 | 1,05 | 0,97 | 1,00 |
| 2 | 1,01 | 0,97 | 1,00 | 0,98 | 1,00 | 0,96 | 1,00 |
| 3 | 1,00 | 1,00 | 1,00 | 0,97 | 1,00 | 1,00 | 1,00 |
| 4 | 1,00 | 0,97 | 1,00 | 0,97 | 1,00 | 1,00 | 1,00 |
| 5 | 1,00 | 0,96 | 1,00 | 0,98 | 1,00 | 1,00 | 1,00 |
| 6 | 0,99 | 0,99 | 1,00 | 0,98 | 1,00 | 1,00 | 1,00 |
| <i>Gain SIMP0</i> | | | | | | | Théorie |
| $p = 1$ | 4 | 5 | 4 | 4 | 5 | 4 | 27 |
| 2 | 24 | 23 | 18 | 19 | 22 | 19 | 216 |
| 3 | 63 | 69 | 59 | 61 | 79 | 62 | 1000 |
| 4 | 176 | 210 | 163 | 180 | 190 | 191 | 3375 |
| 5 | 449 | 467 | 388 | 432 | 473 | 362 | 9261 |
| 6 | 1079 | 1068 | 857 | 969 | 1040 | 944 | 21952 |
| <i>Gain SJ + SIMP0</i> | | | | | | | Théorie |
| $p = 1$ | 7 | 5 | 4 | 4 | 6 | 12 | 27 |
| 2 | 48 | 21 | 18 | 21 | 26 | 24 | 216 |
| 3 | 153 | 69 | 60 | 73 | 85 | 63 | 1000 |
| 4 | 285 | 189 | 162 | 178 | 193 | 150 | 3375 |
| 5 | 688 | 439 | 385 | 424 | 471 | 369 | 9261 |
| 6 | 1450 | 1091 | 830 | 1021 | 1015 | 898 | 21952 |

Tab. 7.4 – (*Matrice de préconditionnement*). Gain en temps CPU, mesuré et théorique, des méthodes SJ, SIMP0 et SJ+SIMP0. Calculs sur le cas test 2, pour six maillages représentatifs où N et RA varient et pour $1 \leq p \leq 6$. Le gain théorique est défini au tableau 3.3 page 88.

7.4.2 Gain sur la matrice de préconditionnement

Le tableau 7.4 donne les gains en temps CPU mesurés et théoriques pour la construction de la matrice \mathbf{P} .

Evaluation de *Gain SJ*.

- Pour chaque maillage et pour chaque degré p , *Gain SJ* est globalement égal à la constante 1,00 comme attendu.

Evaluations de *Gain SIMP0* et de *Gain SJ + SIMP0*.

- Pour chaque maillage, plus p augmente plus les deux gains croissent. Ainsi, les gains dépendent très fortement de p . Cette dépendance est très éloignée de celle attendue en N_p^3 pour $p_s = 0$ fixé. Toutefois, on rappelle que le coût asymptotique en $2N^3/3$ pour la décomposition LU d'une matrice de taille N est valable pour N tendant vers l'infini, et sans changement de pivot partiel.
- Pour chaque degré, les deux gains dépendent du maillage. Cependant, si les gains semblent globalement plus faibles lorsque N est grand et quel que soit p , aucune tendance ne ressort en fonction de RA . Cette dépendance en N est forte et non prédite en théorie. Elle est directement liée à la taille réduite du système linéaire (3.23) à résoudre en $NN_{eq}N_{ps}$ équations. Par ailleurs, elle est sensiblement plus forte pour la méthode SJ+SIMP0.

Le classement des méthodes par rapidité décroissante est : SJ+SIMP0, SIMP0 et SJ. Ainsi, les méthodes SIMP0 et SJ+SIMP0 sont bien les moins sensibles au préconditionneur de Jacobi par blocs en comparaison à la méthode SJ.

7.4.3 Gain sur l'inversion GMRes

Le tableau 7.5 donne les gains en temps CPU mesurés et théoriques pour l'inversion GMRes.

Evaluations de *Gain SJ* et de *Gain SIMP0 + SJ*.

- Pour la majorité des maillages, *Gain SJ* est non monotone avec p , tandis que *Gain SJ + SIMP0* croît avec p . Ainsi, les deux gains dépendent bien faiblement de p .
- Pour chaque degré, les deux gains dépendent du maillage. Par ailleurs, si le gain semble globalement plus fort lorsque N est grand, et quel que soit p pour la méthode SJ ou pour $p \leq 3$ pour la méthode SJ+SIMP0, aucune tendance ne ressort en fonction de RA . Cette dépendance en N est faible et non prédite en théorie. Elle est directement liée à l'approximation du produit \mathbf{Az} par une différence finie d'ordre un, qui requiert la construction d'un second résidu discret. Par ailleurs, elle est plus forte pour la méthode SJ+SIMP0 qui bénéficie en plus d'une réduction de la taille du système linéaire (3.23) à résoudre.

| Détail des maillages du cas test 2 | | | | | | | |
|------------------------------------|-------|-------|--------|-------|-------|-------|---------|
| N | 648 | 2 888 | 10 952 | 2 888 | 2 888 | 2 888 | |
| RA | 1 | 1 | 1 | 20 | 143 | 600 | |
| <i>Gain SJ</i> | | | | | | | Théorie |
| $p = 1$ | 0,90 | 1,70 | 2,54 | 1,72 | 1,79 | 1,63 | 1,15 |
| 2 | 1,16 | 1,61 | 2,35 | 1,62 | 1,67 | 1,53 | 1,43 |
| 3 | 1,37 | 1,65 | 2,20 | 1,62 | 1,63 | 1,63 | 1,67 |
| 4 | 1,63 | 1,74 | 2,20 | 1,77 | 1,82 | 1,82 | 1,87 |
| 5 | 1,83 | 1,85 | 2,34 | 1,91 | 1,96 | 2,66 | 2,06 |
| 6 | 2,14 | 2,17 | 2,46 | 2,16 | 2,23 | 2,24 | 2,22 |
| <i>Gain SIMP0</i> | | | | | | | Théorie |
| $p = 1$ | 2,02 | 1,14 | 1,05 | 1,12 | 1,15 | 1,11 | 9,00 |
| 2 | 5,07 | 1,62 | 1,17 | 1,59 | 1,63 | 1,66 | 36,00 |
| 3 | 11,10 | 2,59 | 1,89 | 2,55 | 2,99 | 3,40 | 100,00 |
| 4 | 25,41 | 4,70 | 2,66 | 5,47 | 7,13 | 7,07 | 225,00 |
| 5 | 49,07 | 8,22 | 3,08 | 11,99 | 12,45 | 12,73 | 441,00 |
| 6 | 87,53 | 15,08 | 6,30 | 21,76 | 20,65 | 22,56 | 784,00 |
| <i>Gain SJ + SIMP0</i> | | | | | | | Théorie |
| $p = 1$ | 1,48 | 1,87 | 2,61 | 1,91 | 1,96 | 1,95 | 1,45 |
| 2 | 1,67 | 1,91 | 2,55 | 1,94 | 2,00 | 2,01 | 1,98 |
| 3 | 2,03 | 2,04 | 3,38 | 2,09 | 2,40 | 2,53 | 2,48 |
| 4 | 2,46 | 2,37 | 3,65 | 2,71 | 3,61 | 3,34 | 2,99 |
| 5 | 2,84 | 2,65 | 3,09 | 4,06 | 4,12 | 3,84 | 3,49 |
| 6 | 3,33 | 3,16 | 4,54 | 4,21 | 4,37 | 4,42 | 3,99 |

Tab. 7.5 – (*Inversion GMRes*). Gain en temps CPU, mesuré et théorique, des méthodes SJ, SIMP0 et SJ+SIMPO. Calculs sur le cas test 2, pour six maillages représentatifs où N et RA varient et pour $1 \leq p \leq 6$. Le gain théorique est défini au tableau 3.3 page 88 et est évalué sur les trois plus grandes valeurs de p , pour $n_f = 3$ et $N_g = p + 1$.

- L'estimation asymptotique retenue de la constante α_1 est 40,06. Elle est à relativiser avec la dépendance en maillage observée d'un côté, et avec les valeurs de n_f et N_g choisies dans les relations théoriques d'un autre côté.
- On remarque que les valeurs moyennes de *Gain SJ* augmentent comme le rapport $N_p/(N_g+0, 20N_p)$, et celles de *Gain SJ+SIMP0* augmentent comme le rapport N_p/N_g .

Evaluation de *Gain SIMP0*.

- Pour chaque maillage, plus p augmente plus *Gain SIMP0* croît. Ainsi, le gain dépend de p . Cette dépendance est très éloignée de celle attendue en N_p^2 pour $p_s = 0$ fixé. Cet écart à la théorie peut s'expliquer comme suit. Premièrement, le coût réel de l'inversion dépend à la fois de la précision requise sur le résidu pour stopper la convergence et de la fréquence s des redémarrages de la méthode GMRes(s) (cf. remarque 3.5 page 71 pour un rappel de cette méthode). Deuxièmement, le temps CPU mesuré ici, considère l'inversion dans sa globalité et pas uniquement les deux multiplications matrice-vecteur \mathbf{Az} et $\mathbf{P}^{-1}\mathbf{z}$ pour un vecteur \mathbf{z} quelconque.
- Pour chaque degré, *Gain SIMP0* dépend du maillage. Cependant, si le gain semble globalement plus faible lorsque N est grand et quel que soit p , aucune tendance ne ressort en fonction de RA . Cette dépendance en N est non prédite en théorie. Elle est directement liée à la taille réduite du système linéaire (3.23) à résoudre.

Le classement des méthodes par rapidité décroissante est alors : SIMP0, SJ+SIMP0 et SJ.

7.4.4 Gain sur la reconstruction des modes d'ordre élevé

La reconstruction des modes d'ordre élevé est une étape de calculs non déterminante pour les méthodes COMP, SJ, et peu déterminante pour les méthodes SIMP0 et SJ+SIMP0 (cf. § 7.2).

Une analyse succincte révèle que la méthode SIMP0 est plus rapide que la méthode SJ+SIMP0. Pour ces deux méthodes, les performances dépendent faiblement de p et du maillage. La dépendance en maillage étant due au produit matrice-vecteur $\hat{\mathbf{U}}\Delta\hat{\mathbf{W}}^{(n)}$ et à la résolution du système triangulaire $(\hat{\mathcal{L}} + \hat{\mathcal{D}})^{-1}\Delta\hat{\mathbf{W}}^{(n+1)}$.

Finalement, conformément aux relations théoriques, la dépendance en p est globalement toujours bien observée. De plus, une dépendance en maillage est également observée. Elle est principalement due à la réduction du système linéaire implicite de Crank-Nicolson à résoudre lors de l'approche SIMP p_s , ou à la construction des résidus supplémentaires lors de l'approche SJ.

Pour compléter l'évaluation des performances de ce chapitre, la dernière analyse porte sur les gains de la résolution totale.

| $N = 648, RA = 1$ | | | $N = 2\,888, RA = 1$ | | | $N = 10\,952, RA = 1$ | | | |
|-------------------|-------------|-------------|----------------------|-------------|-------------|-----------------------|-------------|-------------|-------------|
| p | SJ | SIMP0 | SJ+SIMP0 | SJ | SIMP0 | SJ+SIMP0 | SJ | SIMP0 | SJ+SIMP0 |
| 1 | <i>1,09</i> | 1,29 | 1,18 | 1,43 | <i>1,11</i> | 1,50 | 2,09 | <i>1,03</i> | 2,15 |
| 2 | <i>1,29</i> | 1,55 | 1,44 | 1,38 | <i>1,37</i> | 1,50 | 1,78 | <i>1,19</i> | 1,89 |
| 3 | <i>1,41</i> | 1,65 | 1,54 | <i>1,45</i> | 1,58 | 1,60 | 1,62 | <i>1,44</i> | 1,79 |
| 4 | <i>1,50</i> | 1,72 | 1,67 | <i>1,53</i> | 1,70 | 1,69 | <i>1,62</i> | 1,73 | 1,90 |
| 5 | <i>1,55</i> | 1,76 | 1,72 | <i>1,56</i> | 1,76 | 1,73 | <i>1,61</i> | 1,74 | 1,94 |
| 6 | <i>1,60</i> | 1,77 | 1,76 | <i>1,64</i> | 1,84 | 1,81 | <i>1,64</i> | 1,90 | 1,86 |

| $N = 2\,888, RA = 4$ | | | $N = 43\,808, RA = 1$ | | | |
|----------------------|-------------|-------------|-----------------------|------|-------------|-------------|
| p | SJ | SIMP0 | SJ+SIMP0 | SJ | SIMP0 | SJ+SIMP0 |
| 1 | 1,45 | <i>1,10</i> | 1,49 | 2,63 | <i>1,00</i> | 2,68 |
| 2 | 1,38 | <i>1,37</i> | 1,50 | 2,44 | <i>1,36</i> | 3,25 |
| 3 | <i>1,46</i> | 1,58 | 1,61 | 2,15 | <i>1,46</i> | 2,79 |
| 4 | <i>1,53</i> | 1,70 | 1,69 | 1,95 | <i>1,61</i> | 2,42 |
| 5 | <i>1,57</i> | 1,76 | 1,74 | | | |
| 6 | <i>1,62</i> | 1,84 | 1,81 | | | |

| $N = 2\,888, RA = 20$ | | | $N = 2\,888, RA = 143$ | | | $N = 2\,888, RA = 600$ | | | |
|-----------------------|-------------|-------------|------------------------|-------------|-------------|------------------------|-------------|-------------|-------------|
| p | SJ | SIMP0 | SJ+SIMP0 | SJ | SIMP0 | SJ+SIMP0 | SJ | SIMP0 | SJ+SIMP0 |
| 1 | 1,44 | <i>1,11</i> | 1,52 | 1,39 | <i>1,11</i> | 1,46 | 1,45 | <i>1,10</i> | 1,51 |
| 2 | 1,38 | <i>1,37</i> | 1,50 | 1,37 | <i>1,35</i> | 1,47 | 1,40 | <i>1,31</i> | 1,45 |
| 3 | <i>1,46</i> | 1,59 | 1,61 | <i>1,46</i> | 2,01 | 1,67 | <i>1,45</i> | 1,67 | 1,67 |
| 4 | <i>1,53</i> | 1,70 | 1,69 | <i>1,53</i> | 1,70 | 1,69 | <i>1,52</i> | 1,80 | 1,79 |
| 5 | <i>1,58</i> | 1,75 | 1,76 | <i>1,58</i> | 1,73 | 1,72 | <i>1,62</i> | 1,83 | 1,82 |
| 6 | <i>1,62</i> | 1,79 | 1,80 | <i>1,63</i> | 1,87 | 1,79 | <i>1,60</i> | 1,86 | 1,84 |

(a) Interaction tourbillons-paroi

| $N = 1\,683, RA = 444$ | | | $N = 4\,298, RA = 2\,125$ | | | $N = 7\,368, RA = 1\,685$ | | | |
|------------------------|-------------|-------------|---------------------------|-------------|-------------|---------------------------|-------------|-------------|-------------|
| p | SJ | SIMP0 | SJ+SIMP0 | SJ | SIMP0 | SJ+SIMP0 | SJ | SIMP0 | SJ+SIMP0 |
| 1 | 1,23 | <i>1,17</i> | 1,32 | 1,77 | <i>1,25</i> | 2,11 | 2,06 | <i>1,23</i> | 2,46 |
| 2 | <i>1,28</i> | 1,58 | 1,53 | <i>1,55</i> | 1,61 | 2,02 | 1,77 | <i>1,54</i> | 2,34 |
| 3 | <i>1,39</i> | 1,74 | 1,64 | <i>1,57</i> | 1,79 | 1,98 | <i>1,67</i> | 1,69 | 2,15 |
| 4 | <i>1,50</i> | 1,80 | 1,73 | <i>1,58</i> | 1,94 | 1,93 | <i>1,58</i> | 1,84 | 1,93 |
| 5 | <i>1,55</i> | 1,79 | 1,73 | <i>1,56</i> | 1,90 | 1,86 | <i>1,64</i> | 2,03 | 1,71 |
| 6 | <i>1,58</i> | 1,81 | 1,72 | <i>1,60</i> | 1,89 | 1,88 | <i>1,76</i> | 2,13 | 1,87 |

(b) Profil d'aile NACA0012

| $N = 11\,676, RA = 5\,191$ | | | |
|----------------------------|-------------|-------------|-------------|
| p | SJ | SIMP0 | SJ+SIMP0 |
| 1 | 2,33 | <i>1,27</i> | 3,06 |
| 2 | 2,03 | <i>1,40</i> | 2,69 |
| 3 | 1,82 | <i>1,75</i> | 2,54 |
| 4 | <i>1,72</i> | 1,90 | 2,34 |
| 5 | <i>1,74</i> | 2,02 | 2,13 |
| 6 | <i>1,79</i> | 2,11 | 2,14 |

(c) Profil d'aile SD7003

| $N = 7\,384, RA = 630$ | | | |
|------------------------|-------------|-------------|-------------|
| p | SJ | SIMP0 | SJ+SIMP0 |
| 1 | 2,16 | <i>1,05</i> | 2,47 |
| 2 | 1,91 | <i>1,55</i> | 2,62 |
| 3 | <i>1,73</i> | 1,78 | 2,34 |

(d) Cylindre

Tab. 7.6 – Gain en temps CPU sur la résolution totale. Pour chaque maillage et chaque degré p , les nombres en italique indiquent la méthode ayant le gain le moins important et ceux en gras indiquent la méthode ayant le gain le plus important.

7.5 Estimation des gains sur la résolution totale

Ce paragraphe s'intéresse aux gains en temps CPU sur la résolution totale pour les trois méthodes simplifiées SJ, SIMP0 et SJ+SIMP0. L'objectif est double. Il est tout d'abord question (§ 7.5.1) d'apprécier le lien direct entre les temps de calcul demandés par les étapes intermédiaires (et étudiés au paragraphe 7.3) et les gains intermédiaires associés obtenus (et étudiés au paragraphe 7.4). Il s'agit ensuite de dégager les tendances de chaque méthode en fonction de p , N , RA (§ 7.5.2) et p_s (§ 7.5.3).

L'analyse porte sur l'ensemble des maillages des quatre cas tests (soit treize cas au total) et pour $1 \leq p \leq 6$. Le tableau 7.6 récapitule toutes les mesures des gains en temps CPU obtenus sur la résolution totale. Pour une meilleure vue d'ensemble, ces résultats sont regroupés par cas test, puis par maillage et enfin par stratégie d'intégration temporelle.

7.5.1 Analyse globale de rapidité

Une première analyse consiste à identifier les méthodes en temps arborant les gains les plus faibles ou les plus forts. L'objectif est d'établir un classement de rapidité.

Pour cela, on compte le nombre de fois où une méthode apparaît dans le tableau 7.6 avec les gains sur la résolution totale les plus importants et les moins importants pour chaque degré p . Le tableau 7.7 récapitule ce décompte. On relève les points suivants :

- tous les gains sont compris dans l'intervalle $[1, 00; 3, 25]$ confirmant l'intérêt des simplifications mises en œuvre, notamment pour des applications réelles, par exemple en trois dimensions d'espace ;
- il y a une différence non négligeable entre la méthode SJ dont les gains ne sont jamais les plus importants et les méthodes SIMP0 et SJ+SIMP0 dont les gains sont alternativement les plus importants ;
- le classement de rapidité des méthodes dépend fortement de p et différencie trois cas de figures :
 - (1) pour $p \leq 2$: la méthode SJ+SIMP0 présente les gains les plus importants et la méthode SIMP0 les moins importants ;
 - (2) pour $p = 3$: la méthode SJ+SIMP0 présente les gains les plus importants et la méthode SJ les moins importants ;
 - (3) pour $p \geq 4$: la méthode SIMP0 présente les gains les plus importants et la méthode SJ les moins importants. Toutefois, les gains des méthodes SIMP0 et SJ+SIMP0 restent très proches.

Une seconde analyse consiste à mettre le classement de rapidité en regard simultané avec le temps CPU requis par les étapes de calculs intermédiaires (*cf.* § 7.2) et avec les gains intermédiaires associés obtenus (*cf.* § 7.4). L'objectif est de comprendre quelles étapes (respectivement quels gains) intermédiaires donnent l'avantage à une méthode plutôt qu'à une autre.

| p | SJ | SIMP0 | SJ+SIMP0 |
|-----|----|-------|----------|
| 1 | 2 | 3 | 1 |
| 2 | 2 | 3 | 1 |
| 3 | 3 | 2 | 1 |
| 4 | 3 | 1 | 2 |
| 5 | 3 | 1 | 2 |
| 6 | 3 | 1 | 2 |

Tab. 7.7 – Classement de rapidité des méthodes SJ, SIMP0 et SJ+SIMP0 pour chaque degré $1 \leq p \leq 6$, suite au décompte du nombre de fois où ces méthodes apparaissent avec les gains les moins importants ou les plus importants dans le tableau 7.6.

| | | SJ | SIMP0 | SJ+SIMP0 |
|-------------------------------|-----------|----|-------|----------|
| Matrice implicite | Temps CPU | ++ | ++ | ++ |
| | Rapidité | 2 | 3 | 1 |
| Reste de la résolution | Temps CPU | ++ | ++ | ++ |
| | Rapidité | 3 | 1 | 2 |

(a) $p = 2$

| | | SJ | SIMP0 | SJ+SIMP0 |
|-------------------------------|-----------|-----|-------|----------|
| Matrice implicite | Temps CPU | +++ | +++ | +++ |
| | Rapidité | 2 | 3 | 1 |
| Reste de la résolution | Temps CPU | ++ | + | ++ |
| | Rapidité | 3 | 1 | 2 |

(b) $p = 3$

| | | SJ | SIMP0 | SJ+SIMP0 |
|-------------------------------|-----------|-----|-------|----------|
| Matrice implicite | Temps CPU | +++ | +++ | +++ |
| | Rapidité | 2 | 3 | 1 |
| Reste de la résolution | Temps CPU | ++ | + | + |
| | Rapidité | 3 | 1 | 2 |

(c) $p = 4$

Tab. 7.8 – Bilan recoupant l'analyse des temps CPU intermédiaires du paragraphe 7.2 et celle des gains intermédiaires du paragraphe 7.4 pour l'ensemble des cas tests, pour les méthodes SJ, SIMP0 et SJ+SIMP0, et pour $2 \leq p \leq 4$. Pour « Temps CPU », la signification des symboles est :

- + : le temps CPU est peu significatif, $Temps\ CPU < 25\%$;
- ++ : le temps CPU est moyennement significatif, $25\% \leq Temps\ CPU < 50\%$;
- +++ : le temps CPU est fortement significatif, $50\% \leq Temps\ CPU < 75\%$;
- ++++ : le temps CPU est très fortement significatif, $Temps\ CPU \geq 75\%$.

A cette fin, on se concentre sur l'étape de construction de la matrice implicite et sur celle du reste de la résolution implicite (comprenant les étapes de la construction de la matrice de préconditionnement, de l'inversion GMRes et de la reconstruction des modes d'ordre élevé) pour les trois degrés $p = 2$, $p = 3$ et $p = 4$ représentatifs du classement de rapidité précédemment établi. Le tableau 7.8 récapitule cette mise en regard. On relève les points suivants.

- Le classement de rapidité final indique que la méthode SIMP0 est la moins rapide pour $p \leq 2$. En effet, le gain sur la construction de la matrice \mathbf{A} influence le plus la rapidité de la résolution totale.
- Le classement de rapidité final indique que la méthode SJ est la moins rapide pour $p = 3$. On comprend alors que le gain sur le reste de la résolution implicite devient compétitif par rapport à celui sur la construction de la matrice \mathbf{A} .
- Le classement de rapidité final indique que la méthode SIMP0 est la plus rapide pour $p \geq 4$. De fait, le gain sur le reste de la résolution implicite influence le plus la rapidité de la résolution totale.
- Globalement, quel que soit p , les gains les plus forts observés sur la résolution complète sont en accord avec le temps CPU requis par chacune des étapes principales de calculs et les gains intermédiaires associés.

Finalement, la méthode SJ ne permet pas les gains les plus importants car elle est pénalisée par l'inversion GMRes et la reconstruction. En revanche la méthode SJ+SIMP0 présente les gains les plus importants car elle combine les avantages des deux méthodes SJ et SIMP0.

7.5.2 Analyse en ordre et en maillage

La troisième analyse concerne, à la fois la dépendance en ordre lorsque p varie (soit une analyse verticale du tableau 7.6), et à la fois la dépendance en maillage lorsque N et RA varient (soit une analyse horizontale du même tableau). L'objectif est de bien mettre en avant, non seulement les tendances de chaque méthode séparément, mais aussi l'importance de chaque méthode par rapport aux autres.

Pour des raisons de clarté, cette analyse est reportée en annexe F et seul le bilan des principales observations est exposé au tableau 7.9. On relève les points suivants.

- Le gain de l'ensemble des méthodes explicites-implicites simplifiées dépend de p et N , tandis que la dépendance en RA est soit inexistante (méthode SJ) soit très faible (méthodes SIMP0 et SJ+SIMP0). Ce point est rassurant quant à la robustesse des simplifications envisagées pour diverses applications en CFD.
- Les méthodes SIMP0 et SJ+SIMP0 sont les plus sensibles au degré p du polynôme, comme attendu à cause du découplage des modes (*cf.* hypothèse 3.1 page 81). En particulier, *Gain SIMP0* croît avec p quel que soit N ; tandis que *Gain SJ* et *Gain SJ + SIMP0* croissent avec p lorsque le maillage est grossier ($N \leq 2\,888$) et décroissent sinon.
- Pour les trois méthodes, la variation du gain en fonction de p s'effectue de façon presque monotone. Notamment, *Gain SJ* tend vers la valeur asymptotique de 1,64, *Gain SIMP0* vers 1,89 et *Gain SJ + SIMP0* vers 1,84.

| | SJ | SIMP0 | SJ+SIMP0 |
|------------------------------|---|---------------------------|--|
| Dépendance en p Détail | ++ ↗ si N petit, $\forall RA$ ↘ sinon, $\forall RA$ | ++++ ↗ $\forall N, RA$ | +++ ↗ si N petit, $\forall RA$ ↘ sinon, $\forall RA$ |
| Valeur asymptotique | 1,64 | 1,89 | 1,84 |
| Dépendance en N Détail | +++ ↗ $\forall p, RA$ | ++ ↘ $\forall p, RA$ | ++++ ↗ $\forall p, RA$ |
| Dépendance en RA Détail | NON | + ↗ $\forall p, N$ | + ↗ $\forall p, N$ |

Tab. 7.9 – Bilan de l'analyse des gains sur la résolution totale réalisée pour l'ensemble des cas tests et pour les méthodes SJ, SIMP0 et SJ+SIMP0. La signification des symboles « + » est relative entre les trois méthodes et fonction de ce qui a été observé en annexe F. Les symboles ↗ ou ↘ indiquent que le gain est respectivement croissant ou décroissant.

| p | p_s | Matrice implicite | | Résolution implicite | | Total | |
|-----|-------|-------------------|---------------|----------------------|---------------|-------------|---------------|
| | | SIMP p_s | SJ+SIMP p_s | SIMP p_s | SJ+SIMP p_s | SIMP p_s | SJ+SIMP p_s |
| 2 | 0 | 1,42 | 1,67 | 1,63 | 1,70 | 1,37 | 1,50 |
| 2 | 1 | 1,26 | 1,72 | 1,40 | 1,64 | 1,23 | 1,51 |
| 3 | 0 | 1,54 | 1,70 | 2,62 | 1,81 | 1,58 | 1,60 |
| 3 | 1 | 1,35 | 1,69 | 2,02 | 1,74 | 1,37 | 1,58 |
| 3 | 2 | 1,18 | 1,69 | 1,42 | 1,55 | 1,15 | 1,53 |
| 4 | 0 | 1,62 | 1,71 | 5,04 | 2,45 | 1,70 | 1,69 |
| 4 | 1 | 1,48 | 1,71 | 3,83 | 2,37 | 1,52 | 1,67 |
| 4 | 2 | 1,31 | 1,71 | 2,60 | 2,20 | 1,31 | 1,64 |
| 4 | 3 | 1,14 | 1,71 | 1,68 | 1,88 | 1,07 | 1,58 |
| 5 | 0 | 1,66 | 1,71 | 7,95 | 2,82 | 1,76 | 1,73 |
| 5 | 1 | 1,57 | 1,72 | 6,14 | 2,78 | 1,65 | 1,74 |
| 5 | 2 | 1,43 | 1,71 | 4,16 | 2,63 | 1,47 | 1,72 |
| 5 | 3 | 1,29 | 1,71 | 2,65 | 2,37 | 1,28 | 1,69 |
| 5 | 4 | 1,14 | 1,71 | 1,61 | 1,89 | 1,07 | 1,62 |
| 6 | 0 | 1,73 | 1,76 | 13,93 | 3,53 | 1,84 | 1,81 |
| 6 | 1 | 1,65 | 1,73 | 10,77 | 3,62 | 1,74 | 1,85 |
| 6 | 2 | 1,53 | 1,73 | 7,57 | 3,50 | 1,60 | 1,84 |
| 6 | 3 | 1,40 | 1,73 | 5,04 | 3,26 | 1,44 | 1,83 |
| 6 | 4 | 1,26 | 1,72 | 3,01 | 2,80 | 1,25 | 1,81 |
| 6 | 5 | 1,13 | 1,73 | 1,67 | 2,14 | 1,06 | 1,74 |

Tab. 7.10 – Gain en temps CPU sur les étapes de construction de la matrice implicite (gauche) et du reste de la résolution implicite (milieu), puis sur la résolution totale (droite) en faisant varier p_s . Calculs pour le cas test 2 sur le maillage uniforme tel que $N = 2888$, pour $2 \leq p \leq 6$ et $p_s < p$. Les nombres en gras indiquent la méthode ayant le gain le plus important pour chaque couple de degrés (p, p_s) .

- Les méthodes SJ et SJ+SIMP0 sont les plus sensibles au maillage, surtout au nombre N d'éléments du maillage, comme attendu à cause de la construction des différents résidus discrets supplémentaires caractéristiques de l'approche sans jacobienne. En particulier, Gain SJ et $\text{Gain SJ} + \text{SIMP0}$ augmentent avec N quel que soit p ; tandis que Gain SIMP0 a tendance à diminuer avec N quel que soit p .
- Pour les méthodes SJ et SJ+SIMP0, la variation du gain en fonction de N s'effectue de façon presque monotone tandis que pour la méthode SIMP0 cette variation est erratique.
- Le comportement et la dépendance des gains de la méthode SJ+SIMP0 sont influencés par les deux méthodes SJ et SIMP0.
- Les méthodes SJ et SJ+SIMP0 présentent les gains optimaux pour N grand et p petit, et la méthode SIMP0 pour N petit et p grand.

7.5.3 Aparté sur les simplifications utilisant le découplage

On s'intéresse ici aux méthodes de simplification $\text{SIMP}p_s$ et $\text{SJ+SIMP}p_s$, et plus particulièrement aux gains en fonction de la valeur de l'entier de couplage p_s entre les modes de bas ordre et les modes d'ordre élevé.

Le tableau 7.10 illustre les gains sur les étapes de calculs suivantes : construction de la matrice implicite, reste de la résolution implicite et résolution totale, pour les méthodes $\text{SIMP}p_s$ et $\text{SJ+SIMP}p_s$ et pour différents couples de degrés (p, p_s) sur le maillage uniforme du cas test 2 tel que $N = 2\,888$.

Le comportement du gain sur la résolution totale est différent pour les méthodes $\text{SIMP}p_s$ et $\text{SJ+SIMP}p_s$. En effet, le gain de la méthode $\text{SIMP}p_s$ diminue quand p_s tend vers p , tandis que celui de la méthode $\text{SJ+SIMP}p_s$ ne varie que très peu et semble donc indépendant de la valeur de l'entier p_s . L'explication se trouve dans l'étude suivante concernant les deux gains principaux sur la construction de la matrice implicite et sur le reste de la résolution totale.

Pour le gain sur la construction de la matrice implicite, plus p_s augmente plus la méthode $\text{SIMP}p_s$ requiert le calcul de termes supplémentaires dans les blocs extra-diagonaux et il en résulte un gain décroissant avec p_s ; tandis que la méthode $\text{SJ+SIMP}p_s$ ne requiert aucun calcul supplémentaire.

Pour le gain sur le reste de la résolution implicite, plus p_s augmente et moins il y a de modes d'ordre élevé à reconstruire pour les deux méthodes, mais cela ne suffit pas à compenser le coût de plus en plus grand de l'inversion GMRes. En particulier le produit matrice-vecteur \mathbf{Az} devient alors pénalisant pour la méthode $\text{SIMP}p_s$.

Au final, si l'avantage était à la méthode $\text{SIMP}p_s$ pour $p \geq 4$ et $p_s = 0$ (cf. tableau 7.7), ce n'est plus le cas pour $p \geq 4$ et $p_s > 0$ à cause d'un trop faible gain sur la construction de la matrice implicite. En outre, la méthode $\text{SJ+SIMP}p_s$ se présente comme plus robuste permettant une réduction significative des coûts en temps CPU.

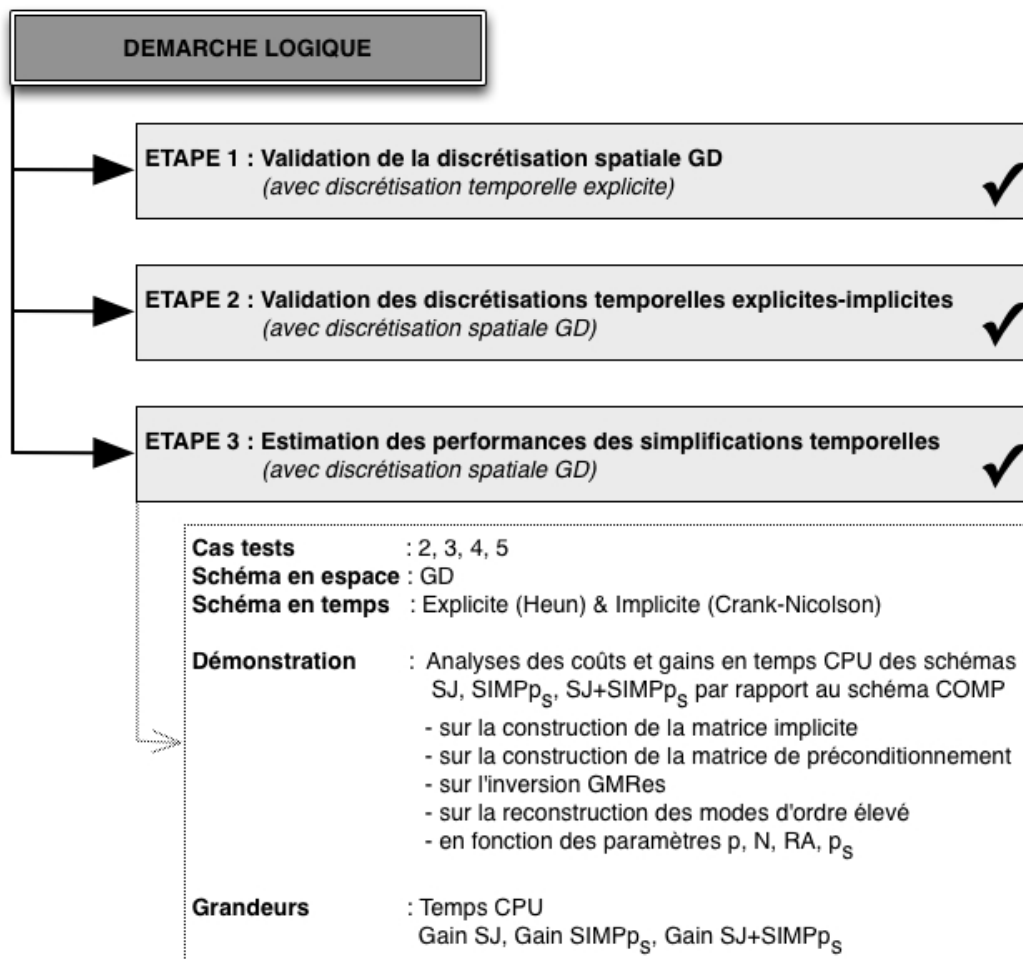


Fig. 7.4 – Schéma récapitulant l'étape 3 de validation du code de calcul développé.

7.6 Conclusion partielle

La figure 7.4 et les points suivants constituent le bilan de ce chapitre.

