Modélisation POD-Galerkine réduite pour le contrôle des écoulements instationnaires

Mathieu Couplet

Université de Paris 13 - Paris Nord
en collaboration avec
l'Office National de Recherches et d'études Aérospatiales

Numéro : |_|_|_|_|_|_|_|_|

THÈSE

pour obtenir le grade de

Docteur de l'Université Paris 13

Discipline : Mathématiques Appliquées
École Doctorale Galilée

présentée et soutenue publiquement par

Mathieu Couplet

le 20 janvier 2005

Titre :

Modélisation POD-Galerkine réduite pour le contrôle des écoulements instationnaires

Directeur de thèse : Claude Basdevant

Encadrant : Pierre Sagaut

Jury :

MM.	Mikhaël Balabane,	Université de Paris 13,	*président*
	Claude Basdevant,	Université de Paris 13,	*examinateur*
	Jacques Blum,	Université de Nice,	*rapporteur*
	Jean-Paul Bonnet,	Université de Poitiers,	*rapporteur*
	Joël Delville,	Université de Poitiers,	*membre invité*
	Thiên-Hiêp Lê,	ONERA,	*examinateur*
	Pierre Sagaut,	Université de Paris 6,	*examinateur*

Remerciements

Ce travail de thèse a été financé par l'Office National d'Études et de Recherches Aérospatiales. Je remercie l'ONERA et M. Philippe Morice, directeur du Département de Simulation Numérique des écoulements et Aéroacoustique, de m'avoir accueilli et d'avoir mis la logistique de l'Office à ma disposition.

Je remercie vivement MM. Claude Basdevant et Pierre Sagaut de m'avoir fait confiance tout au long de ces trois années, d'avoir dirigé efficacement mon travail mais aussi de leur disponibilité et de leurs conseils avisés.

Mes remerciements s'adressent également à MM. Jacques Blum et Jean-Paul Bonnet qui ont accepté de rapporter ce mémoire et dont la célérité m'a permis de soutenir rapidement en dépit des fêtes de fin d'année. Je remercie M. Mikhaël Balabane d'avoir présidé le jury et M. Joël Delville de sa présence au sein du jury.

Je remercie M. Thiên-Hiêp Lê d'être venu représenter l'ONERA dans le jury et, surtout, de son accueil au sein de l'unité ÉTRI.

Mon séjour à l'ONERA s'est déroulé de manière fort agréable, tant sur les plans humain que technique ou scientifique. Pour cela, je tiens à remercier Mmes Brigitte Commelin et Ghislaine Denis de s'être toujours occupées avec efficacité et gentillesse de mes requêtes administratives, MM. Didier Blaise et Philippe Céroni de leur disponibilité pour répondre aux demandes informatiques, MM. Ivan Mary et Emmanuel Montreuil d'avoir mis leurs données numériques à ma disposition et répondu à mes questions.

Merci également aux personnes avec qui j'ai partagé des repas souvent animés, des parties acharnées en réseau, quelques matches de foot, d'agréables pauses cafés et des discussions passionnantes que ce soit à propos de la méthode de Kirchhoff ou des chevaliers-paysants de l'an mil du Lac de Paladru, entre autres et dans le désordre : Erik alias Igor (planophile), Bruno R. (brechi), Noel (pousseur-enflammeur fuxéen), Lionel, Ronan (auteur-interprète), Frédérique (tagadavore), Thorvald, Emmanuel, Lutz (chercheur de mode global), Bruno A. (hérisson rieur), Sébastien (capitaine du FC ÉTRI), Fabrice (contradicteur-maïeuticien à poil dur), Chi-Tuan (Normand pur beurre), Guillaume N. (Adam), Élisa (globe-trotteuse), Frédéric (docteur ès sketches), Guillaume D. (jardinier acousticien), Julien, Kelly et Olivier.

Je remercie tous ceux qui se sont intéressés à ce que je pouvais bien faire pendant toutes ces heures à l'ONERA depuis ces longs mois, à part boire du café et envoyer des courriels, ou qui m'ont soutenu plus activement. Je pense en particulier à ma famille (les remerciements constituent un genre particulier où les derniers cités sont souvent paradoxalement les plus importants aux yeux de l'auteur). Pour finir, je remercie Véronique pour son soutien et son aide très actifs.

Table des matières

Introduction

Le contrôle instationnaire des écoulements fluides, qui a pour but d'optimiser certaines de leurs caractéristiques, est actuellement un des enjeux les plus importants de l'aérodynamique. Par exemple, augmenter la portance et réduire la traînée d'un moyen de locomotion induisent une économie importante de carburant pour les transporteurs. Le problème du contrôle instationnaire fait ainsi l'objet d'études de plus en plus nombreuses, aussi bien dans les laboratoires de recherche que dans l'industrie. Elles ont été rendues possibles par la progression des moyens de calcul informatiques.

La simulation numérique des écoulements instationnaires reste néanmoins extrêmement coûteuse en calcul pour de nombreux cas industriels, et leur optimisation l'est encore plus. En effet, si la théorie de l'optimisation sous contrainte peut être exploitée pour le problème stationnaire de l'optimisation de forme (voir Jameson *et al.* [40] ou Mohammadi et Pironneau [64]), elle ne l'est souvent plus en pratique dans le cadre instationnaire du contrôle : la dimension des problèmes d'optimisation discrets alors obtenus par les méthodes classiques (différences, éléments ou volumes finis) est trop importante au regard des capacités de calcul actuelles. De nouvelles techniques ont donc vu le jour, comme les approches suboptimales (voir Kang *et al.* [42] ou Leclerc [54, 55]) ou les méthodes de construction de modèles réduits.

Les méthodes de modélisation réduite permettent de définir une description simplifiée d'un système physique complexe, en se limitant à un petit nombre de degrés de liberté. En mécanique des fluides, la plus populaire d'entre elles est la méthode POD-Galerkin, introduite dès 1967 par Lumley [62], et qui fait l'objet d'un effort de recherche croissant. Elle consiste à définir par la POD (*Proper Orthogonal Decomposition* ou décomposition orthogonale aux valeurs propres) une base de modes représentant de manière optimale l'état connu d'un système sur un intervalle de temps donné, puis, à partir des équations différentielles qui modélisent le système, à appliquer la méthode de Galerkine en ne considérant que les principaux modes POD. Cette méthode permet ainsi de construire, à partir de données numériques, un système d'EDOs (Équations Différentielles Ordinaires) de petite dimension par rapport à celle des systèmes obtenus par les discrétisations classiques, et conduit donc à un modèle qui peut être optimisé pour un coût informatique relativement faible.

À partir d'une base de modes POD, il est également possible de construire un système d'EDOs en utilisant une approche non variationnelle, appelée *subspace projection method*, ou encore en exploitant la théorie de Volterra : consulter [3], [61, 59] et [60, 58] pour

l'application de ces techniques à la modélisation d'écoulements fluides. Notons que les modes POD permettent également de construire des modèles réduits discrets en temps grâce aux théories statistiques d'identification de systèmes et des réseaux de neurones (voir [90] et [89] pour des illustrations en aérodynamique et sur l'écoulement incompressible 2D de Kolmogorov).

Deux alternatives à la POD peuvent fournir des fonctions spatiales représentatives du système étudié et être combinées avec les méthodes précédentes : des bases obtenues par ondelettes (voir [46] pour le cas des écoulements incompressibles) et des structures spatiales appelées *principal interaction patterns* (voir [50] pour les détails théoriques et une application aux équations de Ginzburg-Landau qui régissent l'évolution de l'amplitude des modes non linéairement stables de nombreux écoulements).

Enfin, d'autres approches permettent également de construire des modèles réduits, par exemple celles qui reposent sur la méthode d'Arnoldi (voir par exemple [53]) ou sur la linéarisation de l'équation d'évolution considérée autour d'un état connu, la solution perturbée étant décomposée dans une base de modes propres de l'opérateur différentiel correspondant (voir [24]) ou une base de modes POD (voir [59]).

La méthode POD-Galerkine est de plus en plus utilisée en mécanique des fluides pour l'analyse physique et le contrôle avec des résultats encourageants. En outre, la POD repose sur des bases mathématiques solides et des résultats théoriques ont montré la pertinence de la méthode POD-Galerkine pour la modélisation et le contrôle des problèmes paraboliques linéaires (voir Kunish *et al.* [49], Henri *et al.* [34, 33, 35]). Cependant, la méthode POD-Galerkine présente certaines difficultés, tant au niveau formel que numérique.

En effet, le traitement de la pression pour un écoulement incompressible pose problème car la modélisation est alors uniquement basée sur une décomposition du champ des vitesses (voir Rempfer [76], Galleti [25] ou Noack *et al.* [67]). De plus, le comportement des modèles réduits peut ne pas être satisfaisant voire être instable, à cause de leur sensibilité aux erreurs numériques ou encore, pour des écoulements de nombre de Reynolds relativement élevé, du fait de l'absence de prise en compte des petites échelles spatiales par les modes POD retenus (consulter Aubry [6], Iollo [38], Rempfer [77] ou Noack *et al.* [66, 67]). Enfin, notons que le coût informatique de la construction de modèles POD-Galerkine réduits devient important pour des écoulements à grand nombre de Reynolds.

La modélisation POD-Galerkine réduite a toutefois déjà été exploitée avec succès pour le contrôle d'écoulements instationnaires, bidimensionnels et laminaires (consulter par exemple [27, 28], [75, 74] ou [19]). En particulier, Graham *et al.* [27, 28] ont mis en évidence l'intérêt mais aussi la difficulté qu'il y avait à définir des stratégies d'enrichissement des bases POD. Ravindran [75, 74] a, quant à lui, obtenu un calcul de commande de contrôle satisfaisant par une stratégie itérative qui alterne simulation, modélisation réduite et optimisation. Fahl *et al.* [19] ont proposé un algorithme itératif plus sophistiqué, basé sur la théorie des méthodes d'optimisation *à région de confiance* (voir aussi [2] et [4]). Les problèmes liés à la modélisation POD-Galerkine mis à part, l'exploitation de ces différentes méthodes dans un contexte industriel n'est néanmoins pas encore d'actualité : leur utili-

sation pratique et leur efficacité restent mal connues, la littérature dans ce domaine étant encore relativement restreinte. Des activités de recherche pour une exploitation optimale des méthodes itératives de contrôle basées sur la modélisation réduite et le développement de stratégies d'enrichissement des bases POD apparaissent donc utiles, sinon nécessaires.

L'objectif qui sous-tend ce travail de thèse est le développement pratique d'algorithmes robustes et efficaces de contrôle instationnaire basés sur la modélisation réduite POD-Galerkine. Pour l'atteindre, deux conditions doivent être préalablement vérifiées : il faut disposer de méthodes numériques efficaces de construction de modèles POD-Galerkine et maîtriser l'exploitation de ces modèles pour le calcul d'une commande de contrôle. Comme nous l'avons vu plus haut, il est apparu qu'aucune de ces conditions n'était pleinement remplie. Il nous a donc semblé nécessaire d'y travailler, et nous nous sommes tout d'abord concentrés sur les problèmes qui se posent lors de l'application de la modélisation POD-Galerkine.

Notre premier objectif était d'analyser et de lever les difficultés inhérentes à la modélisation POD-Galerkine. Il constitue l'axe majeur des recherches qui vont être exposées. Pour cela, nous nous sommes essentiellement intéressés au cas des écoulements incompressibles. En effet, le contrôle instationnaire est le plus souvent appliqué à ces écoulements (c'est même exclusivement le cas dans le cadre de la modélisation POD-Galerkine à notre connaissance) : ils sont moins complexes et moins coûteux à simuler, donc plus faciles à contrôler, que les écoulements compressibles qui ont une variable scalaire d'état supplémentaire et modélisent une physique plus complexe (ils peuvent par exemple présenter des chocs). En outre, nous disposions d'une base de données issue de la simulation numérique d'un écoulement tridimensionnel, turbulent, inhomogène et incompressible, dont l'étude de la modélisation POD-Galerkine paraissait particulièrement intéressante (voir plus bas). Nous nous sommes donc tournés vers le cas des écoulements incompressibles dans le cadre de ce travail.

Tout d'abord, il fallait définir un modèle réduit cohérent qui soit explicitement fonction des mêmes effets environnementaux que l'écoulement, plus précisément des forces extérieures et des conditions aux limites. Or, si le traitement explicite des conditions de Dirichlet sur la vitesse et de certaines conditions de flux est connu, celui d'une condition de flux qui fait intervenir le tenseur des contraintes fluides n'est généralement pas résolu sur le plan formel, mais en négligeant certains termes de bord (voir [66] par exemple), et le traitement formel du terme de pression pose parfois problème. Ici, nous avons souhaité définir une modélisation qui tienne compte des conditions de Dirichlet et de flux usuelles, mais aussi déterminer l'origine du problème du terme de pression, voire le résoudre du point de vue formel. Ceci nécessitait principalement un travail de synthèse bibliographique.

Ensuite, l'objectif était d'étudier la modélisation POD-Galerkine d'un écoulement tridimensionnel et de nombre de Reynolds élevé, à travers l'observation des transferts d'énergie et des interactions entre les modes POD, afin notamment d'analyser les difficultés liées à la réduction du nombre de modes POD qui sont conservés pour construire le modèle. En effet, pour ce type d'écoulement, la construction de modèles réduits fiables est problématique car

les modes POD qui sont retenus ne sont pas en mesure de modéliser l'effet dissipatif des petites structures spatiales de l'écoulement. C'est pourquoi, suivant l'idée d'Aubry *et al.* [6], de nombreux auteurs augmentent artificiellement la viscosité du modèle pour recouvrer l'effet des petites structures (par exemple Podvin [70]). Si cette démarche est bien fondée dans le cadre de ces études (par l'utilisation d'une décomposition hybride POD/Fourier et de par la nature des écoulements étudiés), sa validité dans un cadre plus général reste cependant à être corroborée par des analyses physiques similaires à celles qui ont été réalisées par Webber [96] ou Rempfer *et al.* [78]. C'est ce que nous avons cherché à faire en exploitant les données numériques issues de la simulation d'un écoulement tridimensionnel turbulent qui franchit une marche descendante.

Par la suite, il fallait développer des méthodes robustes qui permettent de calculer des modèles réduits fiables et pour un coût informatique raisonnable, en essayant d'exploiter les informations obtenues lors de l'étude physique précédente. Afin d'atteindre cet objectif, nous avons opté pour une approche similaire à celle utilisée par Galletti [25] ou Delville [17], qui consiste à calibrer un modèle réduit en tirant profit des données temporelles qui sont fournies par la POD, mais qui ne sont pas utilisées dans la méthode de Galerkine.

Enfin, après cette étude de la modélisation POD-Galerkine, nous nous sommes intéressés aux techniques itératives d'utilisation des modèles réduits pour le contrôle des écoulements instationnaires afin de développer des algorithmes efficaces et robustes. Des essais numériques sont présentés à la fin de ce document.

Ce mémoire est organisé en cinq chapitres :

Chapitre 1 Le premier chapitre présente en détail la décomposition orthogonale aux valeurs propres (POD) sur les plans théorique et pratique. Les propriétés essentielles de cette décomposition (optimalité, biorthogonalité et transmission de certaines propriétés) y sont rappelées puis le calcul numérique est abordé.

Chapitre 2 La modélisation POD-Galerkine dans le cadre de la mécanique des fluides est le sujet du second chapitre. Après une présentation générale de cette modélisation et le rappel des équations de Navier-Stokes, nous proposons la construction formelle d'un modèle réduit pour un écoulement incompressible qui tient compte des conditions aux limites usuelles, et en particulier d'une condition de flux faisant intervenir le tenseur des contraintes fluides. Le cas des écoulements compressibles, plus délicat, est évoqué. Enfin, une dernière section illustre la modélisation POD-Galerkine réduite sur un exemple d'écoulement incompressible bidimensionnel laminaire.

Chapitre 3 Dans le troisième chapitre, un écoulement incompressible, tridimensionnel et turbulent qui franchit une marche descendante nous sert de cas d'étude. Une analyse qualitative et quantitative des interactions entre les modes POD au sein d'un modèle POD-Galerkine est proposée afin de cerner les conséquences de la réduction du nombre de modes POD du modèle. Ce travail a été publié dans [16].

Chapitre 4 Nous avons ensuite cherché à définir de nouvelles méthodes performantes de calibration numérique des modèles réduits, en essayant de tenir compte des conclu-

sions du chapitre précédent. La calibration permet une meilleure exploitation de l'information fournie par la POD et peut servir deux objectifs : diminuer les coûts de calculs des modèles réduits et améliorer leur comportement numérique. Trois méthodes originales, basées sur la résolution d'un problème d'optimisation, sont proposées et testées.

Chapitre 5 Enfin, l'optimisation d'un écoulement instationnaire via la définition de la commande d'un actionneur est abordée dans le cinquième chapitre. L'aspect théorique du contrôle d'un écoulement incompressible par un actionneur modélisé par une condition de Dirichlet sur la vitesse est suivi d'une illustration numérique réalisée dans le cas d'un écoulement laminaire, bidimensionnel et décollé.

Chapitre 1

La décomposition orthogonale aux valeurs propres (POD) : résumé théorique et calcul numérique

Sommaire

La décomposition aux valeurs propres, que l'on désignera par la suite par POD (acronyme de *Proper Orthogonal Decomposition*), est un outil théorique et pratique utilisé dans de nombreux domaines de recherche et connu sous plusieurs appellations : décomposition de Karhunen-Loève, fonctions propres de Sobolev, analyse en composantes principales ou encore décomposition en fonctions empiriques propres. Elle permet de définir, à partir d'une donnée $u(t)$ connue pour t parcourant un ensemble $\mathcal{T}$ discret ou continu, une base orthogonale et ordonnée optimale pour "représenter" u. Cette base sera ensuite utilisée pour identifier les degrés de liberté les plus représentatifs d'un écoulement et construire un modèle de petite dimension grâce à la méthode de Galerkin (voir chapitre 2).

Plus formellement, le principe de la POD est le suivant : pour $u(t)$ vivant dans un espace de Hilbert X muni du produit scalaire $(\,.\,,.\,)_X$ et de sa norme $\|.\|_X$ induite, et étant donné un opérateur de moyenne $\langle . \rangle$ sur $\mathcal{T}$, on va considérer les solutions du problème

$$\max_{\varphi \in X^*} \left\langle \left\| (\, u(t), \frac{\varphi}{\|\varphi\|_X}\,)_X \frac{\varphi}{\|\varphi\|_X} \right\|_X^2 \right\rangle,$$

où $X^* = X \setminus \{0\}$. Autrement dit, on maximise la norme de la projection de u sur la droite vectorielle de direction φ en moyenne sur $\mathcal{T}$. Notez qu'en général on préfère la formulation équivalente suivante :

$$\max_{\varphi \in X^*} \frac{\left\langle (\, u(t), \varphi\,)_X^{\,2} \right\rangle}{(\,\varphi, \varphi\,)_X}. \tag{1.1}$$

Comme nous allons le voir, il est possible, partant de ce problème, de définir une base optimale pour représenter u. La décomposition de u dans cette base est sa POD.

La base obtenue, dite *base POD* ou *base modale*, nous fournit dans un certain sens une analyse multirésolution optimale du sous-espace parcouru par $u(t)$ pour $t \in \mathcal{T}$, mais elle n'est *a priori* pas bien adaptée à un élément quelconque de X, comme pourrait l'être une base d'ondelettes par exemple. Il y a donc une distinction épistémologique importante entre, d'une part, certaines bases classiques utilisées en mathématiques appliquées, comme les bases polynômiales, d'éléments finis, de Fourier ou d'ondelettes, qui permettent de définir *a priori* un espace judicieux de travail de dimension fini dans lequel on va par exemple pouvoir résoudre un problème aux EDP, et, d'autre part, une base POD construite *a posteriori* à partir d'une référence u, dont l'intérêt et l'utilisation pratique sont différents.

Ce chapitre est consacré à la description de la décomposition aux valeurs propres et constitue une revue de l'état de l'art de la POD. À l'exception du paragraphe 1.4.4 qui nous semble original, les résultats présentés ici sont déjà connus et proviennent principalement de [34, 33] et [49] pour les résultats théoriques et [19] pour la partie traitant du calcul pratique.

La POD discrète (ensemble $\mathcal{T}$ discret) s'avère être un intermédiaire entre la POD continue (ensemble $\mathcal{T}$ continu), qui correspond au cadre théorique le plus large et le plus satisfaisant - mais aussi le plus laborieux -, et le calcul numérique pratique. Par souci de clarté,

nous avons ainsi choisi de présenter de manière rigoureuse et détaillée la POD discrète pour commencer (section 1.1). Ses principales propriétés sont données (section 1.1.2) ainsi qu'une méthode de construction d'une base POD, la méthode des clichés (section 1.1.3), qui sera utilisée dans les chapitres 3 à 5. La POD sera ensuite étendue au cas continu dans le paragraphe 1.2, et un lien sera établi avec la POD discrète. Le paragraphe 1.3 présente d'autres propriétés de la POD discrète, utilisées notamment pour la modélisation POD-Galerkine (chapitre 2). Le paragraphe 1.4 est consacré à la relation entre POD discrète et *Singular Value Decomposition* (SVD). L'équivalence entre ces deux méthodes est montrée, ce qui permet d'aborder le calcul numérique de la POD pour lequel les méthodes généralement utilisées sont présentées dans la section 1.5.

1.1 POD discrète et méthode des clichés

On suppose ici que l'on connaît u pour $N \in \mathbb{N}^*$ instants t_i : $\mathcal{T} = (t_i)_{i=1..N}$; on va ainsi définir une POD à partir de N *clichés* $u_i = u(t_i)$ de X. On suppose de plus qu'un cliché au moins n'est pas nul.

1.1.1 Définition

La moyenne temporelle considérée est la moyenne arithmétique sur les N instants t_i. Le problème (1.1) s'écrit donc

$$\max_{\varphi \in X^*} \frac{1}{N} \sum_{i=1}^{N} \frac{\left(u_i, \varphi \right)_X^2}{\left(\varphi, \varphi \right)_X}. \tag{1.2}$$

X est un espace réel de Hilbert séparable, en particulier $X = \mathrm{L}^2(\Omega)^n$ ($\Omega \subset \mathbb{R}^d$) ou $X = \mathbb{R}^n$ conviennent.

Proposition 1 *Le problème (1.2) admet au moins une solution.*

La démonstration de ce résultat (donnée dans [34] par exemple) est importante car elle nous donne un premier moyen de déterminer les solutions et va nous permettre de définir la POD, c'est pourquoi on en donne les grandes lignes.

Preuve. On définit l'opérateur

$$\begin{array}{rccl} K : & X & \longrightarrow & X \\ & \varphi & \longmapsto & K\varphi = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} \left(u_i, \varphi \right)_X u_i. \end{array}$$

K est linéaire continu mais aussi compact, auto-adjoint et positif. Son image $\mathrm{Im}(K) = \mathrm{Vect}(u_i)_{i=1..N} = Y$ est de dimension finie $d_Y \geq 1$ (au moins un cliché est non nul), et son

noyau est $\mathrm{Ker}(K) = Y^{\perp}$. K admet donc $d_Y \geq 1$ valeur(s) propre(s) strictement positive(s) que l'on peut ranger par ordre décroissant :

$$\lambda_1 \geq \cdots \geq \lambda_{d_Y} > 0.$$

K nous permet de redéfinir le problème (1.2) comme suit :

$$\text{Trouver } \phi^* \in X^* \text{ tel que } \frac{(\, K\phi^*, \phi^* \,)_X}{(\, \phi^*, \varphi \,)_X} = \max_{\varphi \in X^*} \frac{(\, K\varphi, \varphi \,)_X}{(\, \varphi, \varphi \,)_X}. \tag{1.3}$$

Soient ϕ^* une solution de (1.3) et $\varphi \in X^*$ quelconque. On définit $F_\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ par

$$F_\varphi(\varepsilon) = \frac{(\, K(\phi^* + \varepsilon\varphi), \phi^* + \varepsilon\varphi \,)_X}{(\, \phi^* + \varepsilon\varphi, \phi^* + \varepsilon\varphi \,)_X}.$$

Par définition, $F_\varphi(\varepsilon) \leq F_\varphi(0)$ donc $F'_\varphi(0) = 0$, ce qui nous donne la relation suivante

$$(\, \phi^*, \varphi \,)_X = \frac{(\, K\phi^*, \phi^* \,)_X}{(\, \phi^*, \phi^* \,)_X} \; (\, \phi^*, \varphi \,)_X$$

qui est valable pour tout $\varphi \in X^*$, mais aussi trivialement pour $\varphi = 0$. On obtient donc

$$K\phi^* = \lambda\phi^* \text{ en posant } \lambda = \frac{(\, K\phi^*, \phi^* \,)_X}{(\, \phi^*, \phi^* \,)_X}.$$

Réciproquement, tout vecteur propre φ^* correspondant à la plus grande valeur propre λ_1 de K satisfait (1.3). □

Ainsi, on sait définir l'ensemble des solutions de (1.2) par un problème aux valeurs propres.

Théorème 1 *L'ensemble des solutions de (1.2) est constitué des vecteurs propres de K associés à la plus grande valeur propre λ_1.*

Ce résultat nous pousse à considérer l'ensemble des sous-espaces propres de K et ses propriétés (ou encore à réitérer le processus à partir de u auquel on aurait au préalable soustrait sa projection sur l'espace propre que l'on vient d'obtenir...). D'autant plus que, K étant compact et autoadjoint, le *théorème spectral* ([10, p.97]) nous assure l'existence d'une base hilbertienne de vecteurs propres de $\mathrm{Im}(K) = \mathrm{Vect}(u_i)_{i=1..N} = Y$ dans laquelle on peut décomposer les clichés.

Définition 1 (base POD discrète) *Soit $\lambda_1 \geq \cdots \geq \lambda_{d_Y} > 0$ les valeurs propres strictement positives (les autres sont nulles) de K rangées par ordre décroissant. Pour tout $M \in [\![1, d_Y]\!]$, on appelle base POD (discrète) d'ordre M toute famille $(\varphi_i)_{i=1..M}$ de vecteurs propres orthonormaux associés. Ces vecteurs propres ordonnés sont appelés modes POD.*

"La" base POD est donc unique aux sous-espaces propres de K près. En particulier, si les valeurs propres λ_i sont toutes différentes, deux bases POD (φ_k) et $(\tilde{\varphi}_k)$ sont identiques au signe près : $\varphi_k = \pm\tilde{\varphi}_k$.

On peut bien parler de "décomposition" puisque pour toute base POD d'ordre $M = d_Y$

$$u(t_i) = \sum_{k=1}^{d_Y} (u(t_i), \varphi_k)_X \, \varphi_k \tag{1.4}$$

pour tout $i \in [\![1, N]\!]$. C'est cette équation qui est littéralement "la" POD discrète de u.

1.1.2 Optimalité et spectre POD

L'intérêt d'une base POD est qu'elle permet de représenter les clichés de manière optimale :

Proposition 2 (Propriété fondamentale de la POD discrète) *Soit $M \in [\![1, d_Y]\!]$ et (φ_k) une base POD d'ordre M. Alors pour toute famille orthonormale $(\psi_k)_{k=1..M}$, on a*

$$\frac{1}{N} \sum_{i=1}^{N} \left\| u_i - \sum_{k=1}^{M} (u_i, \varphi_k)_X \, \varphi_k \right\|_X^2 \leq \frac{1}{N} \sum_{i=1}^{N} \left\| u_i - \sum_{k=1}^{M} (u_i, \psi_k)_X \, \psi_k \right\|_X^2 . \tag{1.5}$$

Cette propriété est fondamentale puisque, réciproquement, toute famille orthonormale $(\varphi_k)_{k=1..M}$ la vérifiant est une base POD discrète d'ordre M.

Pour la démonstration de ce résultat, le lecteur pourra consulter [34] et [49].

Notons qu'une autre formulation équivalente, donnée dans [19], est de définir une POD discrète d'ordre M comme une solution de

$$\max_{\varphi_1 \cdots \varphi_M} \sum_{k=1}^{M} \frac{1}{N} \sum_{i=1}^{N} (u_i, \varphi_k)_X$$

sous la contrainte que les φ_k soient orthonormaux.

Il est important de souligner que les valeurs propres λ_k permettent de quantifier l'efficacité de toute POD :

Proposition 3 (Spectre POD) *On appelle spectre POD la famille ordonnée $(\lambda_k)_{k=1..d_Y}$ des valeurs propres non nulles de K, et on a notamment*

$$\frac{1}{N} \sum_{i=1}^{N} \| (u_i, \varphi_k)_X \, \varphi_k \|_X^2 = \lambda_k \tag{1.6}$$

$$et \quad \frac{1}{N} \sum_{i=1}^{N} \left\| u_i - \sum_{k=1}^{M} (u_i, \varphi_k)_X \, \varphi_k \right\|_X^2 = \sum_{k=M+1}^{d_Y} \lambda_k . \tag{1.7}$$

Le spectre POD correspond à la répartition de l'énergie de u (au sens de $\|.\|_X{}^2$) capturée en moyenne sur $\mathcal{T}$ par toute base POD.

1.1.3 Méthode des clichés

Nous donnons maintenant une technique pratique de construction d'une base POD, via un nouveau problème aux valeurs propres. Celle-ci est importante puisqu'elle va aboutir à une méthode de calcul d'une POD discrète pratique et de coût intéressant, qui est utilisée par la majorité des auteurs (voir la section 1.5 pour plus de détails).

Définition 2 (matrice des corrélations temporelles)

$$\tilde{K} = \left(\frac{1}{N} \, (\, u_i, u_j \,)_X \, \right) \in \mathbb{R}^{N \times N}$$

est appelée matrice des corrélations temporelles des clichés.

$\tilde{K}$ est une matrice symétrique réelle positive : elle est diagonalisable en base orthonormale et ses valeurs propres sont positives. Son rang est d_Y et ses valeurs propres sont : $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_{d_Y} \geq \tilde{\lambda}_{d_y+1} = \cdots = \tilde{\lambda}_N = 0$. Soient $v_k = (v_{1,k} \cdots v_{N,k})^T \in \mathbb{R}^N$ des vecteurs propres orthonormaux correspondants.

Proposition 4 (méthode des clichés) *On a $\tilde{\lambda}_k = \lambda_k$ pour tout $k \in [\![1, d_Y]\!]$. De plus, la famille $(\varphi_k)_{k=1..d_Y}$ définie par*

$$\varphi_k = \frac{1}{\sqrt{N \, \lambda_k}} \sum_{i=1}^{N} v_{i,k} \, u_i \tag{1.8}$$

est une base POD des clichés (d'ordre maximal $M = d_Y$). Réciproquement, pour toute base POD (φ_k) d'ordre M, (v_k) définie par

$$v_{i,k} = \frac{1}{\sqrt{N \, \lambda_k}} \, (\, u(t_i), \varphi_k \,)_X \tag{1.9}$$

est une famille orthonormale de vecteurs propres de $\tilde{K}$ associés respectivement aux valeurs propres $\lambda_1, \cdots, \lambda_M$.

Preuve. Partant de la définition (1.8) de (φ_k), on montre facilement que $K\varphi_k = \tilde{\lambda}_k \varphi_k$, que $\varphi_k \neq 0$, et que $(\, \varphi_i, \varphi_j \,)_X = \delta_{i,j}$ pour tout i et j de $[\![1, d_Y]\!]$. Réciproquement, partant de (1.9), les calculs montrent que $\tilde{K} \, v_k = \lambda_k \, v_k$, que $v_k \neq 0$, et que $(v_i)^T \, v_j = \delta_{i,j}$. □

L'équation (1.8) nous donne une technique de construction d'une base POD, appelée *méthode des clichés* depuis les travaux de Sirovich [88]. Elle sera utilisée dans tous les exemples numériques du mémoire (voir aussi les sections 1.4 et 1.5 pour plus de détails).

Une propriété de la POD très importante en pratique concernant la "transmission des conditions homogènes" peut être déduite de cette proposition (théorème 3, section 1.3.1).

1.1.4 Décomposition biorthogonale

Avec les notations de la proposition 4, on peut écrire la POD discrète (1.4) comme suit :

$$u(t_i) = \sqrt{N} \sum_{k=1}^{d_Y} \sqrt{\lambda_k}\, v_{i,k}\, \varphi_k$$

ou encore, en posant $a_k(t_i) = \dfrac{1}{\sqrt{\lambda_k}} \left(u(t_i), \varphi_k \right)_X = \sqrt{N}\, v_{i,k}$:

$$u(t_i) = \sum_{k=1}^{d_Y} \sqrt{\lambda_k}\, a_k(t_i)\, \varphi_k \tag{1.10}$$

où les *coefficients temporels* $a_k(t_i)$ *de la POD* satisfont la condition d'orthogonalité

$$\frac{1}{N} \sum_{k=1}^{N} a_i(t_k)\, a_j(t_k) = (v_i)^T\, v_j = \delta_{i,j}. \tag{1.11}$$

Toute base POD (d'ordre $M = d_Y$) aboutit à une POD sous cette forme "biorthogonale".

1.2 Extension au cas continu et lien avec la POD discrète

L'extension au cas continu s'avère utile pour l'analyse théorique. Elle consiste à considérer une donnée u de l'espace $\mathrm{L}^2(0,T,X)$ pour $0 < T < +\infty$: $\mathcal{T}$ est ici l'ensemble continu $[0,T]$. En outre, il existe un lien intéressant entre POD discrète et continue qui nous permet de faire le pont entre un cadre théorique, qui considère une donnée u définie en tout "temps" t de $\mathcal{T} = [0,T]$, et le calcul numérique pratique (qui utilise la méthode des clichés). Tous les résultats de cette section se trouvent dans [34].

L'opérateur de moyenne considéré est $\langle .(t) \rangle = \int_0^T .(t)\, dt$: le problème (1.1) devient

$$\max_{\varphi \in X^*} \frac{1}{T} \int_{[0,T]} \frac{\left(u(t), \varphi \right)_X{}^2}{\left(\varphi, \varphi \right)_X}\, dt. \tag{1.12}$$

Il est alors possible d'obtenir des résultats similaires à ceux de la POD discrète avec l'opérateur

$$\begin{array}{rccl} K : & X & \longrightarrow & X \\ & \varphi & \longmapsto & K\varphi = \dfrac{1}{T} \displaystyle\int_{[0,T]} \left(u(t), \varphi \right)_X\, u(t)\, dt, \end{array}$$

et l'opérateur des corrélations temporelles

$$\begin{array}{rccc} \tilde{K}: & \mathrm{L}^2([0,T]) & \longrightarrow & \mathrm{L}^2([0,T]) \\ & f & \longmapsto & \tilde{K}f \end{array}$$

défini par

$$(\tilde{K}f)(t) = \frac{1}{T}\int_{[0,T]} (\ u(t), u(s)\ )_X\ f(s)\, ds.$$

Remarque. Dans le cas $X = \mathrm{L}^2(\Omega)^d$ muni de son produit scalaire classique, une écriture matricielle nous donne

$$\begin{aligned} K\varphi &= \frac{1}{T}\int_{[0,T]} u(x',t)\ \int_\Omega u(x,t)^T \varphi(x)\, dx\, dt \\ &= \int_\Omega k(x,x')\,\varphi(x)\, dx \text{ avec } k(x,x') = \frac{1}{T}\int_{[0,T]} u(x',t)\, u(x,t)^T\, dt \end{aligned}$$

si u est suffisamment régulier pour pouvoir intervertir les symboles $\int_0^T . dt$ et $\int_\Omega . dx$. K est alors appelé *opérateur des autocorrélations* ou encore *opérateur des corrélations spatiales.*

$\tilde{K}$ est compact autoadjoint (c'est un opérateur de Hilbert-Schmidt, voir [10, p.99] et [34]) et positif : il admet une famille dénombrable de valeurs propres positives ordonnées, $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \tilde{\lambda}_k \geq \cdots \geq 0$, ainsi qu'une base orthonormale de vecteurs propres associés (v_k) de $\mathrm{L}^2([0,T])$, d'après le théorème spectral.

Dans le théorème suivant, on note $\mathrm{Sp}^*(K)$ la famille des valeurs propres non nulles de K comptées avec leur multiplicité (spectre de K sans 0).

Théorème 2 *Si $u \neq 0$, la plus grande valeur propre $\tilde{\lambda}_1$ de $\tilde{K}$ existe et est non nulle. De plus, $Sp^*(K) = Sp^*(\tilde{K})$ et $(\int_{[0,T]} u(t)\, v_k(t)\, dt)$ est une famille orthogonale de vecteurs propres de K associés aux valeurs propres respectives des (v_k). Enfin, l'ensemble des solutions de (1.12) est formé des vecteurs propres de K associés à $\tilde{\lambda}_1$.*

On peut à nouveau définir une base POD et étendre le principe de construction de la méthode des clichés :

Proposition 5 (base POD continue) *On note maintenant (λ_k) la famille décroissante des valeurs propres de $\tilde{K}$. Une famille orthonormale (φ_k) dans X de vecteurs propres de K associés aux valeurs propres non nulles de (λ_k) est appelée base POD de u. En particulier, la famille définie par*

$$\varphi_k = \frac{1}{\sqrt{T\lambda_k}}\int_{[0,T]} u(t)\, v_k(t)\, dt$$

est une base POD de u.

Les propositions 2 (optimalité) et 3 (spectre POD) se généralisent également :

Proposition 6 *Pour tout $M \geq 1$ et toute famille orthonormale (ψ_k) de X, on a*

$$\frac{1}{T}\int_{[0,T]} \left\| u(t) - \sum_{k=1}^{M} (u(t), \varphi_k)_X \, \varphi_k \right\|_X^2 dt \leq \frac{1}{T}\int_{[0,T]} \left\| u(t) - \sum_{k=1}^{M} (u(t), \psi_k)_X \, \psi_k \right\|_X^2 dt.$$

De plus,

$$\frac{1}{T}\int_{[0,T]} \left\| u(t) - \sum_{k=1}^{M} (u(t), \varphi_k)_X \, \varphi_k \right\|_X^2 dt = \sum_{k=M+1}^{+\infty} \lambda_k \xrightarrow[M\to+\infty]{} 0. \tag{1.13}$$

Une base POD est donc optimale pour représenter u, et la convergence normale (1.13) nous autorise à écrire la POD de u :

$$u(t) = \sum_{k=1}^{+\infty} (u(t), \varphi_k)_X \, \varphi_k$$

pour toute base POD (φ_k). A l'instar de (1.10) et (1.11) obtenues dans le cadre discret, la POD de u peut se mettre sous la forme

$$u(t) = \sum_{k=1}^{+\infty} \sqrt{\lambda_k}\, a_k(t)\, \varphi_k \tag{1.14}$$

avec des coefficients POD temporels $a_k(t) = \dfrac{1}{\sqrt{\lambda_k}} (u(t), \varphi_k)_X$ qui vérifient

$$\frac{1}{T}\int_{[0,T]} a_i(t)\, a_j(t)\, dt = \delta_{i,j}. \tag{1.15}$$

La décomposition (1.14) qui a été obtenue pour des familles (φ_k) et (a_k) orthogonales incite certains auteurs à parler de *décomposition biorthogonale* plutôt que de POD, comme Aubry *et al.* [5] qui définissent la décomposition d'un signal spatio-temporel $u \in \mathrm{L}^2([0,T]\times\Omega)$ ($\Omega \subset \mathbb{R}^d$) - on a $\mathrm{L}^2([0,T]\times\Omega) \subset \mathrm{L}^2(0,T,\mathrm{L}^2(\Omega))$ - d'un point de vue symétrique en temps et en espace, les modes a_k et φ_k étant alors placés sur un pied d'égalité.

Enfin, citons ce résultat [34] qui interprète la POD discrète comme un cas particulier de POD continue :

Proposition 7 *Soit u une fonction constante par morceaux sur une subdivision homogène de $[0,T]$ de pas $\tau > 0$, telle que $u(t) = u_i = u(t_i)$ pour tout $t \in]t_{i-1}, t_i]$ avec $t_i = i\tau$ et $t_N = i\tau = T$. Alors une POD continue de u est une POD discrète des Nclichés u_i.*

Ainsi la POD discrète est équivalente à une POD continue menée sur une approximation "d'ordre zéro" en temps de u définie sur une subdivision temporelle homogène. Il est

d'ailleurs tentant d'essayer de définir une POD discrète "d'ordre supérieur" en utilisant par exemple des fonctions polynômiales par morceaux (en temps) de degré supérieur à zéro pour approcher u (voir la section 1.4.4). En pratique, il est donc cohérent de calculer une base POD à partir de clichés espacés de manière régulière dans le temps.

Enfin, notons que la POD continue et sa relation avec la POD discrète s'avèrent être des outils utiles pour l'analyse théorique, par exemple dans le cadre de l'étude de la convergence de la méthode POD-Galerkine (voir [33]).

1.3 Autres propriétés de la POD

1.3.1 Transmission des conditions homogènes

Citons un résultat connu qui est utilisé pour construire des modèles réduits fluides par la méthode de Galerkine (voir le chapitre suivant) :

Théorème 3 *Soient $(\varphi_k)_{k=1..M}$ une base POD d'ordre M de N clichés $\mathcal{U} = (u(t))_{t\in\mathcal{T}}$ de X, Y un sous-espace de X, Z un espace vectoriel normé et $l \in \mathcal{L}(Y,Z)$. Si $u(t) \in \tilde{Y} = \{v \in Y \,/\, l(v) = 0\}$ pour tout $t \in \mathcal{T}$ alors $\varphi_k \in \tilde{Y}$ pour tout $k \in [\![1, M]\!]$.*

Preuve. D'après la proposition 4, tout mode POD φ_k est combinaison linéaire de clichés $u(t)$. □

Ce théorème est très important en pratique pour la modélisation POD-Galerkine (voir le chapitre suivant, section 2.3.1). Par exemple pour $X = \mathrm{L}^2(\Omega)$, il nous donne les résultats suivants :

- si $Y = \mathrm{H}^1(\Omega)$ et si $l : Y \longrightarrow Z = \mathrm{H}^{1/2}(\Gamma)$ est l'opérateur trace sur un bord $\Gamma \subset \partial\Omega$ de mesure non nulle, il y a **transmission des conditions de Dirichlet homogènes** aux modes POD ;
- de même, il y a **transmission des conditions de Dirichlet périodiques** aux modes POD ;
- si $Y = \mathrm{H}^{\mathrm{div}}(\Omega)$ et $l = \nabla.$ est l'opérateur de divergence, il a **transmission du caractère solénoïdal** aux modes POD.

Remarque. Holmes *et al.* [36] proposent le même résultat pour une POD complexe menée sur l'espace $\mathrm{L}^2(\Omega)$) et pour un opérateur linéaire $\langle . \rangle$ défini sur un ensemble $\mathcal{T} = (t_i)$ dénombrable. Le théorème devrait pouvoir être étendu au cas continu.

1.3.2 Cas d'une direction homogène

Cette section expose la relation entre POD et décomposition de Fourier dans le cas d'une donnée qui est périodique et homogène dans une direction de l'espace. Cette relation n'est pas essentielle dans la mesure où elle est établie dans un cas très particulier. Cependant il est utile de la mentionner puisqu'elle est évoquée dans certains travaux dont on parlera par

la suite, en particulier pour justifier une décomposition hybride POD/Fourier du champ des vitesses d'un écoulement turbulent comportant une direction périodique (et supposé homogène sur un intervalle de temps $[0,T]$ suffisamment long).

Afin de ne pas déborder du cadre théorique qui a été fixé ici, à savoir un espace X de Hilbert réel, ce résultat est présenté pour une POD menée dans l'espace X des fonctions de carré intégrable à valeurs réelles, alors qu'il est en général abordé dans le cas d'une POD définie dans l'espace des fonctions de carré intégrable à valeurs complexes (voir [36] par exemple) ; en effet, la POD pourrait se généraliser aux espaces de Hilbert complexes. De plus, la POD est généralement effectuée sur des données réelles.

Considérons une POD continue d'une référence u de $\mathrm{L}^2(0,T,X)$ où $X = \mathrm{L}^2_{\text{période}}([0,D]) \cap \mathcal{F}(\mathbb{R},\mathbb{R})$ avec $0 < D < +\infty$, $\mathcal{F}(\mathbb{R},\mathbb{R})$ l'ensemble des applications de $\mathbb{R}$ dans $\mathbb{R}$ et $\mathrm{L}^2_{\text{période}}([0,D])$ l'espace de Hilbert

$$\mathrm{L}^2_{\text{période}}([0,D]) = \{f : \mathbb{R} \longrightarrow \mathbb{C} \,/\, f(x) = f(x+D) \text{ p.p. et } \int_0^D |f(x)|^2\,dx < +\infty\}$$

muni du produit scalaire

$$(\ f,g\ )_X = \frac{1}{D}\int_0^D f(x)\overline{g(x)}\,dx.$$

$\mathrm{L}^2_{\text{période}}([0,D])$ est séparable et admet comme base hilbertienne la famille $(x \longmapsto \mathrm{e}^{i\frac{2\pi}{D}k\,x})_{k\in\mathbb{Z}}$ des modes de Fourier. L'espace X, qui est le sous-espace de $\mathrm{L}^2_{\text{période}}([0,D])$ des fonctions à valeurs réelles, est un espace de Hilbert réel séparable qui admet comme base hilbertienne $x \longmapsto 1$, $x \longmapsto \cos(\frac{2\pi}{D}k\,x)$ et $x \longmapsto \sin(\frac{2\pi}{D}k\,x)$, pour $1 \leq k < +\infty$, des modes réels de Fourier (pour le même produit scalaire).

On suppose de plus que u est *homogène*, c'est-à-dire, avec les notations de la remarque de la page 14, que

$$\exists \tilde{k} \in X \quad k(x,x') = \tilde{k}(x-x') \quad \text{p.p. sur } \mathbb{R}$$

($\tilde{k}$ est nécessairement D-périodique puisque u l'est) : les corrélations spatiales sur $[0,T]$ ne dépendent que de la distance entre les deux points considérés (invariance par translation).

On peut décomposer $\tilde{k} \in \mathrm{L}^2_{\text{période}}([0,D])$ en série de Fourier complexe :

$$\tilde{k}(y) = \sum_{p=-\infty}^{+\infty} c_p\,\mathrm{e}^{i\frac{2\pi}{D}p\,y} \quad \text{avec} \quad c_p = \frac{1}{D}\int_0^D \tilde{k}(y)\,\mathrm{e}^{-i\frac{2\pi}{D}p\,y}.$$

On a $c_p = \overline{c_{-p}}$ puisque $\tilde{k}(y)$ est réel ; en fait, k étant symétrique, $\tilde{k}$ est une fonction paire et $c_p = \mathrm{Re}(c_p) = c_{-p} = \dfrac{1}{D}\displaystyle\int_0^D \tilde{k}(y)\cos(\frac{2\pi}{D}p\,y)\,dy.$

Le calcul montre que

$$\int_0^D k(x,x')\,\mathrm{e}^{i\frac{2\pi}{D}l\,x'}\,dx' = \int_0^D \sum_{p=-\infty}^{+\infty} c_p\,\mathrm{e}^{i\frac{2\pi}{D}p\,x}\mathrm{e}^{i\frac{2\pi}{D}(l-p)\,x'}\,dx' = c_l\,\mathrm{e}^{i\frac{2\pi}{D}l\,x}.$$

Ainsi, par linéarité et d'après les formules de Moivre, on obtient, en notant $z = c_l \, \mathrm{e}^{i\frac{2\pi}{D}l\,x}$,

$$\int_0^D k(x,x') \, \cos(\frac{2\pi}{D} l \, x') \, dx' = \frac{1}{2}(z + \overline{z}) = \mathrm{Re}(z) = c_l \, \cos(\frac{2\pi}{D} l \, x)$$

puisque c_l est réel, et de même

$$\int_0^D k(x,x') \, \sin(\frac{2\pi}{D} l \, x') \, dx' = \frac{1}{2\,i}(z - \overline{z}) = \mathrm{Im}(z) = c_l \, \sin(\frac{2\pi}{D} l \, x).$$

Donc les modes $x \longmapsto 1$, $x \longmapsto \sqrt{2} \, \cos(\frac{2\pi}{D} k \, x)$ et $x \longmapsto \sqrt{2} \, \sin(\frac{2\pi}{D} k \, x)$, pour $1 \leq k < +\infty$, forment une famille orthonormale de vecteurs propres de K associée aux valeurs propres c_0, c_k et c_k respectivement. Cette famille engendre l'espace X, donc tous les sous-espaces propres de K.

Proposition 8 *Les modes de Fourier réels $x \longmapsto 1$, $x \longmapsto \sqrt{2} \, \cos(\frac{2\pi}{D} k \, x)$ et $x \longmapsto \sqrt{2} \, \sin(\frac{2\pi}{D} k \, x)$, pour $1 \leq k < +\infty$, forment une base POD pour $X = L^2_{\mathit{période}}([0,D]) \cap \mathcal{F}(\mathbb{R},\mathbb{R})$ pour toute donnée $u \in L^2(0,T,X)$ homogène en espace, à condition de les ordonner en fonction des valeurs propres respectives c_0, c_k et c_k associées.*

On retrouve le résultat qui est proposé habituellement dans le cas complexe (par exemple dans [36]), à savoir que les modes $(x \longmapsto \mathrm{e}^{i\frac{2\pi}{D}k\,x})_{k\in\mathbb{Z}}$ forment une base POD. En effet, le sous-espace engendré par $x \longmapsto \mathrm{e}^{i\frac{2\pi}{D}k\,x}$ et $x \longmapsto \mathrm{e}^{-i\frac{2\pi}{D}k\,x}$ est aussi engendré par combinaison linéaire complexe de $x \longmapsto \cos(\frac{2\pi}{D} k \, x)$ et $x \longmapsto \sin(\frac{2\pi}{D} k \, x)$ et ces deux modes sont associés au même sous-espace propre de valeur propre $c_k = c_{-k}$.

Remarque. La condition de périodicité sur u n'est pas restrictive, car toute donnée qui admet une base POD composée de modes de Fourier est nécessairement périodique : toute fonction non-périodique ne peut avoir une base POD composée de mode de Fourier même si elle est homogène sur $\Omega = \mathbb{R}$.

Il faut également noter que la périodicité ne suffit pas. En effet, $u(x,t) = f(x)\,g(t) \neq 0$ avec $f \in X = \mathrm{L}^2_{\mathrm{période}}([0,D])$ et $g \in \mathrm{L}^2([0,T])$ admet exactement deux bases POD, qui sont $(f/\|f\|_X)$ et $(-f/\|f\|_X)$: si f n'est pas un mode de Fourier, les bases POD de la donnée D-périodique u n'admettent aucun mode de Fourier (par exemple, dans le cas $f(x) = x\,(x-D)$ p.p sur $[0,D]$).

Ainsi, dans le cas où $X = \mathrm{L}^2(\Omega)^n$ avec $\Omega \subset \mathbb{R}^d$ infini dans la direction x_i $(1 \leq i \leq d)$, et où $u \in \mathrm{L}^2(0,T,X)$ est homogène et D-périodique dans la direction x_i, une POD revient moralement à une décomposition de Fourier dans cette direction. Cette propriété explique pourquoi de nombreux auteurs ([6] et [70] par exemple) choisissent une décomposition hybride Fourier/POD pour des écoulements turbulents comportant des directions périodiques supposées quasi-homogènes.

La notion d'homogénéité n'est employée que pour des directions suivant lesquelles l'écoulement se prolonge indéfiniment, en fait pour des directions périodiques dans la pratique : qualifier une direction d'homogène sous-entend que l'écoulement est périodique dans cette direction.

Dans la littérature, le terme de *méthode spectrale* est régulièrement employé pour définir la POD et, bien que ceci ne semble pas rigoureusement exact, de nombreux auteurs décrivent la POD comme une extension de la décomposition de Fourier dans le cas non-périodique.

1.4 POD discrète pour un espace de dimension finie et SVD (*Singular Value Decomposition*)

Dans cette section, nous nous plaçons dans le cadre d'une POD discrète et d'un espace de Hilbert réel X de dimension finie P : c'est le cas pratique où l'on cherche à calculer une POD à partir d'une base de données numériques. Nous verrons que la POD est alors équivalente à une décomposition matricielle classique nommée SVD (*Singular Value Decomposition*). Ceci nous aidera à aborder le calcul pratique de la POD à la section 1.5. En effet, le développement d'algorithmes efficaces de calcul SVD est déjà ancien, en particulier toutes les bibliothèques numériques classiques d'algèbre linéaire proposent des routines de SVD performantes.

1.4.1 Notation

Soit $S \in \mathbb{R}^{n\times n}$ une matrice symétrique définie positive. Par abus de notation, nous noterons $S^{1/2}$ toute matrice de $\mathbb{R}^{n\times n}$ telle que

$$S^{1/2}\left(S^{1/2}\right)^T = S.$$

Il existe au moins une matrice $S^{1/2}$ satisfaisant cette propriété, qui peut être définie par la racine carrée ou le facteur de Cholesky (la racine carrée peut être un facteur de Cholesky, par exemple si S est diagonale). En effet, S peut se diagonaliser dans une base orthonormale d'après le théorème spectral : en notant α_k ses valeurs propres (strictement positives), il existe une matrice orthonormale Q telle que $S = Q^T \operatorname{diag}(\alpha_1, \cdots, \alpha_n)\, Q$ et telle que $S^{1/2} = Q^T \operatorname{diag}(\sqrt{\alpha_1}, \cdots, \sqrt{\alpha_n})\, Q$ soit la racine carrée de S ($\left(S^{1/2}\right)^2 = S$ et $\left(S^{1/2}\right)^T = S^{1/2}$). De plus, d'après le théorème de factorisation de Cholesky (voir [12, p.87]), il existe $L \in \mathbb{R}^{n\times n}$ triangulaire inférieure telle que $L\,L^T = S$: $S^{1/2} = L$ convient.

En pratique, il est possible, en utilisant les routines d'une bibliothèque numérique d'algèbre linéaire, de calculer une diagonalisation de S (donc une racine carrée) ou une factorisation de Cholesky. Cette dernière solution est en général la moins coûteuse en calculs.

1.4.2 La relation entre POD discrète et SVD

Ainsi, on dispose d'une base de données $\mathcal{U} = (u(t_j))_{j=1..N} \in X^N$ de N clichés $u(t_j) = u_j$ connus numériquement par leurs coefficients réels dans une base de X notée $\mathcal{X} = (\mathcal{X}_i)_{i=1..P}$, par exemple une base d'éléments finis. Ces coefficients seront notés $y_{i,j}$:

$$\forall j \in [\![1, N]\!] \qquad u(t_j) = u_j = \sum_{i=1}^{P} y_{i,j}\, \mathcal{X}_i.$$

Soit y_j la représentation vectorielle de u_j : $y_j = (y_{1,j} \cdots y_{P,j})^T \in \mathbb{R}^P$.

Définition 3 (matrice des clichés)

$$U = \begin{pmatrix} y_{1,1} & \cdots & y_{1,N} \\ \vdots & & \vdots \\ y_{P,1} & \cdots & y_{P,N} \end{pmatrix} \in \mathbb{R}^{P \times N}$$

est appelée la matrice des clichés de $\mathcal{U}$ dans la base $\mathcal{X}$.

De plus, on note S_P la matrice symétrique définie positive associée au produit scalaire $(\,.\,,.\,)_X$ de X et à la base $(\mathcal{X}_j)$:

$$S_P = \left(\, (\, \mathcal{X}_i, \mathcal{X}_j \,)_X \, \right) \in \mathbb{R}^{P \times P}.$$

En conséquence,

$$\forall (i,j) \qquad (\, u_i, u_j \,)_X = (\, y_i, y_j \,)_{S_P}$$

avec

$$\forall (y,z) \in (\mathbb{R}^P)^2 \quad (\, y, z \,)_{S_P} = y^T S_P\, z \quad \text{et} \quad \|y\|_{S_P} = \sqrt{(\, y, y \,)_{S_P}} = \left\| (S_P^{1/2})^T\, y \right\|_2.$$

Pour toute matrice A, on note $A_{:,i}$ sa $i$ème colonne. La POD discrète (1.10) peut alors s'écrire sous la forme matricielle suivante :

$$U = \Phi \; \Sigma \; A^T \tag{1.16}$$

où

- $d_Y = \mathrm{Rang}(U)$,
- $\sigma_k = \sqrt{\lambda_k}$ pour tout k, et $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_{d_Y})$,
- $\Phi \in \mathbb{R}^{P \times d_Y}$ est la *matrice des modes POD* dont les colonnes $\Phi_{:,j}$ sont les vecteurs des coefficients des modes POD φ_j dans la base $\mathcal{X}$,
- $A \in \mathbb{R}^{N \times d_Y}$ est la *matrice des coefficients POD temporels* définie par $A_{:,k} = a_k = (a_k(t_1) \cdots a_k(t_N))^T$,

- et $S_N = \dfrac{1}{N}\mathrm{I}_N$.

D'après la propriété de biorthogonalité de la POD, ces matrices vérifient de plus

$$\Phi^T S_P \Phi = \mathrm{I}_{d_Y} \quad \text{et} \quad A^T S_N A = \mathrm{I}_{d_Y}. \tag{1.17}$$

Les équations (1.16) et (1.17) nous montrent donc clairement que la POD donne une SVD généralisée et tronquée de la matrice U :

Proposition 9 (SVD généralisée) *Soient $U \in \mathbb{R}^{P\times N}$ non nulle quelconque de rang d ($1 \leq d \leq \max(P,N)$), $S_P \in \mathbb{R}^{P\times P}$ et $S_N \in \mathbb{R}^{N\times N}$ deux matrices symétriques définies positives. Alors, pour $(M,L) = (P,N)$ ou $(M,L) = (d,d)$, il existe $\Phi \in \mathbb{R}^{P\times M}$, $A \in \mathbb{R}^{N\times L}$ et $\sigma_1 \geq \cdots \geq \sigma_d > 0$ tels que*

$$\Phi^T S_P \Phi = I_M, \qquad A^T S_N A = I_L \tag{1.18}$$

et

$$U = \Phi\ \Sigma\ A^T \tag{1.19}$$

où $\Sigma = \mathrm{diag}_{M\times L}(\sigma_1, \cdots, \sigma_d)$. Cette décomposition est appelée SVD de U, les colonnes de Φ et A respectivement vecteurs singuliers gauches et droits. Les σ_k, appelées valeurs singulières, sont uniques. La décomposition est qualifiée de tronquée pour $(M,L) = (d,d)$, et de généralisée (relativement à S_P et S_N) sauf si $S_P = I_P$ et $S_N = I_N$.

Nous proposons maintenant une démonstration de cette proposition par analyse-synthèse : cela va nous permettre notamment de montrer que toute SVD généralisée correspond réciproquement à une POD, mais aussi d'appréhender le calcul numérique d'une SVD (et donc d'une POD).

Preuve. Le résultat va être démontré pour $M = L = d$: il suffit de compléter les colonnes de Φ, respectivement de A, afin d'obtenir une base de $\mathbb{R}^P$ orthonormale au sens de S_P, respectivement une base de $\mathbb{R}^N$ orthonormale au sens de S_N, et de compléter Σ par des zéros, pour passer d'une SVD tronquée ($M = L = d$) à une SVD ($M = P$ et $L = N$).

Une analyse montre que (1.18) et (1.19) impliquent

$$\tilde{K}\, A = A\, \Sigma^2 \quad \text{avec} \quad \tilde{K} = U^T S_P\, U\, S_N \quad \text{et} \quad \Phi = U\, S_N\, A\, \Sigma^{-1}, \tag{1.20}$$

$$K\, \Phi = \Phi\, \Sigma^2 \quad \text{avec} \quad K = U\, S_N\, U^T S_P \quad \text{et} \quad A = U^T S_P\, \Phi\, \Sigma^{-1}. \tag{1.21}$$

Réciproquement (synthèse à partir de (1.20)), la matrice $B_1 = \left(S_N^{1/2}\right)^T \tilde{K} \left(S_N^{1/2}\right)^{-T} = \left(U\, S_N^{1/2}\right)^T S_P\, U\, S_N^{1/2}$ étant symétrique, réelle, positive et non nulle, elle admet $d > 0$ valeurs propres $\lambda_1 \geq \cdots \geq \lambda_d > 0$ et $N - d$ valeurs propres nulles. Il est également possible de lui associer une matrice $\tilde{A} \in \mathbb{R}^{N\times d}$ de vecteurs propres orthonormaux :

$$B_1\, \tilde{A} = \tilde{A}\, \Sigma^2 \quad \text{et} \quad \tilde{A}^T \tilde{A} = \mathrm{I}_d \quad \text{avec} \quad \Sigma = \mathrm{diag}(\sqrt{\lambda^1}, \cdots, \sqrt{\lambda_N}).$$

Alors, en posant $A = (S_N^{1/2})^{-T}\,\tilde{A}$, $\Phi = U\,S_N\,A\,\Sigma^{-1}$ et $\sigma_k = \sqrt{\lambda_k}$, on obtient bien une SVD de U.

On peut faire le même type de synthèse en partant de (1.21) : soit $\tilde{\Phi}$ une solution orthonormale du problème aux valeurs propres $B_2\,\tilde{\Phi} = \tilde{\Phi}\,\Sigma^2$ avec $B_2 = (S_P^{1/2})^T\,K\,(S_P^{1/2})^{-T} = (S_P^{1/2})^T\,U\,S_N\,U^T\,S_P^{1/2}$. Les matrices $\Phi = (S_P^{1/2})^{-T}\,\tilde{\Phi}$, $A = U^T\,S_P\,\Phi\,\Sigma^{-1}$ et Σ nous donnent une SVD de U. □

Cette démonstration nous donne de plus l'équivalence entre POD et SVD pour un espace X de dimension finie : le problème (1.20) défini par la matrice K équivaut au problème des valeurs propres de l'opérateur K de la section 1.1 et Φ correspond donc bien à une base POD (telle que définie par la définition 1). La construction d'une SVD donnée par (1.20) correspond à la méthode des clichés (proposition 4), puisque $\tilde{K} = U^T\,S_P\,U\,S_N$ est bien la matrice des corrélations temporelles (définition 2).

Théorème 4 (Équivalence POD/SVD) *Soient $\mathcal{U}$ un ensemble de clichés d'un espace réel X de Hilbert séparable de dimension finie, et $\mathcal{X} = (\mathcal{X}_i)$ une base de X. Alors à toute POD discrète de $\mathcal{U}$ correspond une SVD généralisée tronquée de la matrice U des clichés de $\mathcal{U}$ dans $\mathcal{X}$ et réciproquement. Plus précisément, pour toute base POD d'ordre $M \leq Rang(U)$, les modes POD φ_k, leurs coefficients temporels a_k et les valeurs propres associées λ_k du spectre POD correspondent respectivement aux M premiers vecteurs singuliers gauches $\Phi_{:,k}$ et droits $A_{:,k}$, et aux carrés ${\sigma_k}^2$ des M premières valeurs singulières d'une SVD (Φ, Σ, A) généralisée de U pour $S_P = ((\,\mathcal{X}_i, \mathcal{X}_j\,)_X)$ et $S_N = \frac{1}{N} I_N$.*

De plus, avec les notations de la démonstration précédente, on a $\tilde{\Phi}\,\Sigma\,\tilde{A}^T = (S_P^{1/2})^T\,U\,S_N^{1/2}$: $(\tilde{\Phi}, \Sigma, \tilde{A})$ est une SVD de $(S_P^{1/2})^T\,U\,S_N^{1/2}$. En fait, on a les résultats suivants :

Proposition 10 *Une SVD (Φ, Σ, A) généralisée (tronquée ou non) de $U \in \mathbb{R}^{P\times N}$, de rang $d \geq 1$, peut être obtenue par une SVD vérifiant*

$$(S_P^{1/2})^T\,U\,S_N^{1/2} = \tilde{\Phi}\,\Sigma\,\tilde{A}^T \quad avec \quad \tilde{\Phi}^T\,\tilde{\Phi} = \tilde{A}^T\,\tilde{A} = I_d \tag{1.22}$$

en posant

$$\Phi = (S_P^{1/2})^{-T}\,\tilde{\Phi} \quad et \quad A = (S_N^{1/2})^{-T}\,\tilde{A}. \tag{1.23}$$

En outre, si $S_N^{1/2} = \frac{1}{\sqrt{N}} I_N$, alors (Φ, Σ, A) peut être obtenu par une SVD de $\tilde{U} = (S_P^{1/2})^T\,U$ vérifiant

$$(S_P^{1/2})^T\,U = \tilde{\Phi}\,\hat{\Sigma}\,\hat{A}^T \quad avec \quad \tilde{\Phi}^T\,\tilde{\Phi} = \hat{A}^T\,\hat{A} = I_d \tag{1.24}$$

en posant

$$\Phi = (S_P^{1/2})^{-T}\,\tilde{\Phi}, \quad \Sigma = \frac{1}{\sqrt{N}}\,\hat{\Sigma}, \quad et \quad A = \frac{1}{\sqrt{N}}\,\hat{A}. \tag{1.25}$$

Preuve. Les équations (1.22) et (1.23) impliquent $\Phi\,\Sigma\,A^T = U$ et $\Phi^T\,S_P\,\Phi = A^T\,S_N\,A = \mathrm{I}_d$. De même, (1.24) et (1.25) donnent ce résultat pour $S_N = \frac{1}{N}\,\mathrm{I}_N$. □

La deuxième partie de cette proposition sera exploitée dans la section 1.5.1.

1.4.3 Optimalité de la SVD

Habituellement, la relation entre POD et SVD est abordée différemment, via la propriété fondamentale de la POD discrète (proposition 2). Cette manière d'appréhender l'équivalence POD/SVD, qui est par exemple explicitée dans [19], nous donne une interprétation matricielle de la propriété fondamentale de la POD.

En effet, en notant Φ_M la matrice des M premiers modes POD d'une base d'ordre supérieur à M et $\Phi_{:,k}$ sa $k$ème colonne, la propriété (1.5) peut s'écrire :

$$\min_{\Phi_M \in \mathbb{R}^{P\times M}} \quad \sum_{i=1}^{N} \left\| u_i - \sum_{k=1}^{M} (y_i, \Phi_{:,k})_{S_P} \Phi_{:,k} \right\|_{S_P}^2 \tag{1.26}$$

sous la contrainte que $(\Phi_{:,k})$ soit orthonormale au sens de $(. , .)_{S_P}$. En notant $\|.\|_F$ la norme de Frobenius définie par $\|U\|_F^2 = \sum_{k=1}^{N} \|U_{:,k}\|_2^2$ pour $U \in \mathbb{R}^{P\times N}$, il est possible de réécrire (1.26) sous la forme du problème classique de la meilleure représentation de la matrice des clichés modifiée $\tilde{U} = (S_P^{1/2})^T U$ par une matrice de rang M au sens de la norme de Frobenius :

Proposition 11 *Le problème (1.26) est équivalent à*

$$\min_{\Phi_M} \left\| \tilde{U} - B \right\|_F^2 \text{ avec } B = (S_P^{1/2})^T \Phi_M \Phi_M{}^T S_P U \text{ sous la contrainte } \Phi_M{}^T S_P \Phi_M = I_M,$$

ou encore

$$\min_{\tilde{\Phi}_M} \left\| \tilde{U} - \tilde{\Phi}_M \tilde{\Phi}_M^T \tilde{U} \right\|_F^2 \text{ avec } \tilde{\Phi}_M = (S_P^{1/2})^T \Phi_M \text{ sous la contrainte } \tilde{\Phi}_M^T \tilde{\Phi}_M = I_M. \tag{1.27}$$

Preuve. On a $\|.\|_{S_P}{}^2 = \left\| (S_P^{1/2})^T . \right\|_2^2$ et $\left(\Phi_M \Phi_M{}^T S_P U \right)_{:,i} = \sum_{k=1}^{M} (y_i, \Phi_{:,k})_{S_P} \Phi_{:,k}$. □

Le problème (1.27) signifie que l'on recherche un sous-espace de $\mathbb{R}^P$ de dimension M de base orthonormale $(\tilde{\Phi}_{:,k})$ tel que les projections des colonnes de $\tilde{U}$ sur celui-ci, c'est-à-dire les colonnes de $\tilde{\Phi}_M \tilde{\Phi}_M^T \tilde{U}$, donnent une meilleure (ou aussi bonne) approximation de $\tilde{U}$ que tout autre sous-espace de dimension M au sens de $\|.\|_F$.

Ainsi, il apparaît que (1.27) est un cas particulier du problème classique

$$\min_{B \in \mathbb{R}^{P\times N}} \quad \left\| \tilde{U} - B \right\|_F^2 \quad \text{sous la contrainte} \quad \mathrm{Rang}(B) = M,$$

dont la solution est donnée par le théorème suivant :

Théorème 5 (Eckart et Young) *Soit $\tilde{U} \in \mathbb{R}^{P\times N}$ de rang $d > 1$ et $(\tilde{\Phi}, \hat{\Sigma} = \mathrm{diag}(\hat{\sigma}_1, \cdots, \hat{\sigma}_d), \hat{A})$ un triplet donnant une SVD de $\tilde{U}$. Alors*

$$\min_{Rang(B)=M} \quad \left\| \tilde{U} - B \right\|_F = \left\| \tilde{U} - \tilde{U}_M \right\|_F = \sqrt{\sum_{i=k+1}^{d} \hat{\sigma}_i^2}$$

$$où \qquad \tilde{U}_M = \tilde{\Phi}_M \hat{\Sigma}_M \hat{A}_M^T$$

avec $\tilde{\Phi}_M$ *et* $\hat{A}_M$ *les matrices des* M *premières colonnes de* $\tilde{\Phi}$ *et* $\hat{A}$, *et* $\hat{\Sigma}_M = \mathrm{diag}(\hat{\sigma}_1, \cdots, \hat{\sigma}_M)$.

Ce résultat peut se généraliser pour une norme matricielle $\|.\|_{S_P,S_N}$ définie par deux matrices symétriques définies positives S_P et S_N, considérant une SVD généralisée de U associée (voir [30] pour les détails).

La proposition 11 et le théorème 5 sont cohérents avec la proposition 10 et le théorème 4 puisqu'ils aboutissent à la même conclusion :

Proposition 12 *Une base POD d'ordre* M *de* $\mathcal{U}$ *de coefficients* Φ_M *dans* $\mathcal{X}$ *est obtenue par une SVD (non généralisée)* $(\tilde{\Phi}, \hat{\Sigma}, \hat{A})$ *de* $\tilde{U} = \left(S_P^{1/2}\right)^T U$ *en posant* $\Phi_M = \left(S_P^{1/2}\right)^{-T} \tilde{\Phi}_M$.

1.4.4 POD discrète d'ordre supérieur et SVD

Cette section illustre comment, à partir de la POD continue, il est possible d'étendre la notion de POD discrète en essayant d'en augmenter l'ordre : en effet, la proposition 7 nous montre que la POD discrète précédemment définie équivaut à la POD continue d'une approximation "d'ordre zéro" (constante par morceaux) en temps d'une donnée $u(t)$. De plus, dans le cadre d'un espace X de dimension fini, ce nouveau calcul POD revient encore à un calcul SVD.

Nous proposons cette étude afin de prendre un certain recul par rapport à la POD discrète telle qu'elle a été définie, et afin de souligner une nouvelle fois l'importance de la relation qui existe entre POD et SVD. En effet, celle-ci dépasse le cadre qui a été fixé dans la section 1.4.2.

Supposons que $u \in \mathrm{L}^2(0,T,X)$ soit connu sur une subdivision régulière de $[0,T]$, c'est-à-dire en N instants $t_i = \frac{T\,(i-1)}{N-1}$ pour $1 \leq i \leq N$. Alors il est possible de définir une approximation v de u linéaire par morceaux :

$$v(t) = \frac{1}{\tau}((t_{i+1} - t)\, u_i + (t - t_i)\, u_{i+1}) \text{ pour tout } t \in [t_i, t_{i+1}] \text{ avec } \tau = \frac{T}{N-1} \text{ et } u_i = u(t_i).$$

En considérant la POD continue de v, le calcul nous montre alors que, pour tout $\varphi \in X$,

$$\begin{aligned} K\varphi &= \frac{1}{T}\int_0^T (\, v(t), \varphi \,)_X \; v(t)\, dt \\ &= \frac{1}{6(N-1)} \sum_{i=1}^{N-1} [2\,(\, u_i, \varphi \,)_X \; u_i + 2\,(\, u_{i+1}, \varphi \,)_X \; u_{i+1} + (\, u_{i+1}, \varphi \,)_X \; u_i + (\, u_i, \varphi \,)_X \; u_{i+1}]\,. \end{aligned}$$

Il existe donc une matrice $S = \left(\, S_{i,j} \,\right) \in \mathbb{R}^{N \times N}$, non diagonale, telle que

$$\begin{aligned} \forall \varphi \in X \qquad K\varphi &= \sum_{(i,j) \in [\![1,N]\!]^2} S_{i,j}\, (\, u_j, \varphi \,)_X \; u_i \\ &= \left(\, u_1 \cdots u_N \,\right) S \begin{pmatrix} (\, u_1, \varphi \,)_X \\ \vdots \\ (\, u_N, \varphi \,)_X \end{pmatrix}. \end{aligned}$$

Ainsi, dans le cas où X est un espace de dimension finie P, et en reprenant les notations précédentes, toute base POD discrète (d'ordre maximal d_Y) de matrice $\Phi \in \mathbb{R}^{P \times d_Y}$ est, par définition, solution du problème aux valeurs et vecteurs propres suivant

$$\underbrace{\left[U\, S\, U^T\, S_P\right]}_{\text{correspond à } K} \quad \Phi = \Phi\, \Sigma^2, \tag{1.28}$$

avec Σ^2 la matrice diagonale des valeurs propres ordonnées de K, U la matrice des clichés et S_P la matrice transposant le produit scalaire de X dans $\mathbb{R}^P$, sous la contrainte que $\Phi^T S_P\, \Phi = \mathrm{I}_{d_Y}$. L'équation (1.28) est similaire au problème (1.21) aux valeurs et vecteurs propres associé à une famille de vecteurs singuliers gauches de la matrice des clichés.

En conclusion, la nouvelle base POD discrète "d'ordre un" proposée ici peut être obtenue en calculant une SVD généralisée de la matrice des clichés U, relativement à S_P et $S_N = S$, où S n'est pas une matrice diagonale.

Ce résultat est donné pour illustrer les notions de POD et SVD. Dans toute la suite, l'appellation POD (discrète) fait référence à la définition de la section 1.1.

1.5 Calcul numérique pratique d'une POD

Cette section présente différentes manières de calculer une POD à partir de données numériques en s'appuyant sur les résultats de la section précédente. Tout d'abord, le calcul d'une POD à l'aide d'une boîte noire de calcul SVD sera évoqué. Nous nous intéresserons ensuite à des méthodes de calcul plus efficaces qui reposent sur la résolution d'un problème aux valeurs propres, en nous référant à Fahl [19]. Enfin, le calcul POD pratique d'une base de données issue de mesures expérimentales ou de calculs par différences ou volumes finis sera abordé.

1.5.1 Calcul via une routine SVD

D'après le théorème 4, la proposition 10 permet de calculer une POD d'une base de données numérique en utilisant la routine d'une bibliothèque standard d'algèbre linéaire de calcul d'une SVD (non généralisée et non tronquée nécessairement). Le principe est le suivant :

1. Calcul de $\tilde{U} = \left(S_P^{1/2}\right)^T U$, par exemple via une factorisation de Cholesky (voir page 19).
2. Calcul d'une SVD $(\tilde{\Phi}, \hat{\Sigma}, \hat{A})$ de $\tilde{U}$.
3. $\Phi = \left(S_P^{1/2}\right)^{-T} \tilde{\Phi}$, $\Sigma = \dfrac{1}{\sqrt{N}}\, \hat{\Sigma}$, et $A = \dfrac{1}{\sqrt{N}}\, \hat{A}$ donnent la POD suivante :

$$u(t_i) = \sum_{k=1}^{d} \sigma_k\, a_k(t_i)\, \varphi_k \quad \text{avec} \quad a_k(t_i) = A_{i,k} \quad \text{et} \quad \varphi_k = \sum_{j=1}^{P} \Phi_{j,k}\, \mathcal{X}_j.$$

Les principaux désavantages de cette méthode sont : (1) d'imposer le calcul d'une SVD non tronquée, même si on souhaite déterminer une base POD d'ordre $M \ll \text{Rang(U)}$, (2) de devoir stocker en mémoire la matrice $\tilde{U}$ en entier au moment de l'appel à la routine, ce qui peut être problématique pour une base de données numérique volumineuse. Cette méthode n'est donc optimale ni en termes de mémoire, ni en termes de coût de calcul.

1.5.2 Calcul via un problème aux valeurs propres

Un calcul POD plus efficace peut être réalisé via la résolution "manuelle" d'un problème aux valeurs et vecteurs propres équivalent à celui associé soit à la matrice K (voir l'analyse-synthèse de la page 21), soit à la matrice $\tilde{K}$ des corrélations temporelles. Il existe également une troisième possibilité reposant sur le théorème suivant (qui est une version généralisée du théorème proposé dans [19, p.44]) :

Proposition 13 *Soit $U \in \mathbb{R}^{P\times N}$ de rang $d \geq 1$ et $(\Phi, \Sigma = \text{diag}(\sigma_1, \cdots, \sigma_d), A)$ une SVD généralisée tronquée de U relativement à S_P et S_N.*

1. *Les valeurs propres de $B_1 = (S_N^{1/2})^T U^T S_P U S_N^{1/2}$ sont les ${\sigma_k}^2$. Les vecteurs singuliers droits $A_{:,k}$ sont en bijection avec une famille $(S_N^{1/2})^T A_{:,k}$ de vecteurs propres orthonormaux de B_1 pour les valeurs propres respectives ${\sigma_k}^2$.*
2. *Les valeurs propres de $B_2 = (S_P^{1/2})^T U S_N U^T S_P^{1/2}$ sont les ${\sigma_k}^2$. Les vecteurs singuliers gauches $\Phi_{:,k}$ sont en bijection avec une famille $(S_P^{1/2})^T \Phi_{:,k}$ de vecteurs propres orthonormaux de B_2 pour les valeurs propres respectives ${\sigma_k}^2$.*
3. *Les valeurs propres de la matrice symétrique*

$$B_3 = \begin{pmatrix} 0 & (S_P^{1/2})^T U S_N^{1/2} \\ (S_N^{1/2})^T U^T S_P^{1/2} & 0 \end{pmatrix}$$

sont les σ_k, les $-\sigma_k$ et $N + P - 2d$ zéros. Les vecteurs singuliers $\Phi_{:,k}$ et $A_{:,k}$ sont en bijection avec les familles

$$v_k = \frac{1}{\sqrt{2}}\Big((S_P^{1/2})^T \Phi_{:,k} \ (S_N^{1/2})^T A_{:,k}\Big)^T \text{ et } \tilde{v}_k = \frac{1}{\sqrt{2}}\Big((S_P^{1/2})^T \Phi_{:,k} \ -(S_N^{1/2})^T A_{:,k}\Big)^T$$

de vecteurs propres orthonormaux de B_3 pour les valeurs propres respectives σ_k et $-\sigma_k$.

Preuve. Les deux premiers points découlent de l'analyse-synthèse de la page 21. Nous allons maintenant démontré le point 3. Tout d'abord, notons que

$$B_3\, B_3 = \begin{pmatrix} B_2 & 0 \\ 0 & B_1 \end{pmatrix}. \tag{1.29}$$

La famille des carrés des valeurs propres de B_3 (matrice symétrique réelle) correspond donc aux spectres de B_1 et B_2 : B_3 admet $2\,d$ valeurs propres non nulles qui correspondent à $\pm\sigma_k$ et $N+P-2\,d$ valeurs propres nulles. Partant d'une SVD $U=\Phi\,\Sigma A^T$ généralisée, le calcul montre que $B_3\,v_k=\sigma_k\,v_k$, $B_3\,\tilde{v}_k=-\sigma_k\,\tilde{v}_k$, $\tilde{v}_i^T\,\tilde{v}_j={v_i}^T\,v_j=\delta_{i,j}$ et ${v_i}^T\,\tilde{v}_j=0$: le passage des vecteurs singuliers de U à une famille de vecteurs propres orthonormaux de B_3 est correct, et les σ_k et $-\sigma_k$ sont bien les valeurs propres non nulles de B_3. Réciproquement, soit $v_k=\frac{1}{\sqrt{2}}\Big(\big(S_P^{1/2}\big)^T\Phi_{:,k}\ \big(S_N^{1/2}\big)^T A_{:,k}\Big)^T$ une famille orthonormale de vecteurs propres de B_3 associés respectivement aux valeurs propres σ_k de toute SVD de U. Il est alors facile de montrer que $\tilde{v}_k=\frac{1}{\sqrt{2}}\Big(\big(S_P^{1/2}\big)^T\Phi_{:,k}\ -\big(S_N^{1/2}\big)^T A_{:,k}\Big)^T$ est une famille orthonormale de vecteurs propres associés aux valeurs $-\sigma_k$ et orthogonaux aux v_k (forme "diagonale" par blocs de B_3). De plus, $B_3\,B_3\,v_k=B_3\,(\sigma_k\,v_k)={\sigma_k}^2\,v_k$: $\big(S_P^{1/2}\big)^T\Phi_{:,k}$ et $\big(S_N^{1/2}\big)^T A_{:,k}$ sont deux familles de vecteurs propres de B_2 et B_1 d'après (1.29). Ces familles sont orthonormales puisque ${\Phi_{:,i}}^T\,S_P\,\Phi_{:,j}=\frac{1}{2}(v_i+\tilde{v}_i)^T((v_j+\tilde{v}_j))=\delta_{i,j}$ et ${A_{:,i}}^T\,S_N\,A_{:,j}=\frac{1}{2}(v_i-\tilde{v}_i)^T((v_j-\tilde{v}_j))=\delta_{i,j}$. Les deux premiers points de la proposition permettent alors de conclure. □

Comme Fahl [19] l'a souligné, on déduit de cette proposition trois méthodes de calcul d'une base POD d'ordre M à partir de la matrice U des clichés ou de la matrice modifiée $\tilde{U}=\big(S_P^{1/2}\big)^T\,U$ (rappelons que dans le cadre de la POD discrète, $S_N=\frac{1}{N}\,\mathrm{I}_N$). Elles reposent sur la diagonalisation (partielle) d'une matrice symétrique réelle dans une base orthonormale.

