INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES TOULOUSE

CentraleSupélec

MATHEMATIQUES ORSAY

Rapport de fin d'étude

Réduction de modèle non linéaire dans l'espace de Wasserstein et méta-modélisation pour certaines lois de conservation unidimensionnelles

Période de déroulement du stage : Avril-Septembre 2020

Iain HENDERSON, promotion 2020

Département de Mathématiques d'Orsay, Master AMS-AM

CentraleSupélec, Option Mathématiques Appliquées, filière FMR

Tuteur Option : Pauline Lafitte

Tuteur Filière : Sorin Olaru

Sous la direction de

Pascal Noble

Olivier Roustant

INSA Toulouse 135 Avenue de Rangueil, 31400 Toulouse

Contact : iain.pl.henderson@gmail.com

1 Remerciements

Je tiens tout d'abord à remercier mes deux encadrants, Pascal Noble et Olivier Roustant. Ce stage constitue les prémices d'une thèse de doctorat sous leur direction à l'INSA de Toulouse. Je les remercie profondément de la confiance qu'ils me témoignent en m'accueillant parmi eux dans cette aventure ! Malgré les conditions exceptionnelles dues au Covid, ils se sont constamment rendus disponibles pour m'accompagner tout au long de ce stage, et se sont toujours montrés patients face à mes (nombreuses) erreurs et égarements mathématiques. Je me sais être entre de bonnes mains pour les années qui arrivent. Aussi, je remercie tout le département du GMM de l'INSA qui a su m'accueillir très chaleureusement lors de mon arrivée.

Je tiens ensuite à remercier Mme Pauline Lafitte, Professeur à CentraleSupélec. Alors que j'étais en plein questionnements, elle a su prendre tout le temps nécessaire pour écouter mes questions et m'aiguiller afin que je trouve ma voie pendant mes études à CentraleSupélec ; c'est elle qui m'a dirigé vers Pascal alors que j'étais en recherche de doctorat. Enfin, elle a su me redonner le goût des mathématiques lors de ma scolarité, et m'avait déjà aidé à obtenir un stage de recherche à la faculté d'Edimbourg alors que je cherchais à faire l'expérience de ce domaine.

Je tiens à remercier le corps enseignant de la faculté d'Orsay, en particulier du M1 de Mathématiques Fondamentales et du Master AMS, qui a su me présenter de très belles mathématiques de façon passionnante et très pédagogique, contribuant ainsi massivement à ma culture mathématique. Je remercie aussi le corps enseignant de CentraleSupélec, qui a m'a permis d'élargir ma culture scientifique à des domaines très complémentaires des mathématiques.

Évidemment, rien de cela n'aurait été possible sans le soutien sans failles des membres de ma famille. Cela fait des années déjà qu'ils me soutiennent dans mon parcours, tant académique que dans la vie, dans les hauts comme dans les bas, et je sais qu'ils sont là pour moi comme ils l'ont été depuis des années. Enfin, je te remercie Nathalie, pour tout le bonheur et le soutien que tu m'apportes chaque jour. Il m'arrive encore si souvent d'avoir du mal à croire la chance que j'ai de regarder vers le futur avec toi.

2 Résumé

La réduction de modèle vise à approcher un modèle complexe et coûteux à résoudre numériquement à l'aide d'un modèle plus simple et surtout plus rapide à résoudre. De tels modèles prennent souvent la forme d'équations aux dérivées partielles (EDP) à résoudre. Dans ce mémoire, nous nous intéressons à deux techniques de réduction de modèle : la méthode des bases réduites et la méta-modélisation.

La première méthode vise, dans sa formulation classique, à approximer l'espace des solutions du modèle par un espace vectoriel V_n de dimension finie construit à l'aide d'une (petite) base de données de solutions du modèle. Nous nous intéressons ici à une extension récente de cette méthode, introduite en 2019 dans un article de V.Ehrlacher, D.Lombardi, O.Mula et F.-X.Vialard [12]. Cette formulation tire parti de la structure des solutions de certaines lois de conservation et propose une alternative à la structure linéaire de l'espace d'approximation V_n, parfois inadaptée au modèle en question. La deuxième méthode est issue des sciences des données, et consiste à apprendre le modèle de façon statistique (machine learning) à l'aide d'une base de données de solutions. Lorsque l'on connaît un modèle plus simple approximant déjà le modèle en question, la méta-modélisation peut être mise en oeuvre en tant que support afin d'apporter une correction au modèle simplifié. On construit ainsi ce que l'on appelle un "complément de modèle".

Dans une première partie, nous présentons la méthode des bases réduites dans sa version classique et introduisons la largeur de Kolmogorov, une quantité théorique qui permet d'évaluer la qualité de la meilleure approximation linéaire de l'espace des solutions.

Dans une deuxième partie, à la suite de [12], nous interprétons les solutions de lois de conservation comme vivant naturellement dans un espace (non linéaire) de mesures de probabilité muni d'une métrique issue du transport optimal, l'espace de Wasserstein. Nous présentons ensuite une méthode de bases réduites adaptée à cet espace. Cette méthode est construite à partir de transformations intrinsèquement non linéaires, et est accompagnée d'une notion de largeur de Kolmogorov non-linéaire.

Dans une troisième partie, nous présentons des estimations de largeurs de Kolmogorov linéaires et non-linéaires issues de [12] sur deux exemples, l'advection linéaire et l'équation de Burgers. Ces exemples montrent la pertinence de la réduction de modèle non linéaire pour ces lois de conservation.

Dans une dernière partie, nous décrivons la mise en oeuvre d'une méta-modélisation en tant que complément de modèle afin d'approcher les solutions de l'équation de Korteweg-de Vries à partir de solutions de l'équation de Burgers. Cette méta-modélisation prend la forme d'une modélisation par processus gaussiens, et nécessite au préalable l'utilisation de transformées en ondelettes discrètes ainsi qu'une Analyse en Composantes Principales (ACP). Cela permet de manipuler des représentations de solutions suffisamment compactes pour rendre utilisable la modélisation par processus gaussiens.

Table des matières

3 Introduction

Dans les sciences et l'ingénierie, il est devenu extrêmement courant de faire appel à des codes informatiques calculant des résultats de modèles complexes, dont l'exécution peut durer plusieurs heures voire plusieurs jours. La réduction de modèle, qui vise à donner une approximation raisonnable du résultat de tels modèles par le biais de calculs bien moins coûteux, est donc actuellement d'un intérêt évident. Nous nous intéressons ici à deux formes particulières de réduction de modèle, la méta-modélisation et la méthode dite des "bases réduites", qui s'applique lorsque le modèle est une équation aux dérivées partielles paramétrée (EDPP).

La méthode des bases réduites, dans sa version classique, consiste à s'appuyer sur une petite base de donnéee de résultats numériques du modèle pour construire une approximation linéaire de basse dimension de l'espace des solutions. Elle peut donc être comprise comme une méthode numérique de résolution d'équations aux dérivées partielles. Cette méthode a déjà fait ses preuves sur une classe de problèmes, notamment lorsque le l'équation à résoudre est linéaire elliptique [7]. En effet, pour ce type de problème, il existe des techniques efficaces qui permettent de contrôler l'erreur d'approximation induite par la méthode des bases réduites ([7], Chap.4).

Cependant, lorsque l'équation n'est plus linéaire ou lorsque les solutions sont peu régulières, il devient plus difficile de mettre en oeuvre une méthode des bases réduites efficace, et la structure linéaire de l'espace d'approximation doit être remise en question. Ce type de difficulté se rencontre très souvent pour les équations issues de la physique, comme en mécanique des fluides. Un cas extrême est celui des systèmes hyperboliques : de telles équations n'ont pas d'effet régularisant, sont quasi systématiquement non linéaires et peuvent présenter l'apparition de discontinuités ("chocs") en temps fini. Néanmoins, les solutions de nombre d'équations aux dérivées partielles non-linéaires (telles que les systèmes hyperboliques) obéissent à des lois de conservations de quantités positives : masse, énergie... Par conséquent, l'espace des mesures de probabilité semble un espace naturel dans lequel chercher les solutions de l'équation considérée. De plus, cet espace ne présente pas de structure d'espace vectoriel : il permet donc de s'affranchir de la structure linéaire parfois inadaptée de la méthode classique des bases réduites. La théorie du transport optimal, qui connaît un renouveau constant depuis le milieu des années 50, a mis au grand jour la structure riche d'espace métrique dont jouit cet espace. En particulier, cette structure rend possible la formulation d'une méthode des bases réduites non-linéaire à l'aide de transformations inspirées de la géométrie riemanienne.

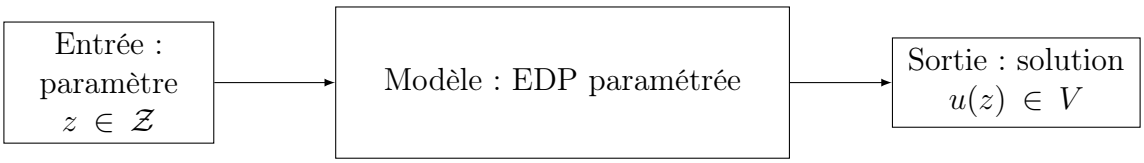
Parallèlement, la méta-modélisation constitue une alternative à la méthode des bases réduites. Dans un contexte industriel où l'analyse des données prend un rôle prépondérant, elle est de plus en plus utilisée. Elle rend possible l'étude statistique du comportement des sorties d'un modèle coûteux, comme son analyse de sensibilité par exemple [6]. Dans le cas où le modèle à approcher est déterministe et particulièrement coûteux à résoudre, une méthode s'est imposée : la modélisation par processus gaussiens. Elle fournit des outils de modélisation flexibles tout en pourvoyant des formules de prédiction explicites. Cette modélisation vise à représenter les sorties du modèle par un processus gaussien indexé par l'espace des paramètres. La loi du processus peut ensuite être conditionnée selon les points d'observations fournis par la base de données. Une conséquence est que le processus gaussien ainsi obtenu interpole le modèle aux points d'observation, exploitant au maximum les informations fournies par la base de donnée. De plus, la prédiction à l'aide d'un tel modèle est accompagnée d'un calcul d'écart-type, fournissant un intervalle de confiance autour de la valeur prédite.

Chapitre 1

Introduction à la méthode des bases réduites

1 Cadre classique de la méthode des bases réduites

Nous cherchons à approcher un modèle qui prend la forme générale suivante :



On a ainsi

- L'espace d'arrivée V dans lequel les solutions $u(z)$ se trouvent. Il est à comprendre comme un espace de fonctions : par exemple, $V = L^2(\Omega)$ où Ω un ouvert de $\mathbb{R}^d$; ou plus généralement un espace de Banach. Dans les faits, comme on travaillera toujours avec un modèle discrétisé, on aura plutôt $V = \mathbb{R}^N$ avec $N \gg 1$.
- L'espace des paramètres : $\mathcal{Z}$, typiquement une sous-partie de $\mathbb{R}^q$ si le modèle prend q paramètres en entrée. En général, si l'EDP est une équation d'évolution, les instants où la solution est observée sont aussi comptabilisés dans l'espace des paramètres.
- Le modèle de type EDP paramétrée, qui visera à résoudre un problème de la forme

$$\mathcal{P}(u(z), z) = 0 \tag{1.1}$$

  où $\mathcal{P}$ est un opérateur différentiel paramétré (au sens large) par z ; le paramétrage peut intervenir dans les coefficients de l'EDP, dans la condition initiale, etc... Ce n'est donc rien d'autre qu'un certain solveur d'EDP.

On introduit aussi l'espace des solutions

$$\mathcal{M} := \{u(z), z \in \mathcal{Z}\} \subset V$$

La méthode des bases réduites (classique) vise, pour une valeur d'entrée z^*, à approcher la sortie théorique du modèle $u(z^*)$ par une approximation cherchée dans un sous espace vectoriel V_n de V de basse dimension, obtenu à partir d'une base de données de sorties déjà calculées, aussi communément appelées "snapshots", $\{u(z_1), ..., u(z_n)\}$:

$$V_n = \text{Span}(u(z_1), ..., u(z_n))$$

Cela revient en quelque sorte à chercher la meilleure approximation linéaire de $\mathcal{M}$ à partir de points échantillonnés de $\mathcal{M}$. Selon la nature du problème (1.1), cette approximation peut être très bien adaptée, par exemple dans certains cas d'EDPs linéaires elliptiques ou paraboliques [7].

La construction d'un espace V_n adapté constitue la difficulté majeure de cette méthode. Dans la section suivante, on introduit la largeur de Kolmogorov, une quantité théorique qui offre un éclairage sur le "meilleur" espace d'approximation V_n possible. Ensuite, nous présentons des algorithmes qui permettent la construction d'un tel espace V_n.

2 Largeur de Kolmogorov

On évalue la qualité de l'approximation de $\mathcal{M}$ offerte par un sous espace $V_n \subset V$ par l'erreur du pire cas ("worst case error")

$$e_{wc}(\mathcal{M}, V, V_n) := \sup_{z \in \mathcal{Z}} \inf_{w \in V_n} ||u(z) - w||_V \tag{1.2}$$

et si $\mathcal{Z}$ est muni d'une mesure dz, par l'erreur moyenne quadratique

$$e_{av}(\mathcal{M}, V, V_n) := \left(\int_{\mathcal{Z}} \inf_{w \in V_n} ||u(z) - w||_V^2 dz \right)^{1/2} \tag{1.3}$$

Ces notions mènent aux définitions suivantes :

Définition 1.1 (Largeur de Kolmogorov)**.** *On définit la largeur de Kolmogorov L^∞ par*

$$d_n(\mathcal{M}, V) := \inf_{\substack{V_n \subset V \\ \dim V_n = n}} e_{wc}(\mathcal{M}, V, V_n) \tag{1.4}$$

et la largeur de Kolmogorov L^2 par

$$\delta_n(\mathcal{M}, V) := \inf_{\substack{V_n \subset V \\ \dim V_n = n}} e_{av}(\mathcal{M}, V, V_n) \tag{1.5}$$

La largeur de Kolmogorov est un nombre décrivant à quel point $\mathcal{M}$ peut être correctement approximé par un sous espace vectoriel de dimension finie. Elle est inhérente au système étudié et ne dépend ainsi d'aucune méthode numérique. Elle donne une borne inférieure théorique sur la qualité de la meilleure approximation linéaire possible de $\mathcal{M}$: si elle décroît assez vite en fonction de n, on peut espérer pouvoir approcher numériquement les solutions de notre problème par une combinaison linéaire d'un petit nombre de snapshots. Dans le cas contraire, si la largeur de Kolmogorov décroît trop lentement, on sera assuré qu'une telle approximation linéaire nécessitera un grand nombre de snapshots pour être satisfaisante.

