Systèmes complexes gouvernés par des flux : schémas de volumes finis hybrides et optimisation numérique

Pascal Jaisson

ÉCOLE CENTRALE DES ARTS ET MANUFACTURES
« ÉCOLE CENTRALE PARIS »

THÈSE

présentée par

Pascal Jaisson

pour l'obtention du

GRADE DE DOCTEUR

Spécialité : Mathématiques

Laboratoire d'accueil : MAS

SUJET :

Systèmes complexes gouvernés par des flux : schémas de volumes finis hybrides et optimisation numérique.

soutenue le : 13 octobre 2006

devant un jury composé de :

M. Grégoire Allaire	**Président**
M. François Alouges	**Examinateur**
M. Fayssal Benkhaldoun	**Rapporteur**
M. Florian De Vuyst	**Directeur**
M. Jérôme Jaffré	**Examinateur**
M. Frédéric Pascal	**Rapporteur**
M. Bruno Scheurer	**Invité**

2006 - 20

Remerciements

Mes remerciements vont tout d'abord à Florian DeVuyst. Il a eu le courage d'encadrer ma thèse pendant ces trois années au cours desquelles je suis resté enseignant au lycée. Ses nombreux conseils, sa disponibilité, sa gentillesse et son optimisme m'ont permis de mener à bien ce travail.

Je remercie toute l'équipe du laboratoire d'Analyse Numérique d'Orsay et plus particulièrement François Alouges qui m'a fait confiance et qui a tout mis en oeuvre pour me permettre de suivre le DEA dans des conditions optimales. Je le remercie aussi de m'avoir guidé vers Florian.

J'exprime ma vive reconnaissance à Jérôme Jaffré pour l'attention qu'il m'a portée lors de nos entretiens.

Je remercie Fayssal Benkhaldoun et Frédéric Pascal d'avoir accepté de rapporter ma thèse.

C'est un grand honneur pour moi que d'avoir Grégoire Allaire, François Alouges, Jérôme Jaffré et Bruno Scheurer comme membres du jury de ma thèse. Je les en remercie.

Je tiens également à remercier tous les membres du laboratoire MAS ainsi que Sylvie Dervin et Géraldine Carbonel grâce à qui les problèmes administratifs sont vite résolus.

Je n'oublie pas mes collègues du Lycée Parc de Vilgénis, avec lesquels j'ai travaillé dans la bonne humeur pendant ces trois années.

Je tiens à remercier ma famille et tout particulièrement Antoine, dont les éclats de rire ont accompagné la rédaction de mon manuscrit de thèse. Enfin, j'aimerais remercier Emmanuelle pour le bonheur qu'elle m'apporte chaque jour.

Table des matières

Introduction générale

Différents phénomènes physiques sont modélisés par des systèmes de lois de conservation de la forme

$$\partial_t u \ + \ \nabla \cdot f(u) \ = \ 0$$

où u est un champ vectoriel à valeurs dans $\mathbb{R}^n$ ([4],[9]). Par exemple, dans le cas d'une seule dimension, on peut définir la densité d'une certaine quantité physique $\rho(x,t)$ par unité de longueur et un flux $f(x,t)$ associé. La vitesse du fluide est alors $v(x,t) = \frac{f(x,t)}{\rho(x,t)}$ (si $\rho \neq 0$) et l'équation peut s'écrire

$$\partial_t \rho \ + \ \partial_x(v(x,t)\rho(x,t)) \ = \ 0.$$

Pour pouvoir résoudre une telle équation, il est alors nécessaire de connaître d'autres relations pour "fermer" le système : par exemple, une expression donnant v explicitement en fonction de ρ ou bien une équation aux dérivées partielles faisant intervenir l'inconnue v.
Une fois le phénomène physique modélisé par un système de lois de conservation, on peut utiliser les équations aux dérivées partielles pour prévoir le comportement du système et éventuellement agir sur des paramètres pour le contrôler. Il s'agit alors d'optimiser ces paramètres afin d'obtenir la solution souhaitée. Il faut évidemment savoir calculer numériquement les solutions du système de lois de conservation. Nous distinguons donc trois étapes : la modélisation du phénomène, le calcul des solutions numériques et l'optimisation. Dans cette thèse, nous nous intéressons à ces trois étapes. Nous considérons plusieurs problèmes d'optimisation faisant intervenir des système de lois de conservation. Le premier problème concerne les flux d'informations. Nous nous concentrons ensuite sur un modèle de trafic routier. Enfin, nous proposons un nouveau type de schémas numériques hybrides dépendant d'un paramètre.

Flux d'informations. En premier lieu, nous nous intéressons aux problèmes de flux d'informations. Par exemple, nous voulons connaître la quantité moyenne de requêtes de types différents que vont traiter un ou plusieurs serveurs informatiques en fonction de la quantité de requêtes fournies par les utilisateurs et des capacités de traitement disponibles du ou des serveurs.
Florian De Vuyst [5, 6, 7] a proposé une modélisation de ces problèmes par des équations de lois de conservation à une dimension. Depuis, d'autres auteurs ont considéré des modèles continus aux équations aux dérivées partielles pour traiter le trafic d'information [1]. Pour simplifier, les équations peuvent s'écrire sous la forme

$$\partial_t \rho + \partial_x v\rho = 0.$$

La variable $\rho(x,t)$ désigne la quantité de requêtes dont le niveau de traitement est compris entre $x\,\%$ et $x + dx\,\%$ et v est la vitesse de traitement qui s'exprime en fonction de x, t et ρ. Nous avons alors cherché à calculer par des équations différentielles les temps de traitement pour chaque type de requêtes dans différents problèmes que nous nous sommes posés. Un des objectifs est par exemple d'optimiser l'allocation des capacités des différents serveurs suivant le type de requêtes afin de satisfaire des critères de qualité de service. Par exemple, nous devrons minimiser les temps d'attente des utilisateurs, en optimisant une fonctionnelle correctement choisie, qui dépendra du problème considéré.

Trafic routier. Ensuite, nous nous intéressons à un système de lois de conservation pour le trafic routier. Le modèle choisi est un modèle du second ordre introduit par Aw-Rascle [2, 3].

Par exemple dans le cas homogène, le système s'écrit sous forme non conservative :

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \partial_t(v + P(\rho)) + v\partial_x(v + P(\rho)) = 0 \end{cases} \tag{1}$$

où ρ désigne la densité des véhicules, v la vitesse des véhicules et P est un terme qui rend compte du comportement des conducteurs face aux conditions de la route devant eux. Le système peut s'écrire sous forme conservative avec les variables ρ et ρw où $w = v - P$, voir [3]. Nous nous intéressons dans cette partie à un problème d'assimilation de données. Nous désirons en effet retrouver une solution (ρ, v) sur une portion de route et un intervalle de temps donné (notre domaine) en connaissant uniquement quelques valeurs mesurées sur ce domaine. Une fois cette solution (ρ, v) calculée en tout (x, t) du domaine, il est possible d'utiliser les valeurs calculées $(\rho(x, T), v(x, T))$ comme "nouvelle" donnée initiale pour prévoir les conditions du trafic routier après le temps T. Le calcul de (ρ, v) sur le domaine nécessite donc d'assimiler ces données observées à notre problème. Il s'agit là encore d'un problème d'optimisation : il faut minimiser l'erreur entre les valeurs observées en certains points et la solution du système de Aw-Rascle calculée en ces mêmes points.

Schémas numériques Nous nous sommes alors naturellement intéressés à l'étude des schémas numériques des lois de conservation sous la forme générale :

$$\partial_t u + \partial_x f(u) = 0$$

afin d'approcher numériquement les solutions entropiques c'est-à-dire les solutions qui vérifient de plus une inégalité d'entropie

$$\partial_t S(u) + \partial_x F(u) \leq 0$$

où (S, F) est une paire entropie-flux d'entropie. Sous certaines conditions, cette inégalité permet d'éliminer les solutions non physiques et d'assurer l'unicité de la solution [8, 9]. Nous avons alors introduit un nouveau schéma hybride dont l'expression dépend d'un paramètre θ. Il s'agit d'optimiser ce paramètre θ afin d'obtenir un schéma à la fois qui fournit des solutions entropiques mais qui permet aussi de capturer correctement les chocs tout en restant TVD, c'est-à-dire en éliminant les oscillations parasites qui apparaissent par exemple avec le schéma de Lax-Wendroff.

Le plan de la thèse est le suivant.

Flux d'informations. Dans une première partie, nous nous intéressons au problème d'optimisation pour le flux d'information. Nous rappelons les diffèrents modèles aux équations aux dérivées partielles proposés par De Vuyst. Ensuite, nous cherchons les équations différentielles concernant les temps de traitement des requêtes (ou temps de service) ainsi qu'une méthode numérique permettant de calculer ces temps. Enfin nous présentons et traitons divers problèmes d'optimisation de l'allocation des capacités de traitement des différents serveurs. Nous considérons notamment plusieurs critères de qualité de service à satisfaire. Nous traitons ces différents problèmes d'optimisation à l'aide d'un algorithme génétique.

Trafic routier. Dans la deuxième partie, nous nous intéressons au problème d'assimilation de données et d'optimisation pour le trafic routier. Après un bref historique des différents modèles, nous présentons le modèle de Aw-Rascle et certaines de ses propriétés. Nous donnons la matrice de Roe du système que nous avons calculée et différentes validations numériques. Ensuite, nous considérons en détail le problème d'optimisation et les deux stratégies que nous avons mises en oeuvre. Cette fois-ci, nous avons choisi une méthode de région de confiance comme algorithme d'optimisation. Nous avons alors calculé le gradient de la fonctionnelle à optimiser par une méthode adjointe. Une attention particulière a été donnée au traitement des conditions aux bords car elles interviennent directement dans le calcul du gradient.

Schémas numériques. Dans la troisième partie, nous introduisons et étudions les schémas hybrides dépendant d'un paramètre θ. Le but est de trouver un paramètre qui rend le schéma entropique. Nous expliquons la construction de ces nouveaux types de schémas dans le cas scalaire puis dans le cas des systèmes. Ensuite, nous déterminons les conditions sur θ pour obtenir des schémas du second ordre en espace possédant la propriété TVD. Nous validons ces schémas par des expériences numériques dans le cas scalaire et le cas des systèmes, en une ou plusieurs dimensions. Cependant, les schémas avec les conditions trouvées sur θ ne sont pas des schémas entropiques et certains tests font apparaître des solutions numériques non entropiques. Nous avons alors greffé à cette phase prédictrice une phase correctrice pour θ afin de rendre le schéma entropique. Nous expliquons en détail la méthode et l'algorithme utilisés.

Bibliographie

[1] Armbruster, D., Degond, P., and Ringhofer, C. A model for the dynamics of large queuing networks and supply chains. *SIAP 66*, 3 (2006), 896–920.

[2] Aw, A., Klar, A., Materne, T., and Rascle, M. Derivation of continuum traffic flow models from microscopic follow-the-leader models. *SIAM J. Appl. Math. 63*, 1 (2000), 259–278.

[3] Aw, A., and Rascle, M. Resurrection of second order models of traffic flow ? *SIAM J. Appl. Math. 60*, 3 (2000), 916–938.

[4] Dautray, R., and Lions, J. *Analyse mathématique et calcul numérique pour les sciences et les techniques*, vol. 9. Masson, 1988.

[5] DeVuyst, F. Feasibility of fluid transport modelling for buffer and processing systems : Information fluid dynamics, 2002. http ://www.math.ntnu.no/conservation/2002/009.html.

[6] DeVuyst, F. *Fluid modelling of buffer and processing systems, finite volume discretization.* The International Symposium on Finite Volume for Complex Applications III (FVCA 3). Herbin and Kröner Eds, Hermes Penton Science, 2002.

[7] DeVuyst, F. Service time assessment in computational information fluid dynamics using level set methods, 2002. http ://www.math.ntnu.no/conservation/2002/010.html.

[8] Godlewski, E., and Raviart, P. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. New York, Springer, 1996.

[9] Serre, D. *Système de Lois de Conservation I : hyperbolicité, entropies, ondes de choc.* Diderot Editeur, 1996.

Première partie

Optimisation des ressources pour les systèmes d'informations

Introduction

Les systèmes d'information et de communication sont de nature complexe et difficile à analyser. Les utilisateurs de ces systèmes sont exigeants sur la qualité des réponses fournies par les serveurs concernant par exemple le temps de traitement d'une requête de service. Il est ainsi nécessaire de connaître ces temps de traitement en fonction d'une part de la capacité disponible du serveur et d'autre part du nombre de requêtes que ce serveur doit traiter. Il est possible d'estimer à l'avance le nombre de requêtes qui devront être traitées (par exemple, voir [5]). Cela permet d'envisager des stratégies pour répartir au mieux les capacités en fonction des demandes.

Le problème que nous allons traiter concerne ainsi le cas de deux types différents de requêtes dont les temps de traitement pour chacun de ces types de requêtes ne doivent pas être dépassés dans la mesure du possible. Il s'agit d'un problème d'allocation optimale de capacité de serveurs.

La première difficulté concerne la modélisation des flux d'informations. Il faut prendre en compte un très grand nombre de requêtes, ce qui conduit rapidement à des tailles et des complexités très importantes du problème. Les modélisations mathématiques standards reposent sur la théorie probabiliste des files d'attentes [3], [4]. Afin de traiter ici un grand volume de requêtes et pouvoir traiter des régimes quasi-continues de requêtes, nous nous intéressons à une modélisation fluide.