La méthode dite *classique*

1. Calculer $\tilde{B}_2=N\,B_2=\tilde{U}\,\tilde{U}^T$.
2. Déterminer une matrice $\tilde{\Phi}$ orthonormale de M vecteurs $\tilde{\Phi}_{:,k}$ associés aux M premières valeurs propres $N\,{\sigma_k}^2$ de $\tilde{B}_2$.
3. $\Phi=\big(S_P^{1/2}\big)^{-T}\,\tilde{\Phi}$ et $A_{:,k}=\dfrac{1}{\sigma_k}\,U^T\,S_P\,\Phi_{:,k}$ nous donnent la base POD (φ_k) d'ordre M suivante :

$$\varphi_k=\sum_{j=1}^{P}\Phi_{j,k}\,\mathcal{X}_j\quad\text{et}\quad a_k(t_i)=\frac{1}{\sigma_k}\,(\,u(t_i),\varphi_k\,)_X=A_{i,k}. \tag{1.30}$$

La méthode classique consiste à calculer des modes POD en partant de leur définition :

$$\Big[\tilde{\Phi}^T\,\tilde{\Phi}=\mathrm{I}_M\ \text{et}\ \tilde{B}_2\,\tilde{\Phi}=\tilde{\Phi}\,\Sigma\Big]\Longleftrightarrow\Big[\Phi^T\,S_P\,\Phi=\mathrm{I}_M\ \text{et}\ K\,\Phi=\Phi\,\Sigma\Big]$$

avec $\tilde{\Phi}=\big(S_P^{1/2}\big)^T\Phi$ et $K=\frac{1}{N}\,U\,U^T\,S_P$ la matrice correspondant à l'opérateur K de la section 1.1. On préfère résoudre le problème aux valeurs et vecteurs propres associé à B_2 plutôt qu'à K, puisque K n'est pas symétrique *a priori*.

La méthode des clichés

1. Calculer $\tilde{B}_1 = N\,B_1 = U^T\,S_P\,U$ ($\tilde{B}_1 = N\,B_1 = \tilde{U}^T\,\tilde{U}$ mais il n'est pas nécessaire de calculer $\tilde{U}$, et donc d'effectuer par exemple une factorisation de Cholesky de S_P, pour obtenir $\tilde{B}_1$).
2. Déterminer une matrice $\hat{A}$ de M vecteurs $\hat{A}_{:,k}$ orthonormaux associés aux M premières valeurs propres $N\,{\sigma_k}^2$ de $\tilde{B}_1$.
3. $A = \dfrac{1}{\sqrt{N}}\,\hat{A}$ et $\Phi_{:,k} = \dfrac{N}{\sigma_k}\,U\,A_{:,k}$ nous donnent la base POD (φ_k) d'ordre M satisfaisant (1.30).

Cette méthode correspond bien à celle des clichés puisque la matrice $\tilde{K}$ des corrélations temporelles est $B_1 = \frac{1}{N}\,\tilde{B}_1$.

Méthode alternative

1. Calculer
$$\tilde{B}_3 = \sqrt{N}\,B_3 = \begin{pmatrix} 0 & \tilde{U} \\ \tilde{U}^T & 0 \end{pmatrix}.$$
2. Calculer M vecteurs $V_{:,k} = \frac{1}{\sqrt{2}}\begin{pmatrix}\tilde{\Phi}_{:,k} \\ \hat{A}_{:,k}\end{pmatrix}$ orthonormaux associés aux M premières valeurs propres $\sqrt{N}\,\sigma_k$ de B_3.
3. $\Phi = \left(S_P^{1/2}\right)^{-T}\tilde{\Phi}$ et $A = \dfrac{1}{\sqrt{N}}\,\hat{A}$ nous donnent la base POD (φ_k) d'ordre M satisfaisant (1.30).

Choix de la dimension M de la base POD

En général, M est défini comme le plus petit entier k satisfaisant un critère $\mathcal{C}(k)$ fonction des λ_i pour $i \leq k$ et de $\sum_{i=1}^{d_Y}\lambda_i$ (d_Y est le rang de la matrice U des clichés), par exemple
$$\mathcal{C}(k) \equiv \left[\sum_{i=1}^{k}\lambda_i > c_1\sum_{i=1}^{d_Y}\lambda_i \quad \text{et} \quad \lambda_k < c_2\sum_{i=1}^{d_Y}\lambda_i\right]$$
avec $c_1 = 0.95$ et $c_2 = 0.01$ (voir Sirovich [88]). En effet, le spectre POD mesure l'efficacité de toute base POD (voir la section 1.1.2).

Dans ce cas, il n'est pas nécessaire de connaître les λ_i pour $i > k$ pour savoir si $\mathcal{C}(k)$ est vrai, puisque $\sum_{i=1}^{d_Y}\lambda_i$ peut être obtenu par la relation
$$\sum_{i=1}^{d_Y}\lambda_i = \frac{1}{N}\sum_{i=1}^{N}{\|u(t_i)\|_X}^2 = \frac{1}{N}\sum_{i=1}^{N}{\|y_i\|_{S_P}}^2 = \left\|\tilde{U}\right\|_F^2,$$
ou encore, dans le cadre de la méthode classique et de la méthode des clichés, par
$$\sum_{i=1}^{d_Y}\lambda_i = \frac{1}{N}\mathrm{Trace}(\tilde{B}_2) = \frac{1}{N}\mathrm{Trace}(\tilde{B}_1).$$

Choix de la méthode et algorithmes de Lanczos

La précision et le coût informatique du calcul d'une POD reposent ainsi sur le choix d'un des trois problèmes aux valeurs et vecteurs propres qui viennent d'être donnés et sur la technique utilisée pour le résoudre numériquement. Cette section donne quelques indications concernant le choix du problème pour le calcul de la POD.

Tout d'abord, notons que les algorithmes de Lanczos (voir [26]) sont particulièrement intéressants en termes de coût, puisqu'ils permettent de calculer précisément les M plus grandes valeurs propres d'une matrice symétrique réelle de dimension L et une famille de vecteurs propres orthonormaux associés, sans avoir nécessairement à diagonaliser entièrement la matrice. Le coût d'un calcul POD à l'aide de ces algorithmes et des méthodes précédentes est le plus souvent nettement moindre que celui induit par une routine SVD classique (qui calcule une SVD non tronquée de la matrice). En effet, la SVD/POD étant optimale, un petit nombre $M \ll L$ de vecteurs propres suffisent en général à bien représenter la matrice des données. Il faut noter qu'il est possible de définir un critère d'arrêt $\mathcal{C}(k)$ qui permette de décider si les k premiers modes POD calculés donnent une suffisamment bonne approximation des données au cours d'un calcul de type Lanczos, c'est-à-dire sans avoir calculé toutes les valeurs propres (voir le paragraphe précédent). Pour plus de détails sur les méthodes de Lanczos dans le cadre de la POD, le lecteur pourra consulter [19, chapitre 4].

La POD est utilisée pour extraire une information pertinente d'un phénomène complexe, dont une évolution (discrète) par rapport à un grand nombre de degrés de liberté est connue, ce qui se traduit en pratique par un espace X de grande dimension par rapport au nombre de clichés connus : $P \gg N$. C'est en général le cas pour des données qui proviennent de la résolution numérique d'une EDP comme celle de Navier-Stokes. En pratique, la méthode classique est alors plus coûteuse que la méthode des clichés : la matrice $\tilde{B}_1$ de dimension N se calcule et se diagonalise plus rapidement en général que la matrice $\tilde{B}_2$ de dimension $P \gg N$.

En effet, supposons (comme dans [94]) que S_P soit une matrice diagonale et que l'on dispose d'une routine de complexité L^3 de diagonalisation d'une matrice symétrique de dimension L. Alors la complexité des deux premières étapes de la méthode classique est de l'ordre de $P^2 \times \max(N, P)$, contre $N^2 \times \max(N, P)$ pour la méthode des clichés.

Même pour une matrice S_P non diagonale, la méthode des clichés reste moins onéreuse dans la pratique, c'est pourquoi la majorité des auteurs y recourent.

La méthode alternative est, quant à elle, intéressante dans la mesure où elle peut conduire à un calcul SVD moins pollué par les erreurs numériques (voir [19]). Par contre, elle ne présente pas d'intérêt pratique en termes de complexité ; elle est même en général plus coûteuse que la méthode des clichés (voir les résultats numériques de [19, chapitre 4]).

Dans les exemples du mémoire, la méthode des clichés, dont le coût de calcul est satisfaisant, a été utilisée sans que les erreurs numériques aient posé problème : la base POD

(φ_k) obtenue est efficace au sens de l'erreur $\frac{1}{N}\sum_{i=1}^{N}\left\|u(t_i)-\sum_{k=1}^{M}\left(\,u(t_i),\varphi_k\,\right)_X\varphi_k\right\|_X^2$, recalculée *a posteriori*, et cette erreur est bien égale à $\sum_{k=M+1}^{d_Y}\lambda_k$ à l'epsilon machine près. En revanche, la méthode n'a pas été implémentée de manière optimale, puisqu'une routine standard de diagonalisation a été utilisée plutôt qu'un algorithme plus performant de type Lanczos : cela n'a pas été gênant puisqu'un seul calcul POD (voir chapitre 3) a été mené sur une base de données très volumineuse.

Remarques. Notons qu'il est encore possible d'améliorer les méthodes proposées. En effet, il est possible de définir des algorithmes de type Lanczos permettant notamment

- de pouvoir résoudre le problème aux valeurs et vecteurs propres associé à une matrice B^TB sans avoir à calculer ce produit,
- de pouvoir effectuer la deuxième étape de la méthode alternative sans jamais calculer le produit $\tilde{U}=\left(S_P^{1/2}\right)^T U$.

Pour ces résultats et pour une discussion plus approfondie du calcul SVD/POD, nous conseillons tout particulièrement la lecture de [19, chapitre 4].

1.5.3 Cas d'une base de données expérimentale ou obtenue par une discrétisation aux différences finies ou aux volumes finis

Le formalisme de la section 1.4 n'est pas valable au sens strict pour des données qui ne sont pas exprimées dans une base $\mathcal{X}$. C'est le cas des données numériques issues de mesures expérimentales, ou encore de simulations par des schémas aux DF (Différences Finies) ou aux VF (Volumes Finis). Or, en mécanique des fluides, la majorité des simulations ont recours à l'un de ces deux types de discrétisation. En effet, il est aisé d'obtenir des schémas d'ordre élevé par DF, ce qui peut être intéressant pour bien prendre en compte l'acoustique d'un problème. De plus les méthodes VF conduisent à des simulations de qualité dont les coûts de calcul sont raisonnables. Pour ce type de données, cependant, il est généralement possible de calculer une POD

- en procédant à une phase "d'interpolation" (ou de moindres carrés), dont l'objectif est de réexprimer ces données dans une base d'un espace de Hilbert de dimension finie,
- ou directement en proposant une matrice des clichés $U\in\mathbb{R}^{P\times N}$ et un produit scalaire $(\,.\,,.\,)_{S_P}$ de $\mathbb{R}^P$ cohérents.

En effet, les données de type DF, VF ou expérimentales sont *localisées*. Par exemple, dans le cas de données scalaires, elles peuvent être stockées sous la forme d'une matrice des clichés

$$U=\left(\;u(x_i,t_j)\;\right)\in\mathbb{R}^{P\times N}$$

où les P points $x_i\in\mathbb{R}^d$ ($d=1,2$ ou 3) sont associés à un "volume" élémentaire δV_i de mesure de Lebesgue $\mathrm{mes}(\delta V_i)$ dans $\mathbb{R}^d$ non nulle, et tels que $x_i\in\delta V_i$, $\bigcup_{i=1}^{P}\delta V_i=\Omega\subset\mathbb{R}^d$ et

$\text{mes}(\delta V_i \bigcap \delta V_j) = 0$ pour $i \neq j$ (en fait les δV_i sont des longueurs pour $d = 1$, des surfaces pour $d = 2$ et des volumes pour $d = 3$). Ceci est illustré par l'exemple 2D de la figure 1.1.

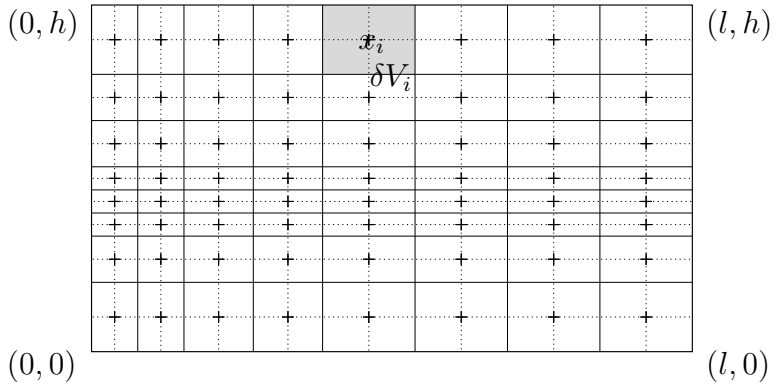


FIG. 1.1 – Exemple de grille cartésienne inhomogène 2D ($\Omega = [0, l] \times [0, h]$) sur laquelle des données DF ou VF peuvent être localisées.

En posant $S_P = \text{diag}(\text{mes}(\delta V_1), \cdots, \text{mes}(\delta V_P))$, la SVD généralisée de U relativement à S_P et $S_N = \frac{1}{N} \text{I}_N$ s'interprète alors comme une POD discrète "de type $\text{L}^2(\Omega)$". Les données numériques exploitées dans la section 2.5 (respectivement 3) provient d'une simulation VF (resp. DF) sur une grille cartésienne inhomogène 2D (resp. 3D) : une POD de type $\text{L}^2(\Omega)$ a pu être calculée en utilisant cette technique.

Il est également possible de définir une POD de type $\text{H}^1(\Omega)$ via un schéma de différentiation spatiale, ou encore de définir une matrice U des clichés et une matrice S_P qui permettent de calculer une POD de type $\text{L}^2(\Omega)^n$ pour des données vectorielles.

Remarque. En pratique, on ne construit jamais la matrice S_P mais une routine capable de calculer le produit scalaire associé $(\,.\,,.\,)_{S_P}$. En effet, la méthode des clichés peut se contenter de cette routine, qui peut d'ailleurs être ensuite réutilisée dans la méthode de Galerkine si elle correspond à un produit scalaire de type $\text{L}^2(\Omega)$ (voir la section 2.1.1). De plus, il serait plus coûteux de calculer $\tilde{B}_1$ en effectuant deux produits matriciels sans tenir compte de la forme particulière de S_P, généralement creuse (matrice-bandes).

1.6 La notion de POD

Même si, dans la suite de ce mémoire, la POD correspondra toujours aux définitions des sections précédentes, il faut noter que la notion de POD peut être étendue à un cadre théorique plus large. Ainsi, la POD pourrait être définie relativement à un espace de Hilbert X complexe (voir [36] pour une définition où X est un espace de fonctions de carré intégrable à valeurs complexes).

De plus, il est possible d'étendre la définition de la POD discrète en redéfinissant l'opérateur $\langle .(t)\rangle = \frac{1}{N}\sum_{i=1}^{N} .(t_i)$ de moyenne temporelle par $\langle .(t)\rangle = \sum_{i=1}^{N} p_i .(t_i)$ où un poids relatif p_i est attribué à chaque cliché. Cette nouvelle définition serait d'ailleurs plus cohérente dans le cas d'une distribution temporelle inhomogène des clichés (voir la proposition 7). De la même manière, on pourrait redéfinir la POD continue relativement à une autre mesure sur $[0, T]$ que la mesure de Lebesgue. En fait, il serait naturel d'étendre la notion de POD pour tout ensemble $\mathcal{T}$ mesurable sur lequel il est possible de définir un opérateur $\langle .\rangle$ de moyenne (voir l'introduction de ce chapitre).

En outre, remarquons que l'équivalence entre SVD et POD discrète, abordée dans la section 1.4, n'a naturellement été établie que pour $S_N = \frac{1}{N}\mathrm{I}_N$. Or la SVD se généralise pour une matrice S_N symétrique définie positive quelconque et conserve alors une propriété d'optimalité (voir [30]). On peut donc se demander s'il est possible de généraliser la notion de POD discrète de façon à englober complètement la notion de SVD. D'autant plus qu'une POD continue "discrétisée" peut amener à effectuer des calculs SVD où S_N n'est pas diagonale (voir la section 1.4.4).

Enfin, notons qu'Henri *et al.*, à la suite des travaux de Kunisch *et al.* [49], proposent d'étendre la définition de la POD à la dérivée temporelle du champ u de référence considéré, dans le cadre de l'analyse de la méthode POD-Galerkine appliquée à un problème parabolique ; voir [33].

Chapitre 2

La modélisation POD-Galerkine pour la mécanique des fluides

Sommaire

Comme nous l'avons vu au chapitre précédent, les modes POD sont optimaux et ordonnés : ils apparaissent donc comme des fonctions de base appropriées pour construire un modèle dynamique réduit représentatif du système physique étudié grâce à la méthode de Galerkine. La modélisation POD-Galerkine qui en résulte permet de définir une description simplifiée d'un système physique, exploitable pour un coût informatique faible. Celle-ci peut servir plusieurs objectifs comme l'analyse physique ou l'optimisation du système considéré, par exemple le contrôle d'un écoulement instationnaire (voir le chapitre 5).

De plus, le calcul d'une base POD est assez simple et son coût est raisonnable. Par ailleurs, les modes POD sont orthogonaux et héritent de certaines propriétés des données utilisées. Ainsi, les modes obtenus à partir de champs de vitesses incompressibles sont de divergence nulle (voir la section 2.3.1), ce qui sera utile lors de la construction des modèles incompressibles.

Ce chapitre est consacré à la modélisation POD-Galerkine. Le principe de cette méthode est rappelé dans le paragraphe 2.1. Elle est ensuite appliquée dans la section 2.3 aux équations de Navier-Stokes préalablement rappelées dans le paragraphe 2.2. La modélisation POD-Galerkine des écoulements incompressibles sera abordée de manière approfondie. Plus précisément, en nous basant sur la littérature existante, nous avons cherché à définir un modèle réduit cohérent qui tienne explicitement compte de toutes les conditions aux limites. De plus, nous montrerons que la modélisation du terme de pression ne pose pas de problème au niveau formel pour les conditions aux limites usuelles, y compris pour des conditions de flux qui font intervenir le tenseur physique des contraintes fluides, ce qui nous semble être un apport par rapport à la littérature existante. En revanche, les conditions aux limites de Dirichlet étant difficiles à traiter en général, le problème de la prise en compte de la pression susbsiste en pratique. Nous présenterons les méthodes que Rempfer [76] et Galletti *et al.* [25] ont proposé pour surmonter cette difficulté. La modélisation POD-Galerkine des écoulements compressibles est évoquée au paragraphe 2.3.2. Les dernières sections traiteront de la stabilité des modèles réduits (section 2.4) et présenteront un exemple de modèle dynamique réduit obtenu à partir de la simulation numérique d'un écoulement incompressible 2D et laminaire (section 2.5) avant de conclure (section 2.6).

2.1 La méthode POD-Galerkine

Le principe général de la modélisation POD-Galerkine réduite est exposé dans cette section.

2.1.1 La modélisation POD-Galerkine réduite

La modélisation POD-Galerkine réduite est schématisée sur la figure 2.1.

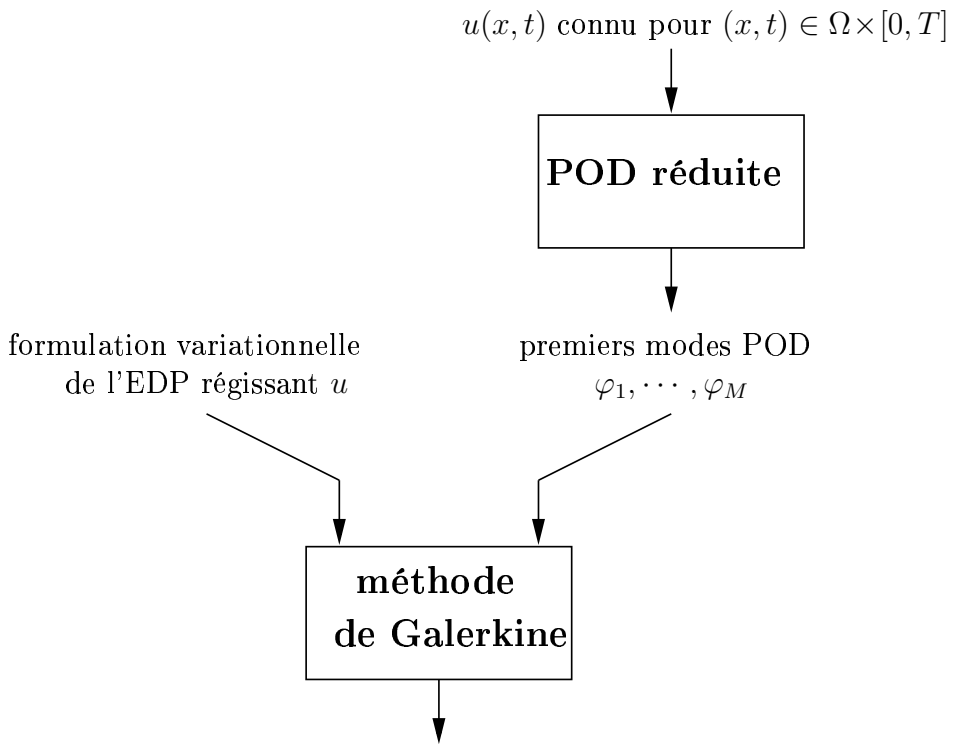


FIG. 2.1 – La modélisation POD-Galerkine réduite.

En pratique, u est connu sous la forme d'une base numérique de N clichés. Dans le cas de la modélisation d'un écoulement incompressible, u correspond au champ des vitesses. Un calcul POD nous permet d'obtenir les M premiers modes d'une base POD. Il y a deux façons de mettre en œuvre cette étape de "POD réduite" : soit on utilise des algorithmes de type Lanczos qui permettent d'obtenir les M premiers modes sans calculer les $N-M$ autres, soit on effectue une SVD complète de la base des données, en général par la méthode des clichés, puis on tronque *a posteriori* la base POD (se reporter à la section 1.5). Qu'ils soient calculés ou non, les modes POD non utilisés, c'est-à-dire ici ceux d'indice strictement supérieur à M, seront qualifiés de *modes négligés* ou *modes tronqués* dans la suite. Les premiers modes

POD représentent de manière optimale la donnée u comme nous l'avons vu au chapitre précédent. Dans le cas d'un écoulement incompressible "organisé", ils permettent de définir ce qu'on nomme les *structures cohérentes*, c'est-à-dire un ensemble fini de structures qui régissent principalement la dynamique de l'écoulement (voir Holmes *et al.* [36]).

La deuxième étape de la modélisation POD-Galerkine réduite consiste à exploiter les M premiers modes de la base POD pour construire un modèle dynamique réduit. Ceci est réalisé en appliquant la méthode de Galerkine à partir d'une formulation variationnelle de l'EDP (Équation aux Dérivées Partielles) qui modélise la dynamique du système physique étudié, par exemple les équations de Navier-Stokes pour un écoulement fluide. Comme nous allons le voir, la méthode de Galerkine permet d'extraire un système d'EDOs de dimension M censé approcher la dynamique du système physique considéré.

2.1.2 Principe de la méthode de Galerkine

Nous allons rappeler le principe bien connu de la méthode de Galerkine. Considérons un système dynamique "de dimension infinie" évoluant dans un espace de Hilbert X, d'état $u(t) \in X$, satisfaisant la condition initiale

$$u(0) = u_0 \in X,$$

et régi par la *formulation variationnelle* suivante

$$\forall (\varphi, t) \in X \times [0,T] \qquad \frac{d}{dt} \left(u(t), \varphi \right)_{H_0} + \sum_{k=1}^{Q} \left(D_k(u(t)), \tilde{D}_k(\varphi) \right)_{H_k} = \left(h(t), \varphi \right)_{H_{Q+1}}, \quad (2.1)$$

avec $X \subset H_0 \cap H_{Q+1}$, et où

- les H_k sont des espaces de Hilbert munis des produits scalaires $(\, . \, , \, . \,)_{H_k}$,
- $h \in H_{Q+1}$ modélise un effet environnemental indépendant de l'état u du système,
- $D_k : X \longrightarrow H_k$ et $\tilde{D}_k : X \longrightarrow H_k$ sont des opérateurs de différentiation spatiale ou de trace.

Le principe de la méthode de Galerkine est alors, partant d'un sous-espace Y de X de dimension finie M, d'approcher l'équation (2.1) par un système de dimension finie évoluant dans Y. Plus précisément, soit $(\varphi_k)_{k=1..M}$ une base de Y orthonormale au sens de $(\, . \, , \, . \,)_{H_0}$ (ce qui est toujours possible puisque $Y \subset X \subset H_0$). Alors, en substituant Y à X dans (2.1), et en notant

$$u(t) = \sum_{j=1}^{M} a_j(t) \, \varphi_j$$

la décomposition de l'état $u(t) \in Y$ du système approché dans cette base Hilbertienne, on obtient, pour tout $i \in [\![1, M]\!]$ et pour la fonction test $\varphi = \varphi_i$,

$$\dot{a}_i(t) = - \sum_{k=1}^{Q} \left(D_k(\sum_{j=1}^{M} a_j(t) \, \varphi_j), \tilde{D}_k(\varphi_i) \right)_{H_k} + \left(h(t), \varphi_i \right)_{H_{Q+1}}.$$

On obtient ainsi un système d'EDOs de dimension M de la forme

$$\dot{a}(t) = f(a(t), t) \quad \text{avec} \quad a(t) = (a_1(t) \cdots a_M(t))^T. \tag{2.2}$$

Il faut donner une condition initiale à ce nouveau système dynamique. Il est cohérent de choisir la projection orthogonale $\sum_{j=1}^{M} (u_0, \varphi_j)_{H_0} \varphi_j$ de u_0 sur Y, ce qui revient à imposer

$$a_i(0) = (u_0, \varphi_i)_{H_0},$$

pour tout $i \in [\![1, M]\!]$.

Remarque. Parfois la méthode de Galerkine est appliquée pour une base de Y qui n'est par orthonormale au sens de H_0. Le système dynamique alors obtenu est de la forme $B\,\dot{a}(t) = f(a(t), t)$ où $B \in \mathbb{R}^{n \times n}$ est la matrice hermitienne $B = \big((\varphi_i, \varphi_j)_{H_O}\big)$; voir par exemple le modèle tourbillonnaire proposé par Rempfer et décrit en page 53. En pratique, on peut orthonormaliser la base par l'algorithme de Gramm-Schmidt ou inverser B pour obtenir un système dynamique équivalent de la forme (2.2).

Cas d'une condition de Dirichlet non homogène

Dans le cas d'un système d'EDPs avec une condition au bord de Dirichlet non homogène (i.e. non nulle), il n'est pas possible d'imposer implicitement cette condition aux limites via la formulation variationnelle. On peut néanmoins l'imposer en prenant

$$u(t) = \bar{u}(t) + \tilde{u}(t)$$

avec $\bar{u}(t)$ choisi de façon à satisfaire les conditions de Dirichlet. En conséquence, $\tilde{u}(t) \in X$ où X est un espace de Hilbert dont les éléments satisfont des conditions de Dirichlet homogènes. Ainsi $u(t)$ évolue dans un espace affine de direction X et $\tilde{u}$ est défini par la formulation variationnelle :

$$\begin{aligned} &\forall (\varphi, t) \in X \times [0, T] \\ &\quad \frac{d}{dt} (\bar{u}(t) + \tilde{u}(t), \varphi)_{H_0} + \sum_{k=1}^{Q} (D_k(\bar{u}(t) + \tilde{u}(t)), \tilde{D}_k(\varphi))_{H_k} = (h(t), \varphi)_{H_{Q+1}}. \end{aligned} \tag{2.3}$$

Considérant un sous-espace Y de X, de dimension $M < +\infty$ et de base (φ_j) orthonormale au sens de $(\,.\,, \,.\,)_{H_0}$, la méthode de Galerkine extrait de (2.3) le système d'EDOs

$$\dot{a}_i(t) = -\frac{d}{dt} (\bar{u}(t), \varphi_i)_{H_0} - \sum_{k=1}^{Q} (D_k(\bar{u}(t) + \sum_{j=1}^{M} a_j(t)\, \varphi_j), \tilde{D}_k(\varphi_i))_{H_k} + (h(t), \varphi_i)_{H_{Q+1}}$$

qui permet de définir l'état du système approché par

$$u(t) = \bar{u}(t) + \sum_{j=1}^{M} a_j(t)\, \varphi_j.$$

Écriture polynômiale du système pour des opérateurs linéaires et quadratiques

En pratique, les opérateurs D_k sont souvent linéaires ou quadratiques. Si D_k est linéaire, on a

$$-\left(D_k(\bar{u}(t) + \sum_{j=1}^{M} a_j(t)\,\varphi_j), \tilde{D}_k(\varphi_i) \right)_{H_k} = \sum_{k=1}^{M} C_i^j\, a_j(t) - \left(D_k(\bar{u}(t)), \tilde{D}_k(\varphi_i) \right)_{H_k}$$

où les $C_i^j = -\left(D_k(\varphi_j), \tilde{D}_k(\varphi_i) \right)_{H_k}$ sont des constantes indépendantes de t. Si D_k est quadratique, il est possible de définir un opérateur $Q_k : X \times X \longrightarrow H_k$ bilinéaire tel que $D_k(u) = Q_k(u,u)$ et on obtient

$$\begin{aligned} -\left(D_k(\bar{u}(t) + \sum_{j=1}^{M} a_j(t)\,\varphi_j), \tilde{D}_k(\varphi_i) \right)_{H_k} &= \sum_{j=1}^{M}\sum_{l=1}^{M} C_i^{j,l}\, a_j(t)\, a_l(t) \\ &- \sum_{j=1}^{M} \left(Q_k(\varphi_j, \bar{u}(t)) + Q_k(\bar{u}(t), \varphi_j), \tilde{D}_k(\varphi_i) \right)_{H_k} a_j(t) \\ &- \left(Q_k(\bar{u}(t), \bar{u}(t)), \tilde{D}_k(\varphi_i) \right)_{H_k} \end{aligned}$$

où les $C_i^{j,l} = -\left(Q_k(\varphi_j, \varphi_l), \tilde{D}_k(\varphi_i) \right)_{H_k}$ sont des constantes indépendantes de t.

Ainsi, en sommant les constantes $C_\times^\times$ indépendantes de t, on aboutit dans de nombreux cas à un système dynamique de dimension M de forme polynômiale de degré inférieur ou égal à deux :

$$\forall i \in [\![1, M]\!] \qquad \dot{a}_i(t) = \sum_{j=1}^{M} C_i^j\, a_j(t) + \sum_{j=1}^{M}\sum_{l=1}^{M} C_i^{j,l}\, a_j(t)\, a_l(t) + C_i^{\bar{u},h}(t), \tag{2.4}$$

où $C_i^{\bar{u},h}$ modélise l'influence du terme source h et des conditions aux limites de type Dirichlet à travers $\bar{u}$. Si $\bar{u}$ et h ne dépendent pas du temps, le système dynamique obtenu est autonome.

Remarque. Ce résultat se généralise facilement : si il existe un opérateur multilinéaire R d'ordre q tel que $D_k(u) = R(u, \cdots, u)$, alors le terme $\left(D_k(u), \tilde{D}_k(\varphi_i) \right)_{H_k}$ se développe en un polynôme de degré q dont les coefficients sont soit constants, soit fonctions de $\bar{u}(t)$. Ainsi, le système d'EDOs obtenu par la méthode de Galerkine a souvent une forme polynômiale aisément manipulable.

La "projection" de Galerkine

Dans certains articles, en particulier dans le cas d'un écoulement compressible qui est particulièrement complexe, la méthode POD-Galerkine est appliquée de manière formelle sans tenir compte des conditions aux limites.

Appliquons cette "projection" de Galerkine à l'équation fonctionnelle

$$\mathcal{D}(u) = 0 \tag{2.5}$$

(par exemple à un système d'EDPs). En admettant que u puisse être décomposé dans une base de fonctions (φ_k)

$$u = \sum_k a_k \, \varphi_k,$$

(ce qui est vrai si u est à tout instant dans un espace de Hilbert séparable X), l'idée est alors de "projeter" l'équation (2.5) sur un ensemble fini de fonctions de base $(\varphi_k)_{k=1..M}$ en considérant que l'espace engendré par cet ensemble est suffisant pour bien représenter u. On obtient alors le système approché de M équations

$$(\mathcal{D}(\sum_{k=1}^{M} a_k \, \varphi_k), \varphi_i \,) = 0 \quad \text{pour} \quad 1 \leq i \leq M$$

où $(\, , \,)$ désigne un produit scalaire (voir [21]). Ces équations peuvent ensuite être manipulées formellement, par exemple en appliquant des intégrations par parties (formule de Green).

Ce mécanisme formel est souvent qualifié de "projection" de Galerkine, et, dans le cadre de la modélisation POD-Galerkine des écoulements, il est souvent appliqué aux EDPs de Navier-Stokes sans tenir compte explicitement des conditions aux limites (comme on le verra dans la section 2.3.2 consacrée aux écoulements compressibles).

2.2 Les équations de Navier-Stokes

Les équations de Navier-Stokes modélisent l'évolution (thermo)dynamique d'un *fluide newtonien* dans une *représentation eulerienne*. Afin de définir leur domaine de validité, nous donnerons succinctement les hypothèses qui conduisent à ces équations avant de les écrire (pour les variables primitives), en nous appuyant principalement sur [69, chapitre 1]. On pourra consulter [52] et [13] pour plus de détails.

Soit $\Omega(t) \in \mathbb{R}^d$ ($d = 2$ ou 3) le domaine physiquement occupé par le fluide à l'instant t. La représentation eulerienne consiste simplement à exprimer toute grandeur thermodynamique v en fonction du doublet $(\boldsymbol{x}, t)$ où $\boldsymbol{x}$ correspond à un point du milieu continu fluide $\Omega(t) : v(\boldsymbol{x}, t)$. On considère un repère orthonormal direct de $\mathbb{R}^d$ dont les vecteurs unitaires sont notés $\boldsymbol{x_i}$ pour tout $i \in [\![1, d]\!]$; les vecteurs $\boldsymbol{v}$ de $\mathbb{R}^d$ seront indiqués en gras et leur composante dans la direction $\boldsymbol{x_i}$ sera notée $v_{x_i} : \boldsymbol{v}.\boldsymbol{x_i} = v_{x_i}$. La dérivation dans la direction $\boldsymbol{x_i}$ sera notée ∂_{x_i} et l'opérateur $\boldsymbol{\nabla} = (\partial_{x_1} \cdots \partial_{x_d})^T$ sera employé pour simplifier l'écriture des équations.

Dans la représentation eulerienne, la dérivée temporelle d'une grandeur thermodynamique v de la particule, suivie dans son mouvement, qui se trouve à t en $\boldsymbol{x}$ est notée

$D_t v(\boldsymbol{x},t)$. Cette *dérivée particulaire*, encore appelée *dérivée totale*, est $D_t v = \partial_t v + (\boldsymbol{u}.\boldsymbol{\nabla})v$, si $\boldsymbol{u}$ est le champ eulerien des vitesses.

Soient $\boldsymbol{u}(\boldsymbol{x},t) \in \mathbb{R}^d$ le champ des vitesses du fluide, $\rho(\boldsymbol{x},t) \in \mathbb{R}$ sa densité (masse volumique), et $p(\boldsymbol{x},t) \in \mathbb{R}$ son champ de pression. Pour introduire la notion de fluide newtonien, considérons l'équation de Newton (qui exprime la conservation de la quantité de mouvement) relative à un élément de volume O quelconque du fluide :

$$\int_O \rho\, D_t \boldsymbol{u}\, d\boldsymbol{x} = \int_0 \boldsymbol{h}\, d\boldsymbol{x} - \int_{\partial O} (p\,\boldsymbol{n} - \sigma_v\,\boldsymbol{n})\; ds \tag{2.6}$$

avec $\boldsymbol{n}$ le vecteur normal unitaire sortant du bord ∂O de O, et où le second membre représente les forces exercées sur O : $\int_0 \boldsymbol{h}\, d\boldsymbol{x}$ les forces extérieures volumiques (magnétisme, force de Coriolis, gravité...), $-\int_{\partial O} p\,\boldsymbol{n}\, ds$ les forces de pression, et $\int_{\partial O} \sigma_v\,\boldsymbol{n}\, ds$ les contraintes de viscosité dues à la déformation du fluide (σ_v est un tenseur d'ordre deux qui peut être interprété comme une matrice de $\mathbb{R}^{d\times d}$). La relation (2.6) étant vraie pour tout $O \subset \Omega(t)$ régulier, et compte tenu de la formule de Stokes, on a ainsi

$$\rho\,(\partial_t \boldsymbol{u} + (\boldsymbol{u}.\boldsymbol{\nabla})\boldsymbol{u}) + \boldsymbol{\nabla} p - \boldsymbol{\nabla}.\sigma_v = \boldsymbol{h}. \tag{2.7}$$

C'est alors qu'intervient l'hypothèse des fluides newtoniens. Elle permet d'exprimer le tenseur des contraintes visqueuses σ_v en fonction de $\boldsymbol{u}$ en proposant une loi linéaire entre σ_v et $\nabla u \in \mathbb{R}^{d\times d}$ défini par $(\nabla u)_{i,j} = \partial_{x_j} u_{x_i}$:

$$\sigma_v = 2\,\mu\, S(\boldsymbol{u}) + \lambda\,(\boldsymbol{\nabla}.\boldsymbol{u})\,\mathrm{I}_d \text{ avec } S(\boldsymbol{u}) = \frac{1}{2}\left(\nabla u + \nabla u^T\right)$$

et μ et λ les première et deuxième viscosités du fluide. On réécrit σ_v en introduisant la *viscosité de dilatation* $\xi = \lambda + \frac{2}{3}\mu$ et en conservant la *viscosité dynamique* μ :

$$\sigma_v = 2\,\mu\, S(\boldsymbol{u}) + (\xi - \frac{2}{3}\mu)\,(\boldsymbol{\nabla}.\boldsymbol{u})\,\mathrm{I}_d \tag{2.8}$$

De (2.7) et (2.8), on déduit l'*équation de conservation de la quantité de mouvement* suivante

$$\rho\;(\partial_t \boldsymbol{u} + (\boldsymbol{u}.\boldsymbol{\nabla})\boldsymbol{u}) + \boldsymbol{\nabla} p - \mu\Delta\boldsymbol{u} - (\xi + \frac{\mu}{3})\,\boldsymbol{\nabla}(\boldsymbol{\nabla}.\boldsymbol{u}) = \boldsymbol{0}. \tag{2.9}$$

Cette dernière équation a une portée importante, mais plusieurs types de fluides échappent à cette modélisation : les fluides pâteux, fibrés (rencontrés en biomécanique), les mélanges fluides-particules (le sang notamment) ou les gaz raréfiés.

Un bilan de masse pour un élément O quelconque du fluide conduit à l'*équation de conservation de la masse* suivante, encore appelée *équation de continuité* :

$$\partial_t \rho + \boldsymbol{\nabla}.(\rho\,\boldsymbol{u}) = 0 \quad \text{ou encore} \quad D_t\rho + \rho\,\boldsymbol{\nabla}.\boldsymbol{u} = 0. \tag{2.10}$$

Le système (2.9)-(2.10), purement dynamique, n'est *a priori* pas satisfaisant puisqu'il est formé de quatre équations pour cinq variables scalaires.

2.2.1 Cas d'un écoulement compressible

Dans le cas d'un écoulement compressible, des considérations thermodynamiques sont nécessaires pour "fermer" le système (2.9)-(2.10). Soit $\theta(\boldsymbol{x},t) \in \mathbb{R}$ la température et $e(\boldsymbol{x},t) \in \mathbb{R}$ l'énergie interne par unité de masse. L'écoulement est supposé satisfaire les hypothèses suivantes :

- le fluide n'est pas réactif, c'est-à-dire que sa composition chimique est figée et qu'il n'y a ni gain ni perte de chaleur suite à un processus thermochimique ;
- il n'y a pas de source de chaleur (par exemple de combustion), et il existe une constante κ, appelée *conductivité thermique*, telle que le taux de chaleur reçu à travers toute surface élémentaire de normale sortante $\boldsymbol{n}$ soit $\kappa \boldsymbol{\nabla}\theta.\boldsymbol{n}$ conformément à la *loi de Fourier* ;
- le fluide est *parfait* : il satisfait la *loi d'état*

$$p = r\,\rho\,\theta, \tag{2.11}$$

où $r = \dfrac{R}{m}$ avec R la constante universelle des gaz parfaits, m la masse molaire du fluide, et il existe une constante C_v telle que $e = C_v\,\theta$.

Alors, le bilan d'énergie d'un volume élémentaire quelconque qui se déplace avec le fluide conduit à l'équation

$$\rho\,D_t\theta + (\gamma-1)\,\rho\,\theta\,\boldsymbol{\nabla}.\boldsymbol{u} - \frac{\kappa}{C_v}\Delta\theta = \frac{1}{C_v}\left[\mu\,Q(\boldsymbol{u},\boldsymbol{u}) + \boldsymbol{h}.\boldsymbol{u}\right] \tag{2.12}$$

avec

$$\gamma = \frac{r}{C_v} + 1 \quad \text{et} \quad Q(\boldsymbol{u},\boldsymbol{u}) = (\frac{\xi}{\mu} + \frac{1}{3})(\boldsymbol{\nabla}.\boldsymbol{u})^2 + \frac{1}{2}\mathrm{Trace}((\nabla u + \nabla u^T)^2).$$

Le système (2.9)-(2.10)-(2.11)-(2.12) est appelé *équations de Navier-Stokes compressibles*. Il est possible de le réécrire de nombreuses façons, en considérant d'autres jeux de variables thermodynamiques ou en le mettant sous la forme de cinq équations scalaires. En particulier, les équations d'état (2.11) et de continuité (2.10) nous donnent la relation $D_t p = r\rho D_t\theta - p\,\boldsymbol{\nabla}.\boldsymbol{u}$, ce qui nous permet de réécrire (2.12) en ne faisant intervenir que les variables primitives $(\rho,\boldsymbol{u},p)$:

$$D_t p + \gamma\,p\,\boldsymbol{\nabla}.\boldsymbol{u} - \kappa\,(\gamma-1)\,\Delta(\frac{p}{\rho}) = (\gamma-1)\,\left[\mu\,Q(\boldsymbol{u},\boldsymbol{u}) + \boldsymbol{h}.\boldsymbol{u}\right]. \tag{2.13}$$

Alors (2.9)-(2.10)-(2.13) forment un nouveau système de Navier-Stokes compressible équivalent sous la condition que la loi d'état (2.11) soit vraie.

Enfin, avec $C_p = \gamma\,C_v = r + C_v$, le système des équations de Navier-Stokes est adimensionné à l'aide d'une vitesse U_∞, d'une longueur L et d'une masse volumique ρ_∞ constantes et représentatives de l'écoulement étudié, et en utilisant le temps L/U_∞, la température U_∞^2/C_p, la pression $\rho_\infty U_\infty^2$ et la force volumique $\rho_\infty U_\infty^2/L$ caractéristiques associés.

Équations de Navier-Stokes compressibles adimensionnées :

$$D_t\rho + \rho\,\boldsymbol{\nabla}.\boldsymbol{u} = 0 \tag{2.14}$$

$$\rho\,D_t\boldsymbol{u} + \boldsymbol{\nabla}p - \frac{1}{\mathrm{Re}}\left[\Delta\boldsymbol{u} + (\frac{\xi}{\mu}+\frac{1}{3})\boldsymbol{\nabla}(\boldsymbol{\nabla}.\boldsymbol{u})\right] = \boldsymbol{h} \tag{2.15}$$

$$\rho\,D_t\theta + (\gamma-1)\,\rho\,\theta\,\boldsymbol{\nabla}.\boldsymbol{u} - \frac{\gamma}{\mathrm{Re}\,\mathrm{Pr}}\Delta\theta = \gamma\,\left[\frac{1}{\mathrm{Re}}\,Q(\boldsymbol{u},\boldsymbol{u}) + \boldsymbol{h}.\boldsymbol{u}\right] \tag{2.16}$$

$$\Updownarrow\; p = \frac{\gamma-1}{\gamma}\,\rho\,\theta$$

$$D_t p + \gamma\,p\,\boldsymbol{\nabla}.\boldsymbol{u} + \frac{1-\gamma}{\mathrm{Re}\,\mathrm{Pr}}\,\Delta(\frac{p}{\rho}) = (\gamma-1)\,\left[\frac{1}{\mathrm{Re}}\,Q(\boldsymbol{u},\boldsymbol{u}) + \boldsymbol{h}.\boldsymbol{u}\right] \tag{2.17}$$

avec

$$Q(\boldsymbol{u},\boldsymbol{u}) = (\frac{\xi}{\mu}+\frac{1}{3})(\boldsymbol{\nabla}.\boldsymbol{u})^2 + \frac{1}{2}\mathrm{Trace}(S(\boldsymbol{u})^2)$$

et les nombres sans dimension

$$\mathrm{Re} = \frac{U_\infty\,L\,\rho_\infty}{\mu}\ \text{et}\ \mathrm{Pr} = \frac{C_p\,\mu}{\kappa} \tag{2.18}$$

appelées respectivement *nombre de Reynolds* et *nombre de Prandtl.*

Dans la pratique, on suppose de plus que le fluide satisfait l'*hypothèse de Stokes*, c'est-à-dire que la viscosité de dilatation ξ est nulle. En effet, les mesures expérimentales montrent que ξ peut être négligé pour de nombreux fluides, l'eau et l'air en particulier.

Remarque. La théorie cinétique des gaz indique, dans le cas du modèle simple de sphère rigide, que le coefficient de viscosité dynamique μ est indépendant de la pression, mais qu'il dépend de la température : $\mu = C_s\,\sqrt{\theta}$ avec C_s constant. Cette approximation est en bon accord avec les données expérimentales mais, afin d'avoir une relation qui corresponde très précisément à ces données, on admet généralement que μ est donné par la *loi de Sutherland* : $\mu = (\beta_s\sqrt{\theta})/(1+\frac{\alpha_s}{\theta})$ avec β_s et α_s constants. En outre, la théorie cinétique des gaz et l'expérience amènent à exprimer le coefficient de conductivité thermique κ proportionnellement à μ, mais toujours pour un nombre de Prandtl constant (voir l'équation (2.18)). Ainsi, il est possible de considérer un modèle de Navier-Stokes qui tient compte de la variation de la viscosité dynamique μ à travers le nombre de Reynolds (qui devient local et non plus global et constant).

2.2.2 Cas d'un écoulement incompressible

L'écoulement est dit *incompressible* lorsque sa masse volumique varie insensiblement : c'est en général le cas pour l'eau, ou pour l'air à basse vitesse, dans des conditions naturelles de température et de pression. Dans ce cas, les dérivées de ρ sont négligées et il est facile d'extraire des équations de Navier-Stokes compressibles un système de quatre équations scalaires qui régit la dynamique de l'écoulement indépendamment de toute considération thermique ($\rho = 1$ après adimensionnement).

Équations de Navier-Stokes incompressibles adimensionnées :

$$\boldsymbol{\nabla}.\boldsymbol{u} = 0 \tag{2.19}$$

$$D_t\boldsymbol{u} + \boldsymbol{\nabla}p - \frac{1}{\mathrm{Re}}\Delta\boldsymbol{u} = \boldsymbol{h} \tag{2.20}$$

L'hypothèse d'incompressibilité aboutit donc à un système aux EDPs parabolique beaucoup plus simple que les équations de Navier-Stokes compressibles. De plus, la pression p ne représente alors plus la pression physique, même si son gradient continue d'avoir une interprétation physique (voir [44]).

À ce système fermé, on ajoute parfois l'équation scalaire passive provenant de (2.16) qui régit la température :

$$D_t\theta - \frac{\gamma}{\mathrm{Re}\,\mathrm{Pr}}\Delta\theta = \gamma\left[\frac{1}{2\mathrm{Re}}\,\mathrm{Trace}(S(\boldsymbol{u})^2) + \boldsymbol{h}.\boldsymbol{u}\right]. \tag{2.21}$$

Par contre, il ne serait pas cohérent d'ajouter une équation régissant la pression puisque seul ses variations $\boldsymbol{\nabla}p$ conservent un sens physique.

Mentionnons aussi l'existence d'une autre formulation de l'équation (2.20) faisant intervenir le vecteur tourbillon $\boldsymbol{w}$ (rotationnel de $\boldsymbol{u}$).

Formulation vitesse-tourbillon de la conservation de quantité de mouvement (cas incompressible) :

$$D_t\boldsymbol{w} - (\boldsymbol{w}.\boldsymbol{\nabla})\boldsymbol{u} - \frac{1}{\mathrm{Re}}\Delta\boldsymbol{w} = \boldsymbol{\nabla}\times\boldsymbol{h} \tag{2.22}$$

avec

$$\boldsymbol{w} = \boldsymbol{\nabla}\times\boldsymbol{u}.$$

2.3 Modélisation POD-Galerkine réduite des équations de Navier-Stokes

La méthode POD-Galerkine va maintenant être appliquée de manière formelle aux équations de Navier-Stokes adimensionnées. Dans un premier temps, nous allons nous placer dans le cas incompressible (paragraphe 2.3.1). La modélisation POD-Galerkine d'un écoulement compressible, rarement présente dans la littérature, sera évoquée dans le paragraphe 2.3.2.

2.3.1 Modèles incompressibles

Partant d'un champ des vitesses $\boldsymbol{u}^e$ d'un écoulement incompressible connu, la méthode POD-Galerkine permet de construire un modèle dynamique réduit sous forme polynômiale et physiquement cohérent.

Cette modélisation POD-Galerkine des équations de Navier-Stokes incompressibles est couramment utilisée. Toutefois, elle est généralement appliquée sans traiter de manière explicite les conditions aux limites de flux, ce qui pose le problème du traitement du gradient de pression (voir la section sur la prise en compte implicite des conditions aux limites et [76]), ou alors pour une condition instationnaire particulière, dite *pseudo-stress-free condition*, qui fait intervenir un pseudo-tenseur des contraintes $\tilde{\sigma}$ différent du tenseur σ des contraintes fluides (voir par exemple [75]).

Notons que pour le contrôle actif par soufflage/aspiration, il est nécessaire de prendre en compte explicitement les conditions aux limites de Dirichlet instationnaires. Pour ce faire, la solution naturelle est de construire des modes POD qui satisfont des conditions de Dirichlet homogènes : voir en particulier la méthode des fonctions de commande utilisée par Graham *et al.* [27, 28], Vigo [94, chapitre 5] ou Fahl [19].

Nous nous proposons ici de synthétiser et d'étendre les travaux précédents : nous allons détailler comment il est possible du point de vue formel de construire un modèle POD-Galerkine incompressible qui prenne en compte explicitement les conditions instationnaires aux limites naturelles du problème de Navier-Stokes incompressible, que ce soit des conditions de Dirichlet sur le champ des vitesses, ou des conditions de flux définies par une combinaison linéaire des tenseurs σ et $\tilde{\sigma}$.

Ainsi, nous verrons que le "problème du terme de pression" n'est pas formel mais purement pratique, et qu'il ne survient que lorsque des conditions de Dirichlet instationnaires complexes ne peuvent pas être explicitement prises en compte lors de la construction du modèle POD-Galerkine. Les méthodes proposées par Rempfer [76] et Galletti [25] pour pallier à ce problème de modélisation lié à la pression sont présentées.

Enfin, la prise en compte facultative d'une équation régissant la température est évoquée.

Conditions aux limites et formulation variationnelle

Le modèle dynamique est construit à partir des équations de Navier-Stokes incompressibles (2.19)-(2.20), considérées pour $(\boldsymbol{x}, t) \in \Omega \times [0, T]$, où $\Omega \subset \mathbb{R}^d$ ($d = 2$ ou 3) est un domaine ouvert et borné dont la frontière $\Gamma = \partial\Omega$ est suffisamment régulière, et où $[0, T]$ est l'intervalle fini de temps considéré ($0 < T < +\infty$). De plus, l'écoulement est supposé satisfaire la condition initiale

$$\forall \boldsymbol{x} \in \Omega \qquad \boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{u_0}(\boldsymbol{x}), \tag{2.23}$$

et, pour tout temps $t \in [0, T]$, la condition aux limites de Dirichlet

$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{g}(\boldsymbol{x}, t) \quad \text{sur} \quad \Gamma_D \tag{2.24}$$

et la condition aux limites de flux

$$[(1 - \omega)\, \tilde{\sigma}(\boldsymbol{x}, t) + \omega\, \sigma(\boldsymbol{x}, t)]\, \boldsymbol{n}(\boldsymbol{x}) = P\, \boldsymbol{n}(\boldsymbol{x}) - \boldsymbol{\beta}(\boldsymbol{x}, t) \quad \text{sur} \quad \Gamma_F, \tag{2.25}$$

avec les tenseurs adimensionnés

$$\sigma = \sigma_v - p\, \mathrm{I}_d = \frac{1}{\mathrm{Re}} \left[\nabla u + \nabla u^T\right] - p\, \mathrm{I}_d \quad \text{et} \quad \tilde{\sigma} = \sigma - \frac{1}{\mathrm{Re}} \nabla u^T = \frac{1}{\mathrm{Re}} \nabla u - p\, \mathrm{I}_d,$$

$\omega \in \mathbb{R}$ un paramètre sans dimension, $P \in \mathbb{R}$ une pression constante quelconque, $\boldsymbol{n}$ la normale unitaire sortante de Ω, $\Gamma = \Gamma_D \cup \Gamma_M$, $\mathrm{mes}(\Gamma_D) > 0$ et $\mathrm{mes}(\Gamma_D \cap \Gamma_M) = 0$ (pour la mesure de Lebesgue de $\mathbb{R}^{d-1}$). Ce sont des conditions aux limites naturelles pour le problème parabolique de Navier-Stokes incompressible.

L'équation (2.25) a une portée très générale ; en effet, elle est nécessairement vraie pour les conditions aux limites couramment utilisées pour simuler un écoulement incompressible ouvert, c'est-à-dire qui correspond à une configuration physique où le fluide peut entrer et sortir du domaine Ω considéré. En particulier, la condition

$$[\nabla u]\, \boldsymbol{n} = \boldsymbol{0} \quad \text{et} \quad p\, \boldsymbol{n} = \boldsymbol{\beta}$$

est parfois utilisée pour simuler un écoulement externe et la condition

$$\tilde{\sigma}\, \boldsymbol{n} = -\boldsymbol{\beta} \quad \Longleftrightarrow \quad p\, \boldsymbol{n} - \frac{1}{\mathrm{Re}} [\nabla u]\, \boldsymbol{n} = \boldsymbol{\beta} \tag{2.26}$$

est souvent utilisée pour $\boldsymbol{\beta} = \boldsymbol{0}$ comme condition de sortie pour les écoulements confinés dans un canal ou une conduite (*pseudo-stress-free condition*, voir [85] ou [75]). De plus, la condition

$$\sigma\, \boldsymbol{n} = -\boldsymbol{\beta}$$

est physiquement plus pertinente puisque σ est le tenseur des contraintes fluides pour un écoulement incompressible. Cette dernière condition est par exemple utilisée dans une étude de la modélisation POD-Galerkine réalisée par Noack *et al.*, [66], mais elle n'est pas prise

en compte explicitement dans leur modèle POD-Galerkine : la modélisation proposée nous semble à ce titre originale. Le traitement implicite des conditions aux limites est discuté un peu plus loin.

La "pression" p n'a pas de sens physique contrairement à $\boldsymbol{\nabla} p$, et il est possible de remplacer p par $p + P$ de manière équivalente dans l'équation (2.20) de conservation de la quantité de mouvement, pour une pression constante P quelconque. Il peut donc paraître anormal que la condition (2.25) fasse intervenir la pression (terme $p\,\boldsymbol{n}$). Afin de montrer que (2.25) est bien une condition naturelle qui ne pose aucun problème quant à l'unicité de p qui ne devrait être garantie qu'à une constante près, nous avons introduit le terme $P\,\boldsymbol{n}$ dans (2.25) : comme nous allons le voir, toute pression P constante conduit à la même formulation variationnelle, et par conséquent au même modèle POD-Galerkine.

Pour une fonction test $\boldsymbol{\varphi}$ de divergence et de trace sur Γ_D nulles, et en notant $\phi_{|\Gamma_\times} \in \mathrm{H}^{1/2}(\Gamma_\times)$ la trace sur le bord $\Gamma_\times$ d'une fonction ϕ quelconque de $\mathrm{H}^1(\Omega)$, les équations de Navier-Stokes incompressibles (2.19)-(2.20) associées aux conditions aux limites (2.24)-(2.25) conduit à la formulation suivante :

Définition 4 (FV des éq. de Navier-Stokes incompressible (FVNSI))

$$\begin{aligned} \frac{d}{dt}\left(\, \boldsymbol{u}(t), \boldsymbol{\varphi} \,\right)_{L^2(\Omega)^d} \;&+\; \mathcal{C}(\boldsymbol{u},\boldsymbol{u},\boldsymbol{\varphi}) + \frac{1}{Re}\left[\mathcal{A}(\boldsymbol{u},\boldsymbol{\varphi}) + \omega\,\mathcal{B}(\boldsymbol{u},\boldsymbol{\varphi})\right] \\ &+\; \left(\, \boldsymbol{\beta}, \boldsymbol{\varphi}_{|\Gamma_F} \,\right)_{L^2(\Gamma_F)^d} = \left(\, \boldsymbol{h}, \boldsymbol{\varphi} \,\right)_{L^2(\Omega)^d} \end{aligned} \tag{2.27}$$

pour toute fonction test $\boldsymbol{\varphi}$ *de*

$$V = \left\{\, \boldsymbol{\varphi} \in H^1(\Omega)^d \;/\; \boldsymbol{\nabla}.\boldsymbol{\varphi} = 0 \;\textit{et}\; \boldsymbol{\varphi}_{|\Gamma_D} = \mathbf{0} \,\right\}$$

avec les formes multilinéaires

$$\mathcal{C}(\boldsymbol{u},\boldsymbol{\psi},\boldsymbol{\varphi}) = \left(\, (\boldsymbol{u}.\boldsymbol{\nabla})\boldsymbol{\psi}, \boldsymbol{\varphi} \,\right)_{L^2(\Omega)^d},$$

$$\mathcal{A}(\boldsymbol{u},\boldsymbol{\varphi}) = \sum_{i,j=1}^{d} \left(\, \partial_{x_j} u_{x_i}, \partial_{x_j}\varphi_{x_i} \,\right)_{L^2(\Omega)} = \sum_{i=1}^{d} \left(\, \boldsymbol{\nabla} u_{x_i}, \boldsymbol{\nabla}\varphi_{x_i} \,\right)_{L^2(\Omega)^d}$$

et

$$\mathcal{B}(\boldsymbol{u},\boldsymbol{\varphi}) = \sum_{i,j=1}^{d} \left(\, \partial_{x_i} u_{x_j}, \partial_{x_j}\varphi_{x_i} \,\right)_{L^2(\Omega)}.$$

Preuve. Pour un écoulement incompressible, le tenseur adimensionné des contraintes fluides est

$$\sigma = \sigma_v - p\,\mathrm{I}_d = \frac{1}{\mathrm{Re}}\left(\nabla u + \nabla u^T\right) - p\,\mathrm{I}_d.$$

Puisque $\boldsymbol{\nabla}.(\nabla u^T) = \boldsymbol{\nabla}(\boldsymbol{\nabla}.\boldsymbol{u})$ pour $\boldsymbol{u}$ suffisamment régulier, on a donc pour un écoulement

incompressible ($\boldsymbol{\nabla}.\boldsymbol{u} = 0$)

$$\boldsymbol{\nabla}.\sigma = \boldsymbol{\nabla}.\tilde{\sigma} = \frac{1}{\mathrm{Re}}\Delta\boldsymbol{u} - \boldsymbol{\nabla}p,$$

d'où l'expression (2.20) de l'équation (2.7) de conservation de la quantité de mouvement, et d'où la réécriture suivante de (2.20) :

$$\partial_t\boldsymbol{u} + (\boldsymbol{u}.\boldsymbol{\nabla})\boldsymbol{u} - \boldsymbol{\nabla}.\left[(1-\omega)\,\tilde{\sigma} + \omega\,\sigma\right] = \boldsymbol{h}. \tag{2.28}$$

En outre, pour un tenseur $T \in \mathbb{R}^{d\times d}$ et une fonction test $\boldsymbol{\varphi}$ suffisamment réguliers, la formule de Green nous donne

$$(\ \boldsymbol{\nabla}.T, \boldsymbol{\varphi}\ )_{\mathrm{L}^2(\Omega)^d} = -\sum_{i,j=1}^{d} (\ T_{i,j}, \partial_{x_j}\varphi_{x_i}\ )_{\mathrm{L}^2(\Omega)} + (\ T\,\boldsymbol{n}, \boldsymbol{\varphi}\ )_{\mathrm{L}^2(\Gamma)^d}\,. \tag{2.29}$$

Pour $\boldsymbol{\varphi}_{|\Gamma_D} = \mathbf{0}$ et $T = (1-\omega)\,\tilde{\sigma} + \omega\,\sigma$, c'est-à-dire

$$T_{i,j} = \frac{1}{\mathrm{Re}}\left[\partial_{x_j}u_{x_i} + \omega\,\partial_{x_i}u_{x_j}\right] + p\,\delta_{i,j},$$

la condition aux limites (2.25) et l'équation (2.29) donnent

$$\begin{aligned}(\ \boldsymbol{\nabla}.\left[(1-\omega)\,\tilde{\sigma} + \omega\,\sigma\right], \boldsymbol{\varphi}\ )_{\mathrm{L}^2(\Omega)^d} &= -\frac{1}{\mathrm{Re}}\left[\mathcal{A}(\boldsymbol{u},\boldsymbol{\varphi}) + \omega\,\mathcal{B}(\boldsymbol{u},\boldsymbol{\varphi})\right] - (\ \boldsymbol{\beta}, \boldsymbol{\varphi}_{|\Gamma_F}\ )_{\mathrm{L}^2(\Gamma_F)^d} \\ &+ (\ p, \boldsymbol{\nabla}.\boldsymbol{\varphi}\ )_{\mathrm{L}^2(\Omega)} + (\ P\,\boldsymbol{n}, \boldsymbol{\varphi}_{|\Gamma}\ )_{\mathrm{L}^2(\Gamma)^d}\,.\end{aligned} \tag{2.30}$$

Enfin, d'après la formule de Stokes, on a

$$(\ p, \boldsymbol{\nabla}.\boldsymbol{\varphi}\ )_{\mathrm{L}^2(\Omega)} + (\ P\,\boldsymbol{n}, \boldsymbol{\varphi}_{|\Gamma}\ )_{\mathrm{L}^2(\Gamma)^d} = (\ p+P, \boldsymbol{\nabla}.\boldsymbol{\varphi}\ )_{\mathrm{L}^2(\Omega)}\,.$$

Ainsi, comme $\boldsymbol{\nabla}.\boldsymbol{\varphi} = 0$, on obtient bien la formulation (2.27) en appliquant une fonction test $\boldsymbol{\varphi} \in V$ à (2.28).
□

La condition (2.26) est équivalente à (2.25) pour $\omega = 0$: le terme non-coercif $\mathcal{B}(\boldsymbol{u},\boldsymbol{\varphi})$ disparaît alors. Cela explique la popularité de cette condition aux limites qui fait intervenir le pseudo-tenseur $\tilde{\sigma}$ plutôt que le tenseur physique σ. En effet, la formulation variationnelle est alors triviale à obtenir. Notons que dans ce cas le modèle POD-Galerkine est moins coûteux à calculer et à simuler, et par ailleurs cette condition donne des résultats physiquement satisfaisants. Dans la suite, les écoulements incompressibles considérés satisferont (2.26) (ou une condition très proche qui aboutit à la même formulation variationnelle, voir la section 2.5).

Nous n'allons pas chercher ici à démontrer la pertinence mathématique de cette formulation, en particulier à rechercher pour quels espaces fonctionnels il y aurait existence de $\boldsymbol{u}$ (ce problème mathématique est d'une grande difficulté). Cependant, il faut noter

qu'il existe plusieurs résultats d'existence et d'unicité pour (FVNSI) dus à Temam [91] et Lions [56], dans le cas particulier de conditions de Dirichlet sur tout le bord, c'est-à-dire pour une formulation variationnelle où le terme $\mathcal{B}(\boldsymbol{u}, \boldsymbol{\varphi})$ n'apparaît pas.

Le premier résultat est obtenu dans le cadre bidimensionnel ($d = 2$) et pour des conditions de Dirichlet homogènes ($\boldsymbol{g} = \boldsymbol{0}$) étendues à tout le bord ($\text{mes}(\Gamma_F) = 0$) :

Théorème 6 (Temam [91], théorème 3.1) *Si $d = 2$, $mes(\Gamma_F) = 0$, $\boldsymbol{h} \in L^2(0,T,V')$ avec V' le dual de V, et $\boldsymbol{u_0} \in H$ avec $H = \{ \boldsymbol{\varphi} \in L^2(\Omega)^d \,/\, \boldsymbol{\nabla}.\boldsymbol{\varphi} = 0 \text{ et } (\boldsymbol{\varphi}.\boldsymbol{n})_{|\Gamma_D=\Gamma} = 0 \}$, alors il existe une unique solution faible $\boldsymbol{u} \in L^2(0,T,V)$ de (2.23)-(2.27), qui vérifie de plus $\boldsymbol{u} \in \mathcal{C}(0,T,H)$.*

Lions s'est quant à lui intéressé au cas où $\boldsymbol{g}(\boldsymbol{x},t)$ est non nul mais stationnaire, en supposant qu'il existait $\boldsymbol{u_g}(\boldsymbol{x})$ tel que

$$\boldsymbol{u}_{|\Gamma}(\boldsymbol{x},t) = \boldsymbol{u_g}(\boldsymbol{x}), \quad \forall (\boldsymbol{x},t) \in \Gamma \times]0,T[$$

($\text{mes}(\Gamma_F) = 0$). Alors en imposant $\boldsymbol{u_0} \in H$, $\boldsymbol{u_g} \in H$ et

$$\boldsymbol{u} - \boldsymbol{u_g} \in \mathrm{L}^2(0,T,H) \cap \mathrm{L}^\infty(0,T,\mathrm{L}^2(\Omega)^d) \tag{2.31}$$

où maintenant $H = \{ \boldsymbol{\varphi} \in \mathrm{H}^1(\Omega)^d \,/\, \boldsymbol{\nabla}.\boldsymbol{\varphi} = 0 \}$, il a montré le résultat suivant (voir [56] et [69]) :

Théorème 7 (Lions [56]) *Pour $d = 2$ ou $d = 3$, si $mes(\Gamma_F) = 0$, le problème (2.23)-(2.27)-(2.31) admet au moins une solution, unique pour $d = 2$. Dans le cas $d = 3$, si la semi-norme $|\boldsymbol{u_0}|_1 = \|\nabla u_0\|_{L^2} = \sqrt{\mathcal{A}(\boldsymbol{u_0}, \boldsymbol{u_0})}$ est assez petite ou si $\boldsymbol{u}$ est suffisamment régulier ($\boldsymbol{u} \in L^\infty(0,T,L^4(\Omega)^3)$), alors la solution est unique.*

Remarquons que (FVNSI) pose problème dans le cadre de la simulation numérique par une méthode variationnelle. Par exemple, dans le cas d'une méthode aux éléments finis, il faudrait être en mesure de définir une base d'éléments formant un sous-espace de H, donc de fonctions de divergence nulle. Cependant, il est possible de construire une base POD composée d'éléments de H. Ainsi, la méthode de Galerkine peut être directement appliquée à (2.27) comme il a été fait à la section 2.1.2.

Construction du modèle POD-Galerkine réduit

Nous allons maintenant expliquer comment construire un modèle POD-Galerkine à partir de la formulation variationnelle (FVNSI) (équation (2.27)).

On suppose que l'on connaît une série de clichés $\boldsymbol{u}^{\boldsymbol{e}}(t_i) \in \mathrm{H}^1(\Omega)$ d'une solution du problème de Navier-Stokes incompressible (2.19)-(2.20)-(2.23)-(2.24)-(2.25) pour des conditions aux limites particulières notées $\boldsymbol{g}^{\boldsymbol{e}}$ et $\boldsymbol{\beta}^{\boldsymbol{e}}$.

La première étape consiste à définir une base POD exploitable (c'est-à-dire d'éléments de V) à partir de $\boldsymbol{u}^{\boldsymbol{e}}$. Pour cela, comme il a été expliqué dans la section 2.1.2, on définit $\bar{\boldsymbol{u}}^{\boldsymbol{e}}(t)$ satisfaisant $\bar{\boldsymbol{u}}^{\boldsymbol{e}}{}_{|\Gamma_D}(t) = \boldsymbol{g}^{\boldsymbol{e}}(t)$. Une base POD de $\tilde{\boldsymbol{u}}^{\boldsymbol{e}} = \boldsymbol{u}^{\boldsymbol{e}} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$ est alors bien une famille de fonctions vérifiant des conditions de Dirichlet homogènes d'après le théorème 3. De même, pour prendre en compte la condition (2.19) d'incompressibilité qui apparaît également dans la définition de l'espace V, il suffit d'imposer $\boldsymbol{\nabla}.\bar{\boldsymbol{u}}^{\boldsymbol{e}} = 0$: dans ce cas, tout mode POD de $\tilde{\boldsymbol{u}}^{\boldsymbol{e}}$ est de divergence nulle. On a le résultat suivant, qui est un corollaire du théorème 3 de la section 1.3.1 :

Théorème 8 *Soit $\bar{\boldsymbol{u}}^{\boldsymbol{e}}(t) \in H^1(\Omega)$ tel que $\bar{\boldsymbol{u}}^{\boldsymbol{e}}{}_{|\Gamma_D} = \boldsymbol{u}^{\boldsymbol{e}}{}_{|\Gamma_D} = \boldsymbol{g}^{\boldsymbol{e}}$ et $\boldsymbol{\nabla}.\bar{\boldsymbol{u}}^{\boldsymbol{e}} = 0$. Alors $\tilde{\boldsymbol{u}}^{\boldsymbol{e}} = \boldsymbol{u}^{\boldsymbol{e}} - \bar{\boldsymbol{u}}^{\boldsymbol{e}} \in V$ et tout mode φ_k de la POD*

$$\tilde{\boldsymbol{u}}^{\boldsymbol{e}}(t) = \sum_k \sigma_k \, a_k^e(t) \, \boldsymbol{\varphi_k}$$

de $\tilde{\boldsymbol{u}}^{\boldsymbol{e}}$ relative à un espace de Hilbert X tel que $H^1(\Omega) \subset X$ ($X = L^2(\Omega)$ ou $H^1(\Omega)$ conviennent) est dans V : $\boldsymbol{\varphi_k} \in H^1(\Omega)$, $\boldsymbol{\varphi_k}_{|\Gamma_D} = \mathbf{0}$ et $\boldsymbol{\nabla}.\boldsymbol{\varphi_k} = 0$.

Par suite, pour tout $a = (a_1 \cdots a_M)^T \in \mathbb{R}^M$,

$$\tilde{\boldsymbol{u}} = \sum_{k=1}^{M} \sigma_k \, a_k \, \boldsymbol{\varphi_k} \in V \ ;$$

ainsi, étant donné $\bar{\boldsymbol{u}}$ de divergence nulle satisfaisant la condition de Dirichlet (2.24) pour $\boldsymbol{g}$ quelconque ($\bar{\boldsymbol{u}}_{|\Gamma_D} = \boldsymbol{g}$ avec $\boldsymbol{g}$ égal ou non à $\boldsymbol{g}^{\boldsymbol{e}}$),

$$\boldsymbol{u}(t) = \bar{\boldsymbol{u}}(t) + \tilde{\boldsymbol{u}}(t) = \bar{\boldsymbol{u}}(t) + \sum_{k=1}^{M} \sigma_k \, a_k(t) \, \boldsymbol{\varphi_k} \tag{2.32}$$

satisfait nécessairement les conditions (2.19) et (2.24) d'incompressibilité et de Dirichlet

$$\boldsymbol{u}_{|\Gamma_D} = \boldsymbol{g} \quad \text{et} \quad \boldsymbol{\nabla}.\boldsymbol{u} = 0$$

pour tout état $a(t) = (a_1(t) \cdots a_M(t))^T$.

Il ne reste donc plus qu'à définir un système d'EDOs régissant $a(t)$ en appliquant la méthode de Galerkine à la formulation variationnelle (FVNSI).

Remarque. Il faut bien distinguer les notations $\boldsymbol{u}^{\boldsymbol{e}}$ et $\boldsymbol{u}$, a_k^e et a_k,... : a_k^e fait référence aux coefficients temporels issus de la POD d'une donnée $\tilde{\boldsymbol{u}}^{\boldsymbol{e}}$, tandis que a_k représente

maintenant les coefficients temporels d'une solution approchée $\boldsymbol{u}$ définie par (2.32). Les coefficients a_k sont donc régis par le problème aux EDOs obtenu par la méthode de Galerkine.

Enfin, la méthode de Galerkine est appliquée à la formulation (FVNSI) en exploitant le sous-espace $Y \subset V$ défini par les M premiers modes d'une POD de $\tilde{\boldsymbol{u}}^{\boldsymbol{e}}$ (voir la section 2.1.2). Formellement, le plus simple est de considérer des modes POD orthonormaux au sens de $\mathrm{L}^2(\Omega)^d$ obtenus avec $X = \mathrm{L}^2(\Omega)^d$, et la méthode de Galerkine conduit alors au modèle dynamique suivant :

Modèle POD-Galerkine réduit des équations de Navier-Stokes incompressibles :

Soit $(\boldsymbol{\varphi_k})$ une base POD de $\tilde{\boldsymbol{u}}^{\boldsymbol{e}} = \boldsymbol{u}^{\boldsymbol{e}} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$ avec

$$\bar{\boldsymbol{u}}^{\boldsymbol{e}}(t) \in \mathrm{H}^1(\Omega)^d, \quad \boldsymbol{\nabla}.\bar{\boldsymbol{u}}^{\boldsymbol{e}}(t) = 0 \quad \text{et} \quad \bar{\boldsymbol{u}}^{\boldsymbol{e}}{}_{|\Gamma_D}(t) = \boldsymbol{u}^{\boldsymbol{e}}{}_{|\Gamma_D}(t) = \boldsymbol{g}^{\boldsymbol{e}}(t) \tag{2.33}$$

relative à $X = \mathrm{L}^2(\Omega)^d$. Alors le modèle réduit issu du problème de Navier-Stokes incompressible (2.19)-(2.20)-(2.23)-(2.24)-(2.25) qui régit une solution approchée $\boldsymbol{u}$ définie par (2.32) et $\bar{\boldsymbol{u}}$ satisfaisant

$$\bar{\boldsymbol{u}}(t) \in \mathrm{H}^1(\Omega)^d, \quad \boldsymbol{\nabla}.\bar{\boldsymbol{u}}(t) = 0 \quad \text{et} \quad \bar{\boldsymbol{u}}_{|\Gamma_D}(t) = \boldsymbol{g}(t), \tag{2.34}$$

est défini, pour tout $i \in [\![1, M]\!]$, par

$$\sigma_i\, \dot{a}_i(t) = C_i^{\bar{u},h}(t) + \sum_{j=1}^{M} \left(C_i^j + C_i^{\bar{u},j}(t)\right)\, a_j(t) + \sum_{j,k=1}^{M} C_i^{j,k}\, a_j(t)\, a_k(t) \tag{2.35}$$

avec

$$C_i^j = -\frac{\sigma_j}{\mathrm{Re}}(\mathcal{A}+\omega\,\mathcal{B})(\boldsymbol{\varphi_j}, \boldsymbol{\varphi_i}), \tag{2.36}$$

$$C_i^{j,k} = -\sigma_j\, \sigma_k\, \mathcal{C}(\boldsymbol{\varphi_j}, \boldsymbol{\varphi_k}, \boldsymbol{\varphi_i}), \tag{2.37}$$

$$\begin{aligned} C_i^{\bar{u},h}(t) = &-\frac{1}{\mathrm{Re}}(\mathcal{A}+\omega\,\mathcal{B})(\bar{\boldsymbol{u}}(t), \boldsymbol{\varphi_i}) - \mathcal{C}(\bar{\boldsymbol{u}}(t), \bar{\boldsymbol{u}}(t), \boldsymbol{\varphi_i}) \\ &+ (\, \boldsymbol{h}(t), \boldsymbol{\varphi_i}\,)_{\mathrm{L}^2(\Omega)^d} - \frac{d}{dt}(\, \bar{\boldsymbol{u}}(t), \boldsymbol{\varphi_i}\,)_{\mathrm{L}^2(\Omega)^d} - (\, \boldsymbol{\beta}(t), \boldsymbol{\varphi_i}{}_{|\Gamma_F}\,)_{\mathrm{L}^2(\Gamma_F)^d}, \end{aligned} \tag{2.38}$$

et

$$C_i^{\bar{u},j}(t) = -\sigma_j\,(\mathcal{C}(\bar{\boldsymbol{u}}(t), \boldsymbol{\varphi_j}, \boldsymbol{\varphi_i}) + \mathcal{C}(\boldsymbol{\varphi_j}, \bar{\boldsymbol{u}}(t), \boldsymbol{\varphi_i}))\,. \tag{2.39}$$

Dans la suite, on choisira la condition initiale $a_i(0) = a_i^e(0)$ et $\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}^{\boldsymbol{e}}$ pour valider les modèles.

Remarque. Le choix de l'espace $X = \mathrm{L}^2(\Omega)^d$ pour définir la POD permet d'obtenir directement un système dynamique de la forme $\dot{a}(t) = f(a(t), t)$. Le recours à l'espace $\mathrm{H}^1(\Omega)^d$ est possible mais il n'est pas pratique, les modes POD obtenus n'étant pas orthogonaux au sens du produit scalaire de $\mathrm{L}^2(\Omega)^d$ qui intervient dans le premier terme de la formulation variationnelle (2.27) : voir la remarque de la section 2.1.2. Par ailleurs, les modes POD obtenus avec $X = \mathrm{H}^1(\Omega)^d$ peuvent conduire à un modèle réduit qui prend mieux en compte les petites échelles des écoulements à grand nombre de Reynolds (voir [38]). Néanmoins cette norme n'est pas dimensionnellement consistante et il faudrait donc pondérer intelligemment les normes L^2 de $\boldsymbol{u}$ et de ses dérivées premières en espace. Notons que ce principe qui consiste à tenir compte dans la définition de la POD des dérivées spatiales a initialement été étudié dans [43] pour la modélisation Galerkine de l'équation de Kuramoto-Sivashinsky. Par ailleurs, le choix de L^2 est justifié d'un point de vue physique puisque **la norme $\mathbf{L}^2(\Omega)^2$ d'un champ de vitesse incompressible adimensionné est égale à deux fois son énergie cinétique** ($\rho = 1$).

Le système d'EDOs (2.35) a une forme polynômiale de degré deux aisément manipulable et prend en compte explicitement l'influence de l'environnement extérieur sur l'écoulement, que ce soit les conditions aux limites de Dirichlet via $\bar{\boldsymbol{u}}$, les forces extérieures via $\boldsymbol{h}$, ou encore les conditions aux limites (2.25) de flux via $\boldsymbol{\beta}$.

Le contrôle actif d'un écoulement se réalise souvent par une action de soufflage/aspiration localisée sur un bord, par une déformation locale d'un obstacle présent dans l'écoulement et modélisable dans une certaine mesure par une condition de Dirichlet sur la vitesse dans le cadre d'actionneurs MEMS (voir [54, p.93]), ou enfin par l'effet d'une force extérieure électromagnétique. La modélisation proposée ici est donc satisfaisante puisque $\bar{\boldsymbol{u}}$ et $\boldsymbol{h}$ permettent de modéliser ce type d'actionneurs.