Remarque : Le calcul d'un espace V_n atteignant la borne inférieure dans (1.4) ou (1.5) n'est pas raisonnable en général. Diverses techniques existent pour construire un sous-espace V_n satisfaisant, les deux principales sont décrites ci-dessous.

3 Construction d'une base réduite

La sélection d'une base réduite se fait usuellement à l'aide d'un algorithme glouton ou d'une Décomposition Orthogonale aux Valeurs Propres ("Proper Orthogonal Decomposition" en anglais, abrégé "POD") aussi connue sous le nom d'Analyse en Composantes Principales (ACP) selon le contexte.

3.1 Décomposition Orthogonale aux Valeurs Propres

On suppose disposer d'une base de donnée de N snapshots $B := \{u_1, ..., u_N\} \subset V$ avec $u_i := u(z_i)$ et N est assez grand. Pour simplifier, on se place dans le vas où $V = L^2(\Omega)$, muni de son produit scalaire usuel et on suppose que B est une famille libre. Soit $n \in \mathbb{N}$, $n \leq N$, on cherche à construire l'espace $V_n \subset V$ de dimension n minimisant la quantité suivante

$$\frac{1}{N}\sum_{i=1}^{N} ||u_i - P_{V_n} u_i||^2_{L^2(\Omega)} \tag{1.6}$$

où P_{V_n} désigne la projection orthogonale sur V_n.

Pour cela, on construit dans un premier temps l'opérateur de corrélation de B. Notons $V_B := \text{Span}(B)$, cet opérateur est défini par

$$\begin{aligned} C : V_B &\longrightarrow V_B \\ f &\longmapsto \frac{1}{N}\sum_{i=1}^{N} \langle f, u_i \rangle u_i \end{aligned}$$

On vérifie facilement qu'il est autoadjoint positif. V_B étant de dimension finie, il est donc diagonalisable dans une base orthonormée et ses valeurs propres sont positives ; on les ordonne selon $\lambda_1 \geq \lambda_2 ... \geq \lambda_N \geq 0$ avec leurs vecteurs propres associés $v_1, ..., v_N$, appelées "composantes principales". Notons que cela revient exactement à diagonaliser la matrice $\tilde{C}$ de taille $N \times N$ dont les coefficients sont donnés par $c_{ij} = \frac{1}{N}\langle u_i, u_j \rangle$. De façon informelle, les n premières composantes principales s'interprètent comme les n directions de $\mathbb{R}^N$ permettant d'expliquer le plus d'information dans les données, selon le critère donné par (1.6).

On prouve que l'espace de dimension n minimisant la quantité (1.6) est engendré par les n premières composantes principales. De plus, la quantité (1.6) associée vaut

$$\frac{1}{N}\sum_{i=1}^{N} ||u_i - P_{V_n} u_i||^2_{L^2(\Omega)} = \sum_{i=n+1}^{N} \lambda_i$$

Ainsi, on choisit comme espace de base réduite $V_n := \text{Span}\{v_1, ..., v_n\}$.

La POD présente de inconvénients majeurs :

— Elle suppose d'avoir à disposition une large base de données de snapshots pour que l'espace d'approximation obtenu soit pertinent. En particulier, si le coût en calcul de chaque solution u_i est trop important, cette approche n'est pas satisfaisante.

— Lorsque l'on rajoute un élément u_{N+1} à la base de donnée $\{u_1, ..., u_N\}$, la procédure de diagonalisation doit être complètement réitérée.

L'algorithme glouton offre une alternative à ces deux facteurs limitants.

3.2 Algorithme glouton

L'algorithme glouton vise à construire un espace de base réduite de façon récursive afin d'éviter le plus possible d'obtenir de l'information redondante lors de l'ajout d'une solution à la base de donnée.

Soit V_n l'espace de base réduite à l'étape n. Un préalable important est d'avoir à disposition une estimation de la forme

$$||u(z) - u_n(z)|| \leq \eta_n(z) \quad \forall z \in \mathcal{Z} \tag{1.7}$$

où $u_n(z) \in V_n$ est la meilleure approximation de $u(z)$ dans V_n ; $u_n(z)$ est calculée d'une façon spécifique à l'EDP étudiée, souvent à partir d'une formulation variationnelle, comme c'est le cas pour un problème linéaire elliptique [7].

On sélectionne z_{n+1} selon

$$z_{n+1} = \arg\max_{z \in \mathcal{Z}} \eta_n(z)$$

et on ajoute ainsi $u(z_{n+1})$ à la base. Ainsi, l'erreur décroît nécessairement au cours de la construction et on est sûr d'enrichir la base au fur et à mesure.

Il existe des méthodes qui permettent d'obtenir de telles fonctions d'estimation η_n, que nous n'évoquerons pas ici ; une introduction assez générale à de telles méthodes se trouve dans [7], Chapitres 4 et 5.

Chapitre 2

Réduction de modèle par plongement dans l'espace de Wasserstein

Pour certains problèmes de bases réduites, l'utilisation de sous espaces vectoriels comme espaces d'approximation est tout à fait naturelle et adaptée. C'est le cas pour certaines EDPs elliptiques et paraboliques. Cependant, si le modèle étudié est non-linéaire comme c'est le cas de nombreuses lois de conservations, cette approche n'est a priori plus pertinente. D'autres espaces (non linéaires) deviennent plus naturels, comme c'est le cas de l'espace des mesures de probabilité sur une partie de $\mathbb{R}^d$: un tel espace paraît en effet adapté pour rendre compte de certaines lois de conservations. En y ajoutant une métrique issue du transport optimal, on obtient l'espace de Wasserstein.

1 Introduction à l'espace de Wasserstein

1.1 Eléments de transport optimal pour l'espace de Wasserstein

Dans cette section, nous introduisons succintement des notions de transport optimal afin de pouvoir définir l'espace de Wasserstein. Cette section est directement inspirée des notes de cours de L.Chizat et L.Nenna de la faculté d'Orsay ; pour plus de détails, on pourra par exemple consulter [5].
La formulation classique du transport optimal se fait à l'aide de mesures de probabilités, que l'on souhaite "transporter" de l'une vers l'autre en minimisant une fonctionnelle représentant le coût du transport.

Définition 2.1 (Mesure push-forward)**.** *Soit (X, $\mathcal{E}$) et (Y, $\mathcal{F}$) deux espaces mesurables et $T : X \longrightarrow Y$ une application mesurable. Soit μ une mesure de probabilité sur X. On définit la mesure "push-forward" de μ par T, notée $T_{\#}\mu$, par*

$$\forall A \in \mathcal{F}, \quad T_{\#}\mu(A) := \mu(T^{-1}(A))$$

C'est la mesure image de T vue comme une variable aléatoire.

Définition 2.2 (Plan de transport)**.** *Soit (X, $\mathcal{E}$) et (Y, $\mathcal{F}$) deux espaces mesurables, μ une mesure de probabilité sur X et ν une mesure de probabilité sur Y. On note π_X la projection canonique $X \times Y \longrightarrow X$, de même pour π_Y. On appelle plan de transport entre μ et ν toute mesure de probabilité γ sur $X \times Y$ telle que*

$$\pi_{X\#}\gamma = \mu \quad \text{et} \quad \pi_{Y\#}\gamma = \nu$$

On note $\Pi(\mu, \nu)$ l'ensemble des plans de transport entre μ et ν.

Remarque : On a toujours $\Pi(\mu,\nu) \neq \emptyset$ car la mesure produit $\mu \otimes \nu$ est dans $\Pi(\mu,\nu)$.

Définition 2.3 (Mesure à moments finis)**.** *Soit (Ω, d) un espace métrique et $x_0 \in \Omega$. On note $\mathcal{P}_2(\Omega)$ l'ensemble des mesures de probabilité sur Ω qui admettent un moment d'ordre 2 :*

$$\mathcal{P}_2(\Omega) := \{\mu \text{ mesure positive } : \mu(\Omega) = 1 \text{ et } \int_\Omega d(x,x_0)^2 \mu(dx) < \infty\}$$

Cette définition est indépendante de $x_0 \in \Omega$.

Remarque : Si Ω est borné alors toute mesure de probabilité sur Ω admet automatiquement tous ses moments finis.

Avec ces définitions, on peut définir une distance sur l'espace $\mathcal{P}_2(\Omega)$ à l'aide d'une formulation issue du transport optimal.

Définition/Théorème 2.1 (Distance de Wasserstein)**.** *Sur l'espace $\mathcal{P}_2(\Omega)$, on définit la métrique de Wasserstein*

$$\forall \mu,\nu \in \mathcal{P}_2(\Omega), \quad W_2(\mu,\nu) := \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int d(x,y)^2 \gamma(dx,dy) \right)^{1/2} \tag{2.1}$$

La métrique de Wasserstein vérifie tous les axiomes d'une distance.

Remarque : Cette métrique L^2 est en fait généralisable en une métrique L^p pour tout $p \in [1, +\infty[$. Quand nous parlerons de distance de Wasserstein, il sera toujours implicite que l'on utilisera le cadre $p = 2$.

La distance de Wasserstein se comprend comme (la racine du) coût minimal, i.e. "optimal", de transport entre la mesure μ et ν, pour le coût "distance au carré". Dans le cas de la dimension1($\Omega \subset \mathbb{R}$) on peut montrer que la borne inférieure dans (2.1) est atteinte en un unique γ appelé *plan de transport optimal*. De plus, ce plan ainsi que la distance de Wasserstein associée peuvent s'exprimer à l'aide des fonctions quantiles.

Définition 2.4 (Fonction de répartition, fonction quantile)**.** *Soit μ une mesure de probabilité sur $\mathbb{R}$. On définit*

- *sa fonction de répartition $F_\mu : \mathbb{R} \ni x \longmapsto \mu(]-\infty, x]) \in [0,1]$*
- *sa fonction quantile $Q_\mu : [0,1] \ni t \longmapsto \inf\{x \in \mathbb{R} | F_\mu(x) \geq t\}$. C'est le pseudo-inverse de F_μ.*

On définit ces mêmes notions pour une mesure de probabilité définie sur $\Omega \subset \mathbb{R}$ en la prolongeant par la mesure nulle sur $\mathbb{R} \setminus \Omega$. On a alors le résultat suivant :

Proposition 2.1. *Si $\Omega \subset \mathbb{R}$, l'unique plan de transport optimal pour le coût "distance au carré" s'exprime comme*

$$\gamma = (Q_\mu, Q_\nu)_\# \lambda_{|[0,1]} \tag{2.2}$$

où $\lambda_{|[0,1]}$ est la mesure de Lebesgue sur $[0,1]$. De plus, la distance de Wasserstein associée est

$$W_2(\mu,\nu) = \left(\int_{[0,1]} \big(Q_\mu(t) - Q_\nu(t)\big)^2 dt \right)^{1/2} = ||Q_\mu - Q_\nu||_{L^2([0,1])} \tag{2.3}$$

1.2 Transformations non-linéaires sur l'espace de Wasserstein : Exponentielle et Logarithme

Notons

$$\mathcal{I} := \{Q_\mu,\ \mu \in \mathcal{P}_2(\Omega)\} \subset L^2([0,1]) \tag{2.4}$$

Alors la proposition (2.1) permet d'affirmer le fait suivant :

Proposition 2.2. *La transformation $Q : \mu \longmapsto Q_\mu$ établit une isométrie d'espaces métriques entre $(\mathcal{P}_2(\Omega), W_2)$ et $(\mathcal{I}, ||\cdot||_{L^2([0,1])})$.*

Remarque : L'inverse de Q est donné par

$$\begin{aligned} Q^{-1} : \mathcal{I} &\longrightarrow \mathcal{P}_2(\Omega) \\ q &\longmapsto \frac{d}{dx}(q^{-1}) \end{aligned}$$

où $q^{-1} : x \mapsto \inf\{t \in [0,1] : q(t) \geq x\}$ désigne l'inverse généralisée de q et où la dérivée est à comprendre au sens des distributions. $\frac{d}{dx}(q^{-1})$ est bien une mesure (de probabilité), étant donné que q^{-1} est croissante donc à variations bornées.

Nous allons maintenant définir une transformation non linéaire sur $(\mathcal{P}_2(\Omega), W_2)$ qui s'apparente à une exponentielle de variété riemannienne. Cette transformation tire parti de l'isométrie précédente.