- Les performances des trois discrétisations temporelles explicites-implicites simplifiées : SJ, SIMP0 et SJ+SIMP0, associées à la discrétisation spatiale GD ont été établies au moyen d'analyses sur :
 - les coûts et les gains mesurés en temps CPU ;
 - les étapes de calculs intermédiaires (construction de la matrice implicite, construction de la matrice de préconditionnement, inversion GMRes et reconstruction des modes d'ordre élevé) et sur la résolution totale ;
 - les paramètres p , N , RA et p_s .
- L'analyse des coûts en temps CPU a mis en avant l'intérêt des méthodes explicites-implicites simplifiées jusqu'à trois fois plus rapides que les mêmes méthodes sans simplification.
- L'analyse des gains en temps CPU (voir également annexe F) a conduit aux constats suivants :
 - il y a une double dépendance des gains mesurés en ordre et en maillage : celle en p est très forte surtout pour la méthode SIMP0, celle en N est forte surtout pour les méthodes SJ+SIMP0 et SJ, et celle en RA est quasiment inexistante ;
 - la dépendance en ordre est en bon accord avec la théorie avec toutefois des gains plus faibles qu'espérés ;
 - la dépendance en maillage est très forte pour les deux étapes de construction de la matrice de préconditionnement et de l'inversion GMRes ;
 - les gains sur la résolution totale sont en bon accord avec le temps CPU requis par chacune des étapes principales de calculs et avec les gains intermédiaires associés ;
 - les méthodes SJ et SJ+SIMP0 présentent les gains optimaux pour N grand et p petit, tandis que la méthode SIMP0 présente les gains optimaux pour N petit et p grand.
- L'analyse des méthodes $SIMPp_s$ ou $SJ+SIMPp_s$, en fonction de différents couples de degrés (p, p_s) , révèle que la méthode $SJ+SIMPp_s$ présente des gains plutôt indépendants de p_s .

Perspectives

Les contributions majeures de ce travail ont été énoncées dans l'introduction générale de ce mémoire (*cf.* § 0.4.1 page 15), et des conclusions partielles ont été recensées à chaque chapitre. Ce dernier volet du mémoire s'intéresse donc aux voies possibles pour la poursuite de ce travail. On souligne en particulier les deux points suivants axés sur l'intégration temporelle simplifiée découplant les degrés de liberté (SIMP_{p_s}) et associée à la méthode spatiale GD pour l'application au système des équations de Navier-Stokes (1.1).

Extension de la méthode SIMP_{p_s} aux termes de convection. Les résultats énoncés au paragraphe 6.2.1 page 134 pour l'ensemble des cas tests 2 à 5 montrent que le pas de temps physique reste fortement contraint par une restriction de stabilité de type CFL associée à la discrétisation explicite en temps des flux convectifs. On note que pour les écoulements subsoniques simulés dans ce travail, la contrainte la plus forte provient des ondes acoustiques. Une première voie consiste donc à implicitiser les termes de convection afin de remédier à cette contrainte.

Des premières expériences, portant à la fois sur une analyse linéaire de stabilité pour l'équation d'advection-diffusion linéaire scalaire et sur la résolution du système des équations de Navier-Stokes [137], indiquent néanmoins que l'application de la méthode SIMP_{p_s} aux flux convectifs dans leur globalité n'est pas aisée. Ces expériences semblent montrer que la simplification de la matrice implicite détruit le décentrement amont des flux numériques et affecte ainsi la stabilité de la méthode. En outre, la simulation d'écoulements instationnaires à convection dominante requiert une résolution précise des phénomènes temporels de transport de matière. Une première approche consiste alors à utiliser une méthode de projection de Lagrange [56, chapitre III] afin de séparer le traitement des phénomènes acoustiques, responsables de la contrainte CFL la plus grande, du traitement des phénomènes de convection, pour lesquels on recherche une méthode précise. On applique alors une discrétisation implicite en temps aux termes acoustiques et une discrétisation explicite aux termes de convection.

Les méthodes de simplification introduites dans ce travail peuvent alors être appliquées à la matrice implicite comportant maintenant la linéarisation des termes acoustiques en plus de ceux de diffusion. Dans le cas favorable où ces méthodes conserveraient les propriétés de stabilité du schéma explicite-implicite sans simplification, cette approche permettrait des gains substantiels en temps de calcul grâce à une augmentation du pas de temps physique à l'échelle des phénomènes de convection. Des applications à des écoulements subsoniques à faible nombre de Mach pourraient alors être envisagées.

Discrétisation temporelle d'ordre élevé. Bien que des simulations spatiales d'ordre très élevé aient été présentées au cours de ce travail, l'approche proposée est limitée à une discrétisation temporelle d'ordre deux, comme le montre par exemple l'application du cas test 2 avec la figure 6.4 page 141. Il est alors important d'envisager des schémas temporels d'ordre élevé adaptés à la résolution spatiale afin de ne pas limiter la précision sur la solution numérique à celle du schéma temporel.

A cette fin, on peut envisager une discrétisation temporelle utilisant par exemple, des schémas RK additifs (ARK) introduits par Kennedy et Carpenter [88]. L'idée de ces schémas est de réaliser une intégration explicite-implicite au sein du même schéma RK, permettant alors une intégration d'ordre élevé en augmentant le nombre de pas.

L'application des schémas ARK à ce travail consisterait en une discrétisation explicite des termes de convection et implicite des termes de diffusion (et éventuellement des termes acoustiques). Elle se substituerait donc à la décomposition d'opérateurs de Strang. En outre, cette approche permet de relaxer la condition de stabilité (1) sur le pas de temps, observée lors de la résolution de l'équation d'advection-diffusion linéaire scalaire avec une discrétisation spatiale GD et une discrétisation temporelle explicite-implicite utilisant un schéma d'Euler sur les termes de convection et d'Euler rétrograde sur les termes de diffusion (*cf.* § 0.4.1.3 page 17).

Troisième partie

Annexes

Annexe A

Adimensionnement

Cette annexe, complément du chapitre 1, revient sur l'adimensionnement, dit « SRVT », utilisé pour le système des équations de Navier-Stokes. Dans une première partie, elle détaille son principe qui est d'appliquer le théorème de Vaschy-Buckingham pour déterminer les nombres sans dimension caractéristiques du système. Dans une seconde partie, elle développe sa mise en œuvre dans un code de calcul afin d'exprimer les variables de l'écoulement sous forme adimensionnée.

A.1 Principe pour le système des équations de Navier-Stokes

L'adimensionnement d'un système d'équations est une des applications de l'analyse dimensionnelle qui consiste à introduire les échelles de grandeurs caractéristiques des variables de l'écoulement pour faire apparaître un ensemble de nombres sans dimension. Cet ensemble est de taille inférieure à celui des variables initiales et suffit à lui-même pour décrire le système. Le processus d'adimensionnement est intéressant du point de vue numérique puisqu'il autorise des calculs entre des nombres du même ordre de grandeur, minimisant alors les erreurs numériques.

L'échelle de grandeur caractéristique (ou de référence), notée z_{ref} , lie une variable dimensionnée (ou physique), notée z , à la variable adimensionnée, notée \bar{z} , par la relation $z = z_{ref}\bar{z}$, avec $\bar{z} = \mathcal{O}(1)$. Dans le cas du système des équations de Navier-Stokes il y a, *a priori*, autant d'échelles que de variables intervenant dans le système.

La détermination des nombres sans dimension repose sur l'application du théorème de Vaschy-Buckingham [22, chapitre XV]. L'application de ce théorème au système des équations de Navier-Stokes (1.1) montre que : l'écoulement du problème considéré est complètement déterminé par les 8 variables (scalaires ou vectorielles) indépendantes¹ ($t, \mathbf{x}, \rho, \mathbf{U}, p, \theta, \mu, k$), lesquelles variables dépendent uniquement de 4 des unités fondamentales du système international : la longueur, la masse, la température et le temps, soit respectivement (L, M, Θ, T), et peuvent être regroupées pour former $8 - 4 = 4$ nombres sans dimension définissant entièrement l'état du système.

1. Comme l'air est assimilé à un gaz parfait : C_p, C_v, r et Pr (donc k) sont fixés. Les variables E, τ et \mathbf{q} sont alors respectivement données par (1.4), (1.5) et (1.7).

L'expression de ces quatre nombres sans dimension dépend alors des quatre échelles des grandeurs caractéristiques retenues pour décrire les quatre unités fondamentales nécessaires. Plusieurs combinaisons d'échelles sont possibles et donc l'adimensionnement d'un système n'est pas unique.

Celui retenu ici, correspond à l'adimensionnement SRVT [30] utilisé dans le logiciel *elsA* [118, 21] développé par l'Onera. Ses échelles de grandeurs caractéristiques sont L_{ref} , ρ_{ref} , U_{ref} et θ_{ref} , toutes les autres pouvant être exprimées en fonction d'elles comme le détaille le tableau A.1. Finalement, avec ce choix d'adimensionnement, les quatre nombres sans dimension qui apparaissent dans le système (1.1) sont les nombres de Mach M_∞ , de Reynolds Re_∞ , de Prandtl Pr et l'angle d'incidence α de l'obstacle soumis à l'écoulement. Leurs définitions sont données par les relations (1.8).

| z | Unités SI | | | | z_{ref} |
|------------------------|-----------|----------|----------|---------------|---|
| \mathbf{x} | | L | | | L_{ref} |
| ρ | M | L^{-3} | | | ρ_{ref} |
| \mathbf{U} (ou c) | | L | T^{-1} | | U_{ref} |
| θ (ou C_S) | | | | Θ | θ_{ref} |
| t | | | T | | $t_{ref} = L_{ref}/U_{ref}$ |
| E | | L^2 | T^{-2} | | $E_{ref} = U_{ref}^2$ |
| p (ou τ) | M | L^{-1} | T^{-2} | | $p_{ref} = \rho_{ref}U_{ref}^2$ |
| μ | M | L^{-1} | T^{-1} | | $\mu_{ref} = \rho_{ref}U_{ref}L_{ref}$ |
| k | M | L | T^{-3} | Θ^{-1} | $k_{ref} = \rho_{ref}U_{ref}^3L_{ref}/\theta_{ref}$ |
| \mathbf{q} | M | | T^{-3} | | $q_{ref} = \rho_{ref}U_{ref}^3$ |
| C_v (ou C_p) | | L^2 | T^{-2} | Θ^{-1} | $C_{v,ref} = U_{ref}^2/\theta_{ref}$ |
| r | | L^2 | T^{-2} | Θ^{-1} | $r_{ref} = U_{ref}^2/\theta_{ref}$ |

Tab. A.1 – Expression des échelles de grandeurs caractéristiques en fonction de L_{ref} , ρ_{ref} , U_{ref} et θ_{ref} dans le cas de l'adimensionnement SRVT.

A.2 Mise en œuvre

Ce paragraphe détermine les trois étapes de mise en œuvre de l'adimensionnement SRVT pour formuler le système des équations de Navier-Stokes (1.1). Pour ce faire, il faut exprimer en tout point de l'écoulement, chaque variable adimensionnée intervenant dans le système en fonction du vecteur des variables conservatives adimensionnées, soit $\bar{\mathbf{w}} = (\bar{\rho}, \bar{\mathbf{U}}, \bar{E})^\top$ et des quatre nombres sans dimension, soit M_∞ , Re_∞ , Pr et α .

La première étape consiste à déterminer les échelles L_{ref} , ρ_{ref} , U_{ref} et θ_{ref} en fonction des caractéristiques connues de l'écoulement. Dans la pratique, ces quatre échelles sont définies à partir de la longueur caractéristique de l'obstacle L_c et, selon les applications, des conditions infinies amont ou de champ lointain de l'écoulement ρ_∞ , U_∞ et θ_∞ , soit :

$$L_{ref} = L_c, \quad \rho_{ref} = \rho_\infty, \quad U_{ref} = \|\mathbf{U}_\infty\| = U_\infty \quad \text{et} \quad \theta_{ref} = \theta_\infty. \quad (\text{A.1})$$

La second étape consiste à initialiser le vecteur des variables conservatives adimensionnées $\bar{\mathbf{w}}^{(0)}$. Dans la pratique il vient : $\bar{\mathbf{w}}^{(0)} = \bar{\mathbf{w}}_\infty = \bar{\rho}_\infty(1, \bar{u}_{1\infty}, \bar{u}_{2\infty}, \bar{E}_\infty)^\top$, avec par définition z_∞ et \bar{z}_∞ telles que, sous une hypothèse de champ lointain :

$$\lim_{|\mathbf{x}| \rightarrow \infty} z = z_\infty \quad \text{et} \quad \lim_{|\bar{\mathbf{x}}| \rightarrow \infty} \bar{z} = \bar{z}_\infty. \quad (\text{A.2})$$

Par conséquent avec (A.1), les relations suivantes sont immédiatement vérifiées :

$$\bar{L}_c = 1, \quad \bar{\rho}_\infty = 1, \quad \bar{U}_\infty = 1 \quad \text{et} \quad \bar{\theta}_\infty = 1,$$

avec les composantes du vecteur vitesse telles que :

$$\begin{aligned} u_{1_{ref}} &= u_{1\infty} = \|\mathbf{U}_\infty\| \cos \alpha, \quad \text{donc} \quad \bar{u}_{1\infty} = \cos \alpha, \\ u_{2_{ref}} &= u_{2\infty} = \|\mathbf{U}_\infty\| \sin \alpha, \quad \text{donc} \quad \bar{u}_{2\infty} = \sin \alpha. \end{aligned}$$

Pour finaliser cette deuxième étape, il reste à définir \bar{E}_∞ par le passage à la limite de \bar{E} (donc de \bar{C}_v) donné par la relation (A.2). Le raisonnement est ici détaillé à titre d'exemple.

Chaleur spécifique à volume constant, où on rappelle que $C_v = r/(\gamma - 1)$:

$$\bar{C}_v = \frac{C_v}{C_{v_{ref}}} = \frac{r/(\gamma - 1)}{U_\infty^2/\theta_\infty} = \frac{\cancel{r}/(\gamma - 1)}{M_\infty^2 \cancel{\gamma} \cancel{\theta}_\infty / \bar{\theta}_\infty} = \frac{1}{\gamma(\gamma - 1) M_\infty^2}.$$

Energie totale, où on rappelle que $E = e + e_c = C_v \theta + \mathbf{U}^2/2$ et où on utilise $C_{v_{ref}} = U_\infty^2/\theta_\infty$:

$$\begin{aligned} \bar{E} &= \frac{E}{E_{ref}} = \frac{\cancel{C_{v_{ref}}} \cancel{C}_v \cancel{\theta}_\infty \bar{\theta}}{\cancel{U}_\infty^2} + \frac{U_\infty^2 \bar{\mathbf{U}}^2/2}{U_\infty^2} = \frac{\bar{\theta}}{\gamma(\gamma - 1) M_\infty^2} + \frac{\bar{\mathbf{U}}^2}{2}, \\ \lim_{|\bar{\mathbf{x}}| \rightarrow \infty} \bar{E} &= \bar{E}_\infty = \frac{1}{\gamma(\gamma - 1) M_\infty^2} + \frac{1}{2}. \end{aligned}$$

A ce stade, le vecteur des variables conservatives adimensionnées est initialisé. La dernière étape consiste alors à déterminer toutes les autres variables adimensionnées de l'écoulement. Le raisonnement est identique à celui présenté pour \bar{C}_v et \bar{E} à l'étape précédente. L'ensemble des résultats est directement résumé dans le tableau A.2 (voir également [30]). On souligne ici, que l'adimensionnement de la viscosité, tel que $\bar{\mu} = \mu/\mu_{ref}$ peut présenter une difficulté. En effet, pour calculer μ , la loi de Sutherland, donnée par la relation (1.6), demande de connaître la viscosité μ_s à une température θ_s donnée. Par exemple, on peut choisir $\theta_s = \theta_\infty$ et il vient alors $\mu_s = \mu(\theta_\infty)$, où l'on utilise la notation $\mu(\theta_\infty) = \mu_\infty$ avec $\mu_\infty = \mu_{ref} \bar{\mu}_\infty$ et $\bar{\mu}_\infty = 1/Re_\infty$.

| z | \bar{z} | $\lim_{ \bar{x} \rightarrow \infty} \bar{z} = \bar{z}_\infty$ |
|--------------|---|--|
| ρ | $\bar{\rho} = \rho/\rho_\infty$ | $\bar{\rho}_\infty = 1$ |
| \mathbf{U} | $\bar{\mathbf{U}} = \mathbf{U}/U_\infty$ | $\bar{\mathbf{U}}_\infty = (\bar{u}_{1\infty}, \bar{u}_{2\infty})^\top, \ \bar{\mathbf{U}}_\infty\ = 1$ |
| u_1 | $\bar{u}_1 = \ \bar{\mathbf{U}}\ \cos \alpha$ | $\bar{u}_{1\infty} = \cos \alpha$ |
| u_2 | $\bar{u}_2 = \ \bar{\mathbf{U}}\ \sin \alpha$ | $\bar{u}_{2\infty} = \sin \alpha$ |
| E | $\bar{E} = \theta/\gamma(\gamma-1)M_\infty^2 + \bar{\mathbf{U}}^2/2$ | $\bar{E}_\infty = 1/(\gamma(\gamma-1)M_\infty^2) + 1/2$ |
| C_v | $\bar{C}_v = 1/\gamma(\gamma-1)M_\infty^2$ | — |
| C_p | $\bar{C}_p = 1/(\gamma-1)M_\infty^2$ | — |
| r | $\bar{r} = 1/\gamma M_\infty^2$ | — |
| θ | $\bar{\theta} = \gamma(\gamma-1)M_\infty^2(\bar{E} - \bar{\mathbf{U}}^2/2)$ | $\bar{\theta}_\infty = 1$ |
| p | $\bar{p} = (\gamma-1)\bar{\rho}(\bar{E} - \bar{\mathbf{U}}^2/2)$ | $\bar{p}_\infty = 1/\gamma M_\infty^2$ |
| c | $\bar{c} = \sqrt{\gamma\bar{p}/\bar{\rho}}$ | $\bar{c}_\infty = 1/M_\infty$ |
| μ | $\bar{\mu} = \bar{\mu}_\infty\sqrt{\bar{\theta}}(1 + \bar{C}_s)/(1 + \bar{C}_s/\bar{\theta})$ avec $\bar{C}_s = C_s/\theta_\infty$ | $\bar{\mu}_\infty = 1/Re_\infty$ |
| k | $\bar{k} = \bar{\mu}/(\gamma-1)M_\infty^2 Pr$ | $\bar{k}_\infty = 1/(\gamma-1)M_\infty^2 Re_\infty Pr$ |
| \mathbf{q} | $\bar{\mathbf{q}} = -(\gamma/Re_\infty Pr)\bar{\nabla}(\bar{E} - \bar{\mathbf{U}}^2/2)$ | $\bar{\mathbf{q}}_\infty = 0$ |

Tab. A.2 – Expression pour l'adimensionnement SRVT, des variables adimensionnées \bar{z} en fonction des composantes du vecteur adimensionné des variables conservatives, soit $\bar{\mathbf{w}} = (\bar{\rho}, \bar{\mathbf{U}}, \bar{E})^\top$, et des quatre nombres sans dimension M_∞ , Re_∞ , Pr et α .

Annexe B

Consistance de la méthode Galerkin Discontinue pour l'équation de la chaleur monodimensionnelle scalaire

Cette annexe, complément du chapitre 2, revient sur le choix du flux numérique pour la discrétisation des problèmes paraboliques ou elliptiques contenant une dérivée seconde en espace. En particulier, un choix de flux similaire à celui effectué pour les termes de convection conduit à un schéma numérique inconsistant portant le nom de « crime variationnel » [28, 183]. L'étude est ici réalisée sur le cas particulier de l'équation de la chaleur monodimensionnelle scalaire.

B.1 Méthode de Galerkin Discontinue

Les notations utilisées sont celles du chapitre 2. Soit $\Omega = [0, 1]$ le segment unité, la méthode GD est appliquée à l'équation de la chaleur monodimensionnelle scalaire :

$$\frac{\partial w}{\partial t} - \nu \frac{\partial^2 w}{\partial x_1^2} = 0, \quad \forall (x_1, t) \in \Omega \times \mathcal{T}, \quad (\text{B.1a})$$

$$w(x_1, 0) = w^{(0)}(x_1), \quad \forall x_1 \in \Omega, \quad (\text{B.1b})$$

complétée par des conditions de périodicité aux bords. Le paramètre ν est une constante strictement positive et représente le coefficient de viscosité. Le terme en dérivée seconde est un terme de diffusion responsable du caractère parabolique de l'équation.

La formulation variationnelle s'écrit : *Trouver w appartenant à \mathcal{V} , telle que pour tout v appartenant à \mathcal{V} , il vienne à chaque instant t dans \mathcal{T}*

$$\int_{\Omega} \frac{\partial w}{\partial t} v dV + \nu \int_{\Omega} \frac{\partial w}{\partial x_1} \frac{dv}{dx_1} dV - \nu \oint_{\partial\Omega} \frac{\partial w}{\partial x_1} v dS = 0. \quad (\text{B.2})$$

Soit Ω_h une discrétisation du domaine Ω qui consiste en N segments réguliers $\kappa = [x_{1,\kappa} - \frac{h}{2}, x_{1,\kappa} + \frac{h}{2}]$, avec $h = 1/N$ la longueur du segment et $x_{1,\kappa}$ son milieu. La figure B.1 illustre cette discrétisation, en particulier la notation choisie aux interfaces.

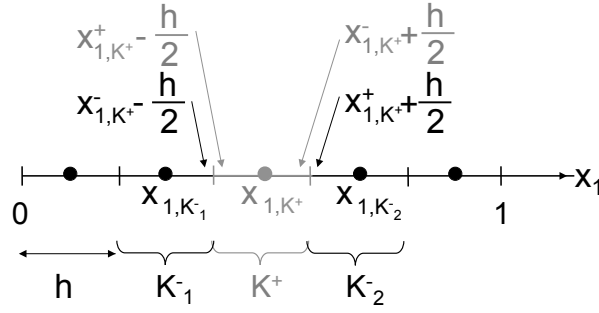


Fig. B.1 – Discrétisation et notations du segment unité.

Ensuite, la base des polynômes de Legendre définis sur l'intervalle $I = [-1, 1]$ est retenue comme base \mathcal{B}_κ^p de l'espace \mathcal{P}^p [2, chapitres 8 et 22]. Soit

$$F_\kappa(\xi_1) = \frac{h}{2}\xi_1 + x_{1,\kappa},$$

la transformation du segment de référence I vers le segment physique κ . Les fonctions tests v_h sur chaque élément κ sont choisies telles que :

$$v_h(x_1) = \phi^k(x_1) = \mathbb{L}^{k-1}(\xi_1), \quad \forall x_1 \in \kappa, \quad \kappa \in \Omega_h, \quad 1 \leq k \leq N_p.$$

Elles vérifient la propriété d'orthogonalité suivante :

$$\int_\kappa \phi^k(x_1) \phi^{k'}(x_1) dV = \frac{h}{2} \int_I \mathbb{L}^{k-1}(\xi_1) \mathbb{L}^{k'-1}(\xi_1) dV = \frac{h}{2k-1} \delta_{kk'},$$

où $\delta_{kk'}$ est le symbole de Kronecker. De plus, les relations suivantes sont vérifiées :

$$\phi^k(x_{1,\kappa} - \frac{h}{2}) = (-1)^k, \quad \text{et} \quad \phi^k(x_{1,\kappa} + \frac{h}{2}) = 1, \quad \forall x_1 \in \kappa, \quad \kappa \in \Omega_h, \quad 1 \leq k \leq N_p.$$

En tenant compte des points précédents, l'approximation interne de (B.2) s'écrit ainsi : Trouver w_h appartenant à \mathcal{V}_h^p , telle que pour tout κ dans Ω_h et ϕ^k élément de \mathcal{B}_κ^p , il vienne à chaque instant t dans \mathcal{T}

$$\int_\kappa \frac{\partial w_h}{\partial t} \phi^k dV + \nu \int_\kappa \frac{\partial w_h}{\partial x_1} \frac{d\phi^k}{dx_1} dV - \nu \oint_{\partial\kappa} \widehat{\frac{\partial w_h}{\partial x_1}} \phi^k dS = 0, \quad (\text{B.3})$$

où w_h est donnée par la relation (2.8). Un premier choix dit « naïf », pour le flux numérique de la dérivée spatiale est un flux numérique centré. Ce dernier est consistant au sens des différences finies et respecte la physique des phénomènes de diffusion, lesquels ne privilégient aucune direction de la propagation de l'information. Par exemple en $x_1 = x_{1,\kappa} + \frac{h}{2}$, il s'écrit :

$$\widehat{\frac{\partial w_h}{\partial x_1}} = \left\{ \frac{\partial w_h}{\partial x_1} \right\} = \frac{1}{2} \left(\left. \frac{\partial w_h}{\partial x_1} \right|_{x_1=x_{1,\kappa}^+ + \frac{h}{2}} + \left. \frac{\partial w_h}{\partial x_1} \right|_{x_1=x_{1,\kappa}^- + \frac{h}{2}} \right).$$

Ainsi, en utilisant (2.8), la forme semi-discrète de (B.3) s'écrit sur chaque élément

$$\begin{aligned} & \frac{h}{2k-1} \frac{dW_{\kappa^+}^k}{dt} + \nu \sum_{l=1}^{N_p} \left(W_{\kappa^+}^l \int_{\kappa} \frac{d\phi^l}{dx_1} \frac{d\phi^k}{dx_1} dV \right) \\ & - \frac{\nu}{2} \left[\sum_{l=1}^{N_p} \left(W_{\kappa^+}^l \frac{d\phi^l}{dx_1} \Big|_{x_1=x_{1,\kappa^+}^- + \frac{h}{2}} \right) + \sum_{l=1}^{N_p} \left(W_{\kappa_2^-}^l \frac{d\phi^l}{dx_1} \Big|_{x_1=x_{1,\kappa^+}^+ + \frac{h}{2}} \right) \right] \\ & + (-1)^k \frac{\nu}{2} \left[\sum_{l=1}^{N_p} \left(W_{\kappa_1^-}^l \frac{d\phi^l}{dx_1} \Big|_{x_1=x_{1,\kappa^+}^- - \frac{h}{2}} \right) + \sum_{l=1}^{N_p} \left(W_{\kappa^+}^l \frac{d\phi^l}{dx_1} \Big|_{x_1=x_{1,\kappa^+}^+ - \frac{h}{2}} \right) \right] = 0, \quad (\text{B.4}) \end{aligned}$$

où $(W_{\kappa}^l)_{1 \leq l \leq N_p}$ représente les composantes de $\mathbf{w}_{h|\kappa}$ dans la base de fonctions.

B.2 Problème de consistance lié à l'approche naïve

L'analyse de la consistance est menée avec la méthode des équations équivalentes. Soit la définition suivante [129] :

Définition B.1. (Système équivalent) *Pour un schéma précis à l'ordre a , le système dit équivalent à l'ordre b ($b > a$) est un système d'équations aux dérivées partielles approché par ce schéma avec une précision d'ordre b . Le système équivalent est obtenu en effectuant des développements de Taylor de la solution numérique autour du centre $x_{1,\kappa}$ de chaque élément κ appartenant à Ω_h . Il comprend les termes du système exact, et des termes complémentaires caractéristiques du schéma.*

La consistance du schéma GD pour $p = 1$ est alors analysée en prenant l'équation (B.4) pour $k = 1$ et $k = 2$, et sachant que $\phi^1(x_1) = \mathbb{L}^0(\xi_1) = 1$ et que $\phi^2(x_1) = \mathbb{L}^1(\xi_1) = 2(x_1 - x_{1,\kappa})/h$. Il vient :

$$\frac{dW_{\kappa^+}^1}{dt} - \frac{\nu}{h^2} [W_{\kappa_2^-}^2 - W_{\kappa_1^-}^2] = 0, \quad (\text{B.5a})$$

$$\frac{dW_{\kappa^+}^2}{dt} - \frac{3\nu}{h^2} [W_{\kappa_2^-}^2 - 2W_{\kappa^+}^2 + W_{\kappa_1^-}^2] = 0. \quad (\text{B.5b})$$

On indique dès à présent, que le terme $W_{\kappa^+}^1$ n'intervient pas dans les quotients différentiels du système (B.5). Cependant, plutôt que de conclure hâtivement, on mène l'analyse de la consistance dans son intégralité.

Ainsi, les degrés de liberté de la solution numérique w_h sont considérés comme étant des fonctions régulières de l'espace et du temps, soit $W_{\kappa}^l = W^l(x_1, t)$ pour $l = 1$ et $l = 2$. Vu le système (B.5), il convient d'utiliser les développements limités de W^2 à l'ordre trois en x_{1,κ_1^-} et en x_{1,κ_2^-} autour de x_{1,κ^+} :

$$W^2(x_{1,\kappa_1^-}, t) = W^2(x_{1,\kappa^+}, t) - h \frac{\partial W^2}{\partial x_1} \Big|_{(x_{1,\kappa^+}, t)} + \frac{h^2}{2} \frac{\partial^2 W^2}{\partial x_1^2} \Big|_{(x_{1,\kappa^+}, t)} - \frac{h^3}{6} \frac{\partial^3 W^2}{\partial x_1^3} \Big|_{(x_{1,\kappa^+}, t)} + \mathcal{O}(h^4),$$

$$W^2(x_{1,\kappa_2^-}, t) = W^2(x_{1,\kappa^+}, t) + h \frac{\partial W^2}{\partial x_1} \Big|_{(x_{1,\kappa^+}, t)} + \frac{h^2}{2} \frac{\partial^2 W^2}{\partial x_1^2} \Big|_{(x_{1,\kappa^+}, t)} + \frac{h^3}{6} \frac{\partial^3 W^2}{\partial x_1^3} \Big|_{(x_{1,\kappa^+}, t)} + \mathcal{O}(h^4).$$

En substituant ces développements dans (B.5), il vient le système équivalent :

$$\frac{\partial W_{\kappa^+}^1}{\partial t} - \frac{\nu}{h^2} \left[2h \frac{\partial W_{\kappa^+}^2}{\partial x_1} + \frac{h^3}{3} \frac{\partial^3 W_{\kappa^+}^2}{\partial x_1^3} \right] = \mathcal{O}(h^4), \quad (\text{B.6a})$$

$$\frac{\partial W_{\kappa^+}^2}{\partial t} - \frac{3\nu}{h^2} \left[h^2 \frac{\partial^2 W_{\kappa^+}^2}{\partial x_1^2} \right] = \mathcal{O}(h^3). \quad (\text{B.6b})$$

Le système (B.5) est dit consistant avec la formulation variationnelle du point de vue de l'analyse du schéma équivalent si et seulement si l'équation (B.6a) du système équivalent contient les termes du problème exact (B.1) et des termes complémentaires caractéristiques du schéma. Autrement dit, si (B.6a) peut s'écrire :

$$\frac{\partial W_{\kappa^+}^1}{\partial t} - \nu \frac{\partial^2 W_{\kappa^+}^1}{\partial x_1^2} + f_{TC} \left(\frac{\partial^i W_{\kappa^+}^m}{\partial x_1^i}, h, \nu \right) = 0,$$

avec f_{TC} une fonction des termes complémentaires, à savoir des dérivées des degrés de liberté W^m d'ordre i , pour i entier positif et $1 \leq m \leq N_p$, et qui tend vers zéro lorsque h tend vers zéro.

Après observation du système équivalent (B.6), la consistance du schéma (B.5) n'est pas établie, puisque le terme du système exact $\partial^2 W_{\kappa^+}^1 / \partial x_1^2$ n'apparaît pas dans l'équation (B.6a). En effet l'équation (B.6b) ne contient ni de terme en $W_{\kappa^+}^1$ ni de terme avec les dérivées spatiales de $W_{\kappa^+}^1$, termes qui permettraient de récrire (B.6a).

De fait, le problème discret (B.4) est inconsistant du point de vue de l'analyse par schéma équivalent. Cette inconsistance vient du choix du flux numérique, à savoir le choix de l'approche « naïve ». Elle porte le nom de « crime variationnel » [28, 183].

Annexe C

Résultats numériques de précision pour la validation spatiale

Cette annexe est un complément du chapitre 5 sur la validation de la discrétisation spatiale GD, soit sur l'étape 1. Plus particulièrement, il s'agit de la validation de la discrétisation des termes de convection par la simulation de l'écoulement subsonique dans un canal de type convergent-divergent (cas test 1). Pour chaque degré p , les valeurs numériques des erreurs globales en norme L^1 , L^2 et L^∞ sur l'entropie, ainsi que le calcul des ordres de convergence associés à ces erreurs sont récapitulés.

On considère les erreurs globales en norme L^1 , L^2 et L^∞ sur l'entropie s définies respectivement par :

$$\begin{aligned} |s - s_\infty|_{L^1(\Omega_h)} &= \frac{\int_{\Omega_h} \left| \frac{p/\rho^\gamma}{p_\infty/\rho_\infty^\gamma} - 1 \right| dv}{\int_{\Omega_h} dv}, \\ \|s - s_\infty\|_{L^2(\Omega_h)} &= \sqrt{\frac{\int_{\Omega_h} \left(\frac{p/\rho^\gamma}{p_\infty/\rho_\infty^\gamma} - 1 \right)^2 dv}{\int_{\Omega_h} dv}}, \\ \|s - s_\infty\|_{L^\infty(\Omega_h)} &= \max \left\{ \left| \frac{p/\rho^\gamma}{p_\infty/\rho_\infty^\gamma} - 1 \right|, \forall \mathbf{x} \in \Omega_h \right\}. \end{aligned}$$

Le tableau C.1 récapitule les valeurs numériques de ces trois erreurs, ainsi que l'ordre de convergence a tel que chacune de ces erreurs soit de l'ordre de $(1/\sqrt{NN_p})^a$ pour chaque degré p de polynôme.

| | $1/\sqrt{NN_p}$ | $ s - s_\infty _{L^1(\Omega_h)}$ | a | $\ s - s_\infty\ _{L^2(\Omega_h)}$ | a | $\ s - s_\infty\ _{L^\infty(\Omega_h)}$ | a |
|-------|-----------------------|----------------------------------|------|------------------------------------|------|---|------|
| $p=0$ | $1/\sqrt{32 * 1}$ | $5,669065E - 03$ | — | $7,604533E - 03$ | — | $2,078750E - 02$ | — |
| | $1/\sqrt{128 * 1}$ | $1,100656E - 03$ | 2,37 | $2,446604E - 03$ | 1,64 | $1,240395E - 02$ | 0,74 |
| | $1/\sqrt{512 * 1}$ | $3,932175E - 04$ | 1,49 | $1,114135E - 03$ | 1,13 | $6,879876E - 03$ | 0,85 |
| | $1/\sqrt{2048 * 1}$ | $1,530614E - 04$ | 1,36 | $4,754828E - 04$ | 1,23 | $3,601676E - 03$ | 0,93 |
| | $1/\sqrt{8192 * 1}$ | $6,371040E - 05$ | 1,27 | $2,009937E - 04$ | 1,24 | $1,608919E - 03$ | 1,16 |
| | $1/\sqrt{32768 * 1}$ | $2,842906E - 05$ | 1,16 | $8,900509E - 05$ | 1,18 | $6,947320E - 04$ | 1,21 |
| $p=1$ | $1/\sqrt{32 * 3}$ | $1,006944E - 03$ | — | $1,802947E - 03$ | — | $6,466806E - 03$ | — |
| | $1/\sqrt{128 * 3}$ | $2,033180E - 04$ | 2,31 | $5,544891E - 04$ | 1,70 | $3,137162E - 03$ | 1,04 |
| | $1/\sqrt{512 * 3}$ | $3,296705E - 05$ | 2,63 | $1,172682E - 04$ | 2,24 | $1,129482E - 03$ | 1,47 |
| | $1/\sqrt{2048 * 3}$ | $6,965145E - 06$ | 2,24 | $2,554189E - 05$ | 2,20 | $3,463907E - 04$ | 1,71 |
| | $1/\sqrt{8192 * 3}$ | $1,423515E - 06$ | 2,29 | $5,134735E - 06$ | 2,31 | $8,224884E - 05$ | 2,07 |
| | $1/\sqrt{32768 * 3}$ | $3,042830E - 07$ | 2,23 | $1,220010E - 06$ | 2,07 | $2,447457E - 05$ | 1,75 |
| $p=2$ | $1/\sqrt{32 * 6}$ | $2,683827E - 04$ | — | $6,283330E - 04$ | — | $3,634015E - 03$ | — |
| | $1/\sqrt{128 * 6}$ | $3,633697E - 05$ | 2,89 | $9,996674E - 05$ | 2,65 | $1,755755E - 03$ | 1,05 |
| | $1/\sqrt{512 * 6}$ | $2,812412E - 06$ | 3,69 | $9,289083E - 06$ | 3,43 | $1,782858E - 04$ | 3,30 |
| | $1/\sqrt{2048 * 6}$ | $1,868771E - 07$ | 3,91 | $7,924597E - 07$ | 3,55 | $2,519317E - 05$ | 2,82 |
| | $1/\sqrt{8192 * 6}$ | $1,692017E - 08$ | 3,47 | $8,057523E - 08$ | 3,30 | $3,288741E - 06$ | 2,94 |
| | $1/\sqrt{32768 * 6}$ | $2,193991E - 09$ | 2,95 | $1,032286E - 08$ | 2,96 | $5,021187E - 07$ | 2,71 |
| $p=3$ | $1/\sqrt{32 * 10}$ | $7,750532E - 05$ | — | $1,457121E - 04$ | — | $1,223694E - 03$ | — |
| | $1/\sqrt{128 * 10}$ | $7,586537E - 06$ | 3,35 | $1,927904E - 05$ | 2,92 | $3,556286E - 04$ | 1,78 |
| | $1/\sqrt{512 * 10}$ | $3,365487E - 07$ | 4,49 | $1,224695E - 06$ | 3,98 | $5,401610E - 05$ | 2,72 |
| | $1/\sqrt{2048 * 10}$ | $1,860070E - 08$ | 4,18 | $8,631557E - 08$ | 3,83 | $5,061909E - 06$ | 3,42 |
| | $1/\sqrt{8192 * 10}$ | $1,146295E - 09$ | 4,02 | $5,713166E - 09$ | 3,92 | $4,960547E - 07$ | 3,35 |
| | $1/\sqrt{32768 * 10}$ | $7,323113E - 11$ | 3,97 | $3,868200E - 10$ | 3,88 | $3,670631E - 08$ | 3,76 |
| $p=4$ | $1/\sqrt{32 * 15}$ | $4,772980E - 05$ | — | $8,072194E - 05$ | — | $5,567984E - 04$ | — |
| | $1/\sqrt{128 * 15}$ | $2,232136E - 06$ | 4,42 | $5,838363E - 06$ | 3,79 | $1,229086E - 04$ | 2,18 |
| | $1/\sqrt{512 * 15}$ | $4,818157E - 08$ | 5,53 | $1,997080E - 07$ | 4,87 | $1,541124E - 05$ | 3,00 |
| | $1/\sqrt{2048 * 15}$ | $1,200615E - 09$ | 5,33 | $5,926043E - 09$ | 5,07 | $4,717242E - 07$ | 5,03 |
| | $1/\sqrt{8192 * 15}$ | $3,450490E - 11$ | 5,12 | $1,918311E - 10$ | 4,95 | $2,180391E - 08$ | 4,44 |
| | $1/\sqrt{32768 * 15}$ | $1,166064E - 12$ | 4,89 | $6,597238E - 12$ | 4,86 | $8,156300E - 10$ | 4,74 |
| $p=5$ | $1/\sqrt{32 * 21}$ | $1,157228E - 05$ | — | $1,974252E - 05$ | — | $6,893504E - 04$ | — |
| | $1/\sqrt{128 * 21}$ | $4,817107E - 07$ | 4,59 | $1,567834E - 06$ | 3,65 | $1,044360E - 04$ | 2,72 |
| | $1/\sqrt{512 * 21}$ | $1,031364E - 08$ | 5,55 | $3,432936E - 08$ | 5,51 | $3,007210E - 06$ | 5,12 |
| | $1/\sqrt{2048 * 21}$ | $2,536195E - 10$ | 5,35 | $8,676548E - 10$ | 5,31 | $1,273561E - 07$ | 4,56 |
| | $1/\sqrt{8192 * 21}$ | $7,234489E - 12$ | 5,13 | $3,271690E - 11$ | 4,73 | $2,071063E - 09$ | 5,94 |
| | $1/\sqrt{32768 * 21}$ | $2,484000E - 13$ | 4,86 | $1,279740E - 12$ | 4,68 | $1,273222E - 10$ | 4,02 |

Tab. C.1 – (Canal). Ordres de précision a en espace tels que les erreurs globales en norme L^1 , L^2 et L^∞ sur l'entropie soient de l'ordre de $(1/\sqrt{NN_p})^a$ pour $0 \leq p \leq 5$.