Cette modélisation est intéressante dès qu'il est peu coûteux de définir des champs $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ et $\bar{\boldsymbol{u}}$ qui satisfont respectivement (2.33) et (2.34). C'est le cas pour des conditions de Dirichlet de la forme

$$\bar{\boldsymbol{u}}^{\boldsymbol{e}}(\boldsymbol{x}, t) = \boldsymbol{g}^{\boldsymbol{e}}(\boldsymbol{x}, t) = \sum_{i=1}^{I} c_i^e(t)\, \boldsymbol{g_i}(\boldsymbol{x}) \quad \text{et} \quad \bar{\boldsymbol{u}}(\boldsymbol{x}, t) = \boldsymbol{g}(\boldsymbol{x}, t) = \sum_{i=1}^{I} c_i(t)\, \boldsymbol{g_i}(\boldsymbol{x}) \tag{2.40}$$

pour des valeurs $c_i^e(t)$ et $c_i(t)$ quelconques en relevant les traces $\boldsymbol{g_i}$ via la résolution d'un problème de Stokes, ceci pour un coût très inférieur à celui de la simulation qui a donné $\boldsymbol{u}^{\boldsymbol{e}}$ pour un nombre I pas trop important (voir le chapitre 5). Néanmoins, pour des conditions aux limites quelconques, calculer $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ revient à calculer $\boldsymbol{u}^{\boldsymbol{e}}$: la modélisation réduite pose problème pour des conditions aux bords "de grande dimension". Cependant, on peut alors essayer de traiter de manière implicite les conditions aux limites (voir la section suivante).

Comme nous avons pu le constater, le traitement formel des conditions aux limites (2.24) et (2.25) du problème de Navier-Stokes incompressible ne pose pas de difficulté. Cependant, il peut exister des situations où les conditions aux limites, définies par $\boldsymbol{g}^{\boldsymbol{e}}$ et $\boldsymbol{\beta}^{\boldsymbol{e}}$,

ne sont pas explicitement connues pour la donnée $\boldsymbol{u^e}$ utilisée : c'est le cas par exemple pour des données expérimentales ou des données calculées par simulation numérique et extraites sur un sous-ensemble du domaine de calcul. Il est alors encore possible de construire le modèle réduit proposé puisque $\boldsymbol{g^e}$ est défini de manière unique par les données. De plus, pour ω fixé (par exemple $\omega = 1$ ou 0 si on veut construire un modèle régi explicitement par le flux de σ ou de $\tilde{\sigma}$), on peut déterminer une valeur approchée de $\boldsymbol{\beta^e}$ si on veut valider le modèle en le simulant pour $\boldsymbol{\beta} = \boldsymbol{\beta^e}$, ou si on veut l'exploiter pour une nouvelle condition de Dirichlet mais dans les mêmes conditions de flux.

Remarque. Dans le cas d'un écoulement périodique par rapport à des bords noté Γ_P^1 et Γ_P^2, on a, pour toute fonction $\boldsymbol{\varphi}$ de

$$V_P = \{\, \boldsymbol{\varphi} \in V \,/\, \boldsymbol{\varphi}_{|_{\Gamma_P^1}} \equiv \boldsymbol{\varphi}_{|_{\Gamma_P^2}} \,\}$$

et avec $\Gamma_P = \Gamma_P^1 \cup \Gamma_P^2$, la relation

$$(\, T\,\boldsymbol{n}, \boldsymbol{\varphi} \,)_{\mathrm{L}^2(\Gamma_P)^d} = 0$$

si le tenseur T est périodique, puisque $\boldsymbol{n}_{|_{\Gamma_P^1}} \equiv -\boldsymbol{n}_{|_{\Gamma_P^2}}$: la formulation (FVNSI) reste donc valable dans le cas d'une direction périodique. Le modèle POD-Galerkine peut alors être construit de la même manière puisque le caractère périodique est transmis aux modes POD qui appartiennent ainsi à V_P (voir la section 1.3.1).

Prise en compte implicite des conditions aux limites et problème du terme de pression

Les conditions naturelles de flux qui apparaissent lorsque l'on applique une fonction test $\boldsymbol{\varphi} \in V$ à l'équation (2.20) puis une intégration par parties ont été explicitement prises en compte dans le modèle réduit POD-Galerkine. Il paraît cohérent que le modèle construit dépende des mêmes effets environnementaux que l'écoulement incompressible considéré, à savoir des forces volumiques $\boldsymbol{h}$ et des conditions aux bords modélisées par $\bar{\boldsymbol{u}}$ et $\boldsymbol{\beta}$. Cependant, plusieurs auteurs ([76] et [25] par exemple) essayent de s'affranchir de la prise en compte des conditions de flux, ce qui peut paraître surprenant mais présente certains avantages pratiques : le modèle obtenu et son utilisation sont simplifiée, puisqu'on n'a alors plus de conditions de flux à lui fournir (elles sont implicitement déterminées par le modèle).

Si la méthode de Galerkine est formellement appliquée à (2.20) sans intégration par parties, considérant le sous-espace engendré par les M premiers modes POD $\boldsymbol{\varphi_k}$ de $\boldsymbol{u}$ relatifs à $X = \mathrm{L}^2(\Omega)^d$, on obtient

$$\sigma_i\, \dot{a}_i(t) = -\sum_{j,k=1}^{M} \sigma_j\, \sigma_k\, \mathcal{C}(\boldsymbol{\varphi_j}, \boldsymbol{\varphi_k}, \boldsymbol{\varphi_i}) a_j(t)\, a_k(t) + \frac{1}{\mathrm{Re}} \sum_{j=1}^{M} \sigma_j\, (\, \Delta\boldsymbol{\varphi_j}, \boldsymbol{\varphi_i} \,)_{\mathrm{L}^2(\Omega)^d} - (\, \boldsymbol{\nabla} p, \boldsymbol{\varphi_i} \,)_{\mathrm{L}^2(\Omega)^d},$$

avec $\boldsymbol{h} = \boldsymbol{0}$ et $\boldsymbol{g} = \boldsymbol{0}$ pour simplifier. Ainsi le terme de pression est le seul qui ne soit pas exprimé en fonction des modes POD $\boldsymbol{\varphi_k}$ et des coefficients temporels a_k. De plus, il n'est pas possible à partir de la base POD d'exprimer simplement la pression en utilisant les modes POD du champ de vitesse, la pression n'étant pas une fonction linéaire de la vitesse d'après l'équation de Poisson sur la pression (voir [76]). Puisque

$$(\boldsymbol{\nabla} p, \boldsymbol{\varphi_i})_{\mathrm{L}^2(\Omega)^d} = (p, \boldsymbol{\varphi_i}.\boldsymbol{n})_{\mathrm{L}^2(\Gamma)} ,$$

une condition de Dirichlet sur la pression ou de la forme (2.25) permettrait d'obtenir un système dynamique exploitable. Remarquons que ce dernier nécessiterait une régularité spatiale supérieure de $\boldsymbol{u}$ par rapport au modèle précédent et en pratique jusqu'à deux différentiation spatiales de $\bar{\boldsymbol{u}}$ et des modes POD contre une, à moins que la formule de Green ne soit appliquée au terme $(\Delta\boldsymbol{\varphi_j}, \boldsymbol{\varphi_i})_{\mathrm{L}^2(\Omega)^d}$ (le calcul du modèle nécessite alors l'évaluation de termes de bord). Ainsi, celui qui ne souhaite pas prendre en compte de manière explicite les conditions de flux (ou une condition de Dirichlet sur la pression) est confronté à un (faux) problème de modélisation du terme de pression.

Pour résoudre ce problème, deux approches ont été proposées. La première, définie par Rempfer dans [76], consiste à construire le modèle dynamique à partir de la formulation vitesse-tourbillon (2.22) où le terme de pression n'apparaît plus. En effet, le rotationnel d'un champ de vitesse approché par une combinaison linéaire d'un champ $\bar{\boldsymbol{u}}$ et de M modes POD sous la forme (2.32) s'exprime par

$$\boldsymbol{w} = \bar{\boldsymbol{w}} + \sum_{k=1}^{M} \sigma_k \, a_k \, \boldsymbol{\varphi_k^w}$$

avec $\boldsymbol{w} = \boldsymbol{\nabla} \times \boldsymbol{u}$, $\bar{\boldsymbol{w}} = \boldsymbol{\nabla} \times \bar{\boldsymbol{u}}$ et $\boldsymbol{\varphi_k^w} = \boldsymbol{\nabla} \times \boldsymbol{\varphi_k}$. Ainsi, la méthode de Galerkine appliquée à (2.22) aboutit au système d'EDOs

$$B \, \Sigma \, \dot{a}(t) = f^w(a(t), t) \quad \text{avec} \quad \Sigma_{i,j} = \sigma_j \, \delta_{i,j} \text{ et } B_{i,j} = \sigma_j \, (\boldsymbol{\varphi_i^w}, \boldsymbol{\varphi_j^w})_{\mathrm{L}^2(\Omega)^d} \tag{2.41}$$

et

$$\begin{aligned} f_i^w(a(t), t) &= C_{w,i}^{\bar{u},\bar{w}}(t) + \sum_{j=1}^{M} \left(C_{w,i}^{\bar{u},\bar{w},j}(t) + C_{w,i}^{j} \right) a_j(t) + \sum_{j,k=1}^{M} C_{w,i}^{j,k} \, a_j(t) \, a_k(t) \\ &- \frac{d}{dt} (\bar{\boldsymbol{w}}, \boldsymbol{\varphi_i^w})_{\mathrm{L}^2(\Omega)^d} + (\boldsymbol{\nabla} \times \boldsymbol{h}, \boldsymbol{\varphi_i^w})_{\mathrm{L}^2(\Omega)^d} \end{aligned}$$

où

$$\begin{aligned} C_{w,i}^{\bar{u},\bar{w}}(t) &= ((\bar{\boldsymbol{w}}.\boldsymbol{\nabla})\bar{\boldsymbol{u}} - (\bar{\boldsymbol{u}}.\boldsymbol{\nabla})\bar{\boldsymbol{w}}, \boldsymbol{\varphi_i^w})_{\mathrm{L}^2(\Omega)^d} + \frac{1}{\mathrm{Re}} (\Delta\bar{\boldsymbol{w}}, \boldsymbol{\varphi_i^w})_{\mathrm{L}^2(\Omega)^d} , \\ C_{w,i}^{\bar{u},\bar{w},j}(t) &= \sigma_j \, ((\bar{\boldsymbol{w}}.\boldsymbol{\nabla})\boldsymbol{\varphi_j} - (\boldsymbol{\varphi_j}.\boldsymbol{\nabla})\bar{\boldsymbol{w}} + (\boldsymbol{\varphi_j^w}.\boldsymbol{\nabla})\bar{\boldsymbol{u}} - (\bar{\boldsymbol{u}}.\boldsymbol{\nabla})\boldsymbol{\varphi_j^w}, \boldsymbol{\varphi_i^w})_{\mathrm{L}^2(\Omega)^d} , \\ C_{w,i}^{j} &= \frac{\sigma_i}{\mathrm{Re}} (\Delta\boldsymbol{\varphi_j^w}, \boldsymbol{\varphi_i^w})_{\mathrm{L}^2(\Omega)^d} , \\ C_{w,i}^{j,k} &= \sigma_j \, \sigma_k \, ((\boldsymbol{\varphi_j^w}.\boldsymbol{\nabla})\boldsymbol{\varphi_k} - (\boldsymbol{\varphi_k}.\boldsymbol{\nabla})\boldsymbol{\varphi_j^w}, \boldsymbol{\varphi_i^w})_{\mathrm{L}^2(\Omega)^d} . \end{aligned}$$

Après inversion de la matrice B, on obtient de nouveau un système dynamique polynômial régissant les a_i. Il est possible de modifier les définitions de $C_{w,i}^{\bar{u},\bar{w}}$ et $C_{w,i}^{j}$ en appliquant la formule de Green : cela permet d'éviter de différentier trois fois $\bar{\boldsymbol{u}}$ et les modes POD (pour calculer $\Delta\bar{\boldsymbol{w}}$ et les $\Delta\boldsymbol{\varphi}_{\boldsymbol{j}}^{\boldsymbol{w}}$), mais nécessite le calcul d'un terme de bord.

La seconde approche, proposée par Galletti *et al.* [25], consiste à approcher le terme de pression par un modèle linéaire

$$-\left(\boldsymbol{\nabla} p, \boldsymbol{\varphi_i}\right)_{\mathrm{L}^2(\Omega)^d} = \sum_{j=1}^{M} P_j^p \, a_j(t)$$

où les coefficients réels P_j^p sont estimés à partir des données temporelles de la POD, c'est-à-dire les a_i^e (voir [25]).

Notons que Noack *et al.* [67] suggèrent un modèle quadratique de prise en compte de la pression.

Les modèles de Rempfer et de Galletti *et al.* définissent un champ des vitesses approché $\boldsymbol{u}$ qui satisfait exactement les conditions de Dirichlet imposées par $\bar{\boldsymbol{u}}$ si les modes POD ont été calculés à partir d'une donnée $\tilde{\boldsymbol{u}}^{\boldsymbol{e}}$ qui vérifie les conditions de Dirichlet homogènes correspondantes : dans le cadre du contrôle actif par soufflage/aspiration, ces modèles semblent donc exploitables même si les conditions aux limites de flux ne sont pas maîtrisées. Cependant, s'il est nécessaire de prendre en compte des conditions de type (2.25) de manière explicite, les modèles de Rempfer, Galletti *et al.* ne sont plus utilisables.

En outre, le problème du terme de pression survient également si on ne tient pas explicitement compte de toutes les conditions de Dirichlet comme précédemment. En effet, la formulation variationnelle (2.27) n'est utilisable que pour des fonctions tests de V ; le cas échéant il n'est pas possible de faire disparaître la pression. Les méthodes de Rempfer, Galletti *et al.* peuvent alors s'avérer utiles.

Il est peu coûteux de calculer des champs $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ et $\bar{\boldsymbol{u}}$ qui satisfont les conditions (2.33) et (2.34) avec des conditions de Dirichlet de la forme (2.40) (pour I petit). Néanmoins, pour des conditions aux limites de Dirichlet complexes, la modélisation de Galletti *et al.* ou de Rempfer est une alternative intéressante qui permet de s'affranchir de $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ et $\bar{\boldsymbol{u}}$ en définissant un modèle autonome, facile à manipuler puisqu'il est alors sous forme polynômiale et à coefficients constants. Le modèle détermine alors lui-même ses conditions aux limites et il est capable de reproduire de manière approchée les conditions aux limites de la donnée $\boldsymbol{u}^{\boldsymbol{e}}$.

Il est important de noter qu'il est envisageable de tenir compte seulement des conditions de Dirichlet sur un sous-ensemble du bord Γ_D en utilisant les modélisations de Rempfer et Galletti pour calculer le modèle. Dans le cadre du contrôle actif, on peut donc construire un modèle qui tiendrait explicitement compte d'un actionneur modélisé par une condition de Dirichlet tout en s'affranchissant des autres conditions aux limites.

Au chapitre suivant, le modèle "tourbillonnaire" de Rempfer est testé pour un écoulement comportant une condition d'entrée turbulente. Cependant, cette condition d'entrée

est quasi-instationnaire et nous avons pu également construire le modèle (2.35) en négligeant certains termes de bord.

Prise en compte passive de la température

Si le but est de contrôler la température de l'écoulement, il est possible de compléter le modèle incompressible par un système d'EDOs de la forme

$$\dot{b}(t) = f(a(t), b(t), t) \tag{2.42}$$

qui permet de prendre en compte la température θ, en appliquant la méthode POD-Galerkine à l'équation (2.21).

Par exemple, considérons pour simplifier des conditions de Dirichlet homogènes sur toute la frontière Γ et une décomposition de θ notée

$$\theta(t) = \sum_k b_k(t)\, \psi_k$$

dans une base (ψ_k) de

$$W = \{\, \phi \in \mathrm{H}^1(\Omega) \,/\, \phi_{|\Gamma} = 0 \,\}$$

orthonormale pour le produit scalaire de $\mathrm{L}^2(\Omega)$. Alors, pour une fonction test ψ de W, l'équation (2.21) conduit à la formulation variationnelle

$$\frac{d}{dt}\,(\,\theta, \psi\,)_{\mathrm{L}^2(\Omega)} + \frac{\gamma}{\mathrm{RePr}}\,\mathcal{A}_\theta(\theta,\psi) = \gamma\left[\frac{1}{2\,\mathrm{Re}}\,\mathcal{C}_\theta(\boldsymbol{u},\boldsymbol{u},\psi) + \mathcal{B}_\theta(\boldsymbol{u},\psi)\right]$$

avec les formes multilinéaires

$$\mathcal{A}_\theta(\theta,\psi) = (\,\boldsymbol{\nabla}\theta, \boldsymbol{\nabla}\psi\,)_{\mathrm{L}^2(\Omega)^d}\,,$$

$$\mathcal{C}_\theta(\boldsymbol{u},\boldsymbol{u},\psi) = (\,\mathrm{Trace}(S(\boldsymbol{u})^2), \psi\,)_{\mathrm{L}^2(\Omega)}\,,$$

et

$$\mathcal{B}_\theta(\boldsymbol{u},\psi) = (\,\boldsymbol{h}.\boldsymbol{u}, \psi\,)_{\mathrm{L}^2(\Omega)}\,.$$

Ainsi, en utilisant la décomposition (2.32) de $\boldsymbol{u}$, la méthode de Galerkine aboutit à des EDOs cubiques de la forme (2.42) en considérant l'espace engendré par les premiers modes POD d'une donnée θ^e relativement à $X = \mathrm{L}^2(\Omega)$. Un modèle intégrant la température est proposé par Ravindran [74] et par Fahl [19].

Remarque. Il existe un système d'EDPs pour les problèmes de convection thermique dans un écoulement incompressible, appelé *équation de Rayleigh-Benard*, où les équations de conservation de la masse et de la quantité de mouvement sont couplées avec l'équation régissant la température via les forces de gravité : $\boldsymbol{h} = -G\,\boldsymbol{z}\,\theta$ dans (2.20) où $-G\,\boldsymbol{z}$ est un vecteur indépendant des grandeurs thermodynamiques du fluide. Il suffit alors de poser

$$(\,\boldsymbol{h}(t), \boldsymbol{\varphi_i}\,) = -\sum_k (\,G\,\psi_k\,\boldsymbol{z}, \boldsymbol{\varphi_i}\,)\,b_k(t)$$

dans (2.38) pour obtenir un modèle POD-Galerkine polynômial.

2.3.2 Modèles compressibles

Comme nous l'avons évoqué précédemment, il n'est pas évident de définir un système d'EDOs simple et physiquement cohérent des équations de Navier-Stokes compressibles. En effet, il existe de nombreuses possibilités quant au choix des variables thermodynamiques sur lesquelles effectuer la POD ou quant à la manière de définir la POD. Faut-il calculer une POD pour chaque variable thermodynamique ou une POD globale ? Quel produit scalaire choisir pour calculer la POD ?

Cette section présente succinctement les solutions apportées par Rowley *et al.* [81, 83] et Vigo *et al.* [94] sur le choix des variables et du produit scalaire, sans se préoccuper des conditions aux limites.

Tout d'abord, pour illustrer le problème du choix des variables, notons que la méthode POD-Galerkine appliquée aux équations de Navier-Stokes compressibles n'aboutit pas à un système dynamique de la forme $\dot{a}(t) = f(a(t), t)$ ni pour le jeu des variables primitives $(\rho, \boldsymbol{u}, p)$, ni pour le jeu $(\rho, \boldsymbol{u}, \theta)$. En effet, pour $q = (\rho\, \boldsymbol{u}\, p)^T$ ou $q = (\rho\, \boldsymbol{u}\, \theta)^T$, les systèmes (2.14)-(2.15)-(2.16) et (2.14)-(2.15)-(2.17) peuvent s'écrire sous la forme

$$A(q)\, \partial_t q = f_1(q) + f_2(q, q) + f_3(q, q, q)$$

où f_1, f_2 et f_3 sont multilinéaires et où

$$A(q) = B + L(q) \quad \text{avec} \quad B = \text{diag}(1, 0, 0) \text{ et } L(q) = \text{diag}(0, \rho, \rho).$$

En conséquence, considérant une décomposition de q notée

$$q = \sum_{k=1}^{M} a_k(t)\, \varphi_k,$$

la "projection" de Galerkine conduit formellement, pour un produit scalaire $(\,.\,,.\,)$, à un système d'EDOs de la forme

$$M(a(t))\, \dot{a}(t) = \sum_{k=1}^{M} \begin{pmatrix} C_k^1 \\ \vdots \\ C_k^M \end{pmatrix} a_k(t) + \sum_{k,l=1}^{M} \begin{pmatrix} C_{k,l}^1 \\ \vdots \\ C_{k,l}^M \end{pmatrix} a_k(t)\, a_l(t) + \sum_{k,l,j=1}^{M} \begin{pmatrix} C_{k,l,j}^1 \\ \vdots \\ C_{k,l,j}^M \end{pmatrix} a_k(t)\, a_l(t)\, a_j(t) \tag{2.43}$$

avec $C_k^i = (\, f_1(\varphi_k), \varphi_i\,)$, $C_{k,l}^i = (\, f_2(\varphi_k, \varphi_l), \varphi_i\,)$, $C_{k,l,j}^i = (\, f_3(\varphi_k, \varphi_l, \varphi_j), \varphi_i\,)$ et

$$M_{i,j}(a(t)) = (\, B\, \varphi_j, \varphi_i\,) + \sum_{k=1}^{M} (\, L(\varphi_k)\, \varphi_j, \varphi_i\,)\; a_k(t).$$

Ce système d'EDOs définit ainsi $\dot{a}(t)$ de manière implicite : il est donc difficilement exploitable.

En outre, le choix des variables conservatives conduit à un système d'EDPs contenant des termes qui ne sont pas multilinéaires mais présentent des fractions rationnelles : le système d'EDOs qui en découle n'est pas pratique (il n'est pas sous forme polynômiale) : voir [94] (ou [59] dans le cas des équations d'Euler, c'est-à-dire pour $\mathrm{Re} = +\infty$).

Dans les deux cas, variables primitives ou conservatives, le problème provient de la variable ρ. Pour pallier cette difficulté, Vigo propose dans [94] d'utiliser le *covolume* (ou *volume spécifique*) $\tau = \frac{1}{\rho}$ plutôt que ρ. En effet, puisque $\partial_v \tau = -\tau^2 \partial_v \rho$ pour toute variable spatio-temporelle $v \in \{x_1, \cdots, x_d, t\}$, l'équation (2.14) donne

$$D_t \tau - \tau \, \boldsymbol{\nabla}.\boldsymbol{u} = 0$$

et les équations de Navier-Stokes compressibles de la page 42 peuvent se réécrire sous la forme

$$\partial_t q = f_1(q) + f_2(q,q) + f_3(q,q,q)$$

où f_1, f_2 et f_3 sont multilinéaires pour les jeux de variables $q = q_p = (\tau \, \boldsymbol{u} \, p)^T$ ou $q = q_\theta = (\tau \, \boldsymbol{u} \, \theta)^T$. Ainsi la méthode POD-Galerkine permet de construire un système d'EDOs de la forme $\dot{a}(t) = f(a(t), t)$ et polynômial de degré trois. Il suffit pour s'en convaincre de poser $L(.) \equiv 0$ et $B = \mathrm{I}_3$ dans (2.43) : $M(a(t)) = \mathrm{I}_3$ pour des fonctions de base φ_k orthonormales.

Cependant, même si q_p et q_θ sont significatifs d'un point de vue physique, il n'est pas possible de construire un produit scalaire de type L^2 consistant avec l'énergie de l'écoulement avec l'un ou l'autre de ces jeux de variable. En effet, $e = \gamma \, \theta$ après adimensionnement et l'énergie totale volumique E de l'écoulement vérifie

$$\tau \, E = \gamma \, \theta + \frac{1}{2} \, \boldsymbol{u}^T \, \boldsymbol{u} = \frac{\gamma^2}{\gamma - 1} \, p \, \tau + \frac{1}{2} \, \boldsymbol{u}^T \, \boldsymbol{u}.$$

Aisni, ni E, ni l'énergie totale massique $\tau \, E$ ne peuvent se mettre sous la forme $q^T \, S_E \, q$ avec S_E une matrice symétrique définie positive, que ce soit pour $q = q_p$ ou $q = q_\theta$. Notons qu'il existe une unique matrice S symétrique telle que $\tau \, E = {q_p}^T \, S \, q_p$ pour tout q_p, mais celle-ci n'est pas définie positive puisqu'elle admet la valeur propre $-\gamma^2/(2\,(\gamma - 1)) < 0$.

C'est pourquoi, pour un écoulement isentropique (les équations de Navier-Stokes peuvent alors se mettre sous la forme de quatre équations scalaires), Rowley *et al.* [83] proposent de construire un système d'EDOs à partir du jeu de variable $q_c = (\boldsymbol{u} \;\; c)^T$, où $c = \sqrt{(\gamma - 1)\,\theta}$ désigne la célérité locale du son adimensionnée. La méthode POD-Galerkine permet alors d'obtenir un système d'EDOs polynômial de degré deux comme dans le cas incompressible, avec de plus ${q_c}^T \, S_E \, q_c = \tau \, E$ pour $S_E = \mathrm{diag}(2/(\gamma\,(\gamma - 1)), 1/2)$ (on peut aussi redéfinir S_E pour obtenir l'enthalpie d'arrêt spécifique, voir [83]).

La modélisation de Rowley est cohérente d'un point de vue physique, et l'utilisation de c est intéressante puisqu'il serait possible de prendre en compte la variation de la viscosité avec la température en conservant un système dynamique polynômial : c est proportionnel à

$\sqrt{\theta}$ donc à $\frac{1}{\mathrm{Re}}$ (voir la remarque de la page 42). Néanmoins, l'utilisation de la célérité du son pose problème dans le cadre compressible plus général des équations (2.14)-(2.15)-(2.16), puisqu'elle ne permet pas d'aboutir à un système polynômial de la forme $\dot{a}(t) = f(a(t), t)$ pour les mêmes raisons que le recours à la variable ρ.

En conclusion, il est possible de construire un modèle POD-Galerkine polynômial compressible en traitant de manière implicite les conditions aux bords, même si dans le cas général du modèle de Vigo l'interprétation physique de la POD soulève des interrogations. Des essais numériques ont montré que le modèle de Vigo pouvait être efficace : [94] et [38].

2.4 Remarques sur la stabilité des modèles POD-Galerkine fluides

La méthode POD-Galerkine permet de définir formellement des systèmes d'EDOs, mais rien n'indique pour l'instant qu'elle soit robuste. En effet, l'expression polynômiale nonlinéaire des modèles soulève la question de leur stabilité : l'unique solution maximale d'un problème de Cauchy associé à un système dynamique polynômial de degré strictement supérieur à un peut exploser en temps fini (voir la section 4.6.2). De plus, un système dynamique polynômial peut être particulièrement sensible à une perturbation infinitésimale de l'un de ses coefficients polynômiaux.

Cette section présente un résultat de stabilité pour le modèle incompressible en reprenant les travaux de Vigo [94] et aborde succinctement les problèmes structuraux intrinsèques à la modélisation (POD-)Galerkine.

2.4.1 Analyse théorique des interactions énergétiques globales du modèle incompressible

Nous allons reprendre l'analyse que Vigo a menée dans [94] sur la formulation variationnelle de Leray afin d'analyser les interactions énergétiques de notre modèle incompressible (2.35) construit pour $\omega = 0$.

Si on multiplie (2.35) par $\sigma_i\, a_i(t)$ et que l'on somme sur i, on obtient pour $\omega = 0$

$$\underbrace{\frac{1}{2}\sum_{i=1}^{M}\frac{d}{dt}(\lambda_i\, a_i(t)^2)}_{= \frac{d}{dt}\|\boldsymbol{u}(t)\|_{\mathrm{L}^2(\Omega)^d}^2} = I_D(a(t)) + I_T(a(t)) + I_E(a(t), \bar{\boldsymbol{u}}(t), \boldsymbol{\beta}(t), \boldsymbol{h}(t))$$

où

- I_D désigne les interactions modales énergétiques diadiques :

$$I_D(a(t)) = \sum_{i,j=1}^{M} \sigma_i \, C_i^j \, a_i(t) \, a_j(t) = -\frac{1}{\mathrm{Re}} \, \mathcal{A}(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}) \; ;$$

- I_T désigne les interactions modales énergétiques triadiques :

$$I_T(a(t)) = \sum_{i,j,k=1}^{M} \sigma_i \, C_i^{j,k} \, a_i(t) \, a_j(t) \, a_k(t) = -\mathcal{C}(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}) \; ; \tag{2.44}$$

- et I_E désigne les interactions énergétiques entre l'écoulement et son environnement :

$$\begin{aligned} I_E(a(t), \bar{\boldsymbol{u}}(t), \boldsymbol{\beta}(t), \boldsymbol{h}(t)) &= \sum_{i=1}^{M} \sigma_i \, C_i^{\bar{u},h}(t) + \sum_{i,j=1}^{M} \sigma_i \, C_i^{\bar{u},j} \, a_i(t) \, a_j(t) \\ &= -\frac{1}{\mathrm{Re}} \, \mathcal{A}(\bar{\boldsymbol{u}}(t), \tilde{\boldsymbol{u}}(t)) - \mathcal{C}(\boldsymbol{u}(t), \bar{\boldsymbol{u}}(t), \tilde{\boldsymbol{u}}(t)) - \mathcal{C}(\bar{\boldsymbol{u}}(t), \tilde{\boldsymbol{u}}(t), \tilde{\boldsymbol{u}}(t)) \\ &\quad + (\, \boldsymbol{h}(t), \tilde{\boldsymbol{u}}(t) \,)_{\mathrm{L}^2(\Omega)^d} - \frac{d}{dt} (\, \bar{\boldsymbol{u}}(t), \tilde{\boldsymbol{u}}(t) \,)_{\mathrm{L}^2(\Omega)^d} - (\, \boldsymbol{\beta}(t), \tilde{\boldsymbol{u}}(t) \,)_{\mathrm{L}^2(\Gamma_F)^d} \end{aligned}$$

Puisque la forme $\mathcal{A}(.,.)$ est positive, on a $I_D(a(t)) \leq 0$: **à tout instant, l'effet global des interactions diadiques est une dissipation de l'énergie**, ce qui normal puisque celles-ci sont d'origine visqueuse. Par ailleurs, on a

$$\begin{aligned} \mathcal{C}(\boldsymbol{\phi}, \boldsymbol{\phi}, \boldsymbol{\phi}) &= \sum_{i,j=1}^{d} (\, \phi_{x_i} \, \phi_{x_j}, \partial_{x_j} \phi_{x_i} \,)_{\mathrm{L}^2(\Omega)} \\ &= -\sum_{i,j=1}^{d} (\, \partial_{x_j}(\phi_{x_i} \, \phi_{x_j}), \phi_{x_i} \,)_{\mathrm{L}^2(\Omega)} + \sum_{i,j=1}^{d} (\, \phi_{x_j} \, n_{x_j}, \phi_{x_i} \, \phi_{x_i} \,)_{\mathrm{L}^2(\Gamma)} \\ &= -\sum_{i,j=1}^{d} (\, \phi_{x_i} \, \phi_{x_j}, \partial_{x_j} \phi_{x_i} \,)_{\mathrm{L}^2(\Omega)} - \sum_{i,j=1}^{d} (\, \phi_{x_i}^2, \partial_{x_j} \phi_{x_j} \,)_{\mathrm{L}^2(\Omega)} + (\, \boldsymbol{\phi}.\boldsymbol{n}, \boldsymbol{\phi}.\boldsymbol{\phi} \,)_{\mathrm{L}^2(\Gamma)} \\ &= -\mathcal{C}(\boldsymbol{\phi}, \boldsymbol{\phi}, \boldsymbol{\phi}) - \sum_{i=1}^{d} (\, \phi_{x_i}^2, \boldsymbol{\nabla}.\boldsymbol{\phi} \,)_{\mathrm{L}^2(\Omega)} + (\, \boldsymbol{\phi}.\boldsymbol{n}, \|\boldsymbol{\phi}\|_2{}^2 \,)_{\mathrm{L}^2(\Gamma)} \, , \end{aligned}$$

donc

$$\mathcal{C}(\boldsymbol{\phi}, \boldsymbol{\phi}, \boldsymbol{\phi}) = -\frac{1}{2} \sum_{i=1}^{d} (\, \phi_{x_i}^2, \boldsymbol{\nabla}.\boldsymbol{\phi} \,)_{\mathrm{L}^2(\Omega)} + \frac{1}{2} (\, \boldsymbol{\phi}.\boldsymbol{n}, \|\boldsymbol{\phi}\|_2{}^2 \,)_{\mathrm{L}^2(\Gamma)} \, .$$

En particulier,

$$\forall \boldsymbol{\varphi} \in V \qquad \mathcal{C}(\boldsymbol{\varphi}, \boldsymbol{\varphi}, \boldsymbol{\varphi}) = \frac{1}{2} (\, \boldsymbol{\varphi}.\boldsymbol{n}, \|\boldsymbol{\varphi}\|_2{}^2 \,)_{\mathrm{L}^2(\Gamma_F)} \, .$$

En conséquence, puisque $\tilde{\boldsymbol{u}} \in V$ et d'après (2.44), **si l'écoulement incompressible ne comporte pas de frontière libre, c'est-à-dire si $\Gamma_D \equiv \Gamma$ (mes$(\Gamma_F) = 0$), alors les**

interactions triadiques globales sont nulles à tout instant.

En conclusion, **si l'écoulement modélisé ne comporte pas de frontière libre, l'énergie cinétique naturelle du modèle ne peut augmenter que par les effets environnementaux** : aspiration/soufflage de fluide, force électromagnétique...

Cette analyse théorique ne permet pas de conclure pour un écoulement décollé comportant une frontière libre sans une nouvelle hypothèse. En outre, dans le cas où $\mathrm{mes}(\Gamma_F) = 0$, il faut noter que les interactions triadiques sont alors nulles puisque $\boldsymbol{\nabla}.\tilde{\boldsymbol{u}} = 0$: dans la pratique, les interactions triadiques numériques peuvent toutefois conduire à une augmentation (ou une diminution) artificielle de l'énergie du système, puisque la divergence des modes POD n'est pas rigoureusement nulle (ou à cause d'autres sources d'erreurs numériques).

Dans le chapitre suivant, une analyse numérique et physique des interactions énergétiques au sein d'une base POD est menée pour un écoulement incompressible comportant une frontière libre en sortie. L'objectif de cette étude sera d'analyser les interactions énergétiques locales, afin notamment d'avoir une meilleure connaissance des interactions à modéliser entre les modes POD utilisés et les modes qui sont négligés.

2.4.2 Difficultés structurelles

Même si le modèle POD-Galerkine incompressible réduit théorique est stable, le comportement numérique des modèles peut être très différent de l'écoulement voire instable. En effet, la méthode POD-Galerkine n'est pas robuste dans la mesure où une perturbation infinitésimale du modèle, donc en pratique les erreurs numériques, peuvent modifier radicalement le comportement du modèle obtenu.

Pour illustrer ce phénomène, nous allons nous appuyer sur un exemple proposé par Noack *et al.* dans [66]. Considérons le système dynamique quadratique d'état $u = (u_1\, u_2\, u_3)^T$ qui suit :

$$\left\{ \begin{array}{rcl} \dot{u}_1 &=& \mu\, u_1 - u_2 - u_1\, u_3 \\ \dot{u}_2 &=& \mu\, u_2 + u_1 - u_2\, u_3 \\ \dot{u}_3 &=& -u_3 + u_1^2 + u_2^2 \end{array} \right.$$

Ce système admet dans le plan $u_3 = \mu$ une solution périodique qui définit un cycle limite : $u_1(t) = \sqrt{\mu}\,\cos(t)$, $u_2(t) = \sqrt{\mu}\,\sin(t)$ et $u_3(t) = \mu$. Les vecteurs $\varphi_1 = (1\,0\,0)^T$ et $\varphi_2 = (0\,1\,0)^T$ forment une base POD dans $X = \mathbb{R}^3$ de l'état fluctuant $\tilde{u} = u - \bar{u}$ de cette solution observée sur une période, où $\bar{u}$ est l'état moyen $(0\,0\,\mu)^T$. Considérant la décomposition $u = \bar{u} + a_1\,\varphi_1 + a_2\,\varphi_2$ et le produit scalaire euclidien de $\mathbb{R}^3$, la méthode de Galerkine appliquée au système précédent conduit au système suivant :

$$\dot{a}_1 = -a_2, \qquad \dot{a}_2 = a_1.$$

Une petite perturbation de ce système autonome, par exemple $\dot{a}_1 = \varepsilon_1\, a_1 - a_2$ et $\dot{a}_2 = \varepsilon_2\, a_2 + a_1$, peut conduire à des solutions qui divergent très rapidement de la solution de référence ; la trajectoire périodique précédente n'est plus stable (dans le plan $u_3 =$

μ) : le modèle POD-Galerkine est *structurellement instable* (consulter [31]). Rempfer avait proposé le même type d'analyse dans [77], également à partir d'un système dynamique de référence de dimension finie mais aussi en analysant le modèle POD-Galerkine numérique d'un écoulement.

Pour éviter ce problème structurel, la seule solution semble de modifier la base des fonctions spatiales qui est exploitée par la méthode de Galerkine, la manière la plus pratique étant de compléter cette base par de nouvelles fonctions préalablement orthonormalisées. Noack *et al.* ont ainsi montré pour un écoulement satisfaisant des conditions aux limites instationnaires l'intérêt d'ajouter à la base modale la solution du problème de Stokes associé. Rempfer suggère également de compléter la base modale par des éléments appartenant au noyau de l'opérateur K associé à la POD (qui a été défini au chapitre précédent). De plus, il conseille d'imposer de manière explicite les conditions aux limites de Dirichlet au niveau structurel comme nous l'avons vu précédemment.

2.5 Un exemple de modélisation d'un écoulement 2D, laminaire et incompressible

Ce premier exemple, qui servira plus tard de configuration de test, est un écoulement bidimensionnel et quasi-incompressible autour d'un obstacle de forme carrée. Le nombre de Reynolds basé sur la vitesse à l'entrée du domaine de calcul (purement longitudinale) et sur la hauteur du carré est de 100 : l'écoulement est laminaire, périodique et comporte en aval de l'obstacle une allée de tourbillons contrarotatifs et alternés en temps et en espace appelée *allée de Von Kármán*.

2.5.1 Base de données et POD

Les données ont été obtenues par un code compressible aux volumes finis développé par Ivan Mary du Département de Simulation Numérique et Aéroacoustique de l'ONERA (voir [63] pour les détails). Elles ont été calculées pour un nombre de Mach de 10^{-3} : l'écoulement peut donc être considéré comme (quasi-)incompressible. La base de données numérique est constituée de $N = 480$ clichés $\boldsymbol{u}^{\boldsymbol{e}}(t_i)$ du champ des vitesses régulièrement répartis dans le temps sur une période $[0, T]$ de l'échappement tourbillonnaire.

Les conditions aux bords étant stationnaires, le champ $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ choisi est simplement le champ moyen obtenu par une moyenne temporelle arithmétique des clichés :

$$\bar{\boldsymbol{u}}^{\boldsymbol{e}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{u}^{\boldsymbol{e}}(t_i). \tag{2.45}$$

Son rotationnel $\bar{w}^e = \partial_{x_1} \bar{u}^e_{x_2} - \partial_{x_2} \bar{u}^e_{x_1}$ et quelques lignes de courant apparaissent sur la figure 2.2. La POD L^2 ($X = \mathrm{L}^2(\Omega)$) du champ fluctuant $\tilde{\boldsymbol{u}}^{\boldsymbol{e}} = \boldsymbol{u}^{\boldsymbol{e}} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$ a été effectuée par la méthode des clichés (voir les sections 1.5 et 1.5.3). Des isocontours du rotationnel

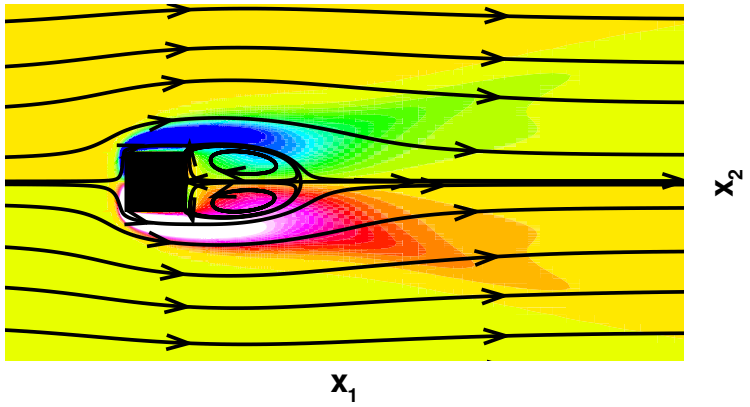


FIG. 2.2 – Carte du rotationnel $\bar{w}^e$ et lignes de courant du champ moyen sur un sous-domaine.

$\tilde{w}^e = \partial_{x_1}\tilde{u}^e_{x_2} - \partial_{x_2}\tilde{u}^e_{x_1}$ du champ fluctuant des vitesses à l'instant $t = T/2$ sont tracés sur la figure 2.3.

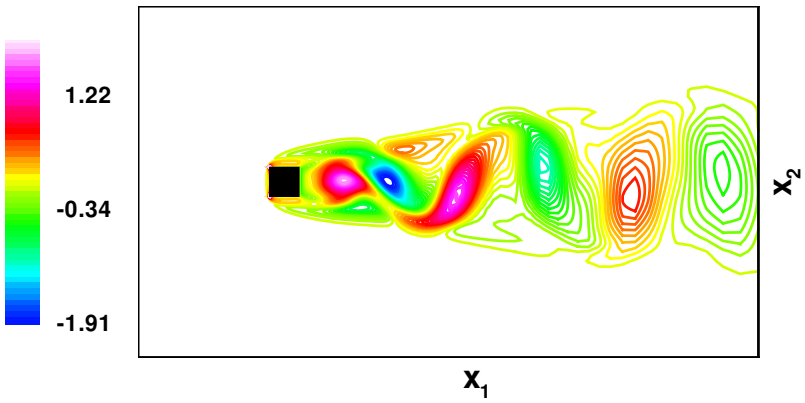


FIG. 2.3 – Isocontours du rotationnel $\tilde{w}^e$ du champ fluctuant à l'instant $t = T/2$ sur la moitié amont du domaine de calcul.

Les six premiers modes POD, $\boldsymbol{\varphi_k}$ pour $1 \leq k \leq 6$, capturent plus de 99.9% de l'énergie cinétique fluctuante $K_N = \sum_{i=1}^{N} \lambda_i$:

i	1	2	3	4	5	6
λ_i/K_N	0.486	0.482	1.214×10^{-2}	1.209×10^{-2}	3.756×10^{-3}	3.744×10^{-3}

Les 16 premières valeurs du spectre POD sont reportées en échelle logarithmique sur la figure 2.4.

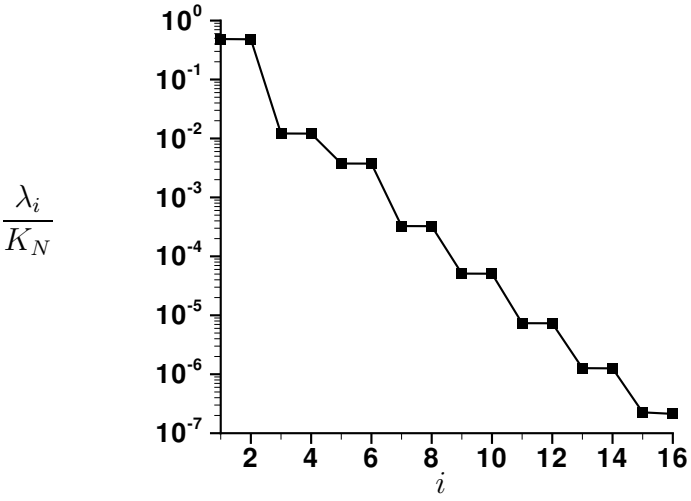


FIG. 2.4 – Les 16 premières valeurs du spectre POD de l'écoulement autour du carré (échelle logarithmique).

Les figures 2.5 et 2.6 représentent des lignes de niveau du rotationnel $\varphi_k^w = \partial_{x_1}(\varphi_k)_{x_2} - \partial_{x_2}(\varphi_k)_{x_1}$ des six premiers modes POD obtenus, ainsi que certaines lignes de courant définies par ces modes. Le spectre POD comme les visualisations des modes POD montrent clairement que les premiers modes peuvent être regroupés par paires. Ceci est cohérent avec la nature de l'allée tourbillonnaire et avec les résultats de nombreuses études, par exemple celles menées par Noack *et al.* [66, 67]. En outre, la topologie des modes POD correspond à celle qui a été observée dans ces travaux pour un obstacle circulaire. Notons que le regroupement naturel des modes par paire peut survenir pour d'autres types d'écoulements, par exemple pour une couche de mélange bidimensionnelle (voir [73]).

2.5.2 Évaluation du modèle POD-Galerkine réduit

Les données ne satisfont pas rigoureusement une condition de la forme (2.25) en sortie mais une condition mixte qui mélange, en quelque sorte, les conditions (2.24) et (2.25). Plus précisément, la vitesse transverse u_{x_2}, la dérivée longitudinale (c'est-à-dire normale) $\partial_{x_1} u_{x_1}$ de la composante longitudinale de la vitesse et la pression dynamique ont été imposées à zéro en sortie durant toute la simulation. On a ainsi

$$u_{x_2} = 0 \quad \text{et} \quad p\,\boldsymbol{n} - \frac{1}{\mathrm{Re}}\,[\nabla u]\;\boldsymbol{n} = -P_s\,\boldsymbol{x_1} - \frac{1}{\mathrm{Re}}\,\partial_{x_1} u_{x_2}\,\boldsymbol{x_2}$$

sur la fontière de sortie où P_s est une pression statique constante (la pression dynamique est $p - P_s$) et où $\boldsymbol{x_1}$ et $\boldsymbol{x_2}$ correspondent respectivement aux directions longitudinale et transverse ($\boldsymbol{n} = \boldsymbol{x_1}$ ou encore $n_{x_1} = 1$ et $n_{x_2} = 0$ sur la frontière de sortie). Bien qu'elle ne soit pas strictement équivalente à la condition (2.25), la formulation variationnelle (2.27)

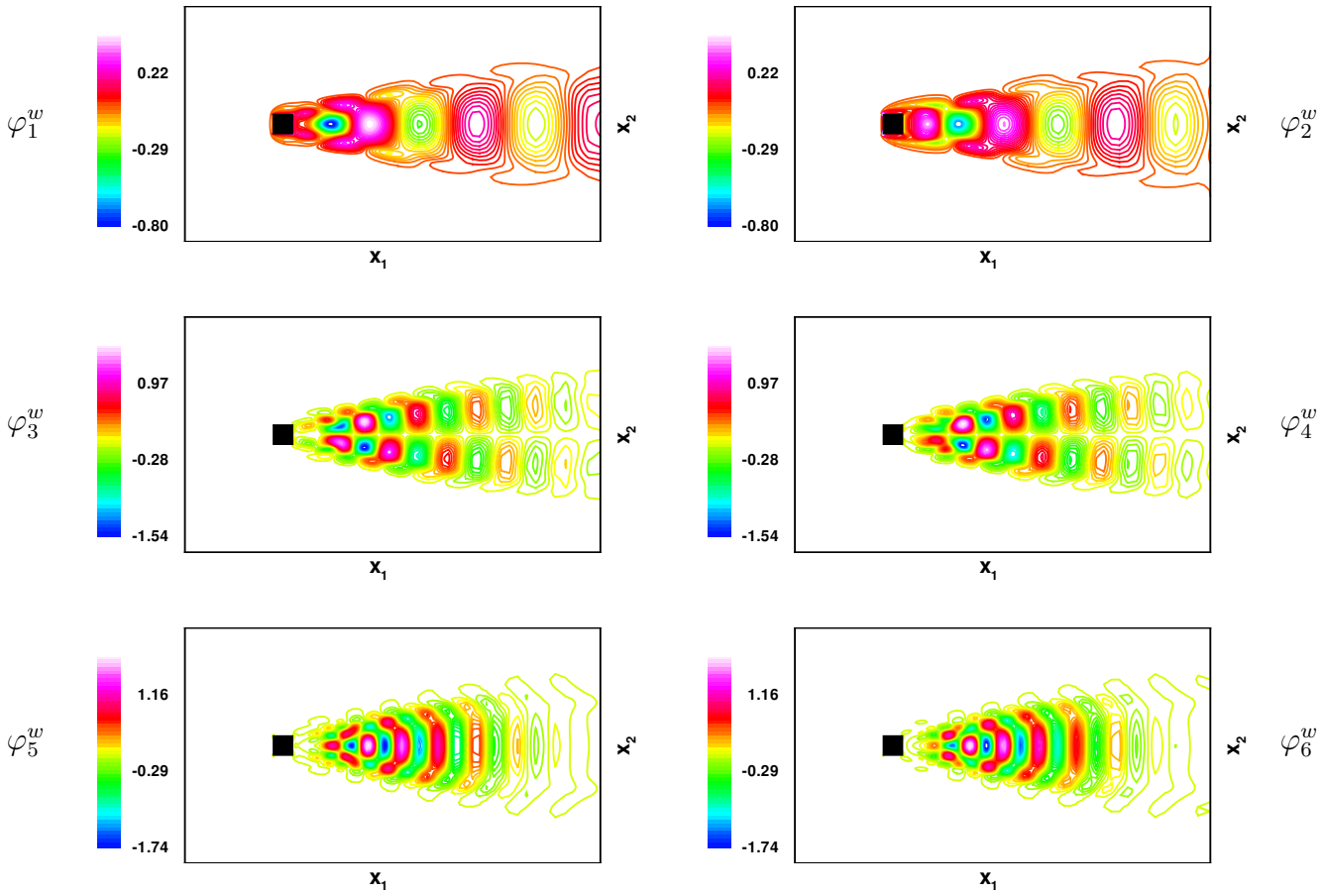


FIG. 2.5 – Isocontours des rotationnels φ_1^w, φ_2^w, φ_3^w, φ_4^w, φ_5^w et φ_6^w des six premiers modes POD sur la moitié amont du domaine de calcul.

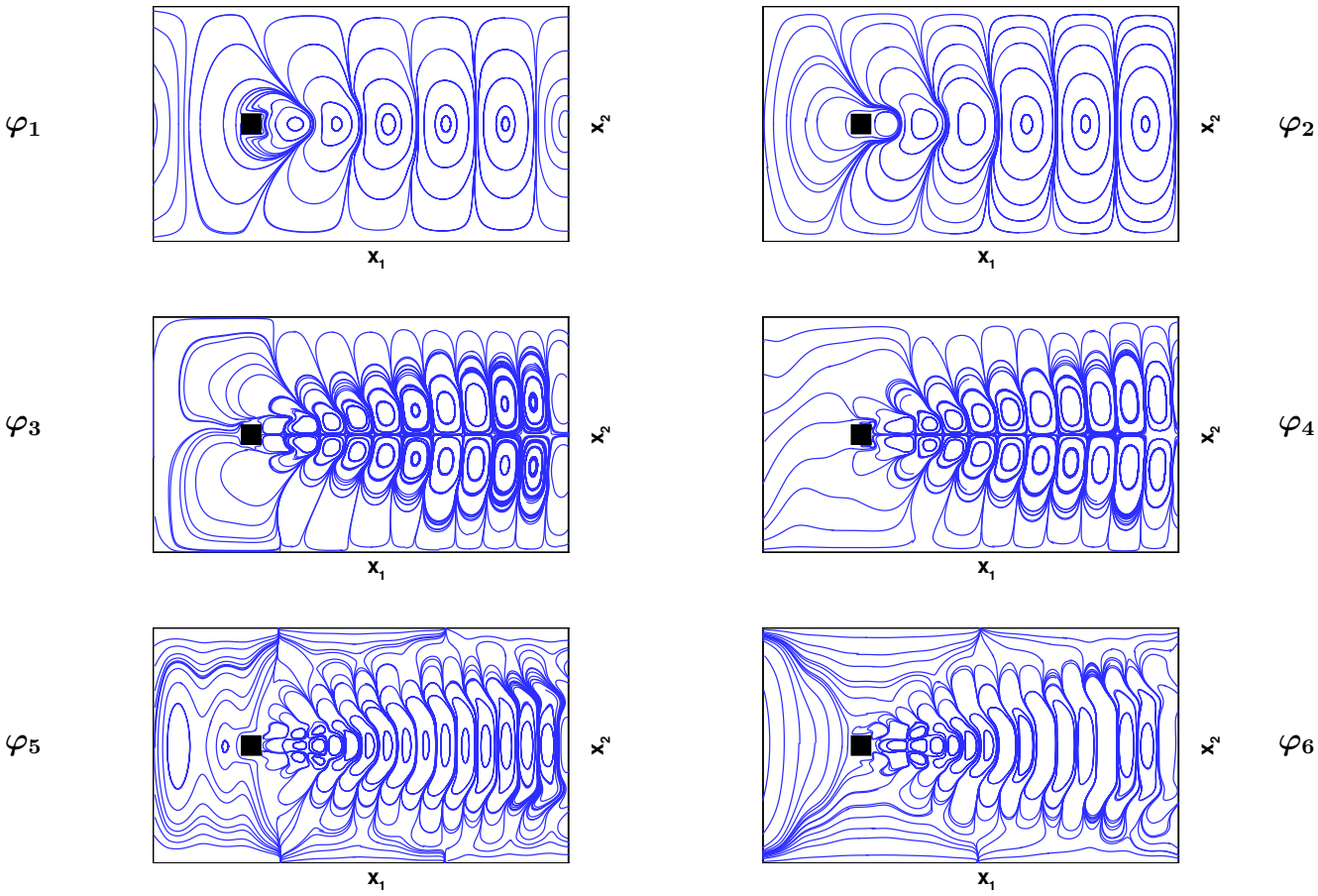


FIG. 2.6 – Lignes de courant des modes φ_1, φ_2, φ_3, φ_4, φ_5 et φ_6 sur la moitié amont du domaine de calcul.

obtenue en posant $\omega = 0$, $\boldsymbol{\beta} = \mathbf{0}$ et $P = P_s$ reste valable pour une fonction test $\boldsymbol{\varphi}$ de divergence nulle et qui satisfait les mêmes conditions aux limites de Dirichlet que $\boldsymbol{u}$, en particulier $\varphi_{x_2} = 0$ en sortie. Le modèle POD-Galerkine obtenu correspond donc encore au modèle (2.35) présenté en page 50 dans le cas $\omega = 0$.

Les coefficients polynômiaux de ce modèle POD-Galerkine réduit ont été calculés grâce à la routine d'intégration spatiale développée pour le calcul POD (voir section 1.5.3) et à un schéma de différentiation spatiale aux différences finies d'ordre deux. Une simulation du modèle a été conduite dans les mêmes conditions que les données utilisées, c'est-à-dire pour $\boldsymbol{\beta} = \mathbf{0}$, $\boldsymbol{h} = \mathbf{0}$ et $\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}^{\boldsymbol{e}}$: le système dynamique est autonome et tous ses coefficients sont constants. Le polynôme correspondant à ce système est noté $f^g = (f_1^g \cdots f_6^g)^T$. Un schéma classique de Runge-Kutta à quatre pas a été utilisé avec un pas de temps $\Delta t = 10^{-4}\, T$.

Les valeurs de $f_1^g(a^e(t))$ et $f_2^g(a^e(t))$ sont comparées à $\dot{a}_1^e(t)$ et $\dot{a}_2^e(t)$ sur la figure 2.7 (les coefficients a_k^e sont les coefficients temporels issus de la POD, leurs dérivées ont été approchées par différences finies). Cette figure représente de plus les histoires de $a_1^g(t)$ et

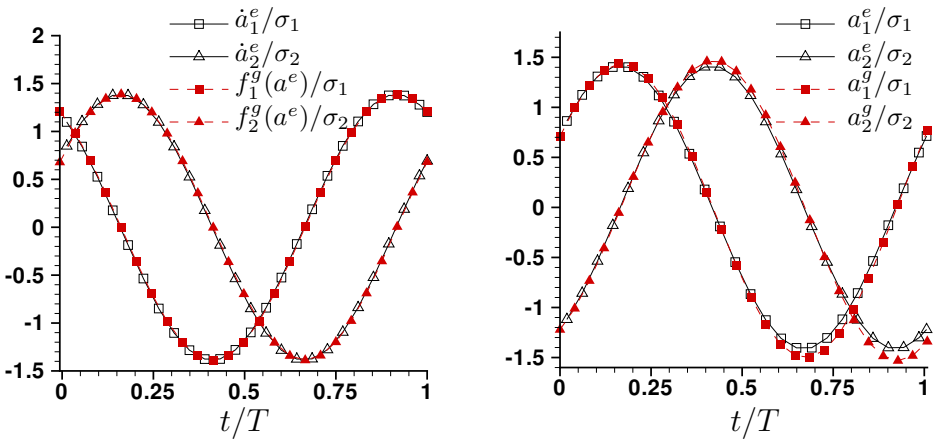


FIG. 2.7 – Comparaison, pour les deux premiers modes POD, de la dynamique (gauche) et de l'histoire (droite) des données et du modèle à six modes.

$a_2^g(t)$ obtenues par simulation du modèle pour $t \in [0, T]$ avec celles de $a_1^e(t)$ et $a_2^e(t)$.

L'évaluation du polynôme f^g montre que le modèle à six modes calculé approche assez précisément la dynamique de l'écoulement. Cependant la simulation du modèle n'est pas pleinement satisfaisante puisqu'il existe un écart perceptible entre le modèle et les données à l'instant final $t = T$, en particulier entre a_2^e et a_2^g.

Si le modèle est simulé sur un intervalle de temps plus long, on observe que la divergence entre les données et le modèle est relativement rapide. La figure 2.8 qui nous permet de comparer le comportement numérique du modèle sur cinq périodes avec la trajectoire périodique de l'écoulement met en évidence cette conclusion.

La différence entre données et modèle a trois causes :

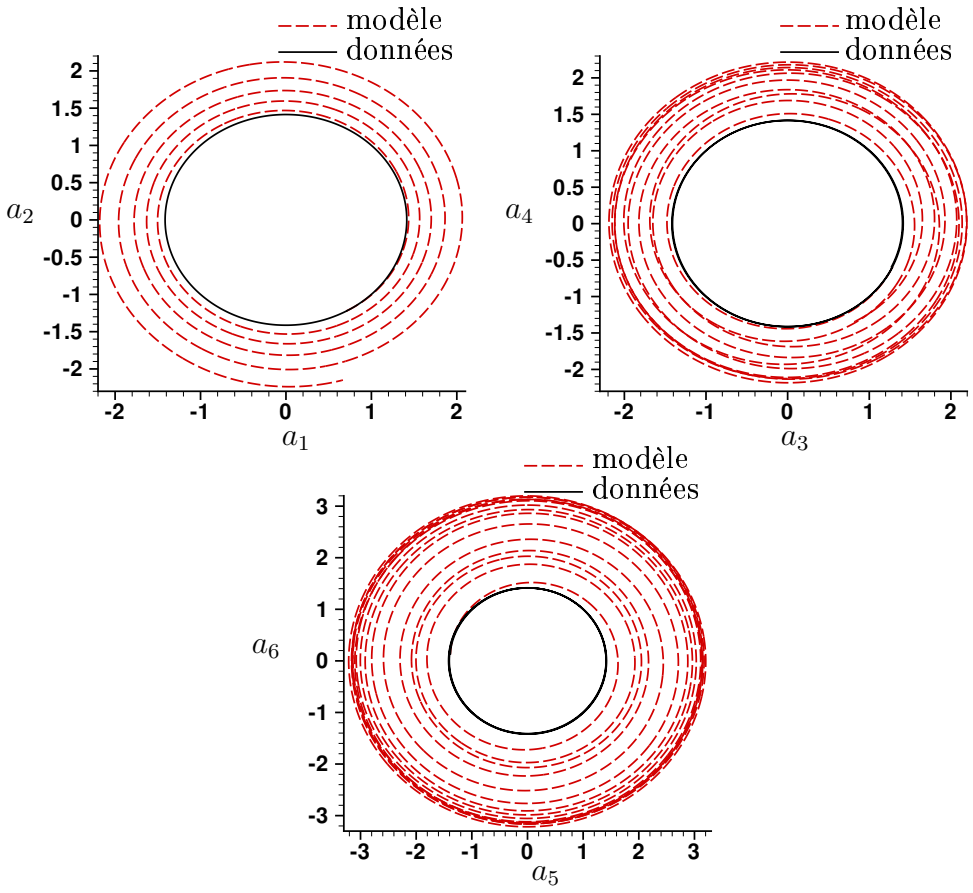


FIG. 2.8 – Trajectoires dans les sous-espaces des phases (a_1, a_2), (a_3, a_4) et (a_5, a_6) des données et du modèle à six modes simulé sur cinq périodes ($t \in [0, 5\,T]$).

- la nature des données qui n'ont pas été calculées à partir de la formulation variationnelle (2.27) des équations de Navier-Stokes utilisée pour construire le modèle,
- la réduction du modèle à six modes POD sur les $N = 480$ qui permettent de reconstruire fidèlement les données,
- et les erreurs numériques.

Comme on l'observe sur la figure 2.8, le modèle réduit voit donc son énergie cinétique croître anormalement : il semblerait qu'un ajout artificiel de viscosité pourrait améliorer le modèle en dissipant ce supplément d'énergie. En pratique dans la littérature, les auteurs cherchent souvent à améliorer leur modèle par de telles viscosités artificielles, en particulier pour les écoulements transitionnels ou turbulents mais aussi pour des écoulements laminaires bidimensionnels : consulter [6] et [87].

Par ailleurs, si la POD avait été calculée à partir de clichés couvrant une fenêtre temporelle de plusieurs périodes (par exemple cinq), les modes POD auraient été les mêmes puisque l'écoulement est périodique : le modèle réduit aurait été identique et divergerait donc des données de manière importante avant l'instant final $t = 5\,T$ comme nous l'avons vu.

Un modèle réduit plus précis peut être obtenu pour ce type d'écoulement laminaire bidimensionnel décollé si les données numériques satisfont plus finement la formulation variationnelle (FVNSI) des équations de Navier-Stokes. En effet, rapelons que les données utilisées ici ont été obtenues par un schéma aux volumes finis compressible. Par exemple, dans le cas d'un écoulement incompressible de Reynolds 100 autour d'un obstacle circulaire, le modèle POD-Galerkin à six modes obtenu par Sirisup *et al.* diverge plus lentement de la trajectoire périodique de l'écoulement que le modèle présenté à la figure 2.7. Précisons au passage que le modèle de Sirisup *et al.* [87] est sensiblement amélioré si vingt modes POD au lieu de six sont utilisés. Enfin, il faut noter que l'utilisation de schémas éléments finis pour les équations de Navier-Stokes incompressibles conduit à la construction de modèles réduits fidèles (si les conditions aux limites sont correctement traitées) pour un écoulement bidimensionnel de nombre de Reynolds relativement faible ($\mathrm{Re} \leq 500$) et un intervalle de temps de longueur raisonnable : voir par exemple [75] ou [7]. Comme ces auteurs, nous avons construit sans difficulté des modèles réduits à moins de dix modes et représentatifs d'un écoulement bidimensionnel de nombre de Reynolds 100 à partir de données éléments finis dans la suite (voir le chapitre 5 consacré au contrôle) pour la fenêtre temporelle $[0, T]$ qui est considérée.

Par cet exemple, nous avons néanmoins mis en évidence certaines limites de la modélisation POD-Galerkine réduite que nous avons cherché à traiter par la suite. Le chapitre 3 est consacré au problème de la modélisation des effets des modes POD tronqués et à leur paramétrisation par des ajouts artificiels de viscosités. Des méthodes numériques de calibration seront proposées dans le chapitre 4 afin de corriger le comportement des modèles dynamiques polynômiaux, et seront notamment testées sur l'écoulement présenté ici.

2.6 Conclusions

La méthode POD-Galerkine permet de définir des systèmes d'EDOs polynômiaux à partir des équations de Navier-Stokes. En particulier, dans le cas incompressible, la méthode de Galerkine est bien maîtrisée du point de vue formel : le système obtenu est quadratique et nous avons montré comment les conditions aux bords de Dirichlet et de flux peuvent être prises en compte explicitement, même si le tenseur des contraintes fluides σ et non le pseudo-tenseur $\tilde{\sigma}$ est utilisé dans la modélisation. Par ailleurs, ce modèle réduit incompressible est physiquement cohérent puisqu'il est basé sur une décomposition optimale du champ des vitesses au sens de l'énergie cinétique moyenne.

En outre, en reprenant l'analyse de Vigo [94], nous avons prouvé que l'énergie du modèle POD-Galerkine réduit ne peut augmenter que par des effets environnementaux dans le cas d'un écoulement incompressible sans frontière libre. Cependant, même dans ce cas particulier, ce résultat théorique ne garantit pas la robustesse du modèle numérique et des problèmes peuvent *a priori* survenir à cause des erreurs numériques et des problèmes structurels.

De plus, l'effet des modes tronqués n'est pas modélisé et les données utilisées pour construire le modèle peuvent ne pas satisfaire finement la dynamique définie par la formulation variationnelle (2.27). Signalons d'ailleurs que même si le code de calcul POD-Galerkine est couplé avec un code de simulation par éléments finis, ce qui est pratique et cohérent, la formulation (2.27) n'est pas utilisée en pratique pour générer les données, car elle nécessiterait l'utilisation d'une base de fonctions de divergence nulle sur les éléments. En conséquence, le modèle peut sensiblement diverger par rapport à l'écoulement, et ce même pour une configuration incompressible, laminaire et bidimensionnelle comme celle de la section 2.5.

Dans le chapitre suivant, nous allons aborder la modélisation visqueuse des modes tronqués. Des méthodes numériques de recalibration des modèles polynômiaux seront ensuite proposées dans le chapitre 4 pour tenter de stabiliser ou d'augmenter la précision des modèles POD-Galerkine réduits.

Chapitre 3

Analyse de la modélisation POD-Galerkine réduite d'un écoulement tridimensionel turbulent

Sommaire

Ce chapitre expose les études qui ont été menées sur un écoulement incompressible "complexe" : tridimensionnel, dont deux directions sont non-homogènes, à haut nombre de Reynolds, décollant au niveau d'une marche descendante et étudié dans son ensemble (la méthode POD-Galerkine n'est pas restreinte à un petit sous-domaine de calcul).

Cet écoulement sort ainsi du cadre qui est habituellement choisi pour mener de telles études. En effet, celles qui portent sur des écoulements en transition ou pleinement turbulents se cantonnent le plus souvent à l'examen d'un "petit" domaine spatial, comme le *minimal channel unit* ([6, 71, 70, 95, 96]) qui comporte deux directions homogènes, ou une couche limite en transition aux abords d'une plaque plane ([79, 78]) à une direction homogène.

En outre, les études qui s'inscrivent dans la résolution d'un problème de contrôle concernent en général des écoulements laminaires bidimensionnels : écoulements contournant un cylindre ([29, 1], $\mathrm{Re} = 100$), franchissant une marche descendante ([75, 74, 39], $\mathrm{Re} = 200$), confinés dans une cavité ([4], $\mathrm{Re} = 200$), ou encore passant au-dessus d'une cavité ([68], $\mathrm{Re} = 1$, ou [82], $\mathrm{Re} = 68.5$).

On notera tout de même les travaux menés à l'INRIA dans le cadre du projet SINUS. L'article [38] nous présente la modélisation POD-Galerkine de deux écoulements compressibles à grand nombre de Reynolds, avec une POD effectuée sur tout le domaine spatial, autour d'un profil d'aile NACA0012 à 20^o d'incidence, et autour d'un obstacle de section carrée ($\mathrm{Re} = 2100$ (laminaire) et $\mathrm{Re} = 22\,000$ respectivement). Bien que relativement complexes, ces écoulements sont bidimensionnels[1] et les conditions d'entrée stationnaires : il n'y a pas de turbulence tridimensionnelle développée.

Par la nature des écoulements choisis et le choix des conditions aux bords qui sont imposées, tous ces travaux conduisent à la construction de modèles de très petite dimension (en général seulement une petite dizaine de modes POD suffisent), ce qui, comme nous le verrons, n'est pas réaliste pour notre écoulement. Celui-ci a été choisi afin d'éprouver la robustesse de la réduction de modèle par une méthode POD-Galerkine, mais aussi pour mener des investigations sur le problème de la "modélisation des petites échelles", pour reprendre la terminologie employée en SGE (Simulation des Grandes Échelles, consulter [84] pour une présentation générale).

En effet, les modes POD tronqués correspondent en pratique à de petites structures spatiales et il semble particulièrement intéressant, sinon indispensable, de modifier les modèles POD-Galerkine réduits des écoulements dont le nombre de Reynolds est relativement important afin de tenir compte des effets des petites échelles qui n'apparaissent pas dans les modes POD conservés et d'aboutir à un modèle précis qui recouvre l'essentiel de la physique de l'écoulement, au minimum l'essentiel de la physique des données utilisées pour construire la base POD et le modèle réduit (cette problématique est discutée à la section 3.4.2).

[1] Notre écoulement est périodique dans la direction transverse, cependant un calcul 3D avec une direction périodique n'est pas équivalent à un calcul 2D.

Les données qui ont été utilisées constituent l'aboutissement du travail de thèse mené par Emmanuel Montreuil à l'ONERA ([65, 51]). Leurs principales caractéristiques sont décrites dans la prochaine section. Les caractéristiques de la décomposition orthogonale propre et les structures cohérentes qui ont ainsi pu être déduites seront présentées dans la section 3.2. Dans toute la suite, le terme de structure cohérente se rapportera aux structures les plus énergétiques de l'écoulement, c'est-à-dire aux premiers modes POD. La section 3.3 présente une évaluation des modèles POD-Galerkine réduits définis par la formulation "classique" et la formulation vitesse-tourbillon (équations (2.35) et (2.41)) et introduit le problème de la modélisation des petites échelles.

La section 3.5 expose une étude des interactions entre les modes de la base POD à partir de l'observation des transferts d'énergie cinétique. De plus, partant d'un système POD-Galerkine fiable de taille M suffisamment grande, des estimations de type pseudo-viscosité, quantifiant les interactions entre les l premiers modes qui sont conservés et les $M - l$ derniers qui sont tronqués, sont présentées pour différentes coupures l. Le travail développé dans cette section, à l'origine de la publication [16], nous éclaire sur le problème de la prise en compte des petites échelles par des modèles de type viscosité artificielle dans le cadre d'une modélisation POD.

3.1 Les données numériques

Nous avons appliqué une méthode POD-Galerkine à $N = 1\,000$ clichés numériques 3D d'un champ $\boldsymbol{u}^e$ des vitesses obtenu par la SGE d'un fluide s'écoulant dans un canal qui comporte une marche descendante (voir [65, 51]). La figure 3.1 décrit la configuration de

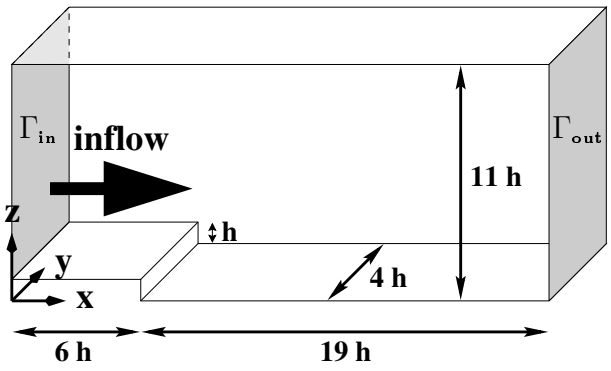


FIG. 3.1 – Géométrie de la marche descendante.

l'écoulement. Dans ce chapitre, les directions spatiales du repère 3D $(\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3})$ seront notées (x, y, z) : x, y et z sont respectivement les directions longitudinale, transverse et verticale de l'écoulement.

La condition de non-glissement ($\boldsymbol{u} = \boldsymbol{0}$) est appliquée au niveau des parois. En entrée, une condition de Dirichlet non homogène est appliquée : les valeurs du champ $\boldsymbol{u}$ sur la

surface Γ_{in} sont imposées par la simulation d'un canal-plan turbulent. Des conditions périodiques dans la direction transverse y sont appliquées. Enfin, les champs de vitesses $\boldsymbol{u}$ et de pression p satisfont la condition (2.26) au niveau de la sortie Γ_{out}.

Le nombre de Reynolds, basé sur la vitesse moyenne U_{in} sur Γ_{in} (constante au cours du temps) et sur la hauteur $10\,h$ du canal en amont de la marche, est de 66 100. Celui basé sur la vitesse en entrée en milieu de canal U_c et sur la hauteur h de la marche vaut 7 432.

Une grille cartésienne, raffinée au niveau de la marche, de $186 \times 40 \times 93$ points dans les directions respectives x, y et z, a été utilisée pour mailler le domaine de calcul $(25h, 4h, 11h)$. Les clichés de $\boldsymbol{u}$ dont nous nous sommes servis, calculés sur cette grille, recouvrent une durée $T = \frac{37.5h}{U_c} = \frac{50h}{U_{\text{in}}}$ (T^{-1} est inférieur aux basses fréquences du battement du bulbe de recirculation de la marche).

3.2 Résultats des calculs POD du champ des vitesses

3.2.1 Étude spectrale de l'influence du choix des clichés

Plusieurs calculs POD ont été menés sur les données décrites dans la section 3.1. Les résultats présentés dans cette section concernent la POD L^2 du champ fluctuant $\tilde{\boldsymbol{u}}^{\boldsymbol{e}} = \boldsymbol{u}^{\boldsymbol{e}} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$, où $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ est un champ moyen obtenu par une moyenne arithmétique des clichés (voir (2.45)). Nous avons fait varier le nombre de clichés sur lesquels a été conduite la POD, afin de voir comment, en terme de spectre, les résultats évoluaient.

Deux séries de spectres ont été obtenues

- en faisant varier la fréquence d'échantillonnage des clichés (on prend tous les clichés, puis un sur cinq, un sur dix et enfin un sur vingt) ;
- en gardant la fréquence d'échantillonnage maximum, mais en prenant de moins en moins de clichés successifs, ce qui revient à considérer l'écoulement sur une durée de plus en plus petite.

Nous avons ainsi obtenu les figures 3.3 et 3.2 où sont présentés les différents spectres POD normalisés par leur énergie cinétique moyenne, c'est-à-dire, pour chaque POD d'énergie cinétique moyenne $K_N = \sum_{i=1}^{N} \lambda_i$ menée sur N clichés, les valeurs $\frac{\lambda_i}{K_N}$ en fonction de l'indice i des modes.

Le spectre normalisé varie beaucoup avec la durée que recouvrent les clichés que l'on utilise. En effet, les spectres de la figure 3.2, déduits pour une même fréquence d'échantilonnage mais de quantités de clichés variables, sont très différents. On observe une décroissance très importante du spectre pour un nombre $N = 250$ de clichés consécutifs, ce qui traduit une variété relativement faible des structures les plus énergétiques de l'écoulement. Quand on augmente le nombre N de clichés consécutifs exploités, donc la durée du phénomène physique que décrivent les clichés, le spectre a tendance à s'aplatir de manière régulière ce qui témoigne, via cette meilleure distribution de l'énergie cinétique au sein des modes

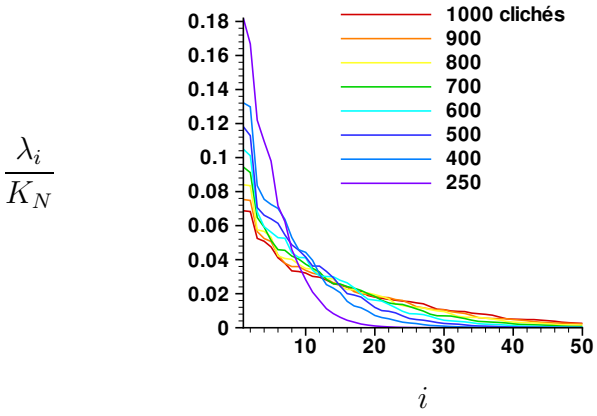


FIG. 3.2 – Spectres POD déduits des N premiers clichés de la base pour N variant de 250 à 1 000.

POD, de la plus grande richesse des structures cohérentes (c'est-à-dire des structures les plus énergétiques ou encore des modes POD prépondérants). Ce résultat est tout à fait normal puisque les spectres correspondent à un phénomène physique complexe évoluant sur des durées de plus en plus importantes.

Cependant, malgré la grande variété des structures cohérentes et leur enrichissement continu, on peut limiter le calcul POD à un nombre de clichés restreint (une centaine par exemple) et obtenir une base POD fiable.

En effet, la figure 3.3 nous montre que le spectre n'évolue que très peu lorsqu'on diminue la fréquence d'échantillonnage, c'est-à-dire lorsqu'on augmente l'intervalle de temps qui sépare les clichés dont on calcule la POD ($F = \frac{N-1}{T}$ est la fréquence de la base complète). En fait, on n'a pas besoin de $N = 1\,000$ clichés pour obtenir de manière précise les principales structures de l'écoulement. Ceci se vérifie d'ailleurs en pratique de manière assez générale : peu de clichés, si ils sont répartis de manière homogène dans le temps, permettent d'extraire une information pertinente (voir [94, p.93]). Il restait cependant intéressant ici d'évaluer la dépendance des modes POD avec la fréquence des clichés, même si cela a été réalisé de manière indirecte via les spectres pour des raisons de coût informatique, puisque, comme il a été souligné en introduction de ce chapitre, l'écoulement étudié est assez complexe (pleinement turbulent et étudié dans tout le domaine de calcul initial).

Puisqu'elle a été calculée pour réaliser cette étude spectrale, on a, par la suite, exploité la POD déduite de la totalité de la base de données. Cette POD est présentée dans la section suivante.

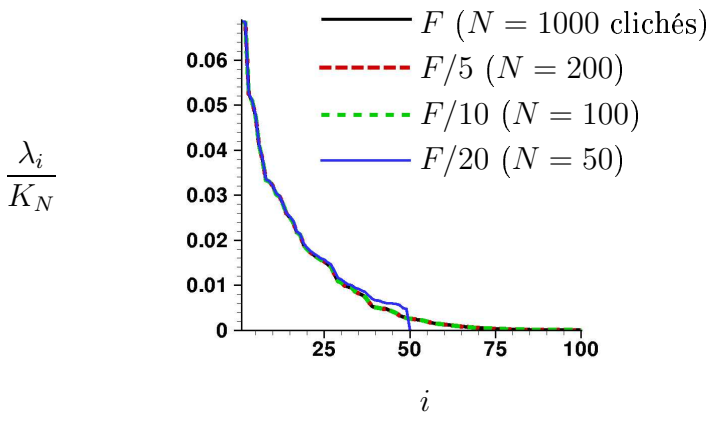


FIG. 3.3 – Évolution du spectre POD pour différentes fréquences d'échantillonnage des clichés.

3.2.2 POD du champ fluctuant

Cette section donne les principales caractéristiques de la POD issue de l'ensemble des clichés du champ fluctuant $\tilde{\boldsymbol{u}}^{\boldsymbol{e}}$ ($N = 1\,000$).

On vérifiera l'efficacité de la décomposition grâce au spectre obtenu, mais aussi en norme L^2. Les modes POD correspondants, que l'on peut assimiler aux structures cohérentes de l'écoulement, ainsi que leur évolution temporelle relative au sein du champ fluctuant total, seront ensuite présentés.

Spectre et efficacité de la décomposition

La figure 3.4 présente les cent premières valeurs du spectre POD, mais aussi le spectre complet en échelle logarithmique. On observe que le spectre décroît rapidement même en échelle logarithmique. Cette décroissance est particulièrement importante pour les premiers modes POD puis elle diminue (enveloppe du spectre concave). Afin de mieux quantifier l'efficacité de la POD, nous définissons l'énergie cinétique moyenne normalisée capturée par les k premiers modes POD :

$$K(k) = \frac{1}{K_N} \sum_{i=1}^{k} \lambda_i = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i}.$$

La partie gauche de la figure 3.5 visualise $K(k)$ et sa partie droite $1 - K(k)$ en échelle logarithmique. Il apparaît clairement que (relativement) peu de modes suffisent à capturer la majeure partie de l'énergie cinétique moyenne K_N : la POD du champ fluctuant $\tilde{\boldsymbol{u}}^{\boldsymbol{e}}$ est efficace. En particulier, pour un critère c fixé à 99.9%, $M = 86$ est le plus petit entier k tel que $K(k) > c$, c'est-à-dire $1 - K(k) < 10^{-3}$. Cela signifie que 86 modes suffisent à bien représenter l'écoulement au sens de 99.9% de l'énergie cinétique fluctuante moyenne.

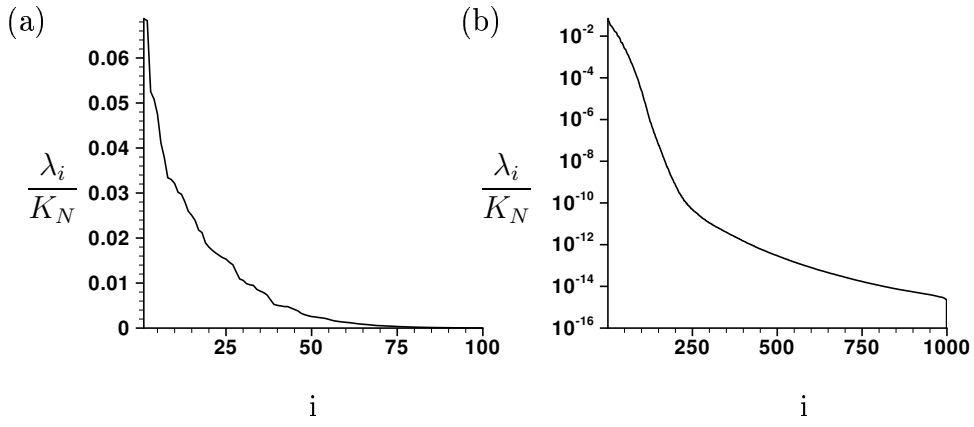


FIG. 3.4 – Spectre POD : (a) les 100 premières valeurs ; (b) Spectre complet en échelle logarithmique.