Notons, pour tout $q_0 \in \mathcal{I}$,

$$\mathcal{K}_{q_0} := \{q \in L^2([0,1]) : q + q_0 \in \mathcal{I}\} \tag{2.5}$$

qui est un ensemble convexe contenant les fonctions constantes égales à $x, \forall x \in \Omega$. Alors, on peut définir les applications suivantes :

Définition 2.5 (Exponentielle et logarithme sur $(\mathcal{P}_2(\Omega), W_2)$, version 1)**.** *Soit $\mu \in \mathcal{P}_2(\Omega)$, au point μ on définit l'exponentielle par*

$$\begin{aligned} \exp_\mu :\ \mathcal{K}_{Q_\mu} &\longrightarrow \mathcal{P}_2(\Omega) \\ q &\longmapsto Q^{-1}(Q_\mu + q) \end{aligned}$$

qui est surjective, et le logarithme par

$$\begin{aligned} \log_\mu :\ \mathcal{P}_2(\Omega) &\longrightarrow \mathcal{K}_{Q_\mu} \\ \nu &\longmapsto Q_\nu - Q_\mu \end{aligned}$$

On tire alors de la proposition (2.2) que

Proposition 2.3. *L'application* $\exp_\mu :\ (\mathcal{K}_{Q_\mu}, ||\cdot||_{L^2([0,1])}) \longrightarrow (\mathcal{P}_2(\Omega), W_2)$ *est une isométrie d'espaces métriques et on a*

$$W_2(\nu_1, \nu_2) = ||\log_\mu(\nu_1) - \log_\mu(\nu_2)||_{L^2([0,1]} \qquad \forall \nu_1, \nu_2 \in \mathcal{P}_2(\Omega) \tag{2.6}$$

Exemples de logarithmes et d'exponentielles sur $(\mathcal{P}_2(\Omega), W_2)$:

— Masses de Dirac : Soit $\Omega = \mathbb{R}$. Soit $x \in \mathbb{R}$, on note δ_x la masse de Dirac en x. Pour tout $s \in (0,1)$, $Q_{\delta_x}(s) = x$, d'où

$$W_2(\delta_{x_1}, \delta_{x_2}) = |x_1 - x_2| \quad \text{et} \quad \log_{\delta_{x_2}}(\delta_{x_1}) = x_1 - x_2$$

— Translation et dilatation : Soit $\mu \in \mathcal{P}_2(\mathbb{R})$ une mesure de probabilité de densité ρ, on définit $\mu_{a,b}$ la mesure de probabilité de densité $\rho_{a,b}(x) = \frac{1}{a}\rho\big(\frac{x-b}{a}\big)$. Alors

$$\forall x \in \mathbb{R}, \quad F_{\mu_{a,b}}(x) = F_\mu(ax+b) \quad \text{et} \quad \forall s \in [0,1], \quad Q_{\mu_{a,b}}(s) = \frac{Q_\mu(s) - b}{a} \tag{2.7}$$

et le logarithme s'écrit

$$\log_\mu(\mu_{a,b}) = \Big(\frac{1}{a} - 1\Big) Q_\mu - \frac{b}{a} \tag{2.8}$$

Remarque : En général, on définit un peu différemment le logarithme et l'exponentielle sur l'espace de Wasserstein. En effet, si Ω est une partie compacte et convexe de $\mathbb{R}^d$, l'espace de Wasserstein devient un espace dit "géodésique" : il est possible de définir des géodésiques sur $(\mathcal{P}_2(\Omega), W_2)$ et on définit l'exponentielle comme la géodésique prise à l'instant $t = 1$. Une façon simplifiée de l'introduire, tirée de [10], équation (2.4), est présentée dans la définition suivante.

Définition 2.6 (Exponentielle et logarithme sur $(\mathcal{P}_2(\Omega), W_2)$, version 2)**.** *Soit $\mu \in \mathcal{P}_2(\Omega)$, au point μ on définit le logarithme par*

$$\begin{aligned} \log_{\mu, ver\ 2} : \ \mathcal{P}_2(\Omega) &\longrightarrow L^2_\mu(\Omega) \\ \nu &\longmapsto Q_\nu \circ F_\mu - id_\Omega \end{aligned}$$

et l'exponentielle par

$$\begin{aligned} \exp_{\mu, ver\ 2} : \ \log_\mu(\mathcal{P}_2(\Omega)) &\longrightarrow \mathcal{P}_2(\Omega) \\ f &\longmapsto (id_\Omega + f)_{\#}\mu \end{aligned}$$

On pourra se référer à [5] (Chap. 2, section 2.3), pour de plus amples détails sur les liens de ces transformations avec une forme de structure riemannienne sur $(\mathcal{P}_2(\Omega), W_2)$.

Remarque : On remarque alors qu'on a simplement

$$\log_\mu(\nu) = \log_{\mu, \text{ver } 2}(\nu) \ \circ \ Q_\mu \ \ \lambda \ - \ p.p \quad \forall \nu \in \mathcal{P}_2(\Omega) \tag{2.9}$$

$$\exp_\mu(q) = \exp_{\mu, \text{ver } 2}(q \circ F_\mu) \qquad \forall q \in \mathcal{K}_{Q_\mu} \tag{2.10}$$

où λ est la mesure de Lebesgue sur $[0,1]$. On montre (2.10) en exprimant par exemple tout membre q de $\mathcal{K}_{Q_\mu}$ en fonction d'un certain $\nu \in \mathcal{P}_2(\Omega)$ et en évaluant les 2 côtés de l'équation. Nous préférons la première version des applications logarithme et exponentielle (définition 2.5) car l'espace d'arrivée du logarithme est indépendant de μ, a contrario de la version de la définition 2.6.

2 Réduction de modèle sur l'espace de Wasserstein

2.1 Interprétation des solutions de lois de conservation comme mesures de probabilité

On considère une EDP admettant une loi de conservation du type "conservation de la masse" :

$$\frac{d}{dt}\int_\Omega u(t,x)dx = 0 \tag{2.11}$$

où u est la solution de l'EDP en question, supposée être une fonction positive et intégrable sur Ω (d'intégrale non nulle...). Alors pour tout t, on interprète la fonction $\tilde{u}(t) : x \mapsto u(t,x)$ comme la mesure de probabilité à densité sur Ω suivante :

$$\tilde{u}(t) \equiv \mu(t) := \frac{u(t,x)}{\int_\Omega u(0,x)dx}dx \tag{2.12}$$

où dx est la mesure de Lebesgue. Nombres d'équations issues de la physique tombent dans cette catégorie, comme certains systèmes hyperboliques de lois de conservation, ou certains modèles de mécanique des fluides.

2.2 Largeur de Kolmogorov non-linéaire

Comme vu dans la section précédente, on peut interpréter les solutions de certaines EDPs comme des mesures de probabilité. Par le biais des applications exponentielle et logarithme introduites dans la proposition (2.5), on peut ainsi se ramener à l'espace vectoriel $L^2([0,1])$ et y calquer les méthodes classiques des bases réduites.

Soit $\mu \in \mathcal{P}_2(\Omega)$: notons

$$\mathcal{T} := \log_\mu(\mathcal{M}) \subset L^2([0,1]) \tag{2.13}$$

On peut maintenant introduire des notions d'erreur qui permettront d'évaluer la qualité de l'approximation non linéaire par rapport à l'approximation linéaire (méthode des bases réduites classique).

Définition 2.7 (Largeur de Kolmogorov sur $(\mathcal{P}_2(\Omega), W_2)$)**.** *Soit $V_n \subset L^2([0,1])$ un sous-espace vectoriel de dimension n.*

— *Dans $(L^2([0,1]), ||\cdot||_2)$, on définit l'erreur du pire cas par*

$$e_{wc}(\mathcal{T}, L^2([0,1]), V_n) := \sup_{f\in\mathcal{T}} ||f - P_{V_n} f||_{L^2([0,1])} \tag{2.14}$$

où $P_{V_n} f$ est la projection orthogonale de f sur V_n pour le produit scalaire canonique de $L^2([0,1])$.

— *Dans $(\mathcal{P}_2(\Omega), W_2)$, on définit l'erreur du pire cas par*

$$e_{wc}(\mathcal{M}, \mathcal{P}_2(\Omega), V_n) := \sup_{u\in\mathcal{M}} W_2(u, \exp_\mu(P_{V_n}\log_\mu(u))) \tag{2.15}$$

pourvu que $P_{V_n}\mathcal{T} \subset \mathcal{K}_{Q_\mu}$, afin que l'exponentielle soit tout le temps bien définie.

Par les propriétés d'isométrie (2.3), on a

$$e_{wc}(\mathcal{T}, L^2([0,1]), V_n) = e_{wc}(\mathcal{M}, \mathcal{P}_2(\Omega), V_n) \tag{2.16}$$

et on définit la largeur de Kolmogorov L^∞ associée

$$d_n(\mathcal{T}, L^2([0,1])) := \inf_{\substack{V_n \subset V \\ \dim V_n = n}} e_{wc}(\mathcal{T}, L^2([0,1]), V_n) \tag{2.17}$$

De façon analogue, on définit les erreurs moyennes $e_{av}(\mathcal{T}, L^2([0,1]), V_n)$ et $e_{av}(\mathcal{M}, \mathcal{P}_2(\Omega), V_n)$ sur $L^2([0,1])$ et $\mathcal{P}_2(\Omega)$; la propriété d'isométrie implique aussi que

$$e_{av}(\mathcal{T}, L^2([0,1]), V_n) = e_{av}(\mathcal{M}, \mathcal{P}_2(\Omega), V_n) \tag{2.18}$$

et on a la largeur de Kolmogorov L^2 associée

$$\delta_n(\mathcal{T}, L^2([0,1])) := \inf_{\substack{V_n \subset V \\ \dim V_n = n}} e_{av}(\mathcal{T}, L^2([0,1]), V_n) \tag{2.19}$$

.

Ainsi, pour comparer les meilleures approximations linéaires et non linéaires, on souhaite comparer $d_n(\mathcal{T}, L^2([0,1]))$ à $d_n(\mathcal{M}, L^2(\Omega))$. Le chapitre suivant vise à fournir des estimations de largeurs de Kolmogorov dans deux cas particuliers : l'équation d'advection linéaire et l'équation de Burgers.

Construction d'une base réduite : les auteurs de [12] proposent deux algorithmes de construction d'une base réduite, qui s'inspirent de ceux exposés dans le chapitre 1. Nous les citons simplement ici, car leur étude sort du sujet de ce mémoire.

- Soit $\mathcal{M}_{tr}$ la base de donnée de sorties du modèle. On effectue une POD de $\mathcal{T}_{tr} = \log_\mu(\mathcal{M}_{tr})$ dans l'espace vectoriel $L^2([0,1])$ et on utilise les n premières composantes principales comme base réduite dans l'espace $L^2([0,1])$. On se ramène ensuite à l'espace de Wasserstein en calculant une exponentielle. Cette méthode possède un désavantage : effectuer un calcul de bases réduites dans $L^2([0,1])$ n'assure pas que le résultat de ce calcul se situe dans K_{Q_μ}. Or, pour retourner dans l'espace $\mathcal{P}_2(\Omega)$, il est nécessaire de composer par l'application $\exp_\mu$, définie uniquement sur K_{Q_μ} ; dans les faits, cela peut résulter en des instabilités numériques et la méthode n'est valide que localement.
- Un algorithme glouton basé sur des calculs de barycentres dans l'espace de Wasserstein. Son avantage principal par rapport à la méthode précédente réside dans le fait que chaque étape de l'algorithme est bien défini, notamment le calcul d'exponentielle.

Chapitre 3

Estimations de largeurs de Kolmogorov pour l'advection linéaire et l'équation de Burgers

Dans cette section sont données des estimations des largeurs de Kolmogorov linéaire et non linéaire pour deux systèmes hyperboliques de loi de conservation : l'advection linéaire et l'équation de Burgers. Ces deux exemples, qui pour certaines conditions initiales possèdent des solutions connues analytiquement, permettent de mener des calculs explicites et justifient l'introduction des changements de variables logarithme-exponentielle présentés dans le chapitre précédent.

1 L'advection linéaire

Étant donnée une vitesse $c \in \mathbb{R}$, l'équation d'advection linéaire unidimensionnelle sur $\mathbb{R}$ se formule de la façon suivante :

$$\begin{cases} \partial_t u + c\partial_x u = 0 & \forall (t,x) \in \mathbb{R}_+ \times \mathbb{R} \\ u(0,x) = u_0(x) & \forall x \in \mathbb{R} \end{cases} \tag{3.1}$$

L'advection linéaire est l'exemple le plus simple de système hyperbolique de loi de conservation. Le test d'une méthode numérique de résolution d'EDP de loi de conservation sur l'advection linéaire est une étape incontournable. Ses solutions sont connues analytiquement si $\Omega = \mathbb{R}$: elles sont données par

$$u(t,x) = u_0(x - ct) \qquad \forall (t,x) \in \mathbb{R}_+ \times \mathbb{R} \tag{3.2}$$

Si u_0 est positive et intégrable, alors $u(t,\cdot)$ l'est aussi pour tout t et on peut l'interpréter comme une mesure de probabilité.

Pour les estimations de largeurs de Kolmogorov, on considère l'équation d'advection sur $\Omega = [-1,1]$ muni de conditions aux bords périodiques, de condition initiale $u_0 = \mathbb{1}_{[-1,0]}$ paramétrée par $c \in \mathcal{Z} = [0,1]$. On observe alors la solution de (3.1) à l'instant $t = 1$: $\mathcal{M} = \{u_c(1,\cdot), c \in [0,1]\}$. Soit $z_0 \in [0,1]$: u_{z_0} s'identifie à la mesure sur Ω de densité $\mathbb{1}_{[z_0-1,z_0]}$. Notons $\mathcal{T} := \log_{u(z_0)}(\mathcal{M})$. On démontre le résultat suivant :

Proposition 3.1. *Il existe une constante $C > 0$ telle que pour tout $n \in \mathbb{N}^*$*

$$d_n(\mathcal{M}, L^2(\Omega)) \geq \delta_n(\mathcal{M}, L^2(\Omega)) \geq Cn^{-1/2} \tag{3.3}$$

et de plus,

$$\forall n \geq 2, \quad d_n(\mathcal{T}, L^2([0,1])) = 0 \tag{3.4}$$

Preuve : On commence par montrer l'égalité (3.4).