Annexe D

Résultats numériques de précision pour la validation temporelle

Cette annexe est un complément du chapitre 6 sur la validation des discrétisations temporelles explicites-implicites, soit sur l'étape 2. Plus particulièrement, elle concerne la simulation de l'écoulement de tourbillons contra-rotatifs (cas test 2). Pour chaque degré p , les valeurs numériques des erreurs globales en norme L^1 et L^∞ sur la quantité de mouvement projetée selon x_1 , ainsi que le calcul des ordres de convergence associés à ces erreurs sont récapitulés pour les quatre méthodes COMP, SJ, SIMP0 et SJ+SIMP0.

On considère les erreurs globales en norme L^1 et L^∞ sur la quantité de mouvement projetée selon x_1 définies respectivement par :

$$|(\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE}|_{L^1(\Omega_h)} = \frac{\int_{\Omega_h} |(\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE}| dv}{\int_{\Omega_h} dv},$$
$$\|(\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE}\|_{L^\infty(\Omega_h)} = \max \left\{ |(\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE}|, \forall \mathbf{x} \in \Omega_h \right\},$$

où

- $(\rho u_1)_{REF}$ est la solution numérique obtenue avec le schéma EXPL pour $p = 4$ et $CFL = 10^{-2}$;
- $(\rho u_1)_{h,AUTRE}$ est la solution numérique obtenue avec un schéma COMP, SJ, SIMP0 ou SJ+SIMP0.

Le tableau D.1a récapitule les valeurs numériques de l'erreur en norme L^1 , ainsi que l'ordre de convergence a tel que $|(\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE}|_{L^1(\Omega_h)}$ soit de l'ordre de $(1/\sqrt{NN_p})^a$ pour chaque degré p de polynôme. Le tableau D.1b fait de même pour l'erreur en norme L^∞ .

| | $1/\sqrt{NN_p}$ | Erreur : $ (\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE} _{L^1(\Omega_h)}$ | | | | a |
|---------|-----------------------|--|----------------|----------------|-------------------|------|
| | | <i>COMP</i> | <i>SJ</i> | <i>SIMP0</i> | <i>SJ + SIMP0</i> | |
| $p = 1$ | $1/\sqrt{200 * 3}$ | 2,922664E - 01 | 2,922664E - 01 | 2,922668E - 01 | 2,9226677E - 01 | – |
| | $1/\sqrt{648 * 3}$ | 2,605908E - 01 | 2,605908E - 01 | 2,605912E - 01 | 2,605912E - 01 | 0,20 |
| | $1/\sqrt{2888 * 3}$ | 1,280510E - 01 | 1,280510E - 01 | 1,280513E - 01 | 1,280513E - 01 | 0,95 |
| | $1/\sqrt{10952 * 3}$ | 3,729526E - 02 | 3,729524E - 02 | 3,729472E - 02 | 3,729472E - 02 | 1,85 |
| | $1/\sqrt{43808 * 3}$ | 8,522203E - 03 | 8,522494E - 03 | 8,522495E - 03 | 8,522493E - 03 | 2,13 |
| $p = 2$ | $1/\sqrt{200 * 6}$ | 2,451241E - 01 | 2,451241E - 01 | 2,451268E - 01 | 2,451268E - 01 | – |
| | $1/\sqrt{648 * 6}$ | 1,089252E - 01 | 1,089252E - 01 | 1,089281E - 01 | 1,089281E - 01 | 1,38 |
| | $1/\sqrt{2888 * 6}$ | 2,233188E - 02 | 2,233187E - 02 | 2,233614E - 02 | 2,233613E - 02 | 2,12 |
| | $1/\sqrt{10952 * 6}$ | 3,012475E - 03 | 3,012489E - 03 | 3,011951E - 03 | 3,011951E - 03 | 3,01 |
| | $1/\sqrt{43808 * 6}$ | 1,583091E - 03 | 1,583266E - 03 | 1,582657E - 03 | 1,582655E - 03 | 0,93 |
| $p = 3$ | $1/\sqrt{200 * 10}$ | 1,801523E - 01 | 1,801523E - 01 | 1,801555E - 01 | 1,801555E - 01 | – |
| | $1/\sqrt{648 * 10}$ | 4,005502E - 02 | 4,005501E - 02 | 4,006198E - 02 | 4,006198E - 02 | 2,56 |
| | $1/\sqrt{2888 * 10}$ | 3,002319E - 03 | 3,002319E - 03 | 3,003693E - 03 | 3,003692E - 03 | 3,47 |
| | $1/\sqrt{10952 * 10}$ | 8,676522E - 04 | 8,676991E - 04 | 8,660343E - 04 | 8,660342E - 04 | 1,86 |
| | $1/\sqrt{43808 * 10}$ | 2,243125E - 04 | 2,243865E - 04 | 2,24864E - 04 | 2,24865E - 04 | 1,95 |
| $p = 4$ | $1/\sqrt{200 * 15}$ | 5,148597E - 03 | 5,148600E - 03 | 5,148600E - 03 | 5,149032E - 03 | – |
| | $1/\sqrt{648 * 15}$ | 3,135385E - 03 | 3,135386E - 03 | 3,135461E - 03 | 3,135461E - 03 | 0,84 |
| | $1/\sqrt{2888 * 15}$ | 7,892238E - 04 | 7,892218E - 04 | 7,893646E - 04 | 7,893644E - 04 | 1,85 |
| | $1/\sqrt{10952 * 15}$ | 1,996538E - 04 | 1,996651E - 04 | 1,996651E - 04 | 2,002930E - 04 | 2,06 |
| | $1/\sqrt{43808 * 15}$ | 4,827460E - 05 | 4,827698E - 05 | 4,819758E - 05 | 4,828526E - 05 | 2,05 |

(a) Erreur en norme L^1

| | $1/\sqrt{NN_p}$ | Erreur : $\ (\rho u_1)_{REF} - (\rho u_1)_{h,AUTRE}\ _{L^\infty(\Omega_h)}$ | | | | a |
|---------|-----------------------|---|-----------------|-----------------|-------------------|-------|
| | | <i>COMP</i> | <i>SJ</i> | <i>SIMP0</i> | <i>SJ + SIMP0</i> | |
| $p = 1$ | $1/\sqrt{200 * 3}$ | 8,861097E + 00 | 8,861097E + 00 | 8,861094E + 00 | 8,861094E + 00 | – |
| | $1/\sqrt{648 * 3}$ | 9,080836E + 00 | 9,080836E + 00 | 9,080823E + 00 | 9,080823E + 00 | –0,04 |
| | $1/\sqrt{2888 * 3}$ | 5,567453E + 00 | 5,567453E + 00 | 5,567499E + 00 | 5,567499E + 00 | 0,65 |
| | $1/\sqrt{10952 * 3}$ | 2,022473E + 00 | 2,022473E + 00 | 2,022647E + 00 | 2,022647E + 00 | 1,52 |
| | $1/\sqrt{43808 * 3}$ | 6,275177E - 01 | 6,275167E - 01 | 6,276723E - 01 | 6,276723E - 01 | 1,69 |
| $p = 2$ | $1/\sqrt{200 * 6}$ | 8,914368E + 00 | 8,914368E + 00 | 8,914455E + 00 | 8,914455E + 00 | – |
| | $1/\sqrt{648 * 6}$ | 7,128707E + 00 | 7,128706E + 00 | 7,128983E + 00 | 7,128983E + 00 | 0,38 |
| | $1/\sqrt{2888 * 6}$ | 3,307237E + 00 | 3,307235E + 00 | 3,308149E + 00 | 3,308149E + 00 | 1,03 |
| | $1/\sqrt{10952 * 6}$ | 5,260844E - 01 | 5,260834E - 01 | 5,268795E - 01 | 5,268791E - 01 | 2,76 |
| | $1/\sqrt{43808 * 6}$ | 4,400972E - 02 | 4,400233E - 02 | 4,393260E - 02 | 4,393256E - 02 | 3,58 |
| $p = 3$ | $1/\sqrt{200 * 10}$ | 11,318641E + 01 | 11,318641E + 01 | 11,318761E + 01 | 11,318761E + 01 | – |
| | $1/\sqrt{648 * 10}$ | 3,360035E + 00 | 3,360034E + 00 | 3,361178E + 00 | 3,361177E + 00 | 2,07 |
| | $1/\sqrt{2888 * 10}$ | 2,291441E - 01 | 2,291433E - 01 | 2,294727E - 01 | 2,294723E - 01 | 3,59 |
| | $1/\sqrt{10952 * 10}$ | 3,226142E - 02 | 3,228012E - 02 | 3,226142E - 02 | 3,226168E - 02 | 2,94 |
| | $1/\sqrt{43808 * 10}$ | 5,013539E - 03 | 5,013596E - 03 | 5,015958E - 03 | 5,015959E - 03 | 2,69 |
| $p = 4$ | $1/\sqrt{200 * 15}$ | 2,746686E - 02 | 2,746683E - 02 | 2,747365E - 02 | 2,747365E - 02 | – |
| | $1/\sqrt{648 * 15}$ | 2,398056E - 02 | 2,398059E - 02 | 2,398020E - 02 | 2,398020E - 02 | 0,23 |
| | $1/\sqrt{2888 * 15}$ | 6,073421E - 03 | 6,073411E - 03 | 6,072986E - 03 | 6,072986E - 03 | 1,84 |
| | $1/\sqrt{10952 * 15}$ | 1,579119E - 03 | 1,577181E - 03 | 1,577181E - 03 | 1,578476E - 03 | 2,02 |
| | $1/\sqrt{43808 * 15}$ | 4,929502E - 04 | 4,929762E - 04 | 4,929763E - 04 | 4,929420E - 04 | 1,68 |

(b) Erreur en norme L^∞

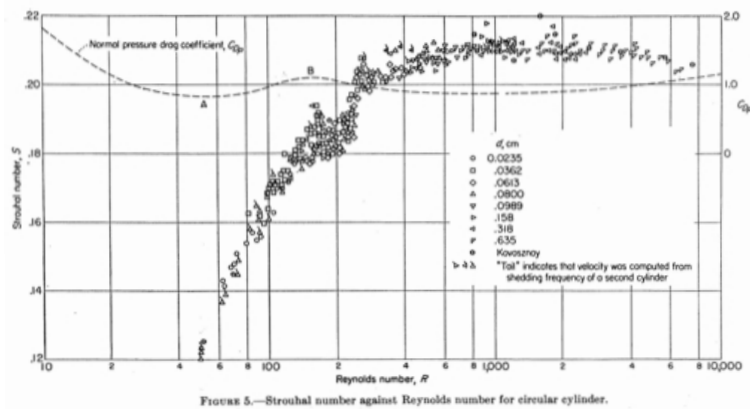
Tab. D.1 – (*Interaction tourbillons-paroi*). Ordres de précision a en espace tels que les erreurs en norme L^1 et L^∞ sur la quantité de mouvement projetée selon x_1 soient de l'ordre de $(1/\sqrt{NN_p})^a$ pour $1 \leq p \leq 4$.

Annexe E

Résultats expérimentaux de référence pour la validation temporelle

Cette annexe est un complément du chapitre 6 sur la validation des discrétisations temporelles explicites-implicites, soit sur l'étape 2. Plus particulièrement, elle concerne la simulation de l'écoulement autour du cylindre (cas test 5). Les copies des courbes expérimentales, où ont été lus le coefficient C_D et le nombre St servant de valeurs de référence, sont reproduites à titre indicatif.

Parmi les nombreux auteurs ayant réalisé des expérimentations sur le cas test 5 (*cf.* § 6.4.2 page 155), on s'est restreint à ceux les plus mentionnés : Roshko [146], Tritton [163] et Williamson [179]. La figure E.1 reproduit alors les courbes où ont été lues les différentes valeurs expérimentales, du coefficient C_D et du nombre St , reportées dans le tableau 6.9 page 159 et servant de valeurs de référence.



(a) Figure 5 extraite de Roshko [146]

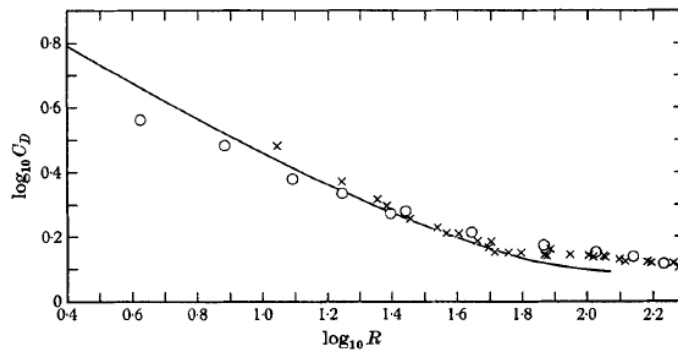
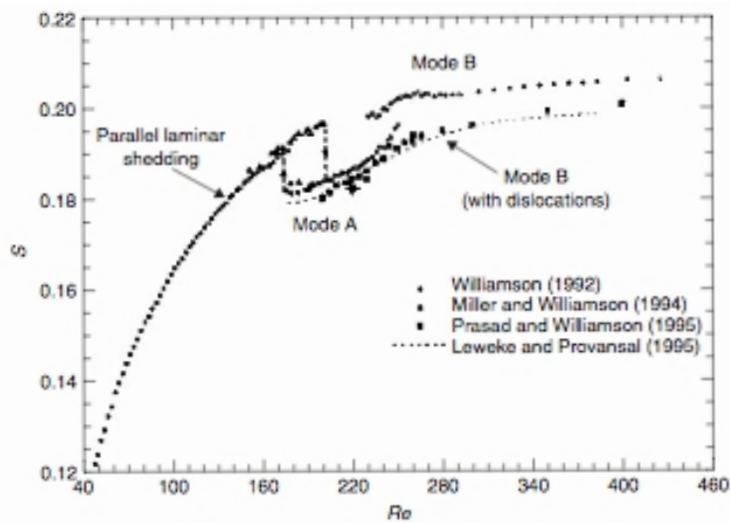


FIGURE 2. Comparison with other experimental measurements of the drag coefficient
x, Relf; O, Wieselsberger; —, Tritton.

(b) Figure 2 extraite de Tritton [163]



(c) Figure 11 extraite de Williamson [179]

Fig. E.1 – (*Cylindre*). Reproduction des courbes expérimentales utilisées pour la lecture graphique du coefficient C_D et du nombre St , tous les deux indiqués en valeurs de référence au tableau 6.9 page 159.

Annexe F

Analyse des gains sur la résolution totale

Cette annexe est un complément du chapitre 7 sur l'estimation des performances des discrétisations temporelles explicites-implicites, soit sur l'étape 3. Plus particulièrement, il est question de l'analyse paramétrique en ordre et en maillage des gains obtenus sur la résolution totale du schéma explicite-implicite, et ce pour les trois méthodes simplifiées SJ, SIMP0 et SJ+SIMP0.

F.1 Prérequis de cette analyse

L'objectif ici, est de bien mettre en avant, non seulement les tendances de chaque méthode séparément, mais aussi l'importance de chaque méthode par rapport aux autres. L'analyse en ordre porte sur p : c'est une analyse verticale du tableau 7.6 page 176, et l'analyse en maillage porte sur N et RA : c'est une analyse horizontale de ce même tableau.

Cette analyse concerne l'intégralité des treize cas tests. Elle est menée en trois paragraphes (§ F.2, § F.3 et § F.4), lesquels correspondent respectivement aux trois méthodes explicites-implicites simplifiées SJ, SIMP0 et SJ+SIMP0. Le bilan de cette analyse est présenté dans le corps du mémoire (*cf.* § 7.5.2 page 179).

Pour chacune des trois méthodes, l'analyse en ordre (p est variable) est menée d'abord par N croissants puis par RA croissants. Suivant le même principe, l'analyse en maillage considère N ou RA variables. Lorsque N varie, l'analyse est menée d'abord par RA croissants puis par p croissants. Lorsque RA varie, l'analyse est menée d'abord par N croissants puis par p croissants.

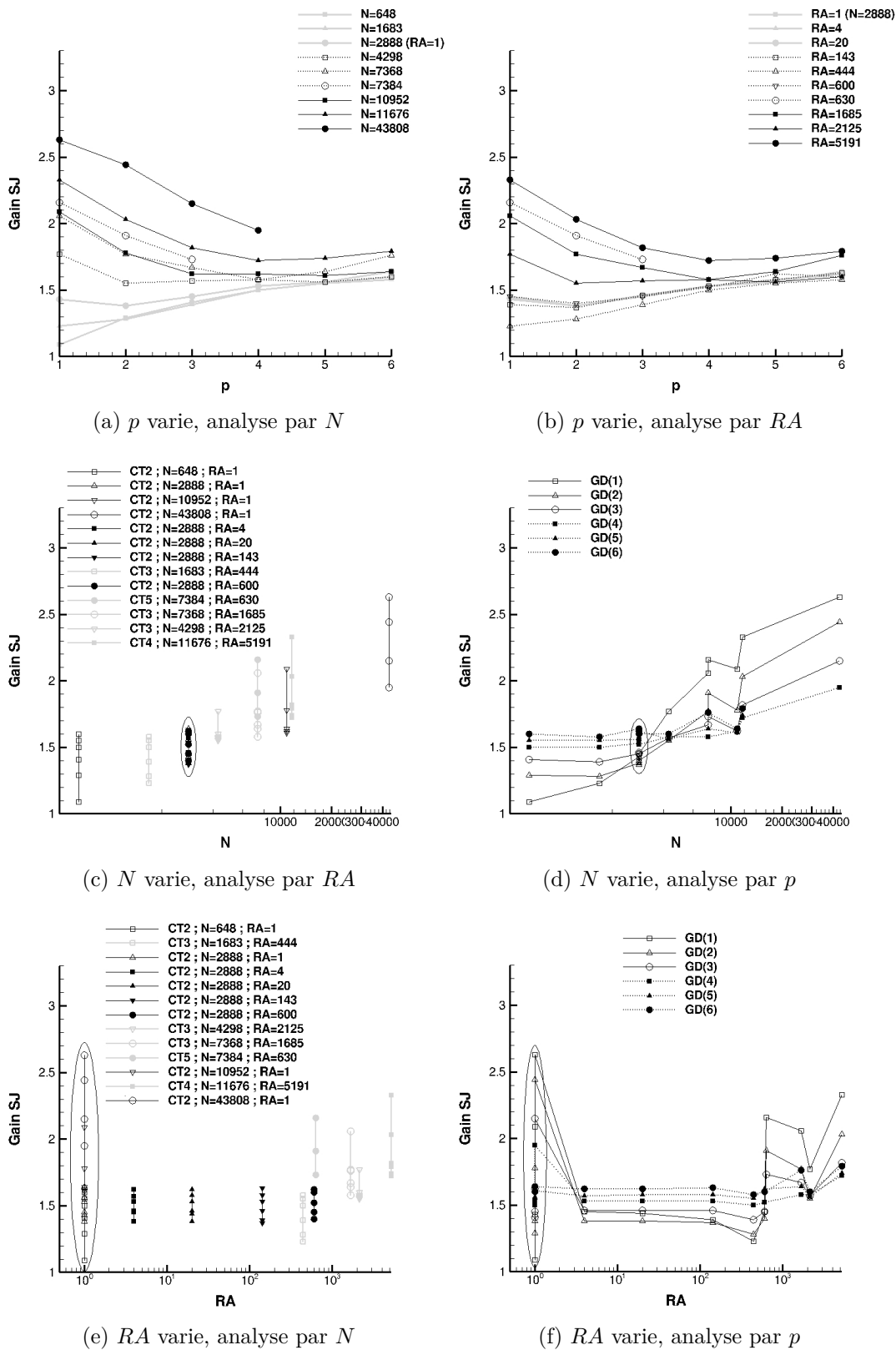


Fig. F.1 – Gain en temps CPU sur la résolution totale pour la méthode SJ lorsque p varie (haut), N varie (milieu) ou RA varie (bas). La référence pour le calcul du gain est la méthode COMP. Sur les figures (c) et (e), les points sont reliés verticalement, et sur les figures (d) et (f), ces mêmes points sont reliés horizontalement. Abréviations : Cas test (CT).

F.2 Analyse pour la méthode sans jacobienne

Pour l'analyse en ordre, on considère les figures F.1a et F.1b :

- (A-i) Fig. F.1a : *Gain SJ* croît faiblement avec p lorsque N est petit ($N \leq 2\,888$) et décroît fortement avec p lorsque N est plus grand, quel que soit RA .
- (A-ii) Fig. F.1a : Pour le cas test 2¹, la non-superposition des courbes traduit une forte dépendance du gain en fonction de N ($RA = 1$) quel que soit p . Ceci est confirmé avec les autres cas tests, où à p fixé, *Gain SJ* augmente globalement avec N quel que soit RA .
- (A-iii) Fig. F.1b : *Gain SJ* croît faiblement avec p lorsque RA est petit ($RA \leq 600$) et décroît fortement avec p lorsque RA est plus grand, quel que soit N .
- (A-iv) Fig. F.1b : Pour le cas test 2², la forte superposition des courbes traduit une indépendance du gain en fonction de RA ($N = 2\,888$) quel que soit p . Ceci n'est pas visible avec les autres cas tests, où à p fixé, *Gain SJ* augmente globalement avec RA quel que soit N , et ce à cause de la dépendance simultanée en N et RA pour les maillages tels que $N \neq 2\,888$.
- (A-v) Dès $p \geq 4$, *Gain SJ* converge vers la valeur plateau de 1,64.

Pour l'analyse en maillage, on considère les figures F.1c à F.1f :

- (A-vi) Fig. F.1c & F.1d : Pour $N = 2\,888$ (cas test 2), la très forte superposition des symboles (*cf.* aussi figures F.1e et F.1f) dans les deux mêmes ellipses confirme l'indépendance de *Gain SJ* en fonction de RA quel que soit p (*cf.* (A-iv)).
- (A-vii) Fig. F.1e & F.1f : Pour $RA = 1$ (cas test 2), la non-superposition des symboles (*cf.* aussi figures F.1c et F.1d) dans les deux mêmes ellipses confirme la forte dépendance de *Gain SJ* en fonction de N quel que soit p (*cf.* (A-ii)).
- (A-viii) Fig. F.1c & F.1e : La longueur moyenne des traits verticaux est petite comparée à celle des autres méthodes, confirmant la faible dépendance en p du gain quels que soient N et RA fixés (*cf.* (A-i) et (A-iii)).
- (A-ix) Les points (A-ii) et (A-vii) justifient une analyse de la figure F.1d : *Gain SJ* croît très fortement avec N lorsque p est petit ($p \leq 4$) et croît faiblement pour p plus grand quel que soit RA (*cf.* (A-ii)). Par ailleurs, à N fixé petit ($N \leq 2\,888$), *Gain SJ* augmente globalement avec p , et à N fixé plus grand, il diminue globalement avec p quel que soit RA (*cf.* (A-i)).
- (A-x) Les points (A-iv) et (A-vi) ne justifient pas une analyse de la figure F.1f.

1. On a $N = 648$, $N = 2\,888$, $N = 10\,952$ ou $N = 43\,808$ et donc $RA = 1$ est fixé.

2. On a $RA = 1$, $RA = 4$, $RA = 20$, $RA = 143$ ou $RA = 600$ et donc $N = 2\,888$ est fixé.

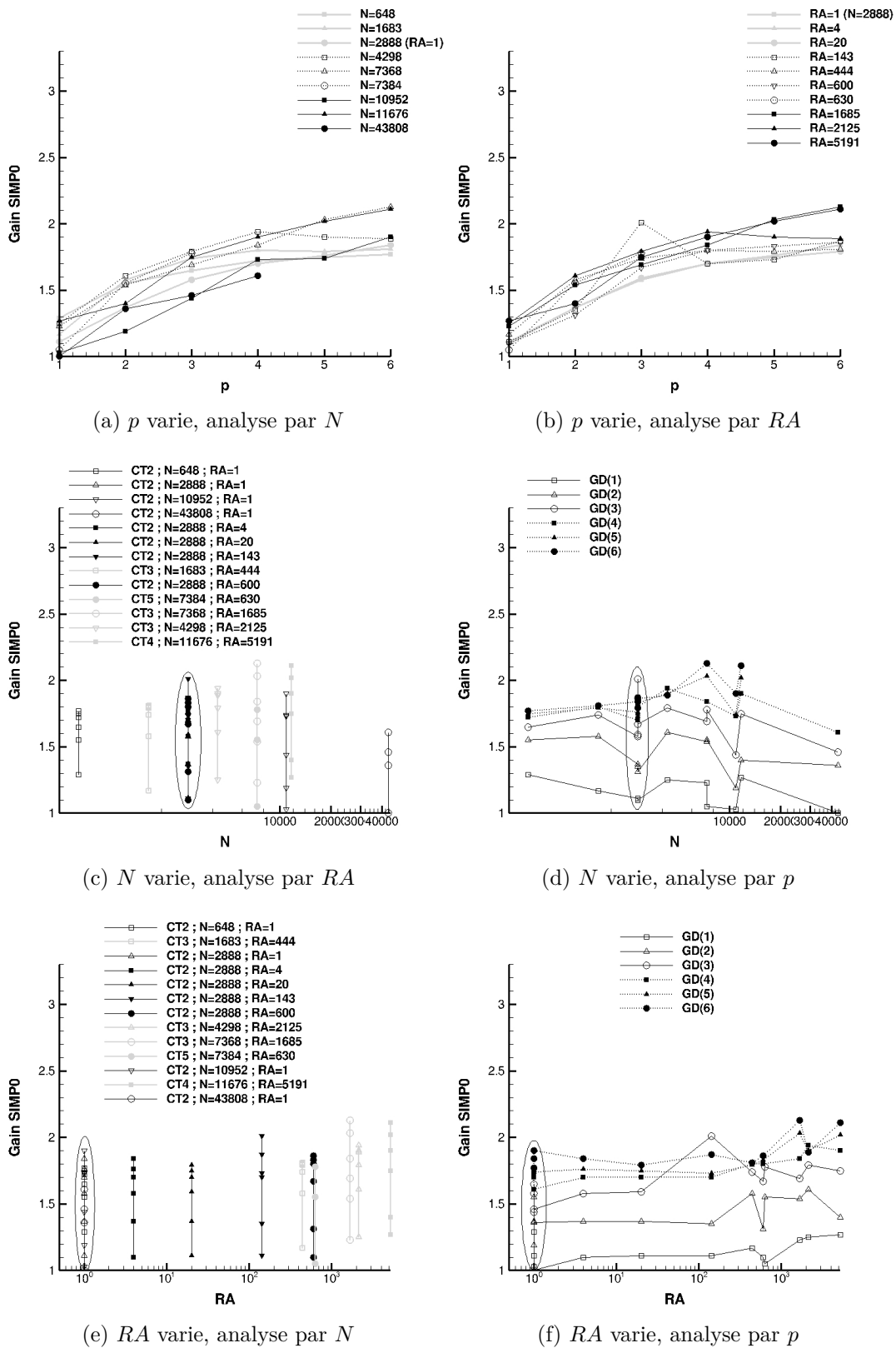


Fig. F.2 – Gain en temps CPU sur la résolution totale pour la méthode SIMPO lorsque p varie (haut), N varie (milieu) ou RA varie (bas). La référence pour le calcul du gain est la méthode COMP. Sur les figures (c) et (e), les points sont reliés verticalement, et sur les figures (d) et (f), ces mêmes points sont reliés horizontalement. Abréviations : Cas test (CT).

F.3 Analyse pour la méthode avec découplage des degrés de liberté

Pour l'analyse en ordre, on considère les figures F.2a et F.2b :

- (B-i) Fig. F.2a : *Gain SIMP0* croît fortement avec p quels que soient N et RA .
- (B-ii) Fig. F.2a : Pour le cas test 2^3 , la non-superposition des courbes traduit une faible dépendance du gain en fonction de N ($RA = 1$) quel que soit p . Ceci est confirmé avec les autres cas tests, où à p fixé, *Gain SIMP0* est sensible à N quel que soit RA mais il ne révèle aucune tendance particulière.
- (B-iii) Fig. F.2b : *Gain SIMP0* croît fortement avec p quels que soient RA et N .
- (B-iv) Fig. F.2b : Pour le cas test 2^4 , la non-superposition des courbes traduit une faible dépendance du gain en fonction de RA ($N = 2\ 888$) quel que soit p . Ceci est confirmé avec les autres cas tests, où à p fixé, *Gain SIMP0* augmente globalement avec RA quel que soit N . Ceci est d'autant plus vrai pour $p \geq 4$.
- (B-v) Dès $p \geq 4$, *Gain SIMP0* converge vers la valeur plateau de 1,89.

Pour l'analyse en maillage, on considère les figures F.2c à F.2f :

- (B-vi) Fig. F.2c & F.2d : Pour $N = 2\ 888$ (cas test 2), la non-superposition des symboles (*cf.* aussi figures F.2e et F.2f) dans les deux mêmes ellipses confirme la faible dépendance de *Gain SIMP0* en fonction de RA quel que soit p (*cf.* (B-iv)).
- (B-vii) Fig. F.2e & F.2f : Pour $RA = 1$ (cas test 2), la non-superposition des symboles (*cf.* aussi figures F.2c et F.2d) dans les deux mêmes ellipses confirme la faible dépendance de *Gain SIMP0* en fonction de N quel que soit p (*cf.* (B-ii)).
- (B-viii) Fig. F.2c & F.2e : La longueur moyenne des traits verticaux est grande comparée à celle des autres méthodes, confirmant la forte dépendance en p du gain quels que soient N et RA fixés (*cf.* (B-i) et (B-v)).
- (B-ix) Les points (B-ii) et (B-vii) justifient une analyse de la figure F.2d : *Gain SIMP0* est sensible à N et il révèle une très faible tendance à décroître de façon très erratique quels que soient p et RA (*cf.* (B-ii)). Par ailleurs, à N fixé, *Gain SIMP0* augmente globalement avec p quel que soit RA (*cf.* (B-i)).
- (B-x) Les points (B-iv) et (B-vi) justifient une analyse de la figure F.2f : *Gain SIMP0* ne révèle pas de tendance particulière en fonction de RA lorsque RA est petit ($RA \leq 600$), et croît faiblement de façon non monotone en fonction de RA lorsque RA est plus grand, quel que soit p (*cf.* (B-vi)). Par ailleurs, à RA fixé, *Gain SIMP0* augmente globalement avec p quel que soit N (*cf.* (B-v)).

3. On a $N = 648$, $N = 2\ 888$, $N = 10\ 952$ ou $N = 43\ 808$ et donc $RA = 1$ est fixé.

4. On a $RA = 1$, $RA = 4$, $RA = 20$, $RA = 143$ ou $RA = 600$ et donc $N = 2\ 888$ est fixé.

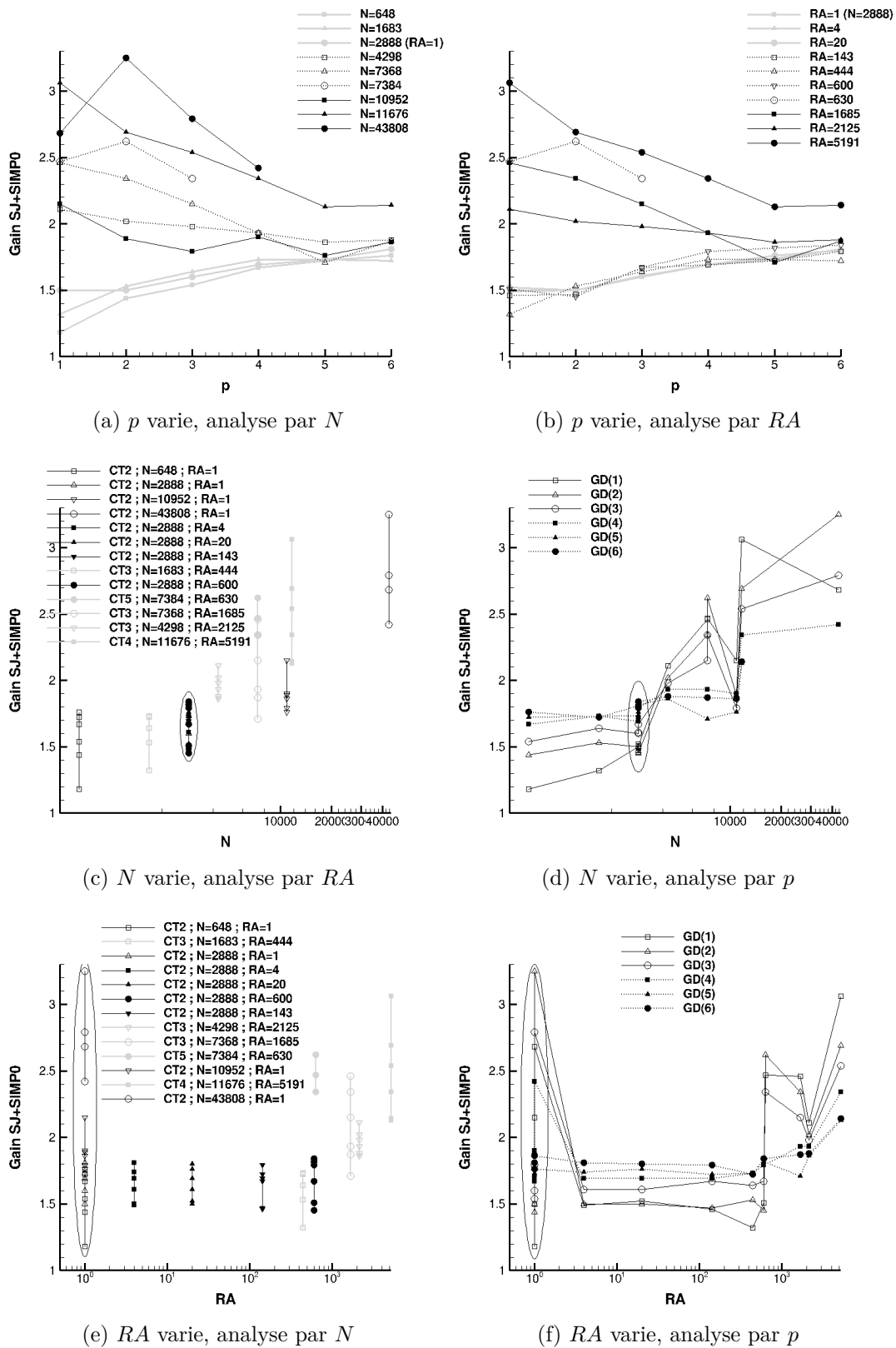


Fig. F.3 – Gain en temps CPU sur la résolution totale pour la méthode SJ+SIMP0 lorsque p varie (haut), N varie (milieu) ou RA varie (bas). La référence pour le calcul du gain est la méthode COMP. Sur les figures (c) et (e), les points sont reliés verticalement, et sur les figures (d) et (f), ces mêmes points sont reliés horizontalement. Abréviations : Cas test (CT).