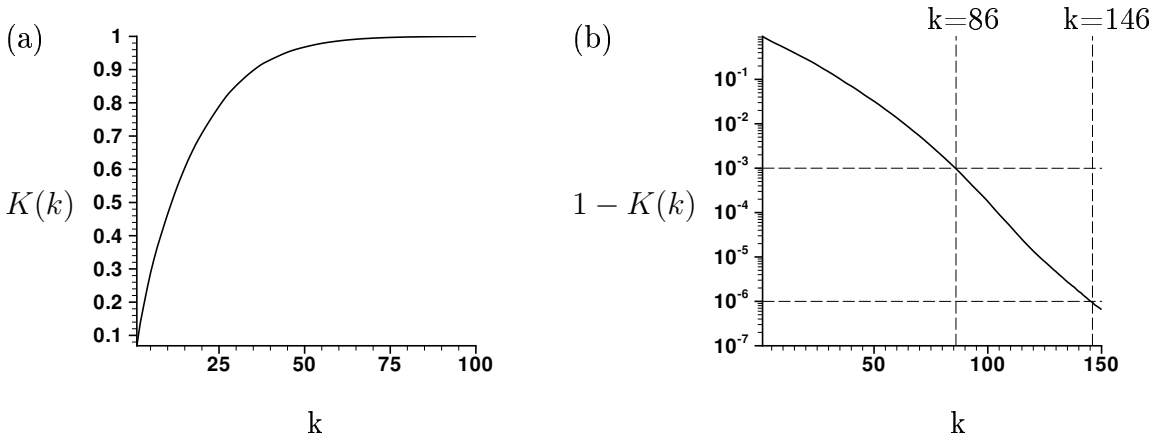


FIG. 3.5 – (a) Énergie cinétique moyenne normalisée $K(k)$ capturée ; (b) énergie $1-K(k)$ (échelle logarithmique) non capturée par les k premiers modes.

Remarquons que c correspond à c_1 dans le critère $\mathcal{C}(k)$ que nous avions donné en page 28. Dans la suite de notre travail, nous n'avons conservé que ces $M = 86$ modes POD prédominants.

Notons que la POD d'un écoulement similaire est présentée dans [41] pour un nombre de Reynolds environ deux fois plus faible (il s'agit d'une analyse physique de l'écoulement qui repose uniquement sur des résultats POD et non sur la construction d'un modèle POD-Galerkine réduit). Contrairement à ce que l'on obtient ici, une faible concentration de la distribution d'énergie cinétique au sein de la base POD y est observée.

Trois faits permettent d'expliquer cette différence. Premièrement, la base de données de [41] a été calculée par Simulation Numérique Directe sur un maillage plus fin et contient donc une plus grande variété de structures spatiales que notre base de données SGE.

Deuxièmement, Jürgens *et al.* [41] combinent la POD avec une décomposition de Fourier, ce qui n'est pas strictement équivalent à un calcul POD pur (qui lui est optimal), même si un spectre proche du spectre POD est généralement attendu (lire le chapitre 1 et en particulier la section 1.3.2).

Troisièmement, l'intervalle de temps $[0, T]$ sur lequel l'écoulement est analysé est douze fois plus long dans [41], si on compare des durées adimensionnées par des unités de temps définies par le rapport entre la vitesse longitudinale maximale dans un plan transverse en amont de la marche (en entrée donc à $6\,h$ de la marche pour notre exemple et à $0.07\,h$ pour l'écoulement de Jürgens *et al.*) et la hauteur h de la marche. Or, la base POD s'enrichit considérablement quand ce type d'écoulement est considéré sur une fenêtre temporelle plus large, comme nous l'avons constaté lors de l'analyse de la figure 3.2.

Nous savons que la POD est la meilleure décomposition au sens de l'énergie cinétique moyenne, ou encore au sens de $\left\langle \|.\|_{\mathrm{L}^2(\Omega)}{}^2 \right\rangle$. Ainsi la base générée par les modes POD est optimale en moyenne sur l'intervalle de temps parcouru par les données, mais on ne sait pas *a priori* si elle est réellement efficace ponctuellement à tout instant (c'est-à-dire pour chaque cliché considéré de manière individuelle).

On a donc calculé, pour chaque cliché, l'erreur e_r relative en norme L^2 entre le champ fluctuant des vitesses et sa projection sur l'espace généré par les $M = 86$ premiers modes POD (figure 3.6). L'erreur est relativement petite, mais n'est pas insignifiante : elle se situe la plupart du temps dans un voisinage de 3% et atteint parfois 9%. La relation $1 - K(M) < 1 - c$ permettait de prédire une valeur moyenne de e_r^2 proche de $1 - c = 0.1\%$, ce qui correspond bien à $e_r \approx 3\%$. On comprend bien que pour avoir une erreur relative e_r de l'ordre de 10^{-3}, il aurait fallu choisir $1 - c$ de l'ordre 10^{-6}, ce qui nous aurait obligé à conserver au minimum 146 modes POD pour que $1 - K(M) < 1 - c$ reste vraie (voir la figure 3.5).

Remarquons enfin que la norme $\|.\|_{\mathrm{L}^2(\Omega)}$ nous permet de définir une erreur globale, définie par rapport à tout le domaine spatial Ω, et cache d'éventuels problèmes locaux. On pourrait mener un calcul d'erreur en norme infinie pour approfondir ce point.

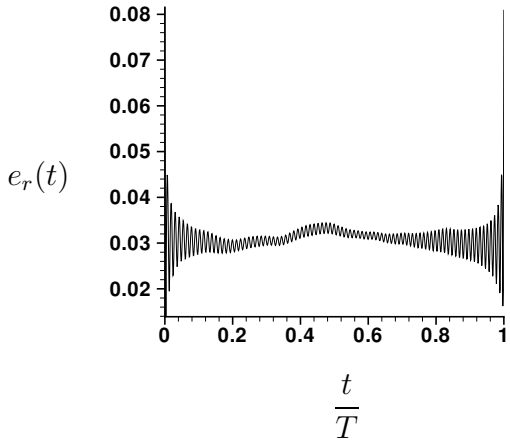


FIG. 3.6 – Erreur relative $e_r(t) = \dfrac{\left\| \tilde{\boldsymbol{u}}^{\boldsymbol{e}}(t) - \sum_{k=1}^{M} \left(\tilde{\boldsymbol{u}}^{\boldsymbol{e}}(t), \boldsymbol{\varphi_k} \right)_{\mathrm{L}^2(\Omega)^d} \boldsymbol{\varphi_k} \right\|_{\mathrm{L}^2(\Omega)^d}}{\| \tilde{\boldsymbol{u}}^{\boldsymbol{e}}(t) \|_{\mathrm{L}^2(\Omega)^d}}$ $(M = 86)$.

Les modes POD

Nous présentons tout d'abord les visualisations 3D de quelques modes POD. Les structures tridimensionnelles que l'on distingue sur les figures 3.7 et 3.8 sont des isocontours du critère Q de Hunt *et al.* (voir [37]) défini par

$$8\,Q \quad = \sum_{i,j=1}^{3} \left(\frac{\partial u_{x_i}}{\partial x_j} - \frac{\partial u_{x_j}}{\partial x_i} \right)^2 - \sum_{i,j=1}^{3} \left(\frac{\partial u_{x_i}}{\partial x_j} + \frac{\partial u_{x_j}}{\partial x_i} \right)^2 .$$

Le critère Q prend des valeurs positives là où le taux de rotation est supérieur au taux de cisaillement. Ici, l'écoulement est confiné dans un canal : ce critère est alors particulièrement bien adapté pour isoler les tourbillons de la pollution due au cissaillement près des parois. Pour plus de lisibilité, ces contours ont ensuite été colorié par le module de la vitesse (l'échelle de valeurs qui correspond à la palette des couleurs est choisie indépendamment pour chaque mode et pour le champ moyen).

Juste en aval de la marche se créent une vorticité transverse et une zone de recirculation. Puis rapidement, l'écoulement a tendance à réorienter le champ de vorticité dans la direction longitudinale x. Ceci explique la topologie du champ moyen et des tous premiers modes POD, auxquels sont associées les principales "structures cohérentes" de l'écoulement (voir le champ moyen et $\boldsymbol{\varphi_1}$ parmi les champs qui sont visualisés ici) : des tourbillons transverses sont créés au niveau de l'arête, ils restent dans le voisinage de la couche limite avant son recollement (c'est-à-dire dans le voisinage du bulbe) tout en se déformant progressivement dans la direction principale de l'écoulement, et produisent plus loin de grosses structures tourbillonnaires purement longitudinales.

Si on continue de parcourir la base modale POD, on observe que les modes, d'énergie cinétique associée λ_k toujours plus faible, correspondent à des structures de plus en

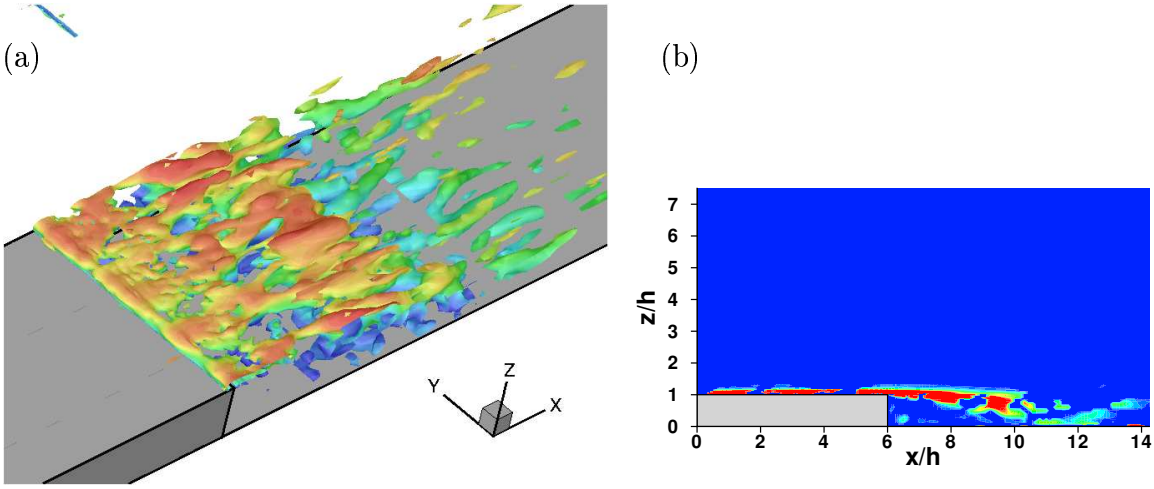


FIG. 3.7 – Champ fluctuant moyen : (a) isosurface $Q = Q_{\bar{u}^e}$; (b) carte du champ $\frac{1+\mathrm{sgn}(Q(\bar{\boldsymbol{u}}^e))}{2}\|\boldsymbol{\nabla}\times\bar{\boldsymbol{u}}^{\boldsymbol{e}}\|_2$ dans le plan $y = 2.1\,h$.

plus petites, ce qui était attendu. Précisons que même si les valeurs de Q choisies pour visualiser les structures diffèrent entre les modes afin d'obtenir des figures lisibles, la taille des structures obtenues, contrairement à leur nombre, évolue très peu avec l'isovaleur : il est donc pertinent de comparer les tailles des structures des modes visualisées sur la figure 3.8. Ces structures révèlent une topologie moyenne de l'écoulement de plus en plus complexe. En particulier, les modes $\boldsymbol{\varphi_{40}}$ et $\boldsymbol{\varphi_{80}}$ nous montrent l'existence de petites structures transverses là où vivent les grosses structures longitudinales (voir le mode $\boldsymbol{\varphi_1}$), qui proviennent des interactions complexes intervenant entre ces grosses structures, ou d'un résidu des structures transverses qui existent en amont. Le mode $\boldsymbol{\varphi_{20}}$ révèle quant à lui des structures intermédiaires complexes à l'orientation indécise.

Distribution des fréquences temporelles dans la base POD

Dans les paragraphes précédents, nous avons présenté le spectre et les modes spatiaux de la POD. Nous allons maintenant nous intéresser à la dernière information déduite de la décomposition orthogonale : les coefficients temporels $a_k^e(t) = \sigma_k^{-1}\,(\,\tilde{\boldsymbol{u}}^{\boldsymbol{e}}(t), \boldsymbol{\varphi_k}\,)$ du champ fluctuant dans la base des modes.

On voit sur la figure 3.9 que ces coefficients ont des allures très régulières. Ils deviennent de plus en plus sinusoïdaux à mesure que l'on parcourt les modes POD : une fréquence semble largement dominer dans chaque signal. Quand on progresse dans la base POD, cette fréquence prépondérante augmente de manière régulière, ce qui est logique puisque les structures spatiales associées aux modes sont de plus en plus petites.

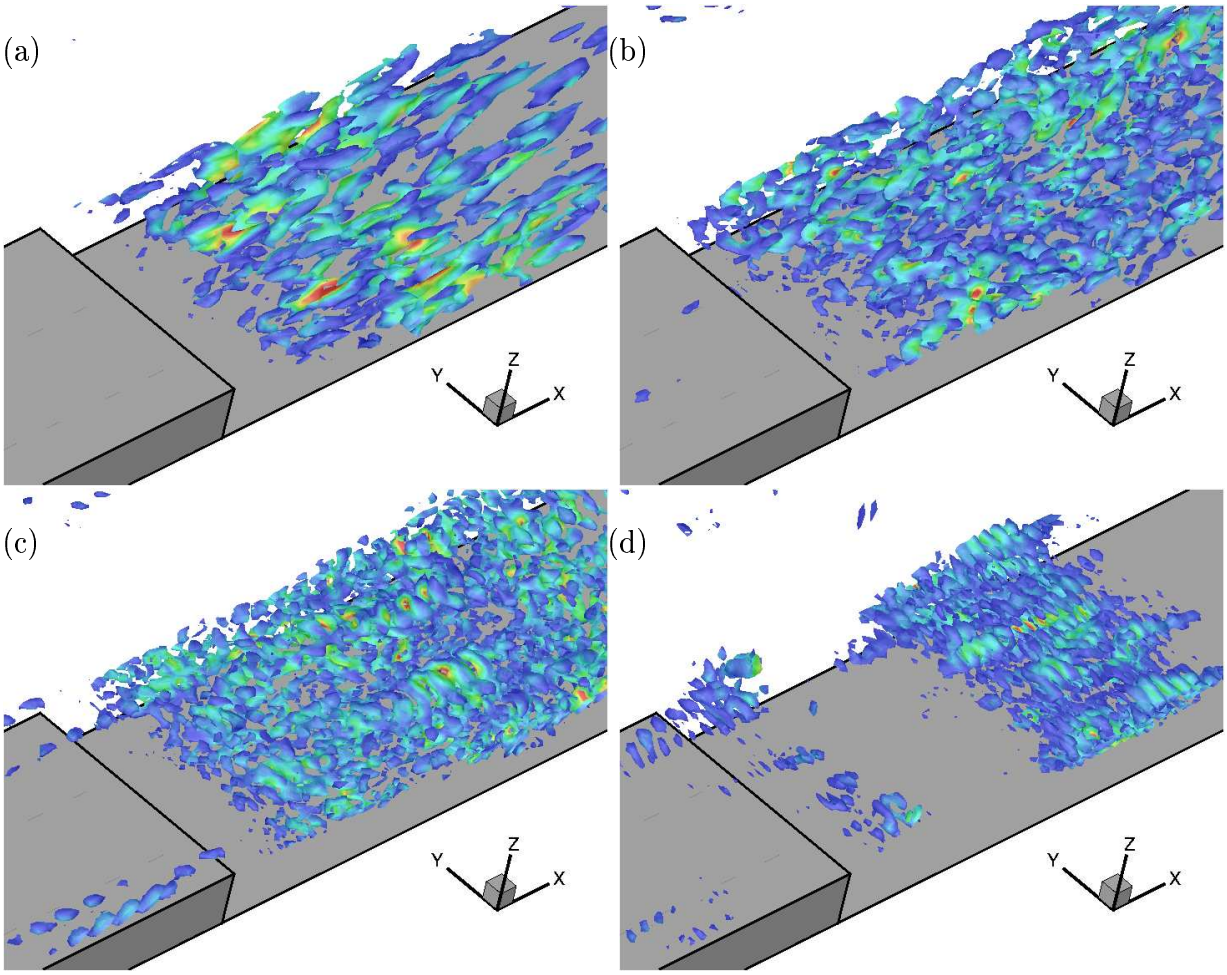


FIG. 3.8 – Q-isosurfaces : (a) $\boldsymbol{\varphi_1}$ avec $Q = 3\,Q_{\bar{u}^e}$; (b) $\boldsymbol{\varphi_{20}}$ avec $Q = 5\,Q_{\bar{u}^e}$; (c) $\boldsymbol{\varphi_{40}}$ avec $Q = 10\,Q_{\bar{u}^e}$; (d) $\boldsymbol{\varphi_{80}}$ avec $Q = 20\,Q_{\bar{u}^e}$.

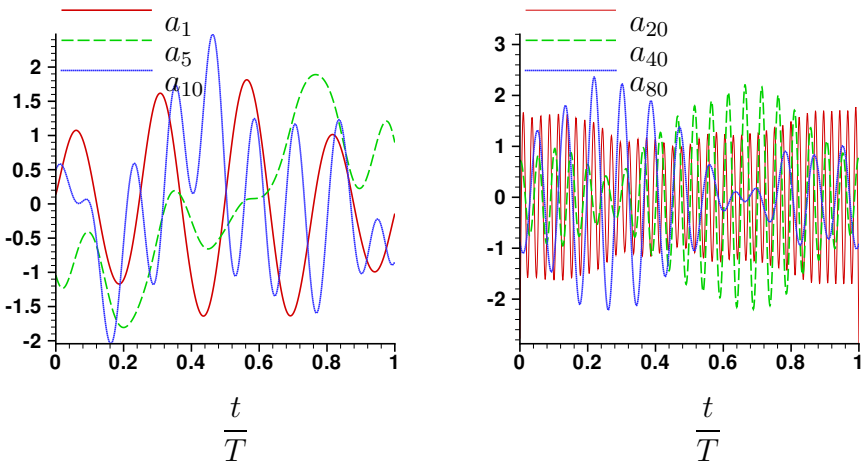


FIG. 3.9 – Coefficients a_1^e, a_5^e, a_{10}^e, a_{20}^e, a_{40}^e et a_{80}^e.

Des calculs de transformée de Fourier ont été réalisés sur la totalité des échantillons des coefficients a_i par DFT (Discrete Fourier Transform), ceci afin de mieux visualiser cette répartition spectrale typique d'une telle POD. La fonction d'apodisation retenue pour préparer les signaux est la fenêtre de Welch définie par

$$g_W(t) = 1 - \left(\frac{t - \frac{1}{2}T}{\frac{1}{2}T} \right)^2 ;$$

aucune technique complémentaire (par exemple un filtrage) n'a été utilisée pour obtenir les spectres.

La figure 3.10 présente les modules $|\hat{a}_i^e(\omega)|$ des DFTs des $a_i^e(t)$ obtenues[2]. Ceux-ci ne présentent qu'un seul pic, bien marqué, à l'exception du spectre de a_{10}^e où deux pics se dessinent : une fréquence dominante assez nette caractérise chaque coefficient a_i^e à l'exception de certains des tous premiers modes où parfois deux fréquences se distinguent.

Pour finir, nous présentons sur la figure 3.11 le graphe des fréquences temporelles dominantes de la POD : pour chaque mode d'indice $i \in [\![1, M]\!]$, on détermine la fréquence ω_i associée à la valeur maximale du module de la DFT de a_i^e. Hormis celles des tous premiers modes (pour lesquels le spectre présente souvent deux pics), les fréquences temporelles dominantes augmentent linéairement quand on parcourt la base ordonnées des modes POD (la pente obtenue par régression linéaire pour les modes d'indice supérieur à 25 est proche de 0.49).

[2]Le pas d'échantillonnage étant de $10^{-3}T$, la bande de fréquence des spectres "complets" est de $\frac{1}{2\Delta t} = \frac{500}{T}$.

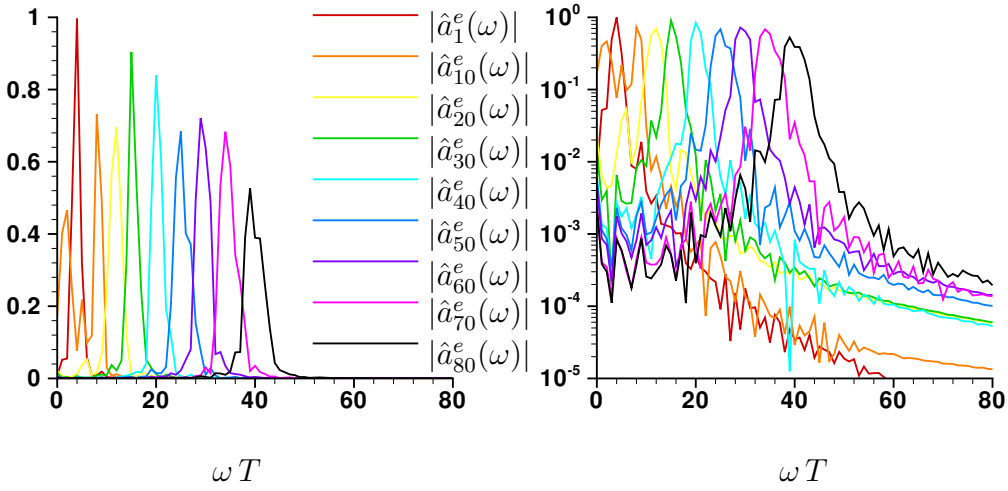


FIG. 3.10 – Module des DFTs de certains coefficients POD temporels en échelle linéaire et logarithmique.

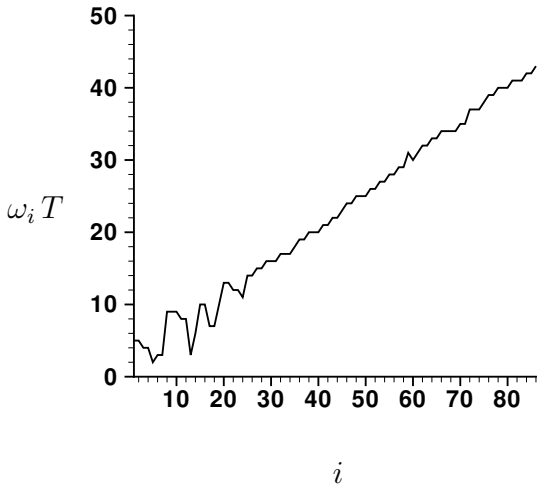


FIG. 3.11 – Fréquences temporelles dominantes associées aux modes POD.

3.3 Calcul et évaluation des modèles POD-Galerkine réduits

Nous allons maintenant utiliser les modes POD obtenus à la section précédente pour construire un modèle POD-Galerkine réduit.

3.3.1 Définitions des modèles réduits

Le modèle POD-Galerkine proposé à la page 50 prend explicitement en compte les conditions aux limites de Dirichlet grâce à $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ (et $\bar{\boldsymbol{u}}$). Cependant, il n'est pas possible ici de calculer un champ $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ qui vérifie (2.33) pour un coût informatique faible à cause de la condition d'entrée instationnaire turbulente de Dirichlet. Néanmoins, les vitesses prescrites en entrée sont quasi-stationnaires dans la mesure où les fluctuations autour de la valeur moyenne sont très faibles :

$$(\boldsymbol{u}^{\boldsymbol{e}} - \bar{\boldsymbol{u}}^{\boldsymbol{e}})_{|\Gamma_{\text{in}}} = \tilde{\boldsymbol{u}}^{\boldsymbol{e}}{}_{|\Gamma_{\text{in}}} \approx \mathbf{0}$$

en définissant $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ comme le champ obtenu par moyenne arithmétique des clichés (équation (2.45)). En conséquence, les modes POD calculés, qui forment une combinaison linéaire des clichés de $\tilde{\boldsymbol{u}}^{\boldsymbol{e}}$, prennent des valeurs très faibles sur le bord Γ_{in} : il est donc envisageable de définir un modèle POD-Galerkine réduit approché en reprenant le modèle de la page 50 qui suppose que les modes POD satisfont des conditions de Dirichlet homogènes sur Γ_D donc sur Γ_{in}.

Les équations de Navier-Stokes adimensionnées donnent, en tenant compte des conditions aux bords et pour une fonction test $\boldsymbol{\varphi}$ qui satisfait

$$\boldsymbol{\nabla}.\boldsymbol{\varphi} = 0 \quad \text{et} \quad \boldsymbol{\varphi}_{|(\Gamma_D \setminus \Gamma_{\text{in}})} = \mathbf{0},$$

la formulation variationnelle suivante

$$\frac{d}{dt} \left(\boldsymbol{u}(t), \boldsymbol{\varphi} \right)_{\mathrm{L}^2(\Omega)^d} + \mathcal{C}(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{\varphi}) + \frac{1}{\mathrm{Re}} \mathcal{A}(\boldsymbol{u}, \boldsymbol{\varphi}) + T_{\Gamma_{\text{in}}}(\boldsymbol{\varphi}) = 0$$

où

$$T_{\Gamma_{\text{in}}}(\boldsymbol{\varphi}) = \left(p\,\boldsymbol{n} - \frac{1}{\mathrm{Re}} \left[\nabla u\right] \boldsymbol{n}, \boldsymbol{\varphi} \right)_{\mathrm{L}^2(\Gamma_{\text{in}})^d}.$$

Ainsi, utiliser le modèle de la page 50 revient à négliger le terme de bord $T_{\Gamma_{\text{in}}}$ pour $\boldsymbol{\varphi} = \boldsymbol{\varphi_k}$, $1 \leq k \leq M$, qui s'annule si $\boldsymbol{\varphi}_{|\Gamma_{\text{in}}} = \mathbf{0}$ exactement.

Nous avons calculé ce modèle, obtenu en négligeant $T_{\Gamma_{\text{in}}}(\boldsymbol{\varphi_k})$ pour tout $k \in [\![1, M]\!]$, pour $M = 86$. En outre, la prise en compte explicite de la condition de Dirichlet d'entrée n'étant pas pratique, nous avons construit le modèle de Rempfer basé sur la formulation vitesse-tourbillon des équations de Navier-Stokes : voir le système (2.41) en page 53.

Dans la suite, le modèle "classique" sera exploité dans les conditions des données SGE, c'est-à-dire pour $\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}^{\boldsymbol{e}}$, $\omega = 0$ et $\boldsymbol{\beta} = \mathbf{0}$. Il est autonome et on le note $\dot{a}(t) = f^g(a(t))$:

dans la suite du chapitre, f^g désignera ainsi le polynôme à coefficients constants associé. Le polynôme du modèle de Rempfer sera noté f^w.

Remarque. Les modèles POD-Galerkine ont été construits sans tenir compte de la viscosité sous-maille de la SGE en connaissance de cause : voir la remarque de la page 93.

3.3.2 Calcul des modèles

Le calcul numérique des modèles POD-Galerkine nécessite des opérations de différentiation et d'intégration spatiale. Comme lors de la modélisation de l'écoulement de la section 2.5, les premières sont réalisées par un schéma aux différences finies d'ordre deux adapté à la grille cartésienne inhomogène utilisée et les secondes en appelant les routines développées pour le calcul POD (voir la section 1.5.3).

En outre, la construction du modèle de Rempfer nécessite la résolution d'un système linéaire associé à la matrice $\tilde{B} = B\,\Sigma$: voir en page 53. Remarquons que la matrice B est symétrique contrairement à $\tilde{B}$ et que le conditionnement de ces matrices pourrait être très différent. En effet, pour $\tilde{B}$ quelconque, le meilleur encadrement du conditionnement $\mathcal{K}(B)$ de B donné par la relation $B = \tilde{B}\,\Sigma^{-1}$ est

$$\frac{1}{\sigma_1}\mathcal{K}(\tilde{B}) \leq \mathcal{K}(B) \leq \frac{1}{\sigma_M}\mathcal{K}(\tilde{B}).$$

Il pourrait donc être beaucoup plus intéressant de résoudre le système linéaire associé à B plutôt que résoudre directement celui associé à $\tilde{B}$. Néanmoins, ces matrices sont petites en pratique (de dimension M) et sur notre exemple les deux matrices sont toutes deux bien conditionnées : $\mathcal{K}(\tilde{B}) = 4.43$ et $\mathcal{K}(B) = 6.39$.

3.3.3 Évaluation efficace des polynômes associés

L'utilisation des modèles POD-Galerkine implique un grand nombre d'évaluations des polynômes correspondants. Le coût d'évaluation de ces polynômes augmente rapidement avec la dimension M du modèle réduit. Or, pour un écoulement dont le nombre de Reynolds est important comme le nôtre, une large gamme de structures cohérentes joue un rôle important et nous sommes amenés à construire un modèle dynamique certes réduit mais de dimension M non négligeable (ici $M = 86$).

À chaque équation des systèmes d'EDOs POD-Galerkine incompressibles correspond un polynôme à M variables de degré deux constitué de $\frac{(M+2)(M+1)}{2}$ monômes, soit 3 828 sur l'exemple de la marche descendante. L'algorithme naturel consistant à évaluer chaque monôme puis à les sommer peut donc s'avérer relativement coûteux. Notons que le modèle POD-Galerkine compressible de Vigo est de degré trois (voir la section 2.3.2) : on manipule alors $\frac{(M+3)(M+2)(M+1)}{6}$ monômes pour chacune des M équations du modèle.

Les algorithmes basés sur le principe de Horner (voir [45]) permettent de diminuer notablement le nombre de multiplications nécessaires à l'évaluation d'un polynôme. En

effet, pour évaluer un polynôme scalaire de degré deux à M variables $a_1, \cdots, a_M$ écrit **sans redondance** des monômes quadratiques

$$C^0 + \sum_{i=1}^{M} C^i \, a_i + \sum_{i=1}^{M} \sum_{j=1}^{i} C^{i,j} \, a_i a_j,$$

l'algorithme naturel d'évaluation nécessite donc $\frac{M(M+3)}{2}$ additions et $M(M+2)$ multiplications. Mais ce polynôme peut se réécrire comme suit :

$$\begin{aligned} C^0 + (C^M + C^{M,M} a_M) a_M \quad &+ \quad (C^{M-1} + C^{M-1,M-1} a_{M-1} + C^{M-1,M} a_M) a_{M-1} \\ &+ \quad \cdots + (C^1 + \textstyle\sum_{i=1}^{M} C^{1,i} a_i) a_1, \end{aligned}$$

ce qui conduit à une évaluation en $\frac{M(M+3)}{2}$ additions et $\frac{M(M+1)}{2}$ multiplications, donc à une diminution de $50 - \frac{100}{2(M+2)}\%$ du nombre des multiplications (49.43% pour $M = 86$).

L'évaluation des modèles nécessite celle d'un nombre M de tels polynômes à valeurs scalaires. Si les évaluations de chaque polynôme scalaire sont réalisées en parallèle (ou vectoriellement), un algorithme de type Horner nécessite quasiment moitié moins de temps de calcul en multiplications que l'algorithme naturel. Par contre, si ces M évaluations sont séquentielles, l'algorithme naturel est plus rapide pour $M > 2$ puisque les multiplications entrant dans le calcul des monômes peuvent être réalisées une fois pour toutes et stockées en mémoire pour servir à l'évaluation individuelle de chaque polynôme scalaire.

Ici nous nous sommes contentés d'un algorithme séquentiel.

3.4 Validation des modèles réduits

Nous allons maintenant tester la validité des modèles réduits obtenus.

3.4.1 Tests numériques

Les polynômes f^g et f^w associés aux deux modèles réduits ont été testés de deux manières. Tout d'abord, nous avons comparé les dérivées $\dot{a}_i^e$ approchées par différences finies avec les valeurs de $f_i^g(a^e)$ et de $f_i^w(a^e)$. La figure 3.12 nous montre que les valeurs données par f^g sont proches des $\dot{a}_i^e$, avec cependant des différences perceptibles au niveau des extrema. Le polynôme f^w est sensiblement moins bon, en particulier pour les dérivées des coefficients qui correspondent aux premiers modes POD (par exemple le cinquième), c'est-à-dire aux modes qui prédominent dans l'écoulement.

Nous avons ensuite simulé les modèles. L'intégration temporelle est réalisée le plus souvent par un schéma multi-pas explicite d'Adams-Bashforth d'ordre quatre (AB4), ce qui est plus efficace en terme de coût de calcul qu'un schéma de type Runge-Kutta de même ordre (RK4). En effet, un schéma AB4 ne nécessite à chaque itération qu'une évaluation du second membre du système EDO contre quatre pour un schéma RK4 et l'évaluation des

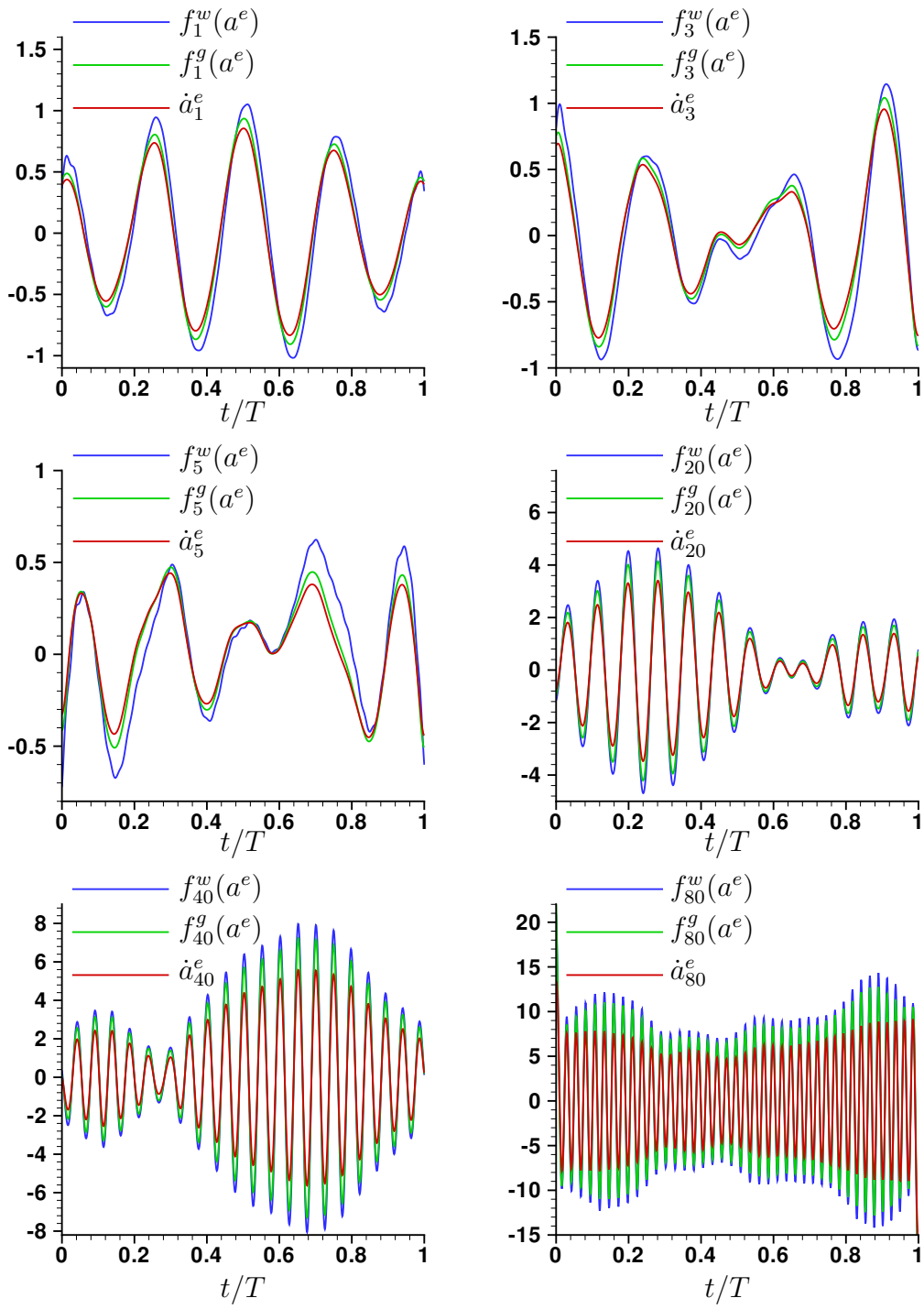


FIG. 3.12 – $\dot{a}_k^e$, $f_k^g(a^e)$ et $f_k^w(a^e)$ pour $k \in \{1, 3, 5, 20, 40, 80\}$.

polynômes est relativement onéreuse (voir la section précédente). Les trajectoires calculées par intégration des systèmes d'EDOs associés à f^g et f^w seront respectivement notées a^g et a^w. Elles sont calculées avec un pas de temps $\Delta t = 2 \times 10^{-6}\, T$.

Le modèle défini par f^w explose assez rapidement (vers $t \approx 0.3\, T$), et cela même pour un schéma RK4 utilisant un pas de temps très petit : le système semble intrinsèquement exploser en un temps fini inférieur à T. La figure 3.13 présente a_1^w : le système devient rapidement instable.

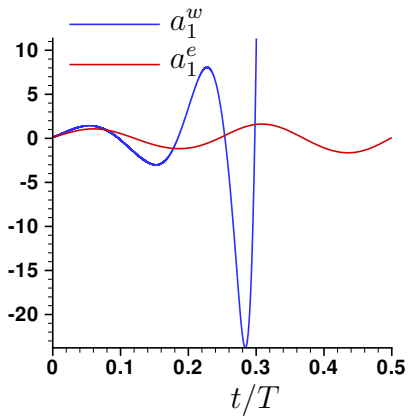


FIG. 3.13 – a_1^e et a_1^w.

Celui défini par f^g se comporte mieux. La simulation du modèle apparaît sur la figure 3.14 avec le comportement réel de l'écoulement. Ce modèle permet de simuler relativement bien l'écoulement sur une durée d'environ $T/4$. Au delà, il s'écarte assez nettement des données, même si le comportement du système reste relativement réaliste.

En fait, le système dynamique défini par f^g permet de prédire l'évolution de l'écoulement sur de petites durées, et de manière beaucoup plus précise que f^w. En effet, si on intégre le système d'EDOs associé à f^g ou f^w par morceaux sur une subdivision de dix intervalles égaux de $[0, T]$, en prenant comme condition initiale sur chaque segment la valeur exacte fournie par les données, on constate l'efficacité du modèle défini par f^g et le manque de précision de f^w. La figure 3.15 présente les résultats ainsi obtenus.

3.4.2 Discussion des problèmes pratiques de la modélisation POD-Galerkine réduite et introduction du problème des petites échelles

À la lecture des résultats précédents, deux questions se posent pour lesquelles nous allons tenter d'apporter des éléments de réponse :

- Comment expliquer le manque de fiabilité des modèles POD-Galerkine ?
- Est-il possible de corriger, d'améliorer le système dynamique ?

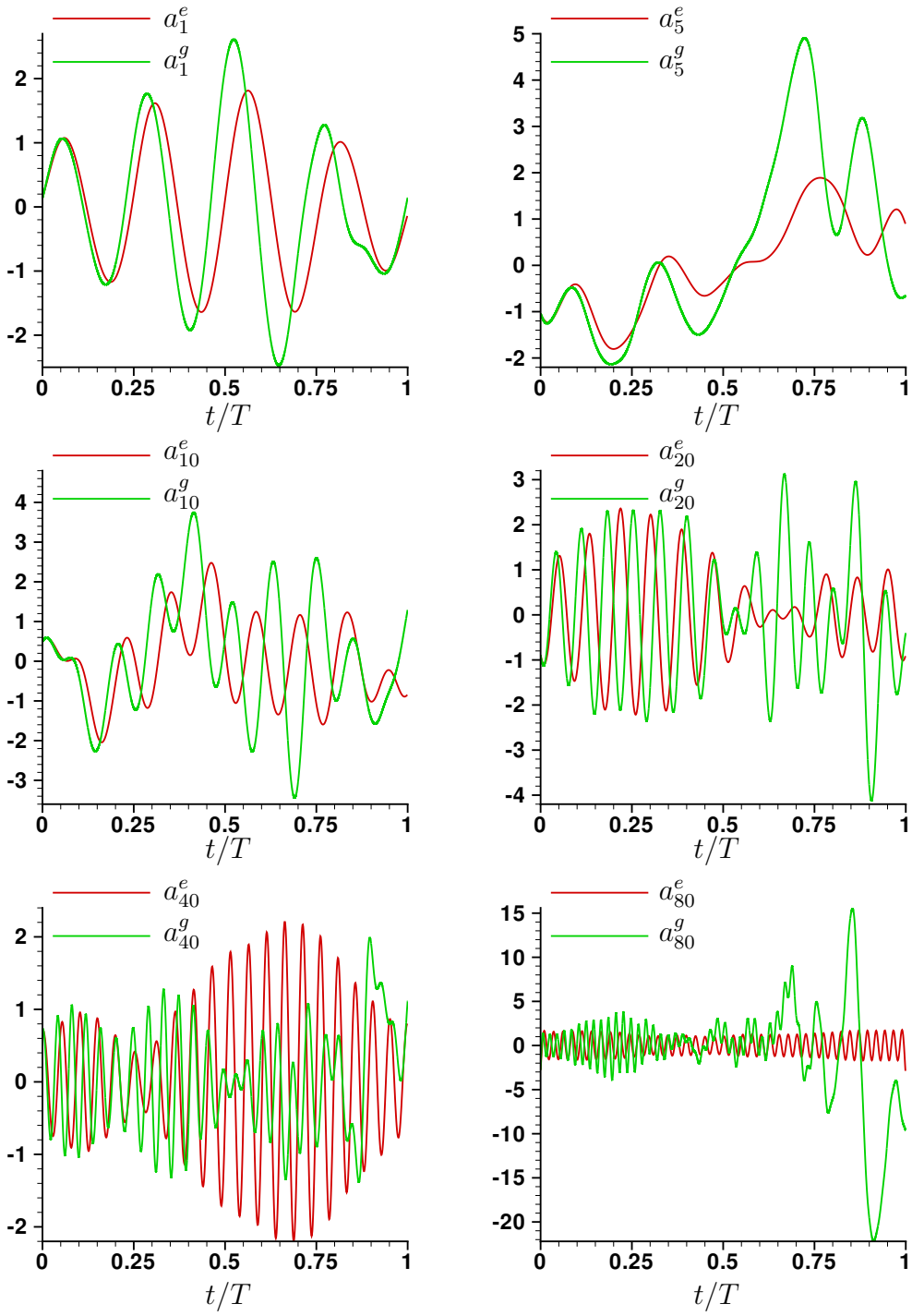


FIG. 3.14 – a_k^e et a_k^g pour $k \in \{1, 5, 10, 20, 40, 80\}$.

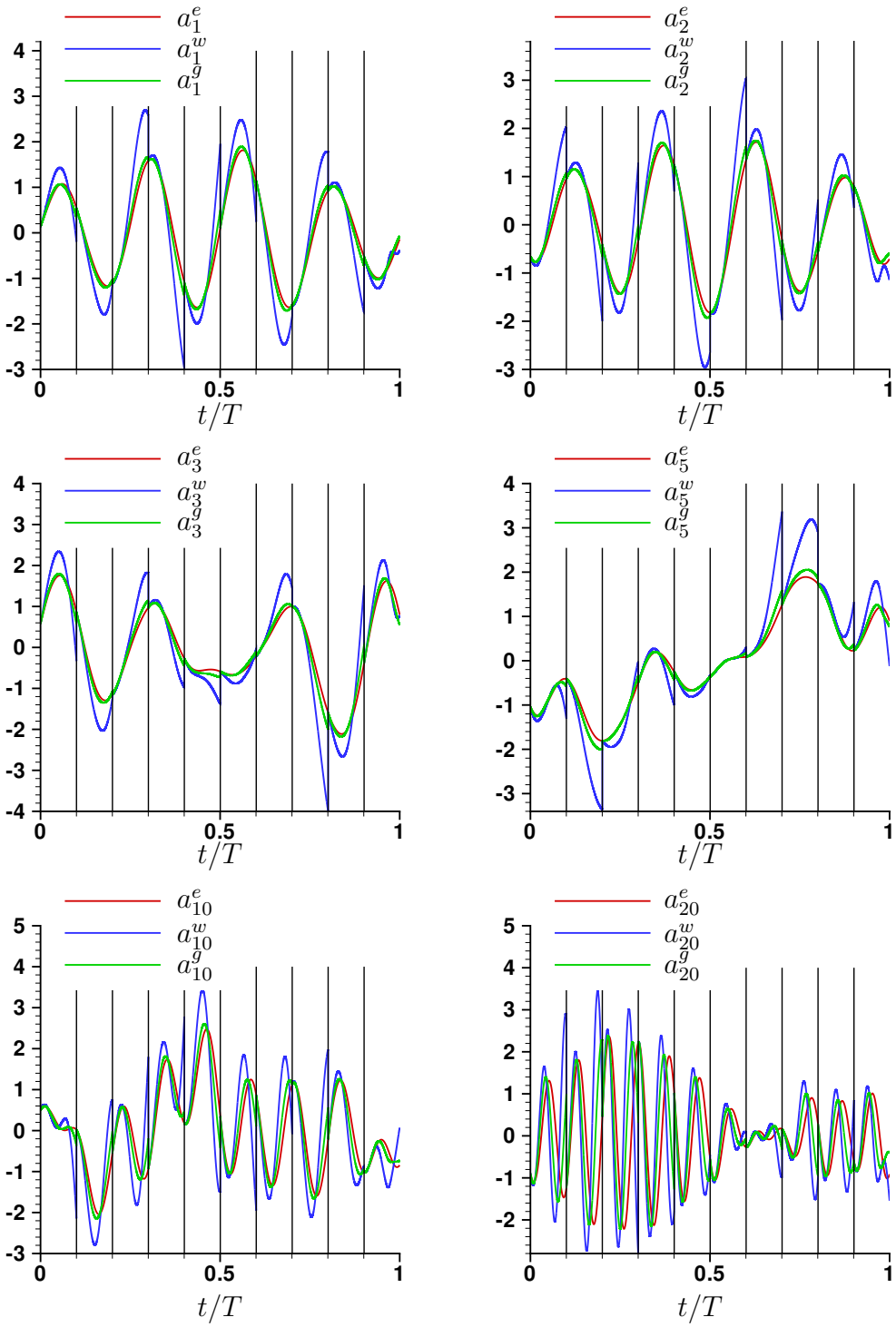


FIG. 3.15 – a_k^e et les solutions numériques a_k^g et a_k^w calculées sur une subdivision de dix segments, pour $k \in \{1, 2, 3, 20\}$.

Le polynôme f^w est moins fiable que f^g d'après les tests numériques précédents. Ce mauvais comportement numérique du modèle issu de la formulation vitesse-tourbillon a déjà été mis en évidence dans les travaux de Cordier ([14, p. 217]). Il est en partie dû à l'erreur qui est engendrée par les opérateurs de différentiation spatiale ; en effet, le calcul de f^w nécessite des différentiations de degré supérieur à celles qui interviennent dans le calcul de f^g.

De plus, il faut noter que, si les $M = 86$ premiers modes POD $\boldsymbol{\varphi}_k$ sont bien représentatifs du champ $\boldsymbol{u^e}$, rien n'assure que les modes $\boldsymbol{\varphi}_k^w = \boldsymbol{\nabla} \times \boldsymbol{\varphi}_k$ permettent d'approcher très précisément $\boldsymbol{w^e} = \boldsymbol{\nabla} \times \boldsymbol{u^e}$. En outre, à la base, si la simulation SGE fournit une bonne approximation du champ $\boldsymbol{u}$ solution des équations de Navier-Stokes (2.19)-(2.20), le schéma utilisé n'a pas été spécifiquement conçu pour donner une approximation très fine de $\boldsymbol{w}$ solution de (2.22) : $\boldsymbol{u^e} \approx \boldsymbol{u}$ n'implique pas nécessairement $\boldsymbol{\nabla} \times \boldsymbol{u^e} \approx \boldsymbol{\nabla} \times \boldsymbol{u}$.

Ainsi la modélisation tourbillonnaire pose des problèmes qui pourraient être éventuellement résolus en s'assurant que les modes $\boldsymbol{\varphi}_k^w$ fournissent une bonne approximation de $\boldsymbol{w}$. En particulier, une POD de type H^1 pourrait être utilisée plutôt qu'une POD de type L^2, d'autant plus que l'orthogonalité des modes POD ne simplifie pas le calcul du modèle dans le cas de la formulation tourbillonnaire (mais cette méthode est plus coûteuse). Une autre idée serait de choisir les modes POD à conserver à partir des valeurs

$$\lambda_i^w = \int_0^T \frac{\left(\boldsymbol{w^e}(t), \boldsymbol{\varphi}_i^w \right)_{\mathrm{L}^2(\Omega)^d}^2}{\left(\boldsymbol{\varphi}_i^w, \boldsymbol{\varphi}_i^w \right)_{\mathrm{L}^2(\Omega)^d}^2} \, dt$$

en plus des λ_i.

De manière plus générale, plusieurs réponses peuvent être apportées à la première question : influence de certains termes de bord qui peuvent avoir été négligés si les conditions aux limites n'ont pas été traitées explicitement (par exemple T_{in} pour le premier modèle, voir la section 2.3.1 pour plus de détails), conséquence des erreurs numériques, sensibilité numérique du système dynamique, nature des données $\boldsymbol{u^e}$ (elles ne satisfont jamais exactement la formulation variationnelle (FVNSI) ou celle obtenue par "projection" de Galerkine, en particulier des données SGE calculées par différences finies).

Cependant, du moins en ce qui concerne f^g, l'une des réponses plus satisfaisantes est l'influence des modes POD d'indice strictement supérieur à M qui ont été tronqués pour obtenir un système de dimension réduite M. Ces derniers, associés à des structures de faible énergie cinétique, correspondent aux petites échelles spatiales de l'écoulement (voir la section 3.2.2).

Cette hypothèse est renforcée par les figures 3.12 et 3.14 qui montrent que l'erreur entre le modèle associé à f^g et l'écoulement apparaît au niveau des modes de plus grand indice, qui correspondent aux plus petites structures. Si on suppose que dans l'écoulement seules les structures de taille proche interagissent entre elles[3], les modes tronqués devraient être

[3]Cette hypothèse est confortée par plusieurs travaux et par notre étude des transferts d'énergie cinétique (section 3.5).

en interaction avec les derniers modes retenus, ce qui rejoint l'observation faite sur les figures (voir notamment le comportement du mode d'indice 80 sur la figure 3.14).

Pour mieux visualiser ce phénomène, nous avons représenté l'erreur $a_i^g(t)^2 - a_i^e(t)^2$ de distribution d'énergie cinétique entre la solution numérique $a^g(t)$ associée à f^g et la distribution correspondant aux données (cette erreur est relative, la différence absolue d'énergie étant $\lambda_i\,(a_i^g(t)^2 - a_i^e(t)^2)$). On obtient alors la figure 3.16 sur laquelle a aussi été reportée, en fonction du temps t et de l'indice i des modes, la distribution d'énergie réelle de l'écoulement dans la base POD, c'est-à-dire $\lambda_i\, a_i^e(t)^2$.

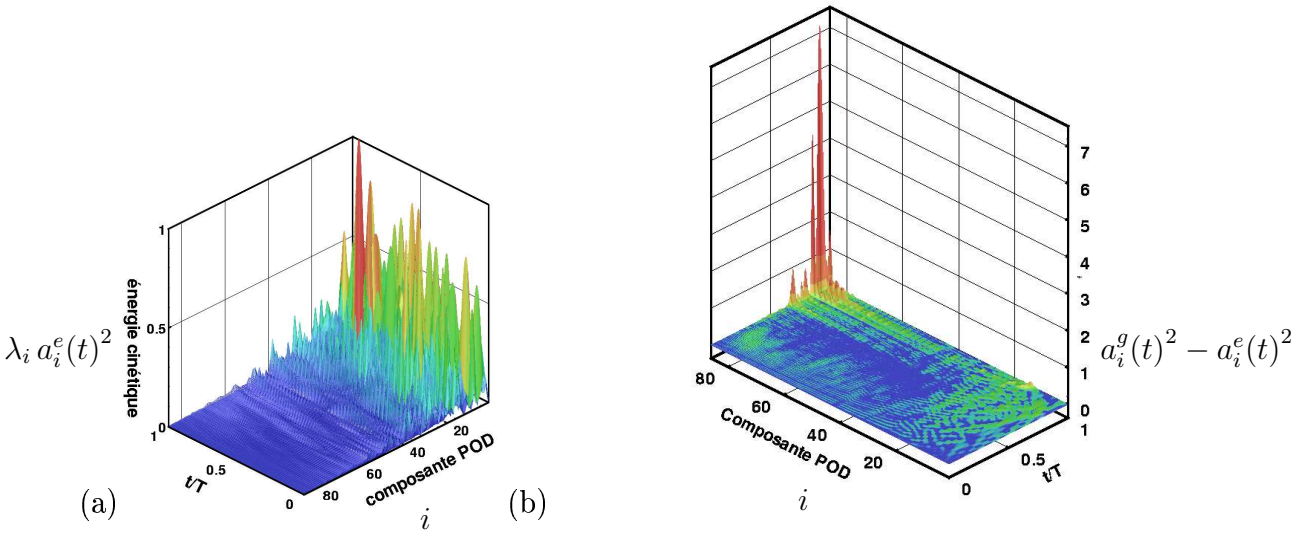


FIG. 3.16 – (a) Distribution d'énergie des données au sein des $M = 86$ premiers modes POD. (b) Différence relative de distribution d'énergie entre modèle et données.

Bien que la visualisation obtenue ne soit pas extrêmement probante, elle est plutôt en accord avec ce mécanisme de transfert d'énergie : un surplus d'énergie apparaît principalement au niveau des derniers modes puis semble être transmis aux autres modes, au tout début de la simulation mais surtout à la fin. Cependant, tous les modes POD étant couplés, des approximations ayant été faites pour définir le modèle et l'analyse mathématique de la section 2.4.1 montrant qu'en théorie le fait de couper la cascade d'énergie d'un écoulement incompressible n'entraîne pas nécessairement une accumulation d'énergie au niveau des derniers modes (le modèle réduit dissipant toujours l'énergie qu'on lui apporte dans le cas d'un écoulement fermé), on comprend que cette visualisation avait peu de chance d'apporter une réponse claire. Une analyse plus pertinente des transferts d'énergie et de leur paramétrisation sera menée dans la section 3.5.

Il faut noter qu'il est possible d'améliorer le comportement d'un modèle POD-Galerkine en augmentant le nombre M de modes exploités (voir [87] par exemple), cependant les coûts de calcul croissent alors extrêmement rapidement puisque le nombre de coefficients polynômiaux d'un modèle de degré deux varie asymptotiquement de l'ordre de $\mathcal{O}(M^3/2)$: pour des nombres de Reynolds relativement grands, les modèles POD-Galerkine véritablement réduits, c'est-à-dire dont la dimension reste très petite par rapport aux systèmes utilisés en simulation numérique et qui sont exploitables pour un coût informatique faible, n'ont le plus souvent pas un comportement physique et leur utilisation nécessite donc de résoudre ce problème de modélisation des modes POD négligés.

Dans le cadre des écoulements turbulents, la majorité des auteurs ayant mené des études grâce à des modèles POD propose de modéliser l'influence des modes tronqués par des ajouts artificiels de viscosité. Ces viscosités artificielles, qui rejoignent l'esprit des modèles utilisés en SGE, sont censées prendre en compte le phénomène de pompage puis de dissipation de l'énergie cinétique des grosses structures par les petites. La seconde question semble ainsi faire echo au problème de la modélisation des petites échelles rencontré en SGE. Restent alors les problèmes de la forme des modèles et de la calibration de leurs paramètres.

La section suivante aborde le problème de la pertinence physique et de la forme d'une modélisation cohérente de type visqueuse des effets des modes POD négligés.

Remarque importante. Rappelons que les données proviennent d'une simulation SGE qui s'attache à la résolution des échelles spatiales de l'écoulement suffisamment grandes pour être prises en compte par le maillage et qui modélise les autres (échelles sous-mailles). Néanmoins, les deux modèles f^g et f^w ont été définis en partant des équations de Navier-Stokes classiques et non des équations filtrées : le modèle sous-maille utilisé lors de la SGE n'est pas pris en compte dans notre modélisation POD-Galerkine. Il y a donc un décalage intrinsèque entre la nature des données et la formulation Navier-Stokes utilisée.

Prendre en considération le modèle sous-maille pourrait être bénéfique, mais l'ignorer simplifie considérablement la modélisation (consulter [94] pour un exemple de modèle POD-Galerkine qui fait intervenir la viscosité sous-maille de la SGE). Il est surtout important de garder à l'esprit que négliger les derniers modes POD, ce qui est le principe même de la modélisation réduite, a intrinsèquement un impact plus important sur la précision du modèle : les structures spatiales des modes POD qui sont négligés sont nécessairement plus grandes que les échelles non-résolues par la SGE qui sont modélisées par le terme sous-maille, et jouent en conséquence un rôle plus important dans la dynamique du fluide. Il apparaît ainsi que la prise ou non-prise en compte du modèle sous-maille est partiellement voire complètement inhibée par le problème de la réduction de la base POD, c'est-à-dire de la modélisation des petites échelles. C'est pourquoi l'approche que nous avons choisie est d'ignorer les termes sous-maille et de se focaliser sur la modélisation des modes POD tronqués (section 3.5) ou encore d'appréhender le problème de la modélisation des "échelles non-résolues", échelles sous-maille ou des modes POD tronqués, de manière globale en recalibrant le modèle (chapitre 4).

3.5 Transferts d'énergie cinétique et paramétrisation visqueuse

Nous allons maintenant étudier les caractéristiques principales des transferts d'énergie cinétique au sein des modes POD. En particulier, une analyse quantitative des interactions entre les modes, basée sur le calcul d'une pseudo-viscosité, est proposée. Le but de ce travail est double.

Tout d'abord, notons que les transferts d'énergie cinétique ont historiquement été étudiés en utilisant la décomposition de Fourier qui n'est appropriée que pour les écoulements périodiques. Ainsi, des résultats pertinents, tant théoriques (voir [47]) que numériques (par exemple [97]), ont permis de mettre en lumière la dynamique des fluctuations turbulentes, établissant l'existence de cascades d'énergie directes et inverses.

Quelques auteurs se sont intéressés au cas des écoulements turbulents non-homogènes, mais le plus souvent pour des configurations locales (voir l'introduction de ce chapitre, page 72). Ici, la méthode POD-Galerkine est appliquée à un écoulement qui occupe un domaine spatial étendu, et qui est turbulent, non-homogène et décollé.

En outre, puisque la POD est moralement équivalente à une décomposition de Fourier dans les directions homogènes (voir la section 1.3.2), la plupart des auteurs appliquent une décomposition explicitement hybride : décomposition de Fourier dans les directions homogènes, puis POD dans les directions restantes. Cette approche est différente de la nôtre, puisque, même si les modes obtenus étaient strictement identiques, le fait d'opérer une POD tridimensionnelle nous impose de classer les modes par leur valeur singulière (i.e. leur énergie cinétique moyenne), alors qu'ils peuvent être ordonnés séparément suivant chaque direction homogène si on utilise une décomposition hybride Fourier/POD.

Comme il avait déjà été proposé par Rempfer *et al.* (consulter [78]), le premier objectif est de dégager les principales caractéristiques des transferts d'énergie cinétique au sein d'une base issue d'une POD tridimensionnelle, et de comparer ces observations avec les résultats donnés par la décomposition de Fourier dans le cas homogène isotrope.

Bien que dans le cas des écoulements turbulents, très peu de modes POD détiennent la majeure part de l'énergie cinétique totale et peuvent servir de base à la construction d'un système dynamique de dimension réduite par la méthode de Galerkine, les modes de faible énergie qui sont tronqués doivent être pris en compte pour retrouver une description précise de la physique. Ceci est vrai sur l'exemple de la marche, comme nous l'avons vu à la section 3.4.

Ce problème s'apparente *a priori* à celui de la SGE : il faut modéliser une partie négligée de l'écoulement qui correspond (en pratique pour la POD) à de petites échelles spatiales dont la contribution physique essentielle est de dissiper de l'énergie cinétique puisée aux plus grandes.

Pour ce faire, la majorité des auteurs, suivant l'idée de Aubry *et al.* [6], recoure à

un modèle diffusif basé sur une extension du modèle de viscosité spectrale[4] proposé par Heisenberg pour les écoulements homogènes : [8, 70, 93]. Comme Aubry *et al.* l'avaient remarqué, cette modélisation semble similaire à celle de Smagorinsky, très utilisée en SGE, qui peut être interprétée comme une extension du modèle de Heisenberg à l'espace physique (et non plus spectral) pour des écoulements décomposés dans des bases locales.

Comme noté plus haut, presque tous les travaux qui adoptent cette approche (études de couche limite turbulente) manipulent une décomposition hybride : celle de Fourier dans deux directions puis la POD. Après avoir tronqué les modes dans cette double représentation Fourier/POD, la validité d'une modélisation de type visqueuse pour tenir compte des modes négligés peut se justifier par l'hypothèse de Kolmogorov, et si on admet qu'à cette troncature correspond, au niveau spectral, un filtre opérant une coupure qui a lieu pour des échelles spatiales suffisamment petites, comme c'est le cas en SGE. Néanmoins cette validité reste à être établie pour des écoulements complexes non-homogènes, puisque l'hypothèse sous-jacente de l'existence et de la dominance d'une cascade d'énergie des premiers modes POD vers les modes de faible énergie (cascade directe) n'a pas encore véritablement été le sujet d'investigations.

Le second objectif de notre étude est donc une analyse quantitative des interactions entre modes POD par une paramétrisation visqueuse, afin de déterminer des lignes directrices pour la définition de modèles de prise en compte des modes tronqués.

On notera que, puisqu'il est connu, dans le simple cadre de la turbulence homogène isotrope, que les caractéristiques des modèles sous-maille utilisés en SGE varient beaucoup avec la forme du filtre et le positionnement de la coupure, le cas d'un modèle POD réduit n'est pas trivial.

Notre étude se base sur le système dynamique défini par f^g, que nous avons extrait de la formulation classique (vitesse-pression) des équations de Navier-Stokes (se référer à la section 3.3.1). En effet, le système réduit de dimension $M = 86$ correspondant est fiable, nonobstant les modes qui n'ont pas été pris en compte et qui permettraient au modèle d'améliorer sa précision. Compte tenu du coût de calcul déjà très important de ce modèle, et de sa relativement bonne fiabilité, nous n'avons pas construit de modèle de plus grande dimension.

Notations

Le système POD-Galerkine réduit étudié est autonome puisqu'il est construit pour $\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}^e$ (donc indépendant de t) et que le fluide n'est sujet à aucune force extérieure ($\boldsymbol{h} = \boldsymbol{0}$) : $\dot{a}(t) = f^g(a(t))$. Nous allons réutiliser par la suite les notations C_i^j et $C_i^{j,k}$ du système (2.35) pour nommer les coefficients polynômiaux (indépendants du temps) de $f^g = (f_1^g \cdots f_M^g)^T$.

[4] Dans cette introduction, "spectral(e)" a le sens habituel donné par la théorie de Fourier (appliquée en espace).

Plus précisément, on obtient, en isolant les coefficients d'origine visqueuse,

$$f_i^g(a(t)) = C_i^0 + \frac{D_i^0}{\mathrm{Re}} + \sum_{j=1}^{M}(C_i^j + \frac{D_i^j}{\mathrm{Re}})\, a_j(t) + \sum_{j=1}^{M}\sum_{k=1}^{j} C_i^{j,k}\, a_j(t)\, a_k(k), \tag{3.1}$$

où les coefficients réels $C_\times^\times$ et $D_\times^\times$ sont indépendants de Re et définis par :

$$\sigma_i\, C_i^0 = -\,\mathcal{C}(\bar{\boldsymbol{u}}, \bar{\boldsymbol{u}}, \boldsymbol{\varphi_i}), \qquad \sigma_i\, D_i^0 = -\,\mathcal{A}(\bar{\boldsymbol{u}}, \boldsymbol{\varphi_i}),$$
$$\sigma_i\, C_i^j = -\sigma_j\, \left[\, \mathcal{C}(j, \bar{\boldsymbol{u}}, \boldsymbol{\varphi}_i) + \mathcal{C}(\bar{\boldsymbol{u}}, \boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i) \,\right], \qquad \sigma_i\, D_i^j = -\sigma_j\, \mathcal{A}(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i)$$
$$\text{et} \quad \sigma_i\, C_i^{j,k} = -\frac{\sigma_j\sigma_k}{1+\delta_{j,k}}\, \left[\, \mathcal{C}(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_k, \boldsymbol{\varphi}_i) + \mathcal{C}(\boldsymbol{\varphi}_k, \boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i) \,\right]$$

(avec $\bar{\boldsymbol{u}}$ est pris égal à $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$) en écrivant les termes quadratiques sans redondance.

Ce système POD-Galerkine permet de régir le champ des vitesses approché $\boldsymbol{u} = \bar{\boldsymbol{u}} + \tilde{\boldsymbol{u}}$ avec $\tilde{\boldsymbol{u}}(t) = \sum_{k=1}^{M} \sigma_k\, a_k(t)\, \boldsymbol{\varphi}_k$.

Dans la suite de cette section, $\langle . \rangle$ désignera l'opérateur de moyenne arithmétique sur les N clichés :

$$\langle g(t) \rangle = \frac{1}{N} \sum_{j=1}^{N} g(t_j). \tag{3.2}$$

On a en particulier $\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}^{\boldsymbol{e}} = \langle \boldsymbol{u}^{\boldsymbol{e}} \rangle$.

3.5.1 Transferts d'énergie cinétique entre modes POD

L'énergie cinétique fluctuante totale par unité de masse est

$$K(t) = \frac{1}{2} \|\tilde{\boldsymbol{u}}\|_{\mathrm{L}^2(\Omega)^d}{}^2 = \sum_i K_i(t) \text{ où } K_i(t) = \frac{1}{2}\lambda_i\, a_i(t)^2.$$

K_i est l'énergie captée par le $i$ème mode. Nous tirons de (3.1) l'expression

$$\dot{K}_i = \lambda_i\, a_i \dot{a}_i = \widetilde{C}_i^0\, a_i + \underbrace{\sum_{j=1}^{M} \widetilde{C}_i^j\, a_j\, a_i}_{\text{interactions diadiques}} + \underbrace{\sum_{j=1}^{M}\sum_{k=1}^{j} \widetilde{C}_i^{j,k}\, a_j\, a_k\, a_i}_{\text{interactions triadiques}} \tag{3.3}$$

avec

$$\widetilde{C}_i^0 = \lambda_i \left(C_i^0 + \frac{D_i^0}{\mathrm{Re}} \right), \quad \widetilde{C}_i^j = \lambda_i \left(C_i^j + \frac{D_i^j}{\mathrm{Re}} \right), \quad \text{et} \quad \widetilde{C}_i^{j,k} = \lambda_i\, C_i^{j,k}. \tag{3.4}$$

L'équation (3.3) montre que l'évolution de l'énergie d'un mode résulte de trois types d'interactions : linéaires avec le champ moyen, diadiques qui proviennent de l'interaction

avec le champ moyen et des termes visqueux, et triadiques dont l'origine est le terme non-linéaire de convection. Dans le cadre d'une décomposition de Fourier, les termes diadiques dégénèrent en terme linéaire, et les interactions triadiques sont nulles sauf pour certaines triades. Ainsi, contrairement aux modes de Fourier, les modes POD sont en interactions via les termes visqueux, et à tout triplet correspond un transfert. En outre, rappelons que l'effet global des interactions diadiques est une dissipation d'énergie à tout instant[5] (voir la section 2.4.1). Il faut néanmoins souligner que les échanges moyens entre modes POD via les interactions diadiques sont nuls, puisque $\langle a_i^e\, a_j^e \rangle = 0$ si $i \neq j$. De plus, l'effet des interactions diadiques est faible par rapport à celui des interactions triadiques pour des nombres de Reynolds importants, puisque les termes diadiques tendent vers 0 lorsque Re augmente[5] : ceci explique que les contributions d'origine visqueuse sont négligeables dans les estimations des paramétrisations visqueuses proposées à la section 3.5.2.

Nous allons maintenant nous focaliser sur les interactions triadiques. Le terme $\widetilde{C}_i^{j,k} a_j\, a_k\, a_i$ est perçu comme l'influence du mode d'indice $\max(j,k)$ sur la variation de K_i (voir la remarque en fin de section). De cette manière, l'influence du $j$ème mode sur l'énergie du $i$ème est

$$\Pi(i|j) = \sum_{k=1}^{j} \widetilde{C}_i^{j,k}\, a_k\, a_j\, a_i.$$

Dans toute la suite, $\Pi(i|j)$ sera estimé à partir des coefficients temporels de référence donnés par la POD : $a_i = a_i^e$.

Le transfert moyen $\langle \Pi(i|j) \rangle$ est présenté figure 3.17, à travers la carte de son module (en échelle logarithmique) et ses profils pour trois valeurs de i en fonction de $(i-j)$.

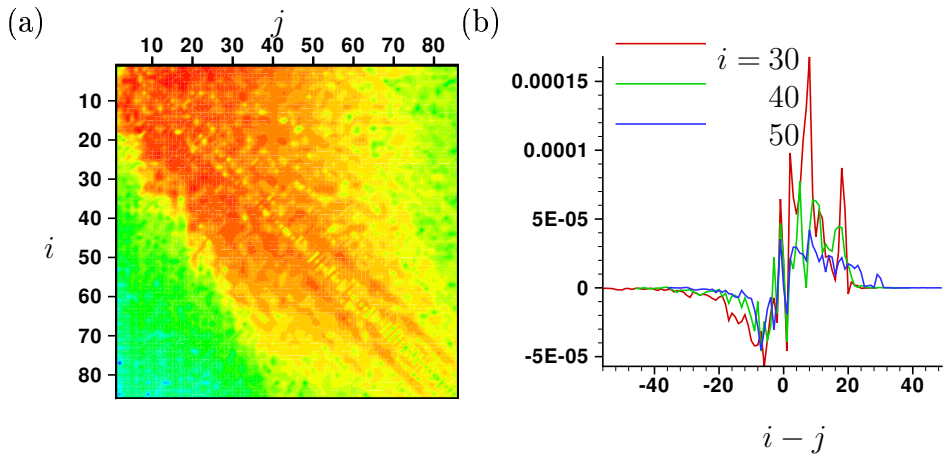


FIG. 3.17 – (a) Carte de $\log(|\langle \Pi(i|j) \rangle|)$. (b) $\langle \Pi(i|j) \rangle$ pour $i \in \{30, 40, 50\}$ en fonction de $(i-j)$.

[5] Dans le cas où aucune interaction entre le champ $\bar{\boldsymbol{u}}$ et les modes POD n'est prise en compte dans les interactions diadiques (purement modales).

On observe que les transferts au sein des modes POD sont locaux ; en effet, $\langle \Pi(i|j) \rangle$ est négligeable pour des indices trop éloignés ($|i-j| \geq 25$). Cette propriété de localité avait déjà été mise en évidence par les résultats de Rempfer *et al.* [78], mais pour un écoulement non turbulent (en transition), et pour une POD effectuée dans un petit sous-domaine de l'espace initial de la simulation numérique. Elle peut être interprétée comme l'extension d'un phénomène mis en évidence dans le cadre de la turbulence homogène isotrope : un mode de Fourier de nombre d'onde k échange la majeure part de son énergie avec les modes de l'intervalle $[k/2, 2k]$, c'est-à-dire que les transferts sont locaux (voir l'analyse théorique de Kraichnan [48] et les résultats numériques de Domaradzki *et al.* [18]).

Cette observation est cohérente avec le fait que les modes POD convergent vers des modes de Fourier (du moins localement) quand leur indice tend vers l'infini (Foias *et al.* [23]), ou autrement dit dans la limite des très grands nombres d'onde (i.e. les échelles dissipatives). Néanmoins, elle n'en était pas une conséquence directe, la coupure implicitement obtenue en ne conservant que les M premiers modes POD ayant lieu dans de plus grandes échelles, et la POD étant extraite de données calculées par SGE. En outre, les analyses menées grâce à une décomposition de Fourier ne traitent que de la zone inertielle du spectre dans le cadre de la turbulence homogène, alors que dans le cas présent, l'écoulement est non-homogène, borné et décollé, et que les modes POD sont globaux au sens où ils intégrent la dynamique sur tout l'espace du domaine de calcul.

La direction principale des transferts est retrouvée en regardant le signe du transfert moyen $\langle \Pi(i|j) \rangle$. La matrice correspondante est tracée sur la figure 3.18. Les régions noires

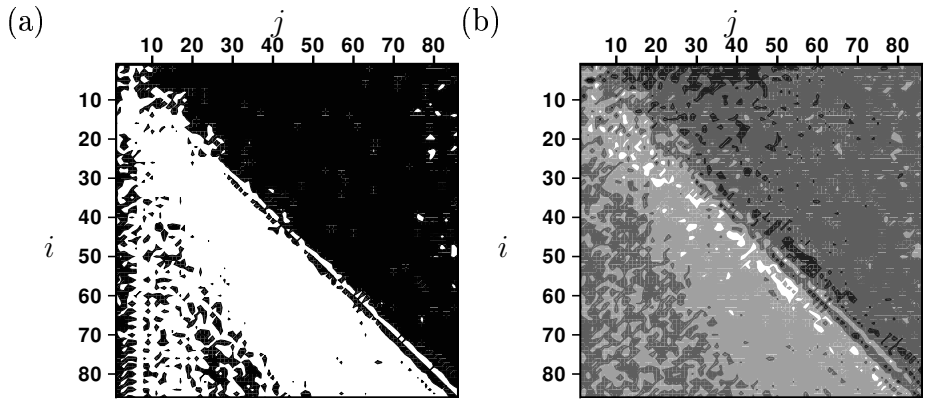


FIG. 3.18 – (a) Signe du transfert triadique moyen $\langle \Pi(i|j) \rangle$. Blanc : positif ; noir : négatif. (b) Pourcentage de T pendant lequelle $\Pi(i|j)$ est positif avec quatre niveaux de gris du noir au blanc : (25,40), (40,50), (50,60) et (60,75).

sont associées à une valeur moyenne négative, donc à un drainage d'énergie cinétique du mode i par le mode j, et les blanches à une valeur positive, c'est-à-dire à un gain net d'énergie par le mode i. Ainsi, de façon similaire aux échanges observés entre les modes de

Fourier, le phénomène prépondérant est une cascade directe d'énergie au sein des modes POD : un mode i draine de l'énergie des modes $j < i$ et redistribue de l'énergie aux modes $j > i$. Dans le cadre de l'analyse de Fourier, ce phénomène de transfert d'énergie vers les grands nombres d'onde apparaît par exemple sur la figure 3.19 qui est proposée par

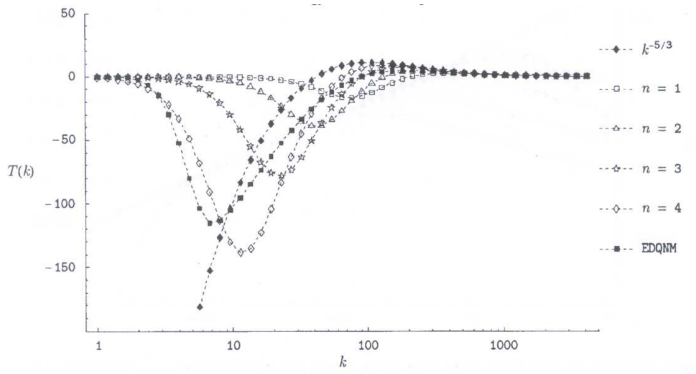


FIG. 3.19 – *Negative values of the kinetic energy transfer term $T(k)$ over a range of k corresponds to a net energy flux out of those modes, and positive value of $T(k)$ over a range of k corresponds to a net energy flux into those modes.* Consulter [86] pour plus de détails.

Schilling *et al.* [86] et qui montre que les grands nombres d'onde k gagnent de l'énergie au détriment des petits nombre d'ondes dans le cadre de la turbulence incompressible, tridimensionnelle et isotrope. Notons que sur la figure 3.18, il existe également de petites régions qui correspondent à une cascade inverse, mais elles sont le plus souvent associées à de très petites valeurs du module du transfert moyen, et ne doivent donc pas être considérées comme la preuve de l'existence d'une cascade inverse en moyenne.

Afin de mieux cerner la dynamique des modes POD, nous avons tracé, figure 3.18, le pourcentage de temps pendant lequel le transfert instantané $\Pi(i|j)$ est positif. Il apparaît que les gains (et pertes) nets sont associés à des régions où le transfert est la plupart du temps dans la même direction que le transfert moyen, mais qu'il prend les deux directions sur tout l'intervalle temporel considéré. Ceci peut apparaître comme une généralisation des cascades inverses que l'on peut détecter dans une base de Fourier, ou encore comme le fait que le signe des transferts locaux à travers un nombre d'onde fixé peut changer de temps à autre et d'un endroit à l'autre dans l'espace physique.

En conclusion, il semble que les transferts d'énergie cinétique fluctuante au sein de modes POD présentent beaucoup de similitudes avec ceux observés entre des modes de Fourier. Ceci n'est pas surprenant à la vue des structures que l'on a pu mettre en évidence pour les modes (section 3.2.2) : plus l'indice d'un mode est grand, plus les structures associées sont petites. Cette observation est cohérente dans la mesure où les modes POD sont ordonnés par leur énergie cinétique moyenne, et où les petites structures sont, de manière assez générale, moins énergétiques que les grosses.

Ainsi, en accord avec l'hypothèse d'isotropie locale de Kolmogorov, les conclusions principales issues d'une analyse par décomposition de Fourier peuvent s'étendre au cadre de la POD. Le phénomène principal est une cascade d'énergie cinétique des modes POD de petit indice vers ceux de grand indice. Cette cascade directe est un effet moyen, des transferts inverses pouvant survenir sur de brèves durées, révélant une cascade inverse.

Remarque. Au premier abord, la définition de $\Pi(i|j)$ peut sembler arbitraire. Cependant elle s'impose naturellement si on se préoccupe des termes qui disparaissent du modèle POD-Galerkine défini par l'expression (3.1) lorsque l'on tronque quelques modes POD.

Pour illustrer ceci, considérons une nouvelle "coupure" $l < M$ (on tronque les $M - l$ derniers modes POD). Les interactions triadiques du nouveau sytème, de dimension l, sont $\sum_{j=1}^{l} \Pi(i|j)$. Autrement dit, pour tout j, $\Pi(i|j)$ représente les interactions triadiques qui sont ajoutées quand la base modale POD est complétée par un nouveau mode.

Notons de plus que

$$\sum_{j=1}^{M} \Pi(i|j) = \sum_{j=1}^{M} \sum_{k=1}^{j} \widetilde{C}_i^{j,k}\, a_j\, a_k\, a_i \ :$$

les $\Pi(i|j)$ correspondent bien à une répartition des interactions triadiques régissant le $i$ème mode dans la base POD.

3.5.2 Paramétrisation visqueuse

À partir du système POD-Galerkine de dimension M, nous allons maintenant analyser les interactions entre les modes grâce à différentes approximations de paramètres, que l'on peut assimiler à des ajouts artificiels de viscosité, qui quantifient la dépendance d'un mode par rapport aux $M - l$ modes d'indice supérieur à un indice de coupure l. En particulier, nous calculerons deux paramétrisations des transferts d'énergie à travers cette coupure l.

Le but de ces estimations est de déterminer les principales caractéristiques des interactions entre modes, afin de proposer à plus long terme des modèles cohérents de prise en compte des modes POD qui sont négligés lorsque l'on construit un modèle dynamique POD réduit.

Cette problématique peut apparaître comme une extension du problème rencontré en SGE, lorsque l'on cherche à modéliser les petites échelles de l'écoulement, trop coûteuses à calculer. Dans le contexte de la SGE, la coupure entre échelles résolues et non résolues correspond à un filtrage spatial, et de nombreuses études reposant sur une analyse de Fourier ont permis de proposer des modèles efficaces, basés sur la définition d'une viscosité artificielle, pour résoudre ce problème de fermeture. Cependant, dans le cadre de la POD, les caractéristiques principales que doit idéalement reproduire un paramètre de type visqueux, pour bien modéliser les modes POD tronqués, restent à être étudiées. Cela d'autant plus que de nombreuses études dans le cas d'une analyse de Fourier ont déjà montré que la paramétrisation visqueuse idéale est très dépendante de nombreux facteurs tels que la

forme du spectre, la forme du filtre ou encore du nombre d'onde de coupure ; par exemple [18, 98].

Une analyse similaire est donc proposée ici, dont les conclusions pourraient permettre d'améliorer un système dynamique POD-Galerkine réduit.