Pour tout $z \in [0,1]$, on a $\mathbb{1}_{[z-1,z]}(x) = \mathbb{1}_{[z_0-1,z_0]}(x - z + z_0)$. D'après l'équation (2.7),

$$Q_{u(z)}(s) = Q_{u(z_0)}(s) - z_0 + z \quad \forall s \in [0,1]$$

d'où,

$$\log_{u(z_0)}(u(z))(s) = Q_{u(z)}(s) - Q_{u(z_0)}(s) = z - z_0 \quad \forall s \in [0,1]$$

Donc l'ensemble $\mathcal{T}$ est contenu dans l'espace vectoriel des fonctions constantes sur $[0,1]$, d'où

$$\forall n \geq 2, \quad d_n(\mathcal{T}, L^2([0,1])) = 0 \tag{3.5}$$

On prouve ensuite l'inégalité (3.3). L'inégalité de gauche est immédiate, étant donné que l'on munit $\mathcal{Z} = [0,1]$ de la mesure de Lebesgue : pour tout sous espace vectoriel $V_n \subset L^2(\Omega)$ de dimension finie,

$$\sup_{z \in \mathcal{Z}} ||u(z) - P_{V_n} u(z)||_{L^2(\Omega)} \geq \left(\int_{\mathcal{Z}} ||u(z) - P_{V_n} u(z)||^2_{L^2(\Omega)} dz \right)^{1/2}$$

ce qui donne $d_n(\mathcal{M}, L^2(\Omega)) \geq \delta_n(\mathcal{M}, L^2(\Omega))$.

Pour la deuxième inégalité, une preuve rapide se trouve dans [9]. On décrit ensuite les débuts d'une preuve alternative présentée dans l'article de Erhlacher et al [12], mais dont nous n'avons pas réussi à justifier la dernière étape. Les outils utilisés sont néanmoins intéressants et s'appuient sur une expression intéressante en soi (3.9) que nous prouvons ci-dessous. Nous esquissons ensuite des pistes possibles de résolution.

On exprime δ_n en fonction des valeurs propres de l'opérateur dit de corrélation K défini à l'aide du noyau κ selon

$$\begin{aligned} K : L^2(\Omega) &\longrightarrow L^2(\Omega) \\ v &\longmapsto \left[x \longmapsto \int_\Omega \kappa(x,x') v(x') dx' = \int_{\mathcal{Z}} \langle u(z), v \rangle u(z)(x) dz \right] \\ \kappa(x,x') &:= \int_{\mathcal{Z}} u(z)(x)\, u(z)(x') dz \end{aligned} \tag{3.6}$$

où κ est définie par (3.6). C'est une fonction (mesurable) symétrique : $\kappa(x,x') = \kappa(x',x)$. Les solutions u sont connues : $u(z) = \mathbb{1}_{[z-1,z]}$. κ est donc bornée par 1, et est donc dans $L^2(\Omega \times \Omega)$, Ω étant borné.

Ainsi, K est un opérateur intégral de Fredholm sur $L^2(\Omega)$ et est donc compact. K est aussi autoadjoint positif puisque

$$\langle w, Kv \rangle = \int_\Omega w(x) K v(x) dx = \int_\Omega \int_\Omega w(x) \kappa(x,x') v(x') dx' dx = \langle Kw, v \rangle \tag{3.7}$$

$$\langle v, Kv \rangle = \int_\Omega v(x) \int_{\mathcal{Z}} \langle u(z), v \rangle u(z)(x) dz dx = \int_{\mathcal{Z}} |\langle u(z), v \rangle|^2 dz \geq 0 \tag{3.8}$$

Ainsi, K admet un ensemble discret de valeurs propres positives qui tendent vers 0, que l'on ordonne $\lambda_1 \geq \lambda_2 \geq ... \geq 0$.

Soit $V_n \subset V$ un sous-espace vectoriel de V de dimension n, $(v_1, ..., v_n)$ une base orthonormée de V_n et $\{v_k, k \geq n+1\}$ une base orthonormée de son supplémentaire orthogonal. On montre facilement à l'aide d'un changement de base que

$$\begin{aligned} &\int_{\mathcal{Z}} ||u(z) - P_{V_n} u(z)||^2_{L^2(\Omega)} dz = \sum_{k=n+1}^{+\infty} \int_{\mathcal{Z}} |\langle u(z), v_k \rangle|^2 dz \\ &= \sum_{k=n+1}^{+\infty} \left\langle \int_{\mathcal{Z}} \langle u(z), v_k \rangle u(z)(\cdot) dz, v_k \right\rangle = \sum_{k=n+1}^{+\infty} \langle K v_k, v_k \rangle = \sum_{k=1}^{+\infty} \lambda_k - \sum_{k=1}^{n} \langle K v_k, v_k \rangle \end{aligned}$$

D'après le lemme (3.1) qu'on démontre dans la suite, on a que

$$\delta_n(\mathcal{M}, L^2(\Omega)) = \sqrt{\sum_{k=n+1}^{\infty} \lambda_k} \tag{3.9}$$

On souhaite donc avoir une estimation du comportement asymptotique de ces valeurs propres pour conclure. A cet effet, on invoque le théorème Max-Min de Courant-Fischer pour les opérateurs autoadjoints compacts :

$$\lambda_k = \max_{\substack{V_k \subset L^2(\Omega) \\ \dim V_k = k}} \min_{\substack{v \in V_k \\ ||v||=1}} \langle Kv, v \rangle$$

Nous allons utiliser cette inégalité sur un sous-espace V_k donné par des modes de Fourier de κ bien choisis. Pour cela, on part de la forme intégrale $\kappa(x, x') = \int_{\mathcal{Z}} u_z(x) u_z(x') dz$ et à l'aide du théorème de Fubini, on peut sans trop de difficultés calculer les coefficients de Fourier de κ pour obtenir

$$\kappa(x, x') = \frac{1}{4} + \sum_{k \in \mathbb{Z}} \frac{1}{\pi^2 (2k+1)^2} \times \left(e^{i\pi(2k+1)x} + e^{i\pi(2k+1)x'} + e^{i\pi(2k+1)(x'-x)} \right)$$

En conséquence, si on note $e_j(x) = \frac{1}{\sqrt{2}} e^{i\pi(2j+1)x}$ le $(2j+1)^{ieme}$ mode de Fourier sur $[-1, 1]$ on obtient

$$\langle K e_i, e_j \rangle = \frac{2}{\pi^2 (2j+1)^2} \delta_{ij} \quad \forall i, j$$

avec δ_{ij} le delta de Kronecker. Et donc en choisissant $V_k = \text{Vect}\{e_1, e_2, ..., e_k\}$, l'égalité Max-Min de Courant Fischer donne que $\lambda_k \geq C/k^2$. Cette inégalité combinée à l'équation (3.9) permet de conclure. □

Lemme 3.1 (Proper Orthogonal Decomposition)**.** *Soit K un opérateur compact autoadjoint positif sur $\mathcal{H} := L^2(\Omega)$, avec $\Omega \subset \mathbb{R}^d$. K admet une base orthonormale de vecteurs propres que l'on note $(e_k)_{k \in \mathbb{N}}$; son spectre est discret, positif, avec 0 comme unique point d'accumulation. On peut l'ordonner en une suite décroissante $\lambda_1 \geq \lambda_2 \geq ... \geq 0$. Alors, pour toute base orthonormée $(v_k)_{k \in \mathbb{N}}$ de $L^2(\Omega)$, pour tout $n \in \mathbb{N}$,*

$$\sum_{i=1}^{n} \langle K v_i, v_i \rangle \leq \sum_{i=1}^{n} \lambda_i \tag{3.10}$$

avec égalité si $Span\{v_1, ..., v_n\}$ est engendré par $(e_1, ..., e_n)$.

Preuve : Soit $n \in \mathbb{N}$, soit $(v_1, ..., v_n)$ une famille orthonormée de $\mathcal{H}$. On note A la matrice symétrique réelle dont les coefficients sont donnés par $a_{ij} := \langle v_i, K v_j \rangle$, et Q la forme quadratique associée :

$$Q : \mathbb{R}^n \ni x \longmapsto x^T A x$$

Dans (3.10), on cherche donc à majorer la trace de A. Soit

$$E := \text{Span}\{v_1, ..., v_n\} \cap \text{Span}\{e_1, ..., e_n\}$$
$$F := \text{Span}\{v_1, ..., v_n\} \cap \text{Span}\{e_k, k \geq n+1\}$$

Soit $p := \dim(E) \leq n$, $(f_1, ..., f_p)$ une base orthonormée de E et $(f_{p+1}, ..., f_n)$ une base orthonormée de F : $(f_1, ..., f_n)$ est une base de $V_n := \text{Span}\{v_1, ..., v_n\}$. L'écriture de Q dans la base

$(f_1, ..., f_n)$ se traduit par une transformation de A en $B = P^T AP$ où les coefficients de B sont $b_{ij} = \langle f_i, K, f_j \rangle$ et P est la matrice de changement de base de $(v_1, ...v_n)$ à $(f_1, ..., f_n)$. Comme $(v_1, ..., v_n)$ et $(f_1, ..., f_n)$ sont des bases orthonormées, P est une matrice orthogonale et A et B ont même trace.

On peut donc écrire :

$$\sum_{i=1}^{n} \langle Kv_i, v_i \rangle = \sum_{i=1}^{n} \langle Kf_i, f_i \rangle = \sum_{i=1}^{p} \langle Kf_i, f_i \rangle + \sum_{i=p+1}^{n} \langle Kf_i, f_i \rangle \tag{3.11}$$

Le deuxième terme de (3.11) se majore comme

$$\begin{aligned} \sum_{i=p+1}^{n} \langle Kf_i, f_i \rangle = \sum_{i=p+1}^{n} \sum_{l=n+1}^{\infty} \lambda_l |\langle f_i, e_l \rangle|^2 &\leq \sum_{i=p+1}^{n} \lambda_{n+1} \sum_{l=n+1}^{\infty} |\langle f_i, e_l \rangle|^2 \\ &\leq (n-p)\lambda_{n+1} \leq \sum_{i=p+1}^{n} \lambda_i \end{aligned} \tag{3.12}$$

Le premier terme de (3.11) se majore en résolvant un problème de maximisation sous contrainte.

En effet, posons $G := \text{Span}\{e_1, ..., e_n\}$, et

$$F : \begin{cases} G^p \longrightarrow \mathbb{R} \\ (u_1, ..., u_p) \longmapsto \sum_{i=1}^{p} \langle Ku_i, u_i \rangle = \sum_{i=1}^{p} \sum_{l=1}^{n} \lambda_l u_{i,l}^2 \end{cases}$$

$$X_{ad} := \{(u_1, ..., u_p) \in G^p :\in \forall i, j\{1, ..., p\}, \langle u_i, u_j \rangle = \delta_{ij}\} \subset G^p$$

où $u_{i,l} = \langle u_i, e_l \rangle$ est la l^{ie} coordonnée de u_i dans la base $(e_1, ..., e_n)$.

La fonction F est clairement continue, et le domaine admissible X_{ad} est fermé borné, donc compact car inclus dans G^p où G est de dimension finie.

Ainsi, le problème de maximisation sous contraintes suivant

$$\max F(u_1, ..., u_p) \quad \text{s.c.} \quad (u_1, ...u_p) \in X_{ad}$$

admet un maximum. On l'étudie par le biais du lagrangien suivant :

$$\mathcal{L}(u_1, ..., u_p, m) := \sum_{i=1}^{p} \sum_{l=1}^{n} \lambda_l x_{i,l}^2 - \sum_{i=1}^{n} m_{i,i}(||u_i||^2 - 1) - 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} m_{i,j} \langle u_i, u_j \rangle \tag{3.13}$$

où $m = (m_{i,j})_{1 \leq i,j \leq n}$ désigne l'ensemble des multiplicateurs de Lagrange.

La dérivée partielle de $\mathcal{L}$ par rapport à $u_{i,l}$ est nulle en $(u_1, ..., u_p)$ si il atteint le maximum de F sur X_{ad} :

$$\partial_{u_{i,l}} \mathcal{L}(u, m) = 2u_{i,l}\lambda_l - 2m_{i,i}u_{i,l} - 2 \sum_{\substack{j=1 \\ j \neq i}}^{n} m_{i,j} u_{i,l} = 0 \tag{3.14}$$

En multipliant (3.14) par e_l et en sommant sur l, on obtient

$$\forall i \in \{1, ..., p\}, \quad Ku_i = m_{i,i}u_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} m_{i,j} u_j = \sum_{j=1}^{n} m_{i,j} u_j \tag{3.15}$$

Et donc $U_p := Vect(u_1, ..., u_p)$ est un sous-espace de G stable par K. Les conditions d'optimalité sur les multiplicateurs de Lagrange imposent que $(u_1, ..., u_p)$ est une famille orthonormale ; U_p est donc de dimension p et est donc engendré par p vecteurs propres (orthonormaux) $(e_{l_1}, ..., e_{l_p})$ de K. Mais alors,

$$\sum_{i=1}^{p} \langle Ku_i, u_i \rangle = \sum_{i=1}^{p} \sum_{j=1}^{p} \lambda_{l_j} u_{i,l_j}^2 = \sum_{j=1}^{p} \lambda_{l_j} \sum_{i=1}^{p} u_{i,l_j}^2 \tag{3.16}$$

et comme $(u_1, ..., u_p)$ est aussi une famille orthonormée, $\sum_{i=1}^{p} u_{i,l_j}^2 = \sum_{i=1}^{p} \langle u_i, e_{l_j} \rangle^2 \leq ||e_{l_j}||^2 \leq 1$ et donc, étant donné qu'on a ordonné les valeurs propres $\lambda_1 \geq \lambda_2 \geq ... \geq 0$,

$$\sum_{i=1}^{p} \langle Ku_i, u_i \rangle = \sum_{j=1}^{p} \lambda_{l_j} \sum_{i=1}^{p} u_{i,l_j}^2 \leq \sum_{j=1}^{p} \lambda_{l_j} \leq \sum_{i=1}^{p} \lambda_i \tag{3.17}$$

Ainsi, en revenant à (3.11) et en utilisant (3.12) et (3.17), on obtient

$$\sum_{i=1}^{n} \langle Kv_i, v_i \rangle = \sum_{i=1}^{p} \langle Kf_i, f_i \rangle + \sum_{i=p+1}^{n} \langle Kf_i, f_i \rangle \leq \sum_{i=1}^{p} \lambda_i + \sum_{i=p+1}^{n} \lambda_i = \sum_{i=1}^{n} \lambda_i \tag{3.18}$$

□

2 L'équation de Burgers

On considère ici l'équation de Burgers non visqueuse sur le domaine $(t, x) \in [0, T] \times \Omega = [0, 5] \times [-1, 4]$. Le paramétrage se fait dans la condition initiale u_0, avec $y \in Y = [\frac{1}{2}, 3]$. L'équation aux dérivées partielles paramétrée que l'on considère est alors

$$\partial_t u + \partial_x \left(\frac{u^2}{2} \right) = 0, \quad u_0(x) = \begin{cases} 0 \text{ si } x \leq 0 \\ y \text{ si } 0 < x \leq \frac{1}{y} \\ 0 \text{ si } x > \frac{1}{y} \end{cases} \tag{3.19}$$

Un exemple d'une telle condition initiale est donnée en Figure 3.1.