L'approche "fluide" classique pour traiter un tel problème conduit à l'utilisation d'équations différentielles ordinaires couplées ou non à des équations stochastiques [9], [12], [13]. Seulement certaines quantités macroscopiques sont prises en considération (taille d'une file d'attente pour une requête avant d'être traitée, à un instant donné par exemple), mais cela permet de rendre compte des principales propriétés du comportement macroscopique. De Vuyst a proposé une modélisation innovante de tels systèmes par des équations aux dérivées partielles. Cette modélisation s'appuie sur des principes de Mécanique des fluides [6], [7], [8]. Une variable d'espace est alors introduite pour modéliser l'état d'avancement du traitement d'une requête particulière. Cela permet de prendre en compte des effets de mémoire et de délai de traitement en utilisant un processus de transport avec une vitesse de propagation de l'information. De Vuyst a proposé de telles modélisations pour des files d'attentes à règle FIFO (first in-first out, c'est-à-dire premier entré-premier sorti), FILO (first in-last out, c'est-à-dire premier entré-dernier sorti), et des serveurs à tâches partagées pour des réseaux devant traiter un type de requêtes ou plusieurs types de requêtes (version multi-fluide) avec des connexions possibles entre différents serveurs. La modélisation fait intervenir des équations de nature hyperbolique

$$\partial_t \rho + \partial_x \rho v = 0, \tag{1}$$

la principale différence entre les trois exemples cités étant l'expression de la vitesse. Citons le

travail de D. Armbruster, P. Degond et C. Ringhofer [2] qui présente une modélisation continue par une loi de conservation du flux d'information en considérant les temps mis par une requête pour aller d'un état de traitement à l'état suivant.

Nous étendons dans cette thèse les travaux de De Vuyst notamment en ce qui concerne le calcul des temps de service. Nous appliquons ensuite ce principe de calcul des temps de service à un problème d'optimisation dynamique des ressources d'un système composé de deux serveurs afin de minimiser le temps d'attente des utilisateurs. Cette partie est organisée comme suit.

Dans le chapitre 1, nous expliquons comment calculer de manière efficace les temps de service moyens par un système d'EDP et d'EDO couplées. A cette occasion, nous rappelons la construction du système d'EDP pour un système multi-tâches. De plus, nous proposons une méthode appropriée pour la discrétisation des équations intervenant dans le calcul des temps de service. Nous proposons dans les deux chapitres suivants de tester la modélisation sur des problèmes d'optimisation en vue d'assurer une certaine qualité de service. Dans le deuxième chapitre, le premier problème d'optimisation proposé concerne le cas de deux types de requêtes dont les temps de traitement doivent rester inférieurs à une durée imposée. Dans le troisième chapitre, la qualité de service concerne essentiellement le premier type de requêtes, le critère pour le deuxième type de requêtes étant considéré comme moins important.

Cette partie a été présentée oralement lors de la conférence IFIP à Sophia Antipolis, en juillet 2003 [10] et a fait l'objet d'un article accepté pour publication dans IJFV [11].

Mots-clés : flux d'information, loi de conservation, système à tâches partagées, modélisation fluide, contrôle optimal.

1

Calcul des temps de service pour un serveur à tâches partagées via un modèle aux EDP

1.1 Modèle continu

Considérons un système capable d'effectuer des tâches (satisfaire à des requêtes d'utilisateurs par exemple) avec une capacité limitée. Nous notons ϕ_o (l'indice o signifie *output* pour sortie) le flux de sortie, c'est-à-dire le taux de requêtes que le système peut traiter par unité de temps. Il est équivalent de considérer un temps de service moyen $D = (\phi_o)^{-1}$ pour une classe donnée de requêtes. Le temps D représente ainsi le temps moyen que met une requête pour être traitée entièrement par le système. Nous supposerons que la capacité du système considéré peut être partagée pour traiter en même temps des requêtes à différents niveaux de traitements. Pour fixer les idées, notons N le nombre de niveaux de traitement du système et numérotons $\frac{k}{N}$, $k = 1, ..., N$ ces niveaux de traitement. Par exemple, dire que la requête est au niveau de traitement $1/N$, $(k = 1)$, signifie qu'elle vient d'arriver dans le système et qu'elle commence à être traitée. De même, dire que la requête est au niveau de traitement N/N, $(k = N)$, signifie que la requête est complétement traitée et que le service est rendu. Supposons qu'au niveau de traitement k/N, le nombre de requêtes est M_k. Lorsque le serveur divise sa capacité de façon homogène (les capacités allouées à chaque niveau de traitement sont égales), le flux instantané de requêtes entre les niveaux k/N et $(k+1)/N$ est

$$\phi_{k,k+1} \;=\; \frac{M_k}{\sum_{l=1}^{N-1} M_l}\,\phi_o, \tag{1.1}$$

de sorte que nous retrouvons la capacité totale lorsque nous considérons tous les flux partiels :

$$\sum_{k=1}^{N-1} \phi_{k,k+1} \;=\; \phi_o.$$

Nous supposons maintenant que le nombre de niveaux de traitements N et le nombre total de requêtes $M = \sum_{l=1}^{N-1} M_l$ (la "masse" de requêtes) sont grands. Nous allons réécrire les équations

d'équilibre dans le cas limite quand N et M tendent tous les deux vers l'infini. Nous introduisons alors une densité de masse $\rho(x,t)$ (nous empruntons volontairement le langage de la mécanique des fluides). Les variables x et t sont respectivement les variables d'espace et de temps. La variable x peut être interprétée ainsi.
Soit $\omega =]\omega_l, \omega_r[$ un intervalle de $]0,1[$. La quantité

$$\int_{\omega_l}^{\omega_r} \rho(x,t)\,dx$$

représente la quantité de requêtes dont le niveau de traitement est entre les niveaux ω_l et ω_r. Le système étant conservatif, nous pouvons écrire une équation d'équilibre sur la masse totale des requêtes entre les niveaux ω_l et ω_r. Cette équation s'écrit

$$\frac{d}{dt}\int_{\omega_l}^{\omega_r} \rho(x,t)\,dx \;=\; (\text{flux en } \omega_l) \;-\; (\text{flux en } \omega_r). \tag{1.2}$$

L'équivalent continu de (1.1) est clairement

$$\frac{\rho(x,t)}{\int_0^1 \rho(y,t)\,dy}\,\phi_o.$$

Nous définissons ainsi un flux local entre les niveaux x et $x+dx$. Par conséquent, l'équation (1.2) est plus précisément

$$\frac{d}{dt}\int_{\omega_l}^{\omega_r} \rho(x,t)\,dx \;=\; \frac{(\rho(\omega_l,t)-\rho(\omega_r,t))\,\phi_o}{\int_0^1 \rho(y,t)\,dy}. \tag{1.3}$$

Cette équation étant vraie pour tout intervalle ω contenant un x donné, nous pouvons faire tendre $mes(\omega)$ vers zéro et nous trouvons ainsi l'Equation aux Dérivées Partielles suivante (EDP) :

$$\partial_t \rho \;+\; \partial_x\left(\frac{\rho\phi_o}{\int_0^1 \rho(y,t)\,dy}\right) \;=\; 0,\; x \in]0,1[,\; t>0. \tag{1.4}$$

Nous rappelons que x est une variable d'espace représentant le niveau de traitement d'une requête courante. Cette équation peut être lue comme une équation de transport

$$\partial_t \rho \;+\; \partial_x \rho v(t) \;=\; 0, \tag{1.5}$$

avec une vitesse dépendant du temps $v(t)$ égale à

$$v(t) \;=\; \frac{\phi_o}{\int_0^1 \rho(x,t)\,dx}. \tag{1.6}$$

Cette équation une fois écrite peut être généralisée au sens faible.

1.2 Domaine de validité du modèle continu et limite stochastique

L'utilisation de ce modèle continu est justifiée lorsqu'il y a suffisamment de requêtes traitées par le système (puisque la construction du modèle implique que le nombre de requêtes tende vers l'infini). De plus, dans ce modèle continu, si la masse totale de requêtes

$$m(t) \;=\; \int_0^1 \rho(x,t)\,dx$$

tend vers zéro, alors la vitesse de propagation de l'information $v(t)$ tend vers l'infini. Nous nous attendrions plutôt dans ce cas à une vitesse de propagation égale à ϕ_o. Le domaine de validité de ce modèle continu est précisément le domaine pour lequel le nombre de requêtes est grand par rapport à 1 et les effets stochastiques sont peu importants. D'un autre côté, il est naturel de penser que si le flux de requêtes en entrée est bruité, alors les effets stochastiques ne sont plus négligeables lorsque m est de l'ordre de 1. Nous aimerions que notre modèle soit encore valable (du moins grossièrement) lorsque le problème considéré se situe à la limite entre le domaine stochastique et le domaine déterministe, car cette configuration apparaît souvent en pratique. Pour éviter la possibilité d'une vitesse de transport infinie, nous donnons une expression plus générale à la vitesse définie par (1.6). Nous introduisons alors dans l'expression de la vitesse une fonction continue croissante φ telle que $\varphi(x) \geq 1$ et $\varphi(x) \sim x$ pour les grandes valeurs de x :

$$v(t) \;=\; \frac{\phi_o}{\varphi\left(\int_0^1 \rho(x,t)\,dx\right)}. \tag{1.7}$$

Ainsi, nous pouvons par exemple utiliser la fonction $\varphi(x) = \max(1, x)$. Dans le cas où m devient plus petit que 1, la vitesse v est exactement ϕ_o et le temps de service ϕ_o^{-1}. Parmi les choix possibles de fonctions pour φ, la famille de fonctions à un paramètre ϵ,

$$\varphi^\epsilon(x) \;=\; \frac{(x+1)+\sqrt{\epsilon^2+(x-1)^2}}{2},\ \epsilon \in \mathbb{R}^+,\ x > 0, \tag{1.8}$$

pour des "petits" ϵ, fournit de bons candidats qui ont l'avantage d'être $\mathcal{C}^\infty$. Il s'agit d'une régularisation de la fonction $\phi(x) = \max(1, x)$. Le paramètre ϵ doit être calibré et optimisé à partir de données réelles et experimentales. Nous nous attendons à trouver une loi qui relie ϵ à des invariants du processus stochastique (par exemple, ϵ pourrait dépendre de la variance du bruit Gaussien du flux d'arrivée). Cependant, nous n'en parlerons pas dans cette thèse et cela pourrait faire l'objet d'un travail futur. Dans ce qui suit, nous considérons donc l'équation suivante comme modèle :

$$\partial_t \rho \;+\; \partial_x \left(\frac{\rho \phi_o}{\max(1, \int_0^1 \rho(x,t)\,dx)} \right) \;=\; 0,\ x \in]0,1[,\ t > 0. \tag{1.9}$$

Nous revenons maintenant sur le calcul de la masse totale $m(t)$ du fluide dans le système. Nous recherchons une équation différentielle d'équilibre concernant cette variable. En intégrant l'équation (1.5) sur l'intervalle $]0,1[$, nous obtenons l'équation différentielle sur m

$$\frac{dm}{dt} \;= \phi_i(t) \;-\; \rho(1,t)v(t), \tag{1.10}$$

qui correspond à un équilibre entre les flux d'entrée et de sortie. Nous pouvons aussi résumer le modèle continu par le système d'équations suivantes :

$$v(t) \;=\; \frac{\phi_o}{\max\left(1, m(t)\right)}, \tag{1.11}$$

$$\partial_t \rho \;+\; \partial_x \rho v(t) \;=\; 0,\ x \in]0,1[,\ t > 0, \tag{1.12}$$

$$\frac{dm}{dt} \;= \phi_i(t) \;-\; \rho(1,t)v(t). \tag{1.13}$$

Pour ce problème avec conditions initiales, nous ajoutons les conditions : $m(0) = m^0 \geq 0$, et $\rho(x,0) = \rho^0(x)$ avec la relation $\int_0^1 \rho^0(x)\,dx = m^0$, $\rho^0 \in BV(0,1)$. L'équation différentielle étant hyperbolique avec une vitesse de transport $v(t)$ positive, nous devons aussi imposer une condition aux bords en amont qui traduit la compatibilité du flux d'entrée $\phi_i(t)$ avec le flux de l'équation (1.12). Il est équivalent de définir des conditions de Dirichlet en amont sur ρ avec

$$\rho(0,t)v(t) \;=\; \phi_i(t). \tag{1.14}$$

1.3 Extension du modèle à un système multi-source

Nous pouvons étendre naturellement ce formalisme lorsque le même serveur à tâches partagées doit traiter en même temps plusieurs sources d'information. Un cas typique d'applications concerne les serveurs web devant traiter plusieurs types de requêtes et de services (consultation de pages web statiques, transactions, commandes...). Dans ce cas, nous considérons K flux d'entrée $\Phi_{i;k}$, $k = 1, ..., K$ (l'indice i désigne ici l'entrée ou input en anglais). Nous noterons le flux d'entrée total

$$\phi_i \;=\; \sum_{k=1}^{K} \phi_{i,k}.$$

Ainsi, en suivant le même procédé que précédemment, nous trouvons un système d'équations aux dérivées partielles d'évolution pour chaque densité partielle ρ_k pour le type de requêtes k :

$$\partial_t \rho_k \;+\; \partial_x \frac{\rho_k \phi_o}{\max(1, \sum_{l=1}^{K} \int_0^1 \rho_l(x,t)\,dx)} \;=\; 0.$$

Les équations sont couplées entre elles par un terme non local qui est la masse totale $m(t)$ en cours de traitement dans le serveur. Cette masse totale est maintenant définie par

$$m(t) \;=\; \sum_{l=1}^{K} \int_0^1 \rho_l(x,t)\,dx.$$

Les conditions de compatibilité concernant la continuité du flux en $x = 0$ donnent les conditions aux bords en amont suivantes

$$\phi_{i;k}(t) \;=\; \frac{\rho_k(0,t)\phi_o}{\max(1, m(t))}.$$

Ces conditions peuvent être interprétées comme des conditions de Dirichlet non homogènes sur chaque densité ρ_k en amont. La vitesse de propagation est la même pour chaque source k ; elle est à nouveau donnée par l'expression (1.11). Nous pouvons aussi reformuler le problème et l'écrire sous la forme d'un système non linéaire couplé d'EDO et d'EDP, de la même manière que (1.11)-(1.13) :

$$v(t) \;=\; \frac{\phi_o}{\max\left(1, m(t)\right)}, \tag{1.15}$$

$$\partial_t \rho_k \;+\; \partial_x \rho_k v(t) \;=\; 0,\; x \in]0,1[,\; t > 0,\; k = 1, ..., K, \tag{1.16}$$

$$\frac{dm}{dt} \;= \phi_i(t) \;-\; \left(\sum_{l=1}^{K} \rho_l(1,t)\right) v(t) \tag{1.17}$$

avec les conditions initiales $m(0) = m^0 \geq 0$ et $\rho_l(x,0) = \rho_l^0(x)$ telles que $\sum_{l=1}^{K} \int_0^1 \rho_l^0(x)\,dx = m^0$.