F.4 Analyse pour la méthode sans jacobienne et avec découplage

Le comportement de $Gain\ SJ + SIMP0$ est un compromis de $Gain\ SJ$ et de $Gain\ SIMP0$.

Pour l'analyse en ordre, on considère les figures F.3a et F.3b :

- (C-i) Comme $Gain\ SJ$, $Gain\ SJ + SIMP0$ croît faiblement avec p lorsque N est petit ($N \leq 2\ 888$) ou lorsque RA est petit ($RA \leq 600$), et décroît fortement avec p lorsque N est plus grand ou lorsque RA est plus grand. Cependant, il est influencé par $Gain\ SIMP0$, la décroissance est alors non monotone lorsque N est grand ($N > 2\ 888$) ou lorsque RA est grand ($RA > 600$) quel que soit p .
- (C-ii) Fig. F.3a : Comme $Gain\ SJ$ et $Gain\ SIMP0$, $Gain\ SJ + SIMP0$ est dépendant de N quel que soit p . En particulier à p fixé, il augmente globalement avec N quel que soit RA , mais de façon plus marquée que $Gain\ SJ$ à cause de la plus grande dépendance en p .
- (C-iii) Fig. F.3b : Comme $Gain\ SJ$ mais influencé par $Gain\ SIMP0$, $Gain\ SJ + SIMP0$ est dépendant de RA quel que soit p . En particulier à p fixé, il augmente avec RA quel que soit N , mais de façon plus marquée que $Gain\ SIMP0$ à cause de la plus grande dépendance en N .
- (C-iv) Dès $p \geq 4$, $Gain\ SJ + SIMP0$ converge vers la valeur plateau de 1, 84.

Pour l'analyse en maillage, on considère les figures F.3c à F.3f :

- (C-v) Fig. F.3c & F.3d : Comme $Gain\ SIMP0$ pour $N = 2\ 888$ (cas test 2), la non-superposition des symboles (*cf.* aussi figures F.3e et F.3f) dans les deux mêmes ellipses confirme la très faible dépendance de $Gain\ SJ + SIMP0$ en fonction de RA quel que soit p (*cf.* (C-iii)).
- (C-vi) Fig. F.3e & F.3f : Comme $Gain\ SJ$ pour $RA = 1$ (cas test 2), la non-superposition des symboles (*cf.* aussi figures F.3c et F.3d) dans les deux mêmes ellipses confirme la très forte dépendance de $Gain\ SJ + SIMP0$ en fonction de N quel que soit p (*cf.* (C-ii)).
- (C-vii) Fig. F.3c & F.3e : La longueur moyenne des traits verticaux se situe entre celle de la méthode SJ et celle de la méthode SIMP0. Ceci confirme la dépendance en p du gain quels que soient N et RA fixés (*cf.* (C-i)).
- (C-viii) Les points (C-ii) et (C-vi) justifient une analyse de la figure F.3d. Cette analyse est identique à celle de $Gain\ SJ$ (*cf.* (A-ix)).
- (C-ix) Les points (C-iii) et (C-v) justifient une analyse de la figure F.3f. Cette analyse est identique à celle de $Gain\ SIMP0$ (*cf.* (B-x)).

Annexe G

Papier & Article

Cette annexe reproduit le papier et l'article écrits grâce aux études menées au cours de ce travail. Ces deux documents portent sur la mise en œuvre de la méthode spatiale GD associée à l'intégration temporelle explicite-implicite simplifiée $SIMP_{p_s}$ basée sur le découplage des degrés de liberté.

Le premier document est le papier de l'AIAA de janvier 2012, il est axé sur l'équation scalaire linéaire d'advection-diffusion en deux dimensions d'espace. En particulier, il mène une analyse de stabilité linéaire du schéma discret retenu.

Le second document est l'article publié dans *Journal of Computational Physics* de juin 2013, il est axé sur le système des équations de Navier-Stokes. Notamment, il réalise une analyse paramétrique des performances du schéma discret retenu.

G.1 Janvier 2012 - Papier AIAA

Référence : Papier AIAA 2012 – 0731, [53]
Auteurs : S. Gérald, F. Renac, C. Marmignon, J. Ryan, F. Coquel
Pages : 1 – 10
Conférence : 50th AIAA Aerospace science meeting
Date : 09 – 12 Janvier 2012

Le contenu de ce papier étant protégé, ce dernier est ici présenté dans sa version de soumission. Le document original reste disponible à l'adresse suivante : <http://arc.aiaa.org/doi/abs/10.2514/6.2012-731>

Fast time implicit-explicit discontinuous Galerkin method for convection-diffusion flow problems

S. Gérald ^{*}, F. Renac [†], C. Marmignon [‡],

Onera - The French Aerospace Lab, F-92322 Châtillon, France

J. Ryan [§],

Onera - The French Aerospace Lab, F-91761 Palaiseau, France

F. Coquel [¶]

CNRS and Ecole Polytechnique, F-91120 Palaiseau, France

An efficient and robust time integration procedure for a high-order discontinuous Galerkin method is introduced for solving unsteady second-order partial differential equations. The time discretization is based on an explicit formulation for the hyperbolic term and an implicit formulation for the parabolic term. The implicit procedure uses a fast iterative algorithm with reduced evaluation cost introduced in [Renac, Marmignon and Coquel, to appear in SIAM J. Sci. Comput.]. The method is here extended to convection dominated flow problems. A second-order discretization in time is achieved by decomposing the integrations of convective and diffusive terms with a splitting method. Numerical examples are presented for the linear convection-diffusion equation in one and two space dimensions. The performance of the present method is seen to be improved in terms of CPU time when compared to a full implicit discretization of the parabolic terms in a wide range of Peclet numbers.

I. Introduction

Discontinuous Galerkin (DG) methods are high-order finite element discretizations and were introduced for the numerical simulation of the first-order hyperbolic neutron transport equation.^{9,10} The method was later extended to nonlinear convection dominated flow problems with the use of a Runge-Kutta method for the time integration.^{5,7} In recent years, there has been a strong interest for these techniques in the field of computational fluid dynamics which has led to the introduction of discretization schemes for parabolic and purely elliptic equations. For more details, the reader is referred to the analysis of existing discretizations in Ref. [2], to the overview of recent progress in DG methods for compressible flows in Ref. [8] and to references cited therein. The success of these methods lies in their high-order of accuracy and flexibility thanks to their high degree of locality. These properties make the DG method well suited to algorithm parallelization, hp -refinement, unstructured meshes, etc.

However, the association of a DG method to an explicit time discretization leads to a strong restriction on the time step due to the so-called Courant-Friedrichs-Levy (CFL) condition for stability of the numerical scheme.^{1,4} Numerical experiments on the resolution of the linear scalar convection equation have shown that for a DG(p) discretization with polynomials of degree p and a $p + 1$ Runge-Kutta method of order $p + 1$, the CFL condition behaves as $\sigma \leq 1/(2p + 1)$ where $\sigma = c\Delta t/h$ denotes the Courant number with $c > 0$ the convection velocity, Δt the time step and h the element size.⁷ This property has been analytically demonstrated for a DG(1) discretization in Ref. [6] when using total variation diminishing Runge-Kutta

^{*}PhD Student, CFD and Aeroacoustics Department DSNA, sophie.gerald@onera.fr.

[†]PhD - Research Engineer, CFD and Aeroacoustics Department DSNA, florent.renac@onera.fr.

[‡]PhD - Research Engineer, CFD and Aeroacoustics Department DSNA, claude.marmignon@onera.fr.

[§]Senior Research Engineer, Applied Aerodynamic Department DTIM, juliette.ryan@onera.fr.

[¶]Research Director, CMAP, frederic.coquel@cmmap.polytechnique.fr.

methods from Ref. [12]. In contrast, the use of a forward Euler method leads to a prohibitive condition for stability $\sigma \leq \mathcal{O}(h^{1/2})$.⁴

Moreover, the discretization of parabolic terms requires an additional stability constraint, the Von Neumann condition, which becomes more restrictive than the CFL condition for flow regions with small Peclet number $Re_h = ch/\nu$, with $\nu > 0$ the diffusion coefficient. In high Reynolds number flows, the large variations in element size required to resolve all the spatial scales make the use of explicit integration techniques prohibitive. On the other hand, the application of fully implicit methods to a system of conservation laws leads to an algebraic system to be solved at each time step. The size of this system is equal to the total number of degrees of freedom (DOF) of the discrete problem and may become very large for practical applications, such as turbulent flow simulations.

A robust and fast time integration procedure for a semi-implicit formulation of a DG discretization of convection-diffusion equations has been introduced in Ref. [11]. The method uses an implicit backward Euler scheme for the viscous terms, while convective terms are treated explicitly. A simplified implicit problem, based on a reduced communication between elements which share a common interface, is used for the resolution of the discrete problem at each time step. This assumption follows from the natural hierarchy of the basis of the discrete function space: the communication between elements is assumed to be dominated by the coupling between DOF associated to the components of the basis of lowest order. After solving an implicit problem of reduced size for the low-order DOF only, a local reconstruction, based on the stencil of the DG discretization, is used for evaluating higher order DOF. Numerical experiments on one-dimensional (1D) and two-dimensional (2D) nonlinear convection-diffusion equations have highlighted better performances of the present method in terms of CPU time when compared to a full implicit discretization of the viscous terms. This simplification also keeps the locality, accuracy and stability properties of the original method. A Von Neumann stability analysis carried out in Ref. [11] for a linear scalar convection-diffusion equation indicated that necessary conditions for stability of the full and simplified methods for $p = 1$ read

$$\sigma Re_h \leq 2, \quad Re_h \leq 6. \quad (1)$$

The first condition corresponds to a restriction on the time step of the form $\Delta t \leq 2\nu/c^2$, while the second condition makes the implicit-explicit time integration inappropriate for convection dominated flow problems. In the present study, we propose to modify the method introduced in Ref. [11] and adapt it to unsteady convection dominated flows in order to remove the stability constraints (1). High-order accuracy in time is achieved through a second-order fast explicit-implicit decomposition method based on a Strang splitting.¹³ The hyperbolic problem is treated explicitly, while the simplified implicit discretization from Ref. [11] is used for the parabolic problem.

II. Model problem and discretization

The discussion in this paper focuses on the discretization of linear scalar convection-diffusion equations with a DG method. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain where d is the space dimension and consider the following problem

$$\partial_t u + \nabla \cdot (\mathbf{c}(\mathbf{x})u) - \nabla \cdot (\mathbf{B}(\mathbf{x})\nabla u) = s(\mathbf{x}, t) \quad \text{in } \Omega, \quad (2)$$

which describes the conservation of the scalar quantity $u(\mathbf{x}, t)$ convected in a velocity field $\mathbf{c}(\mathbf{x}) \in \mathbb{R}^d$ and subject to diffusion modelled by the diffusivity tensor $\mathbf{B}(\mathbf{x}) \in \mathbb{R}^{d \times d}$ which is assumed to be symmetric and positive definite. The forcing function $s(\mathbf{x}, t) \in L^2(\Omega)$ represents production or destruction of the quantity u . The problem is supplemented with the initial condition $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ in Ω , as well as Dirichlet and Neumann boundary conditions $u = u_D$ on Γ_D and $\nabla u \cdot \mathbf{n} = g_N$ on Γ_N with $\partial\Omega = \Gamma_D \cup \Gamma_N$. We assume that $u_D \in L^2(\Gamma_D)$ and $g_N \in L^2(\Gamma_N)$. By \mathbf{n} we denote the unit outward normal vector to $\partial\Omega$.

II.A. Discontinuous Galerkin formulation

The aim of the present work is to derive a fast time marching method for the numerical resolution of problem (2) with a DG space discretization. The domain Ω is partitioned into a shape-regular mesh Ω_h consisting of N nonoverlapping and nonempty simplices κ of characteristic size $h := \min\{\text{diam}(\kappa), \kappa \in \Omega_h\}$ where $\text{diam}(\kappa)$ is the diameter of largest ball included in κ . We define the set of interior faces \mathcal{E}_i and boundary

faces \mathcal{E}_b and further define $\mathcal{E}_h = \mathcal{E}_i \cup \mathcal{E}_b$ the union of all open $(d-1)$ -dimensional faces in Ω_h . The set of boundary faces may be divided into $\mathcal{E}_b = \mathcal{E}_D \cup \mathcal{E}_N$ consisting of faces of Γ_D and Γ_N , respectively.

We look for approximate solutions in the function space of discontinuous polynomials $\mathcal{V}_h^p = \{\varphi \in L^2(\Omega_h) : \varphi|_\kappa \circ F_\kappa \in \mathcal{P}_p(I^d), \kappa \in \Omega_h\}$, where $\mathcal{P}_p(I^d)$ denotes the space of functions formed by tensor products of polynomials and total degree at most p on the master element I^d where $I = [-1, 1]$. Each physical element κ is the image of I^d through the mapping F_κ . The numerical solution of problem (2) is sought under the form

$$u_h(\mathbf{x}, t) = \sum_{l=1}^{N_p} \phi^l(\mathbf{x}) U_\kappa^l(t), \quad \forall \mathbf{x} \in \kappa, \quad \kappa \in \Omega_h, \quad \forall t \geq 0, \quad (3)$$

where U_κ^l are the DOF in the element κ . The subset $(\phi^1, \dots, \phi^{N_p})$ constitutes a hierarchical and orthogonal modal basis of \mathcal{V}_h^p and $N_p = \prod_{i=1}^d (p+i)/i$ is its dimension.

The face-based formulation of the BR2 discretization³ for problem (2) reads: find u_h in \mathcal{V}_h^p such that for all v_h in \mathcal{V}_h^p we have

$$\int_{\Omega_h} v_h \partial_t u_h d\mathbf{x} + \mathcal{L}_c(u_h, v_h) + \mathcal{L}_v(u_h, v_h) + \mathcal{L}_s(v_h) = 0. \quad (4)$$

Hereafter, we will use the following notations $\{v_h\} = (v_h^+ + v_h^-)/2$ and $[[v_h]] = v_h^+ - v_h^-$ which denote the average and jump operators defined for a given interface e . Here, v_h^+ and v_h^- are the traces of v_h on the interface e taken from within the interior of the element κ^+ and the interior of the neighbouring element κ^- , respectively (see Figure 1).

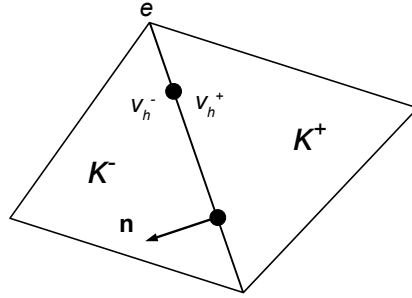


Figure 1. Inner and exterior elements κ^+ and κ^- and definitions of traces v_h^\pm on the interface e and of the unit outward normal vector \mathbf{n} .

The discretization of the hyperbolic term in (4) reads

$$\mathcal{L}_c(u_h, v_h) = - \int_{\Omega_h} u_h \mathbf{c} \cdot \nabla v_h d\mathbf{x} + \int_{\mathcal{E}_i} [[v_h]] h_c(u_h^+, u_h^-, \mathbf{n}) ds + \int_{\mathcal{E}_N} v_h^+ u_h^+ \mathbf{c} \cdot \mathbf{n} ds, \quad (5)$$

where \mathbf{n} denotes the unit outward normal vector to the element κ^+ . We use a standard upwind discretization flux as numerical convective flux: $h_c(u_h^+, u_h^-, \mathbf{n}) = \{u_h \mathbf{c}\} \cdot \mathbf{n} + \frac{|\mathbf{c} \cdot \mathbf{n}|}{2} [[u_h]]$, see for instance.⁷ The bilinear and linear forms in (4) are defined by

$$\begin{aligned} \mathcal{L}_v(u_h, v_h) &= \int_{\Omega_h} \mathbf{B}(\nabla u_h + \mathbf{R}_h) \cdot \nabla v_h d\mathbf{x} - \int_{\mathcal{E}_i} [[v_h]] \{\mathbf{B}(\nabla_h u_h + \mathbf{r}_h^e)\} \cdot \mathbf{n} ds \\ &\quad - \int_{\mathcal{E}_D} v_h^+ \mathbf{B}(\nabla_h u_h^+ + \mathbf{r}_h^{e+}) \cdot \mathbf{n} ds, \end{aligned} \quad (6)$$

and

$$\mathcal{L}_s(v_h) = - \int_{\Omega_h} v_h s(\mathbf{x}) d\mathbf{x} + \int_{\mathcal{E}_D} v_h^+ u_D \mathbf{c} \cdot \mathbf{n} ds - \int_{\mathcal{E}_N} v_h^+ g_N \mathbf{B} ds. \quad (7)$$

The so-called global lifting operator \mathbf{R}_h is weakly defined as the sum of local lifting operators \mathbf{r}_h^e that have support on elements adjacent to e and are solutions of the following problem: for all v_h in \mathcal{V}_h^p we have $\int_{\kappa^+ \cup \kappa^-} v_h \mathbf{r}_h^e d\mathbf{x} = - \int_e \{v_h\} \llbracket u_h \rrbracket \mathbf{n} ds$ on internal faces and $\int_{\kappa^+} v_h \mathbf{r}_h^e d\mathbf{x} = - \int_e v_h^+ (u_h^+ - u_D) \mathbf{n} ds$ on boundary faces. We refer to [3] for further details on the BR2 method.

II.B. Time discretization

The semi-discrete equation (4) is discretized in time by means of a second-order Strang splitting method.¹³ The aim is to solve the explicit and implicit integrations separately. Introducing a time step Δt and defining $t^{(n)} = n\Delta t$, the solution is advanced in time in three substeps on the interval $(t^{(n)}, t^{(n+1)})$. First, the hyperbolic problem is solved for u_h^c on the time interval $(t^{(n)}, t^{(n)} + \frac{\Delta t}{2})$:

$$\int_{\Omega_h} v_h \partial_t u_h^c d\mathbf{x} + \mathcal{L}_c(u_h^c, v_h) = 0, \quad \forall t \in (t^{(n)}, t^{(n)} + \frac{\Delta t}{2}], \quad \forall v_h \in \mathcal{V}_h^p, \quad (8)$$

with initial condition $u_h^c(\cdot, t^{(n)}) = u_h(\cdot, t^{(n)})$. Then, the parabolic problem with source term is solved for u_h^v on the time interval $(t^{(n)}, t^{(n)} + \Delta t]$:

$$\int_{\Omega_h} v_h \partial_t u_h^v d\mathbf{x} + \mathcal{L}_v(u_h^v, v_h) + \mathcal{L}_s(v_h) = 0, \quad \forall t \in (t^{(n)}, t^{(n)} + \Delta t], \quad \forall v_h \in \mathcal{V}_h^p, \quad (9)$$

with $u_h^v(\cdot, t^{(n)}) = u_h^c(\cdot, t^{(n)} + \frac{\Delta t}{2})$. Finally, the hyperbolic problem is solved on the time interval $(t^{(n)} + \frac{\Delta t}{2}, t^{(n)} + \Delta t]$:

$$\int_{\Omega_h} v_h \partial_t u_h^c d\mathbf{x} + \mathcal{L}_c(u_h^c, v_h) = 0, \quad \forall t \in (t^{(n)} + \frac{\Delta t}{2}, t^{(n)} + \Delta t], \quad \forall v_h \in \mathcal{V}_h^p, \quad (10)$$

with $u_h^c(\cdot, t^{(n)} + \frac{\Delta t}{2}) = u_h^v(\cdot, t^{(n)} + \Delta t)$. The solution of equation (4) at time $t^{(n+1)}$ is thus $u_h(\cdot, t^{(n+1)}) = u_h^c(\cdot, t^{(n)} + \Delta t)$.

We use a Heun's method for the time discretization of hyperbolic problems (8) and (10). This method is an explicit two-stage Runge-Kutta scheme and is formally second-order accurate in time.¹² For example for (8), the scheme reads as follow

$$\int_{\Omega_h} \frac{u_h^{c,*} - u_h^{(n)}}{\Delta t/2} v_h d\mathbf{x} = -\mathcal{L}_c(u_h^{(n)}, v_h), \quad \forall v_h \in \mathcal{V}_h^p, \quad (11a)$$

$$\int_{\Omega_h} \frac{u_h^{c,(n+\frac{1}{2})} - u_h^{(n)}}{\Delta t/2} v_h d\mathbf{x} = -\frac{1}{2} \left(\mathcal{L}_c(u_h^{(n)}, v_h) + \mathcal{L}_c(u_h^{c,*}, v_h) \right), \quad \forall v_h \in \mathcal{V}_h^p, \quad (11b)$$

where $u_h^{c,*}$ is the predictive solution at time $t^{(n)} + \frac{\Delta t}{2}$, $u_h^{c,(n+\frac{1}{2})}$ is the approximated solution at time $t^{(n)} + \frac{\Delta t}{2}$ and $u_h^{(n)} = u_h(\cdot, t^{(n)})$. Thus the discret problem for (8) reduces to the following linear system at each time step

$$\mathbf{U}_c^* = \mathbf{U}^{(n)} - \frac{\Delta t}{2} \mathbf{M}^{-1} \mathbf{L}_c(\mathbf{U}^{(n)}), \quad (12a)$$

$$\mathbf{U}_c^{(n+\frac{1}{2})} = \frac{1}{2} \left(\mathbf{U}_c^* + \mathbf{U}^{(n)} - \frac{\Delta t}{2} \mathbf{M}^{-1} \mathbf{L}_c(\mathbf{U}_c^*) \right), \quad (12b)$$

with $\mathbf{U} \in \mathbb{R}^{N_{DOF}}$ the vector of DOF and $N_{DOF} = N \times N_p$, $\mathbf{L}_c(\mathbf{U})$ the residual vector defined by the discrete form (5) evaluated from \mathbf{U} , and $\mathbf{M} \in \mathbb{R}^{N_{DOF} \times N_{DOF}}$ denotes the diagonal mass matrix.

Problem (9) is discretized in time with an implicit second-order Crank-Nicolson method. The approximation of the temporal derivative in (9) leads to

$$\int_{\Omega_h} \frac{u_h^{v,(n+1)} - u_h^{v,(n+\frac{1}{2})}}{\Delta t} v_h d\mathbf{x} = -\frac{1}{2} \left(\mathcal{L}_v(u_h^{v,(n+1)}, v_h) + \mathcal{L}_v(u_h^{v,(n+\frac{1}{2})}, v_h) + 2\mathcal{L}_s(v_h) \right), \quad \forall v_h \in \mathcal{V}_h^p, \quad (13)$$

where $u_h^{v,(n+1)} = u_h^v(\cdot, t(n) + \Delta t)$. The following linear system appears at each time step

$$\mathbf{A}(\mathbf{U}_v^{(n+1)} - \mathbf{U}_c^{(n+\frac{1}{2})}) = -\mathbf{L}_v(\mathbf{U}_c^{(n+\frac{1}{2})}) - \mathbf{L}_s, \quad (14)$$

with $\mathbf{L}_v(\mathbf{U}_c^{(n+\frac{1}{2})}) = \mathbf{L}_v \mathbf{U}_c^{(n+\frac{1}{2})}$ and \mathbf{L}_s the residual vectors defined by the discrete forms (6) and (7). The so-called implicit matrix is defined by $\mathbf{A} = \frac{1}{\Delta t} \mathbf{M} + \frac{1}{2} \mathbf{L}_v$ and is an unsymmetric real matrix of size N_{DOF} since \mathbf{L}_v is. We use a preconditioned biconjugate gradient method for solving the linear system (14) at each time step. The inversion cost of \mathbf{A} may become prohibitive in practical applications where N is large and where accuracy is looked for. To circumvent this shortcoming, we propose to use the modification of the implicit problem (14) introduced in Ref. [11] for simplifying its resolution. This is the subject of the next paragraph.

II.C. Simplification of the implicit procedure

The simplified algorithm for the time integration is here briefly introduced. The resolution of the linear system is simplified by reducing the coupling between modes in adjacent elements. Using the definition (3) of the numerical solution u_h , we denote by mode of order l the component of u_h along the l -th basis element of \mathcal{V}_h^p , *i.e.* $U_\kappa^l(t)\phi^l(\mathbf{x})$. For the sake of clarity, the method is now introduced for a 1D example with a uniform partition of the segment $[0, 1]$ and for polynomials of degree $p = 2$. Thus, in a given element κ^+ , the elementwise numerical solution (3) is a sum of three modes: the lowest order mode is the constant mode $U_{\kappa^+}^0\phi^0$ and higher order modes correspond to the linear mode $U_{\kappa^+}^1\phi^1$ and quadratic mode $U_{\kappa^+}^2\phi^2$.

Every mode in κ^+ is linked to all modes in the neighbouring element κ^- through the numerical fluxes. Here, we assume that the coupling between low-order modes dominates and we choose to neglect coupling with higher order modes in neighbouring elements. For that purpose, let p_s be an integer that separates low-order modes $0 \leq q \leq p_s$ and high-order modes $p_s < q \leq p$. We neglect the coupling between modes in κ^+ and high-order modes in κ^- . For instance, for $p_s = 0$ the coupling between every mode in κ^+ and linear and quadratic modes in κ^- is neglected.

The implicit matrix has a tridiagonal block structure that may be written in a compact form as $\mathbf{A} = (\mathbf{A}_{j-1} | \mathbf{A}_j | \mathbf{A}_{j+1})$, where \mathbf{A}_j and $\mathbf{A}_{j\pm 1}$ are respectively diagonal and off-diagonal blocks of size N_p . The simplification consists in replacing the matrix \mathbf{A} by

$$\mathbf{A}^* = \left(\begin{array}{c|cc} \tilde{\mathbf{A}}_{j-1} & 0 & \tilde{\mathbf{A}}_j & \mathbf{B}_j & \tilde{\mathbf{A}}_{j+1} & 0 \\ \mathbf{C}_{j-1} & 0 & \mathbf{C}_j & \hat{\mathbf{A}}_j & \mathbf{C}_{j+1} & 0 \end{array} \right), \quad (15)$$

where $\tilde{\mathbf{A}}_j \in \mathbb{R}^{N_{p_s} \times N_{p_s}}$, $\hat{\mathbf{A}}_j \in \mathbb{R}^{(N_p - N_{p_s}) \times (N_p - N_{p_s})}$ and $N_{p_s} = \sum_{i=1}^d (p_s + i)/i$ denotes the number of low-order modes. The linear system (12) at a physical time step is then replaced by a two-stage algorithm.

First, an implicit problem for the low-order modes is solved by expliciting terms accounting for higher order modes:

$$\tilde{\mathbf{A}} \Delta \tilde{\mathbf{U}}^{(n+1)} = -\tilde{\mathbf{R}}(\mathbf{U}_c^{(n+\frac{1}{2})}) - \mathbf{B} \Delta \hat{\mathbf{U}}^{(n)}, \quad (16)$$

where $\tilde{\mathbf{A}} = (\tilde{\mathbf{A}}_{j-1} | \tilde{\mathbf{A}}_j | \tilde{\mathbf{A}}_{j+1})$ is a square matrix of size $N_{DOF_s} = N \times N_{p_s}$, $\Delta \tilde{\mathbf{U}}^{(n+1)} = \tilde{\mathbf{U}}_v^{(n+1)} - \tilde{\mathbf{U}}_c^{(n+1/2)}$ and $\Delta \hat{\mathbf{U}}^{(n)} = \hat{\mathbf{U}}_v^{(n)} - \hat{\mathbf{U}}_c^{(n-1/2)}$ where $\tilde{\mathbf{U}}$ and $\hat{\mathbf{U}}$ contain respectively low-order and high-order DOF. By $\tilde{\mathbf{R}}(\mathbf{U}_c^{(n+1/2)}) = -\tilde{\mathbf{L}}_v(\mathbf{U}_c^{(n+1/2)}) - \tilde{\mathbf{L}}_s$ we denote the residuals associated to low-order modes and we set $\mathbf{B} = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_N)$.

Second, higher order modes are reconstructed from the low-order modes and precedent time step solution. We note $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_{j-1} | \hat{\mathbf{A}}_j | \hat{\mathbf{A}}_{j+1})$ the square matrix of size $N_{DOF} - N_{DOF_s}$ and let $\hat{\mathbf{A}} = \hat{\mathbf{L}}_o + \hat{\mathbf{D}} + \hat{\mathbf{U}}_p$ be its decomposition into strictly lower, diagonal and strictly upper parts, respectively. The reconstruction of higher modes reads as follows

$$(\hat{\mathbf{L}}_o + \hat{\mathbf{D}}) \Delta \hat{\mathbf{U}}^{(n+1)} = -\hat{\mathbf{R}}(\mathbf{U}_c^{(n+1/2)}) - \mathbf{C} \Delta \tilde{\mathbf{U}}^{(n+1)} - \hat{\mathbf{U}}_p \Delta \hat{\mathbf{U}}^{(n)}. \quad (17)$$

with $\mathbf{C} = (\mathbf{C}_{j-1} | \mathbf{C}_j | \mathbf{C}_{j+1})$. In the following, the full implicit method (14) will be called the FULL method, and the simplified one (16,17) will be called the SIMP or SIMP p_s method.

Previous investigations¹¹ in the context of nonlinear scalar convection-diffusion equations have shown that the SIMP method allows a saving of $\mathcal{O}(N_p^2/N_{p_s}^2)$ flops compared to the FULL method when using an iterative method for the resolution of the linear system (14) at each time step. Likewise, a Von Neumann

analysis for the DG discretization of the linear scalar convection-diffusion equation with linear polynomials indicates that both schemes present the same stability constraints. In particular, necessary conditions (1) for stability have been demonstrated. We now compare the properties and performances of both methods for convection dominated flow problems in the next sections.

III. Von Neumann analysis

A Von Neumann analysis is here performed in order to derive practical stability constraints on the numerical scheme parameters. For the sake of simplicity, we consider the 1D linear convection-diffusion equation

$$\partial_t u + c\partial_x u - \nu\partial_{xx} u = 0, \quad \text{in } (0, 1] \times (0, +\infty) \quad (18)$$

with $\nu > 0$, $c > 0$, and a periodic boundary condition $u(0, t) = u(1, t)$ for all $t \geq 0$. Equation (18) admits Fourier modes of the form $u(x, t) = e^{i(kx - \omega t)}$ as exact solutions where k is the prescribed real wavenumber and ω is the corresponding complex frequency satisfying the dispersion relation $\omega = kc - i\nu k^2$ with $i^2 = -1$. A similar decomposition is performed for the numerical solution u_h . The domain is first divided into regular elements of size $h = 1/N$. Introducing the reduced wavenumber $\theta = kh$, the condition for stability that any error should remain uniformly bounded for $n \rightarrow \infty$ at fixed time step Δt can be written as $\rho(\mathbf{G}) \leq 1$ for all θ in $(-\pi, \pi]$, where $\rho(\mathbf{G})$ denotes the spectral radius of the amplification matrix between times $t^{(n)}$ and $t^{(n+1)}$.

The eigenspectra of amplification matrices for numerical schemes of the FULL and SIMP0 methods are plotted in Figures 2 and 3 for $p = 1$ and $p = 2$, respectively. Results are obtained for different values of the Courant and Peclet numbers. Both methods present the same stability properties. Indeed, a stability analysis in the low wavenumber limit $\theta = 0$ shows that a necessary condition for stability of both methods for $p = 1$ reads

$$\left| \frac{Re_h - 6\sigma}{Re_h + 6\sigma} \right| \left(1 - 3\sigma + \frac{9\sigma^2}{2} \right)^2 \leq 1. \quad (19)$$

In the convective limit $Re_h \rightarrow \infty$, one recovers the stability condition $\sigma \leq 2/3$ for stability of the Heun's method (12) on half a time step.⁶ Moreover, equation (19) shows that diffusion is seen to slightly enhance the stability of the numerical scheme in agreement with the analysis at low Peclet number in Figures 2c,d. Analytical stability conditions for $p = 2$ are still an open problem but Figure 3 clearly shows that the SIMP method does not alter the spectral properties of the numerical scheme in contrast to what was observed¹¹ in the limit $Re_h \ll 1$.

IV. Numerical experiments

In this section we present numerical experiments in 1D and 2D to illustrate the performances of the simplification of the implicit time discretization.

IV.A. One-dimensional unsteady convection-diffusion equation

In this example we let $\Omega = [0, 1]$ and consider the following problem

$$\partial_t u + c\partial_x u - \nu\partial_{xx} u = 0 \quad \text{in } (0, 1] \times (0, T], \quad (20a)$$

$$u(x, 0) = \sin(2\pi x) \quad \text{in } [0, 1], \quad (20b)$$

with periodic boundary conditions, $c = 1$ and $\nu > 0$. The exact solution reads $u(x, t) = \sin(2\pi(x - ct))e^{-4\pi^2\nu t}$. The problem (20) is solved by using the method introduced in § II. The domain is discretized with a uniform grid $\Omega_h = \cup_{j=1}^N [(j-1)h, jh]$ with $h = 1/N$. We first investigate the asymptotic accuracy of the DG discretization under time step refinement. Table 1 presents the L^2 - and L^∞ -norms of the error $e_h \equiv u - u_h$ as a function of the nondimensional time step σ . To isolate the temporal convergence behaviour as much as possible, effects of spatial discretization errors are minimized by using a sufficiently refined mesh with $N = 640$ elements and a high polynomial degree $p = 3$. The accuracy of the method under space refinement

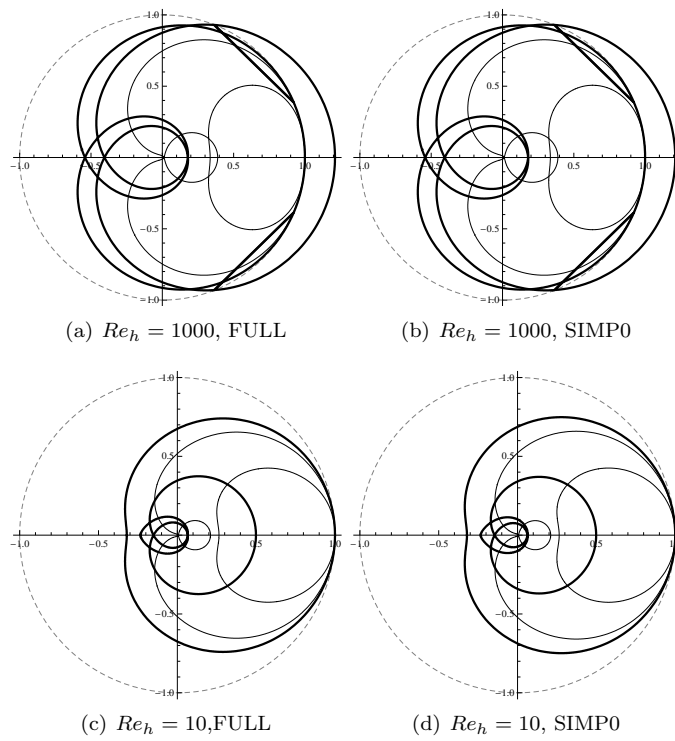


Figure 2. Amplification matrix eigenspectra in the complex plane for $p = 1$ and $\sigma = 0.2$ (thin lines), $\sigma = 0.5$ (normal lines) and $\sigma = 0.7$ (thick lines). The unit circle $\mathcal{C}(0, 1)$ in dashed grey denotes the stability domain.

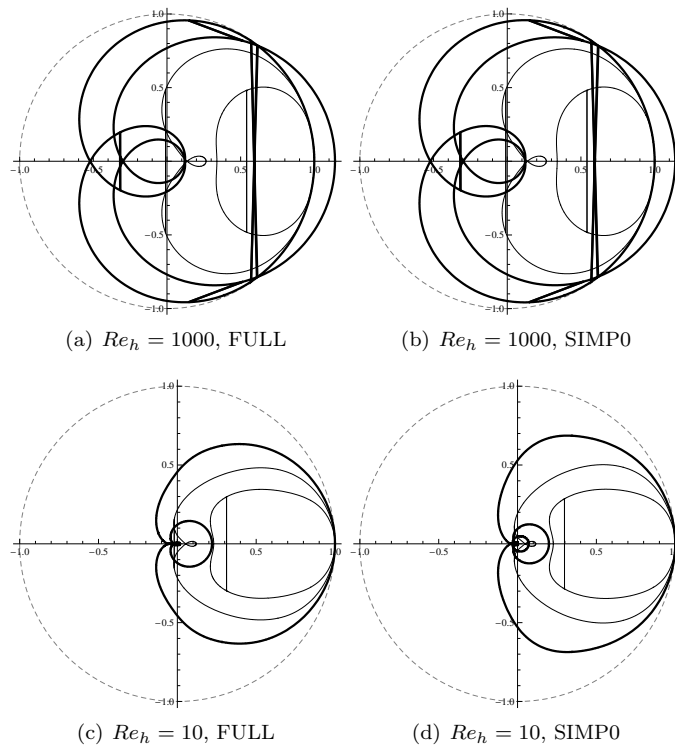


Figure 3. Amplification matrix eigenspectra in the complex plane for $p = 2$ with $\sigma = 0.1$ (thin lines), $\sigma = 0.2$ (normal lines) and $\sigma = 0.35$ (thick lines). The unit circle $\mathcal{C}(0, 1)$ in dashed grey denotes the stability domain.

has been validated in Ref. [11] and was seen to be $\|e_h\|_2 \sim h^{p+1}$ in agreement with previous theoretical study.² As expected, both schemes achieve optimal convergence $\|e_h\|_2 \sim \sigma^2$ as σ tends to zero. Moreover, error levels are identical between both methods indicating that the implicit simplification does not alter the error of the numerical scheme. This property was seen to hold for all experiments presented throughout this study so that the SIMP method may be used for the resolution of unsteady flow problems.