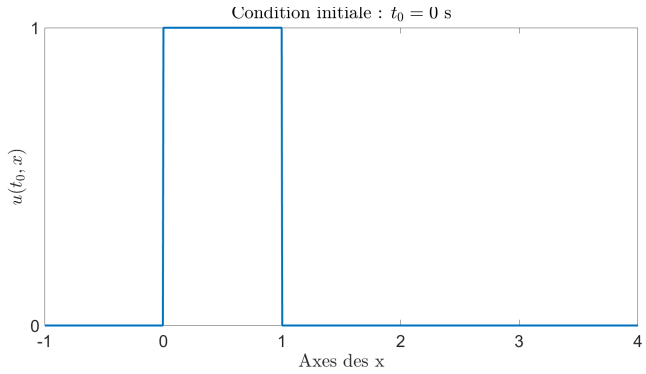


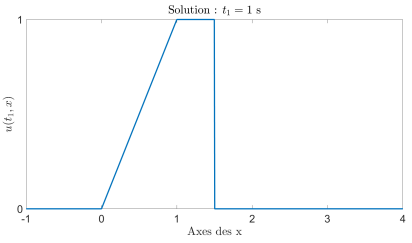
FIGURE 3.1 – Condition initiale paramétrée par $y = 1$

L'ensemble des paramètres est $\mathcal{Z} := Y \times [0, T]$. Muni de conditions aux bords périodiques et d'une telle condition initiale positive d'intégrale 1 (ou strictement positive), l'équation de Burgers admet une loi de conservation "de la masse" et les solutions associées peuvent être interprétées comme des mesures de probabilités.

Le problème (3.19) est hyperbolique, et admet une unique solution entropique connue analytiquement décrite ci-dessous ; des illustrations y sont fournies pour la condition initiale de la Figure 3.1 (i.e. $y = 1$), à $t_1 = 1 < 2/y^2$ et $t_2 = 3 > 2/y^2$.

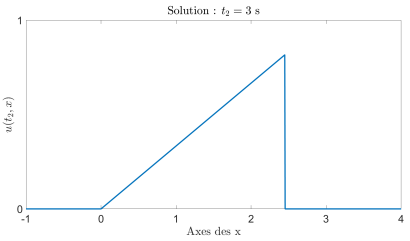
Pour $0 < t < 2/y^2$, une onde de raréfaction se propage à gauche, ainsi qu'une onde de choc à droite :

$$u(t,x) = \begin{cases} 0, & -1 \le x < 0 \\ \frac{x}{t}, & 0 \le x < yt \\ y, & yt \le x < \frac{1}{y} + \frac{yt}{2} \\ 0, & x \ge \frac{1}{y} + \frac{yt}{2} \end{cases}$$



A $t = \frac{y^2}{2}$, l'onde de raréfaction rattrape l'onde de choc et pour $t \ge 2/y^2$, la solution s'écrit

$$u(t,x) = \begin{cases} 0, & -1 \le x < 0 \\ \frac{x}{t}, & 0 \le x < \sqrt{2t} \\ 0, & x \ge \sqrt{2t} \end{cases}$$



Notons $u_{y,t} := u(t,\cdot)dx$ la mesure associée à la solution de (3.19) à l'instant t. On peut calculer analytiquement les fonctions de répartition et les fonctions quantile associées à $u_{y,t}$.

Si $t \le 2/y^2$, les fonctions quantiles sont données par

$$Q_{u_{y,t}}(s) = \begin{cases} -1, & s = 0 \\ \sqrt{2ts} & 0 < s < \frac{y^2 t}{2} \\ \frac{1}{y}\left(s + \frac{y^2 t}{2}\right), & \frac{y^2 t}{2} \le s \le 1 \end{cases} \tag{3.20}$$

Si $t \ge 2/y^2$, elles sont données par

$$Q_{u_{y,t}}(s) = \begin{cases} -1, & s = 0 \\ \sqrt{2ts} & 0 < s \le 1 \end{cases} \tag{3.21}$$

On est en mesure de prouver le théorème suivant :

Théorème 3.2. *Soit $\mu \in \mathcal{M}$ et soit $\mathcal{T} := \log_\mu(\mathcal{M})$. Alors il existe $C > 0$ tel que*

$$d_n(\mathcal{T}, L^2([0,1])) \le C\ n^{-21/10} \tag{3.22}$$

Preuve : Posons $\tilde{\mathcal{T}} := \{Q_u, u \in \mathcal{M}\} = \mathcal{T} + Q_\mu$. On va commencer par montrer que

$$d_n(\tilde{\mathcal{T}}, L^2([0,1])) \le C\ n^{-21/10} \tag{3.23}$$

On en déduira l'inégalité de l'énoncé du théorème.

Pour montrer (3.23), on construit un espace d'approximation de dimension finie adapté aux expressions analytiques (3.20)et (3.21). Pour estimer la qualité de l'approximation offerte, on montre d'abord deux inégalités auxiliaires. Soit $s_0 \in (0,1)$ et $\epsilon > 0$ tel que $I_0 := [s_0 - \epsilon/2, s_0 + \epsilon/2] \subset (0,1)$. On définit aussi, sur I_0

$$f_1(s) := 1, \quad f_2(s)s = s, \quad f_3(s) = \sqrt{s} \quad \forall s \in I_0$$

et on note $W(I_0) := \text{Vect}(f_1, f_2, f_3)$. Soit $P_{W(I_0)}$ la projection orthogonale de $L^2(\Omega)$ sur $W(I_0)$. Comme la projection orthogonale minimise la distance à l'espace de projection, on a

$$||g_z - P_{W(I_0)}g_z||_{L^2(I_0)} \leq ||g_z - \sqrt{2t}f_3 - \frac{1}{y}f_2 + \frac{yt}{2}f_1||_{L^2(I_0)} = ||h_z||_{L^2(I_0)}$$

avec $h_z = g_z - \sqrt{2t}f_3 - \frac{1}{y}f_2 + \frac{yt}{2}f_1$. Sa dérivée vaut

$$h_z'(s) = \begin{cases} -\frac{1}{y} & \text{si} \quad s_0 - \epsilon/2 \leq s < \frac{y^2t}{2} \\ -\sqrt{\frac{t}{2s}} & \text{si} \quad \frac{y^2t}{2} \leq s \leq s_0 + \epsilon/2 \end{cases}$$

La dérivée de h_z est donc bornée par $\frac{1}{y}$ et h_z est $\frac{1}{y}$-Lipschitzienne. Comme $h_z(\frac{y^2t}{2}) = 0$ et $y \in [1/2, 3]$, on obtient que $\sup_{s\in I_0} |h_z(s)| \leq 2\epsilon$, d'où la première inégalité

$$||g_z - P_{W(I_0)}g_z||_{L^2(I_0)} \leq ||h_z||_{L^2(I_0)} \leq 2\epsilon^{3/2} \tag{3.24}$$

Pour la deuxième inégalité, on utilise une inégalité de Taylor-Lagrange : pour toute fonction $f : [0,1] \longrightarrow \mathbb{R}$ dont la dérivée est $M-$Lipschitzienne,

$$|f(s') - f(s) - (s'-s)f'(s)| \leq \frac{M}{2}|s-s'|^2 \quad \forall (s,s') \in [0,1]^2$$

On l'applique à $j_z := g_z - \sqrt{2t}f_3$. D'abord, $j_z(s) = 0$ si $s_0 - \epsilon/2 \leq s \leq \frac{y^2t}{2}$ et

$$j_z'(s) = \frac{1}{y} - \sqrt{\frac{t}{2s}} \quad \text{si} \quad \frac{y^2t}{2} < s \leq s_0 + \epsilon/2$$

Ainsi, j_z' est continue, dérivable sur $I_0 \setminus \left\{\frac{y^2t}{2}\right\}$ et

$$j_z''(s) = \frac{1}{2\sqrt{2}}\sqrt{t}s^{-3/2} \quad \text{si} \quad \frac{y^2t}{2} < s \leq s_0 + \epsilon/2$$

Donc si $t > 0$, la constante de Lipschitz de j_z' est majorée par $\frac{1}{y^3t}$ sur I_0. Au point $s' = \frac{y^2t}{2}$, l'inégalité de Taylor Lagrange donne alors

$$|j_z'(s)| \leq \frac{1}{2y^3t}\epsilon^2$$

ce qui donne, par propriété minimisante de la projection orthogonale, que

$$||g_z - P_{W(I_0)}g_z||^2_{L^2(I_0)} \leq ||j_z||^2_{L^2(I_0)} \leq \frac{1}{4y^6t^2}\epsilon^5$$

D'où la deuxième inégalité,

$$||g_z - P_{W(I_0)}g_z||_{L^2(I_0)} \leq \frac{1}{2y^3t}\epsilon^{5/2} \tag{3.25}$$

On peut maintenant montrer le résultat général. Soit $\beta > 1$ dont on fixera la valeur plus tard. On commence par partitionner l'intervalle $[0,1]$ en $2n$ intervalles $(I_k)_{1\leq k\leq 2n}$, tels que les n premiers soient de taille $\frac{1}{n^\beta}$ et les n derniers soient de taille au plus $\frac{1}{n}$. Pour $1 \leq k \leq n$, on pose

$$x_k := \frac{1}{2n^\beta} + (k-1)\frac{1}{n^\beta} \text{ et } I_k := \left[x_k - \frac{1}{2n^\beta}, x_k + \frac{1}{2n^\beta}\right[$$

et pour $n+1 \leq k \leq 2n$,

$$x_k := \frac{n}{n^\beta} + \frac{1}{2n} + (k-n-1)\frac{1}{n} \text{ et } I_k := \left[\min\left(1, x_k - \frac{1}{2n}\right), \min\left(1, x_k + \frac{1}{2n}\right) \right[$$

Ensuite, on définit

$$V_n := \text{Span}\{\mathbb{1}_{I_k}(s), \mathbb{1}_{I_k}(s)s, \mathbb{1}_{I_k}(s)\sqrt{s}, 1 \leq k \leq 2n\}$$

de dimension $6n$.

Si $t > 2/y^2$, on voit à l'aide de (3.21) que $g_z \in V_n$ et donc $||g_z - P_{V_n} g_z||_{L^2([0,1])} = 0$.

Si $t \leq 2/y^2$, alors $\frac{y^2 t}{2} \in (0,1)$ et on a

$$||Q_{u_z} - P_{V_n} Q_{u_z}||_{L^2([0,1])} = ||g_z - P_{V_n} g_z||_{L^2([0,1])} = ||g_z - P_{W(I_{k_0})} g_z||_{L^2(I_{k_0})}$$

où $k_0 \in \{1, ..., 2n\}$ est l'entier tel que $\frac{y^2 t}{2} \in I_{k_0}$. En effet, l'espace V_n reconstruit parfaitement g_z sur les autres intervalles. Si $k_0 \in \{1, ..., n\}$, alors la première inégalité donne que

$$||g_z - P_{W(I_{k_0})} g_z||_{L^2(I_{k_0})} \leq 2n^{-3\beta/2}$$

Si $k_0 \in \{n+1, ..., 2n\}$, alors on a nécessairement que $\frac{y^2 t}{2} \geq \frac{n}{n^\beta}$ soit $y^3 t \geq 2y n^{1-\beta}$ et la deuxième inégalité donne que

$$||g_z - P_{W(I_{k_0})} g_z||_{L^2(I_{k_0})} \leq \frac{1}{2} n^{-7/2+\beta}$$

On choisit β tel que $\frac{-3\beta}{2} = \frac{-7}{2} + \beta$, soit $\beta = \frac{7}{5}$ et on obtient finalement que

$$||Q_{u_z} - P_{V_n} Q_{u_z}||_{L^2([0,1])} \leq 2n^{-21/10}$$

qui à son tour fournit l'inégalité (3.23).

On termine en montrant que $d_n(\mathcal{T}, L^2([0,1])) \leq 2 d_n(\tilde{\mathcal{T}}, L^2([0,1]))$. Comme $\mu \in \mathcal{M}$, on a que $Q_\mu \in \tilde{\mathcal{T}}$ et ainsi,

$$\begin{aligned} d_n(\mathcal{T}, L^2([0,1])) &= \inf_{\substack{V_n \subset L^2(\Omega) \\ \dim V_n = n}} \sup_{f \in \mathcal{T}} ||f - P_{V_n} f||_{L^2([0,1])} \\ &\leq \inf_{\substack{V_n \subset L^2(\Omega) \\ \dim V_n = n}} \sup_{g \in \tilde{\mathcal{T}}} ||g - P_{V_n} g||_{L^2([0,1])} + \inf_{\substack{V_n \subset L^2(\Omega) \\ \dim V_n = n}} ||Q_\mu - P_{V_n} Q_\mu||_{L^2([0,1])} \\ &\leq 2 d_n(\tilde{\mathcal{T}}, L^2([0,1])). \end{aligned}$$

□

Chapitre 4

Méta-modélisation pour l'équation de Korteweg-de Vries

Dans ce chapitre, nous nous intéressons à la résolution numérique d'une équation issue originellement de la mécanique des fluides : l'équation de Korteweg-de Vries (KdV). C'est une équation d'évolution unidimensionnelle, non-linéaire et dispersive, qui permet de décrire le déplacement de vagues dans la limite des eaux peu profondes, comme c'est le cas dans certains fleuves. Cette équation permet par exemple d'expliquer l'observation d'ondes solitaires se déplaçant sur de grandes distances avant de s'amortir.