1.4 Commentaires à propos de l'extension à un réseau de serveurs multi-tâches

Jusqu'à présent, nous avons modélisé le comportement d'un seul serveur multi-tâches avec sa propre gestion de ressources. Mais en général, plusieurs serveurs multi-tâches sont interconnectés et peuvent échanger des informations pour former un réseau d'information. Ainsi, nous avons besoin de prendre en compte les flux aux interfaces à travers tous les couples de noeuds, c'est-à-dire entre deux serveurs donnés. Nous devons alors ajouter des conditions sur la compatibilité des flux d'entrée et de sortie effectifs entre les différents serveurs interconnectés. Les équations étant hyperboliques avec une vitesse de transport positive, nous devons imposer des conditions aux bords en amont pour chaque noeud (chaque serveur). Cela signifie que le flux d'entrée de chaque noeud est donné. D'un autre côté, le flux de sortie (le taux de départ) est une conséquence de l'état interne du noeud. Si un noeud (1) envoie de l'information vers le noeud (2), alors nous pouvons seulement dire que le taux d'arrivée de (2) est le taux de départ de (1), qui dépend de l'état interne de (1).
Pour résumer, nous dirons que les modèles de continuité donnent pour chaque noeud du réseau des EDP ou des systèmes d'EDP (modèle comportemental), tels que nous les avons décrits précédemment alors que la communication entre les noeuds est modélisée par des conditions sur les flux d'entrée.

1.5 Calcul des temps de service

Il est important pour les concepteurs de connaître le temps mis par le serveur pour traiter les requêtes à chaque instant. Nous utiliserons dans la suite la connaissance de ce temps de service pour optimiser un problème de partage des ressources du serveur afin de minimiser le temps d'attente des utilisateurs lorsque par exemple, il y a plusieurs types de requêtes formulées. Il est équivalent de chercher le temps de passage que met chaque "particule" d'information pour aller de $x = 0$ (requête non traitée) à $x = 1$ (requête entièrement traitée). D'un point de vue différentiel (intégration sur les caractéristiques), ce temps dépend à la fois du passé et du futur : quand une requête entre dans le système, l'état du système qui permet de calculer le temps de passage de cette requête dépend du passé à cause des effets de mémoire (les requêtes entrées avant et qui sont encore dans le système utilisent une certaine capacité du système). Ensuite, lorsque cette requête est entrée, sa trajectoire effective dépend des flux d'arrivée futurs. Par exemple, dans un cas extrême, si les requêtes entrent en trop grand nombre, le système peut devenir congestionné et se bloquer (cas de particules stagnantes). Cette dépendance vis-à-vis du passé et du futur est traduite dans le système d'EDP par la présence de la masse totale $m(t)$. Notons par $d(x, t_0, x_0)$ le temps de trajet d'une requête actuellement à la position x alors qu'elle était à la position $x_0 \in]0, 1[$ au temps t_0. En utilisant la représentation lagrangienne, nous devons résoudre le problème différentiel

$$\frac{D}{Dt}d = 1,$$
$$d(0, t_0, 0) = 0,$$

où $\frac{D}{Dt}$ désigne la dérivée particulaire. Or nous savons aussi que la vitesse de la particule au temps t est donnée par $v(t)$. La représentation eulérienne du problème est alors donnée par le problème d'EDP suivant :

$$\partial_t d \ + \ v(t)\,\partial_x d \ = \ 1, \quad x \in]0,1[, \ t > 0, \tag{1.18}$$

$$d(x=0,t) = 0. \tag{1.19}$$

Nous voulons connaître le temps de passage total de la particule dans le système. La variable qui nous intéresse est alors le temps de service $D(t)$ donné par

$$D(t) \ = \ d(1,t). \tag{1.20}$$

Le problème (1.18),(1.19) est un problème de transport avec un terme source constant. Nous pouvons remarquer qu'un simple changement de variable $u = t - d$ permet de reformuler le problème initialement inhomogène en un problème homogène (sans terme source) :

$$\partial_t u \ + \ v(t)\,\partial_x u \ = \ 0, \quad x \in]0,1[, \ t > 0, \tag{1.21}$$

$$u(x=0,t) = t, \tag{1.22}$$

$$D(t) = t - u(1,t). \tag{1.23}$$

Dans ce qui suit, nous appellerons les équations (1.21)-(1.23), "équations du temps de service" (ETS).

1.6 Discrétisation numérique des équations

Nous considérons une discrétisation uniforme du domaine spatial $]0,1[$ avec un pas d'espace $h = \frac{1}{N}$. Notons $x_j = (j - \frac{1}{2})h$, $j = 1, ..., N$, les points du maillage et $x_{j+\frac{1}{2}} = jh$, $j = 0, ..., N$. Nous considérons alors un pas de temps variable τ^n à l'instant t^n et nous définissons l'instant suivant t^{n+1} par $t^{n+1} = t^n + \tau^n$. Notons enfin par

$$\lambda^n = \frac{\tau^n}{h},$$

le quotient des pas de discrétisation spatiale et temporelle à l'instant t^n. Les équations de transport considérées sont hyperboliques avec une vitesse de transport positive. Nous considérons alors un schéma upwind pour des raisons de stabilité. Nous souhaitons approcher la densité ρ par une fonction constante et égale à ρ_j^n sur chaque cellule $I_j^n = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [t^n, t^{n+1}[$. La donnée initiale ρ^0 est approchée par

$$\rho_j^0 = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \rho^0(x)\,dx.$$

Nous choisissons un schéma implicite :

$$\rho_j^{n+1} \ = \ \rho_j^n \ - \ \lambda^n \frac{(\rho_j^{n+1} - \rho_{j-1}^{n+1})\phi_o}{\max(1, h\,\sum_{l=1}^N \rho_l^n)}. \tag{1.24}$$

Nous obtenons alors un schéma inconditionnellement stable (voir [3]). Du point de vue du calcul, nous devons résoudre un système linéaire triangulaire à deux diagonales pour chaque pas de temps. Ce système peut être résolu explicitement si bien que la complexité numérique est de l'ordre de la complexité d'un schéma explicite.
Dans l'expression (1.24), nous avons discrétisé $m(t^n)$ par une série de Riemann tronquée $m^n = h \sum_{l=1}^{N} \rho_l^n$. Nous pouvons nous demander si cette formule quadratique n'est pas trop grossière ou inadaptée. Au contraire, ce choix nous permet d'obtenir un relation de consistance avec l'équation d'équilibre (1.10) au niveau discret. En effet, puisque le schéma est conservatif, nous pouvons sommer sur tous les indices j dans (1.24) pour obtenir l'expression suivante

$$\sum_{j=1}^{N} \rho_j^{n+1} \;=\; \sum_{j=1}^{N} \rho_j^n \;-\; \lambda^n \left(\frac{\rho_N^{n+1} \phi_o}{\max(1, m^n)} - \phi_i^{n+1} \right)$$

c'est-à-dire

$$m^{n+1} \;=\; m^n + \tau^n \left(\phi_i^{n+1} - \frac{\rho_N^{n+1} \phi_o}{\max(1, m^n)} \right)$$

qui est consistante avec l'équation différentielle de bilan de masse (1.10).
Pour résumer, nous programmons la méthode numérique de la façon suivante :

$$v^n \;=\; \frac{\phi_o}{\max(1, m^n)}, \tag{1.25}$$

$$\rho_1^{n+1} \;=\; \rho_1^n - \lambda^n (\rho_1^{n+1}\, v^n - \phi_i^n), \tag{1.26}$$

$$\rho_j^{n+1} \;=\; \rho_j^n - \lambda^n (\rho_j^{n+1} - \rho_{j-1}^{n+1})\, v^n, \;\; j = 2, N, \tag{1.27}$$

$$m^{n+1} \;=\; m^n + \tau^n \left(\phi_i^n - \rho_N^{n+1}\, v^n \right). \tag{1.28}$$

Nous rappelons que le coût en calcul de ce schéma est de l'ordre de celui d'un schéma explicite (aux divisions scalaires près), bien que les équations (1.27) soient implicites, puisque le système linéaire résultant est triangulaire à deux diagonales.

Enfin, la discrétisation numérique des ETS est simple et la résolution numérique rapide. Nous utilisons la formulation homogène (1.21)-(1.23) et proposons le schéma numérique implicite

$$u_1^{n+1} \;=\; t^{n+1}, \tag{1.29}$$

$$u_j^{n+1} \;=\; u_j^n - \lambda^n \left(u_j^{n+1} - u_{j-1}^{n+1} \right) v^n, \;\; j = 2, N, \tag{1.30}$$

$$D^{n+1} \;=\; t^{n+1} - u_N^{n+1}. \tag{1.31}$$

Remarquons que les temps de service sont calculés correctement, même lorsqu'ils sont très petits par rapport au pas de temps courant. Cela rend cette méthode numérique très puissante pour le calcul sur une grande échelle des systèmes dynamiques avec un trafic intense. Dans nos expériences numériques, nous observons que nous pouvons utiliser des nombres CFL plus grands que 10^6 tout en gardant une très bonne estimation des temps de service pour des flux réguliers.
Dans les chapitres suivants, nous appliquons ces schémas numériques à deux exemples de problèmes d'optimisation.

2

Optimisation de la capacité partagée entre deux sources

2.1 Présentation du problème d'optimisation

Nous considérons ici un serveur dont la capacité totale ϕ_o (homogène à un flux) a été partagée suivant les tâches à traiter. Deux sources de requêtes différentes (1) et (2) doivent être traitées séparément par le serveur et peuvent être traitées en parallèle. Nous modélisons cette situation par un réseau de deux serveurs virtuels numérotés (1) et (2). Ces deux serveurs traitent chacun de leur côté les tâches (1) et (2) respectivement.
Cependant, la capacité totale ϕ_o du serveur doit être partagée entre ces deux serveurs virtuels. Nous introduisons une fonction dépendante du temps $\alpha(t) \in [0,1]$ et les capacités locales $\alpha(t)\phi_o$ et $(1-\alpha(t))\phi_o$ allouées respectivement aux tâches (1) et (2). Les équations résultantes sont alors

$$\partial_t \rho_1 + \partial_x \rho_1 v_1(t) = 0, \tag{2.1}$$

$$\partial_t \rho_2 + \partial_x \rho_2 v_2(t) = 0, \tag{2.2}$$

avec des vitesses de transport et des masses respectives

$$v_1(t) = \frac{\alpha(t)\phi_o}{\max(1, m_1(t))},\ m_1(t) = \int_0^1 \rho_1(x,t)\,dx, \tag{2.3}$$

$$v_2(t) = \frac{(1-\alpha(t))\phi_o}{\max(1, m_2(t))},\ m_2(t) = \int_0^1 \rho_2(x,t)\,dx. \tag{2.4}$$

Pour que le problème soit bien posé, il faut chercher une fonction α appartenant à l'espace fonctionnel $BV(0,T)$ à valeurs dans $[0,1]$.
Les temps de service $D_1(t)$ et $D_2(t)$ pour les serveurs virtuels (1) et (2) calculés par les ETS correspondantes sont donnés respectivement par

$$\partial_t d_\ell + v_\ell \partial_x d_\ell = 1,\ x \in]0,1[,\ d_\ell(0,t) = 0,\ D_\ell(t) = d_\ell(1,t) \tag{2.5}$$

pour $\ell = 1, 2$. La vitesse v_ℓ dépend de la fonction α. Ainsi, il est clair que chaque temps de service D_ℓ dépend aussi de α. Dans la suite, nous soulignerons cette dépendance en utilisant la

notation $D_\ell = D_\ell(t;\alpha)$, $\ell = 1, 2$.
Nous aimerions pouvoir partager la capacité ϕ_o entre les deux serveurs virtuels afin de satisfaire une certaine exigence de qualité de service et minimiser le temps de traitement des requêtes du type (1) et du type (2). Plus précisément, nous nous intéressons à un problème de contrôle optimal sur un intervalle de temps $[0;T]$ (l'intervalle de notre expérience). Nous introduisons deux paramètres constants traduisant la qualité de service du système $D_{max;1}$ et $D_{max;2}$. Ce sont les temps de service prescrits (temps de traitement) à ne pas dépasser. Nous cherchons alors la fonction $\alpha \in BV(0,T;]0,1[)$ qui minimise une fonction de coût J définie par

$$J(\alpha) \;=\; \frac{1}{T}\int_0^T \Big\{(D_1(t;\alpha) - D_{max;1})^2 + (D_2(t;\alpha) - D_{max;2})^2\Big\}\, dt. \tag{2.6}$$

Le problème d'optimisation est formulé ici dans un espace de dimension infinie. Pour le traiter numériquement, il est nécessaire de l'approcher par un problème dont le contrôle appartient à un espace de dimension finie. La fonction α est alors remplacée par le vecteur d'optimisation de dimension finie α^h,

$$\alpha^h \;=\; (\alpha^1, \alpha^2, ..., \alpha^M),$$

où α^n, $n = 1, M$ est une approximation de $\alpha(t^{M;n})$ avec $t^{M;n} = \frac{(n-1)T}{M}$. Nous introduisons un operateur $\mathcal{J}$ qui permet de reconstruire des fonctions dépendant du temps à partir de $\alpha^h \in \mathbb{R}^M$. Cet opérateur a par exemple pour ensemble d'arrivée l'espace des fonctions constantes par morceaux et il est tel que pour tout n,

$$\mathcal{J}\alpha_{|[t^{M;n},t^{M;n+1}[} \;=\; \alpha^n,\; n = 1, M-1.$$

La fonctionnelle résultante discrète est alors donnée par J^M,

$$J^M(\alpha) \;=\; \frac{1}{T}\int_0^T \Big\{(D_1(t;\mathcal{J}\alpha) - D_{max;1})^2 + (D_2(t;\mathcal{J}\alpha) - D_{max;2})^2\Big\}\, dt \tag{2.7}$$

et le problème de contrôle optimal est alors

Problème d'optimisation 1 *Trouver le vecteur $\alpha \in [0;1]^M$ qui réalise*

$$\min_{\alpha \,=\, (\alpha^1,\alpha^2,...,\alpha^M)} \; J^M(\alpha).$$

2.2 Présentation de l'algorithme génétique d'optimisation

Pour résoudre le problème d'optimisation 1, nous décidons d'utiliser un Algorithme Génétique (AG). Nous décrivons ici les principes généraux de l'algorithme génétique.
Il s'agit d'un algorithme probabiliste qui s'inspire du principe biologique de la sélection naturelle. L'initialisation se fait en choisissant une population initiale, formée d'individus. Cette population initiale représente la première génération. Chaque vecteur d'optimisation est un individu et chaque individu est représenté par ses chromosomes. Les composantes d'un vecteur d'optimisation donné représentent les gènes de l'individu correspondant. A chaque itération, différents processus sont mis en oeuvre pour créer une nouvelle génération que l'on espère mieux adaptée à la situation

(ici, nous voulons des individus qui rendent la fonctionnelle minimale). Ces processus sont la sélection, les mutations des chromosomes d'un individu et le croisement entre les chromosomes de deux individus.