Table 1. 1D problem: convergence of the FULL and SIMP methods under time step refinement and orders of accuracy. Results were obtained with $T = 0.25$, $N = 640$, $Re_h = 100$, $p = 3$ and $p_s = 0$.

| | FULL method | | SIMP0 method | |
|----------|----------------------|-------|----------------------|-------|
| σ | $\ u - u_h\ _2$ | order | $\ u - u_h\ _2$ | order |
| 0.2 | 1.78396E - 07 | – | 1.78395E - 07 | – |
| 0.1 | 4.45991E - 08 | 2.00 | 4.46027E - 08 | 2.00 |
| 0.05 | 1.11497E - 08 | 2.00 | 1.11507E - 08 | 2.00 |
| 0.025 | 2.78740E - 09 | 2.00 | 2.78765E - 09 | 2.00 |
| 0.0125 | 6.96821E - 10 | 2.00 | 6.96888E - 10 | 2.00 |
| σ | $\ u - u_h\ _\infty$ | order | $\ u - u_h\ _\infty$ | order |
| 0.2 | 2.52290E - 07 | – | 2.52348E - 07 | – |
| 0.1 | 6.30726E - 08 | 2.00 | 6.30867E - 08 | 2.00 |
| 0.05 | 1.57681E - 08 | 2.00 | 1.57737E - 08 | 2.00 |
| 0.025 | 3.94199E - 09 | 2.00 | 3.94443E - 09 | 2.00 |
| 0.0125 | 9.85454E - 10 | 2.00 | 9.86581E - 10 | 2.00 |

In order to evaluate the relative efficiency of FULL and SIMP methods, we define the speedup as the ratio of CPU time needed by the FULL method to CPU time with the SIMP method:

$$\text{speedup} = \frac{\text{CPU time(FULL)}}{\text{CPU time(SIMP)}}. \quad (21)$$

Results are given in Table 2. The value of p_s that separates low- and high-order DOF is the lowest allowing to recover the same error levels as the FULL method. This choice ensures that a fair comparison between CPU time for all configurations can be made. We note a slight increase of the speedup values as p is increased in agreement with previous experiments for low Peclet number flows¹¹ $Re_h \leq 1$ where speedups were nevertheless seen to reach larger values in the limit $Re_h \ll 1$. However, the present method allows CPU time savings on a wide range of Peclet numbers and speedups are almost independent of Re_h .

Table 2. 1D problem: speedups observed from experiments with $T = 0.25$ and $N = 640$.

| | | $Re_h = 10$ | | $Re_h = 100$ | | $Re_h = 1000$ | | $Re_h = 10000$ | |
|-----|----------|-------------|---------|--------------|---------|---------------|---------|----------------|---------|
| p | σ | p_s | speedup | p_s | speedup | p_s | speedup | p_s | speedup |
| 1 | 0.6 | 0 | 1.36 | 0 | 1.23 | 0 | 1.28 | 0 | 1.38 |
| 2 | 0.3 | 0 | 1.36 | 0 | 1.34 | 0 | 1.40 | 0 | 1.36 |
| 3 | 0.2 | 0 | 1.56 | 0 | 1.41 | 0 | 1.33 | 0 | 1.45 |
| 4 | 0.1 | 1 | 1.47 | 0 | 1.42 | 0 | 1.30 | 0 | 1.37 |
| 5 | 0.1 | 1 | 1.71 | 0 | 1.48 | 0 | 1.39 | 0 | 1.34 |

IV.B. Two-dimensional unsteady convection-diffusion equation

In this second example, we consider the following problem with periodic boundary conditions in $\Omega = [0, 1]^2$:

$$\partial_t u + \nabla \cdot (\mathbf{c}u) - \nabla \cdot (\mathbf{B}(\mathbf{x})\nabla u) = s(\mathbf{x}, t) \quad \text{in } \Omega \times (0, T], \quad (22a)$$

$$u(\mathbf{x}, 0) = \frac{1}{4} e^{-\frac{(\mathbf{x} - \mathbf{x}_0(0))^2}{R^2}} \quad \text{in } \Omega, \quad (22b)$$

with longitudinal advection $\mathbf{c} = (1, 0)^\top$ and a diffusion tensor of the form $\mathbf{B}(\mathbf{x}) = \frac{\nu}{4}(3 + \cos \pi y)\mathbf{I}$ with $\nu > 0$. The initial condition is a Gaussian pulse with $R = 1/7$ and we choose s such that the exact solution of (22) reads $u(\mathbf{x}, t) = \frac{1}{4} \exp((\mathbf{x} - \mathbf{x}_0(t))^2/R^2 - 4\pi\nu t)$ with $\mathbf{x}_0(t) = (1/2 + ct, 1/2)^\top$.

Table 3. 2D problem: speedups and error levels observed from experiments with $T = 0.25$.

| p | N | p_s | Re_h | σ | speedup | $\ u - u_h\ _2$ | |
|-----|------|-------|--------|----------|---------|-----------------|-----------------|
| | | | | | | FULL | SIMP |
| 1 | 3200 | 0 | 100 | 0.2 | 1.61 | $1.80352E - 04$ | $1.80351E - 04$ |
| 2 | 3200 | 0 | 100 | 0.2 | 2.46 | $9.91511E - 06$ | $9.92886E - 06$ |
| 3 | 3200 | 0 | 100 | 0.2 | 3.32 | $7.57723E - 06$ | $7.57993E - 06$ |
| 4 | 3200 | 0 | 100 | 0.2 | 4.06 | $7.52525E - 06$ | $7.53292E - 06$ |
| 5 | 3200 | 1 | 100 | 0.2 | 7.10 | $7.22027E - 06$ | $7.32333E - 06$ |
| 3 | 3200 | 0 | 10 | 0.2 | 6.16 | $5.83359E - 05$ | $5.97768E - 05$ |
| 3 | 3200 | 0 | 100 | 0.2 | 3.32 | $7.57723E - 06$ | $7.57993E - 06$ |
| 3 | 3200 | 0 | 1000 | 0.2 | 2.53 | $5.16744E - 06$ | $5.16745E - 06$ |
| 3 | 3200 | 0 | 10000 | 0.2 | 2.23 | $5.28404E - 06$ | $5.28404E - 06$ |
| 3 | 200 | 0 | 100 | 0.2 | 3.38 | $8.70333E - 05$ | $8.71792E - 05$ |
| 3 | 800 | 0 | 100 | 0.2 | 3.43 | $1.97619E - 05$ | $1.97785E - 05$ |
| 3 | 1600 | 0 | 100 | 0.2 | 3.41 | $1.09399E - 05$ | $1.09451E - 05$ |
| 3 | 3200 | 0 | 100 | 0.2 | 3.32 | $7.57723E - 06$ | $7.57993E - 06$ |
| 3 | 5000 | 0 | 100 | 0.2 | 3.38 | $5.82650E - 06$ | $5.82831E - 06$ |

For the space discretization, we use partitions of the domain built from uniform grids of m -by- m squares split into a total of $N = 2m^2$ isosceles triangles. Table 3 presents the speedup values obtained when varying the polynomial degree, the Peclet number value and the number of elements. Some results are qualitatively similar to those obtained with the 1D problem in § IV.A. The simplified method keeps the accuracy of the numerical scheme and the efficiency of the SIMP method increases with p . Likewise, the speedup is seen to be almost independent of the mesh size and to slightly decrease with Re_h . However, we note that the speedup values are larger than those obtained in the 1D problem and allow significant CPU time savings. Indeed, the computational cost of the SIMP method is less expensive than the FULL method by a theoretical factor $(p/p_s)^{2d}$ for large N and p .¹¹

V. Concluding remarks

An efficient and robust time integration procedure has been successfully applied to linear unsteady convection-diffusion equations. We have used an explicit-implicit time discretization in which the convective term is treated explicitly and the diffusive term implicitly. Integration in time is achieved by applying a Strang splitting method. For the implicit integration we use an iterative algorithm with the simplified method introduced in Ref. [11] and extend it to convection dominated flows. The original method is indeed limited by stability constraints (1) for $p = 1$. The use of a splitting in time of integrations of convective and diffusive terms allows to remove the second condition in (1). When using a second-order accurate integration in time, the first condition in (1) was seen to be converted into a constraint on the CFL similar to what was obtained in the limit of pure convection in Ref. [6]. A Von Neumann analysis shows that the simplification does not alter the stability properties of the numerical scheme for polynomial degrees $p \leq 2$ and numerical experiments support this analysis for larger p values. In terms of CPU time, numerical experiments on 1D and 2D problems demonstrate an acceleration of convergence of the numerical solution over a wide range of polynomial degree, mesh size and Peclet number. The method becomes faster for large number of elements and high polynomial degree. Moreover, it allows significant speedups for 2D problems, thereby confirming that the method is well suited to large-scale problems. The extension of the method to systems of conservation laws is straightforward but its efficiency must yet be analyzed. For future works, we will implement this method to the compressible Navier-Stokes equations.

References

- ¹J.-B. Apoung Kanga & B. Després, *CFL condition and boundary conditions for DGM approximation of convection-diffusion*, SIAM J. Numer. Anal., **44** (2006) 2245–2269.
- ²D. N. Arnold, F. Brezzi, B. Cockburn & L. D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., **39** (2002) 1749–1779.
- ³F. Bassi, S. Rebay, G. Mariotti, S. Pedinotti & M. Savini, *A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows*, In proceedings of the 2nd European Conference on Turbomachinery Fluid Dynamics and Thermodynamics, R. Decuyper & G. Dibelius (eds.), Antwerpen, Belgium, (1997).
- ⁴G. Chavent & B. Cockburn, *The local projection P^0 - P^1 discontinuous Galerkin finite element method for scalar conservation laws*, *M²AN Math. Model. Anal. Numer.*, **23** (1989) 565–592.
- ⁵B. Cockburn & C.-W. Shu, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for scalar conservation laws II: General framework*, *Math. Comp.*, **52** (1989) 411–435.
- ⁶B. Cockburn & C.-W. Shu, *The Runge-Kutta local projection P^1 discontinuous Galerkin method for scalar conservation laws*, *M²AN Math. Model. Anal. Numer.*, **25** (1991) 337–361.
- ⁷B. Cockburn & C.-W. Shu, *Runge-Kutta discontinuous Galerkin methods for convection-dominated problems*, *J. Sci. Computing*, **16** (2001) 173–261.
- ⁸N. Kroll, H. Bieler, H. Deconinck, V. Couaillier, H. van der Ven & K. Sorensen (eds.), *ADIGMA - A European initiative on the development of adaptive higher-order variational methods for aerospace applications*, Notes on Numerical Fluid Mechanics and Multidisciplinary Design, Ed. Springer Verlag, **113** (2010).
- ⁹P. Lesaint & P.-A. Raviart, *On a finite element method for solving the neutron transport equation* in *Mathematical Aspects of Finite Elements in Partial Differential Equations*, Ed. Carl de Boor, Academic Press, New York, (1974) 89–123.
- ¹⁰W. H. Reed & T. R. Hill, *Triangular mesh methods for the neutron transport equation*, Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, NM, (1973).
- ¹¹F. Renac, C. Marmignon & F. Coquel, *Time implicit high-order discontinuous Galerkin method with reduced evaluation cost*, *SIAM J. Sci. Comput.*, to appear, Preprint available at <http://hal.archives-ouvertes.fr/docs/00/56/39/87/PDF/R11004.pdf>.
- ¹²C.-W. Shu & S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes*, *J. Comput. Phys.*, **77** (1988) 439–471.
- ¹³G. Strang, *On the construction and comparison of difference schemes*, *SIAM J. Numer. Anal.*, **5** (1968) 506–517.

G.2 Juin 2013 - Article du Journal of Computational Physics

Référence : Article, [136]
Auteurs : F. Renac, S. Gérald, C. Marmignon, F. Coquel
Journal : Journal of Computational Physics
Volume : 251
Pages : 272 – 291
Date : Juin 2013

*Le contenu de cet article étant protégé, ce dernier est ici présenté dans sa version de soumission. Le document original reste disponible à l'adresse suivante :
<http://www.sciencedirect.com/science/article/pii/S0021999113004105>*

Fast time implicit-explicit discontinuous Galerkin method for the compressible Navier-Stokes equations

Florent Renac^{a,*}, Sophie Gérald^{a,b}, Claude Marmignon^a, Frédéric Coquel^c

^aONERA The French Aerospace Lab, 92320 Châtillon Cedex, France

^bLJLL UMR 7598, Université Paris 6, 75005 Paris, France

^cCMAP-CNRS UMR 7641, Ecole Polytechnique, 91128 Palaiseau Cedex, France

Abstract

An efficient and robust time integration procedure for a high-order discontinuous Galerkin method is introduced for solving the unsteady compressible Navier-Stokes equations. The time discretization is based on an explicit formulation for the convective fluxes and an implicit formulation for the viscous fluxes. The implicit procedure uses a fast iterative algorithm based on a partial uncoupling of variables in neighboring elements introduced in previous works by the authors. In the present work, the method is associated to a Newton-Krylov Jacobian-free method. This association allows to reduce memory requirements and operation counts by avoiding the complete construction of the Jacobian matrix and solving a problem of reduced size. Numerical examples in two space dimensions indicate that the performance of the present method is seen to be significantly improved in terms of CPU time.

Keywords: discontinuous Galerkin method, implicit-explicit time discretization, Newton-Krylov method, Jacobian-free

1. Introduction

Discontinuous Galerkin (DG) methods are high-order finite element discretizations which were introduced in the early 1970s for the numerical simulation of the first-order hyperbolic neutron transport equation [29, 35]. In recent years, these methods have become very popular for the solution of nonlinear convection dominated flow problems [12, 13, 28]. The success of these methods lies in their high-order of accuracy and flexibility thanks to their high degree of locality. These properties make the DG method well suited to parallel computing [8], *hp*-refinement [24], *p*-multigrid [20, 31], unstructured meshes [1], the application of boundary conditions, *etc.*

However, the association of a DG method to an explicit time discretization leads to a strong restriction on the time step due to the so-called Courant-Friedrichs-Levy (CFL) condition for stability of the numerical scheme [2, 10]. The discretization of the parabolic terms requires an additional stability constraint on the von Neumann number, which becomes more restrictive than the CFL condition for flow regions with small cell Reynolds number [26]. In high Reynolds number flows, the large variations in element size required to resolve all the spatial scales make the use of explicit integration techniques prohibitive. Implicit solvers are thus needed with this type of flows. However, the shortcoming of these methods is the extremely high computational cost induced by the large number of degrees of freedom (DOFs) in practical applications (*e.g.* the resolution of Navier-Stokes equations on multi-dimensional grids). Being able to solve these large systems at a reasonable computational cost is therefore essential for the efficiency of the DG methods.

For large size problems, the use of direct solution techniques suffers from high memory requirements and long execution time and iterative methods must be applied. Several authors have applied Newton GMRES and implicit Runge-Kutta methods to the resolution of the compressible Euler, Navier-Stokes and RANS equations in the context of DG discretizations [4, 6, 16, 23, 34]. Various preconditioned iterative algorithms have been introduced showing the

*Corresponding author. Tel.: +33 1 46 73 37 44; fax.: +33 1 46 73 41 66.
Email address: florent.renac@onera.fr (Florent Renac)

potential benefits of taking advantage of the block structure of the Jacobian matrix, of a factorization using reordering [17, 20, 26, 33, 34], or of static condensation [22, 41]. Multigrid approaches where convergence acceleration is achieved through the use of p -multigrid or h -multigrid have also been proposed [17, 20, 31]. Simplifications of the implicit problem have also been introduced. Yasue *et al.* [49] used a pointwise relaxation implicit scheme associated to the resolution of a simplified implicit system for the compressible Euler and Navier-Stokes equations in three-dimensions. Dolejši *et al.* developed a semi-implicit method where the linear terms in the discrete equations are treated implicitly and the nonlinear terms are treated with an explicit extrapolation [15, 16]. In [21], $p_N p_M$ -schemes from [19] with a predictor-corrector formulation were used for an implicit time discretization with polynomials of degree N in the predictor step. A global corrector is then applied where the solution is reconstructed up to an accuracy $M + 1$ with $M > N$.

A robust and fast time integration procedure for an implicit-explicit formulation of a DG discretization has been introduced in [37, 38]. The method uses a time implicit discretization for the viscous terms, while convective terms are treated explicitly. The method is called the SIMP_{p_s} method and consists in a hierarchy of simplified methods. It is based on a reduced communication between elements which share a common interface. This assumption follows from the natural hierarchy of the basis of the discrete function space: the communication between elements is assumed to be dominated by the coupling between DOFs associated to the components of the basis of order lower than a given degree p_s . After solving an implicit problem of reduced size for the low-order DOFs only, a local reconstruction, based on the stencil of the DG discretization, is used for evaluating higher-order DOFs. Numerical experiments on one-dimensional and two-dimensional (2D) nonlinear convection-diffusion equations have highlighted better performances of the SIMP_{p_s} method in terms of CPU time when compared to a full implicit discretization of the viscous terms. This simplification also keeps the locality, accuracy and stability properties of the original method. Theoretical and numerical analyses indicate CPU time savings increasing with the approximation order of the numerical scheme, but also moderate savings for low approximation orders. Let us underline that the assumption of partial coupling between neighboring elements was also used in [7] to neglect components in the preconditioning matrix. The resulting preconditioner corresponds to a LU-SGS method when full coupling is assumed and reduces to a block-Jacobi method when total uncoupling is assumed. As introduced in [37, 38], the present approach uses the partial coupling to simplify the whole implicit problem to be solved at each time step. Therefore, operation reduction can be expected in the construction of the implicit matrix, its factorization, in the preconditioning step, as well as the resolution of the linear system.

The main objective of the present study is to extend the SIMP_{p_s} method to time-dependent nonlinear systems of conservation laws. We propose to use a Jacobian-free (JF) Newton-Krylov method [27] which has been successfully applied in the context of DG discretizations [14, 34] to save storage and computation time. We introduce an algorithm that associates the simplification of the implicit problem with the JF method. The new method is called the JF- SIMP_{p_s} method and allows to reduce the memory requirements and operation counts by avoiding the complete construction of the Jacobian matrix and solving a problem of reduced size.

The paper is organized as follows. Section 2 presents the model problem and the numerical approach for the space-time discretization. In particular, the time discretization and its simplification are introduced in §§ 2.3 and 2.4 with an analysis of computational and memory costs. The overall performance of the method is assessed by several numerical experiments in § 3. Results are obtained for 2D unsteady flow problems. Finally, the conclusions of this work are summarized in § 4.

2. Model problem and discretization

2.1. Navier-Stokes equations

The discussion in this paper focuses on the discretization of the compressible Navier-Stokes equations for gas dynamics with a DG method. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain where $d \geq 1$ is the space dimension and consider the following problem

$$\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}_c(\mathbf{u}) - \nabla \cdot \mathbf{f}_v(\mathbf{u}, \nabla \mathbf{u}) = 0, \quad \text{in } \Omega \times (0, \infty), \quad (1a)$$

$$\mathbf{u}(\cdot, 0) = \mathbf{u}_0(\cdot), \quad \text{in } \Omega, \quad (1b)$$

with appropriate boundary conditions prescribed on $\partial\Omega$. The vector

$$\mathbf{u} = \begin{pmatrix} \rho \\ \rho\mathbf{v} \\ \rho E \end{pmatrix} \quad (2)$$

represents the conservative variables with ρ the density, \mathbf{v} in \mathbb{R}^d the velocity vector and $E = p/(\gamma - 1)\rho + \|\mathbf{v}\|_2^2/2$ the total specific energy where p denotes the static pressure, $\gamma = C_p/C_v > 1$ is the ratio of specific heats, and $\|\cdot\|_2$ denotes the Euclidian norm. The nonlinear convective and diffusive fluxes in (1) are defined by

$$\mathbf{f}_c(\mathbf{u}) = \begin{pmatrix} \rho\mathbf{v}^\top \\ \rho\mathbf{v}\mathbf{v}^\top + p\mathbf{I} \\ (\rho E + p)\mathbf{v}^\top \end{pmatrix}, \quad \mathbf{f}_v(\mathbf{u}, \nabla\mathbf{u}) = \begin{pmatrix} 0 \\ \boldsymbol{\tau} \\ \mathbf{v}^\top\boldsymbol{\tau} - \mathbf{q}^\top \end{pmatrix}, \quad (3)$$

where $\boldsymbol{\tau} = \mu(-\frac{2}{3}(\nabla \cdot \mathbf{v})\mathbf{I} + \nabla\mathbf{v} + \nabla\mathbf{v}^\top)$ denotes the viscous stress tensor with μ the kinematic viscosity coefficient defined by the Sutherland formula, and $\mathbf{q} = -k\nabla T$ is the Fourier's heat conduction law with $k = \mu C_p/Pr$ the thermal conductivity, $Pr = 0.72$ the Prandtl number and T the temperature.

2.2. Discontinuous Galerkin formulation

The DG method consists in defining a discrete weak formulation of the problem (1). The domain Ω is partitioned into a shape-regular mesh Ω_h consisting of nonoverlapping and nonempty simplex elements κ of characteristic size $h := \min\{h_\kappa, \kappa \in \Omega_h\}$ where h_κ is a d -dimensional measure of κ . We further define the sets \mathcal{E}_i and \mathcal{E}_b of interior and boundary faces in Ω_h and set $\mathcal{E}_h = \mathcal{E}_i \cup \mathcal{E}_b$.

We look for approximate solutions in the function space of discontinuous polynomials $\mathcal{V}_h^p = \{\phi \in L^2(\Omega_h) : \phi|_\kappa \circ F_\kappa \in \mathcal{P}^p(I^d), \forall \kappa \in \Omega_h\}$, where $\mathcal{P}^p(I^d)$ denotes the space of functions formed by tensor products of polynomials and of total degree at most p on the master element I^d where $I = [-1, 1]$. Each physical element κ is the image of I^d through the mapping F_κ . A Gram-Schmidt orthonormalization is used to construct the orthonormal basis from an initial basis in order to get a diagonal mass matrix for the explicit time integration. We refer to [3, 36] for details on this procedure. The numerical integration is performed in I^d as proposed in [18], while optimal quadrature for triangles would require numerical integration in a reference triangle [43]. Let us underline that the theoretical analysis of § 2.4.3 assumes a general number of quadrature points and is therefore independent of the present choice.

The numerical solution of equation (1) is sought under the form

$$\mathbf{u}_h(\mathbf{x}, t) = \sum_{l=1}^{N_p} \phi_k^l(\mathbf{x}) \mathbf{U}_k^l(t), \quad \forall \mathbf{x} \in \kappa, \kappa \in \Omega_h, \forall t \geq 0, \quad (4)$$

where $(\mathbf{U}_k^l)_{1 \leq l \leq N_p}$ are the DOFs in the element κ . The subset $(\phi_k^1, \dots, \phi_k^{N_p})$ constitutes a hierarchical and orthonormal modal basis of \mathcal{V}_h^p restricted onto the element κ and $N_p = \prod_{i=1}^d (p + i)/i$ is its dimension.

The semi-discrete form of system (1a) reads: find \mathbf{u}_h in $[\mathcal{V}_h^p]^{d+2}$ such that for all v_h in \mathcal{V}_h^p we have

$$\int_{\Omega_h} v_h \partial_t \mathbf{u}_h dV + \mathcal{L}_c(\mathbf{u}_h, v_h) + \mathcal{L}_v(\mathbf{u}_h, v_h) = 0. \quad (5)$$

Hereafter, we will use the notations $\{\phi\} = (\phi^+ + \phi^-)/2$ and $[\![\phi]\!] = \phi^+ - \phi^-$ which denote the average and jump operators defined for a given interface e in \mathcal{E}_i . Here, ϕ^+ and ϕ^- are the traces of any quantity ϕ on the interface e taken from within the interior of the element κ^+ and the interior of the neighboring element κ^- , respectively (see Fig. 1).

The discretization of the convective terms in equation (5) reads

$$\begin{aligned} \mathcal{L}_c(\mathbf{u}_h, v_h) &= - \int_{\Omega_h} \mathbf{f}_c(\mathbf{u}_h) \cdot \nabla v_h dV \\ &+ \int_{\mathcal{E}_i} [\![v_h]\!] \mathbf{h}_c(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) dS \\ &+ \int_{\mathcal{E}_b} v_h^+ \mathbf{f}_c(\mathbf{u}_b(\mathbf{u}_h^+, \mathbf{n})) \cdot \mathbf{n} dS, \end{aligned} \quad (6)$$

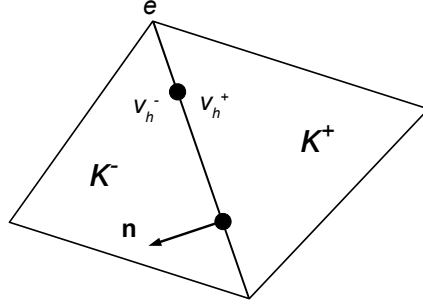


Figure 1: Inner and exterior elements κ^+ and κ^- and definition of traces v_h^\pm on the interface e and of the unit outward normal vector \mathbf{n} .

where \mathbf{n} denotes the unit outward normal vector to an element κ^+ (see Fig. 1) and \mathbf{u}_b is an appropriate operator which allows to impose the boundary conditions on \mathcal{E}_b . The numerical flux \mathbf{h}_c may be chosen to be any monotonic Lipschitz function satisfying consistency and conservativity properties (see [13] for instance). In numerical examples of § 3 we use the Lax-Friedrichs flux $\mathbf{h}_c(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) = \{\mathbf{f}_c(\mathbf{u}_h)\} + \alpha \llbracket \mathbf{u}_h \rrbracket / 2$ with $\alpha = \max\{\rho_s(\nabla_{\mathbf{u}}(\mathbf{f}_c(\mathbf{u}) \cdot \mathbf{n})) : \mathbf{u} = \mathbf{u}_h^\pm\}$ where $\nabla_{\mathbf{u}}(\mathbf{f}_c \cdot \mathbf{n})$ denotes the Jacobian matrix of the convective fluxes in the direction \mathbf{n} and ρ_s its spectral radius.

For the discretization of the diffusive terms in (5), we use the method of Bassi and coworkers [5]:

$$\begin{aligned} \mathcal{L}_v(\mathbf{u}_h, v_h) &= \int_{\Omega_h} \mathbf{f}_v(\mathbf{u}_h, \nabla_h \mathbf{u}_h + \mathbf{R}_h) \cdot \nabla v_h dV \\ &\quad - \int_{\mathcal{E}_i} \llbracket v_h \rrbracket \{\mathbf{f}_v(\mathbf{u}_h, \nabla_h \mathbf{u}_h + \mathbf{r}_h^e)\} \cdot \mathbf{n} dS \\ &\quad - \int_{\mathcal{E}_b} v_h^+ \mathbf{f}_v(\mathbf{u}_b(\mathbf{u}_h^+, \mathbf{n}), \nabla \mathbf{u}_b(\mathbf{u}_h^+, \mathbf{n}) + \mathbf{r}_h^e) \cdot \mathbf{n} dS, \end{aligned} \quad (7)$$

where Neumann boundaries are enforced in the expression of \mathbf{f}_v on \mathcal{E}_b through the operator $\nabla \mathbf{u}_b(\mathbf{u}_h^+, \mathbf{n})$. The so-called global lifting operator \mathbf{R}_h is defined in a weak sense as the sum of local lifting operators \mathbf{r}_h^e : $\mathbf{R}_h = \sum_{e \in \mathcal{E}_h} \mathbf{r}_h^e$, where \mathbf{r}_h^e has support on elements adjacent to e in \mathcal{E}_i and is solution of the following problem for internal faces: for all v_h in \mathcal{V}_h^p we have

$$\int_{\kappa^+ \cup \kappa^-} v_h \mathbf{r}_h^e dV = - \int_e \{v_h\} \llbracket \mathbf{u}_h \rrbracket \mathbf{n}^\top dS.$$

A similar expression consistent with boundary conditions of the problem is used on \mathcal{E}_b . We refer to [4, 5] for further details on this method.

2.3. Time discretization

The semi-discrete equation (5) is discretized in time by means of a second-order Strang splitting method [44]. Introducing a time step Δt and defining $t^{(n)} = n\Delta t$, the solution is advanced in time in three substeps on the interval $(t^{(n)}, t^{(n+1)})$. First, the hyperbolic problem is solved for \mathbf{u}_h^c on the time interval $[t^{(n)}, t^{(n)} + \frac{\Delta t}{2}]$:

$$\int_{\Omega_h} v_h \partial_t \mathbf{u}_h^c dV + \mathcal{L}_c(\mathbf{u}_h^c, v_h) = 0, \quad \forall t \in (t^{(n)}, t^{(n)} + \frac{\Delta t}{2}], \quad \forall v_h \in \mathcal{V}_h^p, \quad (8)$$

with initial condition $\mathbf{u}_h^c(\cdot, t^{(n)}) = \mathbf{u}_h(\cdot, t^{(n)})$. Then, the parabolic problem is solved for \mathbf{u}_h^v on the time interval $[t^{(n)}, t^{(n)} + \Delta t]$:

$$\int_{\Omega_h} v_h \partial_t \mathbf{u}_h^v dV + \mathcal{L}_v(\mathbf{u}_h^v, v_h) = 0, \quad \forall t \in (t^{(n)}, t^{(n)} + \Delta t], \quad \forall v_h \in \mathcal{V}_h^p, \quad (9)$$

with $\mathbf{u}_h^v(\cdot, t^{(n)}) = \mathbf{u}_h^c(\cdot, t^{(n)} + \frac{\Delta t}{2})$. Finally, the hyperbolic problem is solved once again for \mathbf{u}_h^c on the time interval $[t^{(n)} + \frac{\Delta t}{2}, t^{(n)} + \Delta t]$:

$$\int_{\Omega_h} v_h \partial_t \mathbf{u}_h^c dV + \mathcal{L}_c(\mathbf{u}_h^c, v_h) = 0, \quad \forall t \in (t^{(n)} + \frac{\Delta t}{2}, t^{(n)} + \Delta t], \quad \forall v_h \in \mathcal{V}_h^p, \quad (10)$$

with $\mathbf{u}_h^c(\cdot, t^{(n)} + \frac{\Delta t}{2}) = \mathbf{u}_h^v(\cdot, t^{(n)} + \Delta t)$. The solution of equation (5) at time $t^{(n+1)}$ is thus approximated by $\mathbf{u}_h(\cdot, t^{(n+1)}) = \mathbf{u}_h^c(\cdot, t^{(n)} + \Delta t)$. We use a Heun's method for the time discretization of hyperbolic problems (8) and (10). For example for (8) we apply

$$\mathbf{U}_c^* = \mathbf{U}^{(n)} - \frac{\Delta t}{2} \mathbf{M}^{-1} \mathbf{L}_c(\mathbf{U}^{(n)}), \quad (11a)$$

$$\mathbf{U}_c^{(n+\frac{1}{2})} = \frac{1}{2} (\mathbf{U}_c^* + \mathbf{U}^{(n)} - \frac{\Delta t}{2} \mathbf{M}^{-1} \mathbf{L}_c(\mathbf{U}_c^*)), \quad (11b)$$

with \mathbf{U} the vector of DOFs, $\mathbf{L}_c(\mathbf{U})$ the residual vector defined by the discrete form (6) evaluated from \mathbf{U} , and \mathbf{M} denotes the diagonal mass matrix. The method (11) is explicit and formally second-order accurate in time [42].

Problem (9) is discretized in time with an implicit second-order Crank-Nicolson method and reduces to the following linear system

$$\mathbf{A} \Delta \mathbf{U}_v^{(n+1)} = -\mathbf{L}_v(\mathbf{U}_v^{(n)}), \quad (12)$$

with $\Delta \mathbf{U}_v^{(n+1)} = \mathbf{U}_v^{(n+1)} - \mathbf{U}_v^{(n)}$, $\mathbf{U}_v^{(n)} = \mathbf{U}_c^{(n+1/2)}$ and $\mathbf{L}_v(\mathbf{U}_v^{(n)})$ the residual vector defined by the discrete form (7). The so-called implicit matrix is defined by

$$\mathbf{A} = \frac{1}{\Delta t} \mathbf{M} + \frac{1}{2} \frac{\partial \mathbf{L}_v}{\partial \mathbf{U}} \Big|_{\mathbf{U}=\mathbf{U}_v^{(n)}}, \quad (13)$$

and is an unsymmetric real matrix of size $NN_{eq}N_p$ where $N_{eq} = d + 2$ denotes the number of conservative variables and N denotes the number of elements. For the sake of clarity, we assume a uniform distribution of the polynomial degrees in all elements when introducing the present method without loss of generality.

The sparse structure of the implicit matrix and the large number of unknowns motivate the choice of a Krylov subspace method to solve the linear system (12). We use a restarted GMRES method [40] with a block Jacobi preconditioner for its resolution at each time step. The inversion cost of \mathbf{A} may become prohibitive in practical applications where N is large and where accuracy is looked for. To circumvent this shortcoming, we propose the use of the modification of the implicit problem (12) introduced in [37, 38] together with a JF method for simplifying its resolution. This method and different algorithms for the time integration used in this work are detailed in the next section.