Modélisation visqueuse

Après avoir introduit l'indice de coupure l, nous avons différencié les interactions faisant intervenir un terme d'indice supérieur à l, dites non résolues et indiquées par l'exposant $>$, des autres interactions, dites résolues et indiquées par $\leq$. On isole également les termes d'origine viqueuse et on obtient formellement la décomposition suivante des polynômes du système dynamique POD-Galerkine (équation (3.1)) :

$$f_i^g(a) = c_i^{\leq}(a) + c_i^{>}(a) + \frac{1}{\mathrm{Re}} \left(d_i^{\leq}(a) + d_i^{>}(a)\right)$$

avec

$$c_i^{\leq}(a) = C_i^0 + \sum_{j=1}^{l} C_i^j \, a_j + \sum_{j=1}^{l} \sum_{k=1}^{j} C_i^{j,k} \, a_j a_k,$$

$$d_i^{\leq}(a) = D_i^0 + \sum_{j=1}^{l} D_i^j \, a_j,$$

$$c_i^{>}(a) = \sum_{j=l+1}^{M} C_i^j \, a_j + \sum_{j=l+1}^{M} \sum_{k=1}^{j} C_i^{j,k} \, a_j \, a_k$$

et

$$d_i^{>}(a) = \sum_{j=l+1}^{M} D_i^j \, a_j.$$

Le problème de fermeture consiste alors ici à modéliser les termes non résolus grâce aux termes résolus. On cherche ici une fermeture de type visqueuse simple, de paramètre $\nu(i|l)$, nous donnant une approximation de la forme

$$f_i^g(a) \approx c_i^{\leq}(a) + \left(\frac{1}{\mathrm{Re}} + \nu(i|l)\right) d_i^{\leq}(a), \tag{3.5}$$

c'est-à-dire telle que $f_i^{>} \approx \nu(i|l) \, d_i^{\leq}$ avec $f_i^{>} = c_i^{>} + \frac{1}{\mathrm{Re}} d_i^{>}$.

Le coefficient sans dimension $\nu(i|l)$ peut être perçu comme un ajout de viscosité qui permet de modéliser l'influence des modes d'indice supérieur à l sur le mode i. En effet, $U_\infty \, L \, \nu(i|l)$ représente l'augmentation artificielle de la viscosité cinématique pour le $i$ème mode si U_∞ et L sont respectivement la vitesse et la longueur caractéristiques utilisées pour définir le nombre de Reynolds des équations de Navier-Stokes adimensionnées : $\mathrm{Re} = (U_\infty \, L)/\nu$ avec $\nu = \rho/\mu$ la viscosité cinématique des équations de Navier-Stokes et $\rho = \rho_\infty = 1$ (consulter la page 42).

Ce coefficient dépend *a priori* de l'indice i du mode considéré et de l'indice l de coupure, de manière analogue par exemple à la viscosité artificielle de Kraichnan-Chollet-Lesieur définie dans l'espace de Fourier (voir [48] et [11]).

Dans la suite, des estimations par moyenne et moindres carrés seront proposées. Elles seront extraites de la formulation (3.5), mais aussi d'une variante qui prend explicitement en compte la variation $\dot{K}_i$ de l'énergie cinétique du mode. En effet, puisque

$$\frac{1}{2\lambda_i}\dot{K}_i = \tilde{f}_i^g(a) = \tilde{c}_i^{\leq}(a) + \tilde{c}_i^{>}(a) + \frac{1}{\mathrm{Re}}\left(\tilde{d}_i^{\leq}(a) + \tilde{d}_i^{>}(a)\right)$$

avec

$$\widetilde{y_i}(a) = a_i\, y_i(a) \text{ pour tout } y \in \{c^{\leq},\, c^{>},\, d^{\leq},\, d^{>}\},$$

on cherchera également à estimer un paramètre $\tilde{\nu}(i|l)$ nous donnant une approximation de la forme

$$\tilde{f}_i^g(a) \approx \tilde{c}_i^{\leq}(a) + \left(\frac{1}{\mathrm{Re}} + \tilde{\nu}(i|l)\right)\tilde{d}_i^{\leq}(a). \tag{3.6}$$

Bien que les formulations (3.5) et (3.6) se ressemblent, les estimations par moyenne temporelle et moindres carrés de $\nu(i|l)$ et $\tilde{\nu}(i|l)$ seront différentes en pratique, puisqu'il faut multiplier (3.5) par un terme qui dépend du temps (à savoir a_i) pour obtenir (3.6).

Dans les deux sections suivantes, nous garderons toujours une notation sans tilde afin d'exposer les estimations de $\nu(i|l)$ qui ont été calculées, mais ces estimations sont transposables au cas de la formulation (3.6). Puis, nous présenterons les estimations calculées pour les deux formulations : "dynamique" (3.5) et "énergétique" (3.6).

Approximation en moyenne

On dispose des valeurs des coefficients temporels de la POD et de leurs dérivées obtenues par différences finies pour les N clichés : $a_i^e(t_j)$ et $\dot{a}_i^e(t_j)$ pour $1 \leq i \leq M$ et $1 \leq j \leq N$. Le symbole $\approx$ correspond ici à l'égalité en moyenne. Pour une coupure $l \leq M$ quelconque, l'estimation de $\nu(i|l)$, notée $\underline{\nu}_i$, doit alors satisfaire la relation

$$\langle f_i^g(a)\rangle = \left\langle c_i^{\leq}(a) + \left(\frac{1}{\mathrm{Re}} + \underline{\nu}_i\right) d_i^{\leq}(a)\right\rangle,$$

où $\langle . \rangle$ est toujours l'opérateur de moyenne temporelle défini par une moyenne arithmétique (équation (3.2)). On a alors tout simplement

$$\underline{\nu}_i = \frac{\langle f_i^{>}(a)\rangle}{\langle d_i^{\leq}(a)\rangle}.$$

On peut décomposer $\underline{\nu}_i$ en :

- une contribution "visqueuse" $\underline{\nu}_i^{visq.} = \dfrac{\langle d_i^{>}(a)\rangle}{\mathrm{Re}\ \langle d_i^{\leq}(a)\rangle}$;
- et une contribution "non visqueuse" $\underline{\nu}_i - \underline{\nu}_i^{visq.} = \dfrac{\langle c_i^{>}(a)\rangle}{\langle d_i^{\leq}(a)\rangle}$.

Approximation par moindres carrés

Le symbole $\approx$ correspond maintenant à une approximation par moindres carrés. L'estimation de $\nu(i|l)$, notée ν_i, est alors la solution du problème

$$\min_{\alpha\in\mathbb{R}} \quad \sum_{j=1}^{N} \left(f_i^g(a(t_j)) - c_i^{<}(a(t_j)) - \left(\frac{1}{\mathrm{Re}} + \alpha\right) d_i^{<}(a(t_j)) \right)^2, \tag{3.7}$$

ou de manière équivalente

$$\min_{\alpha\in\mathbb{R}} \quad \left\langle \left(f_i^g(a) - c_i^{<}(a) - \left(\frac{1}{\mathrm{Re}} + \alpha\right) d_i^{<}(a) \right)^2 \right\rangle.$$

Elle est donnée par

$$\nu_i = \frac{\displaystyle\sum_{j=1}^{N} d_i^{<}(a(t_j))\, f_i^{>}(a(t_j))}{\displaystyle\sum_{j=1}^{N} d_i^{<}(a(t_j))^2} = \frac{\langle d_i^{<}(a)\, f_i^{>}(a)\rangle}{\left\langle d_i^{<}(a)^2 \right\rangle}.$$

Il est encore possible de décomposer ν_i en :

- une contribution "visqueuse" $\nu_i^{visq.} = \dfrac{\langle d_i^{<}(a)\, d_i^{>}(a)\rangle}{\mathrm{Re}\left\langle d_i^{<}(a)^2 \right\rangle}$;
- et une contribution "non visqueuse" $\nu_i - \nu_i^{visq.} = \dfrac{\langle d_i^{<}(a)\, c_i^{>}(a)\rangle}{\left\langle d_i^{<}(a)^2 \right\rangle}$.

Résultats

Quatre familles d'estimations ont donc été calculées à partir du modèle f^g à $M = 86$ modes et des coefficients temporels $a = a_i^e$ de référence donnés par la POD :

- par moyenne via la formulation "dynamique" :

$$\underline{\nu_i} = \frac{\langle f_i^{>}(a^e)\rangle}{\langle d_i^{<}(a^e)\rangle}\,; \tag{3.8}$$

- par moyenne via la formulation "énergétique" :

$$\underline{\widetilde{\nu_i}} = \frac{\langle a_i f_i^{>}(a^e)\rangle}{\langle a_i d_i^{<}(a^e)\rangle}\,; \tag{3.9}$$

- par moindres carrés via la formulation "dynamique" :

$$\nu_i = \frac{\langle f_i^{>}(a^e) d_i^{<}(a^e)\rangle}{\langle (d_i^{<}(a^e))^2\rangle}\,; \tag{3.10}$$

- par moindres carrés via la formulation "énergétique" :

$$\widetilde{\nu_i} = \frac{\langle a_i^2 f_i^{>}(a^e) d_i^{\leq}(a^e) \rangle}{\langle (a_i d_i^{\leq}(a^e))^2 \rangle}. \tag{3.11}$$

Notons que les modèles usuels de viscosité artificielle pour la SGE sont basés sur une équation de bilan de l'énergie cinétique de la partie résolue de l'écoulement, et qu'il y a donc une analogie avec les fermetures (3.9) et (3.11).

En pratique, les contributions visqueuses sont négligeables (le nombre de Reynolds étant relativement important), et on ne présentera donc pas indépendamment chaque contribution. De plus, les valeurs de $\underline{\nu_i}$ se sont avérées inutilisables car particulièrement erratiques ; elles n'apparaissent pas parmi les résultats présentés en figure 3.20. Pour plus de lisibilité,

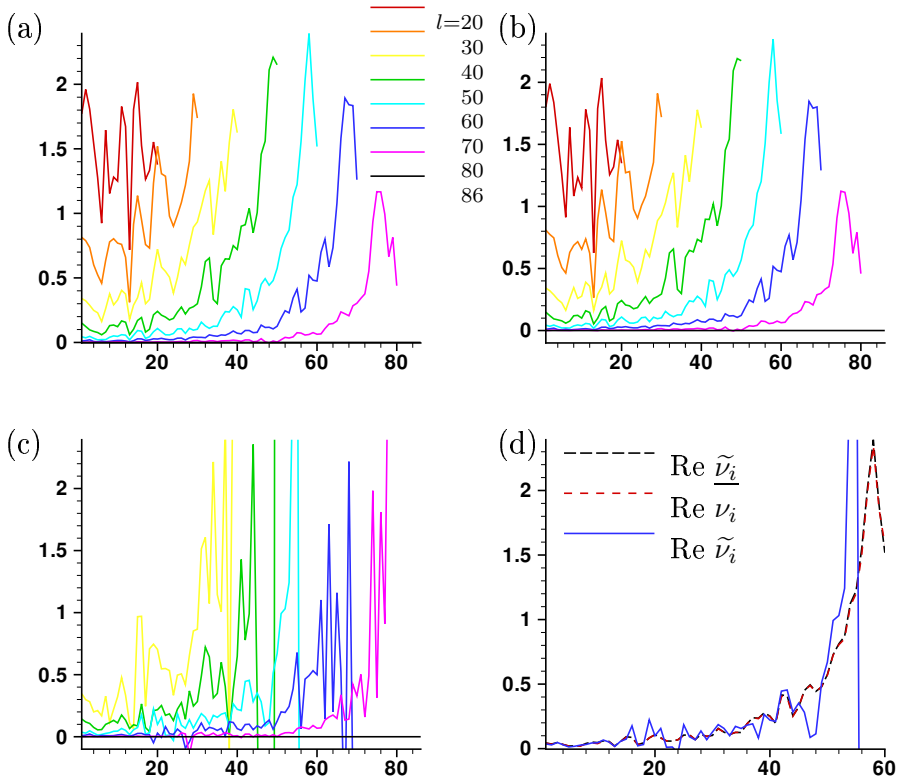


FIG. 3.20 – Viscosités artificielles calculées pour huit indices de coupure (l = 20, 30, 40, 50, 60, 70, 80 et $l = M = 86$) : (a) Re $\underline{\widetilde{\nu_i}}$, (b) Re ν_i et (c) Re $\widetilde{\nu_i}$. (d) Comparaison de Re $\underline{\widetilde{\nu_i}}$, Re ν_i et Re $\widetilde{\nu_i}$ pour $l = 60$.

nous avons conservé pour représenter Re $\widetilde{\nu_i}$ l'intervalle des ordonnées donné par Re $\underline{\widetilde{\nu_i}}$ et Re ν_i.

Pour $l = M$, toutes les viscosités obtenues sont nulles comme il était attendu puisque aucun terme n'est tronqué : $f_i^> = c_i^> = d_i^> = \tilde{f}_i^> = \tilde{c}_i^> = \tilde{d}_i^> = 0$ si $l = M$. Notons que $\underline{\widetilde{\nu}_i}$, ν_i et $\widetilde{\nu}_i$ donnent des profils semblables, qui partagent plusieurs caractéristiques avec la viscosité théorique donnée par l'analyse de Fourier d'un spectre de Kolmogorov sur lequel est opéré un filtre porte (*sharp cut-off filter*) :

- présence d'un pic au niveau de la coupure (pour les derniers indices proches de l) ;
- pour chaque définition de la viscosité artificielle et pour un indice i fixé, les valeurs décroissent quand l'indice l de coupure augmente ;
- les hauteurs des pics pour différentes coupures sont du même ordre de grandeur.

Par exemple, Schilling *et al.* [86] obtiennent les profils de viscosités de la figure 3.21 pour la turbulence incompressible, tridimensionnelle et isotrope pour un filtrage spectral porte.

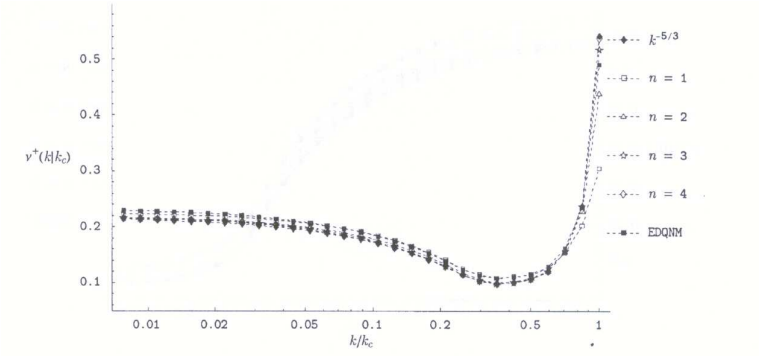


FIG. 3.21 – *Total eddy viscosities $\nu^+(k|k_c)$, as a function of k/k_c with $k_c = 128$ for the Kolmogorov, Leslie-Quarini and EDQNM spectra.* Tiré de [86] : k correspond aux nombres d'onde et k_c à la coupure.

Ces observations sont cohérentes avec les similarités déjà mises en évidence entre modes POD et modes de Fourier.

Les estimations $\widetilde{\nu}_i$ présentent des profils moins lisses que les deux autres définitions avec des variations locales assez marquées pour les indices proches de l, qui peuvent même atteindre des valeurs négatives. Ceci peut s'expliquer par la complexité des interactions modales et notamment par l'existence ponctuelle de petites cascades inverses d'énergie. Toutefois, l'allure globale des profils reste identique.

Il faut souligner que les courbes de $\underline{\widetilde{\nu}_i}$ et ν_i sont quasiment superposables, bien qu'elles correspondent à des définitions très différentes : $\underline{\widetilde{\nu}_i}$ provient d'une approximation moyenne via la formulation "énergétique", tandis que ν_i provient d'une approximation par moindres carrés via la formulation "dynamique". Cette constatation *a posteriori* est très intéressante puisqu'elle renforce rétrospectivement l'intérêt à accorder à l'étude des transferts d'énergie cinétique précédente, en établissant numériquement une correspondance entre, d'une part, les bilans d'énergie entre modes résolus (d'indice $i \leq l$) et non résolus (d'indice $i > l$) qui correspondent à $\underline{\widetilde{\nu}_i}$, et, d'autre part, une viscosité $\nu(i|l)$, qui, une fois injectée, devrait améliorer la dynamique du système à $M - l$ modes de la forme (3.1) au sens de la définition

de ν_i. En effet, l'objectif à plus longue échéance de ce type de travail est de proposer des modèles cohérents et efficaces de prise en compte des modes non résolus dans un système dynamique POD-Galerkine.

Remarque sur le choix de la référence à partir de laquelle a été calculée la paramétrisation

Les viscosités artificielles calculées précédemment ont été définies en prenant pour référence le modèle POD-Galerkine à $M = 86$ modes pour deux raisons : ceci permettait d'une part de valider les résultats obtenus dans la mesure où les viscosités calculées devaient être exactement nulles pour $l = M$. Et surtout, les termes modélisés par les ajouts de viscosités sont parfaitement connus : ce sont les termes tronqués lorsque l'on passe à une modélisation de M modes à l modes, c'est-à-dire $f_i^>$ ou $\tilde{f}_i^>$ pour le $i$ème mode.

Un autre choix possible aurait été de modéliser tout ce que le modèle à l modes ne prend pas en compte, ce qui reviendrait à remplacer $f_i^>$ par $\dot{a}_i^e - f_i^\leq(a)$ dans les équations (3.8) et (3.10) et $\tilde{f}_i^>$ par $a_i\,(\dot{a}_i^e - f_i^\leq(a))$ dans les équations (3.9) et (3.11). Cependant, il n'aurait pas été possible d'analyser les résultats en termes de modélisation des modes tronqués.

En effet, le modèle POD-Galerkine a été calculé en négligeant les termes sous-maille de la SGE et le terme $T_{\Gamma_{\text{in}}}$ (voir la section 3.3.1) : la paramétrisation aurait alors également concerné ces termes, ce qui aurait pu fausser l'analyse.

Notons cependant que la modélisation globale des échelles "non-résolues" (sous-mailles et modes tronqués) et des termes relevant de la pression est intéressante. Elle sera d'ailleurs abordée de manière globale et similaire dans la suite : nous avons ici effectué une recalibration des coefficients d'origine visqueuse par des ajouts de viscosité qui dépendent des modes via notamment la définition d'un problème de minimisation aux moindres carrés (équation (3.7)), tandis que dans le chapitre 4 nous allons nous intéresser à la recalibration globale des coefficients polynômiaux via la définition d'un problème de minimisation plus général.

3.5.3 Lien entre $\langle\Pi\rangle$ (transferts énergétiques moyens) et $\underline{\widetilde{\nu_i}}$ (paramètre visqueux)

On a

$$a_i\, c_i^>(a) = \sum_{k=l+1}^{M} \left(C_i^k\, a_k a_i + \frac{\Pi(i|k)}{\lambda_i} \right).$$

Comme $\langle a_i a_j \rangle = \delta_{i,j}$, on obtient

$$\langle a_i\, c_i^>(a) \rangle = \frac{1}{\lambda_i} \sum_{k=l+1}^{M} \langle \Pi(i|k) \rangle\,.$$

De plus,

$$\langle a_i\, d_i^\leq(a) \rangle = D_i^0\, \langle a_i \rangle + D_i^i.$$

Et finalement, puisque $\langle a_i \, d_i^{>}(a) \rangle = 0$ pour $i \leq l$, on obtient

$$\underline{\widetilde{\nu_i}} = \frac{1}{D(i)} \sum_{k=l+1}^{M} \langle \Pi(i|k) \rangle \, ,$$

où $D(i) = \lambda_i \left(D_i^0 \langle a_i \rangle + D_i^i \right)$ ne dépend pas de l. Ainsi, en admettant que $D(i) < 0$, on déduit que

- $\langle \Pi(i|k) \rangle > 0$ si $i < k \Longrightarrow \underline{\widetilde{\nu_i}} > 0$: puisque le phénomène moyen est une cascade directe d'énergie, l'estimation visqueuse par moyenne via la formulation énergétique est positive ;
- $\langle \Pi(i|k) \rangle > 0$ si $i < k \Longrightarrow \left[l_1 < l_2 \Longrightarrow \underline{\widetilde{\nu_i}}(i|l_1) > \underline{\widetilde{\nu_i}}(i|l_2) \right]$: la direction moyenne des transferts vers les modes d'indice supérieur implique que $\underline{\widetilde{\nu_i}}$ décroît avec l pour un i fixé ;
- $\langle \Pi(i|k) \rangle = 0$ si $|i-k| > p \Longrightarrow \underline{\widetilde{\nu_i}} = 0$ si $l - i > p - 1 (l \geq i)$: la propriété de localité de $\langle \Pi \rangle$ semble liée au profil de $\underline{\widetilde{\nu_i}}$.

Ainsi, les propriétés principales de la paramétrisation visqueuse $\underline{\widetilde{\nu_i}}$ sont cohérentes avec la dominance d'une cascade d'énergie directe au sein des modes POD.

3.6 Conclusions

Dans ce chapitre, nous avons mis en évidence que les transferts d'énergie cinétique entre les modes POD de notre écoulement partagent de nombreuses similarités avec ceux habituellement observés dans une décomposition de Fourier :

- ils sont locaux au sens où $\langle \Pi(i|j) \rangle$ diminue rapidement quand $|i-j|$ augmente ;
- une cascade directe d'énergie domine puisque, en moyenne, un mode d'indice i draine de l'énergie des modes d'indice $j < i$ et en redistribue à ceux d'indice $j > i$;
- néanmoins une cascade inverse a été mise en évidence, le signe de $\Pi(i|j)$ n'étant pas constant au cours du temps.

Ces observations ne sont pas surprenantes, puisque les modes sont classés par énergie cinétique moyenne décroissante, et sont donc en pratique associés à des structures spatiales de plus en plus petites, à l'instar des modes de Fourier. Les interactions entre les modes POD calculés sur l'ensemble du domaine spatial présentent des propriétés analogues au modèle donné par les modes de Fourier, malgré le caractère non homogène de l'écoulement.

En conséquence, la paramétrisation visqueuse $\underline{\widetilde{\nu_i}}$ des transferts entre modes résolus et non résolus, estimée par moyenne via la formulation "énergétique", présente d'importantes similarités avec son analogue dans l'espace de Fourier (par exemple, pic de viscosité au niveau de la coupure dans le cas d'un filtrage spatial de type porte). On retrouve ces profils particuliers de viscosité par deux autres paramétrisations, notamment ν_i qui découle d'une estimation par moindres carrés de la formulation "dynamique".

Une modélisation de prise en compte des effets des modes POD tronqués doit être en mesure de reproduire la complexité des interactions modales mise en évidence, ou tout

du moins son phénomène essentiel : la cascade directe d'énergie. Nos résultats suggèrent, dans l'hypothèse d'une modélisation des effets des modes POD tronqués de type visqueuse (c'est-à-dire intervenant au niveau des termes linéaires d'origine visqueuse du système POD-Galerkine), qu'elle dépende de l'indice i et de la coupure l.

Si une modélisation visqueuse de la forme proposée à la section 3.5.2 paraît pertinente pour le phénomène prédominant de cascade directe d'énergie, il semble utile de développer une modélisation plus complexe, qui soit plus fidèle à la complexité des interactions modales (existence de cascades inverses) et en mesure de rendre compte de l'effet des termes de pression qui peuvent poser problème.

Par ailleurs, des ajouts artificiels de viscosité dépendant de l'indice i du mode considéré permettant de corriger un système POD-Galerkine réduit ont également été proposés par Sirisup *et al.* [87] afin d'améliorer son comportement sur une longue durée, et validés pour un écoulement bidimensionnel laminaire : la recalibration des coefficients polynômiaux qu'ils préconisent est plus complexe puisqu'ils proposent une viscosité artificielle qui est adaptée à chaque mode d'indice $i \leq M$ mais aussi à chaque monôme de f_i^g de degré strictement inférieur à un.

Dans le chapitre suivant, la calibration de tous les coefficients polynômiaux est étudiée et le caractère local des interactions entre les modes POD sera exploitée lors de la définition des méthodes qualifiées de *partial-Galerkin*.

Publication

Les résultats principaux de ce chapitre ont fait l'objet de la publication [16] et ont été exploités par Noack *et al.* [67].

Chapitre 4

Calibration des modèles dynamiques réduits polynômiaux

Sommaire

Motivations et objectifs

Le chapitre précédent a montré qu'il était cohérent de modéliser l'effet des modes POD tronqués par des ajouts artificiels de viscosité. Les pseudo-viscosités calculées dépendaient du mode considéré, étaient indépendantes du temps, n'intervenaient qu'au niveau des coefficients d'origine visqueuse du modèle (leur introduction ne modifie pas les termes quadratiques) et ne tenaient compte que des effets des modes POD négligés. En outre, l'action de l'environnement sur l'écoulement était quasi-indépendant du temps : les conditions de flux étaient homogènes, les forces extérieures $\boldsymbol{h}$ nulles et la condition de Dirichlet en entrée quasi-stationnaire, les modes POD interagissant dans le modèle avec un champ $\bar{\boldsymbol{u}}$ constant. Néanmoins, il faut noter qu'*a priori* les modes POD interagissent soit principalement avec les premiers modes POD et $\bar{\boldsymbol{u}}$, soit avec les modes tronqués dans la modélisation, si $\bar{\boldsymbol{u}}$ ne contient pas de petites structures spatiales : les conclusions de l'étude précédente devraient encore être valables dans le cas d'un champ $\bar{\boldsymbol{u}}$ instationnaire. En définitive, l'introduction de pseudo-viscosités indépendantes du temps et ne modifiant pas les termes quadratiques paraît d'un intérêt pratique limité, même si elle est cohérente avec la cascade directe d'énergie observée et l'effet moyen des modes négligés pour l'écoulement étudié au chapitre 3.

Il apparaît donc plus intéressant de proposer une modélisation plus complexe des modes tronqués et, plus généralement, une modification plus radicale des modèles POD-Galerkine réduits. Ceci va être tenté ici par une calibration de tous les coefficients polynômiaux du modèle POD-Galerkine via un problème d'optimisation défini à partir des données temporelles de la POD.

En effet, une telle calibration présente *a priori* de nombreuses qualités :

- elle permet de traiter de manière globale
  - la modélisation des petites échelles de l'écoulement (échelles sous-maille et modes POD négligés),
  - l'éventuelle prise en compte du terme de pression si toutes les conditions de Dirichlet n'ont pas été traitées explicitement (voir la section 2.3.1)
  - et la calibration des coefficients du modèle afin de corriger les erreurs numériques (le modèle peut être très sensible à une petite perturbation d'un de ses coefficients) ou de combler le décalage qui existe entre la nature des données et la formulation variationnelle (FVNSI) utilisée (équation (2.27)),
- elle est plus générale qu'une simple calibration des coefficients visqueux (ce qui était en quelque sorte le cas lors de la paramétrisation effectuée dans la section 3.5.2), elle devrait donc au moins permettre de recouvrer l'effet moyen de modes POD tronqués,
- le fait de modifier les coefficients quadratiques équivaut à introduire des viscosités instationnaires qui dépendent linéairement des coefficients temporels $a_i(t)$ des modes POD, ce qui paraît nettement plus intéressant que des viscosités constantes vue la complexité des interactions modales et, en particulier, le fait que les transferts d'énergie entre modes varient non seulement d'intensité mais de sens (l'existence ponctuelle de cascades inverses a été mise en évidence dans la section 3.5.1),
- elle est cohérente avec les suggestions de Galletti *et al.* [25] (voir la section 2.3.1) et de Noack *et al.* [67] qui préconisent de modéliser les termes relevant de la pression

(qui n'auraient pas été traités lors de la construction du modèle) par une calibration des coefficients linéaires et, respectivement, quadratiques.

De plus, elle présente deux gros avantages pratiques :

- la calibration du modèle ne modifie pas sa forme polynômiale (elle n'augmente donc pas le coût informatique d'exploitation du modèle) ;
- les méthodes de calibration qui vont être développées pourront être appliquées à la calibration de tout modèle polynômial, par exemple au modèle fluide compressible de Vigo (qui a été présenté dans la section 2.3.2), ou même à un modèle d'EDOs n'entrant pas dans le champ de la mécanique des fluides.

En outre, il faut noter que les méthodes de calibration proposées ici reposent sur le principe suivant : optimiser les coefficients du modèle d'après l'histoire des coefficients temporels a_i^e de la POD. Il faut noter que notre approche est, en ce sens, similaire à celles proposées par Galletti [25] ou Delville *et al.* [17] . Ainsi, la partie de l'information fournie par la POD qui est inutilisée par la méthode de Galerkine est prise en compte et la calibration exploite des données beaucoup moins volumineuses en pratique que la méthode de Galerkine : ceci permettra de définir des méthodes numériques dont le coût informatique est raisonnable, si on le compare au coût global de la méthode POD-Galerkine, et qui pourront même, dans certains cas, permettre de diminuer le coût de construction du modèle à partir de données POD (voir les stratégies dites *partial-Galerkin*, sections 4.2.4 et 4.3.3).

Note

Ce chapitre est en majeure partie constitué d'un article en langue anglaise qui a été accepté pour publication dans *Journal of Computational Physics*. Ce papier est donné dans sa version intégrale, il reprend donc sommairement (sections 4.1 et 4.5.1) plusieurs points qui ont déjà été évoqués dans ce mémoire : la définition et les propriétés principales de la POD, la modélisation POD-Galerkine appliquée aux équations de Navier-Stokes incompressibles et la présentation des modéles POD-Galerkine réduits des écoulements de la section 2.5 et du chapitre 3. Ce chapitre a fait également l'objet d'une communication orale [15].

Le papier est organisé de la manière suivante : dans la section 4.1, la modélisation POD-Galerkine est succintement présentée et appliquée à la modélisation des écoulements de la section 2.5 et du chapitre 3, qui serviront de configurations de test. Les méthodes de calibration sont ensuite définies et testées dans les sections 4.2 puis 4.3, avant la conclusion de la section 4.4 et les annexes de la section 4.5.

La section 4.6 permet de compléter l'article et de donner une vision plus exhaustive du travail qui a été réalisé durant cette thèse.

Calibrated reduced-order POD-Galerkin system for fluid flow modelling

M. Couplet - C. Basdevant - P. Sagaut

Abstract

To improve the behaviour of reduced-order POD-Galerkin systems, two numerical methods are proposed. These methods determine free parameters in the POD-Galerkin system from flow simulations via a minimization problem. They give rise to linear systems and their computational costs are reasonable. Both methods are assessed for two flow configurations : a two-dimensional flow around a square-cylinder for a Reynolds number of 100 and a three-dimensional flow past a backward-facing step for a Reynolds number of 7 432 based on the step height and the streamwise velocity at the middle of the inlet. For both configurations, the methods are effective since accurate calibrated reduced-order POD-Galerkin systems are obtained.

Introduction

The POD (Proper Orthogonal Decomposition, also known as Karhunen-Loève decomposition and principal component analysis) is a theoretical and post-processing tool to educe global coherent structures of flows thus to describe and analyze them. Moreover, since laminar and transitional fluid flows are very often governed by a small number of coherent structures, it is interesting to use the spatial POD functions, called POD modes, as basis functions for a Galerkin method in order to construct a system of ODEs (Ordinary Differential Equations) which approximates the whole flow dynamics. Therefore, after truncating the POD modal basis by keeping only the main POD modes, a ODE system of small dimension can be extracted from numerical data (see [36] for a survey) : this system is called (reduced-order) POD-Galerkin system.

As an effective technique of low-order modelling, the POD-Galerkin method is attractive for flow control (see [75, 74, 19]). However, especially for transitional and turbulent flows, the low-order ODE systems obtained may be barely accurate and even sometimes unstable. A first reason is the intrinsic gap that may exist between the nature of the data whose POD is performed and the variational formulation on which the Galerkin method is based. For instance, experimental data of a flow at low Mach number may not satisfy very accurately the variational formulation derived from the incompressible Navier-Stokes equations, or data provided by Finite-Volume codes are not computed via the numerical discretization of a variational formulation. This can bring about a lack of effectiveness of the POD-Galerkin method in practice in numerous cases.

A second and main reason is that the low-order truncation of the POD basis inhibits

generally all the transfers between the large and the small (unresolved) scales of the fluid flow. In consequence, to recover the effects of the truncated modes, that is generally of the small scales, two different ways have been studied in the literature : the definition of POD in the H^1 Sobolev space rather than in L^2 [38], or the use of "eddy" viscosities [6, 70]. That use of artificial viscosities, whose relevance was investigated for a separated non-homogeneous turbulent flow in [16], amounts to perturbing the viscous terms of the POD-Galerkin system. As mentioned by Sirisup *et al.* [87], the add of artificial viscosities remains interesting for two-dimensional flows with Reynolds numbers of the order of 100 for correcting the long-term behaviour of POD-Galerkin systems. Furthermore, if the boundary conditions are disregarded when the Galerkin method is applied (it can be indeed difficult to deal with complex unsteady Dirichlet conditions in practice, see the appendix 4.5.1 for explanations), the modelling of the pressure term may pose problems. Moreover, the computation of a POD-Galerkin system is subject to numerical errors, which may be very prejudicial since such an ODE system can be very sensitive. So it appears interesting to develop methods to increase the accuracy of reduced-order POD-Galerkin models and to improve the modelling skills of the POD-Galerkin method.

To correct the behaviour of a low-order POD-Galerkin system, two numerical methods are here proposed and assessed. They consist in adjusting the polynomial coefficients which define the POD-Galerkin system by solving a minimization problem : the new ODE system has to recover optimally the dynamics of the data used to construct the POD and is computed in taking the original POD-Galerkin system into account.
A similar principle was investigated by Galletti *et al.* [25] to calculate some linear models of terms relevant to the pressure in the construction of a POD-Galerkin system of laminar flow regimes past a square cylinder. Here, we propose to modify all the coefficients (linear and quadratic) of the POD-Galerkin system to improve it. Moreover, the methods are assessed on a turbulent flow configuration where the main challenge is the modelling of the effect of the truncated POD modes (that is the small scales). Furthermore, it is worthy of note that the cost functions used in the present paper are designed to control the way the initial POD-Galerkin system is modified (thanks to the cost function $\mathcal{D}$, see section 4.2.1), which is very interesting in practice (refer to the observations of sections 4.3.2 and 4.3.3). The computational costs of these methods are reasonable since they use the temporal part of the POD information, whereas the Galerkin method uses the spatial POD information (that is the POD modes themselves), much more voluminous in practice, to construct the ODE system (see section 4.2.4 for further details). This motivates the development of calibration methods : a clever use of the spatial POD information could enable large computational saving for reduced-order system from very demanding computations.

The reduced-order POD-Galerkin fluid flow modelling is described and applied to two configurations in section 4.1 : to a two-dimensional quasi-incompressible laminar flow around a square-cylinder and to a three-dimensional incompressible turbulent flow past a backward-facing step. The general principle of the calibration of low-order POD-Galerkin systems is formally presented then the numerical methods are proposed in section 4.2. These calibration methods are finally assessed in section 4.3 for both flow configurations.

4.1 The reduced-order POD Galerkin modelling

In this section, the POD-Galerkin method is described for the modelling of an incompressible flow (details about treatment of the boundary conditions are displayed in appendix 4.5.1). Then few results for the two test flow configurations which were used in our numerical experiments are presented.

4.1.1 POD-Galerkin method for incompressible flows

The POD is applied to a velocity field $\boldsymbol{u^e} \in \mathbb{R}^d$ which is known over the time interval $[0,T]$ and the physical space $\Omega \subset \mathbb{R}^d$ (d=2 or 3). That is $\boldsymbol{u^e} \in \mathrm{L}^2(0,T,\mathrm{L}^2(\Omega)^d)$ is decomposed into an orthonormal basis of spatial functions $\boldsymbol{\varphi_i}$ of $\mathrm{L}^2(\Omega)^d$, called POD modes, for each time $t \in [0,T]$:

$$\boldsymbol{u^e}(\boldsymbol{x},t) = \sum_i \underbrace{(\ \boldsymbol{u^e}, \boldsymbol{\varphi_i}\ )}_{a_i^e(t)}\ \boldsymbol{\varphi_i}(x) \quad \forall (\boldsymbol{x},t) \in \Omega \times [0,T], \tag{4.1}$$

where $(\ .,.\ )$ is the classical $\mathrm{L}^2(\Omega)^d$ inner product on the flow domain and where the $a_i^e(t)$ are the time-dependent coefficients of the decomposition. For an incompressible flow (with a unitary constant density), the mean kinetic energy per mass unit captured by the ith POD mode $\boldsymbol{\varphi_i}$ is

$$\frac{1}{T}\int_0^T a_i^e(t)^2\, dt = \lambda_i \tag{4.2}$$

and the basis is ordered such that $\lambda_i \geq \lambda_{i+1}$ for all i. In all the following, σ_i will denote $\sqrt{\lambda_i}$. The POD basis is constructed to be optimal in the sense that, for any M and any orthonormal tuple $(\boldsymbol{\psi_1}, \cdots, \boldsymbol{\psi_M})$ of $\mathrm{L}^2(\Omega)^{d\,M}$,

$$\frac{1}{T}\int_0^T \left\| \boldsymbol{u^e} - \sum_{i=1}^{M} (\boldsymbol{u^e}, \boldsymbol{\varphi_i})\ \boldsymbol{\varphi_i} \right\|_{\mathrm{L}^2}^2 dt \leq \frac{1}{T}\int_0^T \left\| \boldsymbol{u^e} - \sum_{i=1}^{M} (\boldsymbol{u^e}, \boldsymbol{\psi_i})\ \boldsymbol{\psi_i} \right\|_{\mathrm{L}^2}^2 dt. \tag{4.3}$$

From a physical point of view, the first M modes, $\boldsymbol{\varphi_i}$ for $i \in \{1, \cdots, M\}$, capture more kinetic energy of $\boldsymbol{u^e}$ on average over $[0,T]$ than any other set of M orthonormal spatial functions. Since the kinetic energy captured by the first M modes is $\sum_{i=1}^{M} \lambda_i$, the decrease of the POD spectrum, that is of the distribution of the λ_i with respect to the index i, quantifies the efficiency of the POD.
This optimality property of the POD explains why the first M POD modes, for M large enough, are interesting candidates for a Galerkin method. The POD-Galerkin system obtained is labeled as reduced-order since the POD basis is truncated by neglecting the POD modes $\boldsymbol{\varphi_i}$ for $i > M$.

The POD-Galerkin system is constructed by applying the Galerkin method, using the space spanned by the first M POD modes. The variational formulation is generally deduced from the velocity-pressure expressions of the non-dimensional Navier-Stokes equations,

considering a solenoidal test function $\boldsymbol{\varphi}$ (indeed, the POD modes keep some properties of $\boldsymbol{u^e}$ as vanishing divergence) :

$$\frac{d}{dt}\left(\boldsymbol{u},\boldsymbol{\varphi}\right)+\left(\left(\boldsymbol{u}.\boldsymbol{\nabla}\right)\boldsymbol{u},\boldsymbol{\varphi}\right)+\frac{1}{\mathrm{Re}}\sum_{i=1}^{d}\left(\boldsymbol{\nabla}u_{x_i},\boldsymbol{\nabla}\varphi_{x_i}\right)+T_{\partial\Omega}=\left(\boldsymbol{h},\boldsymbol{\varphi}\right) \tag{4.4}$$

where $\boldsymbol{u}$ is the velocity field, Re is the Reynolds number, $\boldsymbol{h}$ is a source term (force field independent of the flow), $u_{x_i}=\boldsymbol{u}.\boldsymbol{x_i}$ and $\varphi_{x_i}=\boldsymbol{\varphi}.\boldsymbol{x_i}$ are the component of $\boldsymbol{u}$ and $\boldsymbol{\varphi}$ in the spatial direction of the unitary vector $\boldsymbol{x_i}$ $(1\leq i\leq d)$ and where $T_{\partial\Omega}$ is a boundary term (consult the appendix 4.5.1 for further details about it and the explicit treatment of the boundary conditions).
Therefore, under the assumption that the reduced POD basis is suitable (that is the M first modes are sufficient to accurately represent the flow), a M-dimensional polynomial ODE system is derived by taking the M first POD modes as basis and test functions. More precisely, if $\boldsymbol{h}=\boldsymbol{0}$ and $T_{\partial\Omega}=0$, the POD-Galerkin system is

$$\dot{a}^g(t)=f^g(a^g(t))=\begin{bmatrix} f_1^g(a^g(t)) \\ \vdots \\ f_M^g(a^g(t)) \end{bmatrix} \quad \text{where} \quad a^g(t)=\begin{bmatrix} a_1^g(t) \\ \vdots \\ a_M^g(t) \end{bmatrix} \mathbb{R}^M, \tag{4.5}$$

and each polynomial f_i^g can be expressed as

$$f_i^g(a^g)=\sum_{k=1}^{M}\frac{C_i^k}{\mathrm{Re}}\,a_k^g+\sum_{k=1}^{M}\sum_{j=1}^{M}C_i^{k,j}\,a_k^g a_j^g \tag{4.6}$$

with

$$C_i^k=-\sum_{j=1}^{d}\left(\boldsymbol{\nabla}(\varphi_k)_{x_j},\boldsymbol{\nabla}(\varphi_i)_{x_j}\right) \quad \text{and} \quad C_i^{k,j}=-\left(\left(\boldsymbol{\varphi_k}.\boldsymbol{\nabla}\right)\boldsymbol{\varphi_j},\boldsymbol{\varphi_i}\right). \tag{4.7}$$

Furthermore, when non-homogeneous Dirichlet boundary conditions for the velocity are considered, the POD is generally applied to $\boldsymbol{u^e}(\boldsymbol{x},t)-\bar{\boldsymbol{u}}^{\boldsymbol{e}}(\boldsymbol{x},t)$ instead of $\boldsymbol{u^e}$, with $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ chosen so that $\boldsymbol{u^e}-\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ satisfies homogeneous Dirichlet boundary conditions (see appendix 4.5.1). In that case, the contribution of $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ has to be added to system (4.5) together with the contributions of $T_{\partial\Omega}$ and $\boldsymbol{h}$. In the flow configurations of sections 4.1.2 and 4.1.3, the Dirichlet conditions for the velocity are non-homogeneous yet unsteady and $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ is simply chosen as the mean velocity field : the POD computed corresponds to the fluctuant velocity. In fact, in the case of the turbulent configuration (section 4.1.3), the inlet Dirichlet boundary condition is not strictly unsteady however it is realistic to perform the POD on the fluctuant velocity and to neglect the boundary term since the fluctuant velocity and its POD modes take very small values at the inlet (see [16]).

Finally, the general form of the POD-Galerkin system obtained is

$$\dot{a}^g(t)=f^g(a^g(t))+s(t) \quad \text{and} \quad a^g(0)=a^e(0), \tag{4.8}$$

where $a^e(0)$ are the initial conditions, f^g is a vector polynomial of degree 2 in the components of a^g and s takes the contributions of the source term $\boldsymbol{h}$, of the boundary term $T_{\partial\Omega}$ and of $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ into account.
In the following, the methods are presented for $s = 0$ for the sake of clarity. Moreover, in the modelling of the two test flow configurations used here, $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ is unsteady, $\boldsymbol{h} = 0$ and the boundary term $T_{\partial\Omega}$ vanishes (refer to the appendix 4.5.1), thus s is independent of the time : the constant vector s is simply considered as a term of degree zero of f^g in the numerical experiments of section 4.3.
Notice that a polynomial POD-Galerkin system can also be constructed in the compressible case for a perfect gas by performing the POD on the suitable set of flow variables (the inverse of the density ρ^{-1}, the velocity $\boldsymbol{u}$ and the pressure p) : see [94] or [38]. In consequence, the methods proposed here can be theoretically applied to compressible cases without any modifications although the numerical experiments presented here were only performed for the incompressible configurations of sections 4.1.2 and 4.1.3.
In conclusion, given numerical data for the flow field $\boldsymbol{u}^{\boldsymbol{e}}$, that is, for example, given a set of N snapshots along a time interval $[0, T]$, a POD basis $(\boldsymbol{\varphi_i})_{1\leq i\leq N}$ can be obtained such that $\boldsymbol{u}^{\boldsymbol{e}}(t) = \bar{\boldsymbol{u}}^{\boldsymbol{e}}(t) + \sum_{1\leq i\leq N} a_i^e(t)\ \boldsymbol{\varphi_i}$. The POD-Galerkin system is then formed by evaluating a variational formulation of the governing equations with the M first POD modes. The questions are then : is the solution $a^g(t)$ to (4.8) a good approximation of the true time coefficients $a^e(t)$? And, if not, how to modify (4.8) to improve the model ?

4.1.2 Two-dimensional flow with a Reynolds number of 100

The first test configuration is a quasi-incompressible two-dimensional vortex-shedding flow around a square cylinder for Re = 100. The database was computed by a compressible Finite-Volume code for a Mach number of 10^{-3} (see [63] for details) and is composed of $N = 480$ snapshots of the velocity over one shedding cycle of the Von Kármán vortex street (that is for $[0, T]$). Some vorticity contours of the fluctuant velocity at time $t = T/2$ are plotted on figure 4.1.

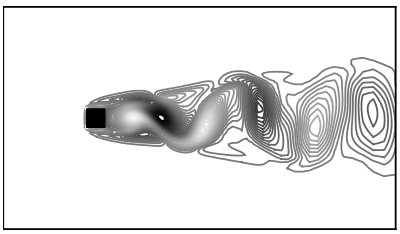


FIG. 4.1 – Iso-lines of the vorticity of the fluctuant velocity at time $t = T/2$ on half the computational domain in the streamwise direction.

From the POD of the fluctuant velocity performed by the snapshot method (see [88]),

a 6-mode POD-Galerkin system was computed. The first six POD modes capture more than 99.9% of the fluctuant kinetic energy of the database $K_N = \sum_{i=1}^{N} \lambda_i$:

i	1	2	3	4	5	6
λ_i/K_N	0.486	0.482	1.214×10^{-2}	1.209×10^{-2}	3.756×10^{-3}	3.744×10^{-3}

The spectrum is plotted in logarithmic scale for the first 16 modes on figure 4.2. Iso-lines

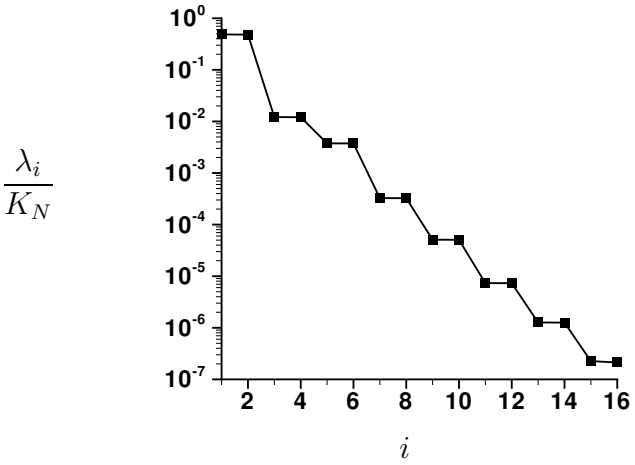


FIG. 4.2 – The 16 first values of the POD spectrum of the square cylinder flow in logarithmic scale.

of the transverse components of the first six modes are plotted on figure 4.3. These figures show that the first POD modes can be naturally grouped by pairs. This is coherent with the nature of the vortex shedding flow and with many preceding works, for instance [66]. Moreover, the flow topology is consistent with the results of the latter paper.

Values of $f_1^g(a^e(t))$ and $f_2^g(a^e(t))$ are compared to $\dot{a}_1^e(t)$ and $\dot{a}_2^e(t)$ in figure 4.4. The histories of $a_1^g(t)$ and $a_2^g(t)$ computed from the 6-mode system are displayed on the same figure with $a_1^e(t)$ and $a_2^e(t)$ (all the simulations of the polynomial POD-Galerkin systems were performed with a classical fourth-order Runge-Kutta scheme).
The 6-mode system gives a good approximation of the dynamics of the two first POD modes, however its simulation shows that the system is not able to reproduce very accurately the history of a^e since there is a noticeable difference between a_2^e and a_2^g at $t = T$. This gap has three causes : the numerical errors, the effects of the truncation of the POD basis and the fact that the data do not perfectly fit the variational formulation (4.4) by nature.

4.1.3 Three-dimensional turbulent flow

The second test configuration is a three-dimensional incompressible turbulent flow past a backward facing step. The database was provided by an incompressible Finite-Difference

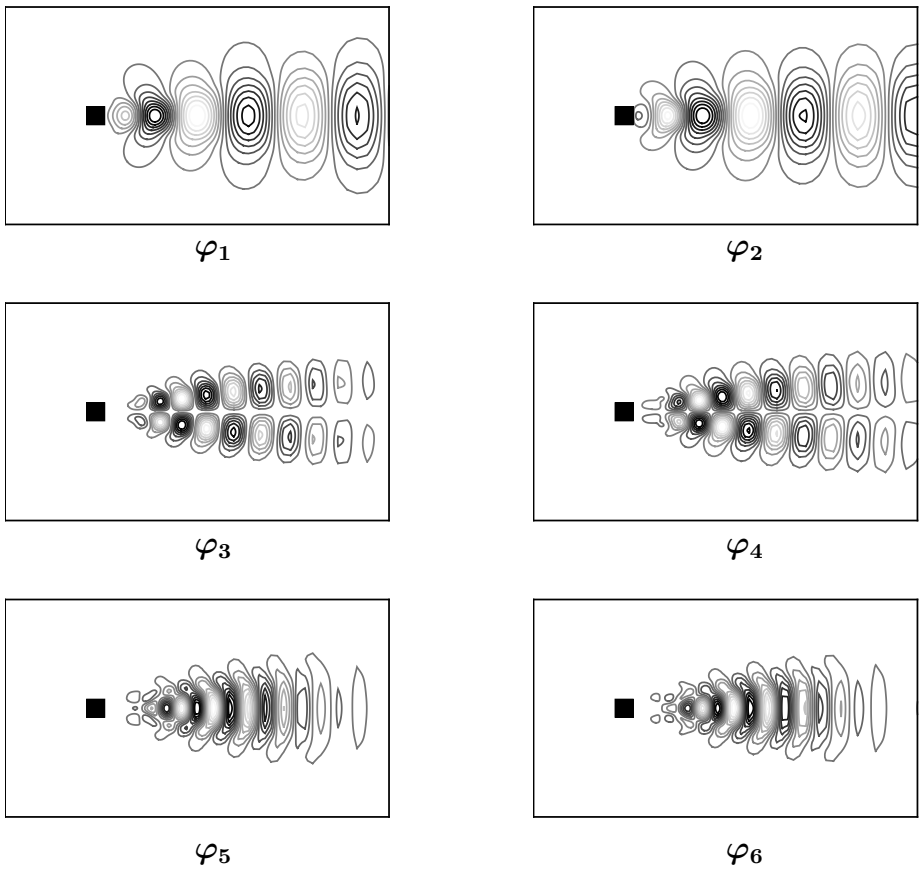


FIG. 4.3 – Iso-lines of the transverse components of φ_1, φ_2, φ_3, φ_4, φ_5 and φ_6 on half the computational domain in the streamwise direction.

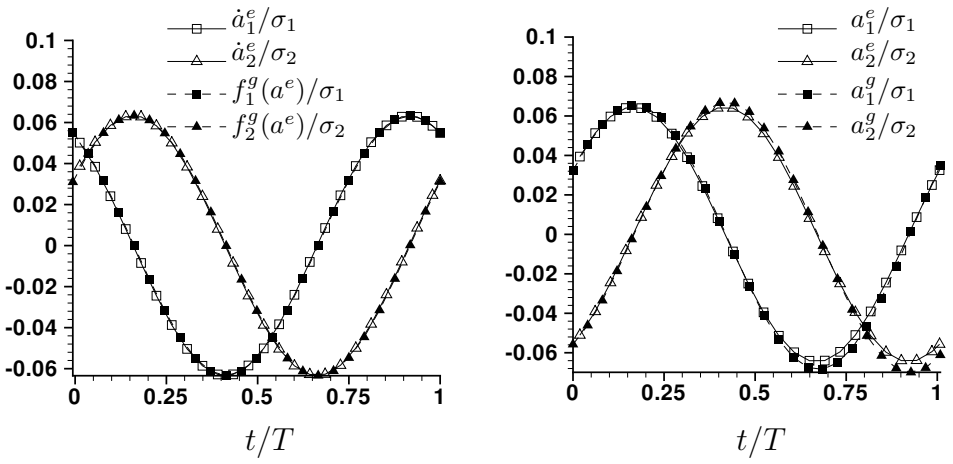


FIG. 4.4 – Comparison of the dynamics (left) and of the history (right) of the data with the behaviour of 6-mode POD Galerkin system for the two first modes.

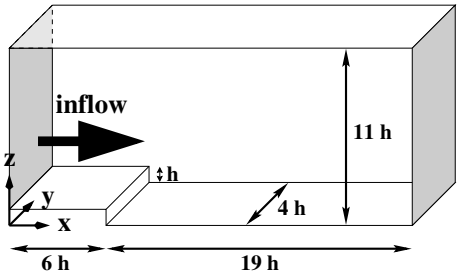


FIG. 4.5 – Geometry of the computational domain which corresponds to the spatial extent of the POD modes.

Large-Eddy Simulation (see [51]) and is formed of $N = 1000$ snapshots of the velocity for a time interval $[0, T]$ long enough to encompass at least one period of the low-frequency breathing mode of the recirculation bubble : $T = 37.5\,h/U = 50\,h/\bar{U}$ where h is the step height, U the streamwise velocity at the middle of the inlet and $\bar{U}$ the mean streamwise velocity at the entrance. The Reynolds number based on U and h is 7 432 and the one based on $\bar{U}$ and the height $10\,h$ of the channel above the step is 66 000. The geometry of the computational domain is presented on figure 4.5.

The POD of the fluctuant velocity was performed by the snapshot method using the full database. Figure 4.6 shows some Q-isosurfaces of the mean velocity $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ and of the POD modes $\boldsymbol{\varphi_1}$, $\boldsymbol{\varphi_{20}}$ and $\boldsymbol{\varphi_{40}}$. The POD spectrum is presented in logarithmic scale in figure 4.7 : the POD is quite efficient according to the decrease of the spectrum.
Notice that the POD of such a step flow is presented in [41] for a Reynolds number about twice smaller and that a low compactness of the kinetic energy distribution within the POD basis is observed. Three facts make it possible to understand this difference. Firstly, the database of [41] was computed by Direct Numerical Simulation so contains a larger

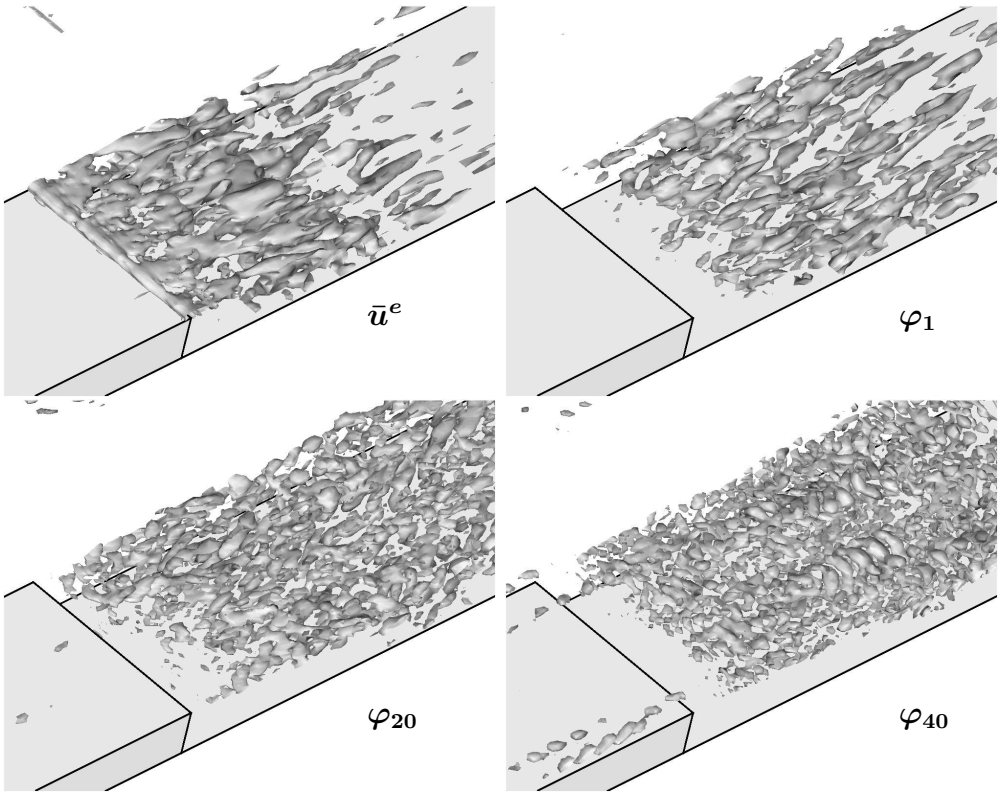


FIG. 4.6 – Visualization of the main structures of $\bar{u}^e$, φ_1, φ_{20} and φ_{40} (Q-isosurface).

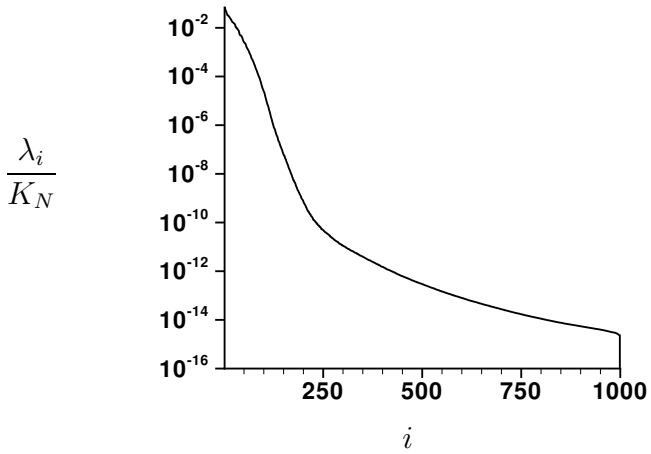


FIG. 4.7 – Logarithmic POD spectrum of the step flow.

range of structures than our LES database. Secondly, the POD is combined with a Fourier decomposition which is not strictly equivalent to a full POD (which is optimal) even if a spectrum close to the POD spectrum is generally expected. Thirdly, the time interval over which the flow is analyzed is twelve times longer in [41], considering some time units based on the step height and on the maximum inlet streamwise velocity in a transverse plane located before the step ($0.07\,h$ and $6\,h$ from the step in [41] and in our case, respectively).

The POD-Galerkin polynomial f^g computed takes the first $M = 86$ modes which capture more than 99.9% of the fluctuant kinetic energy into account.
That POD-Galerkin system was defined as explained in section 4.1.1 and in the appendix 4.5.1 : the system is based on the classical Navier-Stokes equations, not on the filtered Navier-Stokes equations and the subgrid model used to perform the LES of the flow.
Taking the subgrid model into consideration might be a benefit, however disregarding it considerably simplifies the reduced-order modelling ; furthermore and above all, we would like to point out that neglecting the last POD modes, which is the basis of the reduced-order POD-Galerkin modelling, has by nature a stronger impact than neglecting the subgrid model : the structures of the neglected POD modes are necessary larger than the unresolved scales and play in consequence a more important role in the flow dynamics.
That is why the approach chosen here consists in a way in including the modelling of the subgrid scales in the modelling of the neglected POD modes, this will be performed by a calibration of the ODE system.

Figure 4.8 compares the behaviours of the data and of the 86-mode POD Galerkin system for the first and fifth modes. The lack of accuracy of the reduced-order system is obvious for this turbulent three-dimensional configuration. Indeed, a^g diverges from a^e and it seems that the system does not dissipate enough energy.

4.2 Definition of the methods

In the following, the calibration methods designed to improve the POD-Galerkin system are presented for $s = 0$ for the sake of clarity.

4.2.1 The general formulation

The polynomial f^α, which determines the calibrated system we are looking for, will be defined as the solution to an optimization problem. f^α should minimize a functional $\mathcal{J}^\alpha$:

$$\mathcal{J}^\alpha(f) = (1-\alpha)\,\mathcal{E}(f) + \alpha\,\mathcal{D}(f) \tag{4.9}$$

where $\alpha \in [0,1]$ is a weighing parameter, $\mathcal{E}(f)$ measures the "error" between the behaviour of the data, that is the one of $a^e(t)$, and the behaviour of the dynamical system associated to f whose state is $a(t)$ and where $\mathcal{D}(f)$ is a cost linked to the distance between f and f^g. Note that f is a vector polynomial in M variables of degree 2 with M components.

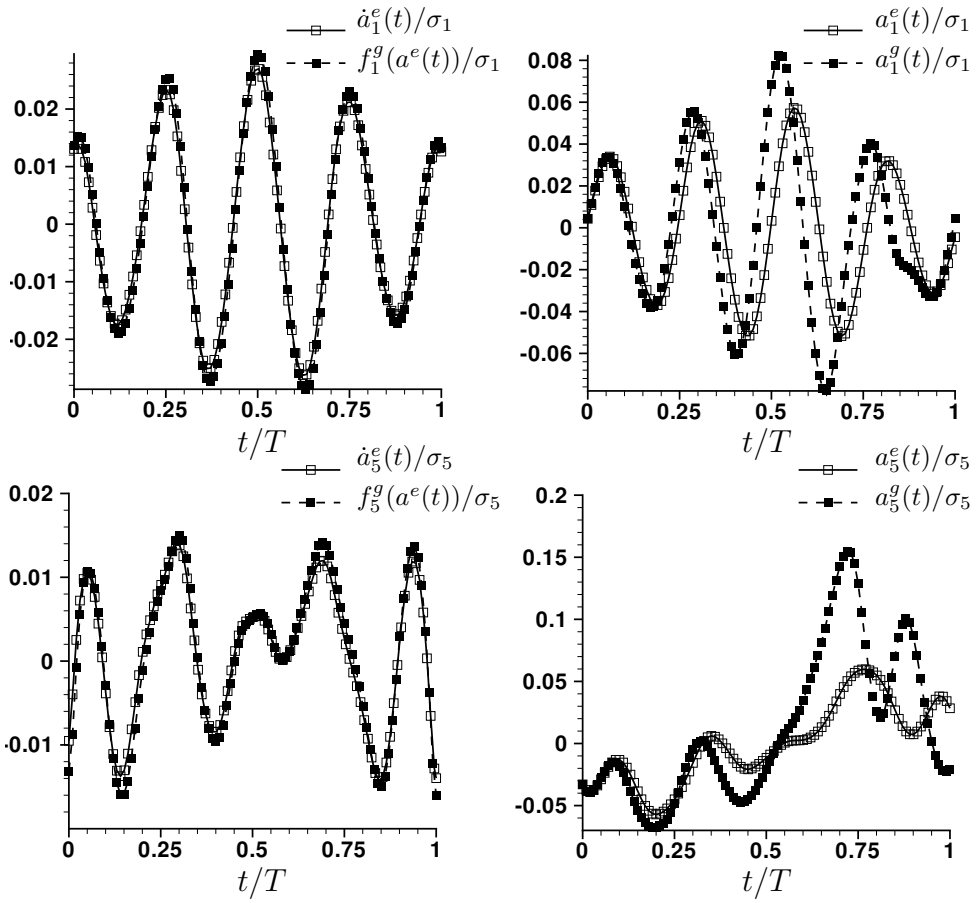


FIG. 4.8 – Comparison of the dynamics (left) and of the history (right) of the data with the behaviour of the POD-Galerkin model of the step flow for the 1st (top) and 5th (bottom) modes.

In the following, $\mathcal{E}$ takes the form

$$\mathcal{E}(f) = \frac{\left\langle \|e(f,t)\|_\Lambda{}^2 \right\rangle}{\left\langle \|e(f^g,t)\|_\Lambda{}^2 \right\rangle} \tag{4.10}$$

where $\langle . \rangle$ is a linear time average operator (discrete or continuous, for instance an integral over $[0,T]$ or an arithmetic average on a subdivision of $[0,T]$), $\|.\|_\Lambda$ is a norm of $\mathbb{R}^M$ and e is an operator with values in $\mathbb{R}^M$. This operator will be defined so that the solution a to the Cauchy problem

$$(\mathcal{P}_f(a)) \quad \left\{ \begin{array}{rcl} \dot{a}(t) & = & f(a(t)) \\ a(0) & = & a^e(0) \end{array} \right. \tag{4.11}$$

is a^e over $[0,T]$ if, and only if, $\|e(f,t)\|_\Lambda = 0$ for all $t \in [0,T]$. In fact, the ith component of $e(f,t)$ quantifies a distance linked to the ith POD mode between the data and the ODE system defined by f at the time t : it depends only on quantities related to the ith mode as a_i^e, $\dot{a}_i^e$, $f_i(a^e)$ or $f_i(a)$ for a satisfying $(\mathcal{P}_f(a))$. Note that $\Lambda \in \mathbb{R}^{M\times M}$ denotes the symmetric definite positive matrix associated to $\|.\|_\Lambda$:

$$\forall z \in \mathbb{R}^M \qquad \|z\|_\Lambda = \sqrt{z^T \Lambda\, z}. \tag{4.12}$$

Changing this matrix enables us to give more or less importance to certain POD components (see the remark below).
Three choices for e are proposed : a constrained non-linear definition (section 4.2.3) and two definitions which are affine with respect to f (sections 4.2.3 and 4.2.3). Appendix 4.5.2 emphasizes the fact that, for e affine, $\mathcal{E}$ is a quadratic function and then the optimization problem reduces to a linear system.

$\mathcal{D}$ is expressed using a semi-norm $\|.\|_\Pi$ on the polynomial vector space :

$$\mathcal{D}(f) = \frac{\|f - f^g\|_\Pi{}^2}{\|f^g\|_\Pi{}^2}. \tag{4.13}$$

If $y \in \mathbb{R}^P$ is the vector of all the coefficients of the vector polynomial f of degree 2 in the natural monomial basis ($P = M\,(1 + M + \frac{M(M+1)}{2}) = \frac{M(M+1)(M+2)}{2}$) $\|f\|_\Pi$ is defined by

$$\|f\|_\Pi = \sqrt{y^T\, \Pi\, y} \tag{4.14}$$

where $\Pi \in \mathbb{R}^{P\times P}$ is a non-negative symmetric matrix. Modifying $\|.\|_\Pi$, that is Π, changes the relative importance of each polynomial coefficient, in particular $\|f\|_\Pi$ may be restricted to a subset of the polynomial coefficients of f such that only this subset is taken into account (partial-Galerkin method, section 4.2.4 and 4.3.3). The semi-norms $\|.\|_\Pi$ used in section 4.3 are Euclidian norms of all or, respectively, a subset of the polynomial coefficients along the natural monomial basis : Π is the identity matrix I_P of dimension P or, respectively, I_P whose some chosen diagonal elements are set to zero.

Whatever the definition of e is, minimizing $\mathcal{J}^\alpha$ is a optimization problem in $\mathbb{R}^P$ since it amounts to find the vector y^α of all the polynomial coefficients of f^α. The proposed methods

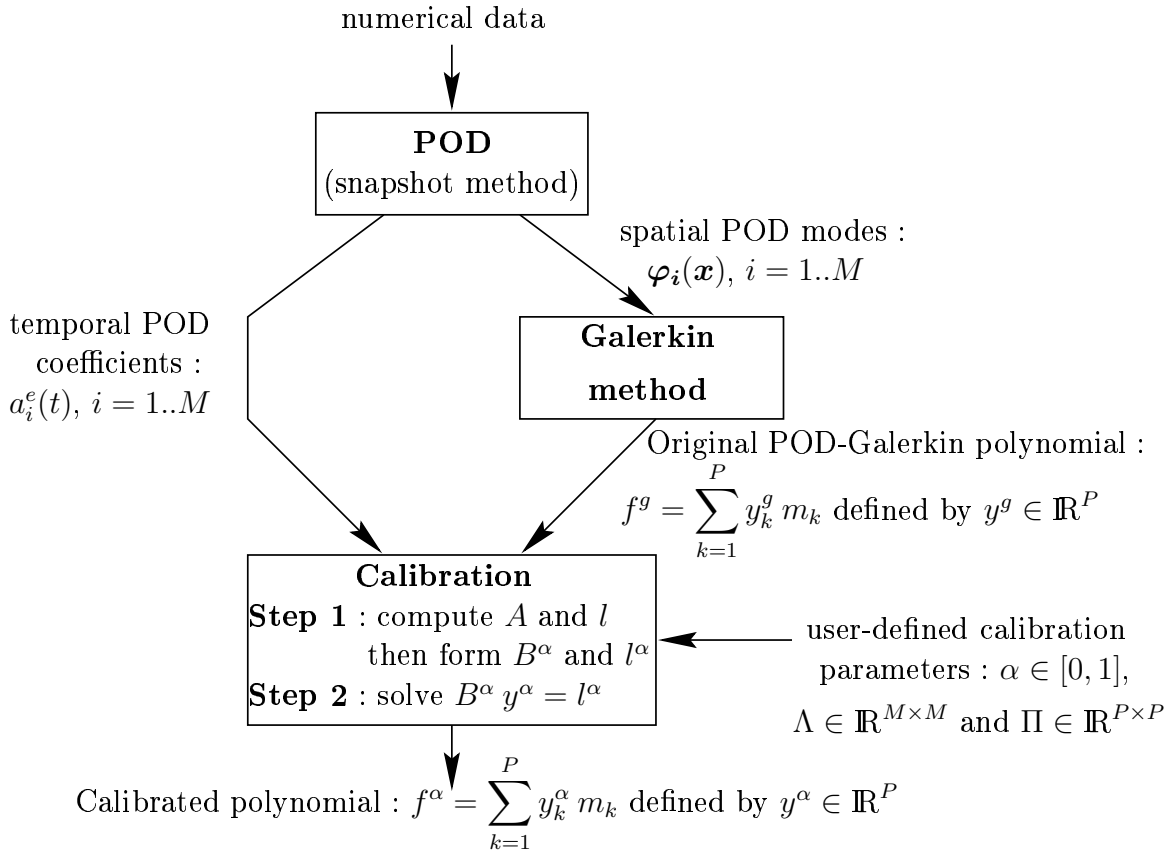
amounts to adding a vector polynomial $f^\alpha - f^g$ to the original POD-Galerkin system. This polynomial can be virtually split into three polynomials : $f^\alpha - f^g = f^\varepsilon + f^> + f^p$, that is into a small perturbation f^ε which corrects the errors of the computation of f^g, a non-linear closure model $f^>$ of the truncated terms which is more general than a linear viscous model and, if the boundary conditions are not explicitely taken into account and in consequence the pressure term poses problems, a polynomial f^p which models the boundary pressure term. Indeed, the calibrated POD-Galerkin system is $\dot{a}(t) = f^\alpha(a(t))$, that is virtually $\dot{a}(t) = (f^g + f^\varepsilon)(a(t)) + f^>(a(t)) + f^p(a(t))$.

Remark

It is worth noting that the meanings of $\mathcal{E}$ and $\mathcal{D}$ depend on $\|.\|_\Lambda$ and $\|.\|_\Pi$ but also on the meaning of the time-dependent coefficients considered. For instance, in our numerical experiments (section 4.3), $\|.\|_\Lambda$ is the Euclidean norm (Λ is the identity matrix I_M of dimension M) hence it keeps the natural hierarchy of the POD modes : since the $\boldsymbol{\varphi_i}$ are normalized, $T^{-1}\int_0^T a_i^e(t)^2\,dt = \sigma_i^2$ is the kinetic energy associated to the ith mode and the a_i^e are "naturally ranked". For the same definitions of $\mathcal{E}$ and $\mathcal{D}$, the interpretation of the results would not be equivalent for polynomial ODE systems constructed to approach the behaviour of the normalized time-dependent coefficients a_i^e/σ_i instead of the a_i^e (Λ should be changed into $\mathrm{diag}(\sigma_1 \cdots \sigma_M)$ for $\mathcal{E}$ to be equivalent).
Furthermore, it is better to calibrate the ODE system whose time-dependent unknowns are the a_i instead of the equivalent one whose unknowns are the a_i/σ_i for the same diagonal matrices $\|.\|_\Lambda$ and $\|.\|_\Pi$ since then minimizing $\mathcal{E}$ amounts to decrease the distance between the ODE system and the data for the first modes in priority and since then it appears less expensive to modify in priority the polynomial coefficients of the last modes if you pay attention to the term $\mathcal{D}$.
However, in our numerical experiments, each component of the optimal vector polynomial f^α is computed independently of the others in practice owing to the particular choices of e, Λ and Π retained (see the introduction of section 4.3). Thus, calibrating either POD-Galerkin systems is equivalent, yet the values of $\mathcal{E}$ and $\mathcal{D}$ are more representative of the effects of the calibrations for the POD-Galerkin system whose unknowns are the a_i.

4.2.2 Synthetic scheme of the calibration POD-Galerkin methods

To give a better idea of the proposed algorithms, the main steps of the calibration POD-Galerkin modelling for an affine definition of $e(.,t)$ are here summed up :



The calibration methods optimize the coefficients of the quadratic vector polynomial f^g associated the reduced-order POD-Galerkin system. The problem is solved in $\mathbb{R}^P$ by considering the P coefficients of the vector polynomials with M components of degree 2 in M variables : y^g and y^α are the vectors of the coefficients of f^g and f^α in the natural monomial basis (m_k).
If $e(f,t)$ is affine with respect to f then the calibration amounts to solve a linear system whose matrix B^α and right-hand side l^α are formed after the computation of the matrix $A = B^0$ and the right-hand side $l = l^0$ ($\alpha = 0$), which is performed during the two steps of the calibration : see the appendix 4.5.2. The step 1 of the calibration, that is the computation of A and l, depends on the definition of the operator e thus is different for the two methods which are proposed in the following. For both of these methods, some expressions of A and l are given in the appendix 4.5.3, moreover the linear systems which are obtained are detailed in the case $M = 2$ in the appendix 4.5.4.

4.2.3 Three definitions for e

Non-linear definition with dynamical constraints

Since the main objective is to obtain an ODE system whose simulation recovers as accurately as possible the evolution of the data, the natural mathematical formulation of the problem would consist in defining $e(f,t)$ as

$$e_1(f,t) = a^e(t) - a(t) \tag{4.15}$$

under the constraint that $a(t)$ satisfies (4.11). Thus, a solution f^α to problem (4.9) would minimize the gap between data and dynamical system states on average.

Unfortunately, that strongly non-linear definition of e is not satisfactory for several reasons. Indeed, since polynomials f may not be globally Lipschitz (only locally), the unique maximal solution $a(t)$ to the Cauchy problem $(\mathcal{P}_f(a))$ can take infinite values in finite time, in particular before the final time T of the experiment. Thus, in general, $\left\langle \|e_1(f,t)\|_\Lambda^{\,2} \right\rangle$ is not defined on all the vector polynomial space but only on an open subset $\mathcal{O}$ (to which at least the polynomials of degree 1 belong) : therefore, for $e = e_1$, $\mathcal{J}^\alpha(f)$ is not defined if $f \notin \mathcal{O}$. Notice that $\mathcal{E}$, as defined in section 4.2.1, has no sense if $f^g \notin \mathcal{O}$ (case where the POD-Galerkin model is unstable). Furthermore, the minimization problem for $e = e_1$ is not well posed : several solutions coexist in general.
However, since $\mathcal{J}^\alpha$ takes finite values and is differentiable in $\mathcal{O}$, it would be possible to look for a local minimum, starting from $f = f^g$ (or from its linear part and redefining $\mathcal{E}$ using for instance $e(0,t)$ instead of $e(f^g,t)$ if $f^g \notin \mathcal{O}$) and applying a non-linear conjugate gradient algorithm. But the computational cost of the gradient evaluation is large, the convergence is not guaranteed and the algorithm has to be designed in considering that a polynomial iterate is not necessary in $\mathcal{O}$.
This non-linear definition of e can yet be modified into an affine one by suppressing the dynamical constraint (4.11) in the definition of e_1 as proposed in the next section.

State calibration method

The operator e_1 defined in the preceding section can be written

$$e_1(f,t) = a^e(t) - a^e(0) - \int_0^t f(a(\tau))d\tau \tag{4.16}$$

where $a(t)$ is given by the constraint (4.11) on f. To suppress this non-linear constraint we can use :

$$e_2(f,t) = a^e(t) - a^e(0) - \int_0^t f(a^e(\tau))d\tau. \tag{4.17}$$

In this manner we test how accurately the data respect the dynamical system, keeping the same point of view than in the previous section where the data history a^e and the state a of the solution to the ODE system defined by f are compared. Clearly the operator e_2 is affine with respect to f : minimizing $\mathcal{J}^\alpha$ gives rise to a linear system.

Flow calibration method

Since the ideal polynomial should satisfy $\dot{a}^e(t) = f(a^e(t))$, we propose as third choice for e the operator

$$e_3(f,t) = \dot{a}^e(t) - f(a^e(t)). \tag{4.18}$$

Therefore we try to minimize the gap between the dynamics of the data and the flow of the ODE system defined by f, that is the gap between the time derivative of a^e and the vector field defined by f along the trajectory covered by $a^e(t)$ for $t \in [0,T]$. In fact, e_3 is the time derivative of e_2 :

$$e_2(f,t) = \int_0^t e_3(f,\tau)\, d\tau \tag{4.19}$$

and e_3 is affine with respect to f as e_2 is.

4.2.4 Computational cost and partial-Galerkin method

The computational cost of the POD-Galerkin polynomial coefficients is large for transitional or turbulent flows. Indeed, if the flow structures cover a large range of scales, the number of POD modes kept to construct the ODE system and also the number of meshes are large. Moreover, the use of the calibration methods to improve the system increases that cost. However, we expect to be able to decrease the cost of the ODE system computation by "mixing" the Galerkin method with a calibration method, that is by calculating only a "minimum" number of POD-Galerkin polynomial coefficients then evaluating the others by the optimization process (see below).
If that technique is effective, its cost is likely to be smaller than the one of a full Galerkin method : the methods we propose only require the temporal information given by the POD (that is $a^e(t)$), whereas the Galerkin method uses the voluminous spatial information (the $\boldsymbol{\varphi_i}$) and leads to spatial operations which are computationally expensive.
And yet, even if the cost of computing a POD-Galerkin system from a POD decreases, it is hard to predict whether the global cost, including the POD, will increase or not when that technique we call partial-Galerkin method is used. Indeed, when the snapshot method is applied to form the POD (see [88]), it is often possible to calculate a suitable set of POD modes from only few snapshots regularly distributed in time. In that case, the $a_i^e(t)$ are only calculated at the corresponding times during the POD computations (they are the eigenvectors of the temporal correlation matrix used to form the $\boldsymbol{\varphi_i}$). Thus, it may be necessary to increase the computational cost of the POD to have enough temporal data for the partial-Galerkin methods to be effective. It is noticeable that the number of temporal POD data (the number of times at which a^e is known) could be artificially increased using cubic spline interpolation.
Nevertheless, that technique remains interesting since this is expected to decrease the computational cost of the calibration POD-Galerkin methods.

Once a subset of the polynomial coefficients have been evaluated by the Galerkin method, two strategies are considered. In all cases, the initial polynomial Galerkin polynomial f^g is defined by setting the other coefficients to zero.

The first strategy is to define $\|.\|_\Pi$ as a semi-norm which takes only the subset of the coefficients computed by the Galerkin method into account. For example, when this strategy is applied in our numerical experiments, Π is defined as the modified identity matrix whose diagonal values whose locations correspond to the coefficients which were not computed are set to zero.
The second strategy is to use a definite norm $\|.\|_\Pi$ on all the coefficients (for example $\Pi = \mathrm{I}_P$ as in our experiments). In that latter case, it is *a priori* important to compute by the Galerkin method the polynomial coefficients whose effects cannot be neglected since the others will be assumed to be zero and taken into account in $\|f - f^g\|_\Pi$.
Some numerical results of partial-Galerkin methods are discussed in section 4.3.3.

4.3 Numerical experiments

In this section, some results for the two calibration methods ($e = e_2$ and e_3) are presented for the fluid flow configurations of sections 4.1.2 and 4.1.3.
$\|.\|_\Lambda$ is the usual Euclidian norm on $\mathbb{R}^M$ ($\Lambda = \mathrm{I}_M$). Moreover, all the results plotted here were computed with $\Pi = \mathrm{I}_P$ (so $\|.\|_\Pi$ is the complete Euclidian norm on the monomial basis). Only the experiments based on the first strategy of partial-Galerkin methods were performed with some singular matrices Π (some of its diagonal elements are set to zero as explained in the previous section), however no results are plotted in this case because the linear system is ill-conditioned : read the section 4.3.3 for the details.
It is worthy of note that for the state and flow calibrations ($e = e_2$ or $e = e_3$) and for the particular forms of Λ and Π used, the resulting linear problem can be naturally split into M similar linear problems thus it requires to solve a linear problem of dimension P/M with M different right-hand sides (where $P/M = \frac{(M+1)(M+2)}{2}$ is the dimension of the scalar polynomial space of degree 2) : each component f_i^α of the optimal polynomial can be computed independently of the others (this is emphasized in the appendix 4.5.4 for $M = 2$).
We remind you that the optimal solution computed is a vector $y^\alpha \in \mathbb{R}^P$ which defines all the coefficients of a vector polynomial f^α of degree 2 : all the coefficients are calibrated in this way, even if only a part of the POD-Galerkin coefficients is computed and taken into account (partial-Galerkin methods).