- Le processus de sélection permet de garder sans aucun changement certains individus pour la nouvelle génération. La probabilité de les garder est d'autant plus grande que ces individus sont mieux adaptés.
- Lors d'une mutation d'un individu choisi au hasard, un de ses gènes subit un changement qui dépend de l'algorithme génétique. Dans notre cas, il s'agira de remplacer un élément de $[0;1]$ par un nombre sélectionné aléatoirement par une loi uniforme sur $[0;1]$.
- Lors d'un croisement, deux parents s'échangent certains de leurs gènes pour donner deux nouveaux individus.

L'algorithme génétique est un algorithme d'optimisation globale.
Nous avons utilisé la boite GAOT (voir [1]) qui est disponible sur le web.

2.3 Expérience et résultats numériques

Pour notre expérience numérique, nous choisissons $\alpha \in \mathbb{R}^{48}$ comme vecteurs d'optimisation. Dans ce cas, les valeurs de $\alpha(t)$ sont rafraîchies toutes les 30 minutes pour une période de 24 heures. La population initiale est composée de 20 individus.Tous les individus de la population initiale (candidats pour le vecteur optimal α) sont choisis aléatoirement. Leurs composantes appartiennent à l'intervalle $[0;1]$. Nous choisissons

$$\Phi_o = 120 \text{ req/sec} \, (D = {\Phi_o}^{-1} = 8.3\ 10^{-3} \text{ sec}), \tag{2.8}$$

$$D_{max;1} = 4 \text{ sec} \tag{2.9}$$

$$D_{max;2} = 3 \text{ sec}. \tag{2.10}$$

La figure 2.1 représente les deux profils de flux d'entrée pour une journée entière d'observation (de 0 heures à 24 heures). Nous avons choisi que le flux d'entrée de la source (1) serait important le matin tandis que celui de la source (2) serait important pendant l'après-midi.
Le deuxième tracé compare la capacité du serveur par rapport à la somme des deux flux d'entrée. Nous remarquons qu'il existe deux périodes pendant lesquelles le système est congestionné (de 10h à 12h30 et de 17h30 à 19h). Enfin, le troisième tracé est le profil du taux optimal α obtenu par l'algorithme d'optimisation. Seulement 60 itérations de l'algorithme génétique ont été nécessaires pour calculer un vecteur très proche du minimum global. Le temps d'optimisation reste ainsi très raisonnable.
Les quatrième et cinquième tracés de la figure 2.1 représentent le profil des temps de service pour les sources (1) et (2) respectivement. Pour la configuration que nous nous sommes donnée, nous pouvons voir que les contraintes concernant la qualité de service sont respectées même si le système est congestionné deux fois dans la journée. Bien entendu, une configuration plus surchargée violerait la qualité de service désirée. Cependant, le procédé d'optimisation trouverait alors une solution qui s'approcherait au mieux de cette qualité.
Il est intéressant de noter que notre modèle peut capturer une solution optimale aussi bien dans le cas d'un flux non congestionné que dans le cas d'un flux congestionné sans problème de transition. En particulier, la simulation calcule des temps de service de l'ordre de $D = 8.3\ 10^{-3}$ seconde (cas

non congestionné) jusqu'à des temps de réponse de l'ordre de 4 secondes (cas congestionné), ce qui représente un rapport de presque 500 entre les deux bornes extrèmes. Le problème d'optimisation peut avoir plusieurs minima et différentes initialisations peuvent conduire à différents minima numériques. Cependant, nous pouvons observer que tous les minima restent proches les uns des autres. Plus précisèment, ces minima sont tous proches de la fonction dépendant du temps :

$$\alpha^c(t) \;=\; \frac{\Phi_1(t)}{\Phi_1(t)+\Phi_2(t)}.$$

Cette fonction α_c correspond au rapport du flux d'entrée de la source (1) sur le flux total au temps t. Par conséquent, cette fonction que nous pouvons considérer comme approchant correctement la solution optimale peut être utilisée pour le contrôle en temps réel de la capacité à allouer à chaque serveur virtuel.

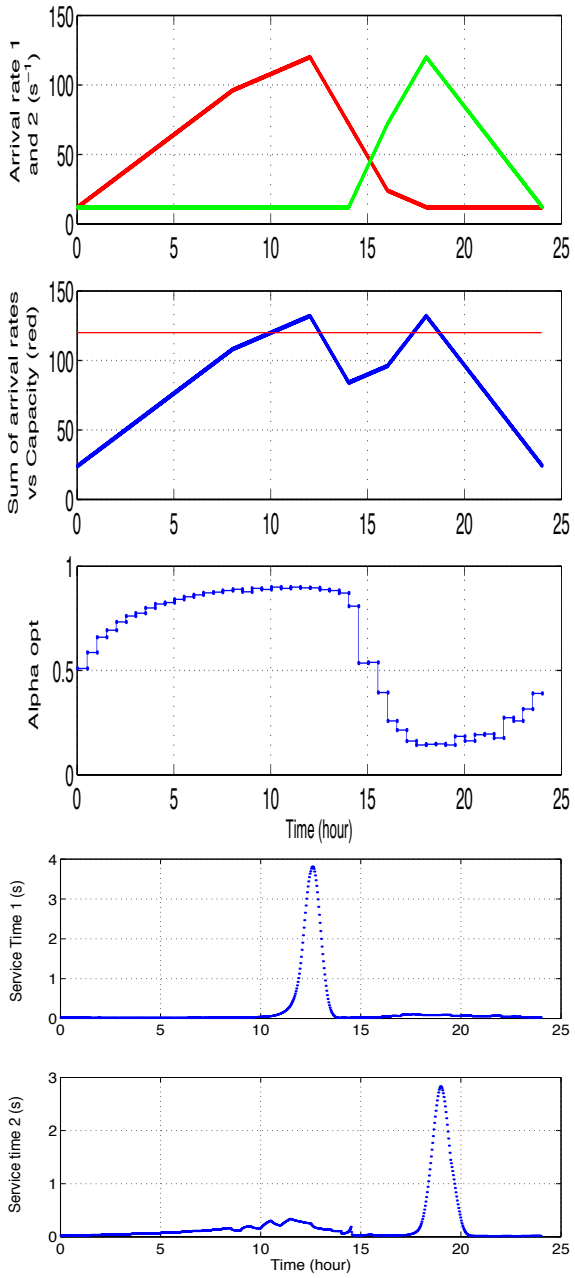

FIG. 2.1 – Optimisation de la capacité partagée entre deux sources. La somme des flux d'entrée comparée au flux de sortie (capacité totale du serveur) montre les deux régions de saturation. Nous avons représenté les profils correspondants de α et les temps de service optimisés pour les sources (1) et (2) pendant une journée.

3

Optimisation de la capacité partagée entre deux sources avec garantie de qualité de service pour la première source

Dans l'exemple du chapitre précédant, nous nous sommes intéressés à l'allocation optimale de la capacité d'un serveur entre deux types de requêtes différentes. Nous devions trouver le partage optimal de la capacité d'un serveur afin de garantir une certaine qualité de service pour les deux types de requêtes. Dans ce chapitre, nous considérons le cas où seule la qualité de service du premier type de requêtes est imposée. La qualité de service du deuxième type de requête n'est ici pas prioritaire.

3.1 Présentation du problème d'optimisation

Nous considérons maintenant le cas où il existe deux sources (1) et (2) telles que la source (1) soit minoritaire par rapport à la source (2). Ainsi les flux $\phi_{i,1}$ et $\phi_{i,2}$ sont tels que $\phi_{i,1} << \phi_o$. Nous n'excluons pas le cas d'un régime générateur de système congestionné : la somme des flux d'entrée $(\phi_{i,1} + \phi_{i,2})$ peut être plus grande que la capacité totale ϕ_o. La première source (1) doit satisfaire la qualité de service exprimant le fait que le temps de service ne doit pas dépasser la constante $D_{max;1}$. De la même manière que dans le problème précédent, nous partageons la capacité du système réel entre deux serveurs virtuels de capacité respective $\alpha(t)\Phi_o$ et $(1 - \alpha(t))\Phi_o$. Les requêtes de la première source sont traitées par le premier serveur virtuel tandis que les requêtes de la seconde source peuvent être traitées à la fois par le premier serveur et le deuxième serveur avec des flux d'entrée valant $\beta(t)\phi_{i,2}(t)$ pour le premier serveur et $(1-\beta(t))\phi_{i,2}(t)$ pour le second. Le problème est schématisé dans la figure 3.1.
La question est ici de garantir la qualité de service sur la première source en donnant une contrainte sur le serveur (1)

$$D_1(t) \ \leq \ D_{max;1} \ \forall t \in]0, T[,$$

tout en acceptant que la capacité du serveur (1), si la contrainte est satisfaite, soit redistribuée au mieux pour la source (2) afin de rendre le temps de service du second serveur le plus petit possible.

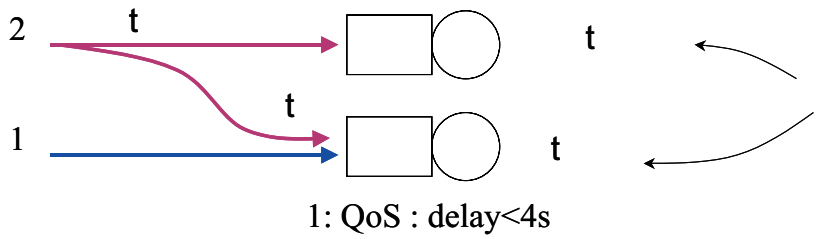

FIG. 3.1 – Description de l'architecture du deuxième problème

Nous décidons ainsi d'introduire la fonctionnelle suivante :

$$J_\sigma(\alpha,\beta) \;=\; \sigma \max_{t\in[0,T]} \left(D_1(t;\alpha,\beta) - D_{max;1}\right)_+ \;+\; \frac{1}{T}\int_0^T \left(D_2(t;\alpha,\beta)\right)^2 \, dt, \tag{3.1}$$

où σ est une constante (suffisamment grande) qui permet au premier terme de la fonctionnelle d'avoir plus d'importance que le second terme. Nous avons relaxé la contrainte de qualité de service en considérant un terme de pénalité important dans la fonctionnelle (3.1). Le second terme de la fonctionnelle a été introduit pour diminuer au mieux le temps de service du serveur (2). Une grande valeur de σ donne la priorité au premier terme de la fonctionnelle et permet ainsi que la qualité de service pour le type de requêtes (1) soit presque surement garantie. L'expression de la fonctionnelle fait intervenir les normes $L^\infty(0,T)$ et $L^2(0,T)$. La fonctionnelle J_σ est très peu régulière à cause de la présence de la norme L^∞. Cela nous dissuade ainsi d'utiliser des méthodes de gradient et nous invite plutot à utiliser à nouveau un algorithme génétique pour le calcul de la solution optimale. Nous avons utilisé la même boîte à outil matlab GAOT [1] pour nos expériences numériques.

3.2 Expérience et résulats numériques

Les vecteurs d'optimisation sont maintenant des vecteurs $\alpha, \beta \in \mathbb{R}^{24}$. Ainsi, les valeurs de $\alpha(t)$ et $\beta(t)$ sont mises à jour toutes les heures. Nous utilisons à nouveau une population initiale de 20 individus. Le flux de sortie est $\Phi_o = 120$ req/sec et la contrainte sur le temps de traitement pour la source (1) est $D_{\max;1} = 4$ sec. Les résultats sont représentés dans la figure 3.2.
Les profils imposés des sources (1) et (2) rendent le système congestionné entre 15h00 et 17h00 (premier et deuxième tracés).
Les troisième et quatrième tracés montrent les profils optimaux des fonctions constantes par morceaux α et β. Les cinquième et sixième tracés représentent les profils des temps de service optimaux pour les serveurs (1) et (2). Nous pouvons remarquer que le temps de traitement des requêtes de la source (1) atteint effectivement la contrainte sur le temps de service de 4 secondes (il est atteint lors d'une période de congestion). D'un autre côté, le temps de service du serveur (2) croît jusqu'à 550 secondes (9 minutes) pendant cette période de congestion du système.