2.4. Time integration strategies

2.4.1. Simplification of the implicit procedure

Here we shortly describe the simplification of the implicit problem (12) for systems of conservation laws with the SIMP p_s method. We refer to [37] for more details and analysis of the method in the context of scalar equations. The implicit matrix has a block-sparse structure with N block rows consisting of one diagonal block and n_f off-diagonal blocks where n_f denotes the number of neighboring elements. Each block is full and has a size of $N_{eq}N_p$ which becomes very large for high-order discretizations since $N_p \sim p^d$ when p is large. For a given element κ in Ω_h , let \mathbf{A}_κ in $\mathbb{R}^{N_{eq}N_p \times N_{eq}N_p}$ denote the diagonal block associated to the discretization in the element, and let $\mathbf{A}_{\kappa, \kappa_i^-}$ in $\mathbb{R}^{N_{eq}N_p \times N_{eq}N_p}$ denote the i -th off-diagonal block corresponding to the coupling of the element with its i -th neighboring element κ_i^- with $1 \leq i \leq n_f$ (see Fig. 2a). The sparse structure of the implicit matrix may be represented in the following compact form

$$\mathbf{A} = (\mathbf{A}_\kappa | \mathbf{A}_{\kappa, \kappa_i^-})_{\kappa \in \Omega_h}.$$

In the following, we will use the shorter notation

$$\mathbf{A} = (\mathbf{A}_\kappa | \mathbf{A}_{\kappa, \kappa^-}). \quad (14)$$

The SIMP_{p_s} method consists in a partial coupling between DOFs in the element κ and DOFs in neighboring elements κ_i^- . This reduction modifies the pattern of the matrices $\mathbf{A}_{\kappa, \kappa^-}$. The simplification indeed introduces a user-defined integer p_s that separates low-order DOFs associated to components of degree $0 \leq q \leq p_s$ and high-order DOFs associated to components $p_s < q \leq p$. Each block is ordered in such a way that the DOFs associated to a variable in \mathbf{u}_h appear contiguously. For instance, the diagonal blocks consist in $N_{eq} \times N_{eq}$ submatrices $\mathbf{A}_\kappa = (\mathbf{A}_\kappa^{i,j})_{1 \leq i, j \leq N_{eq}}$. These submatrices may further be decomposed into parts associated to low- and high-order DOFs in the following way

$$\mathbf{A}_\kappa^{i,j} = \begin{pmatrix} \tilde{\mathbf{A}}_\kappa^{i,j} & \mathbf{B}_\kappa^{i,j} \\ \mathbf{C}_\kappa^{i,j} & \hat{\mathbf{A}}_\kappa^{i,j} \end{pmatrix}, \quad \kappa \in \Omega_h, \quad 1 \leq i, j \leq N_{eq}, \quad (15)$$

where $\tilde{\mathbf{A}}_\kappa^{i,j} \in \mathbb{R}^{N_{p_s} \times N_{p_s}}$ and $\hat{\mathbf{A}}_\kappa^{i,j} \in \mathbb{R}^{(N_p - N_{p_s}) \times (N_p - N_{p_s})}$ with $N_{p_s} = \prod_{i=1}^d (p_s + i)/i$ denoting the number of low-order DOFs. A similar decomposition is used for the off-diagonal blocks. Using the above notations, the implicit matrix (13) may be written as

$$\mathbf{A} = \left(\left(\begin{pmatrix} \tilde{\mathbf{A}}_\kappa^{i,j} & \mathbf{B}_\kappa^{i,j} \\ \mathbf{C}_\kappa^{i,j} & \hat{\mathbf{A}}_\kappa^{i,j} \end{pmatrix} \right)_{1 \leq i, j \leq N_{eq}} \middle| \left(\begin{pmatrix} \tilde{\mathbf{A}}_{\kappa, \kappa^-}^{i,j} & \mathbf{B}_{\kappa, \kappa^-}^{i,j} \\ \mathbf{C}_{\kappa, \kappa^-}^{i,j} & \hat{\mathbf{A}}_{\kappa, \kappa^-}^{i,j} \end{pmatrix} \right)_{1 \leq i, j \leq N_{eq}} \right). \quad (16)$$

Fig. 2b displays the structure of the implicit matrix in the case of $N_{eq} = 2$ and a mesh containing $N = 4$ triangles with a $p = 2$ approximation as depicted in Fig. 2a. Partial coupling of high-order DOFs in neighboring elements amounts to replace the matrix \mathbf{A} by the following matrix

$$\mathbf{A}^* = \left(\left(\begin{pmatrix} \tilde{\mathbf{A}}_\kappa^{i,j} & \mathbf{B}_\kappa^{i,j} \\ \mathbf{C}_\kappa^{i,j} & \hat{\mathbf{A}}_\kappa^{i,j} \end{pmatrix} \right)_{1 \leq i, j \leq N_{eq}} \middle| \left(\begin{pmatrix} \tilde{\mathbf{A}}_{\kappa, \kappa^-}^{i,j} & 0 \\ \mathbf{C}_{\kappa, \kappa^-}^{i,j} & 0 \end{pmatrix} \right)_{1 \leq i, j \leq N_{eq}} \right). \quad (17)$$

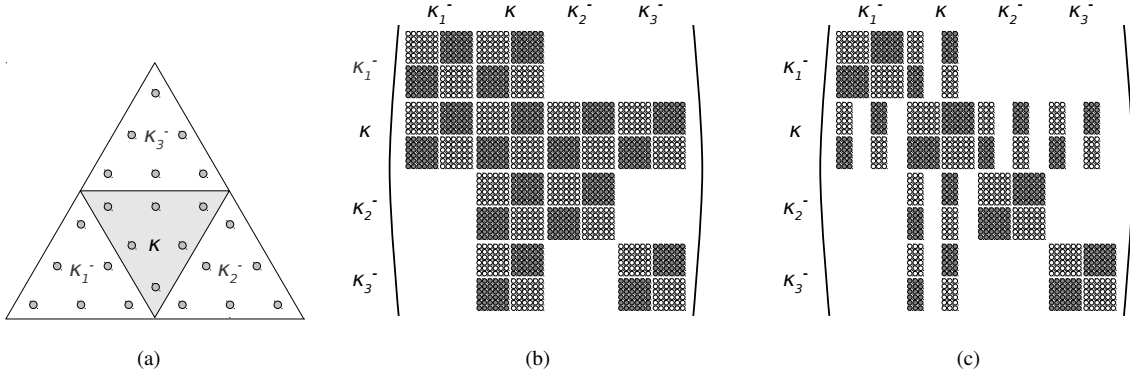


Figure 2: Structure of the implicit matrix for a mesh containing $N = 4$ triangles with $p = 2$ and $N_p = 6$ DOFs per element as illustrated in (a). The non-zero entries of the matrix \mathbf{A} in (16) are represented in (b) by open and full circles according to the $N_{eq} = 2$ variables of the system, while the non-zero entries with $p_s = 1$ of the matrix \mathbf{A}^* in (17) are represented in (c).

The modification of the matrix pattern in off-diagonal blocks is illustrated in Fig. 2c for $p_s = 1$. The SIMP_{p_s} method uses this uncoupling and consists in a two-step algorithm. First, an implicit problem is solved for the low-order DOFs only:

$$\tilde{\mathbf{A}} \Delta \tilde{\mathbf{U}}_v^{(n+1)} = -\tilde{\mathbf{R}}(\mathbf{U}_v^{(n)}) - \mathbf{B} \Delta \hat{\mathbf{U}}_v^{(n)}, \quad (18)$$

where $\tilde{\mathbf{A}} = (\tilde{\mathbf{A}}_\kappa | \tilde{\mathbf{A}}_{\kappa, \kappa^-})$ is a square matrix of size $NN_{eq}N_{p_s}$, $\Delta \tilde{\mathbf{U}}_v^{(n+1)} = \tilde{\mathbf{U}}_v^{(n+1)} - \tilde{\mathbf{U}}_v^{(n)}$ and $\Delta \hat{\mathbf{U}}_v^{(n)} = \hat{\mathbf{U}}_v^{(n)} - \hat{\mathbf{U}}_v^{(n-1)}$ where $\tilde{\mathbf{U}}_v$ and $\hat{\mathbf{U}}_v$ contain low-order and high-order DOFs, respectively. By $\tilde{\mathbf{R}}(\mathbf{U}_v^{(n)})$ we denote the residuals associated to low-order DOFs and we set $\mathbf{B} = \text{diag}(\mathbf{B}_\kappa)_{\kappa \in \Omega_h}$ with $\mathbf{B}_\kappa = (\mathbf{B}_\kappa^{i,j})_{1 \leq i, j \leq N_{eq}}$.

Second, higher-order DOFs are reconstructed from the low-order DOFs and precedent time step solution. We note $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_\kappa | \hat{\mathbf{A}}_{\kappa, \kappa^-})$ the square matrix of size $NN_{eq}(N_p - N_{p_s})$ and let $\hat{\mathbf{A}} = \hat{\mathcal{L}} + \hat{\mathcal{D}} + \hat{\mathcal{U}}$ be its decomposition into strictly lower, diagonal and strictly upper parts, respectively. The reconstruction of higher DOFs reads as follows

$$(\hat{\mathcal{L}} + \hat{\mathcal{D}})\Delta\hat{\mathbf{U}}_v^{(n+1)} = -\hat{\mathbf{R}}(\mathbf{U}_v^{(n)}) - \mathbf{C}\Delta\tilde{\mathbf{U}}_v^{(n+1)} - \hat{\mathcal{U}}\Delta\hat{\mathbf{U}}_v^{(n)}, \quad (19)$$

where $\hat{\mathbf{R}}(\mathbf{U}_v^{(n)})$ corresponds to the vector of residuals associated to high-order DOFs and $\mathbf{C} = (\mathbf{C}_\kappa | \mathbf{C}_{\kappa, \kappa^-})$.

2.4.2. Jacobian-free method and its simplification

The GMRES algorithm requires the evaluation of the product of the implicit matrix with a given vector. For computational efficiency, it is more suitable to store the implicit matrix than to compute the matrix-vector product on the fly [14]. The JF method is an attempt to resolve the huge memory requirements induced by the storage of the implicit matrix [27]. It approximates the product in (12) by

$$\mathbf{A}\Delta\mathbf{U}_v^{(n+1)} \simeq \frac{1}{\Delta t}\mathbf{M}\Delta\mathbf{U}_v^{(n+1)} + \frac{1}{2\delta}\left(\mathbf{L}_v(\mathbf{U}_v^{(n)} + \delta\Delta\mathbf{U}_v^{(n+1)}) - \mathbf{L}_v(\mathbf{U}_v^{(n)})\right), \quad (20)$$

where the product $\mathbf{M}\Delta\mathbf{U}_v^{(n+1)}$ is cheap since \mathbf{M} is diagonal. We use the difference step

$$\delta = \epsilon \frac{\sqrt{1 + \|\mathbf{U}_v^{(n)}\|_2}}{\|\Delta\mathbf{U}_v^{(n+1)}\|_2}$$

as proposed in [14, 32]. This method was successfully applied to the implicit time discretization of DG methods for both inviscid [34] and turbulent flows [14]. In this work, we use the same value $\epsilon = 10^{-8}$ as in [14] for all computations presented in § 3.

In practice, one still forms the diagonal blocks of the matrix for preconditioning purposes so the SIMP_{p_s} method may be easily combined with the JF approach. In this way, the matrix-vector product for the resolution of equation (18) is approximated by

$$\tilde{\mathbf{A}}\Delta\tilde{\mathbf{U}}_v^{(n+1)} \simeq \frac{1}{\Delta t}\tilde{\mathbf{M}}\Delta\tilde{\mathbf{U}}_v^{(n+1)} + \frac{1}{2\delta}\left(\tilde{\mathbf{L}}_v(\mathbf{U}_v^{(n)} + \delta\Delta\tilde{\mathbf{U}}_v^{(n+1)}) - \tilde{\mathbf{L}}_v(\mathbf{U}_v^{(n)})\right), \quad (21)$$

where the variations in the variable vector have been restricted to low-order DOFs. The second step requires another matrix-vector product for the construction of the right-hand side of (19) which is replaced by the restriction to equations corresponding to high-order DOFs of

$$\mathbf{C}\Delta\tilde{\mathbf{U}}_v^{(n+1)} \simeq \frac{1}{2\delta}\left(\hat{\mathbf{L}}_v(\mathbf{U}_v^{(n)} + \delta\Delta\tilde{\mathbf{U}}_v^{(n+1)}) - \hat{\mathbf{L}}_v(\mathbf{U}_v^{(n)})\right). \quad (22)$$

Other products in (18) and (19) are formed directly with the diagonal blocks of the implicit matrix.

In the following we will use the following acronyms to refer to different methods: FULL for the full implicit method (12); JF for the Jacobian-free method (20); SIMP_{p_s} for the simplified implicit method (18)-(19) where p_s denotes the integer that separates the low- and high-order DOFs; and JF- SIMP_{p_s} the method (21)-(22) that associates the two latter methods. We now compare the properties and performances of these methods in the next sections.

2.4.3. Memory and operation counts

The evaluation of memory and computational costs of the linear solver is important since both may become important issues for the resolution of large-scale problems. We first evaluate the memory needed to store the implicit matrix \mathbf{A} and preconditioning matrix \mathbf{P} in each method. For the FULL method, the full implicit matrix is stored, thus using memory for its $N(n_f + 1)N_{eq}^2N_p^2$ non-zero components in the diagonal and the n_f off-diagonal blocks per row. The diagonal blocks are then LU factorized and replaced by their factorization which are used for both the product of diagonal blocks by a vector and for the preconditioning step. As a consequence, the method uses only the memory required to store the implicit matrix with no additional memory storage. For the SIMP_{p_s} method, the submatrices $\mathbf{B}_{\kappa, \kappa^-}$ and $\hat{\mathbf{A}}_{\kappa, \kappa^-}$ of the off-diagonal blocks are neglected by the simplification and are neither computed nor stored. We note that the method requires an additional storage of the vector $\Delta\hat{\mathbf{U}}_v^{(n)}$ in (18) and (19), thus increasing the memory

Table 1: Memory storage estimates per element for the implicit and preconditioning matrices.

| FULL | JF | SIMP p_s | JF-SIMP p_s |
|---------------------------|------------------|---|------------------|
| $(n_f + 1)N_{eq}^2 N_p^2$ | $N_{eq}^2 N_p^2$ | $N_{eq}^2 N_p^2 + n_f N_{eq}^2 N_p N_{p_s}$ | $N_{eq}^2 N_p^2$ |

Table 2: Asymptotic operation counts per element.

| operation | FULL | JF | SIMP p_s | JF-SIMP p_s |
|--|-------------------------------|-------------------------------|--|-------------------------------------|
| residuals | $\alpha_1 N_g N_{eq} N_p$ | $\alpha_1 N_g N_{eq} N_p$ | $\alpha_1 N_g N_{eq} N_p$ | $\alpha_1 N_g N_{eq} N_p$ |
| form A | | | | |
| diagonal blocks | $\alpha_2 N_g N_{eq}^2 N_p^2$ | $\alpha_2 N_g N_{eq}^2 N_p^2$ | $\alpha_2 N_g N_{eq}^2 N_p^2$ | $\alpha_2 N_g N_{eq}^2 N_p^2$ |
| off-diagonal blocks | $\alpha_3 N_g N_{eq}^2 N_p^2$ | 0 | $\alpha_3 N_g N_{eq}^2 N_p N_{p_s}$ | 0 |
| solver | | | | |
| form P | | | | |
| Ax | $\frac{2}{3} N_{eq}^3 N_p^3$ | $\frac{2}{3} N_{eq}^3 N_p^3$ | $\frac{2}{3} N_{eq}^3 N_{p_s}^3$ | $\frac{2}{3} N_{eq}^3 N_{p_s}^3$ |
| P⁻¹x | $2(n_f + 1)N_{eq}^2 N_p^2$ | $\alpha_1 N_g N_{eq} N_p$ | $2(n_f + 1)N_{eq}^2 N_{p_s}^2$ | $\alpha_1 N_g N_{eq} N_p$ |
| P⁻¹x | $2N_{eq}^2 N_p^2$ | $2N_{eq}^2 N_p^2$ | $2N_{eq}^2 N_{p_s}^2$ | $2N_{eq}^2 N_{p_s}^2$ |
| reconstruction | | | | |
| BΔŨ_v⁽ⁿ⁾ | 0 | 0 | $2N_{eq}^2 (N_p - N_{p_s}) N_{p_s}$ | $2N_{eq}^2 (N_p - N_{p_s}) N_{p_s}$ |
| CΔŨ_v⁽ⁿ⁺¹⁾ | 0 | 0 | $2(n_f + 1)N_{eq}^2 (N_p - N_{p_s}) N_{p_s}$ | $\alpha_1 N_g N_{eq} N_p$ |
| ŨΔŨ_v⁽ⁿ⁾ | 0 | 0 | $N_{eq}^2 (N_p - N_{p_s})^2$ | $N_{eq}^2 (N_p - N_{p_s})^2$ |
| (Ũ + D̂)⁻¹ΔŨ_v⁽ⁿ⁺¹⁾ | 0 | 0 | $N_{eq}^2 (N_p - N_{p_s})^2$ | $N_{eq}^2 (N_p - N_{p_s})^2$ |

requirement by $NN_{eq}(N_p - N_{p_s})$. This extra-storage is however negligible compared to the memory needs for the implicit matrix. Finally, only the diagonal blocks of the matrix are stored for the JF and JF-SIMP p_s methods. Table 1 summarizes the memory storage estimates per element for each method.

Table 2 gives details of the asymptotic operation counts per element of the main operations for different methods. For the sake of simplicity, we only indicate the leading orders of magnitude for the construction of explicit residuals and implicit matrix components and assume $\alpha_{1 \leq i \leq 3} = \mathcal{O}(1)$. Indeed, we are only interested in relative costs for one method compared to another. The number of quadrature points in an element is denoted N_g and is assumed to scale with p^d when p is large. Several conclusions may be inferred from this analysis. First, as observed in precedent experiments, the construction of the implicit matrix represents the most expensive operation [14]. We note that the JF and JF-SIMP p_s methods lower its cost by avoiding the construction of the off-diagonal blocks. The SIMP p_s method requires only a fraction N_{p_s}/N_p of these blocks and therefore constitutes an intermediate operation count. Second, the different stages involved in the linear solver constitute another expensive operations. We note that the SIMP p_s and JF-SIMP p_s methods only require the resolution of a linear problem of reduced size $NN_{eq}N_{p_s}$ compared to $NN_{eq}N_p$ for the FULL and JF methods. As a result, the factorization of the implicit matrix is reduced by a factor $(N_p/N_{p_s})^3$ at each time step. Likewise, the repeated matrix-vector products executed iteratively to solve the linear system (12) are reduced by a factor $(N_p/N_{p_s})^2$ for the SIMP p_s method. The JF-SIMP p_s uses a second construction of the residuals for the matrix-vector product as for the JF method, but keeps benefits from the SIMP p_s method for the preconditioning step. Finally, the reconstruction involved in the SIMP p_s and JF-SIMP p_s methods are of moderate costs. The most expensive operations are the matrix-vector product $\hat{\mathbf{U}}\Delta\hat{\mathbf{V}}^{(n)}$ and the triangular solve $(\hat{\mathbf{L}} + \hat{\mathbf{D}})^{-1}\Delta\hat{\mathbf{V}}^{(n+1)}$ for the evaluation of high-order DOFs.

We conclude that both JF and SIMP p_s methods allow important memory and flop savings in the implicit matrix construction and the iterative solution of the linear problem, respectively. The JF-SIMP p_s method using both strategies, one can expect to obtain more significant savings and hence better performances in terms of problem size and execution time.

Remark 1. The present analysis of memory and operation counts assumes a block-Jacobi preconditioning. This method is suitable for time-dependent solutions, but is no longer adapted to steady-state computations [17, 33]. More involved preconditioning techniques such as ILU(0) would require the construction of the full matrix as well as its

factorization. The operation counts of the JF method would hence be increased [14]. In contrast, the SIMP_{p_s} and JF-SIMP_{p_s} methods need a preconditioning matrix of lower size for the problem (18) and their operation counts in Table 2 are therefore less sensitive to the choice of the preconditioner.

3. Numerical experiments

In this section we present numerical experiments for the Navier-Stokes equations (1) in two-space dimensions to illustrate the performances of the time integration strategies derived in §§ 2.3 and 2.4. In order to evaluate the efficiency of the different strategies, we define the speedup as the ratio of CPU time needed to reach a given number of iterations with the FULL method to CPU time for convergence with the three other methods:

$$\text{speedup} = \frac{\text{CPU time(FULL)}}{\text{CPU time(method)}}.$$

In practice, we evaluate the speedup based on 500 time iterations. In order to assess the robustness of the methods, tests are based on solutions restarted at a physical time where gradients in the flow are fully developed. All simulations presented in this work have been carried out using CFL values respectively given by $CFL = 0.5, 0.3, 0.2, 0.1, 0.05$ for $p = 1, 2, 3, 4, 5$ and $CFL = 0.05$ for $p = 6$. The time step is then limited according to

$$\Delta t = CFL \times \min \left\{ \frac{h_\kappa}{|\mathbf{v} \cdot \mathbf{n}| + \sqrt{\gamma p / \rho}}, \kappa \in \Omega_h \right\},$$

where h_κ denotes the diameter of the inscribed circle of the triangular element κ . For the linear solver, we use a restarted GMRES method with a Krylov subspace of 10 eigenvectors and stop the linear iterations when the relative residual convergence is below 10^{-5} . A restart of the solver is applied after a maximum number of 20 iterations.

Following the analysis done in Ref. [17], we can evaluate practical memory savings when using the different methods. The majority of the memory is used for the storage of the implicit matrix and the Krylov vectors (see section 3.4 in [17] and references cited therein) thus leading to a ratio of memory lower than 20% between the memory requirement to store the Krylov vectors to the memory required by the implicit matrix. This indicates practical memory savings between 62% and 73% when using the JF and JF-SIMP0 methods, and between 34% and 70% when using the SIMP0 method, for polynomial degrees $1 \leq p \leq 6$ as used in the present experiments. These memory savings may allow the simulation of problems of larger size on a given machine.

3.1. Wall vortex interaction problem

We first consider the normal collision of a counter-rotating vortex pair with a non-slip boundary [11, 25]. In this example, we let $\Omega = [-1, 1]^2$ and look for solutions of problem (1) with initial condition

$$\begin{aligned} \rho(\mathbf{x}, 0) &= 1, \\ v_x(\mathbf{x}, 0) &= \frac{\omega_e}{2} \left(- (y - y_1) e^{-(r_1/r_0)^2} + (y - y_2) e^{-(r_2/r_0)^2} \right), \\ v_y(\mathbf{x}, 0) &= \frac{\omega_e}{2} \left((x - x_1) e^{-(r_1/r_0)^2} - (x - x_2) e^{-(r_2/r_0)^2} \right), \\ p(\mathbf{x}, 0) &= \frac{1}{\gamma M_\infty^2}, \end{aligned}$$

where r_1 and r_2 denote distances to the vortex axes $(x_1, y_1) = (0, 0.1)$ and $(x_2, y_2) = (0, -0.1)$; $r_0 = 0.1$ and $\omega_e = 299.528385375226$ are the radius and strength of both vortices. We apply non-slip adiabatic condition at the left and right boundaries, and subsonic inflow/outflow condition at the top and bottom boundaries. The Mach and Reynolds numbers, based on the half width of the domain and an average velocity scale, are $M_\infty = 10^{-2}$ and $Re = 10^3$. The average velocity scale is determined by setting the global kinetic energy of the flow

$$E(t) = \frac{1}{2} \int_{\Omega} \|\mathbf{v}\|_2^2 dV \quad (23)$$

to an initial value $E(0) = 2$. We use partitions of Ω built from a grid of m -by- m squares split into a total of $N = 2m^2$ triangles (see Fig. 3). Tests are carried out on uniform meshes with isosceles triangles (Fig. 3a) and refined meshes in the neighborhood of the right boundary where the collision occurs (Fig. 3b).

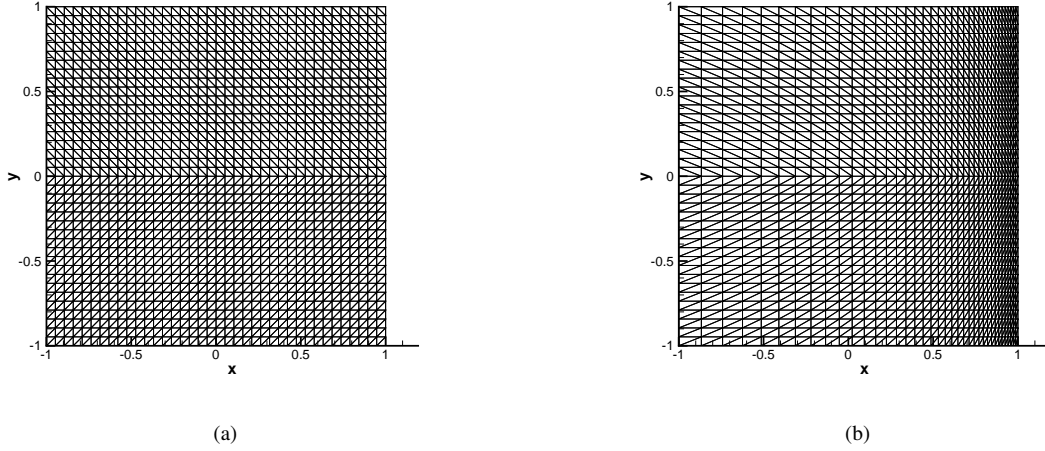


Figure 3: Meshes for the wall vortex interaction problem: (a) uniform mesh with $N = 2888$ elements, (b) stretched mesh with $N = 2888$.

Fig. 4 presents the contour plots of the scalar vorticity field $\omega = \partial_x v_y - \partial_y v_x$ at a physical time $t = 0.8$ obtained with the four different time integration methods on a given uniform mesh with $N = 2888$ elements and a global size of $h = 3.08 \times 10^{-2}$. The solution is post-processed by subdividing each triangle into nine equal sub-triangles. We observe that all results compare qualitatively well. Secondary vortex patches near the wall produced by interaction with the wall are in particular well resolved.

A quantitative comparison is given in Fig. 5(a) where the temporal evolution of the global kinetic energy (23) is plotted in the early stage of the wall vortex interaction $t \leq 0.8$. Results are compared to a well resolved reference solution obtained by [25] using a fast Fourier spectral method with 1024×2048 modes in the x - and y -directions. For comparison, present computations are obtained on a uniform mesh with $N = 2888$ triangles and polynomial degree $p = 4$ corresponding to a global number of $N \times N_p = 43320$ DOFs. We observe that the present results compare well with the reference data. Moreover, all present results clearly coincide and a further analysis of the accuracy of different integration methods is given in Fig. 6 where the L_1 -norm of the error on the axial momentum $e_h = (\rho u)_h - (\rho u)_{ref}$ between each solution and a reference solution is plotted for different mesh refinements and polynomial degrees. The reference solution is defined by an integration with $p = 4$ and a low time step $CFL = 10^{-2}$. We observe that the error converges with the expected convergence rate $|e_h|_1 \sim h^{p+1}$ until a given level $|e_h|_1 \simeq 3 \times 10^{-3}$ where the convergence rates deteriorate. This may be due either to the accuracy of the reference solution, or to the limited convergence rate of the error under time step refinement of the present method. The error levels are however seen to strongly decrease with p . Present results also confirm that there is no alteration of the accuracy by the different time integration strategies. We also refer to [9] for an in-depth analysis of the accuracy of DG methods under hp -refinement for this test case.

Fig. 5(b) presents the evolutions of the kinetic energy as a function of CPU time. The FULL implicit discretization takes the largest CPU time to reach the final time $t = 0.8$. Compared to other strategies, the JF method is 1.64 times faster, the SIMP0 is 1.82 times faster and comparable to the time integration obtained with the JF-SIMP0 method with a speedup of 1.81.

Remark 2. All experiments with the simplified methods $\text{SIMP}p_s$ and $\text{JF-SIMP}p_s$ have been carried out using $p_s = 0$. Solving a simplified implicit problem for the first DOFs only was seen to perform well in every computations up to $p = 6$. We further note that the first DOFs correspond to the mean value of variables \mathbf{u}_h evaluated over the elements. We refer to Table 5 and associated comments for an analysis of the value of p_s on the performances of the methods.

An in-depth analysis of the speedup has been done in Tables 3 and 4 where we evaluate the execution time of

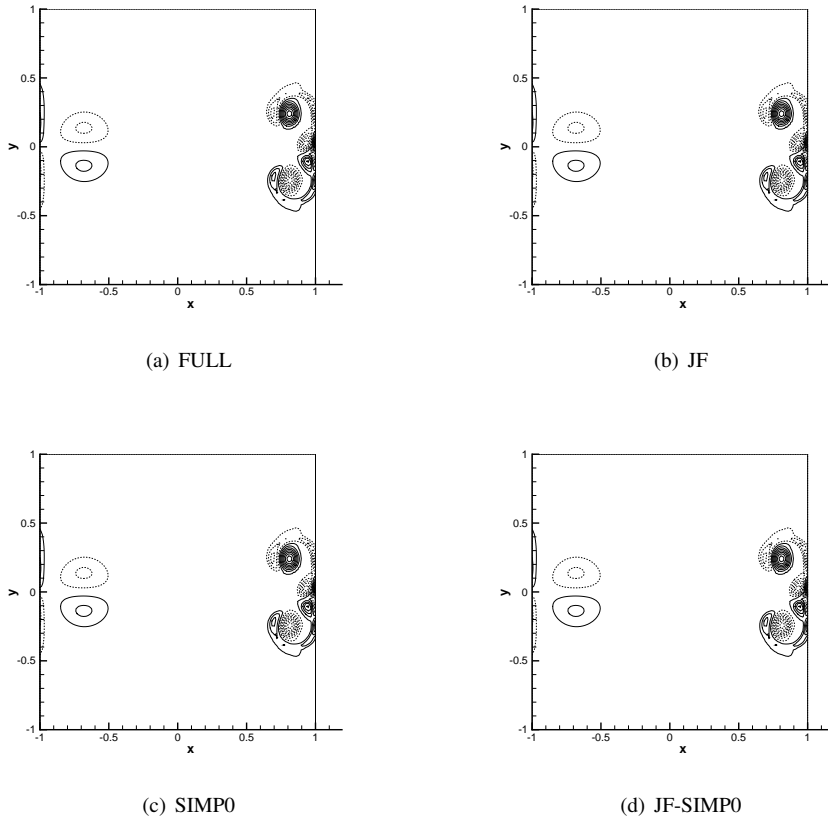


Figure 4: Wall vortex interaction problem: contour plots of the vorticity field with twenty height equally spaced values from -270 to 270 . Dashed lines correspond to negative values.

each method as a function of the polynomial degree, size and anisotropy of the mesh. Table 3 shows that the saving in CPU time is seen to increase with the polynomial degree for all of the three methods except for the JF method and finer mesh where it decreases with p . The gain is always larger with the SIMP0 and JF-SIMP0 methods for moderate and large polynomial degrees $p \geq 2$. In contrast, the speedup values tend to decrease with the number of elements N for the SIMP0 method while they increase with the JF method. The association of both methods allows to keep the acceleration of the computation with both p and N and results in noticeable cost savings for this application. Table 4 presents the speedups computed for a fixed mesh size when varying the mesh anisotropy and polynomial degree. Meshes are refined in the neighborhood of the right boundary using a geometric progression. We aim at assessing the behavior and robustness of the different methods for the simulation of regions of flow with large gradients such as boundary layers which require the use of highly stretched meshes. We thus define the aspect-ratio of a mesh as the ratio of maximum to minimum cell sizes in the domain, $AR = \frac{1}{h} \max\{h_\kappa : \kappa \in \Omega_h\}$, which provides a measure of the anisotropy of the mesh. For instance, meshes in Figs. 3(a) and 3(b) correspond to $AR = 1$ and $AR = 4$, respectively. The aspect-ratio of each mesh is indicated in Table 4 together with the value of the power of the geometric progression used for the refinement. The results are almost insensitive to the mesh aspect-ratio and highlight the robustness of the present methods on anisotropic meshes. In conclusion, this speedup analysis shows that the JF-SIMP0 method allows the better speedup of the computations evaluated over the whole range of mesh properties and approximation order, thereby indicating that the method allows better efficiency over a wide range of problems.

The effect of the coupling level between low- and high-order DOFs on the performances of the $SIMP_{p_s}$ and $JF-SIMP_{p_s}$ methods is illustrated in Table 5 where the value of p_s varies for two different values of the polynomial degree p . Results highlight different behaviors between both methods. Indeed, the speedup of the $SIMP_{p_s}$ method decreases

Table 3: Wall vortex interaction problem: speedups observed from experiments with uniform meshes.

| p | N | JF | SIMP0 | JF-SIMP0 |
|-----|-------|------|-------|----------|
| 1 | 648 | 1.09 | 1.29 | 1.18 |
| 2 | 648 | 1.29 | 1.55 | 1.44 |
| 3 | 648 | 1.41 | 1.65 | 1.54 |
| 4 | 648 | 1.50 | 1.72 | 1.67 |
| 5 | 648 | 1.55 | 1.75 | 1.72 |
| 6 | 648 | 1.60 | 1.77 | 1.76 |
| 1 | 2888 | 1.43 | 1.11 | 1.50 |
| 2 | 2888 | 1.38 | 1.37 | 1.50 |
| 3 | 2888 | 1.45 | 1.58 | 1.60 |
| 4 | 2888 | 1.53 | 1.70 | 1.69 |
| 1 | 10952 | 2.09 | 1.03 | 2.15 |
| 2 | 10952 | 1.78 | 1.19 | 1.89 |
| 3 | 10952 | 1.62 | 1.44 | 1.79 |
| 4 | 10952 | 1.62 | 1.73 | 1.90 |

Table 4: Wall vortex interaction problem: speedups observed from experiments with stretched meshes and a fixed number of elements $N = 2888$.

| p | power | AR | JF | SIMP0 | JF-SIMP0 |
|-----|-------|-----|------|-------|----------|
| 1 | 1 | 1 | 1.43 | 1.11 | 1.50 |
| 2 | 1 | 1 | 1.38 | 1.37 | 1.50 |
| 3 | 1 | 1 | 1.45 | 1.58 | 1.60 |
| 4 | 1 | 1 | 1.53 | 1.70 | 1.69 |
| 1 | 10 | 4 | 1.45 | 1.10 | 1.49 |
| 2 | 10 | 4 | 1.38 | 1.37 | 1.50 |
| 3 | 10 | 4 | 1.46 | 1.58 | 1.61 |
| 4 | 10 | 4 | 1.53 | 1.70 | 1.69 |
| 1 | 100 | 20 | 1.44 | 1.11 | 1.52 |
| 2 | 100 | 20 | 1.38 | 1.37 | 1.50 |
| 3 | 100 | 20 | 1.46 | 1.59 | 1.61 |
| 4 | 100 | 20 | 1.53 | 1.70 | 1.69 |
| 1 | 1000 | 143 | 1.39 | 1.11 | 1.46 |
| 2 | 1000 | 143 | 1.37 | 1.35 | 1.47 |
| 3 | 1000 | 143 | 1.46 | 2.01 | 1.67 |
| 4 | 1000 | 143 | 1.53 | 1.70 | 1.69 |
| 1 | 5000 | 600 | 1.45 | 1.10 | 1.51 |
| 2 | 5000 | 600 | 1.40 | 1.31 | 1.45 |
| 3 | 5000 | 600 | 1.45 | 1.67 | 1.67 |
| 4 | 5000 | 600 | 1.52 | 1.80 | 1.79 |

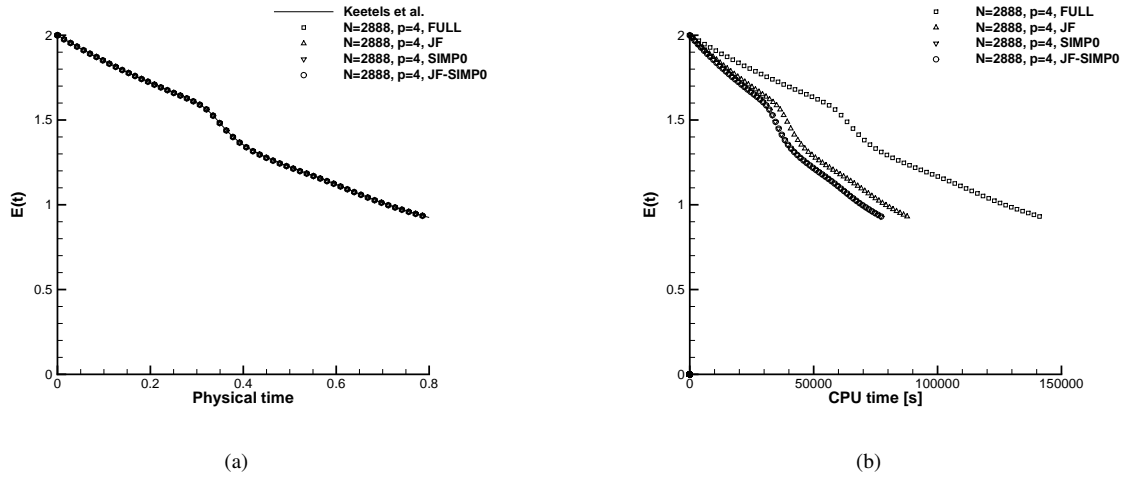


Figure 5: Wall vortex interaction problem: evolutions of the total kinetic energy as a function of physical time (a) and CPU time (b). The present computations (symbols) are compared to the reference solution from [25] (lines).

Table 5: Wall vortex interaction problem: speedups observed from experiments when varying p_s with a fixed number of elements $N = 2888$.

| p | p_s | $\text{SIMP}p_s$ | $\text{JF-SIMP}p_s$ |
|-----|-------|------------------|---------------------|
| 5 | 0 | 1.88 | 1.86 |
| 5 | 1 | 1.77 | 1.87 |
| 5 | 2 | 1.59 | 1.85 |
| 5 | 3 | 1.40 | 1.82 |
| 5 | 4 | 1.19 | 1.75 |
| 6 | 0 | 2.01 | 1.95 |
| 6 | 1 | 1.91 | 1.99 |
| 6 | 2 | 1.77 | 1.98 |
| 6 | 3 | 1.61 | 1.97 |
| 6 | 4 | 1.42 | 1.94 |
| 6 | 5 | 1.23 | 1.88 |

when p_s tends to p , while the speedup of the $\text{JF-SIMP}p_s$ method is almost insensitive to p_s . We underline that the former method requires the construction of an increasing number of components of the off-diagonal blocks when p_s is increased, while the latter does not. The large number of operation required by the matrix construction is thus expected to affect the performances of the $\text{SIMP}p_s$ method when p_s is increased. In contrast, the $\text{JF-SIMP}p_s$ method is seen to be more robust with respect to p_s .

Table 6 presents the speedup values observed from experiments for the main steps of the numerical resolution of the implicit problem. These steps correspond to the construction of the implicit matrix \mathbf{A} , to its factorization to get the preconditioning matrix \mathbf{P} , and to the resolution of the linear system. For this last step, we also add the CPU time used for the reconstruction when solving the simplified implicit problem (see Table 2). First, the speedup for constructing the implicit matrix is independent of the polynomial degree and is equal to $1 + \alpha_3/\alpha_2$ when using the JF and $\text{JF-SIMP}p_s$ methods as expected from Table 2. The experimental speedup confirm this trend and indicate a value $\alpha_3/\alpha_2 \approx 0.77$. The asymptotic speedup is $(1 + \alpha_3/\alpha_2)/(1 + N_{p_s}\alpha_3/N_p\alpha_2) \approx 1 + \alpha_3/\alpha_2$ for the $\text{SIMP}p_s$ method for large p values and a fixed p_s value. Results in Table 6 are again in good agreement with the theoretical prediction. Second, the JF method applies the full factorization of the preconditioning matrix and no speedup for the step “form \mathbf{P} ” is expected. In contrast, large speedups are observed for the $\text{SIMP}p_s$ and $\text{JF-SIMP}p_s$ methods where results suggest a speedup proportional to $(N_p/N_{p_s})^3$ for large p . Finally, the last comparison in Table 6 confirms the larger speedups

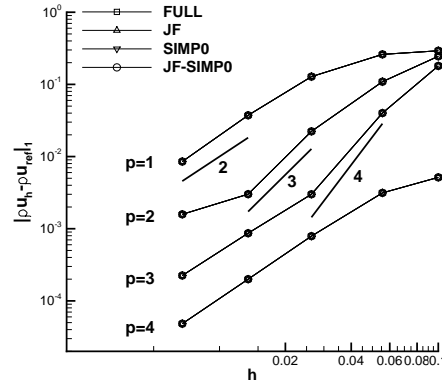


Figure 6: Wall vortex interaction problem: convergence of the different methods under mesh refinement for different polynomial degrees $1 \leq p \leq 4$. Thick lines correspond to indicated convergence rates.