Elle peut être approximée dans une certaine mesure par l'équation de Burgers dans le régime dess ondes longues. Nous nous intéressons ici à la construction d'un complément de modèle à l'équation de Burgers pour approximer les solutions de KdV.

1 Equation de Korteweg-de Vries, équation de Burgers

1.1 L'équation de Burgers visqueuse

Nous avons déjà rencontré l'équation de Burgers dans la section précédente, dans sa forme non visqueuse. Soit $\nu \geq 0$, l'équation de Burgers visqueuse s'écrit comme

$$\partial_t u + \partial_x\Big(\frac{u^2}{2}\Big) = \nu \partial^2_{xx} u$$

On peut obtenir l'équation de Burgers à partir des équations d'Euler en mécanique des fluides, en négligeant certains paramètres ; u représente alors la vitesse du fluide et ν sa viscosité cinématique ; le cas non visqueux correspond au cas $\nu = 0$. Dans ce cas, l'équation de Burgers devient hyperbolique. Comme elle est non-linéaire, ses solutions peuvent présenter l'apparition de discontinuités en temps fini (ondes de chocs), même pour une condition initiale régulière. Néanmoins, elle est facilement et rapidement résoluble numériquement à l'aide de schémas simples de type volumes finis.

L'équation avec viscosité présente l'avantage de remplacer les discontinuités obtenues dans le cas non visqueux par "seulement" des forts gradients dans la solutions, mais ne présentant pas de discontinuités.

1.2 L'équation de Korteweg-De Vries

L'équation de Korteweg-de Vries s'écrit selon

$$\partial_t u + \underbrace{\partial_x\left(\frac{u^2}{2}\right)}_{\text{transport}} + \underbrace{\epsilon^2 \partial^3_{xxx} u}_{\text{terme dispersif}} = 0 \tag{4.1}$$

Ici, on a paramétré l'équation à l'aide d'un paramètre ϵ, mais un changement d'échelle $u(t,x) \longmapsto au(bt,cx)$ permet en fait d'obtenir n'importe quel coefficient devant chacun des termes de (4.1). Elle présente notamment un terme dispersif ; ce dernier permet de rendre compte de l'apparition occasionnelle d'un train d'ondes devant ou derrière un paquet d'onde central. Dans la limite $\epsilon \longrightarrow 0$, on retrouve l'équation de Burgers. Numériquement, cette équation est plus coûteuse à résoudre que l'équation de Burgers. Par exemple, les schémas explicites classiques de résolution comme celui présenté dans [1] ont une condition de stabilité CFL (Courant-Friedrichs-Lewy) restrictive en $\Delta t = O(\Delta x^3)$ où Δt et Δx sont les pas en temps et en espace du schéma numérique, contre $\Delta t = O(\Delta x)$ pour Burgers.

Remarque : Une revue extensive des schémas numériques de résolutions pour l'équation de KdV se trouve dans [11].

1.3 Complément de modèle à l'équation de Burgers

On donne ci-dessous une figure comparant la solution de l'équation de KdV à celle de l'équation de Burgers pour une même condition initiale gaussienne.

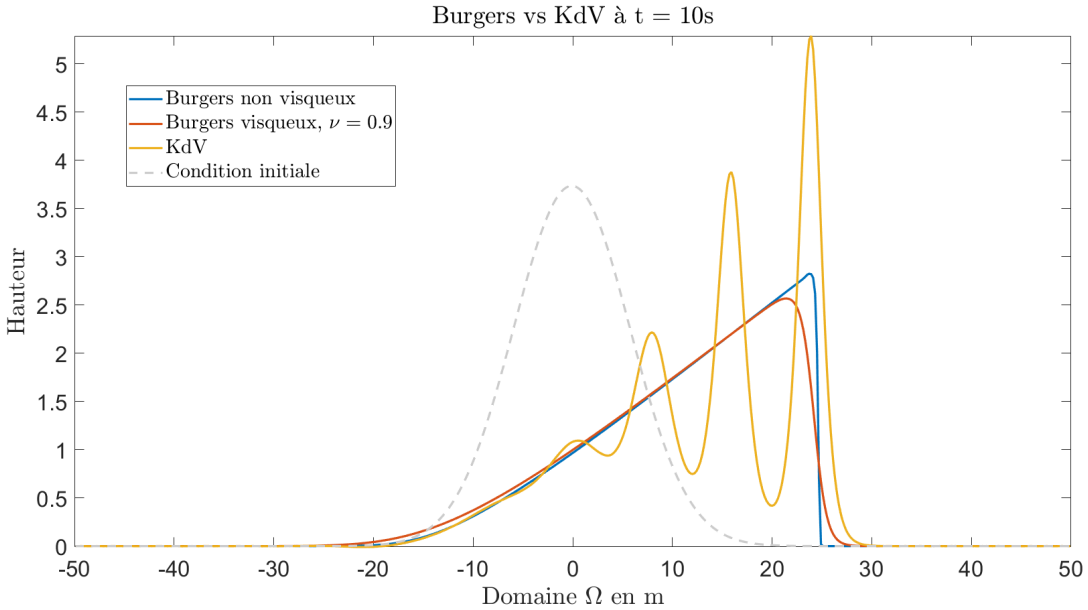


FIGURE 4.1 – Comparaison des solutions de KdV, Burgers et Burgers visqueux

Rappelons qu'à condition initiale fixée, on souhaite approcher la solution de l'équation de KdV par la solution de l'équation de Burgers à laquelle on ajoute un terme correctif. Connaissant la solution de l'équation de Burgers, on peut écrire l'EDP que ce complément de modèle est censé vérifier. Pour cela, notons u_K la solution de l'équation de KdV, u_B la solution de l'équation de

Burgers et u_C la correction apportée par le complément de modèle. On écrit donc $u_K = u_B + u_C$ et l'équation de KdV nous donne

$$\begin{aligned} 0 &= \partial_t u_K + \partial_x\Big(\frac{{u_K}^2}{2}\Big) + \epsilon^2 \partial^3_{xxx} u_K \\ &= \partial_t u_B + \partial_t u_C + u_B \partial_x u_B + u_C \partial_x u_C + (u_B \partial_x u_C + u_C \partial_x u_B) + \epsilon^2 \partial^3_{xxx} u_C + \epsilon^2 \partial^3_{xxx} u_B \end{aligned}$$

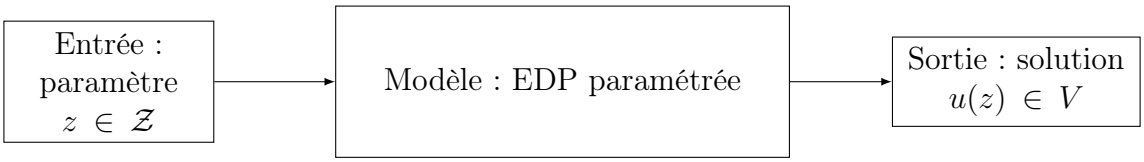
Soit encore, comme $\partial_t u_B + u_B \partial_x u_B = 0$,

$$\partial_t u_C + \partial_x\Big(\frac{{u_C}^2}{2}\Big) + \partial_x(u_B u_C) + \epsilon^2 \partial^3_{xxx} u_C = -\epsilon^2 \partial^3_{xxx} u_B \tag{4.2}$$

Au vu de la complexité de cette équation, on comprend qu'une approche basée sur des données d'observation peut constituer une bonne voie de sortie : plutôt que de résoudre directement l'équation (4.2) ou KdV, on va l'approcher à l'aide de techniques d'apprentissage machine.

2 Quelques techniques et outils des sciences des données

On peut reprendre l'architecture qu'on avait utilisée lors de l'introduction à la méthode des bases réduites : le modèle que l'on cherche à approcher est de la forme ci-dessous.



On se place dans le cas où chaque itération du modèle est très coûteuse : la base de donnée à disposition sera nécessairement très petite, comme de l'ordre de quelques centaines d'observations contre plusieurs millions pour d'autres applications. Dans cette optique, une technique d'apprentissage statistique est utilisée plus que les autres : la modélisation par processus gaussiens. Cette technique a pour avantage d'interpoler aux points d'observation, d'offrir des calculs résolubles analytiquement et de fournir des intervalles de confiance aux points de prédiction.

Néanmoins, dans la grande majorité des cas, la sortie prédite par un processus gaussien est scalaire, i.e. unidimensionnel. Hors, nous avons ici affaire à des sorties fonctionnelles, i.e. de dimension infinie ou très grande si l'on discrétise l'espace d'arrivée. Un travail préliminaire de réduction de la dimension de l'espace d'arrivée doit donc être effectué avant d'utiliser des processus gaussiens.

2.1 Eléments de modélisation par processus gaussiens

Cette section introduit brièvement la modélisation par processus gaussiens. Une présentation extensive de cette technique se trouve dans [3]. On commence par rappeler la définition d'un processus gaussien.

Définition 4.1 (Processus gaussien)**.** *On appelle processus gaussien toute famille de variables aléatoires réelles $(Z_x)_{x\in X}$ définies sur un même espace probabilisé telle que toute sous-famille finie soit un vecteur gaussien :*

$$\forall n \in \mathbb{N},\ \forall \{x_1, ..., x_n\} \subset X,\ \forall (a_1, ..., a_n) \in \mathbb{R}^n,\ \sum_{j=1}^{n} a_j Z_{x_j} \text{ suit une loi normale.}$$

Dans la définition précédente, X peut être un ensemble quelconque mais c'est souvent une sous-partie de $\mathbb{R}^d$. On a alors le théorème suivant :

Théorème 4.1 (Caractérisation par les fonctions moyenne et covariance)**.** *Tout processus gaussien $(Z_x)_{x\in X}$ est déterminé de façon unique par sa fonction moyenne μ et sa fonction de covariance γ, définies par*

— $\forall x \in X, \mu(x) := \mathbb{E}[Z_x]$

— $\forall x, x' \in X, \gamma(x, x') := Cov(Z_x, Z_{x'})$

Et alors la fonction γ est semi-définie positive.

Inversement, pour toute fonction μ mesurable et fonction γ semi-définie positive, il existe un unique processus gaussien $(Z_x)_{x\in X}$ ayant pour fonctions moyenne et covariance les fonctions μ et γ. On note

$$(Z_x)_{x\in X} \sim \mathcal{GP}\big(\mu, \gamma\big) \quad ou \quad (Z_x)_{x\in X} \sim \mathcal{GP}\big(\mu(x), \gamma(x, x')\big)$$

Le processus gaussien le plus fameux est le processus de Wiener, aussi appelé "mouvement brownien", indexé par $\mathbb{R}^+$ dont la fonction moyenne est nulle et dont la fonction de covariance est donnée par $\gamma(x, x') = \min(x, x')$.

Modélisation par processus gaussiens Supposons que la sortie $u(z)$ à modéliser soit scalaire réelle, et que l'espace des paramètres $\mathcal{Z}$ soit une partie de $\mathbb{R}^d$ (on se ramènera par la suite à cette configuration). La modélisation par processus gaussiens consiste à modéliser la fonction $\mathcal{Z} \ni z \longmapsto u(z)$ par un processus gaussien $(U_z)_{z\in\mathcal{Z}}$. Pour chaque nouvelle observation $u(z_0)$ ajoutée dans le jeu de donnée, il est possible de conditionner la loi du processus gaussien (U_z) en $V_z = \big(U_z | U_{z_0} = u(z_0)\big)$ et d'obtenir ainsi un nouveau processus gaussien, qui par définition vaudra $u(z_0)$ presque sûrement au point z_0. Ce processus interpole donc aux points d'observation. De plus, tous les calculs de conditionnement sont réalisables analytiquement.

On obtient donc un nouveau processus gaussien $(V_z)_{z\in\mathcal{Z}} \sim \mathcal{GP}\big(\widetilde{\mu}(x), \widetilde{\gamma}(x, x')\big)$ et la valeur prédite au point z^* sera $\widetilde{\mu}(z^*)$, avec comme écart-type associé $\sqrt{\widetilde{\gamma}(z^*, z^*)}$.

Choix d'une fonction de covariance La fonction de covariance est l'élément central d'un processus gaussien. Elle peut être calibrée selon la fonction que le processus gaussien est censé approximer : stationarité du processus, équations vérifiées par ses trajectoires ainsi que leur régularité, ... La fonction de covariance permet de représenter la physique du phénomène que le processus est supposé modéliser [8].

Le choix d'une fonction de covariance se fait généralement parmi des familles de fonctions semi-définies positives paramétrées. Voici ci-dessous un exemple simple de telle famille, les fonctions gaussiennes, aussi appelées "squared-exponential" dans la littérature :

$$\gamma_{r,l}(x, x') := r^2 e^{-\frac{|x-x'|^2}{2l^2}} \quad \forall x, x' \in X$$

Les paramètres de la fonction de covariance (ci-dessus, les réels r et l) sont appelés "hyperparamètres". Le choix d'une telle famille se fait en général à l'aide de connaissances métier, i.e. de savoirs spécifiques à l'application que l'on fait du processus gaussien.