For the flow calibration ($e = e_3$), $\langle . \rangle$ is the arithmetic time average on the regular subdivision of $[0, T]$ corresponding to the snapshot database used :

$$\langle g(t) \rangle = \frac{1}{N} \sum_{k=0}^{N-1} g(k\Delta t) \quad \text{with} \quad \Delta t = \frac{T}{N-1}. \tag{4.20}$$

For the state calibration ($e = e_2$), $\langle . \rangle$ is defined by the discretization of the integration over $[0, T]$ given by the trapezoidal rule (second-order method), which almost amounts to use an arithmetic time average :

$$\langle g(t) \rangle = \frac{\Delta t}{T} \sum_{k=0}^{N-1} \frac{1}{2} \left[g(k\Delta t) + g((k+1)\Delta t) \right]. \tag{4.21}$$

The implementation of the state calibration method and the calculation of the corresponding cost function $\mathcal{E}$, denoted by $\mathcal{E}_2$ to precise that $e = e_2$, imply the discretization of terms in the form $\int_0^t g(a^e(\tau))d\tau$ and the same trapezoidal rule is applied (these appear in the expressions of A and l defined in appendix 4.5.3).
Moreover, all the simulations of the POD-Galerkin systems are performed by a classical fourth-order Runge-Kutta scheme with a time step of $2 \times 10^{-6}\, T$, in particular to evaluate $\mathcal{E}$ for $e = e_1$ (which is denoted by $\mathcal{E}_1$ in the following).

The numerical code was previously validated on data generated by some simulations of three-dimensional quadratic ODE systems : a prototype one proposed by Rössler in [80] and the Lorenz system (see [57, 22]). Even for $\alpha = 0$, suitable calibrated systems f^α were obtained.

4.3.1 Numerical efficiency and impact on the POD-Galerkin systems

The optimal polynomials f^α and the cost functions $\mathcal{E}$ are indexed by the subscript of the corresponding operator $e : \mathcal{E}_1$ for $e = e_1$, $\mathcal{E}_2$ and $f^{\alpha,2}$ for $e = e_2$, $\mathcal{E}_3$ and $f^{\alpha,3}$ for $e = e_3$: $f^{\alpha,2}$ is obtained by state calibration and $f^{\alpha,3}$ by flow calibration, respectively. More precisely, for all $j \in \{1, 2, 3\}$ and all $k \in \{2, 3\}$,

$$\mathcal{E}_j(f) = \frac{\left\langle \|e_j\left(f, t\right)\|_\Lambda{}^2 \right\rangle}{\left\langle \|e_j\left(f^g, t\right)\|_\Lambda{}^2 \right\rangle}, \text{ in particular } \mathcal{E}_j(f^{\alpha,k}) = \frac{\left\langle \left\|e_j\left(f^{\alpha,k}, t\right)\right\|_\Lambda{}^2 \right\rangle}{\left\langle \|e_j\left(f^g, t\right)\|_\Lambda{}^2 \right\rangle} \tag{4.22}$$

where $f^{\alpha,k}$ is the optimal polynomial which satisfies

$$(1-\alpha)\,\mathcal{E}_k(f^{\alpha,k}) + \alpha\,\mathcal{D}(f^{\alpha,k}) \leq (1-\alpha)\,\mathcal{E}_k(f) + \alpha\,\mathcal{D}(f) \tag{4.23}$$

for all vector polynomial f ($j = k$ in that latter equation) : $f^{\alpha,k}$ depends on α.
The values of $\sqrt{\mathcal{E}_1(f^{\alpha,k})}$, $\sqrt{\mathcal{E}_2(f^{\alpha,k})}$, $\sqrt{\mathcal{E}_3(f^{\alpha,k})}$ and $\sqrt{\mathcal{D}(f^{\alpha,k})}$ are presented for both calibration methods, that is $k = 2$ or $k = 3$, applied to the two flow configurations with respect to α (or a parameter δ in increasing bijection with α on $[0, 1]$, see below).

For $\alpha = 1$, if $\|.\|_\Pi$ is a norm (not just a semi-norm), the unique solution to the minimization problem is $f^\alpha = f^g$, thus $\mathcal{D}(f^\alpha) = 0$ and $\mathcal{E}(f^\alpha) = 1$ since $\mathcal{E}$ is a normalized cost : the POD-Galerkin is not calibrated. For state or flow calibration methods ($k = 2$ or $k = 3$), the original ODE system is more and more calibrated in the sense that $\sqrt{\mathcal{E}_k(f^{\alpha,k})}$ decreases and more and more modified in the sense that $\sqrt{\mathcal{D}(f^{\alpha,k})}$ increases as α tends to zero : see (4.23). It is noticeable that, if $\sqrt{\mathcal{E}_2(f^{\alpha,2})}$, $\sqrt{\mathcal{E}_3(f^{\alpha,3})}$, $\sqrt{\mathcal{D}(f^{\alpha,2})}$ and $\sqrt{\mathcal{D}(f^{\alpha,3})}$ are necessary monotone functions of $\alpha \in [0, 1]$ by definition, the curves of $\sqrt{\mathcal{E}_1(f^{\alpha,2})}$, $\sqrt{\mathcal{E}_1(f^{\alpha,3})}$, $\sqrt{\mathcal{E}_2(f^{\alpha,3})}$ and $\sqrt{\mathcal{E}_3(f^{\alpha,2})}$ may be non-monotone *a priori*.

Few components of the solutions to the original system defined by f^g and of some calibrated systems, which are necessary computed to evaluate $\mathcal{E}_1$, are displayed in the following (in phase-portrait format for the square-cylinder flow configuration and in function

of the time for the step flow configuration) : a_i^g is the ith component of the solution a^g to the original POD-Galerkin system and $a_i^{\alpha,k}$ denotes the ith component of the solution $a^{\alpha,k}$ to the calibrated system defined by $f^{\alpha,k}$. For instance, $a^{1,2}$ and $a^{1,3}$ are respectively the solutions to the state- and flow-calibrated systems computed with $\alpha = 1 : a^{1,2} = a^{1,3} = a^g$ if $\|.\|_\Pi$ is definite (Π non singular).

The square-cylinder configuration

The results were computed varying linearly α, but also varying linearly a parameter δ such that

$$\alpha = \frac{\delta}{\zeta(1-\delta)+\delta} \quad \text{with} \quad \zeta = \frac{\left\langle \|e(f^g,t)\|_\Lambda{}^2 \right\rangle}{\left\langle \|e(0,t)\|_\Lambda{}^2 \right\rangle}. \tag{4.24}$$

α is in increasing bijection with δ on $[0,1]$. In fact, that second way to vary α amounts to minimize

$$\widetilde{\mathcal{J}}^\delta(f) = (1-\delta)\,\widetilde{\mathcal{E}}(f) + \delta\,\mathcal{D}(f) \tag{4.25}$$

with

$$\widetilde{\mathcal{E}}(f) = \zeta\ \mathcal{E}(f) = \frac{\left\langle \|e(f,t)\|_\Lambda{}^2 \right\rangle}{\left\langle \|e(0,t)\|_\Lambda{}^2 \right\rangle}, \tag{4.26}$$

since

$$\widetilde{\mathcal{J}}^\delta(f) = \frac{\delta}{\alpha}\,\mathcal{J}^\alpha(f) = [\zeta(1-\delta)+\delta]\ \mathcal{J}^\alpha(f). \tag{4.27}$$

Indeed, results with linear variation of δ are more readable than ones with linear variation of α since the curves decrease or increase extremely rapidly when α varies between 0.95 and 1 as you can see on figure 4.10. Therefore, two series of results are obtained and plotted for α and δ in $\{0.05, 0.1, \cdots, 1\}$. α is displayed in function of δ for $e = e_2$ and $e = e_3$ on figure 4.9 for $\delta \in [0.05, 1]$. The minimization results are presented on figure 4.10.

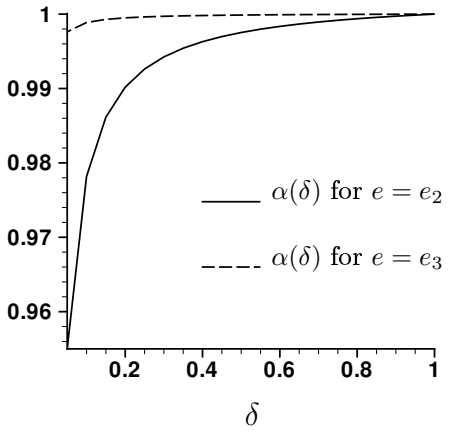


FIG. 4.9 – α function of δ.

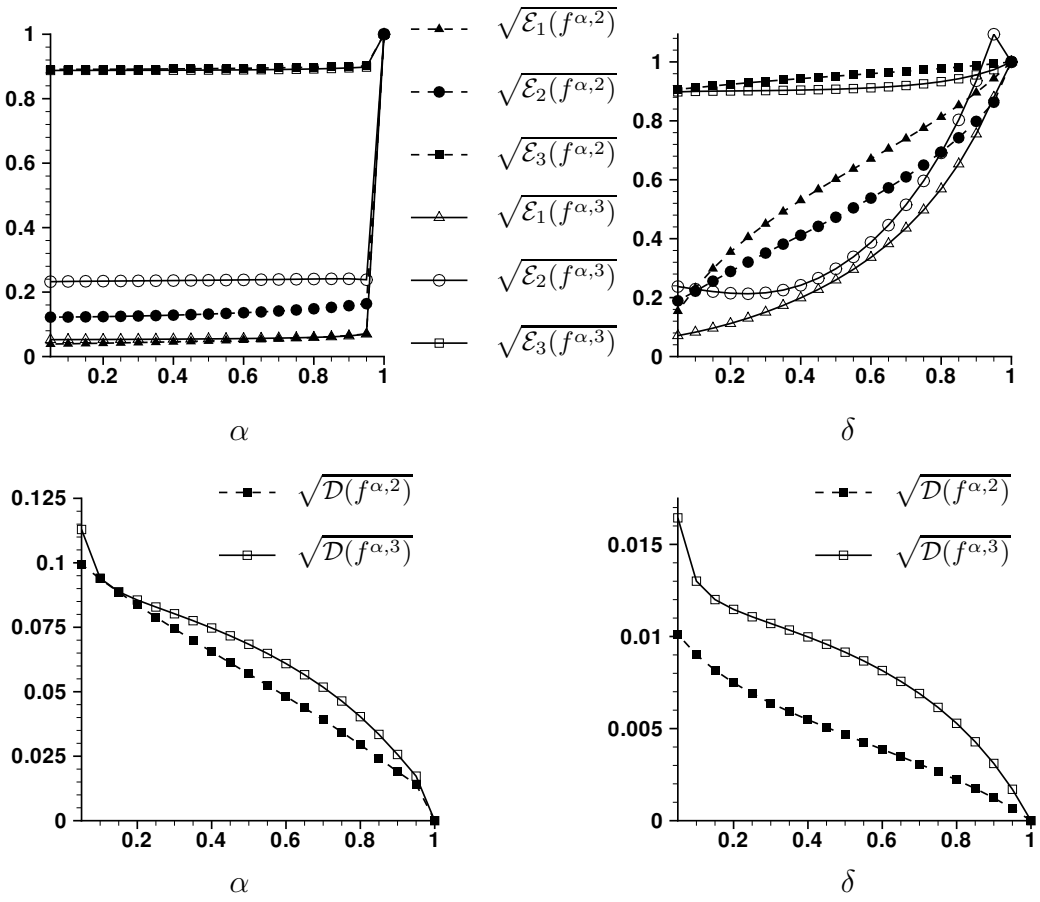


FIG. 4.10 – Minimization results for the square-cylinder configuration.

It is observed that $\sqrt{\mathcal{E}_1(f^{\alpha,2})}$, $\sqrt{\mathcal{E}_3(f^{\alpha,2})}$ and $\sqrt{\mathcal{E}_1(f^{\alpha,3})}$ are increasing functions of α. Only $\sqrt{\mathcal{E}_2(f^{\alpha,3})}$ decreases very slowly for $\delta \in [0.05, 0.25]$ but its global shape is an increase.
For both calibrations ($k = 2$ or $k = 3$), $\sqrt{\mathcal{E}_3(f^{\alpha,k})}$ appears little reduced for small values of α ($\sqrt{\mathcal{E}_3(f^{\alpha,2})}$ and $\sqrt{\mathcal{E}_3(f^{\alpha,3})}$ are about 0.89 for $\alpha = 0.05$) : this is normal since the non-calibrated POD-Galerkin system f^g was relatively accurate (the difference between $\dot{a}^e$ and $f^g(a^e)$ which defines $e_3(f^g, t)$ is very small, look at the figure 4.4) and since the efficiency of the calibration is limited by the machine precision.
The main observation is that $\sqrt{\mathcal{E}_1(f^{\alpha,k})}$ vanishes for $k = 2$ and $k = 3$ as well when α tends to zero : state and flow calibrations are effective on this two-dimensional quasi-incompressible configuration. Moreover, this can be done for very small perturbations of the original POD-Galerkin system since, for instance, $\sqrt{\mathcal{D}(f^{\alpha,2})}$ and $\sqrt{\mathcal{D}(f^{\alpha,3})}$ are less than 2% and correspond to very good improvements of the ODE system for $\alpha = 0.95$: $\sqrt{\mathcal{E}_1(f^{\alpha,2})}$ and $\sqrt{\mathcal{E}_1(f^{\alpha,3})}$ are about 0.07.

To visualize the effectiveness of the calibration of the POD-Galerkin system of the square-cylinder flow, the solution a^g to the non-calibrated system is presented over three periods, that is for $t \in [0, 3\,T]$, with the solution $a^{0.05,3}$ to the system obtained by flow calibration with $\alpha = 0.05$ on figure 4.11. It is observed that the calibrated system gives a

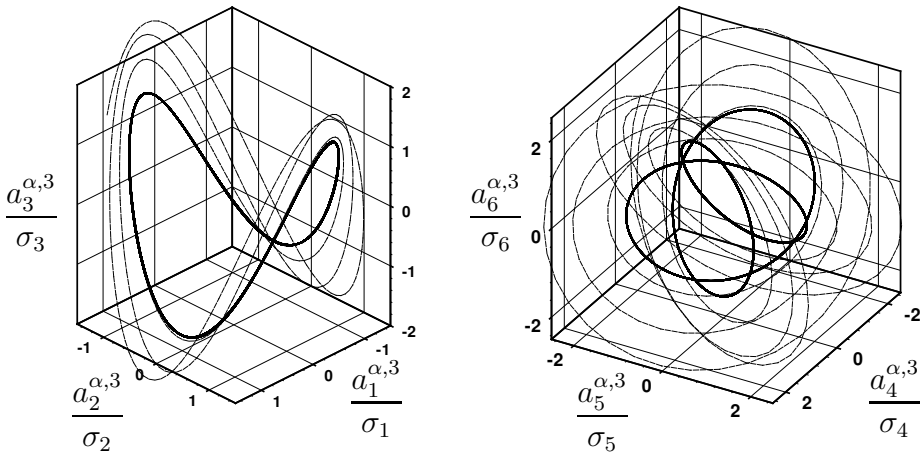


FIG. 4.11 – Solution $a^g = a^{1,3}$ ($\alpha = 1$) to the non-calibrated system (thin dashed line) and solution $a^{0.05,3}$ to the system calibrated with $\alpha = 0.05$ by the flow method (thick line) for $t \in [0, 3\,T]$.

solution which is periodic (and matches the data history a^e) whereas the solution given by the original POD-Galerkin system diverges from that periodic trajectory.

The backward-facing-step configuration

For the second test flow configuration, the calibration of two POD-Galerkin systems are experimented : the 86-mode system and the 45-mode system. The 45 first POD modes

capture more than 95% of the fluctuant kinetic energy of the data; the polynomial coefficients of the 45-mode system are included in the coefficients of the 86-mode system.
Some results for $\alpha = 0.001$ and α in $\{0.05, 0.1, \cdots, 1\}$ are plotted : see figures 4.12 (86-mode system) and 4.13 (45-mode system).

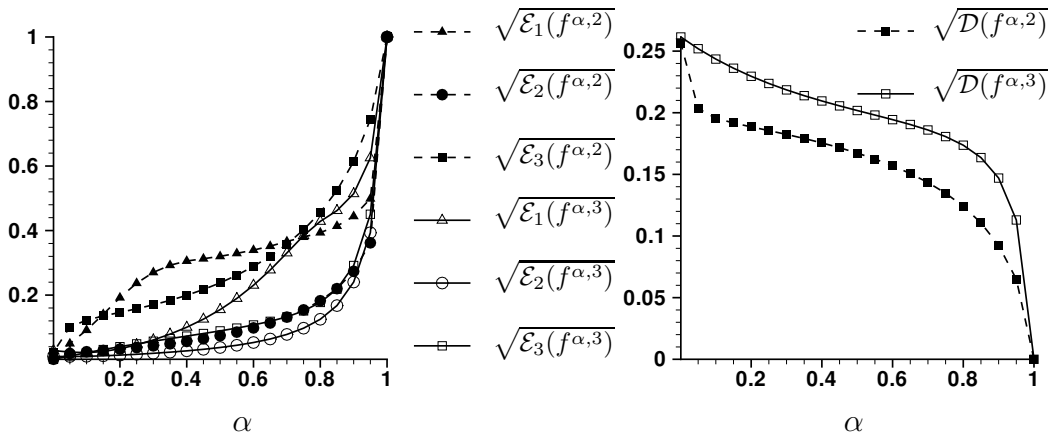


FIG. 4.12 – Minimization results for the 86-mode POD-Galerkin system of the backward-facing-step configuration ($M = 86$).

For the two original POD-Galerkin systems, it is observed that $\sqrt{\mathcal{E}_1(f^{\alpha,2})}$, $\sqrt{\mathcal{E}_3(f^{\alpha,2})}$, $\sqrt{\mathcal{E}_1(f^{\alpha,3})}$ and $\sqrt{\mathcal{E}_2(f^{\alpha,3})}$ are increasing functions of α; furthermore, $\sqrt{\mathcal{E}_1(f^{\alpha,k})}$, $\sqrt{\mathcal{E}_2(f^{\alpha,k})}$ and $\sqrt{\mathcal{E}_3(f^{\alpha,k})}$ take very small values for $\alpha = 0.001$ for both calibration methods ($k = 2$ or $k = 3$).
So the calibrations are effective but the optimal polynomials $f^{\alpha,2}$ and $f^{\alpha,3}$ cannot be regarded as small perturbations of the original POD-Galerkin system f^g for small values of α any more. For instance, for $M = 86$ or $M = 45$, $\sqrt{\mathcal{D}(f^{\alpha,k})}$ reaches about 0.2 when $\sqrt{\mathcal{E}_1(f^{\alpha,k})}$ is around 0.25 for both methods ($k = 2$ or $k = 3$). For $\alpha = 0.001$, the ODE system is more modified for $M = 45$ than for $M = 86$ (the values of $\mathcal{D}$ are greater for $M = 45$), which was expected since more truncated terms have to be modeled.
Furthermore, notice there is a gap between the formal expressions of f^g and the nature of the data used since the data were computed using a Finite-Difference Large-Eddy scheme and the boundary term $T_{\partial\Omega}$ was neglected in the variational formulation (see section 4.1.1 and appendix 4.5.1) : the calibration methods have to model the truncated terms but also to fill this gap.

To display the effects of the calibrations of the 86-mode system, few components of the systems calibrated with $\alpha = 0.95$ or $\alpha = 0.05$ are plotted with the corresponding components of a^g and a^e.
The figure 4.14 presents some components relative to three of the ten first POD modes. For both methods ($k = 2$ and $k = 3$), $a_i^{\alpha,k}$ converges as expected to a_i^e as α tends to zero, for

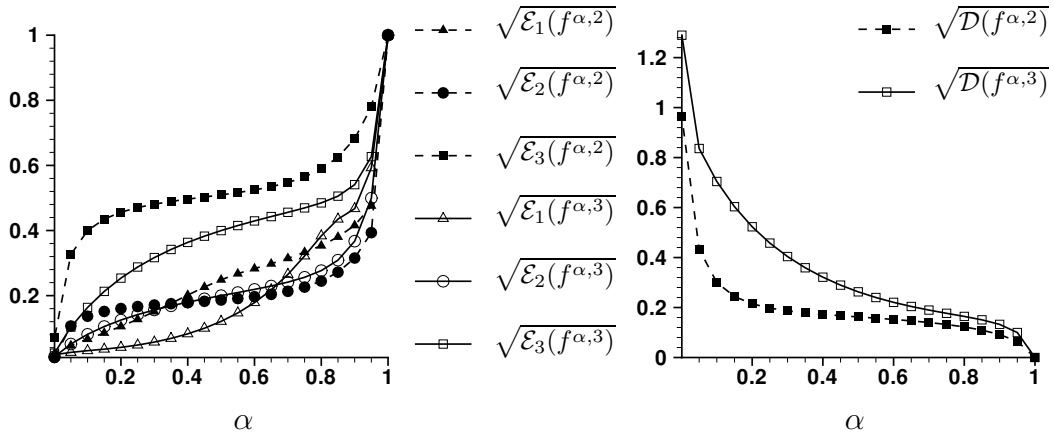


FIG. 4.13 – Minimization results for the 45-mode POD-Galerkin system of the backward-facing-step configuration ($M = 45$).

the three indexes $i = 1$, 5 and 10 : the relative difference $(T\,\sigma_i)^{-1}\int_0^T \left(a_i^e(t) - a_i^{\alpha,k}(t)\right)^2 dt$ decreases, moreover $a_i^{\alpha,2}$ and $a_i^{\alpha,3}$ match a_i^e for $\alpha = 0.05$.
Some components with larger indices, corresponding to less energetic POD modes, are displayed in figure 4.15. It appears that the histories of $a_{40}^{0.05,k}$, $a_{60}^{0.05,k}$ and $a_{80}^{0.05,k}$ are very satisfying for $k = 2$ and $k = 3$ as well, even if $a_i^{0.05,2}$ and $a_i^{0.05,3}$ do not match perfectly a_i^e for $i = 60$ and $i = 80$: the calibrations succeed in modelling the main effects of the truncated POD modes.

4.3.2 Remark on condition numbers

Some computations were also performed with $\alpha = 0$ for the preceding POD-Galerkin systems and for both calibration methods. In those cases, the methods failed to compute a numerically stable ODE system although it was possible for the Rössler and Lorenz systems : taking f^g into account seems compulsory in practice when the number of modes is not very small. This phenomenon can be understood looking at the condition numbers of the matrices of the linear systems. The linear systems are indeed ill-conditioned for $\alpha = 0$ and $\mathcal{D}$ is necessary to form a non-singular problem whose an approximate solution can be computed.

The base-10 logarithm of the condition numbers $\mathcal{K}(B^\alpha)$ of the matrices B^α formed during the calibrations (see appendix 4.5.2) is plotted as a function of α on figure 4.16 for the 6-mode model of the square-cylinder configuration (left) and for the 86-mode model of the backward-facing step configuration (right). It gives estimates of how many base-10 digits are lost when the corresponding linear systems are solved.
The numerical experiments presented were performed by manipulating 64-bit double precision numbers, that is about 15 significative decimal digits, and by using a direct method

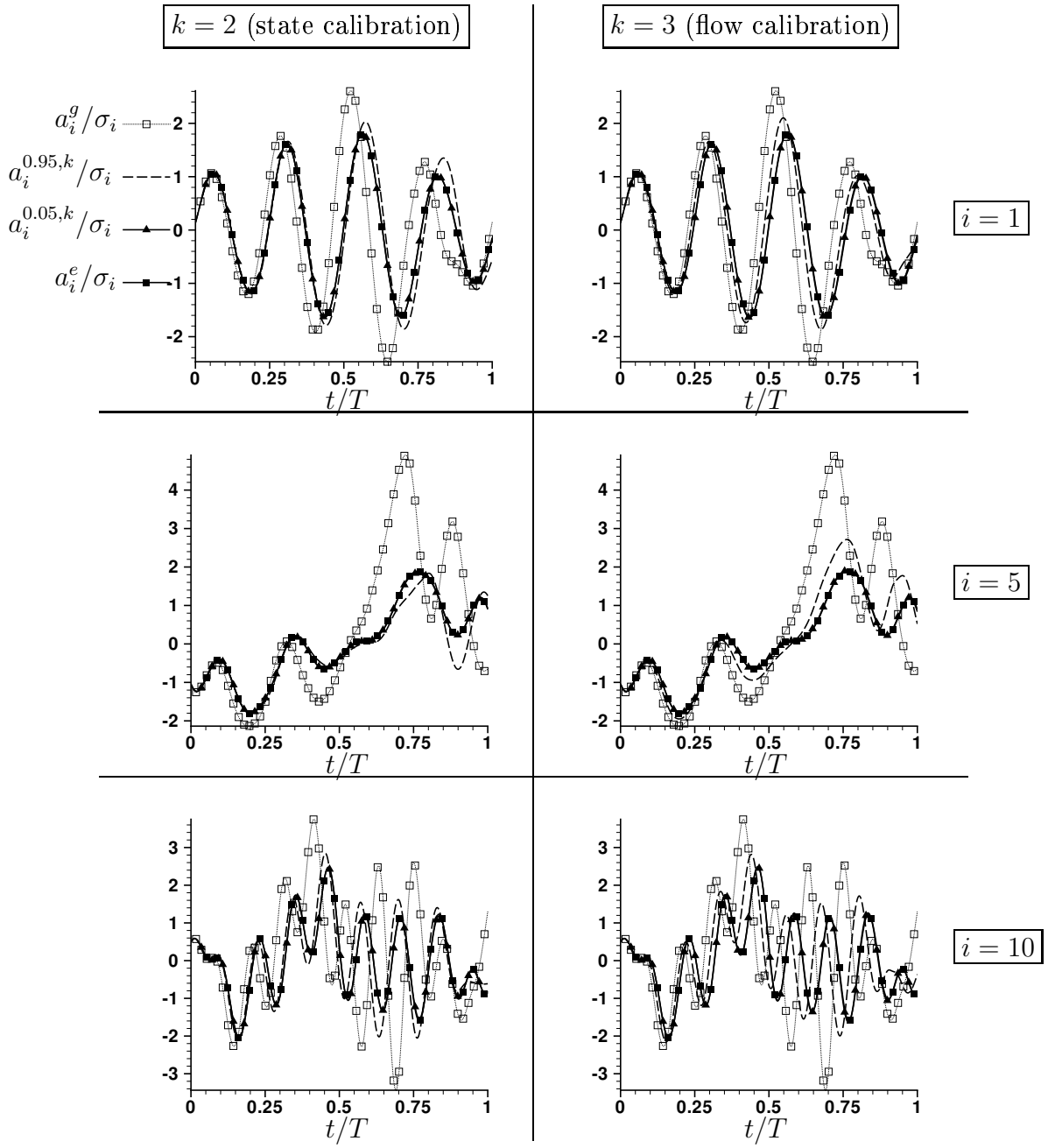


FIG. 4.14 – ith components of the solutions a^g to the original POD-Galerkin system and $a^{\alpha,k}$ to systems calibrated with $\alpha = 0.95$ and $\alpha = 0.05$ by state ($k = 2$) or flow ($k = 3$) method and ith component of the data history a^e, for $i = 1$, 5 and 10.

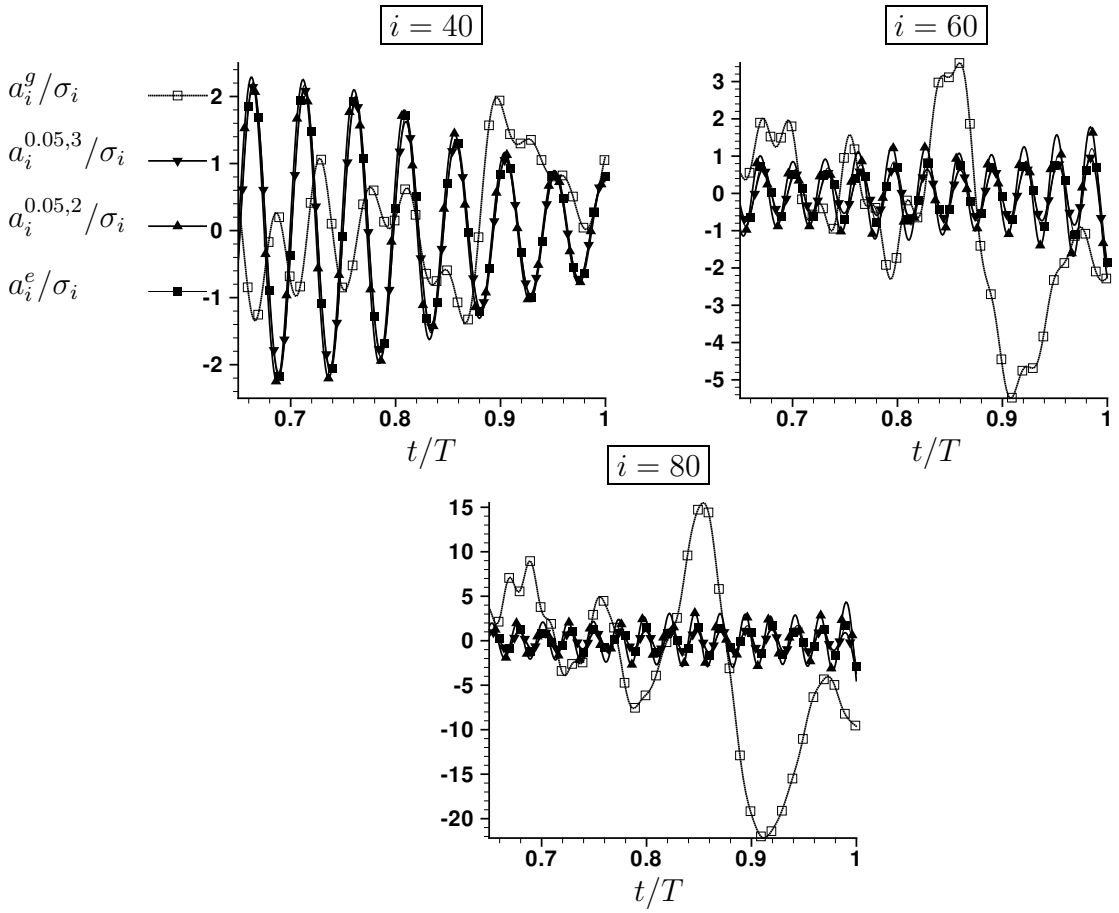


FIG. 4.15 – 40th, 60th and 80th components of the data history a^e, of the solution a^g to the original system and of the solutions $a^{0.05,2}$ and $a^{0.05,3}$ to the state- and flow-calibrated systems for $t \in [0.65\,T, T]$.

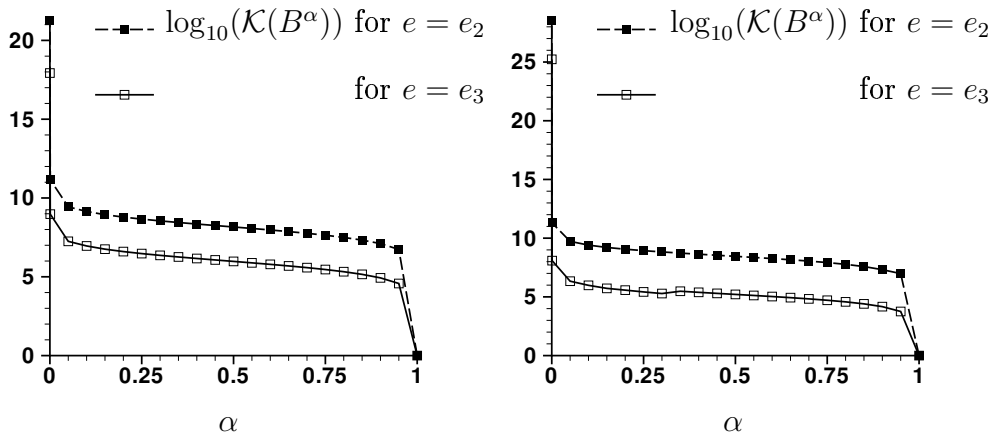


FIG. 4.16 – Condition numbers obtained during the calibrations of the 6-mode system of the square-cylinder flow (left) and of the 86-mode system of the backward facing-step flow (right) for $\alpha = 0.001$ and α in $\{0.05, 0.1, \cdots, 1\}$.

to solve the linear system. Here, $\log_{10}(\mathcal{K}(B^0))$ is always greater than 15 : the problem becomes ill-conditioned for $\alpha = 0$.

4.3.3 Partial-Galerkin methods

In this section, the partial-Galerkin methods are experimented from the Galerkin polynomial coefficients of the 86-mode system of the backward-facing step configuration. All the constant and linear monomials evaluated by the Galerkin method are taken into account but only a subset of the numerous quadratic ones. If the calibration methods remain effective for a small subset, considering that the other coefficients are zero or defining $\|.\|_\Pi$ so that only this subset matters, a computational gain is then obtained and that gain may be greater than the increase of the cost of the POD computations which would be necessary for the calibrations to be effective (see section 4.2.4).
The quadratic terms which are neglected are chosen keeping in mind that the interactions between POD modes are local, hence the "non-local" polynomial coefficients little matter *a priori*. That idea is based on the observations made in [78] and [16]. More precisely, if $C_j^{i_1,i_2} a_{i_1} a_{i_2}$ denotes the quadratic monomials of the jth component of a vector polynomial f as in equation (4.6), the monomial $C_j^{i_1,i_2} a_{i_1} a_{i_2}$ is neglected if, and only if, $|i_1 - i_2|$ is strictly greater than a threshold k. This criterion is the same for all the components of f (it does not depend on j) thus the calibration method still amounts to solving a linear problem of dimension P/M with M different right-hand sides, as in the previous experiments (as precised in the beginning of section 4.3). Notice that more sophisticated criterions with dependence on j could be proposed (for instance $\min(|i_1 - j|, |i_2 - j|) > k_j$ for M thresholds k_j) : these would be closer to the studies of the locality of the kinetic energy transfers of [78] and [16], nevertheless only the first criterion was tested for practical considerations.

Firstly, semi-norms $\|.\|_\Pi$ can be defined so that only the chosen subset of coefficients are taken into account : this is the first strategy where some diagonal elements of Π are set to zero as explained at the end of section 4.2.4. In that way, the methods do not implicitly assume that the neglected Galerkin coefficients are zero.
Unfortunately, this choice is not robust, since it generally leads to ill-conditioned linear systems. Indeed, when $\|.\|_\Pi$ is not definite, the linear system is singular for $\alpha = 1$, whereas it seems that the contribution of $\mathcal{D}$ improves the condition number in the previous experiments (section 4.3.2). In consequence, k must be close to M in our experiments for the calibrated system to be numerically stable and accurate.

Secondly, $\|.\|_\Pi$ is defined as the definite Euclidian norm of the polynomial coefficients along the natural monomial basis and the neglected Galerkin coefficients are assumed to be zero in the definition of f^g. Some numerical experiments based on this second strategy were performed from the partial 86-mode Galerkin system corresponding to $k = 40$ for $\alpha = 0.001$ and α in $\{0.05, 0.1, \cdots, 1\}$.
The results appear satisfying since relatively close to the preceding ones obtained from the full 86-mode Galerkin system ; see figure 4.17. Moreover the conditions numbers, plotted

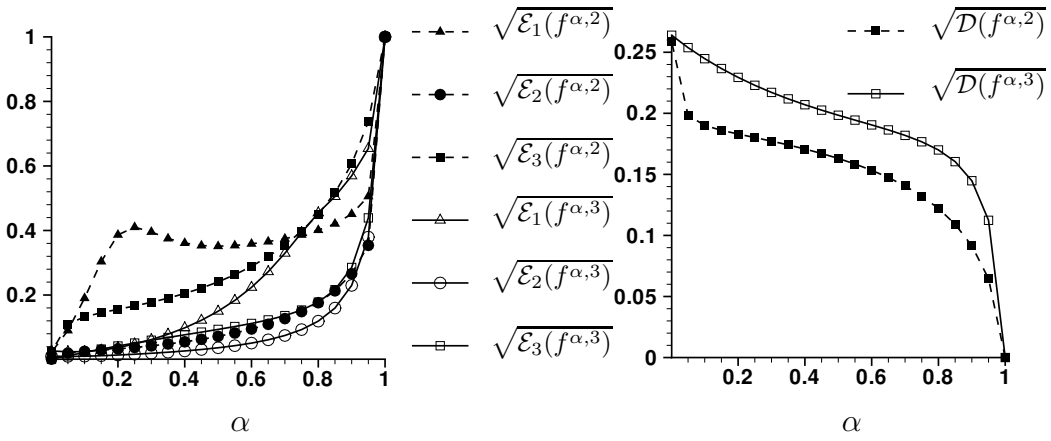


FIG. 4.17 – Minimization results for the partial 86-mode system of the backward-facing-step flow corresponding to a threshold $k = 40$.

on figure 4.18, are not worse than before.

All this leads to conclude that, in the case of transitional or turbulent flow modelling, the partial-Galerkin methods are interesting alternatives for correcting automatically the reduced-order POD-Galerkin system behaviour at relatively low cost, assuming that non-local POD interactions are negligible and using a definite norm $\|.\|_\Pi$.

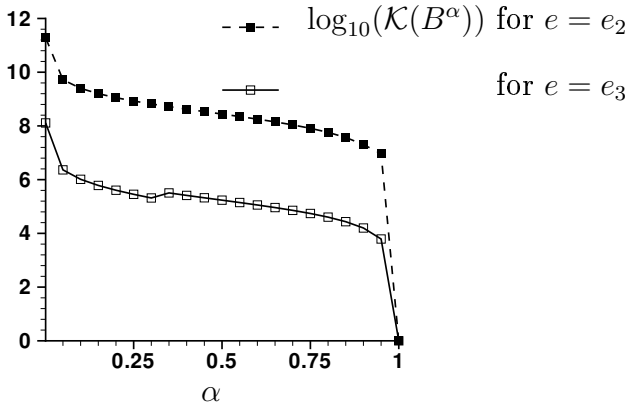


FIG. 4.18 – Condition numbers obtained from the partial 86-mode system of the backward facing-step configuration corresponding to a threshold $k = 40$.

4.4 Conclusions

Two numerical methods which rely on minimization problems and give rise to linear systems were proposed to improve reduced-order POD-Galerkin systems by calibrating their polynomial coefficients. They can be applied to unstable POD-Galerkin systems, that is systems which take infinite values before the time bound T considered. These methods exploit the temporal part of the POD information which is not taken into account in the POD-Galerkin approach.
Some numerical tests were performed on a relevant problem using 1000 velocity snapshots of a three-dimensional turbulent flow : the numerical behaviours of the 86-mode and 45-mode POD-Galerkin systems of a backward-facing-step flow were noticeably improved by calibrating their polynomial coefficients.

The computational cost of the construction and resolution of the linear systems is reasonable for both calibration methods since this cost is generally less than the cost of the POD-Galerkin calculations for transitional or turbulent flows. However, although reasonable, the global cost of the calibration is not insignificant since more POD or interpolation computations could be necessary to increase the number of data for the methods to be accurate enough. In particular, the first method ($e = e_2$) needs close snapshots for the discretization of the integrals over $[0, t]$ to be accurate. Fortunately, large computational gain can be obtained using partial-Galerkin methods in the cases of transitional or turbulent flows.

This work suggests several points to investigate. First, physical interpretation of the calibrations obtained should be studied. Furthermore, the impact of this POD-Galerkin modelling with calibration for the active control of an unsteady flow remains unknown. Moreover, the methods could be immediately applied on compressible flow cases since polynomial POD-Galerkin systems can be derived from the Navier-Stokes equations for a perfect gas (see [94] or [38]). And eventually, the effectiveness of the calibration methods

could be tested for other modelling problems, not necessary in the fluid mechanics field. Indeed, the methods can be easily extended to non-polynomial problems by assuming a new suitable space instead of the quadratic polynomial one which is considered here.

4.5 Appendixes

4.5.1 Treatment of the boundary conditions in the Galerkin method

The boundary term of the variational formulation (4.4) is

$$T_{\partial\Omega} = \int_{\partial\Omega} t_{\partial\Omega}\, ds \quad \text{with} \quad t_{\partial\Omega} = (p\,\boldsymbol{n} - \frac{1}{\mathrm{Re}}\,[\nabla u]\;\boldsymbol{n}).\boldsymbol{\varphi} \tag{4.28}$$

with p the pressure field and $\boldsymbol{n}$ the outward unitary normal at the border $\partial\Omega$ of Ω. That natural flux $t_{\partial\Omega}$ of the incompressible Navier-Stokes equations can be fully defined on $\partial\Omega$ by $\boldsymbol{u}$ for a large class of boundary conditions prescribed for incompressible flows. Thus, the pressure vanishes from the variational formulation and an ODE system only based on the POD of the velocity can be constructed by applying the Galerkin method.

Indeed, let us regard the following conditions on the boundary $\partial\Omega = \Gamma_D \cup \Gamma_N$ with $\Gamma_N = \partial\Omega \setminus \Gamma_D$:

$$\boldsymbol{u} = \boldsymbol{w} \quad \text{on} \quad \Gamma_D, \tag{4.29}$$

$$-\tilde{\sigma}\,\boldsymbol{n} = \boldsymbol{\beta} \quad \text{on} \quad \Gamma_N. \tag{4.30}$$

where

$$\tilde{\sigma} = \frac{1}{\mathrm{Re}}\,[\nabla u] - p\,\mathrm{I}_d \tag{4.31}$$

is a pseudo stress tensor (I_d is the identity matrix of dimension d) and $\boldsymbol{\beta}$ the pseudo flow stress set on Γ_N. (4.30) is a generalization of some boundary conditions usually used in incompressible flow simulations to represent the flow in an unbounded region : for instance

$$[\nabla u]\;\boldsymbol{n} = 0 \quad \text{and} \quad p = q \tag{4.32}$$

which can be applied for external flow (q is the pressure which is prescribed on Γ_N) or

$$p\,\boldsymbol{n} - \frac{1}{\mathrm{Re}}\,[\nabla u]\;\boldsymbol{n} = 0 \tag{4.33}$$

(pseudo stress free condition) which is often used as outlet condition for canal flows : read [85] or [75].
If the Dirichlet condition on the velocity is non homogeneous, that is $\boldsymbol{w} \neq \boldsymbol{0}$ on Γ_D, the POD is performed on $\boldsymbol{u}^e - \bar{\boldsymbol{u}}^e$ with a solenoidal field $\bar{\boldsymbol{u}}^e$ satisfying $\bar{\boldsymbol{u}}^e = \boldsymbol{w}$ on Γ_D : the POD modes are then solenoidal and vanish on Γ_D as $\boldsymbol{u}^e - \bar{\boldsymbol{u}}^e$ and the test function $\boldsymbol{\varphi}$ has

to be considered in a space of solenoidal functions which vanish on Γ_D. Therefore, $t_{\partial\Omega}$ takes zero values on Γ_D.
Moreover, the boundary condition (4.30) implies that $t_{\partial\Omega} = \boldsymbol{\beta}.\boldsymbol{\varphi}$ on Γ_N. Thus, the variational formulation (4.4) becomes

$$\frac{d}{dt}\left(\boldsymbol{u},\boldsymbol{\varphi}\right) + \left((\boldsymbol{u}.\boldsymbol{\nabla})\boldsymbol{u},\boldsymbol{\varphi}\right) + \frac{1}{\mathrm{Re}}\sum_{i=1}^{d}\left(\boldsymbol{\nabla}u_{x_i},\boldsymbol{\nabla}\varphi_{x_i}\right) + \int_{\Gamma_N}\boldsymbol{\beta}.\boldsymbol{\varphi}\,ds = \left(\boldsymbol{h},\boldsymbol{\varphi}\right) \tag{4.34}$$

Finally, the POD-Galerkin system is obtained from (4.34) using the Galerkin method with the POD modes as test functions and basis functions for $\boldsymbol{u} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$.
Notice that $T_{\partial\Omega} = \int_{\Gamma_N}\boldsymbol{\beta}.\boldsymbol{\varphi}\,ds = 0$ if $\boldsymbol{\beta} = -P\,\boldsymbol{n}$ with P a constant pressure for any solenoidal test function $\boldsymbol{\varphi}$ which is zero on Γ_D, which is logical since only the gradient of the pressure and not the pressure is physical in the case of an incompressible flow ($\boldsymbol{\nabla}(p+P) = \boldsymbol{\nabla}p$).

For the backward-facing step flow configuration, the outlet condition (4.33) is used. However, for the square-cylinder flow configuration, the outlet condition is little different : the transverse velocity u_{x_2}, the normal derivative of the streamwise velocity $\partial_{x_1}u_{x_1}$ and the dynamic pressure were set to zero during the computations at the outlet, that is

$$p\,\boldsymbol{n} - \frac{1}{\mathrm{Re}}\left[\nabla u\right]\boldsymbol{n} = -P_s\,\boldsymbol{n} - \frac{1}{\mathrm{Re}}\,\partial_{x_1}u_{x_2}\,\boldsymbol{x_2} \tag{4.35}$$

at the outlet, where P_s is a constant static pressure, $\boldsymbol{x_2}$ is the unitary vector in the transverse direction, the subscripts x_1 and x_2 denote respectively the streamwise and transverse components ($n_{x_1} = 1$ and $n_{x_2} = \boldsymbol{n}.\boldsymbol{x_2} = 0$ at the outlet) and where $\partial_{x_1}u_{x_2}$ is the derivative of $u_{x_2} = \boldsymbol{u}.\boldsymbol{x_2}$ in the normal direction. That condition is not strictly equivalent to (4.30) for $\boldsymbol{\beta} = -P_s\,\boldsymbol{n}$ but leads to the same conclusion : the boundary term $T_{\partial\Omega}$ vanishes if the test function $\boldsymbol{\varphi}$ is solenoidal and satisfies the same Dirichlet conditions than the velocity field, in particular $\varphi_{x_2} = \boldsymbol{\varphi}.\boldsymbol{x_2} = 0$ at the outlet. This is especially true if $\boldsymbol{\varphi}$ is a POD mode of a "fluctuant" velocity $\boldsymbol{u}^{\boldsymbol{e}} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$ where $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ is a solenoidal field defined to satisfy the same Dirichlet boundary conditions than $\boldsymbol{u}^{\boldsymbol{e}}$.

That explicit treatment of the boundary conditions is interesting if constructing such a solenoidal field $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ which satisfies the Dirichlet boundary conditions is easy and cheap. This is especially true for unsteady Dirichlet conditions (a mean velocity field is suitable) or for Dirichlet conditions in the form

$$\bar{\boldsymbol{u}}^{\boldsymbol{e}}(\boldsymbol{x},t) = \boldsymbol{w}(\boldsymbol{x},t) = \sum_{i=1}^{K} w_i(t)\,\boldsymbol{\psi_i}(\boldsymbol{x}) \text{ on } \Gamma_D \tag{4.36}$$

with K small : it suffices to solve K Stokes problems for each velocity profile $\boldsymbol{\psi_i}$. However, for complex unsteady Dirichlet boundary condition as in our step flow configuration, computing $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ amounts to simulating the flow : the boundary term $t_{\partial\Omega}$ poses problems and must be neglected or modeled (or the velocity-vorticity formulation of Rempfer must be tested [76]). Fortunately, the inlet Dirichlet condition of our step flow is quasi-steady and $t_{\partial\Omega}$ can be neglected as first approximation since it takes small values at the inlet.

Important remark

Notice that the variational formulation (4.34) is available for a pseudo stress condition (4.30) but not for the stress condition $-\sigma\,\boldsymbol{n} = \boldsymbol{\beta}$ where

$$\sigma = \tilde{\sigma} + \frac{1}{\mathrm{Re}}[\nabla u]^T = \frac{1}{\mathrm{Re}}\left([\nabla u] + [\nabla u]^T\right) - p\,\mathrm{I}_d \tag{4.37}$$

is the physical stress tensor and now $\boldsymbol{\beta}$ the flow stress on Γ_N. That latter condition is used in particular in the work of Noack *et al.* [66] but is not considered in the variational formulation proposed.
However, that boundary condition could be explicitely taken into account in theory by keeping the zero divergence of ∇u^T in the incompressible expression of the Navier-Stokes equations. Indeed, the incompressible Navier-Stokes equations can be expressed for any $\omega \in \mathbb{R}$ as

$$\begin{gathered} \boldsymbol{\nabla}.\boldsymbol{u} = 0 \\ \partial_t \boldsymbol{u} + (\boldsymbol{u}.\boldsymbol{\nabla})\boldsymbol{u} - \boldsymbol{\nabla}.\left[(1-\omega)\,\tilde{\sigma} + \omega\,\sigma\right] = \boldsymbol{h} \end{gathered} \tag{4.38}$$

since $\boldsymbol{\nabla}.([\nabla u]^T) = \boldsymbol{\nabla}(\boldsymbol{\nabla}.\boldsymbol{u}) = 0$ (the case $\omega = 0$ corresponds to the classical expression of the incompressible Navier-Stokes equations). From the corresponding Navier-Stokes problem for $\omega = 1$ with Dirichlet conditions on Γ_D for the velocity and the condition (4.37) on $\Gamma_N = \partial\Omega \setminus \Gamma_D$, the following variational formulation is deduced for a solenoidal test function $\boldsymbol{\varphi}$ which satisfies homogeneous Dirichlet conditions on Γ_D :

$$\begin{aligned} \frac{d}{dt}(\,\boldsymbol{u},\boldsymbol{\varphi}\,) \;&+\; (\,(\boldsymbol{u}.\boldsymbol{\nabla})\boldsymbol{u},\boldsymbol{\varphi}\,) \\ &+\; \frac{1}{\mathrm{Re}}\left(\sum_{i=1}^{d}(\,\boldsymbol{\nabla}u_{x_i},\boldsymbol{\nabla}\varphi_{x_i}\,) + \sum_{i,j=1}^{d}(\,\partial_{x_i}u_{x_j},\partial_{x_j}\varphi_{x_i}\,)\right) \\ &+\; \int_{\Gamma_N}\boldsymbol{\beta}.\boldsymbol{\varphi}\,ds = (\,\boldsymbol{h},\boldsymbol{\varphi}\,) \end{aligned} \tag{4.39}$$

In conclusion, it seems that the flux boundary condition (4.37) could be explicitely taken into account within a POD Galerkin system as the pseudo stress condition. Such a system has not been constructed yet in the literature.

4.5.2 Case e affine

We suppose that e is an affine function of f, that is of its coefficients $y \in \mathbb{R}^P$ in the natural monomial basis $(m_k)_{1\le k\le P}$ of the vector polynomials in M variables of degree 2 with M components (that natural basis is given for $M = 2$ in the appendix 4.5.4). In the following, y^g and y^α will denote the coefficients of f^g (the original POD-Galerkin system) and f^α (the calibrated system), respectively ; moreover $e(y,t) \equiv e(f,t)$, $\mathcal{E}(y) \equiv \mathcal{E}(f)$ and so on by notation abuse.
Hence, since e is affine,

$$\begin{array}{rccl} e(.,t): & \mathbb{R}^P & \longrightarrow & \mathbb{R}^M \\ & y & \longmapsto & E(t)\,y \;+\; e(0,t) \end{array} \tag{4.40}$$

where the columns of $E(t) \in \mathbb{R}^{M \times P}$ are the vectors $e(m_k, t) - e(0, t)$. Thus,

$$\left\langle \|e(f,t)\|_\Lambda{}^2 \right\rangle = \left\langle e(y,t)^T \, \Lambda \, e(y,t) \right\rangle = c \; - \; 2\, l^T \, y \; + \; y^T \, A \, y \tag{4.41}$$

with

$$c \;=\; \left\langle e(0,t)^T \, \Lambda \, e(0,t) \right\rangle, \tag{4.42}$$

$$l \;=\; -\left\langle E(t)^T \, \Lambda \, e(0,t) \right\rangle, \tag{4.43}$$

$$\text{and} \quad A \;=\; \left\langle E(t)^T \, \Lambda \, E(t) \right\rangle. \tag{4.44}$$

Since A is symmetric, the gradient of $(1-\alpha)\,\mathcal{E}$ at y is $2\,(1-\alpha)\,\chi_\mathcal{E}\;(Ay - l)$ with

$$\chi_\mathcal{E} = \frac{1}{\left\langle \|e(f^g,t)\|_\Lambda{}^2 \right\rangle}. \tag{4.45}$$

Therefore, if Π is the non-negative symmetric matrix associated to $\mathcal{D}$, the differential of $\mathcal{J}^\alpha$ is

$$\begin{array}{rccl} \nabla \mathcal{J}^\alpha(y): & \mathbb{R}^P & \longrightarrow & \mathbb{R} \\ & \delta y & \longmapsto & 2\,[\,(\,(1-\alpha)\chi_\mathcal{E}\,A \,+\, \alpha\,\chi_\mathcal{D}\,\Pi)\;y - (\,(1-\alpha)\chi_\mathcal{E}\,l \,+\, \alpha\chi_\mathcal{D}\,\Pi\,y^g)\,]^T\,\delta y, \end{array} \tag{4.46}$$

where

$$\chi_\mathcal{D} = \frac{1}{\|f^g\|_\Pi{}^2}, \tag{4.47}$$

and

$$\mathcal{J}^\alpha(y^\alpha) = \min_{y \in \mathbb{R}^P} \mathcal{J}^\alpha(y) \Longleftrightarrow B^\alpha \, y^\alpha = l^\alpha \tag{4.48}$$

with

$$B^\alpha = (1-\alpha)\chi_\mathcal{E}\,A \,+\, \alpha\chi_\mathcal{D}\,\Pi \qquad \text{and} \qquad l^\alpha = (1-\alpha)\chi_\mathcal{E}\,l \,+\, \alpha\chi_\mathcal{D}\,\Pi\,y^g. \tag{4.49}$$

If B^α is non-singular, the solution to the problem is unique and can be computed if this matrix is well conditioned.

4.5.3 Expressions of A and l for state and flow calibrations

With the notations of the preceding appendix, the expressions of A and l are with $\Lambda = \mathrm{I}_M$

- for the state calibration ($e = e_2$) :

$$A_{i,j} = \left\langle \left[\int_0^t m_i(a^e(\tau))\,d\tau \right]^T \int_0^t m_j(a^e(\tau))\,d\tau \right\rangle \tag{4.50}$$

$$\text{and} \qquad l_i = \left\langle \left[\textstyle\int_0^t m_i(a^e(\tau))\,d\tau \right]^T (a^e(t) - a^e(0)) \right\rangle \tag{4.51}$$

- and for the flow calibration ($e = e_3$) :

$$A_{i,j} = \left\langle m_i(a^e(t))^T \, m_j(a^e(t)) \right\rangle \tag{4.52}$$

$$\text{and} \qquad l_i = \left\langle m_i(a^e(t))^T \, \dot{a}^e(t) \right\rangle . \tag{4.53}$$

It is worth noting that, for $\Lambda = \mathrm{I}_M$, A can be written in block diagonal form with identical blocks for both calibration methods (it depends on the order of the vector monomial basis). Therefore, if Π is block diagonal with identical blocks as well, the linear system $B^\alpha \, y^\alpha = l^\alpha$ can be split into M similar linear systems of dimension P/M ; in fact, any jth component f_j^α of f^α is solution to an minimization problem defined by α, the diagonal blocks $\widetilde{\Pi}$ of Π, the vector $Y_{:,j}^g$ of coefficients of the jth component of f^g, a^e and $\dot{a}_j^e$: see below for more details in the case $M = 2$.

4.5.4 Linear systems obtained for $M = 2$

If $M = 2$, $P = 12$ and the natural vector monomial basis $(m_k)_{1 \le k \le P}$ is defined, for $a = (a_1 \quad a_2)^T$ and for all $1 \le i \le P/M$, by

$$m_i(a) = \begin{pmatrix} \widetilde{m}_i(a) \\ 0 \end{pmatrix}, \quad m_{i+(P/M)}(a) = m_{i+6}(a) = \begin{pmatrix} 0 \\ \widetilde{m}_i(a) \end{pmatrix} \tag{4.54}$$

where $(\widetilde{m}_k)_{1 \le k \le P/M}$ is the scalar monomial basis : $\widetilde{m}_1(a) = 1$, $\widetilde{m}_2(a) = a_1$, $\widetilde{m}_3(a) = a_2$, $\widetilde{m}_4(a) = {a_1}^2$, $\widetilde{m}_5(a) = a_1 \, a_2$ and $\widetilde{m}_6(a) = {a_2}^2$. Therefore, if $\Lambda = \mathrm{I}_M$,

$$A = \begin{pmatrix} \widetilde{A} & 0 \\ 0 & \widetilde{A} \end{pmatrix} \quad \text{with} \quad \widetilde{A} \in \mathbb{R}^{(P/M)\times(P/M)} \tag{4.55}$$

$$\text{and} \qquad \widetilde{A}_{i,j} = \begin{cases} \left\langle \displaystyle\int_0^t \widetilde{m}_i(a^e(\tau)) \, d\tau \int_0^t \widetilde{m}_j(a^e(\tau)) \, d\tau \right\rangle & \text{for } e = e_2 \\ \langle \widetilde{m}_i(a^e(t)) \, \widetilde{m}_j(a^e(t)) \rangle & \text{for } e = e_3. \end{cases} \tag{4.56}$$

For instance,

$$\widetilde{A}_{2,5} = \widetilde{A}_{5,2} = \left\langle \textstyle\int_0^t a_1^e(\tau) \, d\tau \ \int_0^t a_1^e(\tau) a_2^e(\tau) \, d\tau \right\rangle \quad \text{for} \quad e = e_2 \tag{4.57}$$

$$\text{and} \qquad \widetilde{A}_{2,3} = \widetilde{A}_{3,2} = \widetilde{A}_{1,5} = \widetilde{A}_{5,1} = \langle a_1^e(t) \, a_2^e(t) \rangle \quad \text{for} \quad e = e_3 \tag{4.58}$$

($\widetilde{A}_{2,3} \neq \widetilde{A}_{1,5}$ for $e = e_2$). Thus, if Π is block diagonal with identical blocks as A is, the system $B^\alpha \, y^\alpha = l^\alpha$ of dimension $P = 12$ naturally divides into $M = 2$ similar systems of dimension $P/M = 6$ which can be written as a single linear system :

$$\widetilde{B}^\alpha \, Y^\alpha = L^\alpha \tag{4.59}$$

where

- $\widetilde{B}^\alpha = (1-\alpha)\chi_\varepsilon\,\widetilde{A} + \alpha\chi_\mathcal{D}\,\widetilde{\Pi} \in \mathbb{R}^{(P/M)\times(P/M)}$ with $\Pi = \begin{pmatrix} \widetilde{\Pi} & 0 \\ 0 & \widetilde{\Pi} \end{pmatrix}$ (χ_ε and $\chi_\mathcal{D}$ are defined in appendix 4.5.2),
- $Y^\alpha = \begin{pmatrix} Y^\alpha_{:,1}\, Y^\alpha_{:,2} \end{pmatrix} \in \mathbb{R}^{(P/M)\times M}$ where the jth column $Y^\alpha_{:,j}$ of Y^α contains the coefficients of the scalar polynomial f^α_j in the monomial basis $(\widetilde{m}_k)$ (f^α_j is the jth component of f^α),
- $L^\alpha = (1-\alpha)\chi_\varepsilon\, L + \alpha\chi_\mathcal{D}\,\widetilde{\Pi}\, Y^g$ where $Y^g \in \mathbb{R}^{(P/M)\times M}$ contains the polynomial coefficients of the original POD-Galerkin system f^g as Y^α does for f^α and where $L \in \mathbb{R}^{(P/M)\times M}$ is defined by

$$L_{i,j} == \begin{cases} \left\langle \displaystyle\int_0^t \widetilde{m}_i(a^e(\tau))\,d\tau \quad (a^e_j(t) - a^e_j(0)) \right\rangle & \text{for} \quad e = e_2 \\ \left\langle \widetilde{m}_i(a^e(t)) \quad \dot{a}^e_j(t) \right\rangle & \text{for} \quad e = e_3. \end{cases} \tag{4.60}$$

For instance,

$$L_{6,1} = \left\langle \int_0^t a^e_2(\tau)^2\,d\tau \quad (a^e_1(t) - a^e_1(0)) \right\rangle \quad \text{for} \quad e = e_2 \tag{4.61}$$

and

$$L_{4,2} = \left\langle a^e_1(t)^2\,\dot{a}^e_2(t) \right\rangle \quad \text{for} \quad e = e_3. \tag{4.62}$$

In fact, any solution $Y^\alpha_{:,j}$ to $\widetilde{B}^\alpha\ Y^\alpha_{:,j} = L^\alpha_{:,j}$ corresponds to an optimal scalar polynomial f^α_j :

$$\mathcal{J}^\alpha_j(f^\alpha_j) \le \mathcal{J}^\alpha_j(f) \tag{4.63}$$

for all scalar polynomial f of degree 2 in M variables with

$$\mathcal{J}^\alpha_j(f) = (1-\alpha)\chi_\varepsilon \Lambda_{j,j}\,\mathcal{E}^j(f) + \alpha\chi_\mathcal{D}\,\mathcal{D}^j(f) \tag{4.64}$$

$$\text{where} \qquad \mathcal{E}^j(f) = \begin{cases} \left\langle \left[a^e_j(t) - a^e_j(0) - \displaystyle\int_0^t f(a^e(\tau))d\tau \right]^2 \right\rangle \text{ for } e = e_2 \\ \left\langle \left[\dot{a}^e_j(t) - f(a^e(t)) \right]^2 \right\rangle \text{ for } e = e_3, \end{cases} \tag{4.65}$$

$$\mathcal{D}^j(f) = \left\| f - f^g_j \right\|^2_{\widetilde{\Pi}} = (y - Y^g_{:,j})^T\,\widetilde{\Pi}\,(y - Y^g_{:,j}) \tag{4.66}$$

if $y \in \mathbb{R}^{P/M}$ is the vector of the polynomial coefficients of f in the scalar monomial basis $(\widetilde{m}_k)_{1\le k\le P/M}$.

4.6 Compléments à l'article

4.6.1 Conditionnement et moindres carrés

Dans le cas d'un opérateur $\langle . \rangle$ défini par $\langle . \rangle = \frac{1}{N} \sum_{i=1}^{N} .(t_i)$, d'une matrice Λ diagonale et pour $e = e_2$ ou $e = e_3$, le problème (4.9) est un problème aux moindre carrés classique si $\alpha = 0$.

En particulier, pour $\Lambda = \mathrm{I}_M$, il apparaît clairement que les matrices $\tilde{A} \in \mathbb{R}^{(P/M)\times(P/M)}$ et $L \in \mathbb{R}^{(P/M)\times M}$, définies dans l'appendice 4.5.4, peuvent se décomposer comme suit :

$$\tilde{A} = \frac{1}{N} C^T C \quad \text{et} \quad L = \frac{1}{N} C^T \tilde{L}.$$

avec $C \in \mathbb{R}^{N\times(P/M)}$ et $\tilde{L} \in \mathbb{R}^{N\times M}$. Ainsi résoudre $\tilde{A} Y = L$ équivaut à minimiser $\left\| C\, Y_{:,k} - \tilde{L}_{:,k} \right\|_2^2$ pour $1 \leq k \leq M$.

Puisque $\mathrm{Rang}(C) \leq \min(N, P/M)$, si $N < P/M$ alors $\mathrm{Rang}(C) < P/M$: la matrice $\tilde{A}$ n'est pas inversible si on ne dispose pas d'un nombre suffisant de données, c'est-à-dire si $N < P/M$. C'était le cas lors des expérimentations numériques présentées dans la section 4.3 menées avec le modèle à $M = 86$ modes de l'écoulement franchissant une marche. Ceci explique pourquoi les conditionnements sont énormes pour $\alpha = 0$ dans le cas de ce modèle et pourquoi le terme provenant de l'opérateur $\mathcal{D}$ est alors nécessaire pour obtenir un système bien conditionné.

Si le nombre de données est suffisant ($N \geq P/M$), alors la matrice C peut être de rang maximal P/M. Dans ce cas, $\tilde{A}$ est symétrique définie positive, donc inversible, et il est connu que le problème aux moindres carrés peut être résolu de manière plus efficace en exploitant une factorisation QR de la matrice C (voir [72] par exemple) : cela évite de calculer le produit $C^T C$, source potentielle importante d'erreurs d'arrondi, et de résoudre le système linéaire associé à la matrice $\tilde{A} = C^T C$ qui est souvent mal conditionnée.

Plus précisément, il est possible de calculer $Q \in \mathbb{R}^{N\times N}$ orthogonale et $R \in \mathbb{R}^{N\times(P/M)}$ trapézoïdale supérieure ($R_{i,j} = 0$ pour $i > j$) telles que $C = Q\,R$. Il vient alors, pour $x \in \mathbb{R}^{P/M}$ et $z \in \mathbb{R}^N$ quelconques,

$$\|C\,x - z\|_2^{\,2} = \left\| R\,x - Q^T z \right\|_2^{\,2}$$

et, puisque les $N - (P/M)$ dernières lignes de R sont nulles,

$$\|C\,x - z\|_2^{\,2} = \left\| \tilde{R}\,x - \tilde{Q}^T z \right\|_2^{\,2} + \sum_{k=(P/M)+1}^{N} \left[(Q^T z)_k \right]^2 \tag{4.67}$$

où $\tilde{R} \in \mathbb{R}^{(P/M)\times(P/M)}$ est la matrice des P/M premières lignes de R, $\tilde{Q} \in \mathbb{R}^{N\times(P/M)}$ la matrice des P/M premières colonnes de Q et où $(Q^T z)_k$ désigne le $k$ème élément du vecteur $Q^T z$. Il est clair que la solution du problème (4.67) aux moindres carrés est la solution du

système $\tilde{R}\,x = \tilde{Q}^T z$ qui est le plus souvent mieux conditionné que le système $C^T C\,x = C^T z$.

Ainsi, pour une valeur de α quelconque et un nombre suffisant de données ($N > P/M$), il est envisageable de définir une nouvelle méthode de calibration qui consiste à résoudre le système linéaire suivant :

$$\left[(1-\alpha)\chi_{\mathcal{E}}\,\tilde{R} + \alpha\chi_{\mathcal{D}}\,\tilde{\Pi}\right]\,Y^{\alpha} = (1-\alpha)\chi_{\mathcal{E}}\,\tilde{Q}^T\tilde{L} + \alpha\chi_{\mathcal{D}}\,Y^{g}$$

plutôt que le système (4.59) de l'appendice 4.5.4. Cette méthode devrait alors conduire à des systémes linéaires mieux conditionnés mais qui ne correspondent plus à un problème de minimisation de la forme (4.9) si $\alpha \neq 0$.

4.6.2 Calibration non-linéaire sous contrainte dynamique

Le cas du problème d'optimisation obtenu pour $e = e_1$ n'est abordé que très brièvement dans l'article, au début de la section 4.2.3.

C'est pourquoi nous complétons l'article par cette section qui présente certains aspects mathématiques et numériques du problème : elle décrit comment le gradient peut être calculé puis présente des essais de minimisation par des algorithmes de gradient conjugué.

Expression générale du gradient de $\mathcal{E}_1$

On va reprendre les notations de la section 4.5.1 : la variable y désigne le vecteur de $\mathbb{R}^P$ des coefficients polynômiaux du polynôme vectoriel f optimal recherché ; de plus, on aura $\mathcal{E}_1(y) \equiv \mathcal{E}_1(f)$ et $e_1(y,t) \equiv e_1(f,t)$ par abus de notation.

Ainsi, on cherche à exprimer, si il existe, le gradient de

$$\begin{array}{rccl} \mathcal{E}_1 : & \mathcal{O} \subset \mathbb{R}^P & \longrightarrow & \mathbb{R}^+ \\ & y & \longmapsto & \mathcal{E}_1(y) = \chi_{\mathcal{E}}\,\left\langle \|e_1(y,t)\|_{\Lambda}{}^2 \right\rangle \end{array} \quad \text{avec } \chi_{\mathcal{E}} = \frac{1}{\left\langle \|e(y^g,t)\|_{\Lambda}{}^2 \right\rangle}$$

où y^g est le vecteur des coefficients polynômiaux de f^g, $\langle . \rangle$ un opérateur linéaire continu de "moyenne temporelle" sur l'intervalle $[0,T]$, $\|.\|_{\Lambda}$ la norme induite par la matrice symétrique définie positive $\Lambda \in \mathbb{R}^{M\times M}$ et où l'opérateur e_1 à valeurs dans $\mathbb{R}^M$ est défini par

$$e_1(y,t) = a^e(t) - a(t)$$

sous la contrainte que a soit solution du problème de Cauchy $(\mathcal{P}_{a(y)})$ défini comme suit :

$$(\mathcal{P}_{a(y)}) \quad \left\{ \begin{array}{rcl} \dot{a}(t) & = & F(a(t),y,t) \\ a(0) & = & a^e(0) \end{array} \right. ,$$

avec $F : \mathbb{R}^M \times \mathbb{R}^P \times \mathbb{R} \longrightarrow \mathbb{R}^M$. Dans le cas général où s n'est pas nécessairement nul, F est définie par $F(a(t),y,t) = f(a(t)) + s(t)$ où f est le polynôme associé à y (voir l'équation (4.8)).

Dans la suite, on suppose que F est suffisamment régulier. Les dérivées partielles de F par rapport à a et y seront manipulées sous forme matricielle et notées respectivement $[\partial_a F(a,y,t)]$ de dimension $M \times M$ et $[\partial_y F(a,y,t)]$ de dimension $M \times P$. Il est important de noter que cette hypothèse, qui est vraie pour un système d'EDO polynômial autonome, nous assure que F est (localement) lipschitzienne par rapport à sa première variable a.

Le *théorème de Cauchy* nous assure donc l'existence et l'unicité d'une solution maximale a de $(\mathcal{P}_{a(y)})$ définie sur un ouvert $]\omega^-(a^e(0),y), \omega^+(a^e(0),y)[$. Les temps d'évasion satisfont bien sûr $\omega^-(a^e(0),y) < 0 < \omega^+(a^e(0),y)$. Si $\omega^+(a^e(0),y) \leq T$, la solution maximale a explose en temps fini avant l'instant T : $\mathcal{E}_1(y)$ n'est pas défini (mais vaut moralement $+\infty$). On est ainsi amené à restreindre $\mathcal{O}$ à un sous-ensemble de $\mathbb{R}^P$:

$$\mathcal{O} = \{y \in \mathbb{R}^P / \omega^+(a^e(0),y) > T\}.$$

On admettra que les temps d'évasion ω^- et ω^+ varient continûment avec y (si on les restreint aux sous-ensembles où ils prennent des valeurs finies) : $\mathcal{O}$, en tant qu'image réciproque de l'ouvert $]T,+\infty[$ par ω^+, est un ouvert. En revanche, $\mathcal{O}$ n'est pas nécessairement convexe, ni même connexe sans plus de propriétés sur F.

Afin de pouvoir exprimer clairement la différentielle de $\mathcal{E}_1$, deux fonctions notées $\tilde{\mathcal{E}}_1$ et Υ sont introduites :

$$\Upsilon \ : \ y \longmapsto \ a \text{ solution de } \mathcal{P}_{a(y)}$$

pour $y \in \mathcal{O}$ et

$$\tilde{\mathcal{E}}_1(a) \ : \ a \longmapsto \ \chi_\varepsilon \left\langle \|a^e - a\|_\Lambda^2 \right\rangle. \tag{4.68}$$

On a alors

$$\mathcal{E}_1(y) = \tilde{\mathcal{E}}_1 \circ \Upsilon(y) \tag{4.69}$$

Ainsi, en notant $\mathrm{D}g(x)(\delta x)$ la différentielle de g en x appliquée à δx, si $\tilde{\mathcal{E}}_1$ et Υ sont différentiables,

$$\mathrm{D}\mathcal{E}_1(y)(\delta y) = \mathrm{D}\tilde{\mathcal{E}}_1(\Upsilon(y)) \circ \mathrm{D}\Upsilon(y)(\delta y) = \mathrm{D}\tilde{\mathcal{E}}_1(\Upsilon(y))(\mathrm{D}\Upsilon(y)(\delta y)). \tag{4.70}$$

D'après (4.68), puisque l'opérateur $\langle . \rangle$ est linéaire et continu, la différentielle de $\tilde{\mathcal{E}}_1$ en a est

$$\mathrm{D}\tilde{\mathcal{E}}_1(a)(\delta a) = 2\,\chi_\varepsilon \left\langle (a - a^e)^T \Lambda\, \delta a \right\rangle \tag{4.71}$$

Il reste donc à exprimer la variation $\delta a = \mathrm{D}\Upsilon(y)(\delta y)$ de a, solution de $(\mathcal{P}_{a(y)})$, due à une variation δy de la "commande" y. On admettra que cette variation peut s'exprimer grâce au *système d'équations des perturbations* suivant :

$$(\mathcal{P}_{\delta a(\delta y)}) \quad \begin{cases} \dot{\delta a}(t) = [\partial_y F(a(t),y,t)]\delta y + [\partial_a F(a(t),y,t)]\delta a(t) \\ \delta a(0) = 0 \end{cases} \tag{4.72}$$

sous la contrainte que a satisfasse $(\mathcal{P}_{a(y)})$. Puisque le second membre de la première équation de $(\mathcal{P}_{\delta a(\delta y)})$ est linéaire par rapport à δa, $\delta a(t)$ est bien défini pour tout $t \in [0,T]$ et tout couple (a,y).

Ainsi, les équations (4.70), (4.71) et (4.72) nous donnent une expression de la différentielle $\mathrm{D}\mathcal{E}_1(y)$ appliquée à δy :

$$\mathrm{D}\mathcal{E}_1(y)(\delta y) = 2\,\chi_\varepsilon \left\langle (a - a^e)^T \Lambda\, \delta a \right\rangle \text{ sous } (\mathcal{P}_{a(y)}) \text{ et } (\mathcal{P}_{\delta a(\delta y)}). \tag{4.73}$$

Cependant, cette expression n'est pas pratique au sens où elle n'est pas sous la forme de Riesz, c'est-à-dire que la différentielle n'est pas exprimée à l'aide d'un gradient $\nabla\mathcal{E}_1(y) \in \mathbb{R}^P$ tel que

$$\forall \delta y \in \mathbb{R}^P \qquad \mathrm{D}\mathcal{E}_1(y)(\delta y) = \nabla\mathcal{E}_1(y)^T\, \delta y$$

($\mathrm{D}\mathcal{E}_1(y)$ étant une forme linéaire continue sur $\mathbb{R}^P$, le *théorème de Riesz* nous garantit l'existence d'un tel gradient). Puisque δy vit dans un espace de dimension fini, il est facile d'exprimer ce gradient grâce à (4.73). En effet, on a trivialement par linéarité :

$$\nabla\mathcal{E}_1(y) = \begin{bmatrix} \mathrm{D}\mathcal{E}_1(y)((1\ \ 0\cdots 0)^T) \\ \vdots \\ \mathrm{D}\mathcal{E}_1(y)((0\cdots 0\ \ 1)^T) \end{bmatrix}. \tag{4.74}$$

Cette expression générale du gradient permet de définir une procédure de calcul, cependant il existe, dans certains cas, d'autres expressions du gradient qui conduisent à une évaluation moins onéreuse.

Expressions du gradient à l'aide d'une variable adjointe

On considère d'abord le cas d'une somme continue :

$$\langle . \rangle = \int_0^T .(t)\, dt. \tag{4.75}$$

Le problème entre alors dans le cadre classique du *principe du minimum de Pontryagine* [20] et il est connu qu'une formulation pratique du gradient peut être obtenue en recourant à une variable *adjointe* q.

Notons q la solution maximale du problème de Cauchy

$$(\mathcal{P}_{q(a,y)}) \qquad \begin{cases} -\dot{q}(t) &= 2\Lambda(a(t) - a^e(t)) + [\partial_a F(a(t), y)]^T q(t) \\ q(T) &= 0 \end{cases}$$

Ce problème définit bien une fonction q définie pour tout $t \in [0, T]$ et pour tout couple (a, p). Il est alors possible d'exprimer formellement le gradient de la manière suivante :

$$\forall y \in \mathcal{O} \qquad \nabla\mathcal{E}_1(y) = \chi_\varepsilon \int_0^T [\partial_y F(a(t), y, t)]^T q(t)\, dt, \tag{4.76}$$

sous la condition que a soit la solution de $(\mathcal{P}_{a(y)})$ et q la solution de $(\mathcal{P}_{q(a,y)})$.

Remarque. L'équation d'évolution de l'adjoint peut être obtenue en cherchant à annuler la dérivée du *Lagrangien*

$$\mathcal{L}(\tilde{a}, y, q) = \int_0^T \|a^e(t) - \tilde{a}(t) - a^e(0)\|_\Lambda{}^2 \, dt + \int_0^T q(t)^T (F(\tilde{a}(t) + a^e(0), y, t) - \frac{d}{dt}\tilde{a}(t)) \, dt$$

par rapport à la variable $\tilde{a}$ qui vit dans un espace de fonctions régulières s'annulant en $t = 0$ ($\tilde{a}(0) = 0$). En effet,

$$\partial_{\tilde{a}}\mathcal{L}(\tilde{a}, y, q)(\delta\tilde{a}) = \int_0^T \left[2\Lambda(a - a^e) + [\partial_a F(a, y)]^T q + \dot{q}\right]^T \delta\tilde{a} \, dt + q(T)\delta\tilde{a}(T)$$

pour tout $\delta\tilde{a}$ avec $a(t) = \tilde{a}(t) + a^e(0)$ (et $\delta\tilde{a}(0) = 0$).

On suppose maintenant que $\langle . \rangle$ est défini par une somme discrète :

$$\langle . \rangle = \frac{1}{N} \sum_{i=1}^{N} .(t_i) \, dt. \tag{4.77}$$

Cette définition est intéressante dans les situations où on ne dispose pas de données $a^e(t_i)$ en des instants t_i suffisamment proches pour amener à une évaluation précise de l'opérateur continu (4.75). En pratique, il est souvent possible de calculer des modes POD représentatifs d'un écoulement en recourant à un petit nombre de clichés bien répartis dans l'intervalle de temps $[0, T]$; si la POD est calculée par la méthode des clichés, les coefficients temporels a_i^e de référence ne sont alors connus que pour les instants correspondants aux clichés : les coefficients $a_i^e(t_j)$ peuvent ne pas être connus en des instants très proches.

Ainsi, $\mathcal{E}_1(y) = \chi_\varepsilon \frac{1}{N} \sum_{i=1}^N \mathcal{E}_1^i(y)$ avec $\mathcal{E}_1^i(y) = \|a^e(t_i) - a(t_i)\|_\Lambda{}^2$ sous la contrainte $(\mathcal{P}_{a(y)})$ pour tout i : on se ramène ainsi à N fonctionnelles de coût qui ne dépendent que de l'état final. De nouveau, il est possible d'exprimer le gradient du coût $\mathcal{E}_1^i(y)$ à l'aide d'un adjoint q_i :

$$\forall y \in \mathcal{O} \qquad \nabla\mathcal{E}_1^i(y) = \int_0^{t_i} [\partial_y F(a(t), y, t)]^T q_i(t) \, dt, \tag{4.78}$$

où q_i est défini par

$$(\mathcal{P}_{q_i(a,y)}) \qquad \left\{ \begin{array}{rcl} -\dot{q}_i(t) & = & [\partial_a F(a(t), y)]^T q_i(t) \\ q_i(t_i) & = & 2(a(t_i) - a^e(t_i)) \end{array} \right.$$

Illustration sur un exemple de dimension un

Considérons le problème avec contrainte non-linéaire de dimension un ($M = P = 1$) le plus simple rentrant dans le cadre précédent :

$$\mathcal{E}_1(y) = \int_0^T (a(t) - a_0)^2 \, dt \qquad \text{sous} \qquad (\mathcal{P}_{a(y)}) \quad \left\{ \begin{array}{rcl} \dot{a}(t) & = & -y\, a(t)^2 \\ a(0) & = & a_0 \end{array} \right. ,$$

autrement dit $\chi_\varepsilon = \Lambda = 1$, $\langle . \rangle = \int_0^T .(t)\,dt$ et $F(a,y,t) = -y\,a^2$. Ici, la solution maximale de $(\mathcal{P}_{a(y)})$ est

$$a(t) = \frac{a_0}{a_0\,y\,t+1} \text{ sur }]\omega^-(a_0,y), \omega^+(a_0,y)[\tag{4.79}$$

avec des temps d'évasion qui varient continûment avec y

$$\omega^-(a_0,y) = \begin{cases} -\infty \text{ si } y\,a_0 \leq 0 \\ -\dfrac{1}{a_0\,y} \text{ sinon} \end{cases} \qquad \omega^+(a_0,y) = \begin{cases} -\dfrac{1}{a_0\,y} \text{ si } y\,a_0 < 0 \\ +\infty \text{ sinon.} \end{cases}$$

On a donc $\mathcal{O} =]o^-(a_0), o^+(a_0)[$ avec

$$o^-(a_0) = \begin{cases} -\infty \text{ si } a_0 \leq 0 \\ -\dfrac{1}{a_0\,T} \text{ sinon} \end{cases} \qquad o^+(a_0) = \begin{cases} -\dfrac{1}{a_0\,T} \text{ si } a_0 < 0 \\ +\infty \text{ sinon.} \end{cases}$$

Ainsi, $\mathcal{O} \neq \mathbb{R}$ sauf si $a_0 = 0$ et on voit bien comment construire un problème où $\mathcal{O}$ n'est plus connexe : il suffit de remplacer $\mathcal{E}_1$ par $\mathcal{E}_1 \circ \Psi$ et de choisir $\Psi : \mathbb{R} \longrightarrow \mathbb{R}$ tel que $\Psi^{-1}(\mathcal{O})$ ne soit pas connexe. On pourrait par exemple choisir $\Psi(y) = y^2 - \frac{2}{a_0\,T}$ si $a_0 > 0$.