Table 6: Wall vortex interaction problem: speedups of different steps of the implicit problem resolution observed from experiments with uniform meshes and a fixed number of elements $N = 2888$.

| p | form A | | | form P | | | solver + reconstruction | | |
|-----|---------------|-------|----------|---------------|-------|----------|-------------------------|-------|----------|
| | JF | SIMP0 | JF-SIMP0 | JF | SIMP0 | JF-SIMP0 | JF | SIMP0 | JF-SIMP0 |
| 1 | 1.55 | 1.19 | 1.58 | 1.00 | 4.67 | 5.40 | 1.49 | 1.13 | 1.66 |
| 2 | 1.68 | 1.42 | 1.67 | 0.97 | 22.6 | 20.5 | 1.33 | 1.66 | 1.68 |
| 3 | 1.76 | 1.54 | 1.70 | 1.00 | 68.6 | 68.6 | 1.30 | 2.74 | 1.87 |
| 4 | 1.72 | 1.61 | 1.71 | 0.97 | 210 | 189 | 1.28 | 4.93 | 2.29 |
| 5 | 1.71 | 1.66 | 1.71 | 0.96 | 466 | 439 | 1.28 | 8.28 | 2.73 |
| 6 | 1.77 | 1.73 | 1.76 | 1.00 | 1068 | 1091 | 1.36 | 14.3 | 3.48 |

obtained with the SIMP_{p_s} method for the linear system resolution. The theoretical speedup is $N_{eq}N_p/3$ for large p values, while the experimental values indicate an behavior βN_p with $0.35 \leq \beta \leq 0.5$ comparable to the theoretical value $\beta = 4/3$.

3.2. Flow around a NACA0012 airfoil

We now consider the laminar flow around a NACA0012 airfoil in subsonic condition with a Mach number $M_\infty = 0.5$ and a Reynolds number $Re = 5 \times 10^3$ based on the chord length and freestream velocity. The angle of attack is set at 2° . We use three meshes with $N = 1683, 4298$ and 7368 curved triangles represented by quartic polynomials using Lagrange interpolation; and 134, 91 and 120 nodes on the airfoil, respectively (see Fig. 7). The global aspect-ratio of the meshes are $AR = 444, 2125$ and 1685 , respectively. The medium and fine meshes are refined in the wake for a good capture of the vortex shedding. The boundaries of the computational domain are set at a distance of about 20 chord lengths from the airfoil. The initial condition is a uniform flow $(\rho, v_x, v_y, p) = (1, 1, 0, 1/\gamma M_\infty^2)$. For all computations of open flows we apply non-slip and adiabatic wall boundary condition on the walls and subsonic inflow/outflow boundary condition at the far-field boundaries of the computational domain.

Contours of vorticity obtained with $p = 3$ on the medium mesh are shown in Fig. 8 at a physical time $t = 10$ for each time integration method and highlight the instability of the wake. Results from different tests agree very well. A quantitative comparison is given in Fig. 9 where the temporal evolution of the total drag coefficient in the transient regime $t \leq 10$ is indicated as a function of physical and CPU times. We observe again no discrepancies between different results, thereby confirming that the different strategies do not alter the temporal accuracy of the simulation. The fastest convergence is achieved using the JF-SIMP0 method where a speedup of 1.89 is obtained. The SIMP0 method also performs well with a speedup values of 1.78, whereas the JF method leads to a lower speedup of 1.55.

Table 7: NACA0012 airfoil problem: speedups observed from experiments.

| p | N | JF | SIMP0 | JF-SIMP0 |
|-----|------|------|-------|----------|
| 1 | 1683 | 1.23 | 1.17 | 1.32 |
| 2 | 1683 | 1.28 | 1.58 | 1.53 |
| 3 | 1683 | 1.39 | 1.74 | 1.64 |
| 4 | 1683 | 1.50 | 1.80 | 1.73 |
| 1 | 4298 | 1.77 | 1.25 | 2.11 |
| 2 | 4298 | 1.55 | 1.61 | 2.02 |
| 3 | 4298 | 1.57 | 1.79 | 1.98 |
| 4 | 4298 | 1.58 | 1.94 | 1.93 |
| 5 | 4298 | 1.56 | 1.90 | 1.86 |
| 6 | 4298 | 1.60 | 1.89 | 1.88 |
| 1 | 7368 | 2.06 | 1.23 | 2.46 |
| 2 | 7368 | 1.77 | 1.54 | 2.34 |
| 3 | 7368 | 1.67 | 1.69 | 2.15 |
| 4 | 7368 | 1.58 | 1.84 | 1.93 |

Table 7 gives details on the speedups obtained as a function of the polynomial degree on the three different meshes. The speedups are evaluated from 500 iterations starting from a physical time $t = 50$ where the wake instability is well developed. Clearly, the JF-SIMP0 method outperforms other methods except for large polynomial degrees $p \geq 4$ where speedups are comparable to those obtained from the SIMP0 method. The latter method is the most sensitive to the polynomial degree. We recall that the computational cost of the SIMP0 method has been shown to be less expensive than the FULL method by a theoretical asymptotic factor $(N_p/N_{p_s})^2$ for large N and using a biconjugate gradient method as linear solver [37]. The analysis in § 2.4.3 also indicates a strong reduction of the operation counts. In contrast, the JF method seems to be less sensitive to p and the speedups increase with N . The JF-SIMP0 method is seen to take advantage of both techniques and results in acceleration of the computations over the whole range of numerical conditions.

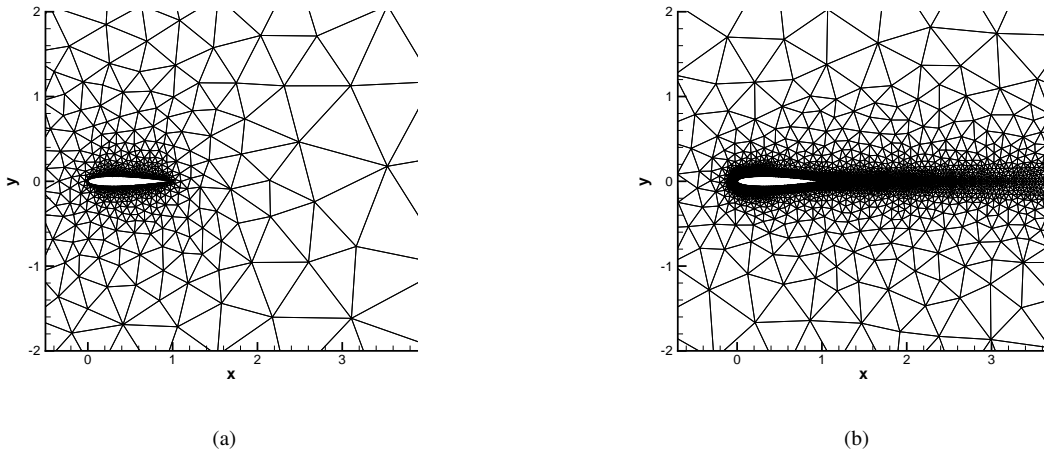


Figure 7: Coarse ($N = 1683$) and fine ($N = 7368$) meshes for the NACA0012 airfoil problem.

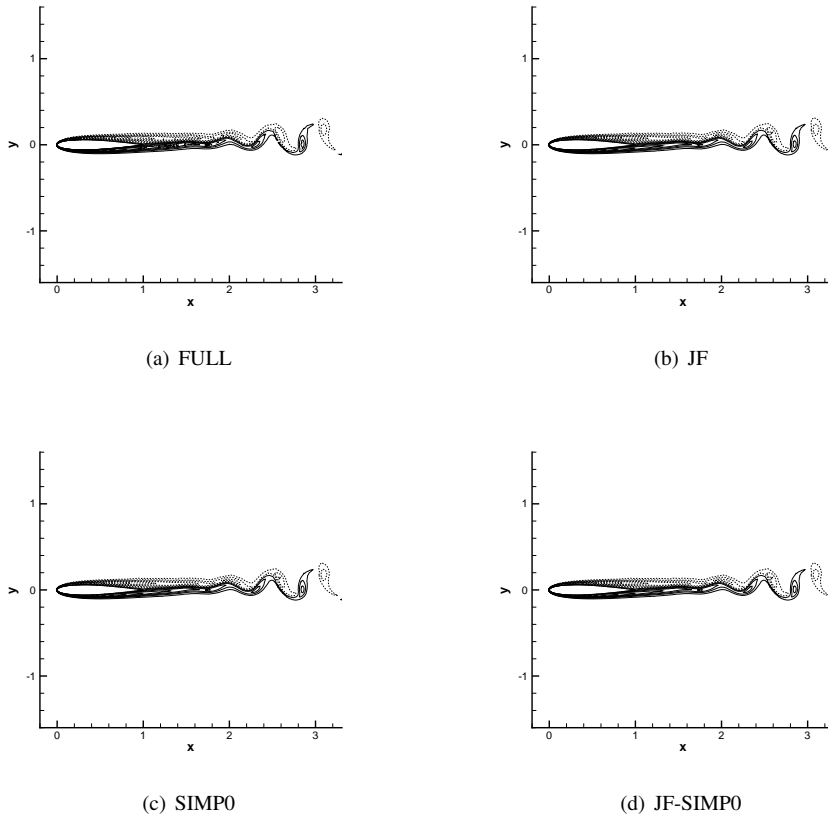


Figure 8: NACA0012 airfoil problem: contour plots of the vorticity field with twenty height equally spaced values from -50 to 50 .

3.3. Flow around a SD7003 airfoil

The next numerical application is also a laminar flow around an airfoil. We consider the SD7003 airfoil in subsonic conditions with a Mach number $M_\infty = 0.2$ and a relatively high Reynolds number $Re = 10^4$ based on the chord length c and freestream velocity. The angle of attack is set at 4° . Under these conditions, the flow that develops around the SD7003 wing is mainly 2D and only small variations occur along the spanwise direction [46], which motivates the present 2D computation. In addition, the flow is characterized by a laminar separation bubble which forms over the airfoil and an associated vortex shedding in the wake. The phenomena associated to this flow are difficult to capture and have been the subject of recent experimental and computational investigations [30, 46, 47]. The airfoil has a maximum thickness of $0.085c$, a maximum camber of $0.0145c$ at $x/c = 0.35$, and a rounded trailing edge with a radius $r/c \simeq 4 \times 10^{-4}$. The far-field boundaries are located at a distance of approximately 20 chords. We use a mesh containing 11676 triangular elements and 158 grid points on the surface of the airfoil for a global aspect ratio of 5190. This mesh is refined near the wall and in the regions of separation and vortex shedding. In addition, high-order boundary representation is used with polynomials of degree 4.

Typical computed Mach number and vorticity contours in the flow field are shown in Fig. 10 at $t = 50$ using polynomials of degree $p = 4$. The streamlines are also drawn and highlight the development of the boundary layer and its separation on lee side of airfoil. One can see that the flow features, such as separation of the flow on the upper surface of the airfoil and periodic shedding of pairs of counter-rotating vortices, are well captured. Notice how the accurate spatial discretization yields a solution that captures the vortex propagation into the wake with high fidelity. Moreover, the streamlines highlight the development of the boundary layers and flow separation on the upper part of the airfoil. All these results compare well with those reported in the literature [30, 46].

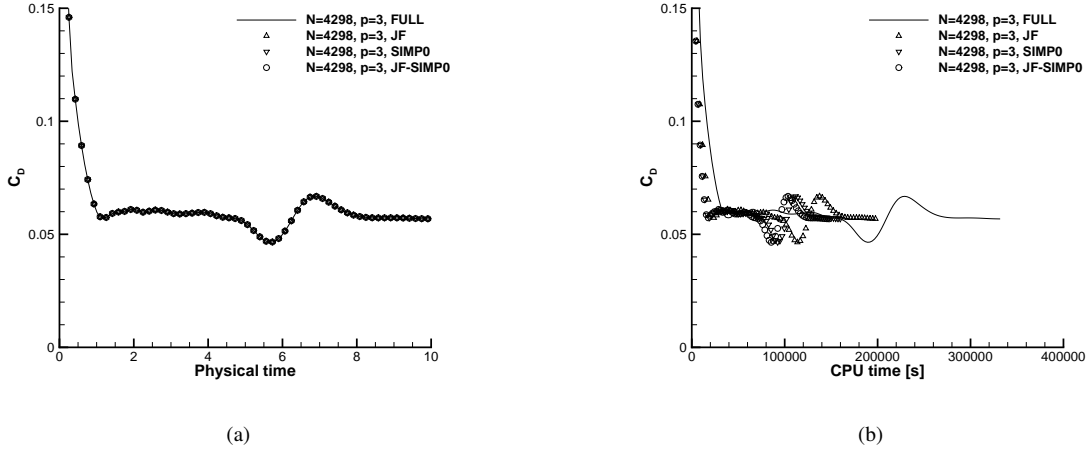


Figure 9: NACA0012 airfoil problem: evolution of the total drag coefficient as a function of physical time (a) and CPU time (b).

Table 8: SD7003 airfoil problem: speedups observed from experiments.

| p | JF | SIMP0 | JF-SIMP0 |
|-----|------|-------|----------|
| 1 | 2.33 | 1.27 | 3.06 |
| 2 | 2.03 | 1.40 | 2.69 |
| 3 | 1.82 | 1.75 | 2.54 |
| 4 | 1.72 | 1.90 | 2.34 |
| 5 | 1.74 | 2.02 | 2.13 |
| 6 | 1.79 | 2.11 | 2.14 |

Table 8 shows the speedups obtained as function of the polynomial degree. We start the computations at a physical time $t = 50$. We observe again the same trends obtained from precedent test cases. The JF method achieves a reduction in CPU time greater than the SIMP0 till $p = 3$ with a corresponding speedup of 1.82. However, the speedups for the JF method tend to decrease with p and those for the SIMP0 method tend to increase, hence the latter method becomes faster for $p \geq 4$. Finally, the JF-SIMP0 method outperforms the other with very large speedup values. The performances of this method tend to slightly decrease with p and seem to reach a plateau for $p \geq 5$.

3.4. Flow past a circular cylinder

Finally, we consider the flow past a circular cylinder in subsonic conditions with a Mach number $M_\infty = 0.2$ and a Reynolds number $Re = 150$ based on the cylinder diameter and freestream velocity. This test case is known to generate a spontaneous unsteadiness under the form of a periodic vortex shedding and is therefore extensively studied in the literature. In particular, at this Reynolds number value, the flow is essentially 2D with periodic vortex pairs shed from the downstream side of the cylinder. We refer to [48] and references cited therein for a review of physical mechanisms at various Reynolds number regimes. The far-field boundaries are set at a distance of thirty diameters from the cylinder. The initial condition is a uniform flow $(\rho, v_x, v_y, p) = (1, 1, 0, 1/\gamma M_\infty^2)$. One mesh consisting of $N = 7384$ curved quadratic triangular elements and 93 grid points on the surface of the cylinder for a global aspect ratio of $AR = 630$ was used for the computation. This mesh is refined near the wall and in the wake where important phenomena occur such as flow separation and vortex shedding.

Mach number contours obtained with $p = 3$ and the FULL method at time $t = 150$ are shown Fig. 11 where the von-Karman vortex street is well developed. Fig. 12 presents the temporal evolution of the pressure and total drag coefficients integrated over the cylinder obtained with third-order polynomials. Results are in good agreement and no differences are observed between the different integration strategies even over a long time period. Some global

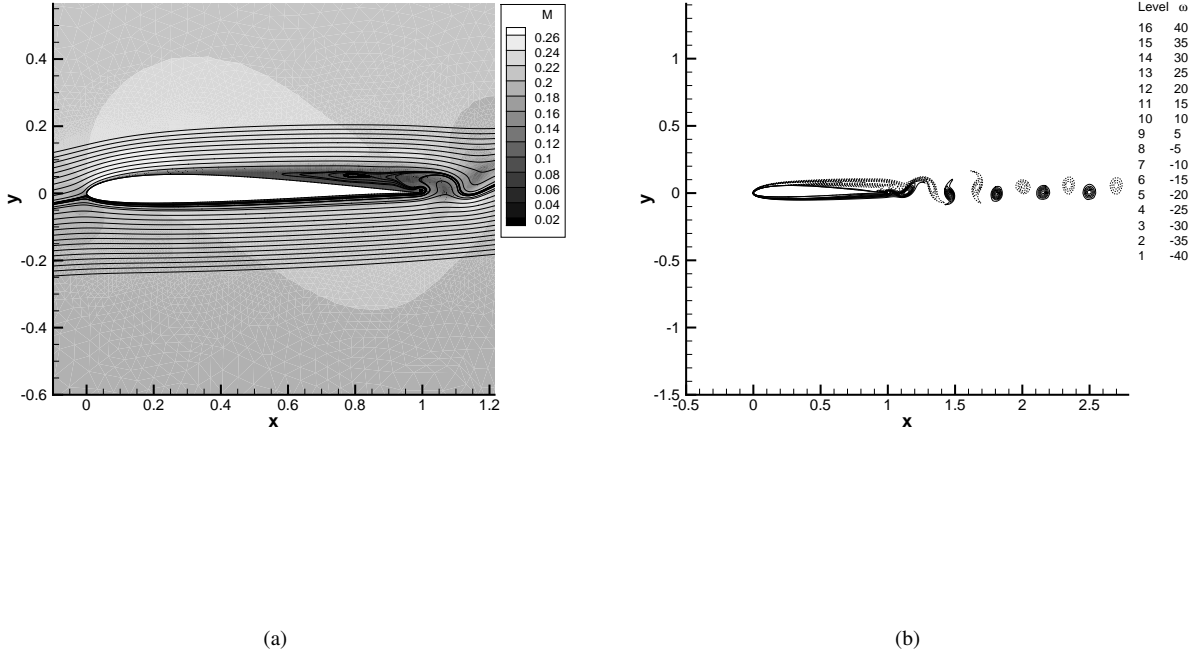


Figure 10: SD7003 airfoil problem: contour plots of local Mach number (a) and vorticity (b).

Table 9: Flow past a circular cylinder: speedups observed from experiments.

| p | JF | SIMP0 | JF-SIMP0 |
|-----|------|-------|----------|
| 1 | 2.16 | 1.05 | 2.47 |
| 2 | 1.91 | 1.55 | 2.62 |
| 3 | 1.73 | 1.78 | 2.34 |

quantities can be evaluated from these results and compared to the literature. We obtain a mean value of the pressure drag coefficient computed over the time range $100 \leq t \leq 120$ of 1.064 very close to experimental results from [39]. Likewise, the Strouhal number evaluated from the time period of pressure fluctuations, cylinder diameter and freestream velocity is $St = 0.183$ and is also in good agreement with the experimental observations of $St = 0.182$ in [39] or numerical evaluation of $St = 0.183$ in [45] for this Reynolds number.

The speedups for $1 \leq p \leq 3$ are summarized in Table 9 where the computations are restarted from a solution at a physical time $t = 150$. Results are very similar to all previous test cases and confirm that the JF-SIMP0 method presents the highest speedup. Here, we again conclude that the present method is therefore an interesting alternative to full implicit time integration of the viscous terms.

4. Concluding remarks

The $SIMP_{p_s}$ method for time integration procedure introduced in [37] has been successfully applied and extended to the resolution of time-dependent nonlinear systems of conservation laws. We have used an explicit-implicit time discretization with a splitting method in which the convective term is treated explicitly and the diffusive term implicitly

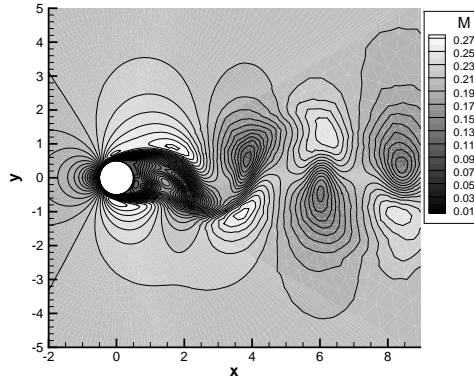


Figure 11: Flow past a circular cylinder: contour plots of local Mach number at $t = 50$.

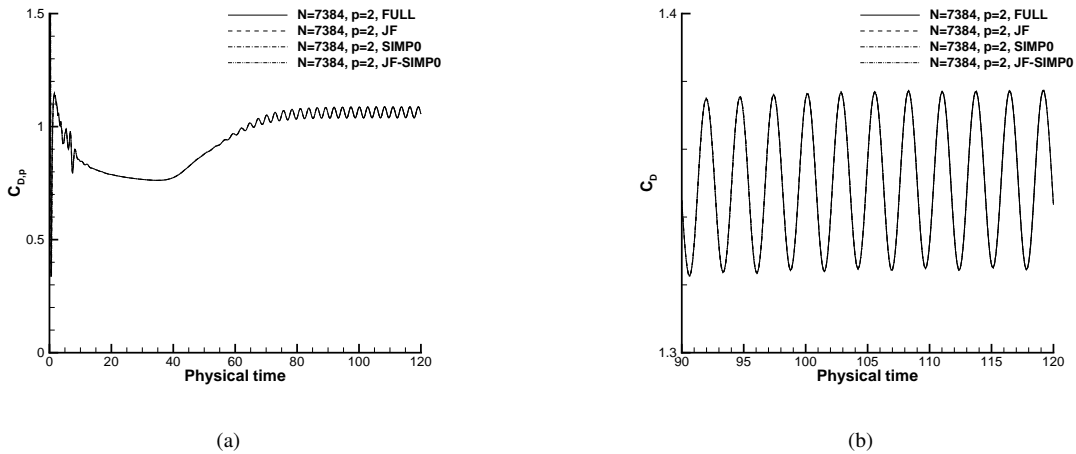


Figure 12: Flow past a circular cylinder: temporal evolutions of pressure (a) and total (b) drag coefficients.

as in [38]. For the implicit time integration we use an iterative algorithm with a Newton-Krylov method and a block-Jacobi preconditioner. The implicit method is here accelerated by combining the SIMP_{p_s} and JF methods to get the JF-SIMP_{p_s} method. The SIMP_{p_s} method is based on a partial coupling of variables in neighboring elements. It allows the resolution of a linear system of lower size for low-order DOFs which is solved with the Newton-Krylov JF method. An explicit local reconstruction is then applied to evaluate higher-order DOFs. Earlier analysis and experiments showed that the original SIMP_{p_s} method becomes efficient when the degree of low-order DOFs p_s becomes low compared to the polynomial degree of the numerical approximation. The JF-SIMP_{p_s} method is however seen to improve the performances of the implicit discretization by keeping advantages of both methods. First, the method only requires the storage of the diagonal blocks in the implicit matrix for preconditioning purposes as in the JF method. This also avoids the construction of the off-diagonal blocks, thus reducing the operation counts of the method. Then, the linear system to be solved at each time step is reduced in size by the simplification SIMP_{p_s} thus lowering the costs of both the factorization and inversion of the preconditioning matrix as well as of the system inversion. In contrast to the full implicit and JF methods, the present JF-SIMP_{p_s} method requires the reconstruction of higher-order DOFs, but it induces relatively moderate extra operation count.

In terms of CPU time, numerical experiments on the compressible Navier-Stokes equations and convection-

dominated flow problems indicate an acceleration of computations with the present method compared to a full implicit discretization of the diffusive terms. The method also outperforms the $SIMP_{p_s}$ and JF methods for a wide range of numerical conditions of mesh size, mesh anisotropy and approximation order, thereby confirming that the method is well suited to a broad range of problems. Future work will consider the adaptation of the present method to the time implicit discretization of convective terms.

Acknowledgments

The author would like to thank the authors from Ref. [25] for providing the reference data used in § 3.1. Jean-Baptiste Chapelier from Onera is warmly acknowledged for his help and fruitful discussions.

References

- [1] R. Abgrall and C.-W. Shu, Development of residual distribution schemes for the discontinuous galerkin schemes: the scalar case with continuous elements, *Commun. Comput. Phys.*, 5 (2009), pp. 376–390.
- [2] J.-B. Apoung Kanga and B. Després, CFL condition and boundary conditions for DGM approximation of convection-diffusion, *SIAM J. Numer. Anal.*, 44 (2006), pp. 2245–2269.
- [3] F. Bassi, L. Botti, A. Colombo, D.A. Di Pietro and P. Tesini, On the flexibility of agglomeration based physical space discontinuous Galerkin discretizations, *J. Comput. Phys.*, 231 (2012), pp. 45–65.
- [4] F. Bassi, A. Crivellini, S. Rebay and M. Savini, Discontinuous Galerkin solution of the Reynolds-averaged Navier-Stokes and $k-\omega$ turbulence model equations, *Comput. Fluids*, 34 (2005), pp. 507–540.
- [5] F. Bassi, S. Rebay, G. Mariotti, S. Pedinotti and M. Savini, A High-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows, In proceedings of the 2nd European Conference on Turbomachinery Fluid Dynamics and Thermodynamics, R. Decuyper, G. Dibelius (eds.), Antwerpen, Belgium, 1997.
- [6] F. Bassi and S. Rebay, Numerical evaluation of two discontinuous Galerkin methods for the compressible Navier-Stokes equations, *Int. J. Numer. Methods Fluids*, 40 (2002), pp. 197–202.
- [7] P. Birken, G. Gassner, M. Haas and C.-D. Munz, A new class of preconditioners for discontinuous Galerkin methods for unsteady 3D Navier-Stokes equations: ROBO-SGS, ECCOMAS 2012, 10-14 September 2012.
- [8] R. Biswas, K. Devine and J.E. Flaherty, Parallel adaptive finite element methods for conservation laws, *Appl. Numer. Math.*, 14 (1994) pp. 255–284.
- [9] J.-B. Chapelier, M. de la Llave Plata, F. Renac and E. Lamballais, Evaluation of a high-order discontinuous Galerkin method for the DNS of turbulent flows, submitted, February 2013.
- [10] G. Chavent and B. Cockburn, The local projection P^0P^1 -discontinuous Galerkin finite element method for scalar conservative laws, *M2AN Math. Model. Anal. Numer.*, 23 (1989), pp. 565–592.
- [11] H.J.H. Clercx and C.-H. Bruneau, The normal and oblique collision of a dipole with a no-slip boundary, *Comput. Fluids*, 35 (2006), pp. 245–279.
- [12] B. Cockburn and C. W. Shu, TVB Runge-Kutta local projection discontinuous Galerkin finite element method for scalar conservation laws II: general framework, *Math. Comp.*, 52 (1989), pp. 411–435.
- [13] B. Cockburn and C. W. Shu, Runge-Kutta discontinuous Galerkin methods for convection-dominated problems, *J. Sci. Computing*, 16 (2001), pp. 173–261.
- [14] A. Crivellini and F. Bassi, An implicit matrix-free discontinuous Galerkin solver for viscous and turbulent aerodynamic simulations, *Comput. Fluids*, 50 (2011), pp. 81–93.
- [15] V. Dolejši, Semi-implicit interior penalty discontinuous Galerkin methods for viscous compressible flows, *Commun. Comput. Phys.*, 4 (2008), pp. 231–274.
- [16] V. Dolejši, M. Holík and J. Hozman, Efficient solution strategy for the semi-implicit discontinuous Galerkin discretization of the Navier-Stokes equations, *J. Comput. Phys.*, 230 (2010), pp. 4176–4200.
- [17] L. T. Diosady and D. L. Darmofal, Preconditioning methods for discontinuous Galerkin solutions of the Navier-Stokes equations, *J. Comput. Phys.*, 228 (2009), pp. 3917–3935.
- [18] M. Dubiner, Spectral methods on triangles and other domains, *SIAM J. Sci. Comput.*, 6 (1991), pp. 345–390.
- [19] M. Dumbser, D. S. Balsara, E. F. Toro and C.-D. Munz, A unified framework for the construction of one-step finite volume and discontinuous Galerkin schemes on unstructured meshes, *J. Comput. Phys.*, 227 (2008), pp. 8209–8253.
- [20] K. J. Fidkowski, T. A. Oliver, J. Lu and D. L. Darmofal, p -Multigrid solution of high-order discontinuous Galerkin discretizations of the compressible Navier-Stokes equations, *J. Comput. Phys.*, 207 (2005), pp. 92–113.
- [21] A. Filimon and M. Dumbser and C.-D. Munz, Discontinuous Galerkin schemes with defect correction based on reconstruction: the $p_N p_M$ -approach, ECCOMAS CFD 2010, 14-17 June 2010.
- [22] R. Hartmann, M. Lukacova-Medvidova and F. Prill, Efficient preconditioning for the discontinuous Galerkin finite element method by low-order elements, *Appl. Numer. Math.* 59 (2009), pp. 1737–1753.
- [23] R. Hartmann, J. Held, T. Leicht, and F. Prill, Discontinuous Galerkin methods for computational aerodynamics - 3D adaptive flow simulation with the DLR PADGE code, *Aerospace Science and Technology*, 14 (2010), pp. 512–519.
- [24] P. Houston and E. Süli, hp -adaptive discontinuous Galerkin finite/element methods for first-order hyperbolic problems. *SIAM J. Sci. Comput.*, 23 (2001), pp. 1226–1252.

- [25] G.H. Keetels, U. D’Ortona, W. Kramer, H.J.H. Clercx, K. Schneider and G.J.F. van Heijst, Fourier spectral and wavelet solvers for the incompressible Navier-Stokes equations with volume-penalization: Convergence of a dipole-wall, *J. Comput. Phys.*, 227 (2007), pp. 919–945.
- [26] C. M. Klaij, J. J. W. van der Vegt and H. van der Ven, Space-time discontinuous Galerkin method for the compressible Navier-Stokes equations, *J. Comput. Phys.*, 217 (2006), pp. 589–611.
- [27] D. A. Knoll and D. E. Keyes, Jacobian-free Newton-Krylov methods: a survey of approaches and applications, *J. Comput. Phys.*, 193 (2004), pp. 357–397.
- [28] N. Kroll, H. Bieler, H. Deconinck, V. Couaillier, H. van der Ven and K. Sorensen (eds.), ADIGMA - A european initiative on the development of adaptive higher-order variational methods for aerospace applications, *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, 113 (2010), Springer Verlag.
- [29] P. Lesaint and P.-A. Raviart, On a finite element method for solving the neutron transport equation, in *Mathematical Aspects of Finite Elements in Partial Differential Equations*, de Boor ed., Academic Press, New York, 1974, pp. 89–123.
- [30] H. Luo, H. Segawa and M. R. Visbal, An implicit discontinuous Galerkin method for the unsteady compressible Navier-Stokes equation, *Comput. Fluids*, 53 (2012), pp. 133–144.
- [31] C. R. Nastase and D. J. Mavriplis, High-order discontinuous Galerkin methods using an *hp*-multigrid approach, *J. Comput. Phys.*, 213 (2006), pp. 330–357.
- [32] M. Pernice and H. F. Walker, NITSOL: a Newton iterative solver for nonlinear systems, *SIAM J. Sci. Comput.*, 19 (1998), pp. 302–318.
- [33] P. O. Person and J. Peraire, Newton-GMRES preconditioning for discontinuous Galerkin discretizations of the Navier-Stokes equations, *SIAM J. Sci. Comput.*, 30 (2008), pp. 2709–2733.
- [34] P. Rasetarinera and M. Y. Hussaini, An efficient implicit discontinuous spectral Galerkin method, *J. Comput. Phys.*, 172 (2001), pp. 718–738.
- [35] W. H. Reed and T. R. Hill, Triangular mesh methods for the neutron transport equation, Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, NM, 1973.
- [36] J. F. Remacle, J.E. Flaherty and M.S. Shephard, An Adaptive Discontinuous Galerkin Technique with an Orthogonal Basis Applied to Compressible Flow Problems, *SIAM Review*, 45 (2003), pp. 55–73.
- [37] F. Renac, C. Marmignon and F. Coquel, Time implicit high-order discontinuous Galerkin method with reduced evaluation cost, *SIAM J. Sci. Comput.*, 34 (2012), pp. 370–394.
- [38] F. Renac, C. Marmignon and F. Coquel, Fast time implicit-explicit discontinuous Galerkin method for convection dominated flow problems, *Commun. Math. Sci.*, 10 (2012), pp. 1161–1172.
- [39] A. Roshko, On the development of turbulent wakes from vortex streets, NACA Report 1191, 1954.
- [40] Y. Saad and M. Schultz, A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM Journal Sci. Stat. Comput.*, 7 (1986), pp. 856–869.
- [41] S. J. Sherwin, R. M. Kirby, J. Peiró, R. L. Taylor and O. C. Zienkiewicz, On 2D elliptic discontinuous Galerkin methods, *Int. J. Numer. Methods Fluids*, 65 (2006), pp. 752–784.
- [42] C. W. Shu and S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, *J. Comput. Phys.*, 77 (1988), pp. 439–471.
- [43] P. Solin, K. Segeth and I. Dolezel, Higher-order finite element methods, Chapman & Hall/CRC Press, 2003.
- [44] G. Strang, On the construction and comparison of difference schemes, *SIAM J. Numer. Anal.*, 5 (1968), pp. 506–517.
- [45] M. L. Ould Salihi, G.-H. Cottet and M. El Hamraoui, Blending finite-difference and vortex methods for incompressible flow computations, *SIAM J. Sci. Comput.*, 22 (2000), pp. 1655–1674.
- [46] A. Uranga, P.-O. Persson, M. Drela and J. Peraire, Implicit large eddy simulation of transition to turbulence at low Reynolds numbers using a discontinuous Galerkin method, *Int. J. Numer. Meth. Fluid*, 87 (2011), pp. 232–261.
- [47] M. R. Visbal, R. E. Gordnier and M. C. Galbraith, High-fidelity simulations of moving and flexible airfoils at low Reynolds numbers, *Exp. Fluids*, 46 (2009), pp. 903–922.
- [48] C.H.K Williamson, Vortex dynamics in the cylinder wake, *Annu. Rev. Fluid. Mech.*, 28 (1996), pp. 477–539.
- [49] K. Yasue, M. Furudate, N. Ohnishi and K. Sawada, Implicit discontinuous Galerkin method for RANS simulation utilizing pointwise relaxation algorithm, *Commun. Comput. Phys.*, 7 (2010), pp. 510–533.

Bibliographie

La validité des liens pour les sites internet est garantie à la date d'écriture de ce manuscrit.

- [1] R. Abgrall, A. Larat, and M. Ricchiuto. Construction of very high order residual distribution schemes for steady inviscid flow problems on hybrid unstructured meshes. J. Comput. Phys., 230 : 4103–4136, 2011.
- [2] M. Abramowitz and I.A. Stegun, editors. HandBOOK of mathematical functions with formulas, graphs, and mathematical tables. Dover Publications, 10th edition, 1972.
- [3] S.R. Allmaras. A coupled Euler/Navier-Stokes algorithm for 2D unsteady transonic shock/boundary-layer interaction. PhD thesis, MIT, 1989.
- [4] D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. SIAM J. Numer. Anal., 39 : 1749–1779, 2002.
- [5] F. Bassi, L. Botti, A. Colombo, A. Ghidoni, and S. Rebay. Discontinuous Galerkin for turbulent flows. In Z.J. Wang, editor, Adaptive high-order methods in computational fluid dynamics, pages 1–32. World Scientific Publishing, 2011.
- [6] F. Bassi, L. Botti, D.A. Colombo, D.A. Di Pietro, and P. Tesini. On the flexibility of agglomeration based physical space discontinuous Galerkin discretizations. J. Comput. Phys., 231 : 45–65, 2012.
- [7] F. Bassi, A. Crivellini, S. Rebay, and M. Savini. Discontinuous Galerkin solution of the Reynolds-averaged Navier-Stokes and $k-\omega$ turbulence model equations. Comput. Fluids, 34 : 507–540, 2005.
- [8] F. Bassi, A. Ghidoni, and S. Rebay. Optimal Runge-Kutta smoothers for the p-multigrid discontinuous Galerkin solution of the 1D Euler equations. J. Comput. Phys., 230 : 4153–4175, 2011.
- [9] F. Bassi and S. Rebay. A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations. J. Comput. Phys., 131 : 267–279, 1997.
- [10] F. Bassi and S. Rebay. High-order accurate discontinuous finite element solution of the 2D Euler equations. J. Comput. Phys., 138 : 251–285, 1997.
- [11] F. Bassi and S. Rebay. GMRES discontinuous Galerkin solution of the compressible Navier–Stokes equations. In B. Cockburn, G.E. Karniadakis, and C.-W. Shu, editors, Discontinuous Galerkin methods, theory, computation and applications, volume 11, pages 197–208. Springer-Verlag, 2000.