Bien que les conditions de ce test ne soient pas apparemment trop difficiles, des tentatives humaines pour choisir des profils raisonnables des fonctions α et β conduisent souvent à des temps de service de l'ordre de 10000 secondes avec une violation de la qualité de service sur la source (1) ! Cela justifie le besoin de contrôle par des procédés numériques de ces systèmes.

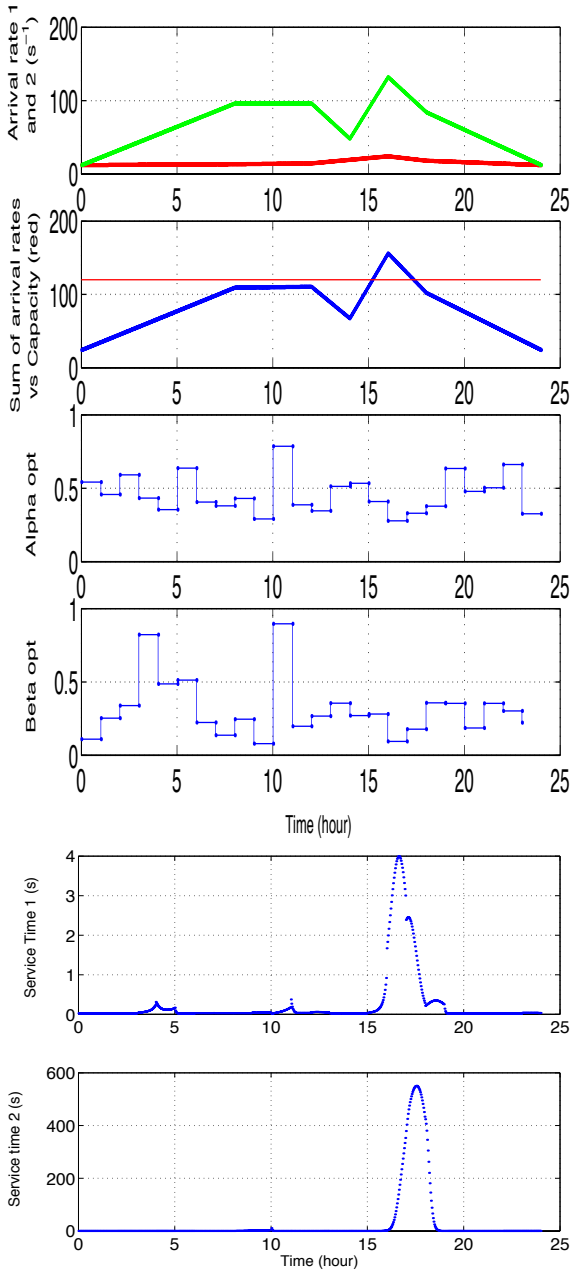

FIG. 3.2 – Optimisation du partage de la capacité d'un système dans le cas de deux sources et de deux serveurs virtuels avec une garantie de service pour la première source. La somme des flux d'entrée comparée au flux de sortie (capacité totale du serveur) montre la région de saturation. Nous avons représenté les profils correspondants de α et β et les temps de service optimisés pour les sources (1) et (2) pendant une journée.

Bibliographie

[1] GAOT, the genetic algorithm optimization toolbox for matlab 5. http ://www.ie.ncsu.edu/mirage/GAToolBox/gaot/.

[2] Armbruster, D., Degond, P., and Ringhofer, C. A model for the dynamics of large queuing networks and supply chains. *SIAP 66*, 3 (2006), 896–920.

[3] Baccelli, F., McDonald, D., and Reynier, J. A mean-field model for multiple tcp connections through a buffer implementing red. Tech. Rep. Tech. Rep. RR-4449, INRIA France, April 2002.

[4] Banks, J., Carson, J., Nelson, B., and Nicol, D. *Discrete Event System Simulation*, 3 ed. Prentice Hall, ISBN : 0130887021, 2000.

[5] Boel, R., and Vuyst, S. D. Prediction based resource allocation, a simulation experiment. In *Proceedings of the COST 279 3rd Management Commitee Meeting* (Leidschendam, The Netherlands, 2002), vol. COST279TD(03)17.

[6] DeVuyst, F. Feasibility of fluid transport modelling for buffer and processing systems : Information fluid dynamics, 2002. http ://www.math.ntnu.no/conservation/2002/009.html.

[7] DeVuyst, F. *Fluid modelling of buffer and processing systems, finite volume discretization.* The International Symposium on Finite Volume for Complex Applications III (FVCA 3). Herbin and Kröner Eds, Hermes Penton Science, 2002.

[8] DeVuyst, F. Service time assessment in computational information fluid dynamics using level set methods, 2002. http ://www.math.ntnu.no/conservation/2002/010.html.

[9] Figueiredo, D., Liu, B., Guo, Y., Kurose, J., and Towsley, D. On the efficiency of fluid simulation of networks. *Computer Networks Journal (to appear)* (2000).

[10] Jaisson, P., and DeVuyst, F. Pde fluid modeling of multithread systems and optimal control. *IFIP, Sophia Antipolis, France, juillet 2003.*

[11] Jaisson, P., and DeVuyst, F. Pde fluid modeling of multithread systems and optimal control. *accepté à IJFV.*

[12] Liu, B., Figueiredo, D., Guo, Y., Kurose, J., Towsley, D., and Gong, W. A study of networks simulation efficiency : Fluid simulation vs. packet-level simulation. In *Proceedings of IEEE INFOCOM 2001* (Apr. 2001).

[13] Liu, B., Guo, Y., Kurose, J., Towsley, D., and Gong, W. Fluid simulation of large scale networks : issues and tradeoffs. In *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'99)* (Las Vegas, 1999), vol. IV, pp. 2136–2142.

Deuxième partie

Assimilation de données pour des modèles de trafic routier et algorithmes d'optimisation associés

Introduction

La prévision du trafic routier à plus ou moins long terme est indispensable aujourd'hui pour la gestion des conditions de circulation urbaine ou extra-urbaine. Pour cela, il est nécessaire de pouvoir modéliser correctement le flux routier. Depuis plus de cinquante, de nombreux modèles ont été développés [14]. Ces modèles peuvent prendre en compte le comportement de chaque véhicule et ses intéractions avec les véhicules voisins (modèles microscopiques, par exemple [4]). D'autres modèles prennent en compte les comportements probabilistes des véhicules (modèles mésoscopiques, par exemple [20], [21]). Enfin, certains modèles sont dérivés de l'analogie entre le flux des véhicules et le flux d'un gaz ou d'un fluide ; ils décrivent le trafic en utilisant des quantités macroscopiques comme la densité et la vitesse (modèles macroscopiques, par exemple [3], [17], [22]). Le modèle de trafic routier choisi va permettre de calculer les conditions de circulation sur une certaine portion de route pendant un certain temps. Mais cela nécessite la connaissance d'un maximum d'observations sur le terrain, par exemple pour fournir des conditions initiales de circulation avant la prévision. Malheureusement, la quantité disponible d'observations est souvent limitée. Il s'agit alors d'un problème d'assimilation de données.

Les procédés d'assimilation de données sont fréquemment utilisés en météorologie et en océanographie. Les méthodes utilisées sont soit séquentielles soit variationnelles. Dans le premier cas, elles reposent sur des estimations statistiques et des modèles probabilistes. On cherche à trouver l'état du système le plus vraisemblable connaissant la dynamique du système, un ensemble d'observations et les lois de probabilité des erreurs portant sur la loi d'évolution ainsi que sur les observations [8]. La deuxième méthode consiste à trouver par contrôle optimal la solution du modèle considéré qui minimise une fonction de coût connue dépendant des observations. En particulier nous citons la méthode adjointe utilisée en météorologie [10], [9] et en océanographie [18].

Dans cette partie, nous nous restreignons aux modèles macroscopiques pour le trafic routier et nous nous intéressons uniquement à la méthode d'assimilation de données variationnelle et plus particulièrement au procédé d'optimisation de la fonction de coût. Nous avons choisi d'utiliser une méthode adjointe.

Il est possible de prévoir les états macroscopiques du trafic routier sur une section de route à l'aide de modèles continus aux EDP qui font l'analogie entre le langage de la physique des fluides et celui du trafic routier. Nous décidons en particulier d'utiliser le modèle hyperbolique du second ordre de Aw-Rascle [3]. Ce modèle fournit de bons résultats qualitatifs et quantitatifs [7], [14] par rapport aux autres modèles antérieurs de Ligthill-Whitham-Richards [17], [22], de Payne-Whitham [19] ou d'autres modèles [16].

Nous désirons ainsi prévoir la densité et la vitesse des véhicules sur une section de route en utilisant le modèle de Aw-Rascle. Nous devons en particulier définir des données initiales adéquates qui serviront au calcul de la solution du système de Aw-Rascle et qui permettront de simuler les conditions de circulation futures. En pratique, les informations concernant les données initiales sont seulement partiellement connues (échantillon de données), voire totalement inconnues. Par exemple, les capteurs ne sont pas installés sur la route de manière continue tandis que des photographies de l'état de la route ne sont prises qu'à des instants particuliers. Il est ainsi possible de ne pas disposer des données voulues à l'instant où nous désirons débuter notre prévision. Dans ce contexte, nous avons besoin de mettre en oeuvre un procédé complexe d'interpolation et d'extrapolation. De plus, même si nous disposons de suffisamment de données à l'instant voulu, il est en général nécessaire de procéder à des vérifications sur ces données. Ainsi, l'utilisation de données antérieures permet de construire la donnée initiale avec plus de sûreté. Nous devons donc assimiler ces données sur un domaine spatio-temporel intégrant la section de route et l'intervalle de temps où sont connues les données observées. De plus, nous devons prendre en compte la dynamique de notre système.

Pour résoudre ce problème d'assimilation de données, nous cherchons la solution du système de Aw-Rascle qui est la plus proche (dans un certain sens que nous définirons) des observations passées et connues. Nous pouvons alors connaître la condition initiale utilisée pour la prévision en calculant les valeurs de la solution précédente à l'instant correspondant au début de la prévision. Ainsi, le problème qui nous intéresse ici est de résoudre un problème de contrôle optimal en utilisant le modèle aux EDP de Aw-Rascle.

La solution du problème hyperbolique de Aw-Rascle est une fonction régulière de la condition initiale et des conditions aux bords. Il semble donc naturel de considérer les conditions initiales et les conditions aux bords comme paramètres de contrôle pour le problème d'optimisation. Il est nécessaire de faire attention aux nombres de variables de contrôle utilisant les conditions aux bords et au nombre d'informations sortant du domaine. Nous introduisons une fonction de coût qui dépend des conditions initiales et aux bords. Elle prendra en compte l'écart entre les données observées et les valeurs calculées. Ensuite, un algorithme d'optimisation utilisant le gradient minimisera cette fonctionnelle afin de donner les conditions initiales et aux bords optimales. Une fois que nous aurons trouvé des conditions correctes, nous pourrons calculer la densité et la vitesse jusqu'au temps initial de notre prévision.

Nous explorerons deux stratégies différentes. La première stratégie est la plus naturelle. Les paramètres d'optimisation seront les conditions initiales et les conditions aux bords en amont. Dans la seconde stratégie, notre algorithme d'optimisation utilisera uniquement les conditions initiales comme paramètres. Cependant, nous devrons considérer une section de route plus longue (en amont) que la portion sur laquelle nous voulons connaître la condition initiale nécessaire à la prévision. Notre analyse montrera quelle longueur de route est suffisante pour obtenir assez d'informations et rendre le problème bien posé.

Nous utiliserons la méthode adjointe pour calculer le gradient de la fonctionnelle nécessaire pour l'algorithme d'optimisation. Lors du calcul numérique du gradient, nous rencontrerons un problème récurrent : le calcul des conditions aux bords semble très important pour avoir une bonne

optimisation. Nous avons donc porté une attention particulière au calcul des conditions aux limites.

Cette partie est organisée comme suit. Dans le premier chapitre, nous expliquons brièvement le contexte de notre problème d'assimilation de données. Nous présentons dans le second chapitre différents modèles de trafic routier avec une attention particulière pour le modèle de Aw-Rascle et son traitement numérique. Dans le troisième chapitre, nous expliquons les deux stratégies utilisées pour notre problème d'optimisation. Dans le chapitre 4, nous utilisons la première stratégie et nous décrivons l'algorithme d'optimisation ainsi que le calcul du gradient de la fonctionnelle par la méthode adjointe. Le cinquième chapitre est relatif à la seconde stratégie. Enfin, le dernier chapitre donne la preuve des résultats utilisés pour le calcul du gradient de la fonctionnelle.

Cette deuxième partie a été présentée oralement lors du congrès IFIP, à Turin en juillet 2005 [15] et fait l'objet d'un article qui est en cours de rédaction.

Mots-clés : trafic routier, assimilation de données, modèle de Aw-Rascle, loi de conservation, contrôle optimal, schéma de Roe.

1

Présentation du contexte

La figure 1.1 présente le problème de prévision et d'assimilation de données que nous nous sommes imposé. Nous désirons prévoir la vitesse et la densité des voitures après un certain temps T sur une portion de route $[0; L]$ (zone 1 de la figure). Pour y parvenir, nous disposons de différentes mesures collectées à des temps et des endroits fixés. Nous imposons une condition aux limites en aval (à droite sur le schéma) : la vitesse à la fin de la portion de route considérée est égale à une fonction supposée connue dépendant du temps. Cette situation peut modéliser par exemple la présence d'un feu tricolore en $x = L$. La donnée d'une fonction déterminant le flux au début de la portion de route ($x = 0$) nous donne une condition aux limites en amont. Cette fonction n'est pas connue et devra être calculée. Nous verrons dans la suite que la donnée de ces deux conditions aux bords ainsi que la donnée des conditions initiales ($t = 0$) suffisent pour calculer la densité et la vitesse des voitures sur le domaine $Q = [0; L] \times [0; T]$.