Le calcul formel montre qu'alors, considérant la solutions $\delta a(t) = -\frac{a_0^2\,t}{(a_0\,y\,t+1)^2}$ du système d'équations des perturbations pour $\delta y = 1$ ($\dot{\delta a}(t) = -a(t)^2 - 2\,y\,a(t)\,\delta a(t)$ et $\delta a(0) = 0$) et l'adjoint $q(t) = 2\frac{ya_0^2((1+2ya_0T)t^2 - 2ya_0T^2t - T^2)}{(2ya_0T+1)^2}$ (solution de $-\dot{q}(t) = 2(a(t)-a_0) - 2\,y\,a(t)^2\,q(t)$ et $q(T) = 0$), les deux manières précédentes de définir le gradient (équations (4.74) et (4.76)) conduisent au même résultat.

Calcul numérique du gradient

Dans le cas d'un opérateur $\langle . \rangle$ de la forme discrète (4.77), le gradient peut être calculé par l'évaluation des expressions (4.74) ou (4.78). Celles-ci conduisent à des coûts informatiques qui peuvent être très différents : on préfèrera évaluer l'expression générale (4.74) si P est petit devant N, mais utiliser le calcul via l'adjoint dans le cas contraire.

Si $\langle . \rangle$ est défini par une somme continue (équation (4.75)), alors le calcul par l'adjoint est plus efficace. Le gradient est alors évalué en trois étapes :

1. Calcul de l'état a, solution de $(\mathcal{P}_{a(y)})$.

2. Calcul de l'adjoint q, solution de $(\mathcal{P}_{q(a,y)})$.

3. Calcul de $\nabla\mathcal{E}_1$ via l'équation (4.76) (puis calcul de $\nabla\mathcal{J}^\alpha$).

Des tests, présentés dans la section suivante, ont été menés dans le cas d'une somme continue. Des schémas Adams-Bashforth d'ordre quatre ont été utilisés pour les deux premières étapes. Il faut noter que l'utilisation d'une méthode de Runge-Kutta d'ordre quatre n'est pas possible pour la deuxième étape puisque le second membre de l'EDO ne peut être évalué qu'aux instants où a et a^e sont connus. De plus, il faut nécessairement calculer a

aux mêmes instants que a^e. Lors de la troisième étape, l'intégration temporelle du terme $[\partial_y F(a(t), y, t)]^T q(t)$ (équation (4.74)) a été effectuée à l'aide d'un schéma de Newton-Cotes à quatre points (*Simpson's 3/8 rule*).

4.6.3 Tests numériques

Il a été choisi d'essayer de déterminer un minimum local de la fonction coût $\mathcal{J}^\alpha$ du problème (4.9) par des algorithmes de gradient conjugué. Ceux-ci sont souvent plus efficaces que les algorithmes de plus grande pente, qui ne requièrent également que des évaluations de la fonction à minimiser et de son gradient, mais moins que les algorithmes de type Newton. Ces derniers nécessitent le calcul de la hessienne de la fonction coût (ou une approximation), ce qui demanderait un très gros travail de programmation ; de plus, le coût d'une seule évaluation de la hessienne serait très important pour notre problème.

Les méthodes de Fletcher-Reeves, Polak-Ribière ainsi que les recherches linéaires de Goldstein, de Wolfe et une recherche de Wolfe avec ajustement cubique ont donc été codées : le lecteur trouvera toutes les informations nécessaires sur ces méthodes dans l'ouvrage de Bonnans *et al.* [9].

Il est important de noter que la fonction $\mathcal{J}^\alpha$ du problème de calibration obtenu pour $e = e_1$ n'est en général pas définie dans $\mathbb{R}^P$ au complet, mais dans un sous-ensemble ouvert $\mathcal{O}$. En conséquence, **les algorithmes de recherche linéaire doivent tenir compte du fait que le coût $\mathcal{J}^\alpha(y)$ peut prendre la valeur $+\infty$** : il est nécessaire de pouvoir gérer cette situation du point de vue informatique par des tests appropriés.

Les tests concerneront des systèmes dynamiques quadratiques de dimension $M = 1$:

$$\dot{a}_1(t) = F(a(t), y, t) = y_1 + y_2\, a_1(t) + y_3\, a_1(t)^2, \quad P = 3 \quad \text{et} \quad y = (y_1 \;\; y_2 \;\; y_3)^T \; ; \qquad (4.80)$$

ou $M = 3$: $\dot{a}(t) = F(a(t), y, t)$, $P = 30$ et $y = (y_1 \cdots y_{30})^T$. Dans la suite, les vecteurs de $\mathbb{R}^{30}$ des coefficients polynômiaux des systèmes dynamiques quadratiques de dimension $M = 3$ de Lorenz et de Rössler seront notés respectivement y^L et y^R (ces systèmes ont été proposés dans [57] et [80]). Le système de Lorenz provient d'une réduction à trois modes spatiaux obtenue par une projection de Galerkine d'EDP qui proviennent d'une simplification des équations de Rayleigh-Bénard pour un problème de convection thermique : il est proche par nature des modèles POD-Galerkine réduits incompressibles ; consulter [22].

Les résultats présentés ici ont été obtenus en initiant l'algorithme de gradient conjugué à partir du polynôme de référence y^g. La donnée a^e de référence à partir de laquelle le système d'EDOs polynômial est calibré est obtenue par la simulation numérique d'un système polynômial dont le vecteur des coefficients sera noté y^e : y^e est donc une solution optimale du problème (4.9) si $\alpha = 0$.

Validations

Les routines de calcul du gradient ont été validées en comparant leurs sorties avec des approximations par DF (Différences Finies) du gradient (les DF sont beaucoup plus coûteuses et sont très sensibles aux valeurs des "perturbations" δy_i de y qui sont employées). Ainsi, dans le cas d'une solution de référence a^e obtenue par simulation du système de Rössler ($y^e = y^R$) sur l'intervalle de temps $[0, 5]$ ($T = 5$) pour un pas de temps $\Delta t = 10^{-3}$ et pour la condition initiale $(a_1(0), a_2(0), a_3(0)) = (0, 15, 0)$, les routines donnent les dérivées partielles par rapport aux coefficients polynômiaux de a_1 suivantes :

	$\alpha = 0$ ($\nabla \mathcal{J}^\alpha = \nabla \mathcal{E}_1$)		$\alpha = 1$ ($\nabla \mathcal{J}^\alpha = \nabla \mathcal{D}$)	
$\partial_{y_i} \mathcal{J}^\alpha(y)$	approx. par DF	calcul via adjoint	approx. par DF	calcul via adjoint
$i = 1$	-8.43671713E-02	-8.43673726E-02	2.01184914E-02	2.01184754E-02
$i = 2$	0.30904991	0.30920889	-2.11244045E-02	-2.11243992E-02
$i = 3$	-1.18540688E-04	-1.18136576E-04	-1.00602749E-03	-1.00592377E-03
$i = 4$	-0.40759064	-0.40759105	0.	0.
$i = 5$	1.32721359	1.32721443	0.	0.
$i = 6$	-3.93613478E-04	-3.93573981E-04	0.	0.
$i = 7$	-4.51386755	-4.51625682	0.	0.
$i = 8$	1.93756197E-03	1.93791038E-03	0.	0.
$i = 9$	-5.24019465E-07	-3.75362675E-07	0.	0.
$i = 10$	-2.07331670E-02	-2.07442659E-02	0.	0.

au point $y = \frac{1}{2} y^R$ et pour le polynôme de référence $y^g = y^L$ de Lorenz (avec des perturbations de l'ordre de 10^{-9} pour les composantes de y lors des DFs). Il y a donc une très bonne correspondance entre les valeurs de la routine de calcul de la fonction coût (qui est utilisée lors des DFs) et la routine de calcul du gradient par l'adjoint.

Les algorithmes de gradient conjugué et de recherche linéaire ont été testés pour plusieurs fonctions, en particulier la fonction banane (encore appelée Rosenbrock valley et deuxième fonction de De Jong)

$$f_B(y_1, \cdots, y_P) = \sum_{j=1}^{P-1} ((100(y_{j+1} - y_j^2))^2 + (1 - y_j^2)).$$

de dimension $P = 2$. Alors, on obtient les nombres d'itérations de l'algorithme de gradient conjugué suivants pour f_B

	# itér. avec Fletcher-Reeves	# itér. avec Polak-Ribière
Goldstein	572	182
Wolfe	377	189
Wolfe avec ajustement cubique	253	219

en partant de $(-2, 2)$ (le minimum est 0 en $(1, 1)$), pour le critère d'arrêt $\|\nabla f_B\|_2 < 10^{-3}$

et un nombre maximum d'itérations de la recherche linéaire de 15 pour chaque itération de l'algorithme de gradient conjugué. Les tests ont montré que les méthodes de gradient conjugué étaient satisfaisantes. Néanmoins pour certaines fonctions, comme f_B qui a été créée spécialement afin d'éprouver l'efficacité des algorithmes d'optimisation, la convergence peut être lente.

Résultats

La série de tests concerne la calibration d'un modèle quadratique de dimension un : $M = 1$ et $P = 3$, voir l'équation (4.80). La donnée a_1^e de référence a été obtenue par simulation du système polynômial de coefficients $y^e = (0\ \ 1\ \ -1)^T$ sur l'intervalle de temps $[0, 5]$ $(T = 5)$, pour un pas de temps de temps $\Delta t = 10^{-3}$ et la condition initiale $a_1(0) = 0.1$. Le vecteur du polynôme de référence est $y^g = (0\ \ -1\ \ 0)^T$, c'est aussi le vecteur à partir duquel les algorithmes de descente sont initiés. Le critère d'arrêt choisi est $\|\nabla \mathcal{J}^\alpha\|_2 < 10^{-3}$.

Les résultats ont été obtenus par la méthode de Polak-Ribière combinée avec une recherche linéaire de Wolfe qui utilise un ajustement cubique (25 itérations de la recherche linéaire sont autorisées au maximum).

La minimisation de $\mathcal{J}^\alpha$ donne alors les valeurs finales suivantes :

α	$\mathcal{E}_1(y^\alpha)$	$\mathcal{D}(y^\alpha)$	$\|y^\alpha - y^e\|_2 / \|y^e\|_2$
1	1.	0.	1.58113883
0.75	0.38585634	8.12335496E-02	1.55343399
0.5	0.10302424	0.241454665	1.51448254
0.25	2.92527154E-02	0.355726667	1.49539568
0.1	1.23352440E-02	0.439113371	1.47922029
10^{-2}	5.61546530E-03	0.612715152	1.30186631
10^{-3}	1.07894743E-03	2.291886778	0.56122027
0	3.94936817E-09	4.998650007	2.55735314E-04

où y^α désigne le vecteur final obtenu. Ainsi, quand α varie de 1 à 0, les coûts $\mathcal{D}(y^\alpha)$ et $\mathcal{E}_1(y^\alpha)$ sont bien décroissants et croissants, de plus le vecteur "optimal" y^α tend vers y^e : la calibration est efficace.

Intéressons-nous maintenant au coût informatique de ces calibrations en observant le nombre d'itérations de descente effectuées et le nombre d'évaluations de la fonction coût et de son gradient qui ont été réalisées lors des recherches linéaires :

α	# itér. de descente	# calculs $\mathcal{J}^\alpha$ dans rech. lin.	# calcul $\nabla\mathcal{J}^\alpha$ dans rech. lin.
1	0	0	0
0.75	6	26	26
0.5	10	38	38
0.25	15	55	55
0.1	28	138	138
10^{-2}	42	182	182
10^{-3}	106	528	527
0	160	735	732

Il apparaît que le nombre de calculs nécessaire à la calibration varie beaucoup avec α et peut devenir particulièrement important pour de petites valeurs de α.

Un essai a ensuite été mené dans le cas $M = 3$ avec les paramètres suivants : $y^e = y^L$, $T = 10$, $\Delta t = 10^{-3}$ et $a^e(0) = (1\ 1\ 1)^T$ (données a^e générées par simulation du modèle de Lorenz sur l'intervalle de temps $[0, 10]$), $y^g = 0.9\, y^L$. Puisque y^g sert aussi de vecteur initial, la descente commence donc *a priori* d'un vecteur proche d'une solution optimale, du moins si $\alpha = 1$ et $\alpha = 0$ car alors y^g et, respectivement, y^L sont optimaux. Alors, le critère d'arrêt $\|\nabla\mathcal{J}^\alpha\|_2 < 10^{-3}$ n'est toujours pas atteint après 100 itérations.

4.6.4 Discussion

Il apparaît donc que le choix $e = e_1$ conduit à une calibration qui est particulièrement coûteuse en temps de calcul pour de petits systèmes ($M = 1$ ou $M = 3$), puisque le calcul du gradient est relativement coûteux et que la convergence vers un minimum local peut être très lente.

La méthode de calibration associée à $e = e_1$ ne semble donc pas pratique, d'autant plus qu'une calibration menée avec $e = e_2$ ou $e = e_3$ offre des résultats satisfaisants pour un coût informatique raisonnable. Il faut aussi noter que le codage de la calibration pour $e = e_2$ et $e = e_3$ est moins fastidieux.

En conséquence, nous déconseillons le choix $e = e_1$ d'un point de vue pratique et c'est pourquoi nous ne l'avons pas traité plus en détails dans l'article.

Conclusions et perspectives

Trois méthodes de calibration ont été définies à partir du même principe : optimiser les coefficients polynômiaux (indépendants du temps) du modèle réduit en exploitant les données temporelles de la POD.

La première méthode, qui consiste à minimiser une fonction coût définie via une contrainte dynamique non linéaire ($e = e_1$), est limitée par les coûts importants de calculs qu'elle entraîne, du moins si des algorithmes de gradient conjugué sont employés. De plus, l'utilisation d'algorithmes de type Newton n'est pas aisée, car le calcul de la hessienne de la

fonction coût s'avère coûteux et difficile à implémenter. Cependant, il ne faut pas exclure que la convergence vers un minimum local puisse être obtenue de manière efficace par une autre méthode numérique d'optimisation.

Les deux autres méthodes (*state and flow calibration*) sont efficaces : pour les deux écoulements étudiés, des modèles réduits satisfaisants sont obtenus pour un coût raisonnable (en particulier pour $e = e_3$).

Il faut noter que, dans tous les cas, la calibration est très sensible à la valeur de α : le problème du choix automatique de ce paramètre se pose.

Les méthodes de calibration devraient permettre d'exploiter la modélisation POD-Galerkine réduite pour une plus grande variété d'écoulements qu'auparavant, en particulier des écoulements dont le nombre de Reynolds est relativement grand et qui menaient à des modèles non calibrés très peu précis, voire instables, pour un nombre de modes POD raisonnable. Elles sont effectivement capables de stabiliser un modèle dont la solution numérique explose avant l'instant final T, comme des essais sur le modèle f^w de Rempfer de l'écoulement franchissant la marche l'ont montré (voir la section 3.4).

On peut néanmoins penser que l'intérêt pratique de la calibration est limité dans la mesure où celle-ci est effectuée sur un jeu de données spécifique : le modèle est calibré pour des conditions environnementales précises, c'est-à-dire pour une seule valeur de $\boldsymbol{h}$, $\boldsymbol{g}$ et de $\boldsymbol{\beta}$ (et $\boldsymbol{u_0}$).

Cependant, cette limitation ne pose pas de problème car elle est intrinsèque à la méthode POD-Galerkine : les modes POD, donc le modèle réduit, ne sont pas universels mais simplement représentatifs d'un jeu de données. En outre, même si les modes sont issus de plusieurs jeux, qui correspondent chacun à des conditions environnementales particulières, il est alors possible :

- d'effectuer une calibration globale du modèle en redéfinissant $\mathcal{E}$ comme la somme des fonctions coûts définies sur chaque jeu de données ;
- de calibrer le modèle pour chaque jeu, puis de construire un modèle global par une interpolation des modèles calibrés obtenus (c'est ce qu'ont fait Galletti *et al.* [25] pour construire un modèle valable pour différents nombres de Reynolds).

D'ailleurs, dans le cadre d'un problème de contrôle (voir le chapitre suivant), il est préférable sinon indispensable de n'exploiter un modèle réduit que dans une *région de confiance* limitée, c'est-à-dire pour des conditions environnementales proches de celles par lesquelles les modes POD puis le modèle ont été obtenus. Ceci peut être notamment réalisé en utilisant les méthodes POD-Galerkine à région de confiance proposées et testées lors du travail de thèse de Fahl : consulter [19] et [4].

Chapitre 5

La modélisation POD-Galerkine et le contrôle actif d'écoulements

Sommaire

Il a été montré au chapitre 4.1.1, du moins en détails dans le cas d'un écoulement incompressible, que la méthode POD-Galerkine permettait de construire un modèle dynamique approché qui prenne en compte explicitement et de manière exacte l'action de l'environnement sur l'écoulement, à travers les conditions aux bords ou le terme $\boldsymbol{h}$ qui modélise les forces extérieures. Ainsi, il apparaît que la modélisation proposée est tout à fait capable de tenir compte de l'effet d'un actionneur sur l'écoulement, par exemple un système générant un champ de force électromagnétique ou un actionneur qui souffle ou aspire du fluide.

En outre, la dimension des modèles POD-Galerkine est en pratique très petite par rapport à celle des systèmes d'EDOs qui sont obtenus après une discrétisation spatiale "classique" des équations de Navier-Stokes par différences, éléments ou volumes finis. La modélisation POD-Galerkine est donc très attractive en termes de coût de calculs pour le contrôle, c'est-à-dire pour la définition de la commande d'un actionneur afin d'optimiser, sinon améliorer, les caractéristiques d'un écoulement instationnaire.

Ce chapitre est consacré au contrôle des écoulements instationnaires basé sur la modélisation POD-Galerkine et il est divisé en deux sections.

Tout d'abord, la section 5.1 expose comment la modélisation POD-réduite peut être utilisée pour définir une fonction "modèle" réduite à partir de la fonction coût originale d'un problème de contrôle instationnaire d'un écoulement incompressible. Comme nous allons le voir, cela consiste essentiellement à remplacer la contrainte du problème de contrôle, c'est-à-dire les équations de Navier-Stokes et leurs conditions aux limites, par un modèle POD-Galerkine réduit. Ensuite, les stratégies itératives qui permettent de tirer partie de cette modélisation pour minimiser la fonction coût originale sont présentées. Cette section reprend principalement les travaux de Graham [27, 28], Ravindran [75, 75] et Fahl [19].

La section 5.2 présente ensuite des essais numériques qui ont été menés sur un problème de contrôle de type *flow tracking* pour un écoulement laminaire ($\mathrm{Re} = 100$) et bidimensionnel. Un actionneur modélisant un système de soufflage/aspiration est placé sur un obstacle circulaire, du côté du sillage.

5.1 Modélisation POD-Galerkine réduite pour le contrôle actif

Cette section présente le problème de contrôle considéré, puis montre comment la méthode POD-Galerkine permet de le modéliser par un problème d'optimisation dynamique de dimension réduite.

5.1.1 Définition et modélisation réduite du problème de contrôle

Le problème étudié ici est celui du contrôle d'un écoulement régi par les équations de Navier-Stokes incompressibles (2.19)-(2.20), par un actionneur qui peut être modélisé par une condition de Dirichlet sur la vitesse. Ce type d'actionneur peut correspondre à un système de soufflage/aspiration ou à un déplacement de parois auxquelles le fluide adhère,

mais sans modification du domaine occupé par le fluide (par exemple la rotation d'un obstacle de forme circulaire).

Définition du problème de contrôle

Les actionneurs sont modélisés par une condition de Dirichlet instationnaire sur la vitesse de la forme

$$\boldsymbol{u}(\boldsymbol{x},t) = \sum_{i=1}^{I} c_i(t)\,\boldsymbol{g_i}(\boldsymbol{x}) \quad \text{sur} \quad \Gamma_A$$

où les $c_i(t)$ sont les fonctions de commande des actionneurs, $\boldsymbol{g_i}$ leurs profils spatiaux de vitesses et $\Gamma_A \subset \Gamma_D$ le sous-ensemble du bord du domaine fluide Ω où les actionneurs agissent. Les autres conditions de Dirichlet sont supposées stationnaires pour simplifier la présentation :

$$\boldsymbol{u}(\boldsymbol{x},t) = \boldsymbol{g}_s(\boldsymbol{x}) \quad \text{sur} \quad \Gamma_D \setminus \Gamma_A.$$

Ainsi les conditions de Dirichlet sur Γ_D s'écrivent

$$\boldsymbol{u}(\boldsymbol{x},t) = \boldsymbol{g}(\boldsymbol{x},t) = \begin{cases} \displaystyle\sum_{i=1}^{I} c_i(t)\,\boldsymbol{g_i}(\boldsymbol{x}) & \text{pour} \quad \boldsymbol{x} \in \Gamma_A \\ \\ \boldsymbol{g}_s(\boldsymbol{x}) & \text{pour} \quad \boldsymbol{x} \in \Gamma_D \setminus \Gamma_A. \end{cases} \tag{5.1}$$

L'état de l'écoulement incompressible est donné par $\boldsymbol{u}$ et $\boldsymbol{\nabla} p$, donc, dans le cadre le plus général, les fonctions coûts dépendent de ces deux variables (éventuellement de la pression p moyennant une condition qui permette de la définir de manière unique). Les modèles POD-Galerkine incompressibles ne régissent cependant que la vitesse $\boldsymbol{u}$ du fluide : il n'est pas possible de considérer une fonction coût qui dépende de la pression sans compliquer sérieusement la modélisation. Ainsi, comme tous les auteurs ayant menés des investigations sur l'utilisation de la méthode POD-Galerkine pour le contrôle d'un écoulement incompressible, nous considérerons ici une fonction coût à minimiser où n'intervient pas la pression, de la forme

$$\mathcal{J}^{\alpha}(\boldsymbol{u}(\boldsymbol{x},t), c(t)) = (1-\alpha) \int_0^T \left\| \mathcal{G}(\boldsymbol{u}(\boldsymbol{x},t),t) \right\|_{\mathrm{L}^2}{}^2 dt + \alpha \int_0^T \left\| c(t) \right\|_2{}^2 dt \tag{5.2}$$

avec α un paramètre de $[0,1]$, $c(t) = (c_1(t) \cdots c_I(t))^T$ le vecteur de dimension I des commandes de contrôle et $\mathcal{G}(.,t)$ un opérateur affine, par exemple

$$\mathcal{G}(\boldsymbol{u}(\boldsymbol{x},t),t) = \boldsymbol{\nabla} \times \boldsymbol{u}(\boldsymbol{x},t) \tag{5.3}$$

ou, pour une fonction coût de type *flow tracking*,

$$\mathcal{G}(\boldsymbol{u}(\boldsymbol{x},t),t) = \boldsymbol{u}(\boldsymbol{x},t) - \boldsymbol{u}^R(\boldsymbol{x},t) \tag{5.4}$$

où $\boldsymbol{u}^R(\boldsymbol{x},t)$ est un champ de vitesses de référence.

Ainsi, le problème de contrôle actif se définit comme suit :

Définition 5 (Formulation forte du problème de contrôle actif instationnaire) *Pour α, ω, $\boldsymbol{\beta}$, $\boldsymbol{h}$, $\boldsymbol{g_S}$ et $(\boldsymbol{g_i})_{1\leq i\leq I}$ donnés, trouver la commande $c(t)$ qui minimise $\mathcal{J}^\alpha(\boldsymbol{u},c)$ sous la contrainte que $\boldsymbol{u}$ satisfasse les équations de Navier-Stokes incompressibles (2.19)-(2.20) et les conditions aux limites (2.24)-(2.25) (avec $P=0$) où $\boldsymbol{g}$ est définie en fonction de $\boldsymbol{g_S}$, $(\boldsymbol{g_i})_{1\leq i\leq I}$ et $c(t)$ par la relation (5.1).*

Prise en compte de la commande des actionneurs dans le modèle POD-Galerkine et problème de contrôle réduit

Le modèle POD-Galerkine réduit est construit sous l'hypothèse que les données $\boldsymbol{u}^e$ proviennent d'une simulation des équations de Navier-Stokes incompressibles pour des conditions aux limites de Dirichlet similaires au problème de contrôle (définition 5) et pour une commande $c^e(t) = (c_1^e(t)\cdots c_I^e(t))^T$ connue.

Afin de pouvoir appliquer la modélisation POD-Galerkine exposée dans la section 2.3.1, il faut être en mesure de définir des champs $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ et $\bar{\boldsymbol{u}}$, de divergence nulle, qui vérifient les conditions

$$\bar{\boldsymbol{u}}^{\boldsymbol{e}}(\boldsymbol{x},t) = \begin{cases} \displaystyle\sum_{i=1}^{I} c_i^e(t)\,\boldsymbol{g_i}(\boldsymbol{x}) \text{ sur } \Gamma_A \\ \\ \boldsymbol{g}_s(\boldsymbol{x}) \text{ sur } \Gamma_D \setminus \Gamma_A \end{cases} \quad \text{et} \quad \bar{\boldsymbol{u}}(\boldsymbol{x},t) = \begin{cases} \displaystyle\sum_{i=1}^{I} c_i(t)\,\boldsymbol{g_i}(\boldsymbol{x}) \text{ sur } \Gamma_A \\ \\ \boldsymbol{g}_s(\boldsymbol{x}) \text{ sur } \Gamma_D \setminus \Gamma_A \end{cases}$$

pour $c(t)$ quelconque.

Pour ce faire, il suffit de considérer la solution $(\boldsymbol{\psi}, \boldsymbol{\nabla} p)$ du problème de Stokes

$$\left.\begin{aligned} \boldsymbol{\nabla}.\boldsymbol{\psi} &= 0 \\ -\Delta\boldsymbol{\psi} + \boldsymbol{\nabla} p &= 0 \end{aligned}\right\} \text{ sur } \Omega \tag{5.5}$$

pour des conditions de Dirichlet sur le bord $\partial\Omega = \Gamma$ bien choisies. En effet, si $\boldsymbol{\psi_s}$ est solution de (5.5) pour la condition

$$\boldsymbol{\psi_s}(\boldsymbol{x}) = \begin{cases} \boldsymbol{g}_s(\boldsymbol{x}) \text{ sur } \Gamma_D \setminus \Gamma_A \\ \boldsymbol{0} \text{ sur } \Gamma \setminus (\Gamma_D \setminus \Gamma_A) \end{cases} \tag{5.6}$$

et si, pour tout $i \in [\![1, I]\!]$, $\boldsymbol{\psi_i}$ est solution de (5.5) pour la condition

$$\boldsymbol{\psi_i}(\boldsymbol{x}) = \begin{cases} \boldsymbol{g_i}(\boldsymbol{x}) \text{ sur } \Gamma_A \\ \\ \boldsymbol{0} \text{ sur } \Gamma \setminus \Gamma_A, \end{cases} \tag{5.7}$$

alors

$$\bar{\boldsymbol{u}}^{\boldsymbol{e}} = \boldsymbol{\psi_S} + \sum_{i=1}^{I} c_i^e(t)\,\boldsymbol{\psi_i} \qquad \text{et} \qquad \bar{\boldsymbol{u}} = \boldsymbol{\psi_S} + \sum_{i=1}^{I} c_i(t)\,\boldsymbol{\psi_i} \tag{5.8}$$

conviennent. Alors, en effectuant la POD de $\boldsymbol{u}^{\boldsymbol{e}} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$, une base de modes $\boldsymbol{\varphi}_i$ de trace sur le bord Γ_D et de divergence nulles est obtenue : la formulation variationnelle (FVNSI) peut être exploitée car les modes POD sont dans l'espace V (voir en page 46).

Sous les contraintes du problème de contrôle (définition 5), si les modes POD sont dans l'espace V et si $\bar{\boldsymbol{u}}$ est défini par l'équation (5.8), le modèle POD-Galerkine incompressible réduit de la page 50 devient

$$\begin{aligned} \sigma_i\,\dot{a}_i(t) + \sum_{j=1}^{I} H_{i,j}\dot{c}_j(t) = C_i^h(t) \;\; &+ \;\; \sum_{j=1}^{M} C_i^j\,a_j(t) + \sum_{j,k=1}^{M} C_i^{j,k}\,a_j(t)\,a_k(t) \\ + \sum_{j,k=1}^{M,I} F_i^{j,k}\,a_j(t)\,c_k(t) \;\; &+ \;\; \sum_{j=1}^{I} G_i^j\,c_j(t) + \sum_{j,k=1}^{I} G_i^{j,k}\,c_j(t)\,c_k(t) \end{aligned} \tag{5.9}$$

avec

$$H_{i,j} = (\ \boldsymbol{\psi_j}, \boldsymbol{\varphi_i}\ )_{\mathrm{L}^2(\omega)^d}\,, \tag{5.10}$$

$$C_i^h(t) = (\ \boldsymbol{h}(t), \boldsymbol{\varphi_i}\ )_{\mathrm{L}^2(\Omega)^d} - (\ \boldsymbol{\beta}(t), \boldsymbol{\varphi_i}_{|\Gamma_F}\ )_{\mathrm{L}^2(\Gamma_F)^d}\,, \tag{5.11}$$

$$F_i^{j,k} = -\mathcal{C}(\boldsymbol{\psi_k}, \boldsymbol{\varphi_j}, \boldsymbol{\varphi_i}) - \mathcal{C}(\boldsymbol{\varphi_j}, \boldsymbol{\psi_k}, \boldsymbol{\varphi_i}), \tag{5.12}$$

$$G_i^j = -\frac{1}{\mathrm{Re}}(\mathcal{A} + \omega\,\mathcal{B})(\boldsymbol{\psi_j}, \boldsymbol{\varphi_i}), \tag{5.13}$$

et

$$G_i^{j,k} = -\mathcal{C}(\boldsymbol{\psi_j}, \boldsymbol{\psi_k}, \boldsymbol{\varphi_i}). \tag{5.14}$$

Dans la suite, $s(t)$ et $\tilde{s}(t)$ correspondent aux termes qui sont fonction de $\boldsymbol{h}$ et de $\boldsymbol{\beta}$. $s(t)$ est défini par

$$s(t) = \left(C_1^h(t) \cdots C_M^h(t)\right)^T.$$

Le système d'EDOs (5.9) n'est pas pratique car il fait apparaître les dérivées $\dot{c}_j$ des commandes de contrôle. Il y a trois manières de le transformer pour obtenir un système d'EDOs de dimension $Q \geq M$ de la forme $\dot{b}(t) = f(b(t), v(t)) + \tilde{s}(t)$ où f est une fonction polynômiale et $\tilde{s}(t)$ un terme indépendant de b et de v :

(M1) augmenter la taille du système en considérant $b(t)^T = (a(t)^T\ c(t)^T)$ et la nouvelle variable de contrôle $v(t) = \dot{c}(t)$ (on a alors $Q = M+I$) et poser $\tilde{s}(t) = \left(s(t)^T\,0 \cdots 0\right)^T$;

(M2) définir $b(t)$ par

$$b(t) = a(t) + \Sigma^{-1}\,H\,c(t) \quad \text{avec} \quad \Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_M) \quad \text{et} \quad H = (H_{i,j})$$

et remplacer $a(t)$ par $b(t) - \Sigma^{-1}\,H\,c(t)$ dans le second membre, ce qui permet de conserver la même commande de contrôle : $v(t) = c(t)$ et $\tilde{s}(t) = s(t)$ conviennent ;

(M3) orthogonaliser les fonctions $\boldsymbol{\psi_j}$ et $\boldsymbol{\psi_S}$ par rapport aux modes POD par l'algorithme de Gramm-Schmidt (entre le calcul des modes POD de $\boldsymbol{u^e} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$ et le calcul du système d'EDOs (5.9)) ; alors, puisque les traces sur Γ_D et les divergences des modes POD sont nulles, celles de $\boldsymbol{\psi_S}$ et des $\boldsymbol{\psi_j}$ sont inchangées et le champ $\bar{\boldsymbol{u}}$ défini par l'équation (5.8) est satisfaisant : cela implique $H = 0$ et fait disparaître les $\dot{c}_j$ du système, ainsi $b(t) = a(t)$, $v(t) = c(t)$ et $\tilde{s}(t) = s(t)$ conviennent.

La première manière de procéder a été proposée, dans ce contexte, par Ravindran [75, 74], la seconde a été définie et utilisée par Fahl [19] et la troisième est originale. Dans le cas de la méthode (M3), l'étape d'orthogonalisation peut être réalisée comme suit :

Pour i allant de 1 à I

$$\boldsymbol{\psi_i} := \boldsymbol{\psi_i} - \sum_{j=1}^{M} \left(\boldsymbol{\psi_i}, \boldsymbol{\varphi}_j \right)_{\mathsf{L}^2(\Omega)^d} \boldsymbol{\varphi}_j$$

Fin pour

$$\boldsymbol{\psi_S} := \boldsymbol{\psi_S} - \sum_{j=1}^{M} \left(\boldsymbol{\psi_S}, \boldsymbol{\varphi}_j \right)_{\mathsf{L}^2(\Omega)^d} \boldsymbol{\varphi}_j$$

Cette orthogonalisation n'est donc pas beaucoup plus coûteuse que le calcul de la matrice H qui est réalisé si les méthodes (M1) ou (M2) sont employées.

En conséquence, la contrainte du problème initial de contrôle, à savoir les équations de Navier-Stokes et leurs conditions aux limites, est maintenant modélisée par le système d'EDOs

$$\dot{b}(t) = f(b(t), v(t)) \; + \; \tilde{s}(t). \tag{5.15}$$

Ce système régit un champ de vitesse $\boldsymbol{u}$ approché défini par

$$\boldsymbol{u} = \boldsymbol{\psi_S} + \sum_{i=1}^{I} c_i(t)\, \boldsymbol{\psi_i} + \sum_{i=1}^{M} \sigma_i\, a_i(t)\, \boldsymbol{\varphi}_i \tag{5.16}$$

où les a_i et les c_i sont des combinaisons linéaires des composantes b_i et v_i de b et de v :

- $a_i(t) = b_i(t)$ et $c_i(t) = b_{i+M}(t)$ pour la méthode (M1),
- $a_i(t) = b_i(t) - \sigma_i^{-1} \sum_{j=1}^{I} H_{i,j}\, c_j(t)$ et $c_i = v_i$ pour la méthode (M2),
- et $a_i(t) = b_i(t)$ et $c_i(t) = v_i(t)$ pour la méthode (M3).

Le champ des vitesses $\boldsymbol{u}(t)$ est donc défini de manière unique par $b(t)$ et $v(t)$:

$$\boldsymbol{u} = \boldsymbol{\psi_S} + \sum_{i=1}^{Q} b_i(t)\, \boldsymbol{\phi_i} + \sum_{i=1}^{I} v_i(t)\, \boldsymbol{\phi_{i+Q}} \tag{5.17}$$

avec $\boldsymbol{\phi_i} = \boldsymbol{\varphi_i}$ pour $i \in [\![1, M]\!]$ et

- pour (M1) : $\boldsymbol{\phi_{i+M}} = \boldsymbol{\psi_i}$ pour $i \in [\![1, I]\!]$ et $\boldsymbol{\phi_{i+Q}} = \mathbf{0}$ pour $i \in [\![1, I]\!]$ $(Q = M + I)$;
- pour (M2) : $\boldsymbol{\phi_{i+M}} = \boldsymbol{\phi_{i+Q}} = \boldsymbol{\psi_i} - \sum_{j=1}^{I} H_{i,j}\, \boldsymbol{\varphi_j}$ pour $i \in [\![1, I]\!]$;
- pour (M3) : $\boldsymbol{\phi_{i+M}} = \boldsymbol{\phi_{i+Q}} = \boldsymbol{\psi_i}$ pour $i \in [\![1, I]\!]$.

Il faut maintenant définir la condition initiale du système d'EDOs en considérant que $\boldsymbol{u}(0) = \boldsymbol{u}_0$. Soit c^{u_0} l'état des actionneurs qui correspond à l'état initial $\boldsymbol{u}_0$ de l'écoulement, la décomposition (5.16) nous donne :

$$a_i(0) = \sigma_i^{-1}\ (\ \boldsymbol{u_0} - \boldsymbol{\psi_S} - \sum_{j=1}^{I} c_j^{u_0}\, \boldsymbol{\psi_j}, \boldsymbol{\varphi_i}\ )_{\mathrm{L}^2(\Omega)^d} = \sigma_i^{-1}\ (\ \boldsymbol{u_0} - \boldsymbol{\psi_S}, \boldsymbol{\varphi}_i\ )_{\mathrm{L}^2(\Omega)^d} - \sum_{j=1}^{I} c_j^{u_0}\, H_{i,j} \tag{5.18}$$

(pour la méthode (M3), les $\boldsymbol{\psi_j}$ sont, dans cette dernière équation, les fonctions orthogonalisées qui ont été utilisées pour la construction du modèle). Pour les méthodes (M1), (M2) et (M3), on choisira donc, respectivement, les conditions initiales

$$b(0) = \begin{pmatrix} a(0) \\ c^{u_0} \end{pmatrix},\ b(0) = a(0) + \Sigma^{-1}\, H\, c^{u_0},\ \text{et}\ b_i(0) = a_i(0) = \sigma_i^{-1}\ (\ \boldsymbol{u_0} - \boldsymbol{\psi_S}, \boldsymbol{\varphi_i}\ )_{\mathrm{L}^2(\Omega)^d} \tag{5.19}$$

où $a(0)$ est définie par l'équation (5.18). Ces trois conditions initiales, particulières à chaque méthode, sont synthétisées comme suit :

$$b(0) = b^0. \tag{5.20}$$

Ainsi, le problème de contrôle (définition 5) est transformé en le problème suivant :

Définition 6 (Problème de contrôle réduit) *Pour α, ω, $\boldsymbol{\beta}$, $\boldsymbol{h}$, $\boldsymbol{g_S}$ et $(\boldsymbol{g_i})_{1 \le i \le I}$ donnés, trouver la commande $v(t)$ qui minimise $\mathcal{J}^\alpha(\boldsymbol{u}(t), v(t))$ où $\boldsymbol{u}$ est défini par (5.17) et régi par le système d'EDOs (5.15) et la condition initiale (5.20).*

Pour la méthode (M1), cela revient en fait à remplacer le terme $\int_0^T \|c(t)\|_2^{\,2}\, dt$ de la fonction coût (équation (5.2)) par $\int_0^T \|\dot{c}(t)\|_2^{\,2}\, dt$, ce qui n'est pas équivalent mais beaucoup plus pratique pour la résolution du problème.

En pratique, la fonction coût $\mathcal{J}^\alpha$ est reformulée en fonction de $b(t)$ et $v(t)$ en injectant l'expression (5.17) dans le premier terme $\int_0^T \|\mathcal{G}(\boldsymbol{u}, t)\|_{\mathrm{L}^2}^{\,2}\, dt$ de l'équation (5.2). Le problème de contrôle réduit (définition 6) est alors écrit sous la forme d'un problème classique d'optimisation dynamique :

$$\min_v \quad \mathcal{M}(v) = \int_0^T \mathcal{N}(b(t), v(t), t)\, dt \ \text{sous la contrainte} \ \begin{cases} \dot{b}(t) &= F(b(t), v(t), t) \\ b(0) &= b^0 \end{cases} \tag{5.21}$$

avec $F(b(t), v(t), t) = f(b(t), v(t)) + \tilde{s}(t)$. Notons que, si $\mathcal{G}$ est affine, $\mathcal{N}(b(t), v(t), t)$ peut se mettre sous une forme quadratique facilement manipulable et qu'il est relativement simple de calculer la fonction $\mathcal{M}$ et son gradient (voir plus bas).

Dans la suite, nous utiliserons une variante de la méthode (M3) : cela nécessite l'orthogonalisation des fonctions $\boldsymbol{\psi_S}$ et $\boldsymbol{\psi_i}$, mais simplifiera la construction du modèle et permettra d'obtenir une expression simple de la fonction coût pour un problème de *flow tracking* (voir la section 5.1.3).

Remarque. Dans le cadre de la modélisation POD-Galerkine des écoulements, la décomposition (5.16) du champ des vitesses, qui permet de tenir explicitement compte des conditions aux limites de Dirichlet donc de l'actionneur, est parfois appelée *méthode des fonctions de commandes* (*control function method*) en référence au papier de Graham *et al.* [27]. De plus, les fonctions $\boldsymbol{\psi_i}$ sont généralement appelées *fonctions de commande*. Ce papier propose également une autre méthode qui consiste à pénaliser une condition de Dirichlet sur $\boldsymbol{u}$ par la dérivée normale $[\nabla u]\,\boldsymbol{n}$, ce qui permet ainsi de la prendre en compte de manière implicite et approchée dans le modèle POD-Galerkine par un terme de bord approprié.

Résolution numérique du problème de contrôle réduit

Comme nous venons de le voir, la modélisation POD-Galerkine aboutit à un problème de contrôle réduit qui se met sous la forme (5.21) d'un problème classique d'optimisation dynamique. Il est donc possible de minimiser $\mathcal{M}$ par des méthodes numériques en calculant son gradient à l'aide d'une variable adjointe.

En effet, pour des fonctions $\mathcal{N}$ et F suffisamment régulières, le gradient de $\mathcal{M}$ en v est :

$$\nabla\mathcal{M}(v) = \partial_v\mathcal{N}(b, v, t) - [\partial_v F(b, v, t)]^T q,$$

où l'adjoint q est défini par

$$\begin{cases} -\dot{q} = [\partial_b F(b, v, t)]^T q - \partial_b\mathcal{N}(b, v, t) \\ q(T) = 0 \end{cases} \tag{5.22}$$

(on pourra consulter [20] pour plus de détails). Notons que cette expression de l'adjoint peut être obtenue via le lagrangien

$$\mathcal{L}(b, v, q) = \int_0^T \mathcal{N}(b, v, t)\,dt + \int_0^T q(t)^T \left(\dot{b}(t) - F(b, v, t)\right)\,dt,$$

en cherchant à annuler sa dérivée par rapport à la variable b :

$$\partial_b\mathcal{L}(b, v, q)(\delta b) = \int_0^t \left[\partial_b\mathcal{N}(b, v, t) - \dot{q} - \partial_b F(b, v, t)^T q\right]^T \delta b\,dt + q(T)^T\,\delta b(T)$$

(en considérant $\delta b(0) = 0$).

Ravindran [75, 74] a réalisé l'optimisation du problème de contrôle réduit en résolvant le système obtenu par la discrétisation des conditions d'optimalité par une méthode de Newton (les conditions d'optimalité expriment la stationnarité du lagrangien, c'est-à-dire que son gradient est nul). Cette technique permet de converger vers la solution optimale et le multiplicateur de Lagrange associé à la contrainte, c'est-à-dire l'adjoint q pour un état b et une commande v qui correspondent à un minimum local de $\mathcal{M}$.

En revanche, Bergmann *et al.* [7] ont déterminé la commande optimale de leur problème de contrôle réduit par une méthode de plus grande pente en exploitant directement cette expression pour évaluer le gradient.

Enfin, précisons que Fahl [19] cherche à minimiser la fonction réduite $\mathcal{M}$ dans une boule dont le centre est une valeur particulière de v à l'aide d'un algorithme spécifique, proposé par Toint [92] (voir plus bas). Cet algorithme ne nécessite que des évaluations de la fonction à minimiser et de son gradient, cependant Fahl a simplement évalué le gradient de la fonction coût par différences finies lors de ses essais numériques (ce qui est beaucoup plus coûteux en calculs que de passer par l'adjoint, mais nécessite moins de travail de programmation).

Dans la suite, la résolution des problèmes de contrôle réduits est effectuée par une méthode de gradient conjugué et une évaluation du gradient via le calcul de l'adjoint q.

5.1.2 Algorithmes itératifs de contrôle

Deux approches peuvent être *a priori* distinguées pour le calcul d'une loi de contrôle à partir d'une méthode de réduction de modèle :

- une approche "globale" qui consiste à construire un modèle réduit mais néanmoins fidèle au système physique (l'écoulement) pour un ensemble relativement important de commandes de contrôle, le principe étant de calculer une commande efficace à partir de ce modèle seulement : elle a été étudiée par Graham *et al.* [27, 28] ;
- une approche itérative qui alterne simulation, modélisation réduite et optimisation du modèle afin de converger vers une commande de contrôle satisfaisante : elle a été appliquée avec succès par Ravindran [75, 74] et a conduit Fahl *et al.* [19] à développer un algorithme sophistiqué qui étend le principe des méthodes de minimisation à région de confiance au cadre de la modélisation POD-Galerkine en s'appuyant sur l'algorithme étudié par Toint [92].

La première approche a été étudiée par Graham *et al.* pour le contrôle du sillage d'un écoulement bidimensionnel de nombre de Reynolds $\mathrm{Re} = 100$ derrière un cylindre et par une commande de rotation du cylindre.

Graham *et al.* ont voulu construire un modèle réduit le plus satisfaisant possible en complétant la base de modes POD ou en la déduisant de données obtenues par simulation de l'écoulement pour une commande de l'actionneur (le cylindre) judicieusement choisie.

Ils proposent notamment de définir une commande dont le spectre fréquentiel est formé d'une bande relativement large qui contient les fréquences les plus intéressantes *a priori* pour la commande de contrôle et d'effectuer une simulation de l'écoulement sur une durée suffisamment longue : le nombre de modes POD représentatifs de la simulation est alors beaucoup plus important que celui de l'écoulement non contrôlé, ainsi le modèle est fidèle à la dynamique de l'écoulement pour une relativement grande variété de commandes. L'optimisation du modèle conduit finalement à une commande de contrôle efficace. Cette méthode a été réemployée avec succès par Bergmann *et al.* [7] pour un écoulement de nombre de Reynolds $\mathrm{Re} = 200$.

Il faut noter que Graham *et al.* pallient le problème de la définition d'une fonction coût qui permette de contrôler le sillage tout en n'étant pas définie par la pression (en effet, le modèle POD-Galerkine ne définit pas de manière simple le champ de pression et il n'est pas trivial de définir la portance ou la traînée à partir des variables $b(t)$ et $v(t)$ du modèle). Pour cela, ils définissent le problème de contrôle directement à partir de la décomposition POD du champ $\boldsymbol{u}$ des vitesses. En conséquence, il n'est pas cohérent d'appliquer une stratégie itérative à partir de cette fonction coût, puisqu'elle dépend intrinsèquement des modes POD donc du modèle POD-Galerkine.

Cette première stratégie présente le désavantage de devoir être adaptée spécifiquement au système physique considéré, le problème étant de savoir *a priori* comment construire un modèle réduit fiable pour une grande variété de commandes de contrôle. De plus, elle ne semble pas pouvoir être appliquée à des systèmes complexes, car il paraît peu réaliste de pouvoir construire un modèle significativement réduit et néanmoins suffisamment fidèle pour de tels systèmes.

Notons enfin que le travail de Graham *et al.* montre qu'il est difficile de définir une stratégie efficace pour enrichir la base des modes POD afin d'améliorer le modèle qui en résulte.

Notations $\mathcal{J}$ et $\mathcal{M}^c$. Afin de simplifier la présentation de l'approche itérative, on considérera que $v \equiv c$, ce qui est trivialement vraie pour les méthodes (M2) et (M3).

La fonction coût du problème initial (définition 5), que l'on souhaite minimiser et qui dépend de la commande c des actionneurs, est notée $\mathcal{J}$; elle est définie par

$$\mathcal{J}(c) = \mathcal{J}^\alpha(\boldsymbol{u}, c)$$

sous la contrainte que $\boldsymbol{u}$ soit régi par les équations de Navier-Stokes incompressibles (2.19)-(2.20) et les conditions aux limites (2.24)-(2.25) (avec $P = 0$) où $\boldsymbol{g}$ est définie en fonction de $\boldsymbol{g_S}$, $(\boldsymbol{g_i})_{1\leq i\leq I}$ et $c(t)$ par la relation (5.1).

Comme il a été décrit précédemment, la modélisation POD-Galerkine permet de définir une fonction réduite $\mathcal{M}$ qui "modélise" $\mathcal{J}$. Cette fonction a été définie à l'aide de données numériques qui proviennent de la simulation de l'écoulement pour une commande $c^e(t)$ particulière des actionneurs. Afin de préciser à partir de quelle commande la fonction modèle est définie, un exposant lui sera ajouté :

$$\mathcal{M}^{c^e}(c) = \mathcal{J}^\alpha(\boldsymbol{u}, c) \text{ sous les contraintes (5.17)-(5.15)-(5.20)}$$

où les modes POD ont été définis à l'aide de l'écoulement $\boldsymbol{u}^e$ contrôlé par la commande c^e.

L'approche itérative utilisée par Ravindran, qui sera reprise dans la section 5.2.3, repose sur l'algorithme suivant :

Algorithme itératif de contrôle primitif

Étape 0 (calcul des fonctions de commandes) Résolution des $I+1$ problèmes de Stokes associés aux conditions de Dirichlet : les fonctions $\boldsymbol{\psi_S}$ et de commande $\boldsymbol{\psi_i}$ sont ainsi obtenues.

Étape 1 (simulation) Résolution des équations de Navier-Stokes incompressibles pour la commande $c^k(t)$ des actionneurs : un champ de vitesse instationnaire $\boldsymbol{u}^e$ est calculé.

Étape 2 (POD) Calcul POD de $\boldsymbol{u}^e - \bar{\boldsymbol{u}}^e$ où $\bar{\boldsymbol{u}}^e$ est défini par (5.8) et $c^e(t) = c^k(t)$: une base $(\boldsymbol{\varphi}_i)$ de modes POD de trace sur Γ_D et de divergence nulle est obtenue ;

Étape 3 (construction du problème de contrôle réduit) Calcul du modèle POD-Galerkine (5.9) et construction de la fonction "modèle" $\mathcal{M}^{c^k}$.

Étape 4 (Calcul de la commande "optimale") Résolution numérique du problème de contrôle réduit (5.21) (minimisation de $\mathcal{M}^{c^k}(c)$) : une nouvelle commande $c^{k+1}(t)$ est obtenue.

Étape 6 (Test d'arrêt) Si la commande c^{k+1} est satisfaisante (par exemple si $\int_0^T \left\|c^{k+1}(t) - c^k(t)\right\|_2^{\,2} dt \leq \varepsilon$), alors c^{k+1} est la commande finale, sinon on incrémente k ($k := k+1$) et on retourne à l'étape 1.

Cet algorithme primitif donne des résultats relativement satisfaisants pour des configurations simples : voir [75, 74] et la section 5.2.3. Il ne garantit néanmoins ni la convergence de l'algorithme, ni même que la commande c^{k+1} soit meilleure que la commande c^k précédente. C'est la raison pour laquelle il est particulièrement intéressant de combiner cet algorithme primitif avec les méthodes de minimisation dites *à région de confiance* afin d'obtenir un algorithme robuste et efficace en terme de coût de calculs : consulter [2] et [19].

Ces méthodes sont basées sur l'optimisation itérative de fonctions "modèles" supposées représentatives de la fonction à minimiser dans une région locale qui engloble le dernier itéré, la *région de confiance*. Dans le cas de l'algorithme précédent, l'itéré est c^k et la fonction modèle est $\mathcal{M}^{c^k}$: cette fonction est censée être représentative du coût $\mathcal{J}$ à minimiser pour un ensemble de commandes c "proches" de c^k (en pratique une boule de centre c^k) : en effet, le modèle POD-Galerkine a été construit à partir de données obtenues pour $c = c^k$.

Afin d'éclaircir notre propos, nous donnons ici l'algorithme qui a été proposé par Fahl [19, p.80] :

Algorithme itératif de contrôle par région de confiance

Étape 0 (Initialisation) Choisir $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 \leq \gamma_2 < 1 \leq \gamma_3$, un rayon de confiance initial r^0 et une commande initiale $c^0(t)$. Calculer le champ de référence $\boldsymbol{u^0}$ pour la commande c^0 et $\mathcal{J}_0 = \mathcal{J}(c^0)$. Poser $k := 0$.

Étape 1 Construire le modèle POD-Galerkine et la fonction modèle $\mathcal{M}^{c^k}$ à partir des données $\boldsymbol{u^e} = \boldsymbol{u^k}$.

Étape 2 Calculer la commande c^{k+1} qui minimise $\mathcal{M}^{c^k}(c)$ dans la boule de confiance définie par $\|c - c^k\| \leq r^k$.

Étape 3 Calculer le champ de vitesse $\boldsymbol{u^{k+1}}$ pour la commande c^{k+1}, $\mathcal{J}_{k+1} = \mathcal{J}(c^{k+1})$ et le rapport

$$\rho_k = \frac{\mathcal{M}^{c^k}(c^{k+1}) - \mathcal{M}^{c^k}(c^k)}{\mathcal{J}_{k+1} - \mathcal{J}_k}.$$

Étape 4 Mettre à jour le rayon de confiance :

- si $\rho_k \geq \eta_2$ (la minimisation du modèle a été très efficace) : choisir $r^{k+1} \in [r^k, \gamma_3\, r^k]$ (le rayon de confiance est augmenté), poser $k := k + 1$ et retourner à l'étape 1 ;
- si $\eta_1 \leq \rho_k < \eta_2$ (la minimisation du modèle a été moyennement efficace) : choisir $r^{k+1} \in [\gamma_2\, r^k, r^k[$ (le rayon de confiance est légèrement diminué), poser $k := k + 1$ et retourner à l'étape 1 ;
- si $\rho_k < \eta_1$ (la minimisation du modèle n'est pas satisfaisante) : choisir $r^{k+1} \in [\gamma_1\, r^k, \gamma_2\, r^k]$ (le rayon de confance est diminué), poser $k := k + 1$ et retourner à l'étape 2.

L'étape 2, c'est-à-dire la minimisation du problème de contrôle réduit dans une boule de confiance de rayon r^k, peut être réalisée à l'aide de l'algorithme proposé par Toint [92]. Il faut noter que ici, contrairement au cadre classique des méthodes à région de confiance, la fonction modèle ne coïncide en général pas exactement avec la fonction coût à minimiser pour la valeur de l'itéré autour duquel la région de confiance est définie : $\mathcal{M}^{c^k}(c^k) \neq \mathcal{J}(c^k)$ en général. Pour plus de détails, le lecteur est invité à se reporter à la thèse de Fahl [19].

5.1.3 Modélisation particulière du problème de *flow tracking*

La modélisation POD-Galerkine va être explicitée sur un problème de *flow tracking* (voir l'équation (5.4)) et pour une variante de la méthode (M3) : c'est le problème et la méthode qui correspondent aux expérimentations numériques qui sont proposées dans la suite.

Le modèle POD-Galerkine est construit comme suit :

étape 0 calcul des fonctions de commande $\boldsymbol{\psi_i}$, pour $1 \leq i \leq I$, via la résolution numérique des I problèmes de Stokes (5.5) associés aux conditions de Dirichlet (5.7) ;

étape 1 calcul de la fonction $\boldsymbol{\psi_S}$ de prise en compte des conditions de Dirichlet station-

naires définie par

$$\boldsymbol{\psi_S} = \frac{1}{T}\int_0^T \left[\boldsymbol{u^e} - \sum_{i=1}^{I} c_i^e(t)\,\boldsymbol{\psi_i}\right] dt$$

(en pratique la moyenne est effectuée par une moyenne arithmétique sur les clichés) puis calcul des M premiers modes POD du champ $\boldsymbol{u^e} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$ de divergence et de trace sur Γ_D nulles qui est obtenu en définissant $\bar{\boldsymbol{u}}^{\boldsymbol{e}}$ par l'équation (5.8) ;

étape 2 orthonormalisation des fonctions $\boldsymbol{\psi_i}$ et $\boldsymbol{\psi_S}$ par l'algorithme de Gramm-Schmidt :

Pour i allant de 1 à I

$$\left|\begin{array}{l} \boldsymbol{\psi_i} := \boldsymbol{\psi_i} - \displaystyle\sum_{j=1}^{M} (\,\boldsymbol{\psi_i}, \boldsymbol{\varphi}_j\,)_{\mathsf{L}^2(\Omega)^d}\ \boldsymbol{\varphi}_j - \sum_{j=1}^{i-1} (\,\boldsymbol{\psi_i}, \boldsymbol{\psi_j}\,)_{\mathsf{L}^2(\Omega)^d}\ \boldsymbol{\psi_j} \\ \boldsymbol{\psi_i} := (\,\boldsymbol{\psi_i}, \boldsymbol{\psi_i}\,)^{-1/2}_{\mathsf{L}^2(\Omega)^d}\ \boldsymbol{\psi_i} \end{array}\right.$$

Fin pour

$$\boldsymbol{\psi_S} := \boldsymbol{\psi_i} - \sum_{j=1}^{M} (\,\boldsymbol{\psi_S}, \boldsymbol{\varphi}_j\,)_{\mathsf{L}^2(\Omega)^d}\ \boldsymbol{\varphi}_j - \sum_{j=1}^{I} (\,\boldsymbol{\psi_S}, \boldsymbol{\psi_j}\,)_{\mathsf{L}^2(\Omega)^d}\ \boldsymbol{\psi_j}$$

$$\boldsymbol{\psi_S} := (\,\boldsymbol{\psi_S}, \boldsymbol{\psi_S}\,)^{-1/2}_{\mathsf{L}^2(\Omega)^d}\ \boldsymbol{\psi_S}$$

étape 4 construction du modèle (5.9) : $H = 0$ et les $\dot{c}_j$ disparaissent naturellement.

À propos de cet algorithme de construction de modèle réduit, notons les points suivants.

- L'étape 3 ne correspond pas exactement à la méthode (M3) puisque les fonctions $\boldsymbol{\psi_S}$ et $\boldsymbol{\psi_i}$ sont non seulement orthogonalisées par rapport aux modes $\boldsymbol{\varphi}_j$ mais également orthogonalisées entre elles et normalisées : ceci implique un changement de la valeur des traces de $\boldsymbol{\psi_S}$ et des $\boldsymbol{\psi_i}$ dont il faut tenir compte (voir ci-dessous).
- L'étape 0 est réalisée une fois pour toutes au tout début d'un algorithme itératif de contrôle, la construction de tous les modèles qui se succèdent reposant sur les mêmes fonctions de commande $\boldsymbol{\psi_i}$. Cependant, il faut alors sauvegarder la valeur initiale des $\boldsymbol{\psi_i}$, puisque l'étape 3 modifie leur trace (en outre, ce n'est pas indispensable mais recommandé si la méthode (M3) décrite précédemment est effectivement employée).
- $\boldsymbol{\psi_S}$ est recalculée à partir de $\boldsymbol{u^e}$ pour chaque modèle POD-Galerkine : elle est donc plus représentative des données et capte en pratique une grande partie de l'énergie cinétique de $\boldsymbol{u^e}$.
- L'étape 3 a été employée plutôt que la méthode (M3) afin de simplifier l'expression de la fonction coût et de son gradient.

La modification de la trace des fonctions $\boldsymbol{\psi_S}$ et $\boldsymbol{\psi_i}$ est facilement prise en compte par une transformation linéaire de la commande après la résolution du problème de contrôle réduit. En effet, notons $\boldsymbol{\psi_S}$ et $\boldsymbol{\psi_i}$ les fonctions avant leur orthonormalisation par l'étape 3

et $\tilde{\boldsymbol{\psi}}_{\boldsymbol{S}}$ et $\tilde{\boldsymbol{\psi}}_{\boldsymbol{i}}$ celles obtenues après le processus de Gramm-Schmidt. La relation

$$\boldsymbol{\psi}_{\boldsymbol{S}} + \sum_{j=1}^{I} c_j(t)\,\boldsymbol{\psi}_{\boldsymbol{j}} + \sum_{j=1}^{M} \sigma_j\, a_j(t)\,\boldsymbol{\varphi}_j = \tilde{\boldsymbol{\psi}}_{\boldsymbol{S}} + \sum_{j=1}^{I} \tilde{c}_j(t)\,\tilde{\boldsymbol{\psi}}_{\boldsymbol{j}} + \sum_{j=1}^{M} \sigma_j\, \tilde{a}_j(t)\,\boldsymbol{\varphi}_j \tag{5.23}$$

correspond, en tenant compte des propriétés d'orthogonalité et d'orthonormalité, à la transformation linéaire suivante :

$$\tilde{c}(t) = P^c\, c(t) + l^c \quad \text{avec} \quad P^c_{i,j} = (\,\tilde{\boldsymbol{\psi}}_{\boldsymbol{i}}, \boldsymbol{\psi}_{\boldsymbol{j}}\,)_{\mathrm{L}^2(\Omega)^d} \text{ et } l^c_i = (\,\tilde{\boldsymbol{\psi}}_{\boldsymbol{i}}, \boldsymbol{\psi}_{\boldsymbol{S}}\,)_{\mathrm{L}^2(\Omega)^d}.$$

Ainsi, si les fonctions $\boldsymbol{\psi}_{\boldsymbol{i}}$ sont linéairement indépendantes, la matrice P^c est inversible et la variable de contrôle $\tilde{c}(t)$ du modèle POD-Galerkine construit avec les fonctions $\tilde{\boldsymbol{\psi}}_{\boldsymbol{i}}$ et $\tilde{\boldsymbol{\psi}}_{\boldsymbol{S}}$, définit une unique commande $c(t) = (P^c)^{-1}\,(\tilde{c}(t) - l^c)$ des actionneurs au sens des conditions aux limites de Dirichlet (5.1).

Après la résolution numérique du problème de contrôle de variable $\tilde{c}$, on est donc amené à effectuer la transformation $c(t) = (P^c)^{-1}(\tilde{c}(t) - l^c)$ pour obtenir la véritable commande des actionneurs. Cette transformation est faite avant d'effectuer une nouvelle simulation numérique dans le cas d'un algorithme itératif de contrôle. Avec la méthode (M3) originale proposée dans la section 5.1, le mécanisme de passage de c à $\tilde{c}$ n'est pas nécessaire, cependant le fait d'avoir des fonctions orthonormales va simplifier le calcul de la fonction coût de *flow tracking* et de son gradient.

Avec la décomposition orthonormale

$$\boldsymbol{u} = \tilde{\boldsymbol{\psi}}_{\boldsymbol{S}} + \sum_{j=1}^{I} \tilde{c}_j(t)\,\tilde{\boldsymbol{\psi}}_{\boldsymbol{j}} + \sum_{j=1}^{M} \sigma_j\, \tilde{a}_j(t)\,\boldsymbol{\varphi}_j,$$

on obtient l'expression suivante de la fonction coût pour $\alpha = 0$

$$\begin{aligned} \mathcal{J}^0(\boldsymbol{u}, c(t)) &= \int_0^T \|\mathcal{G}(\boldsymbol{u}, t)\|_{\mathrm{L}^2}{}^2\, dt = \int_0^T \left\|\boldsymbol{u}(t) - \boldsymbol{u}^{\boldsymbol{R}}(t)\right\|_{\mathrm{L}^2}{}^2 dt \\ &= \int_0^T \left[\left\|b(t) - b^R(t)\right\|_2{}^2 + \left\|v(t) - v^R(t)\right\|_2{}^2\right] dt + C \end{aligned}$$

avec $b(t) = \tilde{a}(t)$, $v(t) = \tilde{c}(t)$, $b^R_i(t) = (\,\boldsymbol{u}^{\boldsymbol{R}}(t), \boldsymbol{\varphi}_i\,)_{\mathrm{L}^2(\Omega)^d}$, $v^R_i(t) = (\,\boldsymbol{u}^{\boldsymbol{R}}(t), \tilde{\boldsymbol{\psi}}_{\boldsymbol{i}}\,)_{\mathrm{L}^2(\Omega)^d}$ et C une constante. En omettant la constante C, la fonction coût modèle est donc

$$\mathcal{M}(v) = \int_0^T \left[\left\|b(t) - b^R(t)\right\|_2{}^2 + \left\|v(t) - v^R(t)\right\|_2{}^2\right] dt$$

sous la contrainte des équations (5.15) et (5.20) obtenues pour les fonctions de commande $\tilde{\boldsymbol{\psi}}_{\boldsymbol{i}}$ et $\tilde{\boldsymbol{\psi}}_{\boldsymbol{S}}$. Son gradient peut être efficacement calculé grâce à l'adjoint q défini par

$$\begin{cases} -\dot{q}(t) = 2\,(b(t) - b^R(t)) + [\partial_b f(b(t), v(t))]^T\, q(t) \\ q(T) = 0, \end{cases}$$

et la relation

$$(\nabla \mathcal{M}(v))(t) = \left[\partial_v f(b(t), v(t))\right]^T q(t) + 2(v(t) - v^R(t)).$$

Des essais numériques de *flow tracking* sont menés dans la suite en minimisant itérativement cette fonction coût modèle par des algorithmes de gradient conjugué et des évaluations du gradient via cet adjoint (le calcul de l'adjoint est effectué par un schéma d'Adams-Bashforth d'ordre quatre).

5.2 Illustration du contrôle d'un écoulement laminaire bidimensionnel décollé

Cette section propose un exemple de calcul de contrôle actif basé sur la modélisation réduite POD-Galerkine. La fonction coût est de type *flow tracking* (voir l'équation (5.4)) et l'écoulement de référence $\boldsymbol{u}^{\boldsymbol{R}}$ est calculé à partir d'une commande de contrôle connue : cela va permettre de valider l'algorithme itératif employé et d'évaluer son efficacité en termes de convergence vers une commande optimale connue.

5.2.1 Description de la configuration

La configuration considérée est celle d'un écoulement laminaire et bidimensionnel autour d'un obstacle circulaire présentant un actionneur. Le nombre de Reynolds, basé sur la vitesse en entrée qui est constante et purement longitudinale et sur le diamètre de l'obstacle, est $\mathrm{Re} = 100$.

La géométrie du domaine fluide Ω, qui correspond au domaine de calcul, est visualisée sur la figure 5.1. Les directions longitudinale et transverse sont notées x_1 et x_2, Γ_C est la

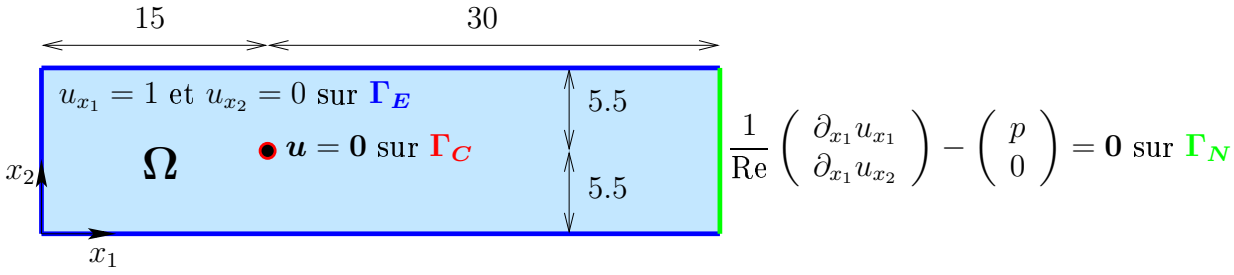


FIG. 5.1 – Géométrie du domaine fluide Ω et conditions aux limites de l'écoulement non contrôlé.

frontière circulaire de l'obstacle, $\Gamma_N = 45 \times [0, 11]$ et $\Gamma_E = ([0, 45] \times \{0, 11\}) \cup (0 \times [0, 11])$.

Cette figure décrit aussi les conditions aux limites, de Dirichlet sur $\Gamma_D = \Gamma_E \cup \Gamma_C$ et de flux sur Γ_N, pour un actionneur inactif : la vitesse transverse u_{x_2} est nulle et la vitesse longitudinale u_{x_1} est unitaire sur Γ_E, c'est-à-dire sur les bords droit, supérieur et inférieur du pavé $[0, 45] \times [0, 11]$, la condition de flux (2.25) est appliquée sur le bord gauche de sortie avec $\boldsymbol{\beta} = \mathbf{0}$ et $\omega = 0$ et enfin une condition d'adhésion $\boldsymbol{u} = 0$ est imposée au niveau de l'obstacle (si l'actionneur est inactif).

L'écoulement est contrôlé grâce à un actionneur qui est positionné en deux endroits de la partie droite de l'obstacle, près des zones de décollement de la couche limite, comme l'indique la figure 5.2 : l'actionneur agit au niveau du cercle pour la gamme d'angles $[-\pi/2, -\pi/4] \times [\pi/4, \pi/2]$. Les conditions de Dirichlet sont ainsi stationnaires sur le bord

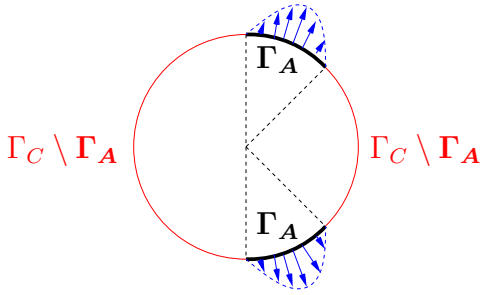


FIG. 5.2 – Actionneur de soufflage/aspiration du fluide de profil parabolique positionné sur l'obstacle.

$\Gamma_D \setminus \Gamma_A = \Gamma_E \cup (\Gamma_C \setminus \Gamma_A)$:

$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{g_S}(\boldsymbol{x}) \quad \text{sur} \quad \Gamma_D \setminus \Gamma_A$$

avec

$$\boldsymbol{g_S}(\boldsymbol{x}) = \begin{cases} (1\ 0)^T & \text{sur} \quad \Gamma_E \\ \mathbf{0} & \text{sinon} \end{cases} .$$

Elles sont instationnaires sur Γ_A :

$$\boldsymbol{u}(\boldsymbol{x}, t) = c_1(t)\, \boldsymbol{g_1}(\boldsymbol{x}, t) \quad \text{sur} \quad \Gamma_A$$

où $c_1(t)$ est la commande de l'actionneur et $\boldsymbol{g_1}(\boldsymbol{x})$ correspond à des profils paraboliques de vitesses. Plus précisément, $\boldsymbol{g_1}(\boldsymbol{x})$ est définie comme suit :

$$\boldsymbol{g_1}(\boldsymbol{x}) = pf(z)\, \boldsymbol{n}(\boldsymbol{x}) \text{ avec } \frac{\pi^2}{64} pf(z) = \begin{cases} (\frac{\pi}{2} - z)(z - \frac{\pi}{4}) & \text{pour} \quad z \in [\frac{\pi}{4}, \frac{\pi}{2}] \\ -(z + \frac{\pi}{2})(z + \frac{\pi}{4}) & \text{pour} \quad z \in [-\frac{\pi}{2}, -\frac{\pi}{4}] \end{cases}$$

où $z \in]-\pi, \pi]$ est tel que $x_1 = 15 + \cos z$ et $x_2 = 5.5 + \sin z$ ($pf(-3\pi/8) = pf(3\pi/8) = 1$).

5.2.2 Simulation de l'écoulement

La simulation de l'écoulement est effectuée à l'aide du logiciel freefem développé par F. Hecht, O. Pironneau et K. Ohtsuka [32]. Ce logiciel interprète un langage de type C qui permet de gérer des maillages et des bases d'éléments finis et d'effectuer toutes sortes de calcul (résolution de système linéaire, différentiation et intégration spatiales notamment) : la génération du maillage du domaine fluide Ω, la résolution des problèmes de Stokes (5.5) associés aux conditions de Dirichlet (5.6) et (5.7) mais aussi de nombreuses opérations nécessaires à la construction du modèle POD-Galerkine sont effectuées grâce à freefem.

Schémas numériques

Les solutions approchées des problèmes de Stokes sont obtenues par une discrétisation spatiale par éléments finis mixtes, basée sur la formulation variationnelle suivante : trouver $(\boldsymbol{u}_h, p_h) \in X_h \times M_h$ tels que $\boldsymbol{u_h} = \boldsymbol{g_h}$ sur $(\Gamma_D)_h$,

$$\forall \boldsymbol{v_h} \in X_{0h} \qquad \sum_{i=1}^{d} \int_{\Omega_h} \boldsymbol{\nabla}(u_h)_{x_i}.\boldsymbol{\nabla}(v_h)_{x_i}\, d\boldsymbol{x} - \int_{\Omega_h} p_h(\boldsymbol{\nabla}.\boldsymbol{v_h})\, d\boldsymbol{x} = 0,$$

$$\text{et} \qquad \forall q_h \in M_h \qquad \int_{\Omega_h} (\boldsymbol{\nabla}.\boldsymbol{u_h}) q_h\, d\boldsymbol{x} = 0 \tag{5.24}$$

où Ω_h est la triangulation du domaine fluide Ω, $\boldsymbol{g_h}$ définit une condition de Dirichlet sur $(\Gamma_D)_h$, où

$$X_h = \{\boldsymbol{v_h} \in \mathcal{C}^0(\Omega_h)^2 \ / \ \boldsymbol{v_h}_{|_{\mathcal{T}_k}} \in (P^2)^2 \quad \forall k\} \text{ et } M_h = \{q_h \in \mathcal{C}^0(\Omega_h) \ / \ q_{h|_{\mathcal{T}_k}} \in P^1 \quad \forall k\}$$

(éléments conformes P^1/P^2 de Hood-Taylor sur les triangles $\mathcal{T}_k$ de Ω_h) et où $X_{0h} \subset X_h$ est l'espace des éléments de X_h de trace nulle sur $(\Gamma_D)_h$. La pression p_h n'étant unique qu'à une constante près, ce système conduit à un système linéaire non inversible et c'est pourquoi le problème est modifié par pénalisation : le terme $\varepsilon \int_{\Omega_h} p_h\, q_h\, d\boldsymbol{x}$ est ajouté au premier membre de l'équation (5.24) afin d'obtenir un système inversible (nous avons fixé ε à 10^{-15}).

La simulation numérique de l'écoulement, qui est régi par les équations de Navier-Stokes incompressible pour Re $= 100$ et $\boldsymbol{h} = 0$, est réalisée en résolvant, via une formulation de type éléments finis mixtes similaire, le problème obtenu par une discrétisation temporelle d'ordre un de la dérivée particulaire par la méthode des caractéristiques :

$$\partial_t \boldsymbol{u}(\boldsymbol{x}, t) + (\boldsymbol{u}(\boldsymbol{x}, t).\boldsymbol{\nabla})\boldsymbol{u}(\boldsymbol{x}, t) \approx \frac{1}{\Delta t}\left[\boldsymbol{u}(\boldsymbol{x}, t + \Delta t) - \boldsymbol{u}(\boldsymbol{x} - \Delta t\, \boldsymbol{u}(\boldsymbol{x}, t), t)\right].$$

Ainsi, en notant $\boldsymbol{u_h^n}$ et p_h^n les vitesses et pression approchées au temps t^n, le problème obtenu, d'inconnues $\boldsymbol{u_h^{n+1}}$ et p_h^{n+1}, a un second membre non nul défini par le champ des vitesses $\boldsymbol{u_h^n}$ précédemment calculé. Plus précisément, ce second membre, qui est noté $\boldsymbol{h^n}$, est

$$\boldsymbol{h^n} = \frac{1}{\Delta t} \boldsymbol{u_h^n}(\boldsymbol{x} - \Delta t\, \boldsymbol{u_h^n}(\boldsymbol{x})).$$

Il est estimé à partir de $\boldsymbol{u_h^n}$ à l'aide d'un opérateur numérique de convection qui a été spécialement implémenté dans freefem pour la résolution des EDPs comportant des termes non-linéaires de convection. Ainsi, au temps t^{n+1}, le problème est : trouver $(\boldsymbol{u_h^{n+1}}, p_h^{n+1}) \in X_h \times M_h$ tels que $\boldsymbol{u_h} = \boldsymbol{g_h^n}$ sur $(\Gamma_D)_h$ et, pour tout $(\boldsymbol{v_h}, q_h) \in X_{0h} \times M_h$,

$$\frac{1}{\Delta t}\int_{\Omega_h} \boldsymbol{u_h^{n+1}}.\boldsymbol{v_h}\, d\boldsymbol{x} + \frac{1}{\mathrm{Re}}\sum_{i=1}^{d}\int_{\Omega_h} \boldsymbol{\nabla}(u_h^{n+1})_{x_i}.\boldsymbol{\nabla}(v_h)_{x_i}\, d\boldsymbol{x} + \int_{\Omega_h} p_h^{n+1}(\boldsymbol{\nabla}.\boldsymbol{v_h})\, d\boldsymbol{x} = \int_{\Omega_h} \boldsymbol{h^n}.\boldsymbol{v_h}\, d\boldsymbol{x},$$

$$\text{et} \qquad \int_{\Omega_h} (\boldsymbol{\nabla}.\boldsymbol{u_h^{n+1}}) q_h\, d\boldsymbol{x} + \varepsilon \int_{\Omega_h} p_h\, q_h\, d\boldsymbol{x} = 0.$$

Le schéma obtenu est stable, même pour des pas de temps Δt importants, ce qui est pratique pour obtenir rapidement le régime périodique de type Von Kármán de l'écoulement précédent.

On pourra consulter [32] et [69] pour plus de détails.

Le maillage Ω_h qui a été généré grâce à freefem comporte 13 535 triangles et 6 935 sommets ; il est visualisé sur la figure 5.3. Les parties droite, supérieure et inférieure de

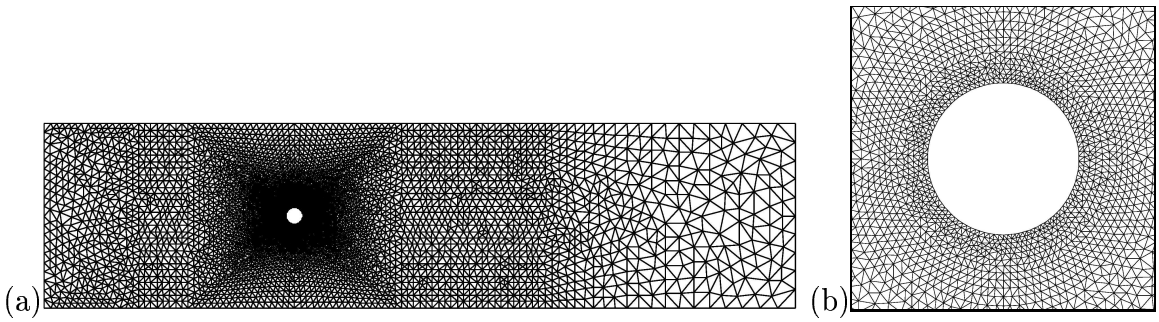


FIG. 5.3 – (a) Maillage global Ω_h ; (b) maillage de la boîte $[14, 16] \times [4.5, 6.5]$

Γ_E sont respectivement décomposées en 20, 100 et 100 arêtes ; Γ_N en 10 arêtes ; le bord discret $(\Gamma_A)_h$ correspondant à l'actionneur est décomposé en 2×15 arêtes et le reste du bord $(\Gamma_C)_h \setminus (\Gamma_A)_h$ de l'obstacle en 75 arêtes. Ainsi, le bord $(\Gamma_D)_h = \Gamma_E \cup (\Gamma_C)_h$ où sont imposées les conditions de Dirichlet comprend 325 des 335 arêtes du bord Γ_h de Ω_h.

Les fonctions spatiales manipulées sont définies grâce à des bases d'éléments finis P^1 et P^2 sur Ω_h de dimensions respectives le nombre 6 935 de sommets et 27 405 de nœuds. Ainsi, les systèmes linéaires à inverser pour résoudre un problème de Stokes ou effectuer une itération de Navier-Stokes sont de dimension $2 \times 27\,405 + 6\,935 = 61\,745$, si on ne tient pas compte des valeurs imposées par les conditions aux limites de Dirichlet. La matrice de masse est stockée sous un format *sky-line* et la résolution des systèmes linéaires est effectuée par une inversion directe à l'aide d'une décomposition LU (la décomposition n'est réalisée qu'une seule fois et réutilisée à chaque itération des différentes simulations).

Résultats pour un actionneur inactif

Une simulation a été menée pour un actionneur inactif en initiant le champ des vitesses à partir de la solution du problème de Stokes associé et sur une durée suffisamment longue pour obtenir un écoulement en régime périodique de type Von Kármán.

Le champ des vitesses obtenu à la fin de ce calcul nous a ensuite servi de champ initial $\boldsymbol{u_0}$ pour toutes les simulations effectuées par la suite. Ses composantes longitudinale et transverse sont visualisées avec son rotationnel sur la figure 5.4. L'allée de tourbillons

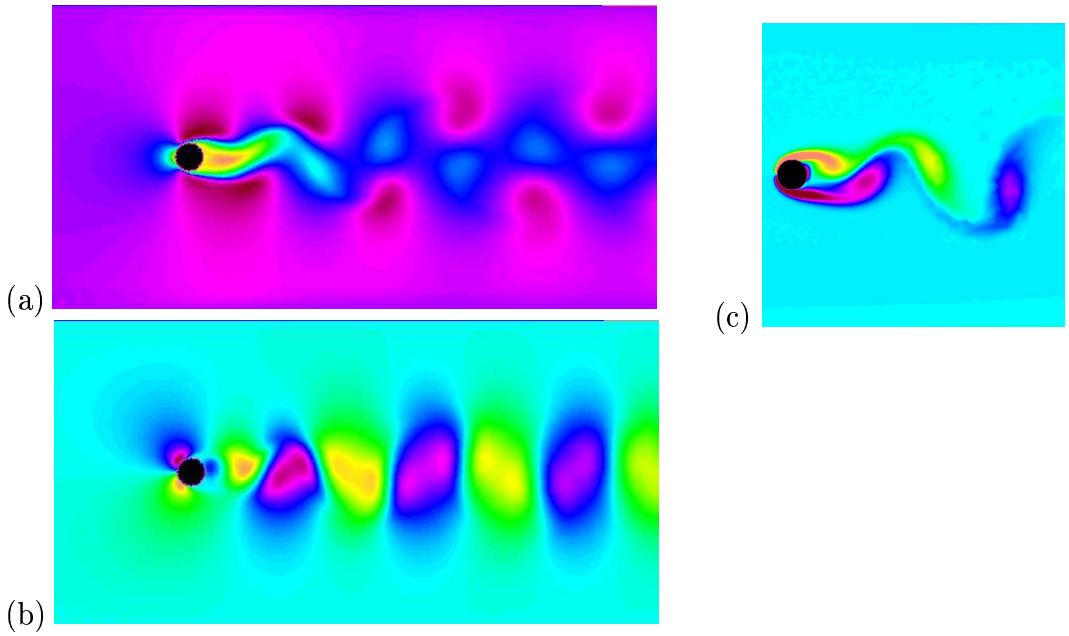


FIG. 5.4 – Visualisations de $\boldsymbol{u_0}$: (a) $(u_0)_{x_1}$ sur $[10,32]\times[0,11]$, (b) $(u_0)_{x_2}$ sur $[10,32]\times[0,11]$ et (c) $w_0 = \partial_{x_1}(u_0)_{x_2} - \partial_{x_2}(u_0)_{x_1}$ sur $[14,24]\times[0.5,10.5]$.

alternés de Von Kármán semble bien établie.