Apprentissage d'un processus gaussien Ayant choisi une famille de fonctions de covariance $\gamma(x, x', \theta)$ paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on peut maintenant calibrer les hyperparamètres θ en fonction des données d'observation : c'est la phase d'apprentissage. Notons

$$\boldsymbol{z} = (z_1, ..., z_n) \qquad \text{les points d'observation}$$
$$\boldsymbol{u} = (u(z_1), ..., u(z_n)) \qquad \text{la base de données}$$

Le calibrage se fait en maximisant par rapport à θ la fonction de vraisemblance marginale, qui reflète la probabilité qu'on ait observé $\boldsymbol{u}$ aux points $\boldsymbol{z}$ étant donné un hyperparamètre θ. On cherche donc les hyperparamètres θ pour lesquels les observations sont les plus vraisemblables. Notons K_θ la matrice dont les coefficients sont donnés par $\gamma(z_i, z_j, \theta), \forall i, j \in \{1, ..., n\}$. La (log-)vraisemblance marginale s'exprime comme

$$p(\boldsymbol{u}|\boldsymbol{z}, \theta) = -\frac{1}{2}\boldsymbol{z}^T K_\theta^{-1}\boldsymbol{z} - \frac{1}{2}\log\det(K_\theta) - \frac{n}{2}\log 2\pi \tag{4.3}$$

(voir [3], Chap.2, eq 2.29) et on souhaite donc résoudre

$$\theta^* = \arg\max_{\theta\in\Theta} p(\boldsymbol{u}|\boldsymbol{z}, \theta) \tag{4.4}$$

Le coût numérique d'une telle opération est dominé par le calcul de l'inverse de K_θ dans (4.3). Dans nos simulations numériques, nous avons utilisé le package GPML sous MATLAB qui effectue une décomposition de Cholesky de K_θ plutôt que de l'inverser directement ; c'est une solution plus rapide et plus stable numériquement. Le package utilise ensuite un algorithme de gradient conjugué non linéaire (version de Polak-Ribière) afin de résoudre le problème d'optimisation (4.4).

2.2 Réduction de la dimension

Nous avons utilisé deux techniques afin de réduire la dimension de l'espace d'arrivée : une décomposition en ondelettes et une Analyse en Composantes Principales (ACP) sur les coefficients d'ondelette. Cette articulation particulière de ces deux techniques s'appelle "Functional Principal Componant Analysis" ou FPCA [2].

Transformée en ondelettes

Les ondelettes sont des familles de fonctions tests dont certaines forment des bases orthogonales de L^2 et sont conçues entre autres pour capturer les propriétés locales du signal tout en rendant compte de son contenu fréquentiel, ce que l'analyse de Fourier peine à faire. Pour plus de détails, nous nous référons à [4].

Une propriété clé est qu'une représentation sur une base d'ondelettes aboutit de façon générale à une représentation parcimonieuse : tous sauf quelques coefficients d'ondelettes sont presque nuls (en général, environ 90%). En particulier, deux propriétés fondamentales des ondelettes sont fondamentales pour obtenir une telle représentation : la taille de leur support et leur nombre de moments nuls. Si le signal est irrégulier, on choisira des ondelettes à petit support : les irrégularités locales ne viendront pas polluer beaucoup de coefficients d'ondelettes et on pourra quand même espérer une représentation parcimonieuse. si le signal est régulier, on choisira des ondelettes avec beaucoup de moments nuls : par approximation de Taylor-Young, une grande partie des coefficients d'ondelette seront nuls. Sans surprises, ces deux aspects sont antagonistes (théorème de Debauchies).

Comme la transformée de Fourier, la transformée en ondelettes admet des équivalents discrets (Discrete Wavelet Transform, DWT) et multidimensionnels.

Analyse en Composantes Principales

Supposons que l'on ait un jeu de données avec n individus $(x_1, ..., x_n)$ à valeur dans $\mathbb{R}^p$. L'analyse en composantes principales vise à trouver les p vecteurs $(v_1, ..., v_p)$ de $\mathbb{R}^p$ qui expliquent chacun indépendamment le plus d'information dans les données. Le critère d'information utilisé

par l'ACP est le critère "d'inertie" $\mathcal{I}$: c'est la distance moyenne des données à leur centre de gravité. Si V est un sous-espace vectoriel de $\mathbb{R}^p$, on appelle P_V la projection orthogonale sur V et on a les définitions suivantes de centre de gravité, d'inertie et d'inertie projetée :

$$g := \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \mathcal{I} := \frac{1}{n}\sum_{i=1}^{n} ||x_i - g||^2_{\mathbb{R}^p} \quad \text{et} \quad P_V\mathcal{I} := \frac{1}{n}\sum_{i=1}^{n} ||P_V(x_i - g)||^2_{\mathbb{R}^p} \quad \text{(inertie projetée)}$$

On commence par chercher la droite vectorielle D sur laquelle l'inertie projetée est maximale, et on obtient ainsi une première "composante principale" $v_1 : D = \text{Span}(v_1)$. on réitère ce procédé sur l'orthogonal de D pour obtenir v_2 et ainsi de suite jusqu'à obtenir une base orthonormée de $\mathbb{R}^p$.

On montre que cela revient à trouver les vecteurs propres de la matrice de covariance empirique des données. Si ses valeurs propres sont $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$, alors $\mathcal{I} = \sum_{i=1}^{p} \lambda_i$ et chaque direction principale v_i permet d'expliquer $\frac{\lambda_i}{\mathcal{I}}$ pourcents de l'inertie totale $\mathcal{I}$.

2.3 Stratégie de résolution

Génération de la base de données On résout les équations de Burgers visqueux et de KdV avec le paramètre $\epsilon = 1$ de la façon suivante :

- Domaine de résolution : $\Omega \times [0, T] = [-50, 50]m \times [0, 10]s$ où Ω est muni de conditions aux bords périodiques.
- Discrétisation : $J = 512$ points en espace pour Ω et 256 points en temps pour $[0, T]$.
- Conditions initiales : gaussiennes centrées paramétrées par leur hauteur $h \in [1, 5]$ et leur écart type $\sigma \in [0.1, 7]$: $u(0, x) = u_0(x) = he^{-\frac{x^2}{2\sigma^2}}$
- Echantillonnage de l'espace des paramètres $[1, 5] \times [0.1, 7]$: hypercube latin avec remplissage d'espace ("space-filling algorithm").
- Solveurs utilisés : En espace : MUSCL d'ordre 2 (Kurganov et Tadmor semi-discret). En temps : Runge-Kutta d'ordre 2 (Heun).

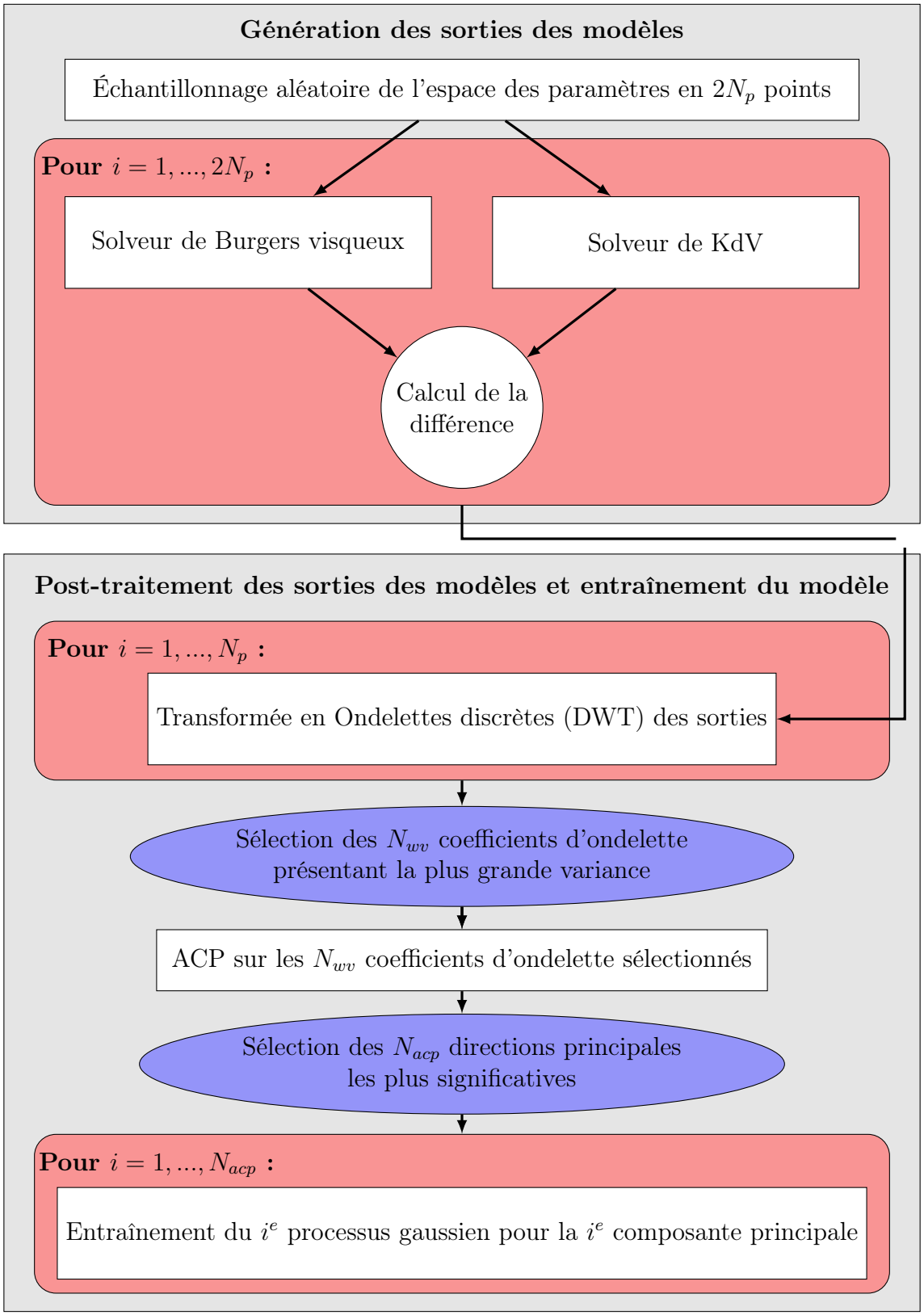
On capture toute la solution générée qu'on stocke en une image de taille 512×256. On effectue 600 runs, 300 pour l'apprentissage et 300 pour le test (validation). Ensuite, on calcule la différence entre chaque image puisqu'on s'intéresse à modéliser la différence entre les solutions de KdV et de Burgers. On construit ainsi une base de données d'apprentissage $B_0 = (u(z_1), ..., u(z_n))$ avec $n = 300$, de taille $300 \times (512 \times 256)$.

Réduction de la dimension On a choisi de résoudre l'équation de Burgers visqueuse car elle permet d'assurer une certaine régularité de ses solutions. Comme les solutions de KdV sont aussi régulières, on utilise des ondelettes avec de nombreux moments nuls. Nous avons choisi la base d'ondelettes de Debauchies 2D à 20 moments nuls, abrégée "db20".

1. On effectue une transformée en ondelettes discrète 2D sur chaque observation et on sélectionne les $N_{wv} = 300$ coefficients présentant la plus grande variance empirique. On a donc une nouvelle base de données B_1 de taille 300×300.
2. On effectue une ACP sur B_1 et on sélectionne les $N_{acp} = 30$ premières composantes principales : elles expliquent 99.8% de l'inertie totale. On a donc une dernière base de données B_2 de taille 300×30, où chaque individu $y_i \in \mathbb{R}^{N_{acp}}$ (i^e ligne) contient les 30 premières coordonnées de l'individu $x_i \in \mathbb{R}^{N_{wv}}$ de la base de données B_1 dans la base des directions principales (qui est une base de $\mathbb{R}^{N_{wv}}$).

Modélisation par processus gaussiens On suppose que le comportement "moyen" de l'équation de KdV est dicté par l'équation de Burgers. On utilisera donc uniquement des processus gaussiens centrés : $\mu \equiv 0$. On choisit une fonction de covariance de Matérn ; elles sont

stationnaires et généralisent les fonctions de covariance gaussiennes. On utilise la base de données B_2, et on entraîne donc 30 processus gaussiens indépendants, chacun correspondant à une composante principale. Remarquons que l'étape d'ACP légitime une modélisation de processus gaussiens indépendants : cette hypothèse et beaucoup plus discutable si l'on travaille directement sur les coefficients d'ondelette. Or, la modélisation de processus gaussiens non indépendants est beaucoup plus fastidieuse... Ci-dessous se trouve un diagramme résumant tout le processus effectué.



3 Quelques résultats numériques

Pour chaque élément de l'ensemble de données de test, on effectue une prédiction à l'aide du modèle qu'on compare à la vraie valeur que le modèle est censé prédire. Ci-contre se trouve un exemple où le modèle est performant. Après cette prédiction, un changement de base est effectué pour passer de la base des directions principales à la base des ondelettes 'db20'. Enfin, on effectue une transformée en ondelettes discrètes inverse (IDWT). On se retrouve finalement avec une image de taille 512×256.