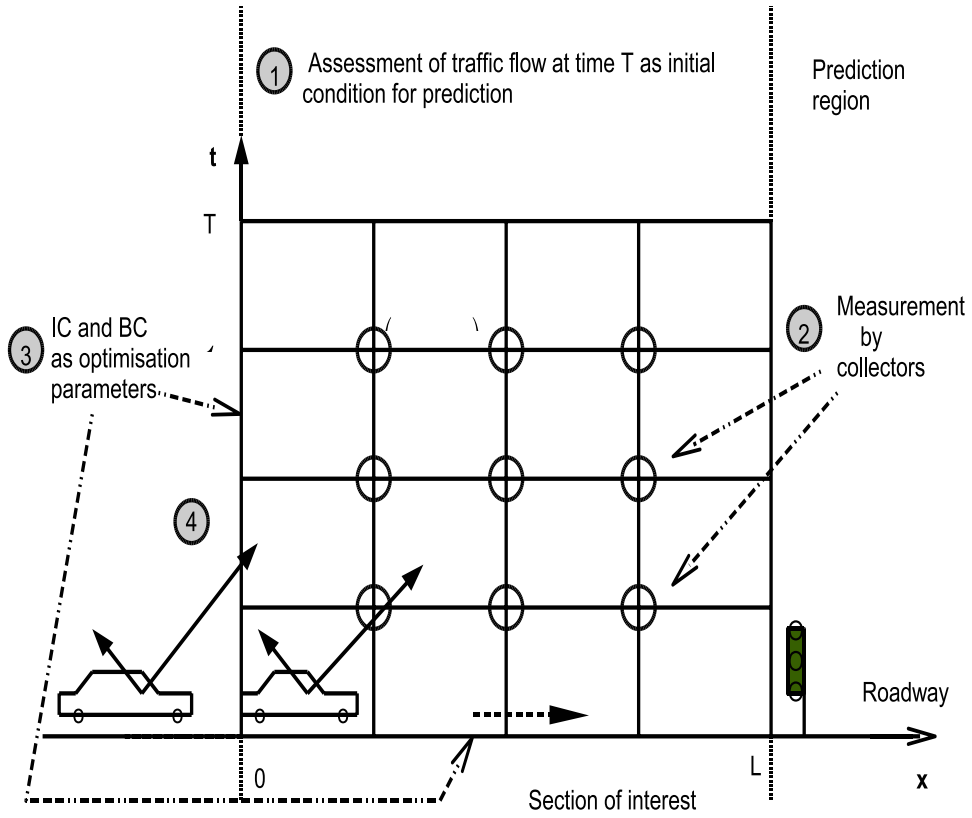

FIG. 1.1 – Problème d'assimilation de données : les symboles "o" représentent les points de mesures en (x_j, t^n). Les quantités observables en trafic routier sont le débit et le taux d'occupation. A partir du taux d'occupation, on peut déduire une densité moyenne des véhicules.

2

Quelques exemples de modèles de trafic routier

2.1 Modèles du premier ordre

Les premiers modèles de trafic routier sont des modèles du premier ordre. Ils ne font intervenir qu'une seule équation à une inconnue ρ qui représente la densité du trafic routier (nombre de véhicules par unité de longueur). Cette équation est une loi de conservation scalaire sur les véhicules. Des expériences ont montré que, sous certaines conditions, le comportement global des conducteurs pouvait être décrit par une vitesse moyenne connue. La vitesse des voitures V est alors donnée par une fonction arbitraire de ρ :

$$V(\rho) = \frac{Q(\rho)}{\rho}$$

où Q désigne le débit. Le modèle de Lighthill-Whitham-Richards (modèle LWR, [17], [22], [24]) est apparu dans les années 50 (1955 et 1956). L'équation de continuité intervenant dans ce modèle peut s'écrire

$$\partial_t \rho + \partial_x(\rho\, V(\rho)) = 0. \tag{2.1}$$

Dans la réalité, plus le trafic se densifie, plus la vitesse des véhicules diminue. La vitesse V est donc une fonction positive et décroissante de ρ. Lorsque la densité est faible ($\rho = 0$), la vitesse est maximale. Inversement, lorsque la densité maximale est atteinte ($\rho = \rho_{emb}$) lors d'un embouteillage où les voitures sont pare-chocs contre pare-chocs, la vitesse est nulle. Ainsi la fonction Q vaut zéro lorsque $\rho = 0$ et $\rho = \rho_{emb}$. On a représenté l'allure typique du débit dans la figure 2.1. Dans ce modèle, la vitesse de propagation des ondes, c'est-à-dire de l'information perçue par les conducteurs, est égale à $Q'(\rho) = V(\rho) + \rho V'(\rho)$. La vitesse V étant décroissante, l'information voyage moins vite que les véhicules.

Cependant, ce modèle présente certains inconvénients majeurs [7], [14]. Par exemple, il ne rend pas compte correctement du comportement des conducteurs dans certaines conditions de circulation, comme la présence de feux tricolores ou une circulation peu dense. De plus, lorsque la densité des véhicules change instantanément (onde de choc due à un embouteillage) le modèle LWR prévoit que la vitesse change aussi instantanément ; cela n'est pas en accord avec la réalité. Les modèles de second ordre ont alors été introduits pour tenter d'éliminer ces inconvénients.

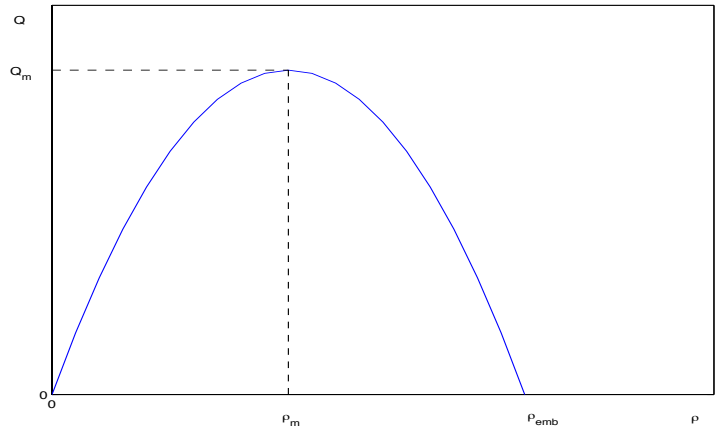

FIG. 2.1 – Débit en fonction de la densité dans le modèle LWR.

2.2 Les "anciens" modèles du second ordre

Les modèles du second ordre font intervenir deux équations différentielles et deux inconnues : la vitesse n'est plus une fonction connue de ρ mais devient dans ces modèles une inconnue. Les variables primitives sont alors la densité ρ et la vitesse v. Nous présentons ici le modèle de Payne [19] et Whitham [24] apparu dans les années soixante-dix :

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \partial_t v + v\partial_x v + \dfrac{1}{\rho}\partial_x P = \dfrac{1}{\tau}(V(\rho) - v) + \nu\partial_x^2 v. \end{cases} \tag{2.2}$$

Le terme P est une fonction de ρ, les quantités τ et ν sont des constantes. Pour reprendre le vocabulaire issu de la mécanique, ces équations sont respectivement appelées équation de continuité et équation de conservation du moment. Cependant, les systèmes du second ordre présentent encore des inconvénients importants pour la modélisation du trafic routier. Par exemple, l'analyse de certains problèmes montre que leurs résolutions conduisent parfois à des vitesses négatives. De plus, on peut montrer que l'information se propage à des vitesses valant respectivement $v - c$ et $v + c$, où c est une constante positive. Ainsi, la vitesse de propagation de l'information pourrait être plus élevée que celle des véhicules. Or il est clair que le comportement des conducteurs est essentiellement influencé par ce qui se passe devant eux (par exemple, un embouteillage qui se créerait après le passage d'un véhicule ne doit pas avoir d'influence sur lui). Daganzo [7] décrit en détails les inconvénients des systèmes du second ordre. Ces inconvénients seraient tels que les modèles du second ordre antérieurs à son article n'améliorent pas véritablement les modèles du premier ordre.

C'est pourquoi, Aw et Rascle [3] ont récemment proposé un nouveau modèle du second ordre qui permet d'éviter ces inconvénients.

2.3 Présentation du modèle de Aw et Rascle

Le système de deux équations différentielles aux deux inconnues ρ et v s'écrit dans leur modèle :

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \partial_t(v + P(\rho)) + v\partial_x(v + P(\rho)) = C\frac{\rho}{T_r}(V(\rho) - v) \end{cases} \tag{2.3}$$

où C est une constante qui vaut 0 dans le cas homogène. Ici le terme P est une fonction de la densité ρ. Aw et Rascle ont montré que leur modèle était qualitativement plus acceptable que les modèles du second ordre précédents. En particulier, la vitesse et la densité restent bien positives au cours du temps. De plus, l'information se propage à une vitesse inférieure à celle des véhicules. Aw et Rascle ont montré une autre propriété interessante de leur modèle : leur modèle macroscopique est la limite d'un modèle microscopique "Follow The Leader" [2].

Le système hyperbolique non homogène de Aw-Rascle

Le modèle de Aw-Rascle non homogène peut s'écrire sous la forme d'un système hyperbolique conservatif avec un terme source [3] :

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, \\ \partial_t(\rho w) + \partial_x(v\rho w) = C\frac{\rho}{T_r}(V(\rho) - v), \end{cases} \tag{2.4}$$

où ρ désigne la densité et v la vitesse des véhicules. Les variables conservatives sont ρ et ρw avec $w = v + P(\rho)$. Dans ce qui suit, nous considérerons les fonctions P définies sur $]0; +\infty[$ par

$$P(\rho) = \frac{v_{ref}}{\gamma}\left(\frac{\rho}{\rho_m}\right)^{\gamma} \text{ pour } \gamma > 0, \text{ modèle } \mathbf{M}_\gamma \tag{2.5}$$

et

$$P(\rho) = v_{ref}\ln\left(\frac{\rho}{\rho_m}\right) \text{ pour } \gamma = 0, \text{ modèle } \mathbf{M_0}. \tag{2.6}$$

La loi P traduit l'anticipation des conducteurs par rapport à ce qui se passe sur la route devant leurs véhicules. Si l'on s'intéresse à la dérivée de P par rapport à ρ, nous remarquons que le modèle M_0 est la limite du modèle M_γ lorsque γ tend vers 0. Le paramètre γ est une constante dépendant du modèle choisie. La densité maximale ρ_m correspond à la densité d'un embouteillage (où les voitures seraient "pare-chocs contre pare-chocs"). La vitesse v_{ref} est une vitesse de référence donnée.

Nous introduisons le vecteur U composé des variables conservatives

$$U = (\rho, \rho w)^t$$

où l'exposant t signifie transposé. Le système (2.4) peut s'écrire sous la forme vectorielle :

$$\partial_t U + \partial_x f(U) = \sigma(U), \tag{2.7}$$

où $f(U) = (v\rho\,; v\rho w)^t$ est le flux et $\sigma(U) = \left(0, C\frac{\rho}{T_r}(V(\rho) - v)\right)^T$ est le second membre.
L'espace des états admissibles Ω^{ad} est ici

$$\Omega^{ad} = \left\{(\rho, \rho w) \in \mathbb{R}^2 / \, 0 < \rho < \rho_m,\ 0 < v < v_m \text{ où } v = w - P(\rho)\right\} \tag{2.8}$$

où v_m est la vitesse maximale des véhicules. Nous notons $A(U) = \partial_U f(U)$ la matrice jacobienne de f par rapport aux variables conservatives ρ et ρw. Nous obtenons après un rapide calcul

$$A(U) = \begin{pmatrix} -\rho P'(\rho) - P(\rho) & 1 \\ -w^2 - \rho w P'(\rho) & 2w - P(\rho) \end{pmatrix}. \tag{2.9}$$

Le système de Aw-Rascle est strictement hyperbolique sauf pour $\rho = 0$ et $v = 0$. En effet, les valeurs propres de $A(U)$ sont :

$$\gamma_1 = v - \rho P'(\rho) = w - P(\rho) - \rho P'(\rho), \tag{2.10}$$

$$\gamma_2 = v = w - P(\rho). \tag{2.11}$$

Aw et Rascle ont montré dans leur article [3] que la vitesse v restait dans l'intervalle $[0; v_m]$ tandis que la densité ρ restait positive dès que $\rho(x,0) \geq 0$ et que la donnée initiale $v(x,0)$ appartenait à $[0; v_m]$ et à $[0; v_m - P(\rho(x,0))]$. Cette dernière condition est automatiquement vérifiée ici puisque P est une fonction positive. En particulier, la valeur propre γ_2 est toujours positive et $\gamma_1 < \gamma_2$ d'après les choix possibles pour la fonction P. Les vecteurs propres associés respectivement à γ_1 et γ_2 sont

$$r^1 = \begin{pmatrix} 1 \\ w \end{pmatrix} \text{ et } r^2 = \begin{pmatrix} 1 \\ w + \rho P' \end{pmatrix}.$$

Dans la suite, nous décidons d'écrire les valeurs propres sous la forme

$$\gamma_1 = v - c, \quad \gamma_2 = v,$$

avec $c = \rho P'(\rho) > 0$. Nous appellerons c la vitesse du son du fluide. Il n'y a pas de raison physique particulière de nommer cette quantité ainsi, ce vocabulaire est uniquement une référence à la dynamique des gaz. De même, nous dirons que le flux est supersonique lorsque $\gamma_1 > 0$ et nous dirons que le fluide est subsonique lorsque $\gamma_1 < 0$. Enfin, nous dirons qu'un état est sonique lorsque $\gamma_1 = 0$. Il est facile de montrer qu'il existe des configurations qui permettent au flux de rester subsonique. La proposition qui suit nous sera utile pour traiter le problème d'assimilation de données. La proposition 1 donne un exemple de choix possible pour le paramètre v_{ref}.