Le coefficient de portance est défini par

$$C_p = -\int_{\Gamma_C} (\sigma\,\boldsymbol{n}).\boldsymbol{x_2}\,ds$$

où $\sigma = \frac{1}{\mathrm{Re}}\left[\nabla u + \nabla u^T\right] - p\,\mathrm{I}_d$ est le tenseur des contraintes fluides et $\boldsymbol{x_2}$ le vecteur unitaire orienté dans la direction verticale x_2. Son histoire, calculée à partir de $\boldsymbol{u_0}$ pour un actionneur inactif, est donnée sur la figure 5.5. L'intervalle de temps qui sépare les deux maxima et les deux minima est d'environ 5.82 : le nombre de Strouhal de l'écoulement simulé (fréquence du régime périodique) est $\mathrm{St} = 0.172$. Cette valeur est proche de celles qui ont été obtenues pour de nombreuses simulations (voir [54] par exemple).

Enfin, des visualisations du rotationnel $w = \partial_{x_1}u_{x_2} - \partial_{x_2}u_{x_1}$ sont proposées à différents instants sur la figure 5.6. Le calcul correspondant à cette figure a été réalisé en effectuant

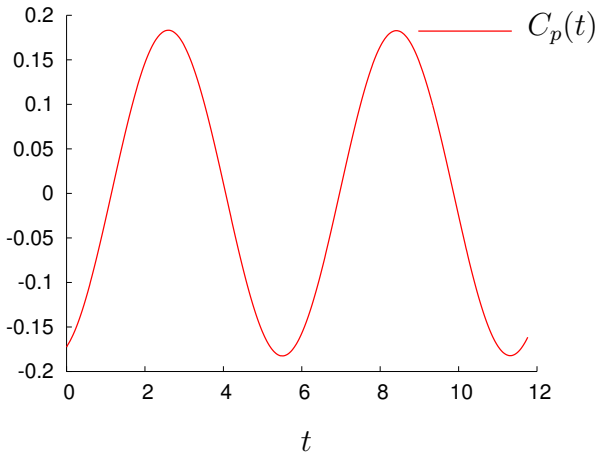


FIG. 5.5 – Évolution du coefficient C_p de portance pour $t \in [0, 11.8]$.

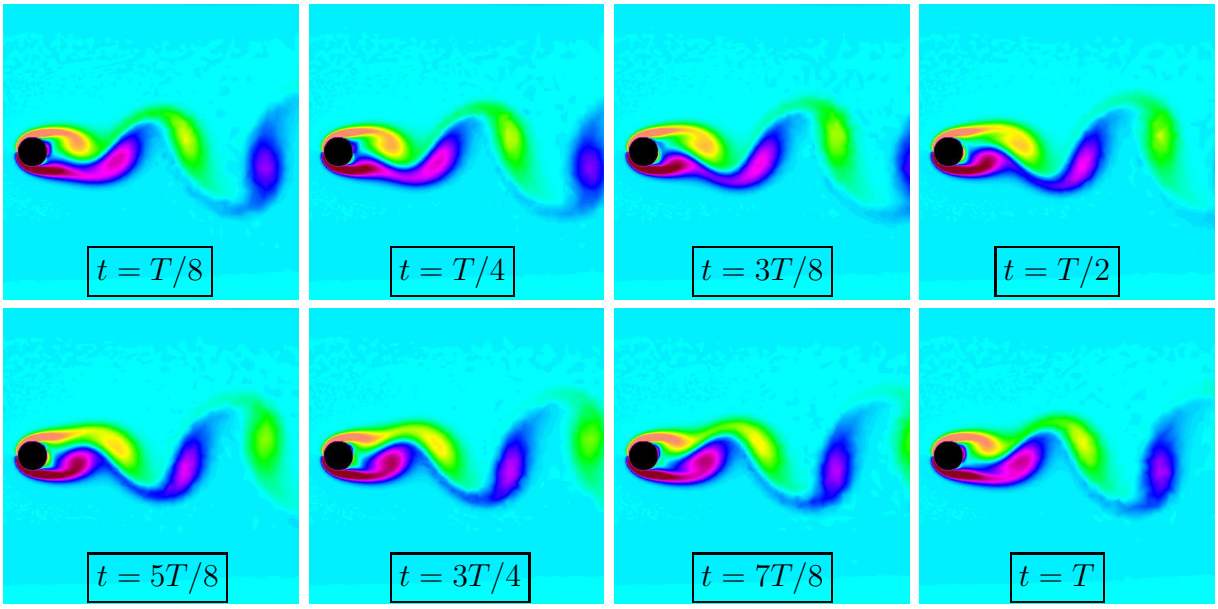


FIG. 5.6 – Rotationnel w de l'écoulement non contrôlé ($T = 5.82$).

240 itérations sur une période, c'est-à-dire pour $\Delta t = 5.82/240$ (30 itérations séparent chaque image). La simulation semble convenable, d'autant plus que le champ final au temps $T = 5.82$ est (quasiment) identique au champ initial (comparer les rotationels des figures 5.4 et 5.6).

5.2.3 Tests de l'algorithme itératif primitif

Cette section présente deux tests pour le problème de *flow tracking*. L'écoulement de référence $\boldsymbol{u}^R$ est calculé pour une commande $c^R(t)$ connue de l'actionneur sur une durée $T = 6$: par définition, $c^R(t)$ est donc une commande optimale pour le problème, ce qui va nous permettre d'analyser l'efficacité de l'algorithme de contrôle primitif.

Le premier test concerne l'écoulement de référence obtenu par le soufflage $c^R(t) = t/T$ et le second celui obtenu par la commande $c^R(t) = h(t)$ avec

$$h(t) = -\frac{e^{-1/t}}{4}\left[x\,(x-5)\,\cos(x^2/2) + 5\right]. \tag{5.25}$$

Cette commande est tracée sur la figure 5.7.

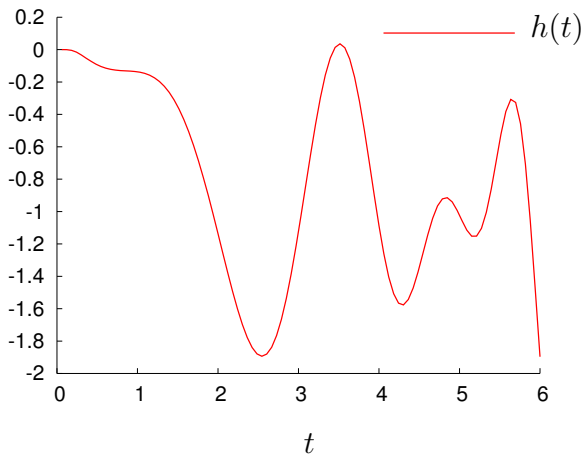


FIG. 5.7 – Commande $h(t)$.

Afin de visualiser l'effet de l'actionneur sur l'écoulement, la figure 5.8 montre l'évolution du rotationnel de l'écoulement de référence $\boldsymbol{u}^R$ qui est contrôlé par la commande $c(t) = h(t)$ (il a été simulé pour le pas de temps $\Delta t = T/240$). L'actionneur effectue alors essentiellement une aspiration du fluide au niveau du décollement de la couche limite, car $h(t) < 0$ la plupart du temps. L'effet de cette aspiration est de bloquer la génération du tourbillon qui aurait dû survenir en haut de l'obstacle, mais aussi d'empêcher le départ du tourbillon qui se situe en bas et qui se trouve aspiré. Les premiers tourbillons qui suivent subissent également l'effet de l'aspiration.

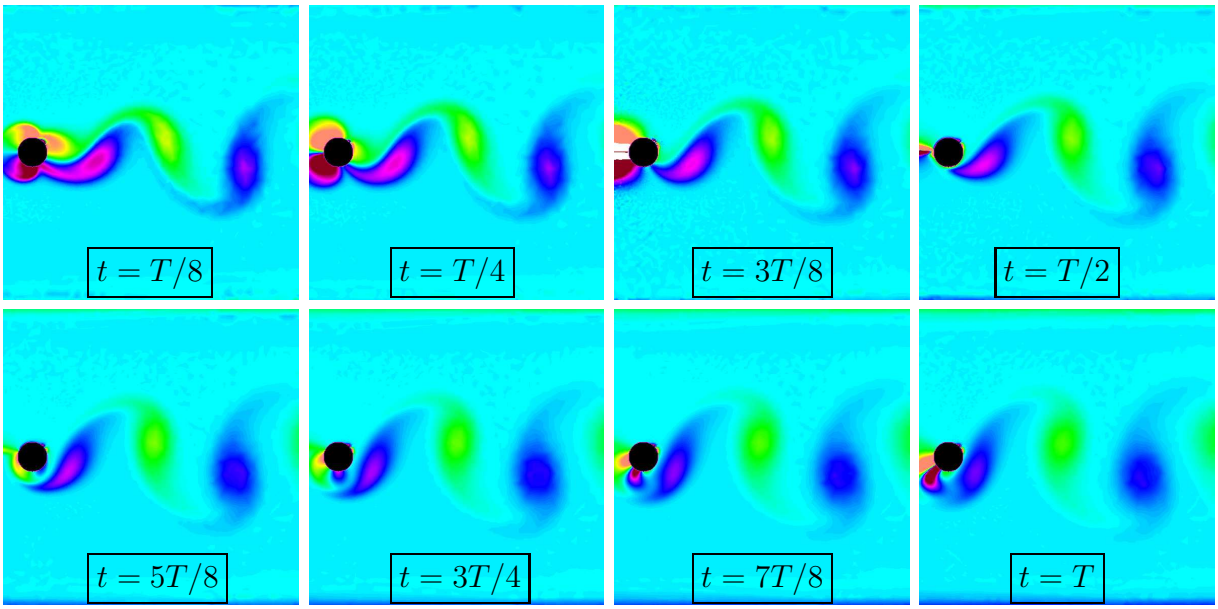


FIG. 5.8 – Rotationnel w^R de l'écoulement contrôlé avec la commande $c(t) = h(t)$ $(T = 6)$.

Lors de l'algorithme de contrôle, les simulations sont réalisées en 250 itérations et un champ de vitesse sur cinq est conservé : la base POD est déduite de 50 clichés. Pour les deux cas tests, les modèles POD-Galerkine sont construits en utilisant la même fonction de commande $\boldsymbol{\psi_1}$ et les M premiers modes POD qui capturent plus de 99% du champ de référence $\boldsymbol{u^e} - \bar{\boldsymbol{u}}^{\boldsymbol{e}}$, en fait entre 6 et 10 modes en pratique. La figure 5.9 propose une

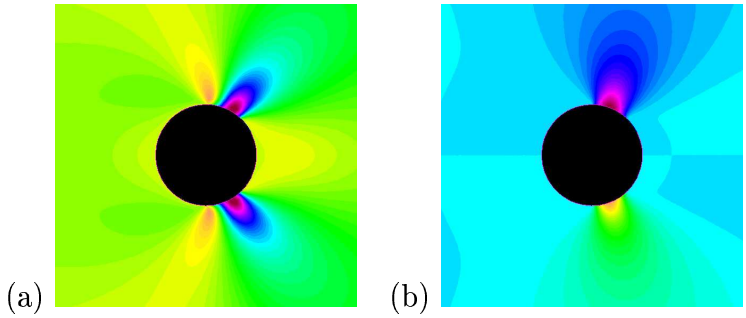


FIG. 5.9 – Composantes de la fonction de commande $\boldsymbol{\psi_1}$ sur $[13.5, 16.5] \times [4, 7]$: (a) longitudinale et (b) transverse.

visualisation dans le voisinage de l'obstacle des composantes longitudinales et transverses de la fonction de commande $\boldsymbol{\psi_1}$ qui permet de prendre en compte la condition de Dirichlet sur Γ_A, c'est-à-dire le soufflage ou l'aspiration effectués par l'actionneur.

L'algorithme itératif primitif décrit dans la section 5.1.2 est utilisé en partant de la commande $c^0(t) = 0$. De plus, l'étape 4 de calcul de la commande optimale du problème réduit est effectuée par l'algorithme de gradient conjugué de Polak-Ribière combiné avec une recherche linéaire de Wolfe (voir [9] pour ces méthodes) : la convergence vers un minimum local est obtenue pour un coût informatique faible. Les routines de calcul de la fonction coût réduite et de son gradient ont été validées par des comparaisons avec des approximations du gradient par différences finies.

Les premières commandes $c^k(t)$ obtenues au cours de l'algorithme sont tracées sur la figure 5.10. Dans les deux cas, la commande $c^1(t)$ qui a été calculée par optimisation du

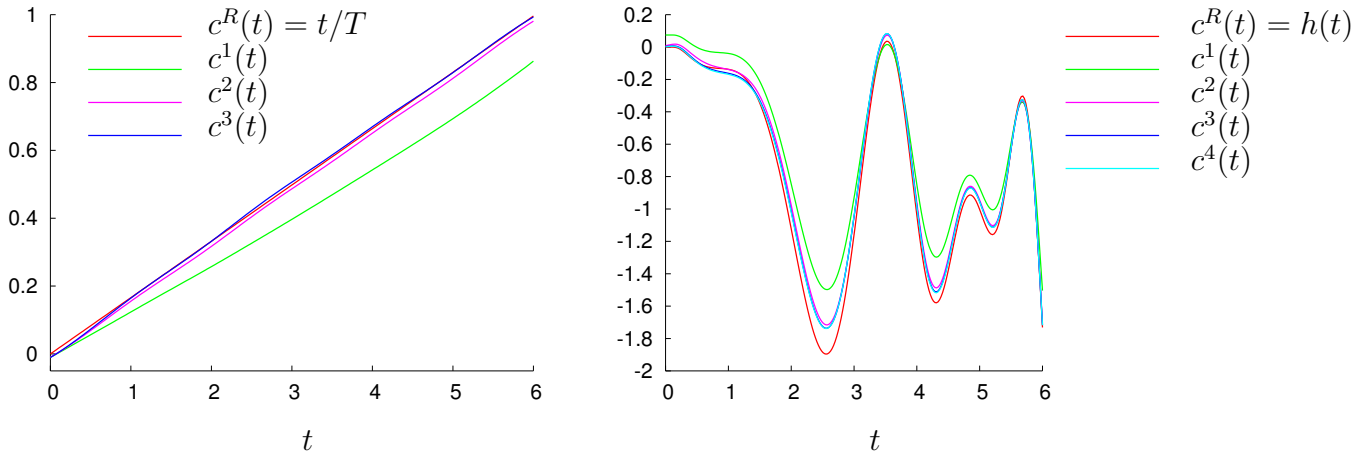


FIG. 5.10 – Commandes $c^k(t)$, obtenues au cours de l'algorithme itératif primitif, et $c^R(t)$ de référence.

premier modèle réduit est très proche de la commande optimale. Cela est dû au fait que l'écoulement de référence $\boldsymbol{u^R}$ est relativement proche de l'écoulement non contrôlé, qui sert à construire le premier modèle POD-Galerkine (en effet, $c^0(t) = 0$).

Pour le premier cas test ($c^R(t) = t/T$), les itérés c^k convergent rapidement vers $c^R(t)$, la commande optimale étant quasiment obtenue au bout de trois itérations : cela a permis de valider le code de contrôle.

Pour le second cas test, si les trois premières itérations s'approchent toujours plus près de la commande optimale c^R, on remarque que les itérés c^3 et c^4 sont quasiment identiques : l'algorithme primitif converge mais vers une commande qui est sensiblement différente de c^R.

5.2.4 Conclusions et perspectives

L'algorithme primitif proposé par Ravindran a été testé sur deux problèmes simples de *flow tracking*. Le second problème a mis en évidence le fait qu'il était difficile d'obtenir une approximation très précise de la commande optimale, l'algorithme pouvant converger vers une autre valeur.

Cependant, même dans ce cas de figure, l'utilisation des modèles réduits reste très prometteuse puisqu'elle permet en général de minimiser la fonction coût de manière significative pour un coût informatique faible. De plus, il est envisageable de raffiner *a posteriori* la commande calculée à l'aide d'une méthode de contrôle plus coûteuse mais plus précise.

Le code de contrôle qui a été développé devrait permettre de mener des investigations numériques plus appronfondies.

Il serait notamment intéressant de comparer les algorithmes itératifs primitif et à région de confiance sur un problème de contrôle plus délicat. Ce problème pourrait être un problème de *flow tracking* complexe, c'est-à-dire mené pour un champ de référence $\boldsymbol{u}^R$ très différent de l'écoulement non contrôlé et sur une fenêtre temporelle $[0, T]$ plus large (une dizaine de périodes de l'écoulement non contrôlé). En outre, il est également envisageable de tester une nouvelle fonction coût, par exemple celle basée sur le rotationnel (équation (5.3)).

Enfin, il pourrait être bénéfique de tester des stratégies d'enrichissement de la base POD. Un moyen simple d'y parvenir serait par exemple de sauvegarder certains modes POD au cours des itérations de la boucle de contrôle, puis de les réutiliser ultérieurement lors de la construction des modèles POD-Galerkine, ou encore d'ajouter aux bases POD quelques modes calculés au préalable par des simulations judicieusement choisies.

Conclusions et perspectives

Ce travail de thèse a concerné l'étude de la modélisation POD-Galerkine réduite en vue de son utilisation pour le contrôle actif instationnaire d'écoulements. Il se situe essentiellement dans le cadre des écoulements incompressibles.

La construction de modèles fiables, de dimension réduite, et représentatifs de la dynamique d'un écoulement, a constitué l'axe principal des recherches qui ont été menées. L'exploitation des modèles réduits pour le contrôle d'un écoulement instationnaire a néanmoins été abordée.

Tout d'abord, pour le cas d'un écoulement incompressible, un modèle POD-Galerkine qui prend explicitement en compte toutes les actions de l'environnement sur le fluide, y compris via des conditions aux limites de flux basées sur le tenseur physique des contraintes fluides, a été défini. En reprenant les travaux de Vigo [94], il a été montré que ce modèle réduit est théoriquement stable pour un écoulement fermé. Cependant, la question de la stabilité reste posée dans le cas des écoulements ouverts, donc en général pour ceux qui interviennent dans les applications concrètes (ailes d'avion, conduites) et qu'il serait intéressant de contrôler. De plus, ce résultat n'assure pas la stabilité des systèmes obtenus, puisque leur calcul implique nécessairement des erreurs numériques (en particulier au niveau des interactions triadiques globales). En outre, la modélisation repose sur une formulation variationnelle qui n'est exploitable que si les modes POD satisfont les conditions de Dirichlet homogènes appropriées. Le problème de la modélisation du terme de pression subsiste donc en pratique pour des conditions de Dirichlet complexes.

La modélisation réduite d'un écoulement incompressible, tridimensionnel, ouvert, décollant au niveau d'une marche descendante, à haut nombre de Reynolds et non homogène a ensuite été étudiée : une analyse qualitative des transferts énergétiques au sein de la base des modes POD, puis une étude quantitative des effets de la réduction de la base modale sur le modèle, via la définition et l'estimation de pseudo-viscosités, ont été menées. Il est apparu que les interactions modales présentent des propriétés similaires à celles qui ont été mises en évidence pour les modes de Fourier d'un écoulement homogène : une cascade directe d'énergie prédomine en moyenne et les transferts sont essentiellement effectués au niveau local, c'est-à-dire que les modes interagissent principalement avec des modes d'indice proche. Les profils de pseudo-viscosité obtenus rejoignent ces observations et confortent l'idée introduite par Aubry *et al.* [6] d'une amélioration des modèles réduits par des ajouts adéquats de viscosité qui sont alors censés modéliser les modes POD tron-

qués (en pratique les petites structures spatiales de l'écoulement). Ce travail a fait l'objet d'une publication dans la revue intitulée *Journal of Fluid Mechanics* [16]. Par la suite, nous avons réussi à tirer parti de la propriété de localité des interactions modales, bien que le problème de la modélisation des modes POD tronqués, qui est important en pratique pour ce type d'écoulements, n'ait pas été traité de manière spécifique mais globale, c'est-à-dire avec les problèmes du décalage entre nature des données et formulation variationnelle, des erreurs numériques ou de la modélisation du terme de pression.

En effet, des méthodes de calibration ont été définies afin de répondre à ces différentes difficultés inhérentes à la modélisation POD-Galerkine réduite. Pour cela, à l'instar de Delville *et al.* [17], l'idée a été d'optimiser le modèle en exploitant l'information temporelle qui est fournie par la POD, mais qui n'est pas utilisée par la méthode de Galerkine. Le principe général de la calibration a été décliné en trois méthodes : la première consiste à résoudre un problème d'optimisation sous contrainte dynamique non-linéaire, tandis que les autres conduisent à la résolution d'un système linéaire. La première méthode s'est avérée trop coûteuse pour pouvoir être exploitée pour des applications en mécanique des fluides, du moins pour les méthodes de descente par gradient conjugué qui ont été testées. En revanche, les deux autres ont permis une optimisation efficace du modèle réduit à six modes d'un écoulement bidimensionnel laminaire décollé et des modèles à 86 et 45 modes issus de l'écoulement tridimensionnel turbulent qui avaient servi de sujet d'étude précédemment. Des stratégies "de Galerkine partielles" ont ainsi pu être testées pour cet écoulement tridimensionnel : le nombre de calculs nécessaires à la construction d'un modèle réduit s'est avéré sensiblement inférieur à celui induit par la méthode de Galerkine "complète". Ces techniques de calibration devraient donc permettre d'étendre le champ d'application de la modélisation POD-Galerkine réduite à des écoulements complexes.

Dans un dernier temps, un code de calcul pour le contrôle instationnaire d'un écoulement bidimensionnel laminaire a été développé et testé pour un problème de *flow tracking*. Il permettra de mener des investigations plus approfondies sur les algorithmes itératifs de calcul de commande.

De nombreux axes de recherche peuvent permettre de prolonger ce travail. Tout d'abord, comme la section 1.4.4 le souligne, il est possible de définir une POD "d'ordre supérieur". La question se pose de savoir si cette POD pourrait permettre de calculer des modes encore plus représentatifs de l'écoulement étudié, ou encore de diminuer le nombre de clichés minimal nécessaire au calcul d'une base POD correcte. La POD temporelle, proposée par Kunisch *et al.* [49] et étudiée mathématiquement par Henri [33], nécessiterait également des investigations numériques approfondies.

Ensuite, nous avons montré que la prise en compte explicite des conditions de flux basées sur le tenseur des contraintes fluides est possible, mais elle doit encore être testée par des essais numériques.

Par ailleurs, il serait intéressant de tester les méthodes de calibration pour des écoulements compressibles. En effet, ces méthodes sont directement transposables au cas de ces écoulements, puisque Vigo [94] a montré qu'il était possible de construire des modèles réduits polynômiaux à partir des équations de Navier-Stokes compressibles pour un jeu de

variables approprié. De plus, il est envisageable de diminuer le coût de ces méthodes en ne choisissant qu'un ensemble restreint de coefficients à calibrer, ou de mieux tenir compte de la physique du problème, par exemple en imposant à la calibration de ne conduire qu'à une variation positive des coefficients d'origine visqueuse afin de modéliser l'effet des petites échelles. En outre, nous avons vu que la détermination d'un minimum local pour le problème d'optimisation non-linéaire associé à la calibration est parfois difficile à obtenir. Il pourrait être intéressant de rechercher des méthodes d'optimisation plus performantes, qui seraient à même de résoudre ce problème non-linéaire efficacement pour une calibration portant sur un nombre relativement important de coefficients. Par ailleurs, notons que le problème du choix automatique du paramètre α, auquel la calibration est très sensible, se pose.

Enfin, le code de contrôle doit être complété afin d'évaluer les performances de l'algorithme itératif à région de confiance qui a été proposé par Fahl [19] et de pouvoir développer des stratégies efficaces de calcul de commande. Une fois que seront définis des algorithmes de contrôle robustes pour des configurations laminaires et bidimensionnelles, les méthodes de calibration pourront alors être évaluées pour des configurations complexes dans le cadre du contrôle des écoulements instationnaires.

Annexe

Article présentant l'analyse des interactions modales et des effets de la réduction de la base POD au sein de la modélisation POD-Galerkine réduite de l'écoulement tridimensionnel turbulent (référence [16])

J. Fluid Mech. (2003), *vol.* 491, *pp.* 275–284.
DOI: 10.1017/S0022112003005615 Printed in the United Kingdom

Intermodal energy transfers in a proper orthogonal decomposition–Galerkin representation of a turbulent separated flow

By M. COUPLET[1], P. SAGAUT[2,1] AND C. BASDEVANT[3,1]

[1]ONERA, DSNA, 29 av. de la Division Leclerc, 92322 Châtillon, France

[2]Laboratoire de Modélisation en Mécanique, Université Pierre et Marie Curie, Boite 162, 4 place Jussieu, 75252 Paris Cedex 5, France

[3]Laboratoire de Météorologie Dynamique, École Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 5, France

(Received 28 January 2003 and in revised form 18 June 2003)

Energy transfers between modes obtained from the proper orthogonal decomposition (POD) of a turbulent flow past a backward-facing step are analysed with the aim of providing guidelines for modelling unresolved modes in truncated POD–Galerkin models. It is observed that energy transfers are local in the POD basis, and that the Fourier-decomposition-based concepts of forward and backward energy cascades are also valid in the POD basis, the net effect being a forward energy cascade. General features of the eddy-viscosity representation of kinetic energy transfers are investigated through *a priori* tests. It is observed that the ideal eddy-viscosity model should exhibit a cusp behaviour near the cutoff mode.

1. Introduction

The proper orthogonal decomposition (POD, also known as the Karhunen–Loève decomposition, see Holmes, Lumley & Berkooz 1996 for a survey) is a convenient tool for describing non-homogeneous turbulent flows. Indeed, it makes it possible to educe global coherent structures of the flow in an uniquivocal way. This decomposition being optimal, it became a popular way to construct dynamical systems representing turbulent flows.

The present paper aims to describe the global features of energy transfers between POD modes in a turbulent, wall-bounded, non-homogeneous separated flow. In particular, a quantitative analysis of the modal interactions, based on computations of eddy-viscosity-like parameters, is proposed. The selected configuration is the turbulent flow past a backward-facing step, with a turbulent inlet. The purpose is twofold.

First, energy transfers have historically been studied using Fourier decomposition, which is convenient for homogeneous flows only; thus, both theoretical (e.g. Kraichnan 1971) and numerical studies (e.g. Yeung, Brasseur & Wang 1995) have provided deep insight into the dynamics of turbulent fluctuations, assessing the existence of forward and backward energy cascades. A few studies based on the POD approach have addressed non-homogeneous turbulent flows, but generally in local configurations such as the minimal channel unit (Aubry *et al.* 1988; Podvin & Lumley 1998; Podvin 2001; Webber, Handler & Sirovich 1997, 2002), or a transient flat-plate boundary layer (Rempfer & Fasel 1994). Here, the POD–Galerkin method is applied to a

spatially extended, non-homogeneous separated flow. Moreover, as POD is equivalent to the Fourier decomposition in homogeneous directions, most authors apply an explicit hybrid POD/Fourier decomposition (Fourier decomposition in homogeneous directions followed by POD in the remaining directions). Such an approach is different from the full three-dimensional POD adopted in this study as far as mode selection is concerned, since the way the modes will be ordered are different: in the full POD, all modes are uniquely ordered (by their eigenvalues), while they can be ordered separately in each homogeneous direction in hybrid Fourier/POD analysis. As previously proposed in Rempfer & Fasel (1994), one goal here is to analyse the energy transfers within a full POD basis, looking at their main characteristics and comparing them with results drawn from the Fourier analysis in the isotropic case.

Secondly, when turbulent flows are considered, although very few POD modes contain most of the total turbulent kinetic energy and can be kept to construct a reduced-order dynamical system, the low-energetic modes, which drop out, must be taken into account in the POD–Galerkin approach to recover an accurate description of the flow. This problem is formally equivalent to that of large-eddy simulation (LES, see Sagaut 2002 for a general presentation). Following the proposal of Aubry *et al.* (1988), most authors use a diffusive model based on an extension of the Heisenberg spectral viscosity model for homogeneous flows: Berkooz *et al.* (1990); Podvin (2001); Ukeiley *et al.* (2001). As quoted by Aubry *et al.* (1988), such a model is very similar to the well-known Smagorinsky one, which can be interpreted as an extension of the same Heisenberg model in physical space for flows represented using a local basis. However, as previously mentioned, almost all works using that approach rely on a POD/Fourier decomposition (e.g. two-homogeneous-direction plane channel flow). Truncating the basis in both the Fourier and POD representation, the validity of the parameterization of unresolved modes using a viscosity-type model may be understood by invoking the Kolmogorov hypothesis and assuming that the cutoff occurs at small enough scales, as is done in the usual LES framework. The validity of this model for the minimal channel flow unit was assessed by Podvin (2001). However, when considering complex non-homogeneous flows, the question may arise of the validity of this type of eddy-viscosity model, since the underlying assumptions dealing with both the existence and the dominance of a forward energy cascade remain to be investigated. The second goal of the present study is therefore to provide an analysis of the representation of the energy transfers between modes via an eddy-viscosity assumption, in order to provide guidelines for the definition of models for the unresolved modes. Since it is known that even in the simple case of the Fourier representation of isotropic turbulence the global features of subgrid models are very dependent on the spectrum shape, the filter shape and the cutoff length, the case of truncated POD models is non-trivial.

2. Numerical database and POD–Galerkin method

The POD modes are computed using $M = 1000$ three-dimensional instantaneous snapshots obtained by performing a LES of a turbulent incompressible flow past a backward-facing step. The Reynolds number based on the mean streamwise velocity at the entrance and the height of the channel above the step is 66 100. That based on the inflow velocity Γ_{in} and the height h of the step is 7432. The three velocity components are stored over the full computational domain every 50 time steps. The computational domain (see figure 1) contains the turbulent inlet channel and the flow past the step. The sampling time is chosen so as to encompass at least one period of

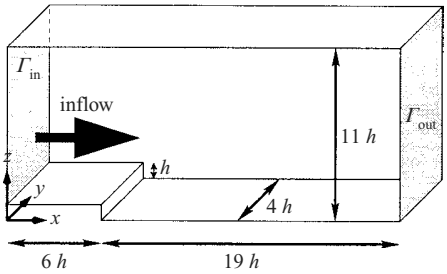


FIGURE 1. Geometry of the computational domain, corresponding to the spatial extent of the POD modes.

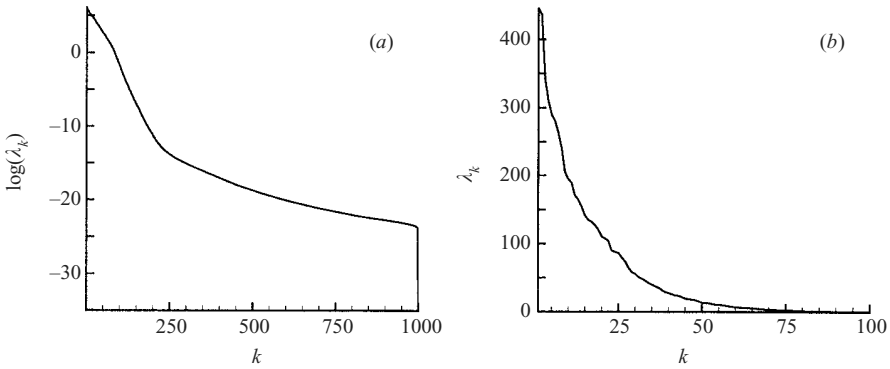


FIGURE 2. (*a*) Full POD spectrum; (*b*) first 100 POD modes.

the low-frequency breathing mode of the recirculation bubble. The reader is referred to Labbé, Sagaut & Montreuil (2002) for details on the LES.

The instantaneous velocity field $\boldsymbol{u}$ is split into a mean part $\bar{\boldsymbol{u}}(x, y, z) = \langle \boldsymbol{u}(x, y, z, t) \rangle$ and a fluctuating part $\tilde{\boldsymbol{u}} = \boldsymbol{u} - \bar{\boldsymbol{u}}$, where $\langle . \rangle$ denotes the average over the M snapshots. The POD decomposition is applied to $\tilde{\boldsymbol{u}}$, yielding

$$\tilde{\boldsymbol{u}}(x, y, z, t) = \sum_k \underbrace{(\tilde{\boldsymbol{u}}, \boldsymbol{\phi}_k)}_{\tilde{a}_k(t)} \boldsymbol{\phi}_k(x, y, z) \tag{2.1}$$

where $(.,.)$ is the classical L^2 inner product on the flow domain. $(\boldsymbol{\phi}_k)_{k \in [\![1,M]\!]}$ is the orthonormal POD basis and the $\tilde{a}_k(t)$ the time-dependent coefficients of the decomposition. They have the following orthogonality property:

$$\langle \tilde{a}_k \, \tilde{a}_j \rangle = \lambda_k \delta_{j,k} = {\sigma_k}^2 \delta_{j,k} \quad \forall k, j \tag{2.2}$$

and the basis is ordered so that $\lambda_k \geqslant \lambda_{k+1} \forall k$.

This decomposition is optimal in the sense that, for all n, the first n POD modes capture more kinetic energy on the average than any other set of n spatial functions. Figure 2 displays the POD spectrum on a logarithmic scale computed from the database. One of the main advantage of the POD basis, thanks to its optimality property, is that it allows very large data compression factors. In the present case, the first 87 modes contain 99.9% of the mean turbulent kinetic energy.

To analyse energy transfers among POD modes, the evolution equation for each mode is obtained by applying the Galerkin method on the space spanned by the

first $n \leqslant M$ POD modes. Considering periodic boundary conditions in the spanwise direction y, the no-slip condition at solid walls, a Dirichlet condition on velocity at the inflow plane, Γ_{in}, where values were prescribed using a precursor simulation of a turbulent incompressible plane channel flow, and the zero-stress outflow boundary condition on the exit plane, the following weak formulation is obtained from the non-dimensional-Navier–Stokes set of equations:

$$\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}, \boldsymbol{\phi}\right) + \frac{1}{Re}\left[\sum_{v\in\{x,y,z\}} (\nabla u_v, \nabla \phi_v)\right] + \underbrace{\int_{\Gamma_{\text{in}}} \left(p\boldsymbol{n} - \frac{1}{Re}\nabla\boldsymbol{u}\cdot\boldsymbol{n}\right)\cdot\boldsymbol{\phi}\,\mathrm{d}\sigma}_{\text{inlet boundary term}} = 0 \tag{2.3}$$

for all solenoidal test functions $\boldsymbol{\phi}$ satisfying periodic boundary conditions in spanwise direction and the no-slip condition at solid walls. The POD–Galerkin system is obtained by taking the first n modes as basis and test functions. Making the assumption that the inlet boundary term can be neglected (this is realistic since modes arising from the decomposition of $\tilde{\boldsymbol{u}}$ have very small contributions on the inlet plane Γ_{in}), the following n-dimensional polynomial dynamical system is derived:

$$\dot{a}_i(t) = p_i(A(t)) \quad \forall i \in [\![1, n]\!] \tag{2.4}$$

where $A(t) = (a_1(t), \cdots, a_n(t))$ and $a_k = \tilde{a}_k/\sigma_k$ for all k. Each p_i can be expressed as

$$p_i(A) = C_i^0 + \frac{D_i^0}{Re} + \sum_{k=1}^{n}\left(C_i^k + \frac{D_i^k}{Re}\right)a_k + \sum_{k_1=1}^{n}\sum_{k_2=1}^{k_1} C_i^{k_1,k_2}\, a_{k_1}a_{k_2} \tag{2.5}$$

with

$$C_i^0 = -((\bar{\boldsymbol{u}}\cdot\nabla)\bar{\boldsymbol{u}}, \boldsymbol{\phi}_i), \quad D_i^0 = -\sum_{v\in\{x,y,z\}} (\nabla \bar{u}_v, \nabla(\phi_i)_v), \tag{2.6}$$

$$C_i^k = -\sigma_k((\boldsymbol{\phi}_k\cdot\nabla)\bar{\boldsymbol{u}} + (\bar{\boldsymbol{u}}\cdot\nabla)\boldsymbol{\phi}_k, \boldsymbol{\phi}_i), \quad D_i^k = -\sigma_k \sum_{v\in\{x,y,z\}} (\nabla(\phi_k)_v, \nabla(\phi_i)_v), \tag{2.7}$$

and

$$C_i^{k_1,k_2} = -\frac{\sigma_{k_1}\sigma_{k_2}}{1+\delta_{k_1,k_2}}[((\boldsymbol{\phi}_{k_1}\cdot\nabla)\boldsymbol{\phi}_{k_2}, \boldsymbol{\phi}_i) + ((\boldsymbol{\phi}_{k_2}\cdot\nabla)\boldsymbol{\phi}_{k_1}, \boldsymbol{\phi}_i)]. \tag{2.8}$$

3. Intermodal kinetic energy transfers

The total fluctuating kinetic energy per mass unit is $K(t) = \frac{1}{2}\,\|\tilde{\boldsymbol{u}}\|^2 = \sum_i K_i(t)$ where $K_i(t) = \frac{1}{2}\,\lambda_i\, a_i(t)^2$ is the energy captured by the ith mode. From (2.5), we obtain

$$\dot{K}_i = \lambda_i a_i \dot{a}_i = \widetilde{C}_i^0\, a_i + \underbrace{\sum_{k=1}^{n} \widetilde{C}_i^k\, a_k a_i}_{\text{diadic interactions}} + \underbrace{\sum_{k_1=1}^{n}\sum_{k_2=1}^{k_1} \widetilde{C}_i^{k_1,k_2}\, a_{k_1}a_{k_2}a_i}_{\text{triadic interactions}} \tag{3.1}$$

where

$$\widetilde{C}_i^0 = \lambda_i\left(C_i^0 + \frac{D_i^0}{Re}\right), \quad \widetilde{C}_i^k = \lambda_i\left(C_i^k + \frac{D_i^k}{Re}\right), \quad \widetilde{C}_i^{k_1,k_2} = \lambda_i C_i^{k_1,k_2}. \tag{3.2}$$

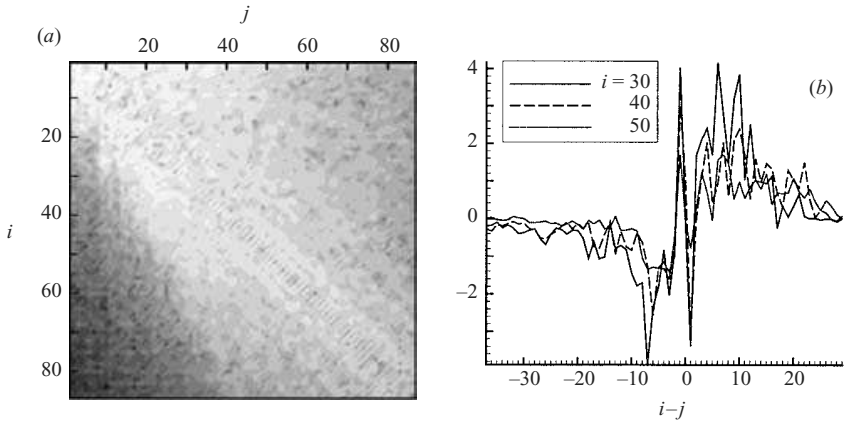


FIGURE 3. (*a*) Map of $\log(|\langle \Pi(i|j)\rangle|)$: iso-levels, ranging from white (maximum) to black (minimum). (*b*) $\langle \Pi(i|j)\rangle$ as a function of $(i-j)$ for three values of i.

Equation (3.1) shows that the evolution of the energy of a mode results from three kinds of interactions: a linear interaction with the mean flow, diadic terms arising from the interaction with the mean flow and viscous terms, and triadic interactions which account for the nonlinear inviscid interactions between modes. With the Fourier basis, diadic terms degenerate into simple linear terms, while triadic interactions are non-zero only for specific triads. Therefore differences between POD and usual Fourier decomposition are that viscous terms yield interactions between modes and that all groups of three modes lead to energy transfers. However it is worth noting that mean energy exchanges between POD modes through diadic interactions are zero since $\langle a_i a_j\rangle = \delta_{i,j}$.

Energetic exchanges via triadic interactions are now considered. The triadic term $\widetilde{C}_i^{k_1,k_2} a_{k_1} a_{k_2} a_i$ is regarded as the influence of the mode whose index is $\max(k_1, k_2)$ on the variation of K_i. Thus, the influence of the jth mode on the energy of the ith mode is

$$\Pi(i|j) = \sum_{k=1}^{j} \widetilde{C}_i^{j,k} a_k a_j a_i. \tag{3.3}$$

The absolute value of the mean transfer $\langle \Pi(i|j)\rangle$ for the first 87 modes is presented in figure 3; both the global energy transfer map, using a logarithmic scale for a clarity, and profiles for three values of i are plotted. It is observed that energy transfers among POD modes are local; indeed, the mean transfer is negligible for modes (i, j) such that $|i-j| \geqslant 25$. That property of locality was raised by the results of Rempfer & Fasel (1994), but for a transitional flow and in a small three-dimensional window of the whole initial computational domain. It can be seen as an extension of the well-known result that a Fourier mode with wavenumber k will exchange most of its energy with modes within the range $[k/2, 2k]$, i.e. that energy transfers are local, Kraichnan (1971, 1976). This observation is consistent with the result that POD modes converge toward Fourier modes in the limit of very high wavenumber, i.e. for dissipative scales (see Foias, Manley & Sirovich 1990), but it was not *a priori* obvious in the present case since the cutoff occurs within larger scales. Moreover, analyses carried out on the Fourier basis deal only with homogeneous turbulence and with the inertial range of the spectrum, whereas in the present case, the flow is non-homogeneous, wall

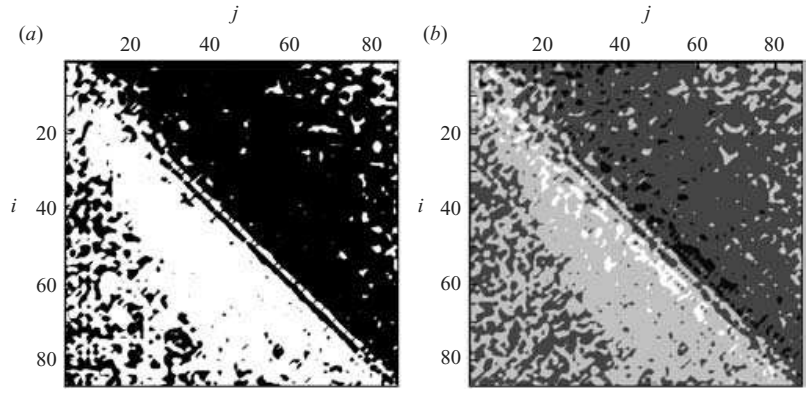


FIGURE 4. (*a*) Sign of the mean triadic transfer term $\langle \Pi(i|j) \rangle$. Black: negative; white: positive. (*b*) Percentage of time during which $\Pi(i|j)$ is positive (minimum: 27.5%; maximum: 77.7%) with four grey levels ranging from black to white: [27.5, 40], [40, 50], [50, 60] and [60, 77.7].

bounded and separated, and POD modes are global, i.e. they integrate the dynamics over the whole computational domain.

The direction of transfer is recovered by looking at the sign of the mean transfer term $\langle \Pi(i|j) \rangle$. The corresponding map is displayed in figure 4. Black regions correspond to a negative mean value, i.e. to a net drain of the kinetic energy of mode i by mode j, and white regions to a positive sign, i.e. to a net gain of kinetic energy by mode i. In good agreement with results from Fourier decomposition, it is observed that the main phenomenon is a forward energy cascade among the POD modes: a mode i drains energy from modes $j < i$ and redistributes energy toward modes $k > i$. Some small regions corresponding to an inverse cascade are also detected, but they are associated with very small absolute values of the mean transfer, and thus should not be considered as evidence for the existence of an inverse cascade in the mean.

To obtain a deeper insight into the dynamics of POD modes, the percentage of time during which the instantaneous transfer term $\Pi(i|j)$ is positive is plotted in figure 4. It is seen that net energy gains (resp. losses) are associated with regions where the transfer is most of the time in the same direction as the mean transfer, but that the transfer between two modes is in both directions during the full integration time. This can be interpreted has a generalization of the classical finding of the existence of a backward energy transfer among Fourier modes, or the fact that the sign of local transfers across a cutoff wavenumber may change from time to time and point to point in physical space.

It was observed that the fluctuating kinetic energy transfer among POD modes exhibits many features already observed using a Fourier-mode decomposition. This similarity can be explained by looking at the flow structures associated with each POD mode. Some of them are shown in figure 5, where isosurfaces of the Q-criterion (Hunt, Wray & Moin 1988) are displayed. It is seen that the higher the index of the POD mode is, the smaller are the flow structures. This is consistent with the fact that POD modes are sorted by decreasing order of kinetic energy, and that small structures are less energetic than larger ones. So, in agreement with Kolmogorov's local isotropy hypothesis, the main conclusions from the Fourier analysis can be extended to the POD decomposition framework. Consequently, the main phenomenon is a cascade

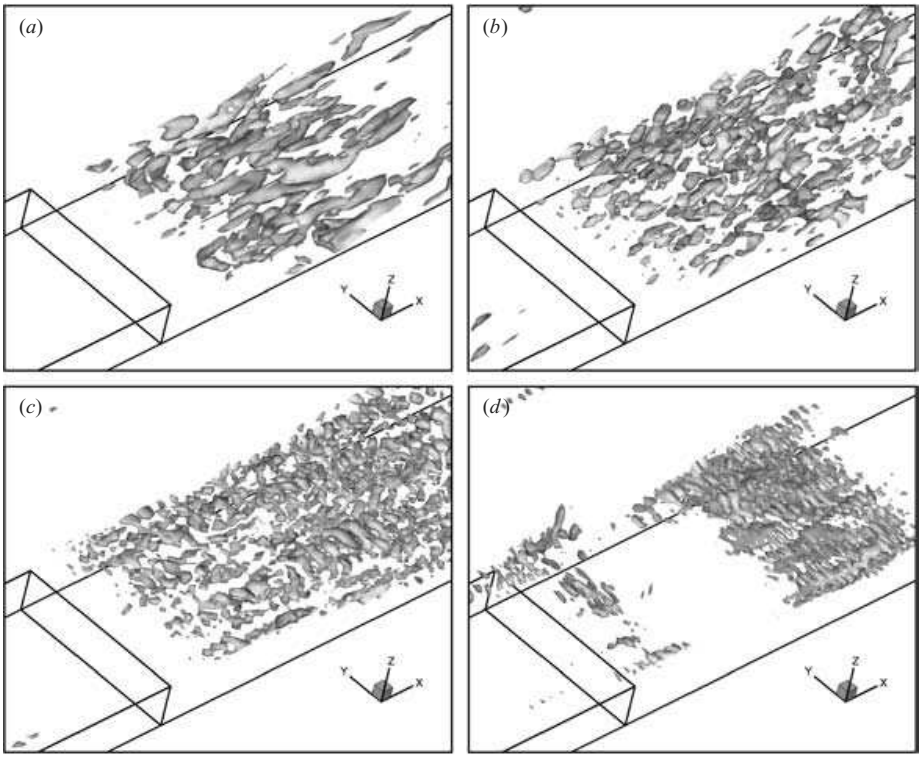


FIGURE 5. Isosurfaces of the Q criterion for some POD modes: (*a*) ϕ_1 with $Q = Q_1$; (*b*) ϕ_{20} with $Q = \frac{2}{3}Q_1$; (*c*) ϕ_{42} with $Q = \frac{10}{3}Q_1$; (*d*) ϕ_{87} with $Q = \frac{10}{3}Q_1$.

of kinetic energy from low- toward high-index modes. This forward cascade is a net effect, and some inverse cascade transfers occur over short durations, yielding an inverse cascade.

4. On eddy-viscosity parameterization of kinetic energy transfers

We now analyse the POD modal interactions through the computations of pseudo-eddy-viscosities, with emphasis on the parameterization of the energy transfers across a cutoff index l. This, which is an extension of the usual LES closure problem, is encountered when a truncated POD basis is used, i.e. when some POD modes are discarded. Since the use of POD leads to optimal data compression, very-low-order models can be obtained from the POD–Galerkin method, which will mimic either DNS or LES. In the case of LES-like dynamical systems, the problem of taking into account interactions with discarded modes should be solved. The global features of eddy-viscosity type models for POD need to be determined, since many studies carried out in the Fourier framework (e.g. Domaradzki *et al.* 1993, Zhou & Vahala 1993) have shown that the eddy-viscosity representation is very dependent on many parameters (e.g. spectrum shape, filter shape, cutoff wavenumber). A similar analysis is provided here which is expected to give useful informations to improve an LES-like POD–Galerkin model.

Introducing the cutoff index l in (2.5), the resolved (denoted by the $\leqslant$ exponent) and unresolved interactions ($>$ exponent) terms are

$$p_i(A) = c_i^{\leqslant}(A) + c_i^{>}(A) + (d_i^{\leqslant}(A) + d_i^{>}(A))/Re \tag{4.1}$$

with

$$c_i^{\leqslant}(A) = C_i^0 + \sum_{k=1}^{l} C_i^k\, a_k + \sum_{k_1=1}^{l}\sum_{k_2=1}^{k_1} C_i^{k_1,k_2}\, a_{k_1}a_{k_2}, \quad d_i^{\leqslant}(A) = D_i^0 + \sum_{k=1}^{l} D_i^k\, a_k, \tag{4.2}$$

$$c_i^{>}(A) = \sum_{k=l+1}^{n} C_i^k\, a_k + \sum_{k_1=l+1}^{n}\sum_{k_2=1}^{k_1} C_i^{k_1,k_2}\, a_{k_1}a_{k_2} \quad \text{and} \quad d_i^{>}(A) = \sum_{k=l+1}^{n} D_i^k\, a_k. \tag{4.3}$$

Corresponding terms arising from (3.1) are $a_i c_i^{\leqslant}(A)$, $a_i d_i^{\leqslant}(A)$, $a_i c_i^{>}(A)$ and $a_i d_i^{>}(A)$. The closure problem consists in expressing unresolved terms in the momentum equation and/or the kinetic energy equation as functions of the resolved modes. We consider here eddy-viscosity-type closures, which can be expressed as

$$p_i(A) \approx c_i^{\leqslant}(A) + (1/Re + \nu(i|l))d_i^{\leqslant}(A), \quad \text{i.e.} \quad p_i^{>} \approx \nu(i|l)\, d_i^{\leqslant} \tag{4.4}$$

with $p_i^{>} = c_i^{>} + d_i^{>}/Re$. The pseudo eddy-viscosity $\nu(i|l)$ *a priori* depends on the index of the mode considered and on the cutoff index, as is the case for the Kraichnan–Chollet–Lesieur subgrid viscosity in Fourier space (see Kraichnan 1976; Chollet & Lesieur 1981). Two ways computing values of this closure parameter $\nu(i|l)$ are investigated below, corresponding to the mean value over the set of snapshots or the least-square approximation. The three following values are obtained by applying these approximations to the momentum equation and to the kinetic energy equation (mean value computed from the momentum equation yields irrelevant results):

$$\nu_i = \frac{\langle p_i^{>}(A) d_i^{\leqslant}(A)\rangle}{\langle (d_i^{\leqslant}(A))^2\rangle} \text{ (least-square approximation from momentum equation);} \tag{4.5}$$

$$\underline{\widetilde{\nu_i}} = \frac{\langle a_i\, p_i^{>}(A)\rangle}{\langle a_i d_i^{\leqslant}(A)\rangle} \text{ (mean value from kinetic energy equation);} \tag{4.6}$$

$$\widetilde{\nu_i} = \frac{\langle a_i^2\, p_i^{>}(A) d_i^{\leqslant}(A)\rangle}{\langle (a_i d_i^{\leqslant}(A))^2\rangle} \text{ (least-square approximation from kinetic energy equation).} \tag{4.7}$$

Usual subgrid-viscosity models for LES are based on the budget equation for the resolved kinetic energy, and thus are similar to closures defined by (4.6) and (4.7). Results obtained for several cutoff values are presented in figure 6. Only modes $1 \leqslant i \leqslant 87$ are used to compute the eddy viscosity, the other modes being unimportant. The locality of the transfers observed in the preceding section also indicates that most transfers occur within this part of the POD spectrum for the cutoff values considered here.

It is observed that the three definitions yield similar eddy-viscosity profiles which share several properties with the theoretical eddy viscosity in Fourier space for a Kolmogorov spectrum and a sharp-cutoff filter: (i) a cusp is observed near the cutoff as i tends to l (zero viscosity is recovered for $i = l$, since modes higher than 87 are discarded), (ii) $\nu(i|l)$ is a decreasing function of l for fixed i and (iii) the maximum value of the cusp is a nearly constant function of the cutoff index l. These three observations are consistent with the previous statement that the global interactions

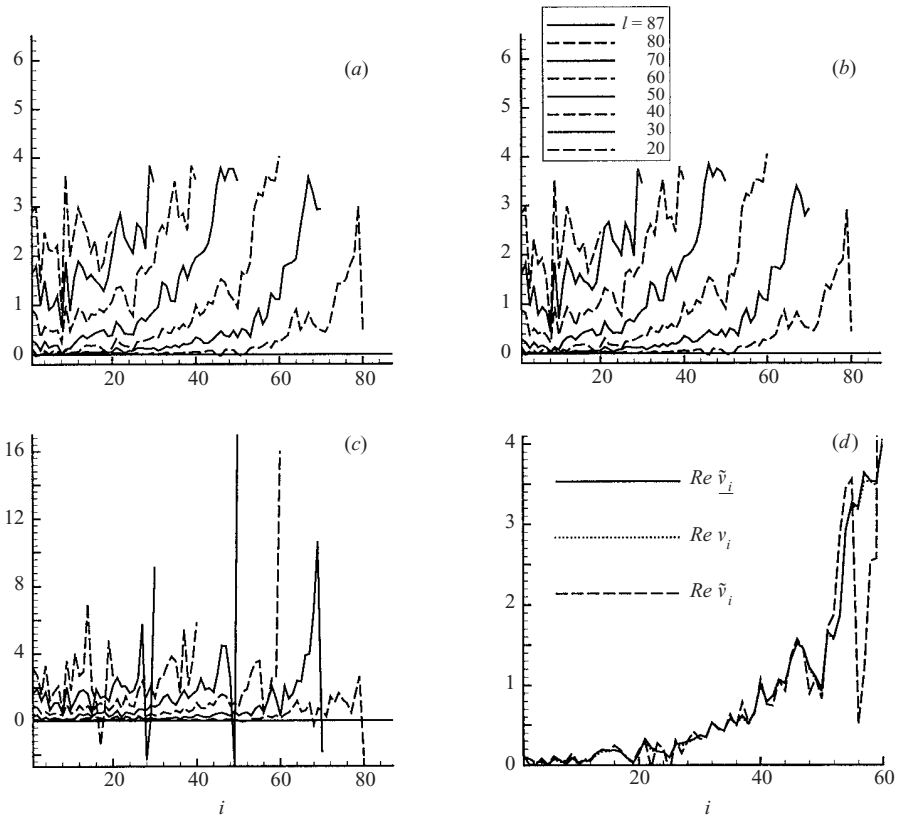


FIGURE 6. Artificial viscosities computed with different cutoff index ($l = 87$, 80, 70, 60, 40, 30 and 20): (*a*) *Re* $\underline{\tilde{\nu}_i}$, (*b*) *Re* ν_i and (*c*) *Re* $\tilde{\nu}_i$. (*d*) Comparison of *Re* $\underline{\tilde{\nu}_i}$, *Re* ν_i and *Re* $\tilde{\nu}_i$ for $l = 60$ (the scale was chosen with respect to the maximum value of *Re* $\underline{\tilde{\nu}_i}$).

between POD modes of increasing index number (i.e. of decreasing energy) are similar to interactions between Fourier modes of increasing wavenumber. The recovery of a cusp-like behaviour near the cutoff on the POD basis is very interesting, since it implies that consistent eddy-viscosity-type models should be mode-dependent. Note that the existence of this cusp is not a straightforward extension of classical Fourier-space results, since the latter disappear for realistic spectrum shapes or smooth filters (Leslie & Quarini 1979).

5. Conclusion

Energy transfers between POD modes representing a turbulent incompressible flow over a backward-facing step have been studied. Using 1000 instantaneous three-dimensional snapshots, the interaction terms have been reconstructed using the Navier–Stokes equations and a Galerkin approach. Energy transfers among POD modes are found to share several properties with their counterparts in Fourier space: (i) they are local in the POD basis in the sense that the transfer term $\Pi(i|j)$ is a rapidly decreasing function of $|i-j|$, (ii) a net forward energy cascade exists, i.e. mode i drains kinetic energy from modes $j < i$ and redistributes it to modes $j > i$ and (iii) a backward energy cascade was observed, since the sign of $\Pi(i|j)$ is not constant

over the 1000 snapshots. These similarities could be explained by POD modes being ranked in decreasing kinetic energy order, and thus associated with smaller and smaller vortical structures. As a consequence, the integrated interactions over the whole computational domain exhibit the same properties as the model interactions within Fourier modes, despite the non-homogeneous character of the present flow.

An interesting consequence is that an eddy-viscosity parameterization of transfers between resolved and unresolved modes in a truncated POD basis shares many properties with its counterpart in Fourier space including the existence of a cusp near the cutoff. Present results suggest that the pseudo-eddy-viscosity $\nu(i|l)$ should explicitly depend on both i and l. Future works will deal with the development of closed truncated POD–Galerkin models.

REFERENCES

Aubry, N., Holmes, P., Lumley, J. & Stone, E. 1988 The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.* **192**, 115–173.

Berkooz, G., Guckenheimer, J., Holmes, P., Lumley, J., Marsden, J., Aubry, N. & Stone, E. 1990 Dynamical-systems-theory approach to the wall region. In *AIAA 21st Fluid Dynamics, Plasma Dynamics and Laser Conference*.

Chollet, J. & Lesieur, M. 1981 Parametrization of small scales of three-dimensional isotropic turbulence utilizing spectral closures. *J. Atmos. Sci.* **38**, 2747–2757.

Domaradzki, J., Metcalfe, R., Rogallo, R. & Riley, J. 1993 An analysis of subgrid-scale interactions in numerically simulated isotropic turbulence. *Phys. Fluids* A **5**, 1747.

Foias, C., Manley, O. & Sirovich, L. 1990 Empirical and stokes eigenfunctions and the far-dissipative turbulent spectrum. *Phys. Fluids* A **2**, 464–467.

Holmes, P., Lumley, J. & Berkooz, G. 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press.

Hunt, J., Wray, A. & Moin, P. 1988 Eddies, stream, and convergence zones in turbulent flows. In *Proc. 1988 Summer Program*, pp. 193–208. CTR, Stanford University.

Kraichnan, R. 1971 Inertial-range transfer in two- and three-dimensional turbulence. *J. Fluid Mech.* **47**, 525–535.

Kraichnan, R. 1976 Eddy viscosity in two and three dimensions. *J. Atmos. Sci.* **33**, 1521–1536.

Labbé, O., Sagaut, P. & Montreuil, E. 2002 Large-eddy simulation of heat transfert over a backward-facing step. *J. Numer. Heat Transfer* A **42**, 73–90.

Leslie, D. & Quarini, G. 1979 The application of turbulence theory to the formulation of subgrid modelling procedures. *J. Fluid Mech.* **91**, 65–91.

Podvin, B. 2001 On the adequacy of the ten-dimensional model for the wall layer. *Phys. Fluids* **13**, 210–224.

Podvin, B. & Lumley, J. 1998 A low-dimensional approach for the minimal flow unit. *J. Fluid Mech.* **362**, 121–155.

Rempfer, D. & Fasel, H. 1994 Dynamics of three-dimensional coherent structures in a flat-plate boundary layer. *J. Fluid Mech.* **275**, 257–283.

Sagaut, P. 2002 *Large Eddy Simulations for Incompressible Flows*, 2nd Edn. B. Springer.

Ukeiley, L., Cordier, L., Manceau, R., Delville, J., Glauser, M. & Bonnet, J. 2001 Examination of large-scale structures in a turbulent plane mixing layer. Part 2. Dynamical systems model. *J. Fluid Mech.* **441**, 67–108.

Webber, G., Handler, R. & Sirovich, L. 1997 The Karhunen-Loève decomposition of minimal channel flow. *Phys. Fluids* **9**, 1054–1066.

Webber, G., Handler, R. & Sirovich, L. 2002 Energy dynamics in a turbulent channel flow using the Karhunen-Loève approach. *Intl J. Numer. Meth. Fluids* **40**, 1381–1400.

Yeung, P., Brasseur, J. & Wang, Q. 1995 Dynamics of direct large-small scale couplings in coherently forced turbulence: concurrent physical and Fourier-space views. *J. Fluid Mech.* **283**, 43–95.

Zhou, Y. & Vahala, G. 1993 Reformulation of recursive-renormalization-group based subgrid modeling of turbulence. *Phys. Rev.* E **47**, 2053.

Bibliographie

[1] K. Afanasiev and M. Hinze. Adaptive control of a wake flow using proper orthogonal decomposition. In *Shape Optimization & Optimal Design : Proceedings of the IFIP Conference*. Marcel Dekker, 2000.

[2] N. Alexandrov, J.E. Dennis, R.M. Lewis, and V. Torczon. A trust region framework for managing the use of approximation models in optimization. ICASE Report No. 1997-50, NASA Langley Research Center, Hampton, Virginia, 1997.

[3] J.S.R. Anttonen, P.I. King, and P.S. Beran. The accuracy of pod-based reduced-order models with deforming grids. In *15th AIAA Computational Fluid Dynamics Conference*, 2001. AIAA Paper 2001-2541.

[4] E. Arian, M. Fahl, and E.W. Sachs. Trust-region proper orthogonal decomposition for flow control. ICASE Report No. 2000-25, NASA Langley Research Center, Hampton, Virginia, 2000.

[5] N. Aubry, R. Guyonnet, and R. Lima. Spatiotemporal analysis of complex signals : theory and applications. *Journal of Statistical Physics*, 64(3/4) :683–739, 1991.

[6] N. Aubry, P. Holmes, J.L. Lumley, and E. Stone. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.*, 192 :115–173, 1988.

[7] M. Bergmann, L. Cordier, and J.-P. Brancher. Contrôle optimal d'un modèle réduit du sillage d'un cylindre circulaire. In *Colloque de l'Association Aéronautique et Astronautique de France*, 2003.

[8] G. Berkooz, J. Guckenheimer, P. Holmes, J. Lumley, J. Marsden, N. Aubry, and E. Stone. Dynamical-systems-theory approach to the wall region. In *AIAA 21st Fluid Dynamics, Plasma Dynamics and Laser Conference*, 1990. AIAA Paper 90-1639.

[9] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Optimisation numérique*. Springer-Verlag, 1997.

[10] H. Brézis. *Analyse fonctionnelle : théorie et applications*. Dunod, 1999.

[11] J.P. Chollet and M. Lesieur. Parametrization of small scales of three-dimensional isotropic turbulence utilizing spectral closures. *J. Atmos. Sci.*, 38 :2747–2757, 1981.

[12] P.G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson, 1982.

[13] P.G. Ciarlet. *Élasticité tridimentionnelle*. Masson, 1986.

[14] L. Cordier. *Étude de systèmes dynamiques basés sur la décomposition orthogonale aux valeurs propres, application à la couche de mélange turbulente et à l'écoulement entre deux disques contra-rotatifs.* PhD thesis, Université de Poitiers, 1996.

[15] M. Couplet, C. Basdevant, and P. Sagaut. Stabilized POD-Galerkin models for turbulent flows. In *European Congress on COmputational Methods in Applied Sciences and Engineering*, 2004.

[16] M. Couplet, P. Sagaut, and C. Basdevant. Intermodal energy transfers in a proper orthogonal decomposition-Galerkin representation of a turbulent seperated flow. *J. Fluid Mech.*, 491 :275–284, 2003.

[17] J. Delville, E. Lamballais, C. Braud, and P. Coiffet. Rapport final de la convention PEA 982610 ayant pour thème la génération de conditions amont. CEAT, Poitiers, France, 2001.

[18] J.A. Domaradzki, R.W. Metcalfe, R.S. Rogallo, and J.J. Riley. An analysis of subgrid-scale interactions in numerically simulated isotropic turbulence. *Phys. Fluids* A, 5 :1747, 1993.

[19] M. Fahl. *Trust-region methods for flow control based on reduced order modelling.* PhD thesis, Universität Trier, 2000.

[20] P. Faurre. *Notes d'optimisation.* École Polytechnique, Ellipse, France, 1988.

[21] C.A.J. Fletcher. *Computational Galerkin methods.* Springer-Verlag, Berlin, 1984.

[22] C. Foias, I. Kukavica, M. Jolly, and E.S. Titi. The Lorenz equations as a metaphore for the Navier-Stokes equations. *Discrete and Continuous Dynamical Systems*, 7(2) :403–429, 2001.

[23] C. Foias, O. Manley, and L. Sirovich. Empirical and Stokes eigenfunctions and the far-dissipative turbulent spectrum. *Phys. Fluids* A, 2 :464–467, 1990.

[24] E. Gadoin, P. Le Quéré, and O. Daube. Vers un système réduit basé sur les modes propres de l'opérateur de Navier-Stokes. In *XVème Congrès Français de Mécanique*, 2001.

[25] B. Galletti, C.H. Bruneau, L. Zannetti, and A. Iollo. Low-order modelling of laminar flow regimes past a confined square cylinder. *J. Fluid Mech.*, 503 :161–170, 2004.

[26] G.H. Golub and C.F. Van Loan. *Matrix computations.* John Hopkins University Press, Baltimore, 1989.

[27] W.R. Graham, J. Peraire, and K.T. Tang. Optimal control of vortex shedding using low order models. Part 1. Open-loop model development. *Int. J. for Numer. Meth. in Engrg.*, 44(7) :945–972, 1999.

[28] W.R. Graham, J. Peraire, and K.T. Tang. Optimal control of vortex shedding using low order models. Part 2. Model-based control. *Int. J. for Numer. Meth. in Engrg.*, 44(7) :973–990, 1999.

[29] W.R. Graham, J. Peraire, and K.Y. Tang. Optimal control of vortex shedding using low-order models. *Int. J. Numer. Meth. Engineering*, 44 :945–990, 1999.

[30] M.J. Greenacre. *Theory and applications of correspondence analysis.* Academic Press, 1984.

[31] J. Guckenheimer and P. Holmes. *Nonlinear oscillations, dynamical systems and bifurcation of vector fields.* Springer, 1986.

[32] F. Hecht, O. Pironneau, and K. Ohtsuka. *Freefem++ manual*, 2004. http ://www.freefem.org.

[33] T. Henri and J.-P. Yvon. Convergence estimates of POD Galerkin methods for parabolic problems. Prépublication 02-48, Institut de Recherche Mathématique de Rennes, octobre 2002.

[34] T. Henri and J.-P. Yvon. Stability of the POD and convergence of the POD Galerkin method for parabolic problems. Prépublication 02-40, Institut de Recherche Mathématique de Rennes, septembre 2002.

[35] T. Henri and J.-P. Yvon. A result on using proper orthogonal decomposition for controlled parabolic problems. Prépublication 03-03, Institut de Recherche Mathématique de Rennes, janvier 2003.

[36] P. Holmes, J.L. Lumley, and G. Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry.* Cambridge University Press, 1998.

[37] J.C.R. Hunt, A.A. Wray, and P. Moin. Eddies, stream, and convergence zones in turbulent flows. In *Proc. 1988 Summer Program*, pages 193–208. CTR, Standford University, 1988.

[38] A. Iollo, S. Lanteri, and J.A. Désidéri. Stability properties of POD-Galerkin approximations for the compressible Navier-Stokes equations. *Theoret. Comput. Fluid Dynamics*, 13 :377–396, 2000.

[39] K. Ito and S.S. Ravindran. A reduced-order method for simulation and control of fluid flows. *J. Comput. Phys.*, 143 :403–425, 1998.

[40] A. Jameson, L. Martinelli, and N.A. Pierce. Optimum aerodynamic design using the Navier-Stokes equations. *Theor. Comput. Fluid Dyn.*, 10 :213–237, 1998.

[41] W. Jürgens and H.-J. Kaltenbach. Eigenmode decomposition of turbulent velocity fields behind a swept, backward-facing step. *Journal of Turbulence*, 4(018), 2003.

[42] S. Kang and H. Choi. Suboptimal feedback control of turbulent flow over a backward-facing step. *J. Fluid Mech.*, 463 :201–227, 2002.

[43] M. Kirby. Minimal dynamical systems from PDEs using Sobolev eigenfunctions. *Physica D*, 57 :466–475, 1992.

[44] S. Klainerman and A. Majda. Compressible and incompressible fluids. *Comm. Pure Appl. Math.*, 35 :629–651, 1982.

[45] D.E. Knuth. *The Art of Computer Programming*, volume 2, Seminumerical Algorithms. Addison-Wesley, third edition, 1997.

[46] J. Ko, A.J. Kurdila, and O.K. Rediniotis. Divergence-free bases and multiresolution methods for reduced-order flow modeling. *AIAA Journal*, 38(12) :2219–2232, 2000.

[47] R.H. Kraichnan. Inertial-range transfer in two- and three-dimensional turbulence. *J. Fluid Mech.*, 47 :525–535, 1971.

[48] R.H. Kraichnan. Eddy viscosity in two and three dimensions. *J. Atmos. Sci.*, 33 :1521–1536, 1976.

[49] K. Kunish and S. Volkwein. Galerkin proper orthogonal decomposition for parabolic problems. *Numerische Mathematik*, 90 :117–148, 2001.

[50] F. Kwasniok. Low-dimensional models of the Ginzburg-Landau équation. *SIAM J. Appl. Math.*, 61(6) :2063–2079, 2001.

[51] O. Labbé, P. Sagaut, and E. Montreuil. Large-Eddy Simulation of heat transfer over a backward-facing step. *J. Numer. Heat Transfer* A, 42 :73–90, 2002.

[52] L. Landau and E. Lifschitz. *Mécanique des fluides.* MIR Moscou, 1953.

[53] G. Lassaux and K. Willcox. Model reduction for active control design using multiple-point Arnoldi methods. In *41st Aerospace Sciences Meeting and Exhibit*, 2003.

[54] E. Leclerc. *Contrôle sub-optimal pour les écoulements instationnaires turbulents.* PhD thesis, Université Pierre et Marie Curie, 2003.

[55] E. Leclerc, P. Sagaut, and B. Mohammadi. On the use of incomplete sensitivities for feedback control of compressible flows. *Computers & Fluids.* To appear.

[56] J.L. Lions. *Quelques méthodes de résolution des problèmes aux limites nonlinéaires.* Dunod, 1968.

[57] E.N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2) :130–148, 1963.

[58] D.J. Lucia and P.S. Beran. Aeroelastic system development using proper orthogonal decomposition and Volterra theory. AIAA Paper 2003-1922, 2003.

[59] D.J. Lucia and P.S. Beran. Projection methods for reduced order models of compressible flows. *J. Comput. Phys.*, 188 :252–280, 2003.

[60] D.J. Lucia and P.S. Beran. Reduced-order model development using proper orthogonal decomposition and Volterra theory. *AIAA Journal*, 42(6) :1181–1190, 2004.

[61] D.J. Lucia, P.I. King, P.S. Beran, and M.E. Oxley. Reduced order modeling for a one-dimensional nozzle flow with moving shocks. In *15th AIAA Computational Fluid Dynamics Conference*, 2001. AIAA Paper 2001-2602.

[62] J.L. Lumley. The structure of inhomogeneous turbulent flows. *In A.M. Yaglom & V.I. Tatarski, editors, Atmospheric Turbulence and Radio Wave Propagation, Nauka, Moscow*, pages 166–178, 1967.

[63] I. Mary, P. Sagaut, and M. Deville. An algorithm for unsteady viscous flows at all speeds. *Int. J. Numer. Meth. Fluids*, 34 :371–401, 2000.

[64] B. Mohammadi and O. Pironneau. *Applied Shape Optimization for Fluids.* Oxford University Press, 2002.

[65] E. Montreuil. *Simulations Numériques pour l'Aérothermique avec des Modèles Sous-Maille.* PhD thesis, Université Pierre et Marie Curie, 2000.

[66] B.R. Noack, K. Afanasiev, M. Morzynski, G. Tadmor, and F. Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. Fluid Mech.*, 497 :335–363, 2003.

[67] B.R. Noack, G. Tadmor, M. Morzyński, and S. Siegel. Low-dimensional models for feedback flow control. In *2nd AIAA Flow Control Conference*, 2004. AIAA Paper 2004-2408.

[68] H.M. Park and M.W. Lee. Control of Navier-Stokes equations by means of mode reduction. *Int. J. Numer. Meth. Fluids*, 33 :535–557, 2000.

[69] O. Pironneau. *Méthodes des éléments finis pour les fluides.* Masson, 1988.

[70] B. Podvin. On the adequacy of the ten-dimensional model for the wall layer. *Phys. Fluids*, 13 :210–224, 2001.

[71] B. Podvin and J.L. Lumley. A low-dimensional approach for the minimal flow unit. *J. Fluid Mech.*, 362 :121–155, 1998.

[72] A. Quarteroni, R. Sacco, and F. Saleri. *Méthodes numériques pour le calcul scientifique.* Springer-Verlag, Paris, 1986.

[73] M. Rajaee, S.F.K. Karlsson, and L. Sirovich. Low-dimensional description of free-shear-flow coherent structures and their dynamical behavior. *J. Fluid Mech.*, 258 :1–29, 1994.

[74] S.S. Ravindran. Reduced-order adaptive controllers for fluid flows using POD. *J. Sci. Comput.*, 15(4) :457–478, 2000.

[75] S.S. Ravindran. A reduced-order approach for optimal control of fluids using proper orthogonal decomposition. *Int. J. Numer. Meth. Fluids*, 34 :425–448, 2000.

[76] D. Rempfer. Investigations of boundary-layer transition via Galerkin projections on empirical eigenfunctions. *Phys. Fluids*, 8 :175–188, 1996.

[77] D. Rempfer. On low-dimensional Galerkin models for fluid flow. *Theoret. Comput. Fluid Dynamics*, 14 :75–88, 2000.

[78] D. Rempfer and H.F. Fasel. Dynamics of three-dimensional coherent structures in a flat-plate boundary layer. *J. Fluid Mech.*, 275 :257–283, 1994.

[79] D. Rempfer and H.F. Fasel. Evolution of three-dimensional coherent structures in a flat-plate boundary layer. *J. Fluid Mech.*, 260 :351–375, 1994.

[80] O.E. Rössler. An equation for continuous chaos. *Physics Letters*, 57A(5) :397–398, 1976.

[81] C.W. Rowley. *Modeling, simulation and control of cavity flow oscillations.* PhD thesis, California Institute of Technology, 2002.

[82] C.W. Rowley, T. Colonius, and R.M. Murray. Dynamical models for control of cavity oscillations. In *7th AIAA/CEAS Aeroacoustics Conference*, 2001. AIAA Paper 2001-2126.

[83] C.W. Rowley, T. Colonius, and R.M. Murray. Model reduction for compressible flows using POD and Galerkin projection. *Physica D*, 189(1-2) :115–129, 2004.

[84] P. Sagaut. *Large Eddy Simulations for Incompressible flows.* B. Springer, second edition, 2002.

[85] R.L. Sani and P.M. Gresho. Resume and remarks on the open boundary condition mini-symposium. *Int. J. Numer. Meth. Fluids*, 18 :983–1008, 1994.

[86] O. Schilling and Y. Zhou. Analysis of spectral eddy viscosity and backscatter in incompressible, isotropic turbulence using statistical closure theory. *Phys. Fluids*, 14(3) :1244–1258, 2002.

[87] S. Sirisup and G.E. Karniadakis. A spectral viscosity method for correcting the long-term behavior of POD models. *J. Comput. Phys.*, 194 :92–116, 2004.

[88] L. Sirovich. Turbulence and the dynamics of coherent strutures. *Quart. Appl. Math.*, XLV(3) :561–590, 1987.

[89] N. Smaoui. A model for the unstable manifold of the bursting behavior in the 2D Navier-Stokes flow. *SIAM J. Sci. Comput.*, 23(3) :824–840, 2001.

[90] D. Tang, D. Kholodar, J.-N. Juang, and E.H. Dowell. System identification and pod method applied to unsteady aerodynamics. NASA/TM-2001-211243, Hampton, Virginia, 2001.

[91] R. Temam. *Navier-Stokes equations and nonlinear functional analysis.* SIAM, Philadelphia, 1995.

[92] P.L. Toint. Global convergence of a class of trust-region methods for nonconvex minimization in Hilbert space. *IMA J; Numer. Anal.*, 8(2) :231–252, 1988.

[93] L. Ukeiley, L. Cordier, R. Manceau, J. Delville, M. Glauser, and J.P. Bonnet. Examination of large-scale structures in a turbulent plane mixing layer. part 2. Dynamical systems model. *J. Fluid Mech.*, 441 :67–108, 2001.

[94] G. Vigo. *Méthodes de décomposition orthogonale aux valeurs propres appliquées aux écoulements instationnaires compressibles complexes.* PhD thesis, Paris IX Dauphine, 2000.

[95] G.A. Webber, R.A. Handler, and L. Sirovich. The Karhunen-Loève decomposition of minimal channel flow. *Phys. Fluids*, 9 :1054–1066, 1997.

[96] G.A. Webber, R.A. Handler, and L. Sirovich. Energy dynamics in a turbulent channel flow using the Karhunen-Loève approach. *Int. J. Numer. Meth. Fluids*, 40 :1381–1400, 2002.

[97] P.K. Yeung, J.G. Brasseur, and Q. Wang. Dynamics of direct large-small scale couplings in coherently forced turbulence : concurrent physical and Fourier-space views. *J. Fluid Mech.*, 283 :43–95, 1995.

[98] Y. Zhou and G. Vahala. Reformulation of recursive-renormalization-group based subgrid modeling of turbulence. *Phys. Rev.*, E 47 :2053, 1993.

Résumé. Ce mémoire présente une étude de la modélisation POD-Galerkine réduite dans l'objectif du développement d'algorithmes robustes et efficaces de contrôle actif d'écoulements. Les deux premiers chapitres sont consacrés à la définition et au calcul numérique d'une base de modes POD et à la construction formelle de modèles dynamiques réduits par la méthode de Galerkine à partir des équations de Navier-Stokes. Le cas des écoulements incompressibles est traité en détail. Une étude de la modélisation d'un écoulement tridimensionnel turbulent, pour lequel le problème de la modélisation des petites échelles se pose, est ensuite menée : les interactions entre modes POD et les effets de la réduction de la base modale sont analysés qualitativement et quantitativement, à l'aide notamment d'une paramétrisation visqueuse. Dans le chapitre suivant, des méthodes de calibration, qui reposent sur la résolution d'un problème d'optimisation, sont développées afin de pouvoir calculer automatiquement des modèles réduits fiables pour un coût informatique raisonnable. Enfin, le dernier chapitre est consacré au problème du contrôle d'écoulements par des actionneurs de soufflage ou d'aspiration : les stratégies d'exploitation de la modélisation POD-Galerkine pour le contrôle sont abordées, puis des investigations numériques sont présentées.

Mots clés : méthode POD-Galerkine, modèle réduit, équations de Navier-Stokes, écoulement tridimensionnel turbulent, calibration, contrôle actif.

Discipline : Mathématiques Appliquées.

Reduced-order POD-Galerkin modelling for the control of unsteady flows

Abstract. This Ph. D. thesis presents a study of the reduced-order POD-Galerkin modelling with the objective to develop robust and efficient algorithms for the flow control. The first and second chapters deal with the definition and the computation of POD modal basis, then with the formal construction of reduced-order dynamical models from the Navier-Stokes equations. The case of incompressible flows is presented in detail. In the third chapter, a study of the modelling of a three-dimensional turbulent flow, for which the small-scale modelling problem arises, is carried out : the interactions between POD modes and the effects of the POD basis reduction are qualitatively and quantitatively analysed, by a viscous parameterisation in particular. In the fourth chapter, some calibration methods, whose principle is to solve a minimization problem, are designed to be able to construct automatically some reliable reduced-order models for a reasonable computational cost. They are assessed for two flow configurations. In the fifth chapter, the problem of flow control by blowing/sucking actuators is considered : some strategies of the use of the POD-Galerkin modelling are described, then some numerical investigations are presented.

Key words : POD-Galerkin method, reduced-order model, Navier-Stokes equations, three-dimensional turbulent flow, calibration, active control.

Département de Simulation Numérique des écoulements et Aéroacoustique, Office National d'études et de Recherches Aérospatiales, 29 avenue de la Division Leclerc, F92322 Châtillon.