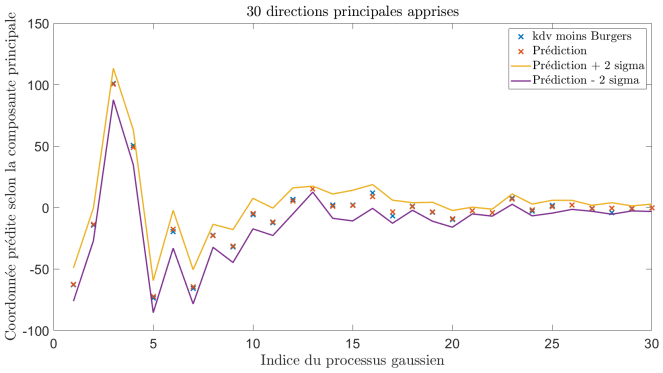


FIGURE 4.2 – Exemple de prédiction des 30 processus gaussiens

Ci-dessous se trouve la comparaison entre les solutions produites par (1) Burgers, (2) KdV, (3) le complément de modèle correspondant aux prédictions de la Figure 4.2 et (4) le complément de modèle si ses prédictions étaient parfaites (i.e. si dans la Figure 4.2, les croix rouges et bleues étaient identiques). On observe qu'il y a très peu de différences entre la prédiction du complément de modèle (Fig. 4.2, en bas à gauche) et la meilleure prédiction possible (Fig. 4.2, en bas à droite). La différence entre la figure en bas à droite et la solution de KdV correspondante est donc due (modulo l'étape d'ACP) à la troncature que nous effectuons dans la représentation en ondelettes : rappelons que nous souhaitons représenter la différence entre les solutions des équations KdV et de Burgers à l'aide d'uniquement $N_{wv} = 300$ coefficients d'ondelettes.

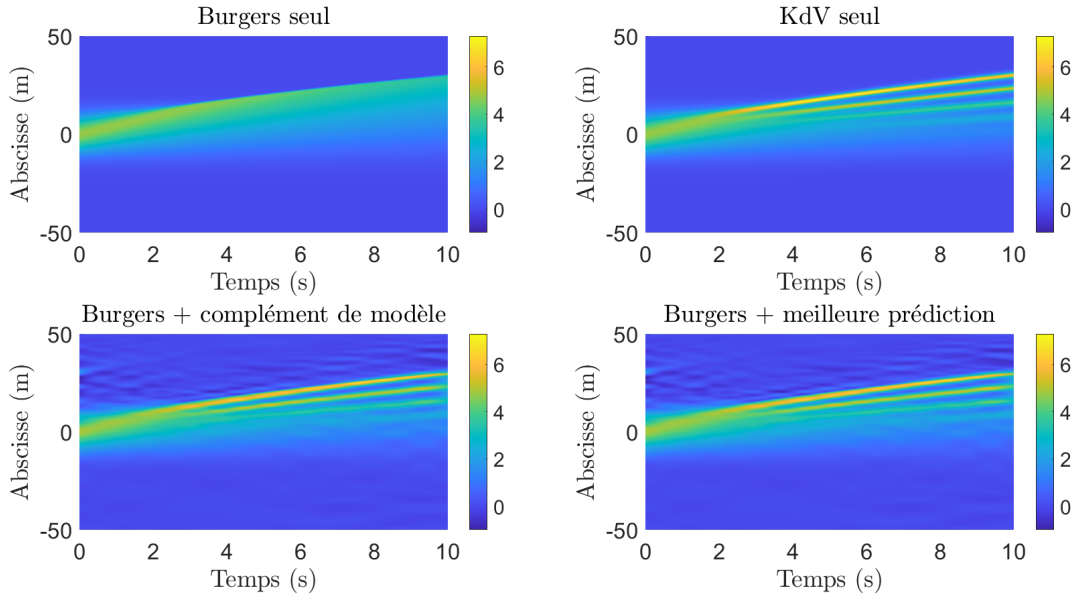


FIGURE 4.3 – Exemple de prédiction de solutions, vue supérieure

Une inspection des Figures 4.3 et 4.4 montre que le modèle semble capable de prédire le bon nombre de vaguelettes produites par KdV. Cependant, la "crête maximale" de la solution de KdV n'est pas reproduite par le métamodèle (Figure 4.4). Les irrégularités observées en Figure 4.4 sont dues à la troncature que nous effectuons dans la base d'ondelettes. Sur l'ensemble des données de test ($N_{test} = 300$), nous avons calculé des erreurs normalisées du métamodèle contre

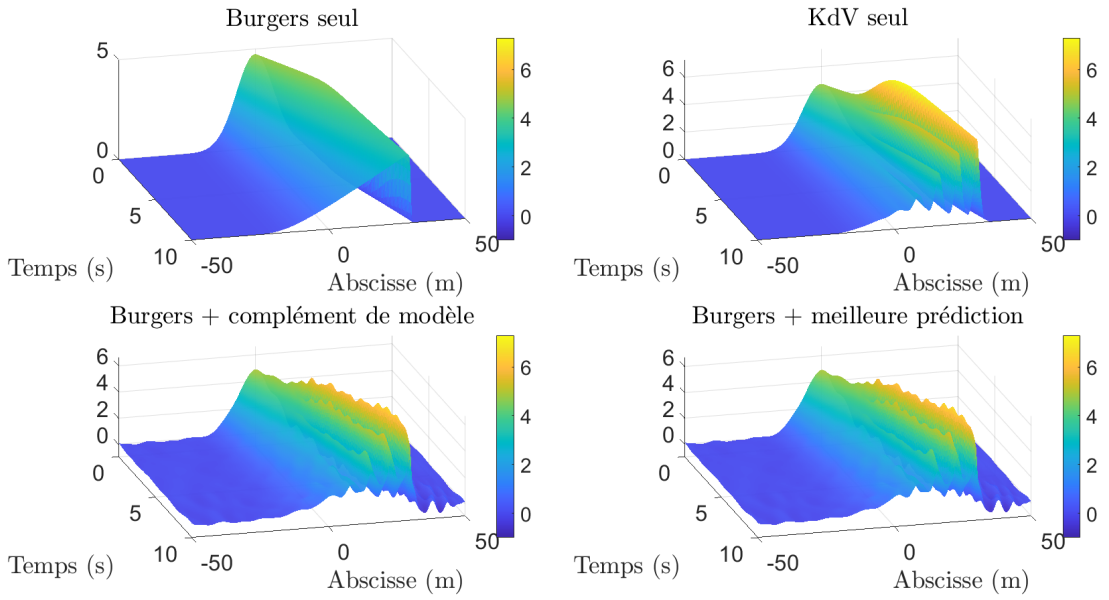


FIGURE 4.4 – Même exemple qu'en Figure 4.3, vue latérale

la meilleure prédiction et la solution de KdV en guise de premiers indicateurs statistiques. Ils sont résumés dans les boîtes à moustache en Figure 4.5. Ces erreurs sont calculées selon

1. Erreur 1 : l'erreur L^2 par rapport à la meilleure prédiction : $e_1 = \frac{||u_{pred} - u_{\text{best pred}}||_{L^2}}{||u_{\text{best pred}}||_{L^2}}$
2. Erreur 2 : l'erreur en norme ∞ par rapport à la meilleure prédiction : $e_2 = \frac{||u_{pred} - u_{\text{best pred}}||_{\infty}}{||u_{\text{best pred}}||_{\infty}}$
3. Erreur 3 : l'erreur L^2 normalisée par rapport à la solution de KdV : $e_3 = \frac{||u_{pred} - u_{KdV}||_{L^2}}{||u_{KdV}||_{L^2}}$
4. Erreur 4 : l'erreur en norme ∞ par rapport à la solution de KdV : $e_4 = \frac{||u_{pred} - u_{KdV}||_{\infty}}{||u_{KdV}||_{\infty}}$

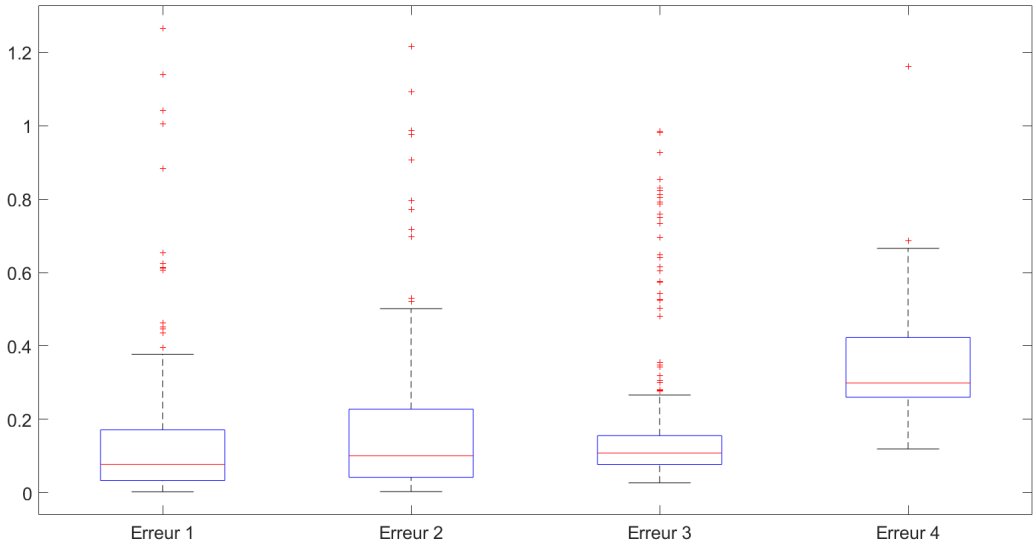


FIGURE 4.5 – Boîtes à moustache des erreurs normalisées

On observe que de façon générale, les processus gaussiens prédisent relativement bien par rapport à leur objectif (médiane à 10% d'erreur) ; L'erreur en norme L^2 est du même ordre par rapport à KdV, bien que présentant plus de nombreuses valeurs extrêmes ; enfin, l'erreur 4 est la plus grande en moyenne, comme on aurait déjà pu le deviner sur la Figure 4.4.

Conclusion et perspectives

La réduction de modèle est un sujet de recherche actuel très fécond et peut être abordée selon des points de vue très variés : les questions qu'elle soulève concernent des champs très diverses des mathématiques. La formulation classique de la méthode des bases réduites utilise le langage et les outils standards de l'analyse des EDPs ; mais d'autres formulations sont possibles. Nous avons vu qu'en plongeant les solutions de certaines lois de conservation dans un espace de mesures de probabilité, il était possible d'obtenir une nouvelle méthode des bases réduites non-linéaire qui était, au moins dans certains cas, bien plus adaptée aux structures des solutions de l'EDP en question. Cependant, cette formulation présente des faiblesses. Notamment, les transformations non-linéaires sont locales : l'exponentielle $\exp_\mu$ n'est définie que sur le convexe fermé $K_{Q_\mu} \subset L^2([0,1])$. Cet aspect restrictif doit être pris en compte lors de la conception pratique d'algorithmes de bases réduites, comme observé dans [12]. De plus, le cas de la dimension 1 permet de rendre explicites toutes les transformations non-linéaires en question. Dans le cas de lois de conservations multidimensionnelles, ces formules explicites n'ont plus lieu et compliquent grandement la mise en pratique d'une telle méthode des bases réduites.

L'abord de la réduction de modèle du point de vue de la métamodélisation constitue une alternative populaire. Nous avons exposé une mise en oeuvre de métamodélisation par processus gaussiens et obtenu des premiers résultats satisfaisants d'un point de vue qualitatif. Il faut maintenant affiner le modèle et quantifier la qualité de ces résultats à l'aide d'indicateurs statistiques comme le RMSE (Root Mean Square Error) ou d'autres indicateurs plus sophistiqués. Une tâche plus ardue consiste en la construction d'un tel métamodèle pour des données en entrée plus générales, c'est-à-dire ne se limitant pas à des familles de gaussiennes.

Ces pistes et questions ouvertes constituent autant de directions possibles à suivre en thèse. La trame de fond de la thèse visera à la mise en place de ces techniques de réduction de modèle pour des modèles de mécanique des fluides de plus en plus sophistiqués : l'objectif à terme sera d'obtenir des modèles réduits efficaces pour la prédiction de phénomènes marins côtiers.

Bibliographie

[1] N.J. Zabusky et M.D. Kruskal. « Interaction of "solitons" in a collisionless plasma and the recurrence of initial states ». In : *Phys. Rev. Lett.* (1965), 15(6) :240-243.

[2] J.O. Ramsay. *Functional Data Analysis*. Wiley Online Library, 2006.

[3] C. E. Rasmussen et C.K.I. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006. isbn : 026218253X. url : www.GaussianProcess.org/gpml.

[4] S Mallat. *A Wavelet Tour of Signal Processing, 3rd Edition*. Academic Press, 2008.

[5] L Ambrosio et N Gigli. *A user's guide to optimal transport*. Italy : CIME summer school, 2009. url : https://hal.archives-ouvertes.fr/hal-00769391.

[6] A Marrel et al. « Global sensitivity analysis for models with spatially dependent outputs ». In : *Environmetrics* 22 (2011), p. 383-397. url : https://hal.archives-ouvertes.fr/hal-00430171.

[7] J.S. Hesthaven, G. Rozza et B Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. SpringerBriefs in Mathematics. Springer International Publishing, 2015. isbn : 978-3-319-22469-5. doi : 10.1007/978-3-319-22470-1.

[8] D Ginsbourger, O Roustant et N Durrande. « On degeneracy and invariances of random fields paths with applications in Gaussian process modelling ». In : *Journal of Statistical Planning and Inference* (2016), 170 :117 -128.

[9] M Ohlberger et S Rave. « Reduced Basis Methods : Success, Limitations and Future Challenges ». In : *Proceedings of the Conference Algoritmy* (2016), p. 1-12. url : http://www.iam.fmph.uniba.sk/amuc/ojs/index.php/algoritmy/article/view/389.

[10] J. Bigota et al. « Geodesic PCA in the Wasserstein space by convex PCA ». In : *Annales de l'institut Henri Poincaré, Probabilités et Statistiques* (2017), volume 53, pages 1-26.

[11] C Courtès. *Analyse numérique de systèmes hyperboliques-dispersifs*. Thèse de doctorat de l'Université Paris-Saclay, 2017. url : http://irma.math.unistra.fr/~courtes/72024_COURTES_2017_archivage.pdf.

[12] V. Ehrlacher et al. « Nonlinear model reduction on metric spaces. Application to one-dimensional conservative PDEs in Wasserstein spaces ». working paper or preprint. Sept. 2019. url : https://hal.inria.fr/hal-02290431.