Proposition 1 *Considérons le modèle* $\mathbf{M_0}$, (2.6). *Si la vitesse de référence* v_{ref} *est telle que*

$$v_{ref} = \mu\, v_m, \quad \mu > 1.$$

alors le flux reste subsonique ($\gamma_1 < 0$).

Preuve. Pour le modèle $\mathbf{M_0}$, (2.6), $\rho P'(\rho) = v_{ref}$. Ainsi $\gamma_1 = v - v_{ref} < 0$ car $v \leq v_m < v_{ref}$.

2.4 Approximation numérique

Dans cette section, nous désirons construire une approximation discrète U^h de la solution U pour le système (2.4). Nous considérons un maillage en espace et en temps constitué de cellules

$I_i^n =]x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[\times]t^n, t^{n+1}[$, $x_{i-\frac{1}{2}} = ih$, $t^{n+1} = t^n + \Delta t^n$ et de points (x_i, t^n). Nous approchons une fonction intégrable f sur chaque cellule I_i^n par sa valeur moyenne sur cette cellule :

$$f_i^n = \frac{1}{|I_i^n|} \int\int_{I_i^n} f(x,t)\, dx\, dt.$$

La notation $|I_i^n|$ désigne la mesure de Lebesgue de la cellule I_i^n. Le pas d'espace h est supposé constant ; le pas de temps $\Delta t^n = t^{n+1} - t^n$ peut être variable. Nous notons

$$\lambda^n = \frac{\Delta t^n}{h},$$

le quotient du pas d'espace et du pas de temps. Nous avons décidé d'utiliser le schéma de Roe [23] comme schéma numérique conservatif du fait de sa simplicité à implémenter.

Schéma de Roe

Le schéma de Roe s'écrit sous sa forme conservative :

$$U_i^{n+1} = U_i^n - \lambda^n \left(\phi_{i+\frac{1}{2}}^n - \phi_{i-\frac{1}{2}}^n \right). \tag{2.12}$$

Son flux numérique est alors

$$\phi_{i+\frac{1}{2}}^n = \phi(U_i^n, U_{i+1}^n)$$

avec

$$\phi(U,V) = \frac{f(U) + f(V)}{2} - \frac{1}{2} |\bar{A}(U,V)|(V-U) \tag{2.13}$$

où $|A|$ est définie par $|A| = R\, \mathrm{diag}(|\gamma_i|)\, R^{-1}$ lorsque A se diagonalise sous la forme $\Gamma = R^{-1} A R$. Dans la suite, nous utiliserons les notations

$$\bar{A}_{i+\frac{1}{2}}^n = \bar{A}(U_i^n, U_{i+1}^n), \quad f_i^n = f(U_i^n).$$

La matrice linéarisée de Roe $\bar{A}(U,V)$ est une matrice qui remplit les trois conditions données par Roe [23] : une conditon de consistance, une condition d'hyperbolicité et une condition permettant d'obtenir un schéma de type Godunov :

$$\forall U \in \Omega^{ad},\ \bar{A}(U,U) = A(U), \tag{2.14}$$

$$\forall (U,U') \in (\Omega^{ad})^2,\ \bar{A}(U,U') \text{ est diagonalisable}, \tag{2.15}$$

$$\forall (U,U') \in (\Omega^{ad})^2,\ \bar{A}(U,U')(U'-U) = f(U') - f(U). \tag{2.16}$$

En particulier, la condition (2.16) permet de réinterpréter le schéma de Roe comme un schéma de Godunov. En effet, en notant $\bar{A}^+$ (*respectivement* $\bar{A}^-$) la partie positive (*respectivement* la partie négative) de $\bar{A}$, alors le schéma de Roe peut s'écrire :

$$U_i^{n+1} = U_i^n - \lambda^n \left(\bar{A}_{i+\frac{1}{2}}^{n,-} (U_{i+1}^n - U_i^n) + \bar{A}_{i-\frac{1}{2}}^{n,+} (U_i^n - U_{i-1}^n) \right). \tag{2.17}$$

Le schéma de Roe est ainsi un schéma "upwind". La condition de stabilité pour ce schéma est contrôlée par une condition classique de Courant-Friedrichs-Lewy (CFL). La proposition suivante fournit un candidat pour la matrice de Roe.

Proposition 2 *La matrice suivante satisfait aux trois conditions* (2.14)-(2.16) *:*

$$\bar{A}(U,U') = \begin{pmatrix} -\frac{\rho P(\rho)-\rho' P(\rho')}{\rho-\rho'} & 1 \\ -ww' - \frac{\rho w+\rho' w'}{2}\frac{P(\rho)-P(\rho')}{\rho-\rho'} & w+w' - \frac{P(\rho)+P(\rho')}{2} \end{pmatrix} \text{ pour } \rho \neq \rho' \tag{2.18}$$

et

$$\bar{A}(U,U') = \begin{pmatrix} -\rho P'(\rho) - P(\rho) & 1 \\ -ww' - \rho\frac{w+w'}{2}P'(\rho) & w+w' - P(\rho) \end{pmatrix} \text{ pour } \rho = \rho'. \tag{2.19}$$

Preuve. Remarquons d'abord que

$$\lim_{U'\to U} \bar{A}(U,U') = \bar{A}(U,U) = A(U).$$

Nous prouvons à présent que $\bar{A}(U,V)$ est diagonalisable. Notons Δ le discriminant du polynôme caractéristique pour la matrice $\bar{A}(U,V)$ avec $U \neq V$. Nous l'écrivons après calculs sous la forme d'une somme de deux carrés dans le cas $\rho \neq \rho'$:

$$\Delta = \rho\rho'\frac{(P(\rho)-P(\rho'))^2}{(\rho-\rho')^2} + \left(\frac{P(\rho)-P(\rho')}{2} - (w-w')\right)^2. \tag{2.20}$$

Ainsi le discriminant est strictement positif si ρ et ρ' sont tous les deux différents de zéro. Nous en déduisons que $\bar{A}$ a deux valeurs propres distinctes qui s'écrivent

$$\begin{aligned} \gamma_1 &= \tfrac{1}{2}\left(-\tfrac{\rho P(\rho)-\rho' P(\rho')}{\rho-\rho'} + w + w' - \tfrac{P(\rho)+P(\rho')}{2} - \sqrt{\Delta}\right), \\ \gamma_2 &= \tfrac{1}{2}\left(-\tfrac{\rho P(\rho)-\rho' P(\rho')}{\rho-\rho'} + w + w' - \tfrac{P(\rho)+P(\rho')}{2} + \sqrt{\Delta}\right). \end{aligned} \tag{2.21}$$

De la même manière, nous obtenons dans le cas $\rho = \rho', w \neq w'$:

$$\Delta = \rho^2 P'(\rho)^2 + (w-w')^2 \tag{2.22}$$

qui est encore différent de zéro.

Remarque 1 *Nous pouvons aussi calculer* $|\bar{A}(U,V)|$ *explicitement. Nous obtenons en effet facilement les vecteurs propres* r_1 *et* r_2 *de* $\bar{A}(U,V)$ *associés aux valeurs propres* γ_1 *et* γ_2 *:*

$$r_1 = \begin{pmatrix} 1 \\ \gamma_1 - a \end{pmatrix}, \; r_2 = \begin{pmatrix} 1 \\ \gamma_2 - a \end{pmatrix}, \tag{2.23}$$

où $a = -\frac{\rho P(\rho)-\rho' P(\rho')}{\rho-\rho'}$ *si* $\rho \neq \rho'$ *et* $a = -\rho P'(\rho) - P(\rho)$ *si* $\rho = \rho'$*. Nous trouvons alors après calculs*

$$|\bar{A}(U,V)| = \frac{1}{\gamma_2-\gamma_1}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{2.24}$$

avec

$$\begin{aligned} a_{11} &= |\gamma_1|(\gamma_2 - a) + |\gamma_2|(-\gamma_1 + a), \\ a_{12} &= |\gamma_2| - |\gamma_1|, \\ a_{21} &= |\gamma_1|(\gamma_1 - a)(\gamma_2 - a) + |\gamma_2|(\gamma_2 - a)(-\gamma_1 + a), \\ a_{22} &= -|\gamma_1|(\gamma_1 - a) + |\gamma_2|(\gamma_2 - a). \end{aligned}$$

Nous présentons ici un autre schéma numérique conservatif.

Schéma à volumes finis à flux caractéristique (VFFC)

Le schéma VFFC (volumes finis à flux caractéristique) a été introduit par Ghidaglia, Kumbaro et Le Coq [12] pour des calculs complexes à volumes finis. En particulier, ce schéma évite le calcul des matrices linéarisées de Roe. Pour ce schéma, le flux numérique s'écrit

$$\phi(U,V) = \frac{f(U)+f(V)}{2} - \frac{1}{2}\,\mathrm{sgn}(\bar{A}(U,V))(f(V)-f(U)) \tag{2.25}$$

où $sgn(A)$ est la fonction signe étendue aux matrices. Elle est définie par

$$sign(A) = R\,\mathrm{diag}(\mathrm{sgn}(\gamma_i))\,R^{-1}$$

lorsque A se diagonalise sous la forme $\Gamma = R^{-1}\,A\,R$, où $sgn(a)$ est la fonction signe usuelle définie pour un réel a. Dans ce cas, la matrice $\bar{A}(U,V)$ est une matrice diagonalisable qui vérifie la seule condition de consistance $\bar{A}(U,U) = A(U)$. Nous considérerons la matrice VFFC suivante

$$\bar{A}(U,V) = A\left(\frac{U+V}{2}\right).$$

Le calcul de la matrice $\bar{A}$ est alors plus simple.

2.5 Traitement numérique des conditions aux bords

Dans cette partie, nous nous restreignons au cas intéressant des conditions aux bords subsoniques. Dans ce cas, il existe une caractéristique entrante qui apporte de l'information venant de l'extérieur et une caractéristique sortante qui fournit de l'information venant de l'intérieur.

Conditions aux bords en amont $x = 0$

Nous introduisons une cellule "fantôme" fictive contenant un état virtuel. Cet état nous permet de calculer un flux à la frontière à l'aide de la formule du flux numérique. L'avantage de cette approche est d'obtenir un flux numérique régulier à la frontière. Notons U_1^n l'état U dans la première cellule du domaine $[x_{-\frac{1}{2}}, x_{\frac{1}{2}}]$, où $x_{-\frac{1}{2}} = 0$. Nous devons trouver un état fantôme U_g^n à gauche de l'état U_1^n, que nous notons $U_g^n = (\rho_g^n, \rho_g^n w_g^n)$. Pour un flux subsonique, nous imposons un débit en amont q^n, de sorte que

$$\rho_g^n v_g^n = q^n.$$

La seconde condition qui permet de définir complètement U_g^n doit être compatible avec le fait qu'une caractéristique doit sortir du domaine. Dans ce contexte, des conditions aux bords fortement non linéaires peuvent être traitées en utilisant les demi-problèmes de Riemann (Dubois et Le Floch [11]). Le problème de Riemann ayant pour données initiales

$$U(x,0) = \begin{cases} U_g \text{ pour } x < 0 \\ U_1^n \text{ pour } x > 0 \end{cases} \tag{2.26}$$

doit avoir une 1-onde sortante et une 2-onde entrante. En général, l'analyse de tels demi-problèmes de Riemann montre qu'il existe une famille de solutions dépendant d'un paramètre.

Pour notre problème, nous utiliserons plutôt une approche simplifiée utilisant une linéarisation locale du système. La compatibilité concernant le nombre d'ondes sortantes est ainsi plus simple à exprimer : les valeurs propres $\gamma_1(\bar{A})$ et $\gamma_2(\bar{A})$ pour la matrice calculée à la frontière $\bar{A}(U_g, U_1)$ doivent être négative et positive respectivement. Nous considérons ici une linéarisation locale de A sous la forme $A(U')$ où $U' = (\rho', \rho' w')$ est l'état à la frontière dépendant de U_g et U_1 avec

$$\rho' = \frac{\rho_g + \rho_1}{2}, \quad w' = v' - P(\rho'), \quad \rho' v' = \frac{\rho_g v_g + \rho_1 v_1}{2}.$$

Ainsi nous cherchons l'état fantôme U_g tel que

$$\gamma_2(U') = \frac{\rho_g v_g + \rho_1 v_1}{\rho_g + \rho_1} > 0 \tag{2.27}$$

$$\gamma_1(U') = \frac{q + \rho_1 v_1}{\rho_g + \rho_1} - \frac{\rho_g + \rho_1}{2} P'\left(\frac{\rho_g + \rho_1}{2}\right) < 0. \tag{2.28}$$

Nous obtenons ainsi le résultat suivant :

Proposition 3 *Pour le modèle* $\mathbf{M_0}$, (2.6), *une densité possible permettant d'avoir un flux subsonique à la frontière est*

$$\rho_g = \max\left(\rho_1, \frac{q + \rho_1 v_1}{v_{ref}} - \rho_1\right). \tag{2.29}$$

Preuve. La première condition (2.27) est toujours vérifiée. La condition (2.28) est obtenue dès que :

$$\frac{q + \rho_1 v_1}{\rho_g + \rho_1} - v_{ref} < 0$$

Ainsi

$$\rho_g > \frac{q + \rho_1 v_1}{v_{ref}} - \rho_1. \tag{2.30}$$

Conditions aux bords en aval $x = L$

Nous recherchons maintenant un état fantôme U_g^n en fonction de l'état U_I^n. La condition en aval imposée est la vitesse du fluide. Nous procédons de la même façon que précédemment. Nous cherchons une linéarisation de la matrice jacobienne locale sous la forme $A(U')$. L'état à la frontière U' tel que

$$\rho' = \frac{\rho_g + \rho_I}{2}, \quad w' = v' - P(\rho'), \quad \rho' v' = \frac{\rho_g v_g + \rho_I v_I}{2},$$

doit vérifier les conditions de compatibilité sur le signe des valeurs propres

$$\gamma_1(U') = \frac{\rho_g v_g + \rho_I v_I}{\rho_g + \rho_I} > 0, \tag{2.31}$$

$$\gamma_2(U') = \frac{\rho_g v_g + \rho_I v_I}{\rho_g + \rho_I} - \frac{\rho_g + \rho_I}{2} P'\left(\frac{\rho_g + \rho_I}{2}\right) < 0. \tag{2.32}$$

Nous obtenons la proposition suivante.