-
- [12] F. Bassi and S. Rebay. Numerical evaluation of two discontinuous Galerkin methods for the compressible Navier–Stokes equations. Int. J. Numer. Meth. Fluids, 40 : 197–207, 2002.
- [13] F. Bassi, S. Rebay, G. Mariotti, S. Pedinotti, and M. Savini. A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows. In R. Decuyper and G. Dibelius, editors, Proceedings of the 2nd European Conference on Turbomachinery Fluid Dynamics and Thermodynamics, Antwerpen, Belgium, 1997.
- [14] B. Berde. Etude d’une méthode de volumes finis de type Galerkin discontinu : résolution des équations d’Euler et de Navier-Stokes sur des maillages irréguliers. PhD thesis, Université de Pierre et Marie Curie - Paris VI, 1995.
- [15] P. Birken, G. Gassner, M. Haas, and C.-D Munz. Preconditioning for modal discontinuous Galerkin methods for the unsteady 3D Navier-Stokes equations. J. Comput. Phys., 240 : 20–35, 2013.
- [16] M. Borrel and J. Ryan. The elastoplast discontinuous Galerkin (EDG) method for the Navier-Stokes equations. J. Comput. Phys., 231 : 1–22, 2012.
- [17] F. Brezzi, G. Manzini, D. Marini, P. Pietra, and A. Russo. Discontinuous Galerkin approximations for elliptic problems. Numer. Methods Partial Differential Eqns., 16 : 365–378, 2000.
- [18] M.O. Bristeau, R. Glowinski, J. Periaux, and Viviani H., editors. A GAMM workshop, volume 18 of Notes on Numerical Fluid Mechanics and Multidisciplinary Design. Springer Vieweg, 1987.
- [19] A. Burbeau. Méthodes de Galerkin discontinu d’ordre élevé pour la simulation instationnaire en maillage non structuré. PhD thesis, Université de Bordeaux 1, 2000.
- [20] A. Burbeau and P. Sagaut. Simulation of a viscous compressible flow past a circular cylinder with high-order discontinuous Galerkin methods. Comput. Fluids, 31 : 867–889, 2002.
- [21] L. Cambier, S. Heib, and S. Plot. The Onera *elsA* CFD software : input from research and feedback from industry. Mechanics & Industry, 14 : 159–174, 2013.
- [22] S. Candel. Mécanique des fluides. Cours. Dunod, 2nd edition, 1995.
- [23] J.-B. Chapelier, M. de la Llave Plata, F. Renac, and E. Lamballais. Evaluation of a high-order discontinuous Galerkin method for the DNS of turbulent flows. Comput. Fluids (in review), 2014.
- [24] H.J.H. Clercx and C.-H. Bruneau. The normal and oblique collision of a dipole with a no slip boundary. Comput. Fluids, 35 : 245–279, 2006.
- [25] B. Cockburn, S. Hou, and C.-W. Shu. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws IV : the multidimensional case. Math. Comp., 54 : 545–581, 1990.
- [26] B. Cockburn, G.E. Karniadakis, and C.-W. Shu, editors. Discontinuous Galerkin methods, theory, computation and applications, volume 11. Springer-Verlag, 2000.
- [27] B. Cockburn and C.-W. Shu. The local discontinuous Galerkin method for time-dependent convection-diffusion systems. SIAM J. Numer. Anal., 35 : 2440–2463, 1998.
- [28] B. Cockburn and C.-W. Shu. Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. J. Sci. Computing, 16 : 173–261, 2001.

- [29] B. Cockburn and C.-W. Shu. Foreword. *J. Sci. Comput.*, 40 : 1–3, 2009.
- [30] Y. Colin and J.-P. Boin. Driving state definition and normalization within elsA-Python script. Cerfacs, Technical Report (TR/CFD/06/29), 2006.
- [31] G.-H. Cottet, P. Koumoutsakos, and M.L. Ould Salihi. Vortex methods with spatially varying cores. *J. Comput. Phys.*, 162 : 164–185, 2000.
- [32] V. Couaillier, A. Gazaix-Jollès, M. Gazaix, and C. Marmignon. PRC ANANAS - Tranche 2 - Tâches A.2.2, B.2.2.2 et C - Méthodes numériques pour la simulation des écoulements instationnaires. Onera, Technical Report (RT 10/11457), 2007.
- [33] V. Couaillier, F. Renac, and M.C. Le Pape. Hybrid multigrid DG/FV methods for viscous turbulent flows. In N. Kroll, H. Bieler, H. Deconinck, V. Couaillier, H. van der Ven, and K. Sorensen, editors, ADIGMA - A European initiative on the development of adaptive higher-order variational methods for aerospace applications, Notes on Numerical Fluid Mechanics and Multidisciplinary Design, pages 225–242. Springer-Verlag, 2010.
- [34] J. Crank and P. Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Mathematical Proceedings of the Cambridge Philosophical Society*, 43 : 50–67, 1947.
- [35] A. Crivellini and F. Bassi. An implicit matrix-free Discontinuous Galerkin solver for viscous and turbulent aerodynamic simulations. *Comput. Fluids*, 50 : 81–93, 2011.
- [36] G. Dahlquist. A special stability problem for linear multistep methods. *BIT Numerical Mathematics*, 3 : 27–43, 1963.
- [37] T. Dairay. Développement et évaluation d’une méthode implicite à bas coût appliquée aux schémas numériques de type Galerkin discontinu. Master’s thesis, Ecole Centrale Paris, 2010.
- [38] D.A. Di Pietro and A. Ern. Mathematical aspects of discontinuous Galerkin methods, volume 69 of Mathematics and Applications. Springer Verlag, 2012.
- [39] L.T. Diosady and D.L. Darmofal. Preconditioning methods for discontinuous Galerkin solutions of the Navier-Stokes equations. *J. Comput. Phys.*, 228 : 3917–3935, 2009.
- [40] V. Dolejší, M. Holík, and J. Hozman. Efficient solution strategy for the semi-implicit discontinuous Galerkin discretization of the Navier-Stokes equations. *J. Comput. Phys.*, 230 : 4176–4200, 2011.
- [41] C. Drozo. Méthode de type Galerkin discontinu pour la résolution des équations de Navier-Stokes en compressible. PhD thesis, Ecole Nationale Supérieure d’Arts et Métiers, 1998.
- [42] M. Dubiner. Spectral methods on triangles and other domains. *J. Sci. Computing*, 6 : 345–390, 1991.
- [43] J.A. Ekaterinaris. High-order accurate, low numerical diffusion methods for aerodynamics. *Progress in Aerospace Sciences*, 41 : 192–300, 2005.
- [44] A. Ern. Aide-mémoire : éléments finis. Dunod, 2005.
- [45] M. Feistauer and V. Kučera. On a robust discontinuous Galerkin technique for the solution of compressible flows. *J. Comput. Phys.*, 224 : 208–221, 2007.
- [46] K.J. Fidkowski, T.A. Oliver, J. Lu, and D.L. Darmofal. p -Multigrid solution of high-order discontinuous Galerkin discretizations of the compressible Navier-Stokes equations. *J. Comput. Phys.*, 207 : 92–113, 2005.

-
- [47] K.J. Fidkowski and P.L. Roe. An entropy adjoint approach to mesh refinement. *SIAM J. Sci. Comput.*, 32 : 1261–1287, 2010.
- [48] M.C. Galbraith. Implicit large eddy simulation of low-Reynolds-number transitional flow past the SD7003 airfoil. Master’s thesis, University of Cincinnati, 2009.
- [49] M.C. Galbraith, P.D. Orkwis, and J.A. Benek. Automated quadrature-free discontinuous Galerkin method to viscous flows. AIAA-2011-283. In conference paper of the 48th AIAA Aerospace Sciences Meeting, Orlando, Florida, 2011.
- [50] H. Gao, Z.J. Wang, and H.T. Huynh. Differential formulation of discontinuous Galerkin and related methods for the Navier-Stokes equations. *Commun. Comput. Phys.*, 13 : 1013–1044, 2013.
- [51] G. Gassner, F. Lörcher, and C.-D. Munz. A discontinuous Galerkin scheme based on a space-time expansion II. Viscous flow equations in multi dimensions. *J. Sci. Comput.*, 34 : 260–286, 2008.
- [52] C.W. Gear. Numerical initial value problems in ordinary differential equations. Prentice-Hall, 1971.
- [53] S. Gérald, F. Renac, C. Marmignon, J. Ryan, and F. Coquel. Fast time implicit-explicit discontinuous Galerkin method for convection-diffusion flow problems. AIAA-2012-0731. In conference paper of the 50th AIAA Aerospace Sciences Meeting, Nashville, Tennessee, 2012.
- [54] C. Geuzaine and J.-F. Remacle. Gmsh : a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. *Int. J. Numer. Meth. Engng.*, 79 : 1309–1331, 2009.
- [55] C. Geuzaine and J.-F. Remacle. Gmsh Reference Manual : the documentation for Gmsh 2.5, a finite element mesh generator with built-in pre- and post-processing facilities. Technical Report, 2010.
- [56] E. Godlewski and P.-A. Raviart. Numerical approximation of hyperbolic systems of conservation laws, volume 118. Applied Mathematical Sciences, Springer-Verlag, 1996.
- [57] R. Hartmann, J. Held, T. Leicht, and F. Prill. Discontinuous Galerkin methods for computational aerodynamics - 3D adaptive flow simulation with the DLR PADGE code. *Aero. Sci. Tech.*, 14 : 512–519, 2010.
- [58] R. Hartmann and P. Houston. Symmetric interior penalty DG methods for the compressible Navier-Stokes equations I : method formulation. *Int. J. Numer. Anal. Model.*, 3 : 1–20, 2006.
- [59] R. Hartmann, M. Lukáčová-Medvid’ová, and F. Prill. Efficient preconditioning for the discontinuous Galerkin finite element method by low-order elements. *Appl. Numer. Math.*, 59 : 1737–1753, 2009.
- [60] J.K. Harvey and F.J. Perry. Flowfield produced by trailing vortices in the vicinity of the ground. *AIAA Journal*, 9 : 1659–1660, 1971.
- [61] F. Heimann, C. Engwer, O. Ippisch, and P. Bastian. An unfitted interior penalty discontinuous Galerkin method for incompressible Navier-Stokes two-phase flow. *Int. J. Numer. Meth. Fluids*, 71 : 269–293, 2013.
- [62] J.S. Hesthaven and T. Warburton. Nodal discontinuous Galerkin methods : Algorithms, analysis and applications, volume 54 of Texts in Applied Mathematics. Springer, 2006.

- [63] K. Heun. Neue Methode zur approximativen Integration der Differentialgleichungen einer unabhängigen Veränderlichen. *Z. Math. u. Phys.*, 45 : 23–38, 1900.
- [64] K. Hillewaert, J.-F. Remacle, N. Cheveaugeon, P.-E. Bernard, and P. Geuzaine. Analysis of a hybrid p -multigrid method for the discontinuous Galerkin discretisation of the Euler equations. In P. Wesseling, E. Oñate, and J. Périaux, editors, *Proceedings of the ECCOMAS 2006*, TU Delft, The Netherlands, 2006.
- [65] C. Hirsch. *Numerical computation of internal & external flows. Vol. 1 : The fundamentals of computational fluid dynamic*. John Wiley & Sons, 2007.
- [66] H. Holden, K.H. Karlsen, K.-A. Lie, and N.H. Risebro. *Splitting methods for partial differential equations with rough solutions. Analysis and Matlab programs*, volume 69 of *Series of lectures in mathematics*. European Mathematical Society, 2010.
- [67] P. Houston, C. Schwab, and E. Süli. Stabilized hp -finite element methods for first-order hyperbolic problems. *SIAM J. Numer. Anal.*, 37 : 1618–1643, 2000.
- [68] P. Houston, C. Schwab, and E. Süli. Discontinuous hp -finite element methods for advection-diffusion reactions problems. *SIAM J. Numer. Anal.*, 39 : 2133–2163, 2002.
- [69] A. Huerta, E. Casoni, and J. Peraire. A simple shock-capturing technique for high-order discontinuous Galerkin methods. *Int. J. Numer. Meth. Fluids.*, 69 : 1614–1632, 2012.
- [70] H.T. Huynh. A flux reconstruction approach to high-order schemes including discontinuous Galerkin methods. AIAA-2007-4079. In *conference paper of the 18th AIAA Computational Fluid Dynamics Conference*, Miami, Florida, 2007.
- [71] Internet. Acare. <http://www.acare4europe.org/>.
- [72] Internet. AeroSol. <http://raweb.inria.fr/rapportsactivite/RA2011/bacchus/uid28.html>.
- [73] Internet. AFOSR. <http://www.wpafb.af.mil/afrl/afosr/>.
- [74] Internet. COOLFluid. <http://coolfluidsrv.vki.ac.be/trac/coolfluid>.
- [75] Internet. DLR Germany. TauBench - IPACS. <http://www.ipacs-benchmark.org>.
- [76] Internet. Dune. <http://www.dune-project.org/dune.html>.
- [77] Internet. Idihom. <http://www.idihom.de/>.
- [78] Internet. NACA. <http://naca.central.cranfield.ac.uk/>.
- [79] Internet. NPARC. <http://www.grc.nasa.gov/WWW/wind/valid/archive.html>.
- [80] Internet. Onera. <http://www.onera.fr/fr/missions-et-objectifs>.
- [81] Internet. Xflow. <http://xflow.engin.umich.edu/>.
- [82] Internet. 2nd international workshop on high-order CFD methods, Cologne, Germany. http://www.dlr.de/as/desktopdefault.aspx/tabid-8170/13999_read-35550/, 27–28 may 2013.
- [83] Internet. 1st international workshop on high-order CFD methods, Nashville, Tennessee. <http://www.people.ku.edu/~z651w035/hiocfd.html>, 7–8 janvier 2012.
- [84] E.N. Jacobs. *Tests of six symmetrical airfoils in the variable density wind tunnel*. NACA, Technical Report (385), 1931.
- [85] C. Johnson and J. Pitkäranta. An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.*, 46 : 1–26, 1986.
- [86] G. Kanschat. *Discontinuous Galerkin methods for viscous incompressible flow*. Advances in numerical Mathematics, Vieweg-Teubner, 2007.

-
- [87] G.H. Keetels, U. D’Ortona, W. Kramer, H.J.H. Clercx, K. Schneider, and G.J.F. Van Heijst. Fourier spectral and wavelet solvers for the incompressible Navier-Stokes equations with volume-penalization : convergence of a dipole-wall collision. *J. Comput. Physics*, 227 : 919–945, 2007.
- [88] C.A. Kennedy and M.H. Carpenter. Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Appl. Numer. Math.*, 44 : 139–181, 2003.
- [89] C.M. Klaij, J.J.W. van der Vegt, and H. van der Ven. Space-time discontinuous Galerkin method for the compressible Navier-Stokes equations. *J. Comput. Phys.*, 217 : 589–611, 2006.
- [90] C.M. Klaij, M.H. van Raalte, H. van der Ven, and J.J.W. van der Vegt. h-multigrid for space-time discontinuous Galerkin discretizations of the compressible Navier-Stokes equations. *J. Comput. Phys.*, 227 : 1024–1045, 2007.
- [91] R. Klofko. Efficient matrix-free implementation of discontinuous Galerkin methods for compressible flow problems. In A. Handlovičová, Z. Minarechová, and D. Ševčovič, editors, *Proceedings of the ALGORITHM 2012*, 2012.
- [92] D.A. Knoll and D.E. Keyes. Jacobian-free Newton-Krylov methods : a survey of approaches and applications. *J. Comput. Phys.*, 193 : 357–397, 2004.
- [93] D.A. Kopriva. *Implementing spectral methods for partial differential equations*. Prentice-Hall, 2009.
- [94] W. Kramer, H.J.H. Clercx, and G.J.F. Van Heijst. Vorticity dynamics of a dipole colliding with a no-slip wall. *Phys. Fluids*, 19 (126603) : 1–13, 2007.
- [95] A.G. Kravchenko, P. Moin, and R. Moser. Zonal embedded grids for numerical simulations of wall-bounded turbulent flows. *J. Comput. Phys.*, 127 : 412–423, 1996.
- [96] N. Kroll, H. Bieler, H. Deconinck, V. Couaillier, H. van der Ven, and K. Sorensen, editors. *ADIGMA - A European initiative on the development of adaptive higher-order variational methods for aerospace applications*, volume 113 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*. Springer-Verlag, 2010.
- [97] S. Laizet and E. Lamballais. High-order compact schemes for incompressible flows : a simple and efficient method with quasi-spectral accuracy. *J. Comput. Phys.*, 228 : 5989–6015, 2009.
- [98] B. Landmann. *A parallel discontinuous Galerkin code for the Navier-Stokes and Reynolds-averaged Navier-Stokes equations*. PhD thesis, Universität Stuttgart, 2008.
- [99] S. Langer. Investigation and application of point implicit Runge-Kutta methods to inviscid flow problems. *Int. J. Numer. Meth. Fluids.*, 69 : 332–352, 2011.
- [100] S. Langer. Hierarchy of preconditioning techniques for the solution of the Navier-Stokes equations discretized by 2^{nd} order unstructured finite volume methods. In J. Eberhardsteiner, H.J. Böhm, and F.G. Rammerstorner, editors, *Proceedings of the ECCOMAS 2012*, Vienna, Austria, 2012.
- [101] T.-H. Lê, J.-M. Le Gouez, and E. Garnier. High accuracy flow simulations : advances and challenges for future needs in aeronautics. *Comput. Fluids*, 43 : 90–97, 2011.
- [102] R. Léger. *Couplage pour l’aéroacoustique entre méthode de Galerkin discontinue en maillage non-structuré et méthode de différences finies en maillage cartésien*. PhD thesis, Université Paris-Est Marne-la-Vallée, 2011.
- [103] R. Léger, C. Peyret, and S. Piperno. Coupled discontinuous Galerkin/finite difference solver on hybrid meshes for computational aeroacoustics. *AIAA Journal*, 50 : 338–349, 2012.

- [104] P.G. Lemarié-Rieusset. Recent developments in the Navier-Stokes problem. Chapman & Hall / CRC, 2002.
- [105] P. Lesaint and P.-A. Raviart. On a finite element method for solving the neutron transport equation. In Mathematical aspects of finite elements in partial differential equations, pages 89–123. Publication No. 33., Math. Res. Center, Univ. of Wisconsin–Madison, Academic Press, New York, 1974.
- [106] E.H.-L Liu. Optimization and validation of discontinuous Galerkin code for the 3D Navier-Stokes equations. Master’s thesis, MIT, 2011.
- [107] I. Lomtev, C.B. Quillen, and G.E. Karniadakis. Spectral/hp methods for viscous compressible flows on unstructured 2D meshes. J. Comput. Phys., 144 : 325–357, 1998.
- [108] F. Lörcher, G. Gassner, and C.-D. Munz. An explicit discontinuous Galerkin scheme with local time-stepping for general unsteady diffusion equations. J. Comput. Phys., 227 : 5649–5670, 2008.
- [109] H. Luo, J.D. Baum, and R. Löhner. A p-multigrid discontinuous Galerkin method for the Euler equations on unstructured grids. J. Comput. Phys., 211 : 767–783, 2006.
- [110] H. Luo, H. Segawa, and M.R. Visbal. An implicit discontinuous Galerkin method for the unsteady compressible Navier-Stokes equations. Comput. Fluids, 53 : 133–144, 2012.
- [111] A.J. Majda and A.L. Bertozzi, editors. Vorticity and incompressible flow. Cambridge texts in applied Mathematics, 2002.
- [112] K. Michalak and C. Ollivier-Gooch. Globalized matrix-explicit Newton-GMRES for the high-order accurate solution of the Euler equations. Comput. Fluids, 39 : 1156–1167, 2010.
- [113] D. Moro, N.C. Nguyen, and J. Peraire. Navier-Stokes solution using hybridizable discontinuous Galerkin methods. AIAA-2011-3407. In conference paper of the 20th AIAA CFD Conference, Honolulu, Hawaii, 2011.
- [114] C.R. Nastase and D.J. Mavriplis. High-order discontinuous Galerkin methods using an hp-multigrid approach. J. Comput. Phys., 213 : 330–357, 2006.
- [115] R.H. Nichols and B.D. Heikkinen. Validation of implicit algorithms for unsteady flows including moving and deforming grids. Journal of Aircraft, 43 : 1341–1351, Sept.-Oct. 2006.
- [116] J. Nordström, J. Gong, E. van der Weide, and M. Svärd. A stable and conservative high order multi-block method for the compressible Navier-Stokes equations. J. Comput. Phys., 228 : 9020–9035, 2009.
- [117] C. Ollivier-Gooch, A. Nejat, and K. Michalak. On obtaining and verifying high-order finite-volume solutions to the Euler equations on unstructured meshes. AIAA Journal, 47 : 2105–2120, 2009.
- [118] Onera. elsA : Theoretical Manual, /ELSA/STB-97020, 2.0, 2007.
- [119] P. Orlandi. Vortex dipole rebound from a wall. Phys. Fluids, 2 : 1429–1436, 1990.
- [120] M.L. Ould Salihi, G.-H. Cottet, and M. El Hamraoui. Blending finite-difference and vortex methods for incompressible flow computations. SIAM J. Sci. Comput., 22 : 1655–1674, 2000.

-
- [121] H. Park, R.R. Nourgaliev, R.C. Martineau, and D.A. Knoll. Jacobian-free Newton-Krylov discontinuous Galerkin method and physics-based preconditioning for nuclear reactor simulations. In R. Chawla, V.N. Dang, and K. Mikityuk, editors, Proceedings of the PHYSOR08, Interlaken, Switzerland, 2008.
- [122] H. Park, R.R. Nourgaliev, R.C. Martineau, and D.A. Knoll. On physics-based preconditioning of the Navier-Stokes equations. J. Comput. Phys., 228 : 9131–9146, 2009.
- [123] H. Park, R.R. Nourgaliev, V. Mousseau, and D.A. Knoll. Recovery discontinuous Galerkin Jacobian-free Newton-Krylov method for all-speed flows. In H. Choi, H.G. Choi, and J.Y. Yoo, editors, Proceedings of the ICCFD5, Seoul, South Korea, 2008.
- [124] J. Peraire and P.-O. Persson. The compact discontinuous Galerkin (CDG) method for elliptic problems. SIAM J. Sci. Comput., 30 : 1806–1824, 2008.
- [125] M. Pernice and H.F. Walker. NITSOL : a Newton iterative solver for nonlinear systems. SIAM J. Sci. Comput., 19 : 302–318, 1998.
- [126] P.-O. Persson. A sparse and high-order accurate line-based discontinuous Galerkin method for unstructured meshes. J. Comput. Phys., 233 : 414–429, 2013.
- [127] P.-O. Persson and J. Peraire. Newton-GMRES preconditioning for discontinuous Galerkin discretizations of the Navier-Stokes equations. SIAM J. Sci. Comput., 30 : 2709–2733, 2008.
- [128] L. Pesch and J.J.W. van der Vegt. A discontinuous Galerkin finite element discretization of the Euler equations for compressible and incompressible fluids. J. Comput. Phys., 227 : 5426–5446, 2008.
- [129] R. Peyret. Résolution numérique des systèmes hyperboliques. Application à la dynamique des gaz. Onera, Technical Report (RT 1977-5), 1977.
- [130] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. Numerical recipes in Fortran 77 : the art of scientific computing (Volume 1). Cambridge University Press, 2nd edition, 1992.
- [131] J. Qiu, B.C. Khoo, and C.-W. Shu. A numerical study for the performance of the Runge-Kutta discontinuous Galerkin method based on different numerical fluxes. J. Comput. Phys., 212 : 540–565, 2006.
- [132] P. Rasetarinera and M.Y. Hussaini. An efficient implicit discontinuous spectral Galerkin method. J. Comput. Phys., 172 : 718–738, 2001.
- [133] W.H. Reed and T.R. Hill. Triangular mesh methods for the neutron transport equation. Los Alamos Scientific Laboratory, Technical Report (LA-UR-73-479), 1973.
- [134] J.-F. Remacle, J.E. Flaherty, and M.S. Shepard. An adaptive discontinuous Galerkin technique with an orthogonal basis applied to compressible flow problems. SIAM Review, 45 : 53–72, 2003.
- [135] F. Renac and V. Couaillier. Synthesis report on MTC high order scheme results. Adigma project No. AST-CT-2006-030719, Technical Report (ADIGMA/WP6/ONERA/D6.2.1), 2008.
- [136] F. Renac, S. Gérald, C. Marmignon, and F. Coquel. Fast time implicit-explicit discontinuous Galerkin method for compressible Navier-Stokes equations. J. Comput. Phys., 251 : 272–291, 2013.

- [137] F. Renac, S. Gérald, C. Marmignon, and F. Coquel. Stability analysis of a fast time implicit-explicit discontinuous Galerkin method. Application to compressible Navier-Stokes equations. In European workshop HONOM, Bordeaux, France, 2013.
- [138] F. Renac, C. Marmignon, and F. Coquel. Fast-time implicit discontinuous Galerkin method for convection dominated flow problems. Commun. Math. Sci, 10 : 1161–1172, 2012.
- [139] F. Renac, C. Marmignon, and F. Coquel. Time implicit high-order discontinuous Galerkin method with reduced evaluation cost. SIAM J. Sci. Comput., 34 : 370–394, 2012.
- [140] F. Renac and E. Martin. Le code AGHORA : un démonstrateur intégrant des méthodes numériques d’ordre élevé pour la simulation d’écoulements compressibles en aérodynamique. Onera, Technical Report (RT 1/20130), 2012.
- [141] L.F. Richardson and J.A. Gaunt. The deferred approach to the limit. Part I : single lattice. Part II : interpenetrating lattices. Phil. Trans. R. Soc. Lond. A, 226 : 299–361, 1927.
- [142] G.R. Richter. An optimal-order error estimate for the discontinuous Galerkin method. Math. Comp., 50 : 75–88, 1988.
- [143] B. Rivière. Discontinuous Galerkin methods for solving elliptic and parabolic equations : Theory and implementation. SIAM, Frontiers in applied Mathematics, 2008.
- [144] A. Rizzi and H. Viviand, editors. Numerical methods for the computation of inviscid transonic flows with shock waves : a GAMM workshop, volume 3 of Notes on numerical fluid mechanics. Vieweg, 1981.
- [145] P.J. Roache. Perspective : a method for uniform reporting of grid refinement studies. J. Fluids Eng., 116 : 405–413, 1994.
- [146] A. Roshko. On the development of turbulent wakes from vortex streets. NASA, Technical Report (1191), 1954.
- [147] Y Saad. Iterative methods for sparse linear systems. SIAM, 2nd edition, 2003.
- [148] Y. Saad and M.H. Schultz. GMRES : a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comput., 7 : 856–869, 1986.
- [149] H. Schlichting. Boundary-layer theory. McGraw-Hill, 6th edition, 1968.
- [150] M.S. Selig, J.J. Guglielmo, A.P. Broeren, and P. Giguère. Summary of Low-Speed Airfoil Data, Volume 1. SoarTech Publication, Virginia Beach, Virginia, 1995.
- [151] R. Sevilla, S. Fernández-Méndez, and A. Huerta. Comparison of high-order curved finite elements. Int. J. Numer. Meth. Engng., 87 : 719–734, 2011.
- [152] K. Shahbazi, P.F. Fischer, and C.R. Ethier. A high-order discontinuous Galerkin method for the unsteady incompressible Navier-Stokes equations. J. Comput. Phys., 222 : 391–407, 2007.
- [153] S.J. Sherwin and G.E. Karniadakis. Spectral/hp element methods for computational fluid dynamics. Numerical mathematics and scientific computation. Oxford University Press, 2nd edition, 2005.
- [154] S.J. Sherwin, R.M. Kirby, J. Peiró, R.L. Taylor, and O.C. Zienkiewicz. On 2D elliptic discontinuous Galerkin methods. Int. J. Numer. Meth. Engng., 65 : 752–784, 2006.
- [155] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys., 77 : 439–471, 1988.

-
- [156] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes, II. J. Comput. Phys., 83 : 32–78, 1989.
- [157] T. Sonar. Turbulences sur les équations des fluides. Dossier Pour la Science, 74 : 36–44, janvier - mars 2012.
- [158] P.R. Spalart. Strategies for turbulence modelling and simulations. Int. J. Heat Fluid Flow, 21 : 252–263, 2000.
- [159] B. Sportisse. An analysis of operator splitting techniques in the stiff case. J. Comput. Phys., 161 : 140–168, 2000.
- [160] G. Strang. On the construction and comparison of difference schemes. SIAM J. Numer. Anal., 5 : 506–517, 1968.
- [161] R. Temam. Navier-Stokes equations. North-Holland Pub. Company, 3th revised edition 1984.
- [162] E.F. Toro, editor. Riemann solvers and numerical methods for fluid dynamic. A practical introduction. Springer-Verlag, 2nd edition, 1999.
- [163] D.J. Tritton. Experiments on the flow past a cylinder at low Reynolds numbers. J. Fluid Mech., 6 : 547–567, 1959.
- [164] P. Trontin. Isogeometric analysis of Euler compressible flow. application to aerodynamics. AIAA-2012-0295. In conference paper of the 50th AIAA Aerospace Sciences Meeting, Nashville, Tennessee, 2012.
- [165] S. Turek. A comparative study of time-stepping techniques for the incompressible Navier-Stokes equations : from fully implicit non-linear schemes to semi-implicit projection methods. Int. J. Numer. Meth. Fluids, 22 : 987–1011, 1996.
- [166] A. Uranga, P.-O. Persson, M. Drela, and J. Peraire. Implicit large eddy simulation of transition to turbulence at low reynolds numbers using a discontinuous Galerkin method. Int. J. Numer. Meth. Fluids, 87 : 232–261, 2011.
- [167] J.J.W. van der Vegt and H. van der Ven. Space–time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows. I. General formulation. J. Comput. Phys., 182 : 546–585, 2002.
- [168] G.J.F. Van Heijst and J.B. Flor. Laboratory experiments on dipole structures in a stratified fluid. Mesoscale/Synoptic coherent structures in geophysical turbulence, 50 : 591–608, 1989.
- [169] B. van Leer and S. Nomura. Discontinuous Galerkin for diffusion. AIAA-2005-5108. In conference paper of the 17th AIAA CFD Conference, Toronto, Ontario Canada, 2005.
- [170] M. van Raalte and B. van Leer. Bilinear forms for the recovery-based discontinuous Galerkin method for diffusion. Commun. Comput. Phy, 5 : 683–693, 2009.
- [171] J.C. Vassberg. Expectations for computational fluid dynamics. Int. J. Comput. Fluid Dyn., 19 : 549–558, 2005.
- [172] C. Villani. Il faudra toujours de l’intelligence derrière l’ordinateur. La lettre de BULL : Bull Direct, 50 : 3–4, 2011.
- [173] M.R. Visbal, R.E. Gordnier, and M.C. Galbraith. High-fidelity simulations of moving and flexible airfoils at low Reynolds numbers. Exp. Fluids, 46 : 903–922, 2009.
- [174] L. Wang and D. Mavriplis. Implicit solution of the unsteady Euler equations for high-order accurate discontinuous Galerkin discretizations. J. Comput. Phys., 225 : 1994–2015, 2007.

- [175] Z.J. Wang. High-order methods for the Euler and Navier-Stokes equations on unstructured grids. Progress in Aerospace Sciences, 43 : 1–41, 2007.
- [176] Z.J. Wang, K. Fidkowski, R. Abgrall, F. Bassi, D. Caraeni, A. Cary, H. Deconinck, R. Hartmann, K. Hillewaert, H.T. Huynh, N. Kroll, G. May, P.-O. Persson, B. van Leer, and M. Visbal. High-order CFD methods : current status and perspective. Int. J. Numer. Meth. Fluids, 72 : 811–845, 2013.
- [177] Z.J. Wang and H. Gao. A unifying lifting collocation penalty formulation including the discontinuous Galerkin spectral volume/difference methods for conservation laws on mixed grids. J. Comput. Phys., 228 : 8161–8186, 2009.
- [178] T.C. Warburton, I. Lomtev, Y. Du, S.J. Sherwin, and G.E. Karniadakis. Galerkin and discontinuous Galerkin spectral/hp methods. Comput. Methods Appl. Mech. Engrg., 175 : 343–359, 1999.
- [179] C.H.K Williamson. Vortex dynamics in the cylinder wake. Annu. Rev. Fluid. Mech., 28 : 477–539, 1996.
- [180] M. Yano and D.L. Darmofal. An optimization framework for anisotropic simplex mesh adaptation : application to aerodynamic flows. AIAA-2012-0079. In conference paper of the 50th AIAA Aerospace Sciences Meeting, Nashville, Tennessee, 2012.
- [181] K. Yasue, M. Furudate, N. Ohnishi, and K. Sawada. Implicit discontinuous Galerkin method for RANS simulation utilizing pointwise relaxation algorithm. Commun. Comput. Phys., 7 : 510–533, 2010.
- [182] Q. Yuan and C.-W. Shu. Discontinuous Galerkin method based on non-polynomial approximations space. J. Comput. Phys., 218 : 295–323, 2006.
- [183] M. Zhang and C.-W. Shu. An analysis of three different formulations of the discontinuous Galerkin method for diffusion equations. Math. Models Methods Appl. Sci., 13 : 395–413, 2003.
- [184] J. Zhu, Y.-T. Zhang, S.A. Newman, and M. Alber. Application of discontinuous Galerkin methods for reaction-diffusion systems in developmental biology. SIAM J. Sci. Comput., 40 : 391–418, 2009.

**Méthode de Galerkin Discontinue et intégrations explicites-implicites en temps basées sur un découplage des degrés de liberté.
Applications au système des équations de Navier-Stokes.**

En mécanique des fluides numérique, un enjeu est le développement de méthodes d'approximation d'ordre élevé, comme celles de Galerkin Discontinues (GD). Si ces méthodes permettent d'envisager la simulation d'écoulements complexes en alternative aux méthodes usuelles d'ordre deux, elles souffrent cependant d'une forte restriction sur le pas de temps lorsqu'elles sont associées à une discrétisation explicite en temps. Ce travail de thèse consiste à mettre en œuvre une stratégie d'intégration temporelle explicite-implicite efficace, associée à une discrétisation spatiale GD d'ordre élevé, pour les écoulements instationnaires à convection dominante de fluides visqueux compressibles modélisés par le système des équations de Navier-Stokes.

La discrétisation spatiale de la méthode GD est associée à des flux numériques de fluides parfaits et visqueux à stencil compact. En présence de frontières matérielles courbes, l'ordre élevé est garanti par la discrétisation du domaine de calcul à l'aide d'une représentation iso-paramétrique. La stratégie d'intégration temporelle repose sur une décomposition d'opérateurs de Strang, où les termes de convection sont résolus explicitement et ceux de diffusion implicitement. Son efficacité résulte d'une simplification du schéma implicite, où le calcul de la matrice implicite est approché avec une méthode sans jacobienne et où les degrés de liberté du schéma sont découplés. De fait, la taille du système linéaire à résoudre et le temps de calcul de la résolution sont significativement réduits. Enfin, la validation et l'évaluation des performances du schéma numérique sont réalisées à travers cinq cas tests bien référencés en deux dimensions d'espace.

***Mots-clés :** Méthode de Galerkin Discontinue ; Intégration temporelle explicite-implicite ; Ecoulement à convection dominante ; Stencil compact ; Frontière courbe ; Décomposition d'opérateurs de Strang ; Méthode sans jacobienne ; Découplage des degrés de liberté.*

**Discontinuous Galerkin method and implicit-explicit time integrations based on a decoupling of degrees of freedom.
Applications to the system of Navier-Stokes equations.**

In computational fluid dynamics, one challenge is the development of high order approximation methods, such as those of Discontinuous Galerkin (DG). If such methods allow the simulation of complex flows to be considered as an alternative to usual methods of order two, they nevertheless suffer from a strong restriction on the time step when combined with an explicit time discretization. The thesis consists in implementing an efficient strategy of implicit-explicit time integration, associated with a DG space discretization of high order, for convection dominated unsteady flows of viscous compressible fluids modelled by the system of Navier-Stokes equations.

The spatial discretization of DG method is associated with numerical fluxes of inviscid and viscous fluids with compact stencil. In the presence of curved material boundaries, high order is guaranteed by the discretization of the computational domain using iso-parametric representation. The temporal integration strategy is based on a Strang splitting where the convective terms are solved explicitly and the diffusive ones implicitly. Its efficiency results from a simplification of the implicit scheme where the computation of the implicit matrix is approximated with a Jacobian-free method and where the degrees of freedom of the scheme are decoupled. Thereby, the size of the linear system to be solved and the computational time of the resolution are significantly reduced. Finally, validation and performance assessment of the numerical scheme are made through five test cases properly referenced in two-dimensional space.

***Keywords:** Discontinuous Galerkin method ; Implicite-explicit time integration ; Convection dominated flow ; Compact stencil ; Curved boundary ; Strang splitting ; Jacobian-free method ; Decoupling of degrees of freedom.*