Proposition 4 *Pour le modèle* $\mathbf{M_0}$, (2.6), *si* $v_{ref} > \mu\, v_m$, $\mu > 1$ *alors une densité possible permettant d'avoir un flux subsonique à la frontière est*

$$\rho_g^n = \rho_I^n. \tag{2.33}$$

Preuve. La condition (2.31) est toujours vérifiée. La seconde condition est vérifiée dès que

$$\frac{\rho_g v_g + \rho_I v_I}{\rho_g + \rho_I} - v_{ref} < 0. \tag{2.34}$$

Or $v_g < v_{ref}$ et $v_I < v_{ref}$. Ainsi cette condition est toujours vérifiée.

2.6 Expérimentation numérique

Nous désirons valider les schémas numériques que nous avons programmés avant de les utiliser pour le problème d'optimisation. Nous considérons les cas test proposés dans [2]. Il s'agit d'un problème de Riemann avec une condition initiale ($t = 0$) constituée de deux états constants. Les simulations concerneront le cas homogène $C = 0$ puis le cas non homogène $C = 1$. Les états constants U_L (à gauche, $x < 0$) et U_R (à droite, $x > 0$) sont définis par la donnée des valeurs des variables primitives :

$$\begin{cases} \rho_L = 0.05,\ \rho_L v_L = 0.0025,\ v_L = 0.05 \\ \rho_R = 0.05,\ \rho_R v_R = 0.025,\ v_R = 0.5. \end{cases} \tag{2.35}$$

Le pas d'espace est $h = \frac{1}{40}$ et le pas de temps est choisi afin de satisfaire la condition CFL avec un nombre de Courant inférieur à 1. Nous considérons tout d'abord le modèle $\mathbf{M_0}$, (2.6) avec $\gamma = 0$ et $P(\rho) = 2\ln(\frac{\rho}{\rho_m})$, puis le modèle $\mathbf{M}_\gamma$, (2.5) avec $\gamma = 1$ et $P(\rho) = 6\frac{\rho}{\rho_m}$. Dans le cas non homogène, nous prenons $C = 1$, $T(\rho) = 20$ et

$$V(\rho) = v_m \frac{\pi/2 + \arctan(11\frac{\rho - 0.22}{\rho - 1})}{\pi/2 + \arctan(11 \cdot 0.22)}. \tag{2.36}$$

Résultats numériques sans le terme de relaxation

Les figures 2.2 (schéma de Roe) et 2.3 (schéma VFFC) d'une part et les figures 2.4 (schéma de Roe) et 2.5 (schéma VFFC) d'autre part montrent les solutions au temps final $t = 100$ dans les cas $\gamma = 0$ et $\gamma = 1$ respectivement. Dans ce test, nous pouvons observer la présence d'un choc entropique qui rend compte d'une solution discrète non physique. Cela montre que le schéma de Roe et le schéma VFFC ne sont pas des schémas entropiques. Pour la méthode de Roe, on peut suivre les arguments de Harten [13] en échangeant la matrice $|A|$ par $R\,\mathrm{diag}(\psi_\epsilon(\gamma_i))\,R^{-1}$ où $\psi_\epsilon(x) = \sqrt{x^2 + \epsilon^2}$. Tout se passe comme si nous remplacions les valeurs absolues $|\gamma_1|$ et $|\gamma_2|$ par $\psi_\epsilon(\gamma_1)$ et $\psi_\epsilon(\gamma_2)$ dans (2.24). Nous devons choisir ϵ assez petit, $\epsilon << \gamma_i$ afin d'obtenir une bonne approximation de la valeur absolue des valeurs propres. Or les valeurs propres sont du même ordre de grandeur que la vitesse des véhicules. Ainsi nous pouvons choisir $\epsilon = \frac{v_{max}}{100}$. La solution numérique présentée dans la figure (2.6) ne fait plus apparaître de choc d'entropie. Les résultats sont satisfaisants.

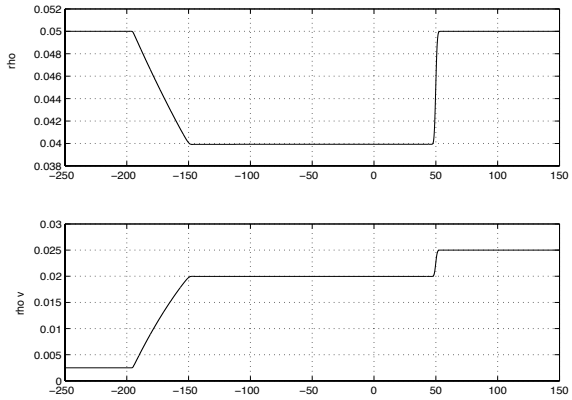

FIG. 2.2 – Schéma de Roe, $\gamma = 0$, système homogène sans relaxation.

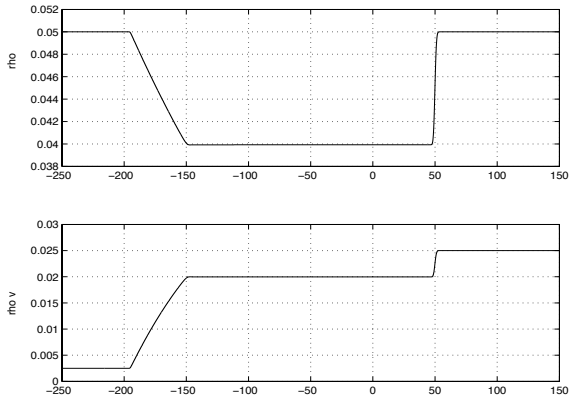

FIG. 2.3 – Schéma VFFC, $\gamma = 0$, système homogène sans relaxation.

Résultats numériques avec un terme de relaxation

Enfin, les figures (2.7) et (2.8) représentent les cas $\gamma = 0$ et $\gamma = 1$ respectivement avec un terme de relaxation pour le schéma de Roe. Les résultats sont là encore très satisfaisants. Dans les paragraphes suivants, nous utiliserons essentiellement le modèle $\mathbf{M_0}$, (2.6) pour plus de simplicité.

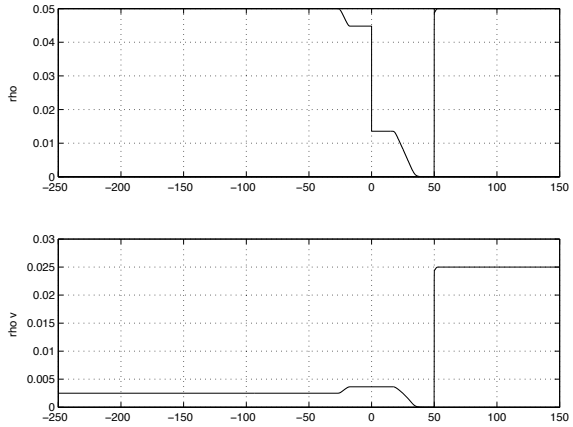

FIG. 2.4 – Schéma de Roe, $\gamma = 1$, système homogène sans relaxation. Un choc d'entropie apparaît.

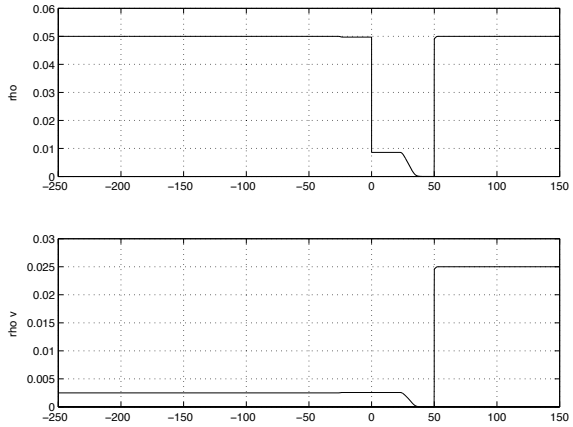

FIG. 2.5 – Schéma VFFC, $\gamma = 1$, système homogène sans relaxation. Un choc d'entropie apparaît.

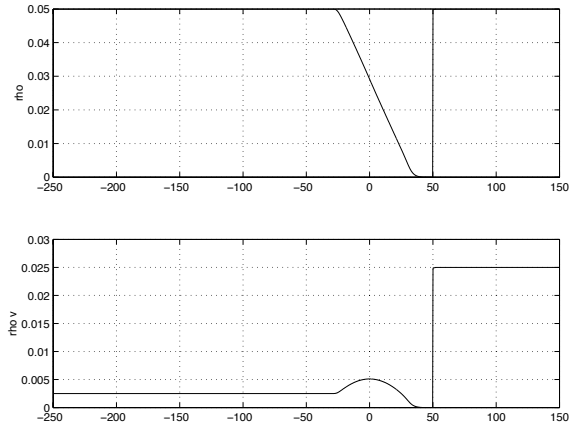

FIG. 2.6 – Schéma de Roe, $\gamma = 1$, système homogène sans relaxation, avec la régularisation de la valeur absolue.

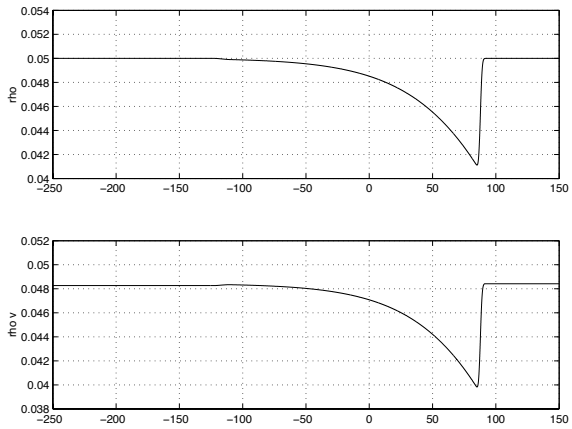

FIG. 2.7 – Schéma de Roe, $\gamma = 0$, système non homogène avec relaxation.

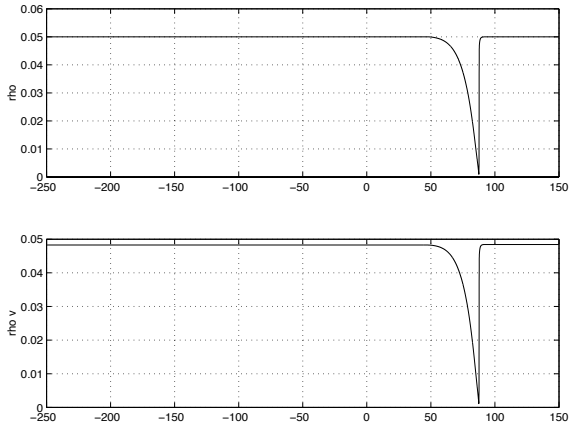

FIG. 2.8 – Schéma de Roe, $\gamma = 1$, système non homogène avec relaxation.

3

Données observables en trafic routier, problème d'assimilation de données

3.1 Données observables en trafic routier

Nous présentons dans ce paragraphe les différentes variables de trafic macroscopiques. Les variables utiles pour les modèles de trafic routier sont la vitesse et la densité. Cependant les données observables en réalité sont le débit et le taux d'occupation.

3.1.1 Le débit

Le débit moyen $q(x, t_1, t_2)$ au point d'abscisse x et entre les instants t_1 et t_2 est défini par

$$q(x, t_1, t_2) = \frac{n(x, t_1, t_2)}{t_2 - t_1} \tag{3.1}$$

où $n(x, t_1, t_2)$ désigne le nombre de véhicules passés en x entre les instants t_1 et t_2.
En pratique, le débit peut être déterminé par des comptages sur la route.
Il est nécessaire de connaître le débit $q(x, t)$ en tout point d'abscisse x et en tout instant t pour les modèles "fluide" de trafic routier. On définit alors

$$q(x, t) = \lim_{\Delta t \longrightarrow 0} q(x, t - \frac{\Delta t}{2}, t + \frac{\Delta t}{2}). \tag{3.2}$$

En pratique, pour ne pas avoir de débit infini, on pose

$$q(x, t) = q(x, t - \frac{\Delta t}{2}, t + \frac{\Delta t}{2}) \tag{3.3}$$

pour de petites valeurs de Δt.

3.1.2 La densité

La densité moyenne des véhicules $\rho(x_1, x_2, t)$ entre les points d'abscisses respectives x_1 et x_2 et à l'instant t est définie par

$$\rho(x_1, x_2, t) = \frac{n(x_1, x_2, t)}{x_2 - x_1}, \tag{3.4}$$

où $n(x_1, x_2, t)$ est le nombre de véhicules présents sur la section de route entre les points d'abscisses respectives x_1 et x_2 et à l'instant t. La densité moyenne peut être obtenue par photographie aérienne ou par caméra vidéo. Comme pour le débit, il est nécessaire pour les modèles continus de trafic routier de définir la densité $\rho(x,t)$ en tout point d'abscisse x et à tout instant t. On définit alors :

$$\rho(x,t) = \lim_{\Delta x \longrightarrow 0} \rho(x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2}, t). \tag{3.5}$$

En fait, on pose

$$\rho(x,t) = \rho(x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2}, t). \tag{3.6}$$

pour de petites valeurs de Δx.

3.1.3 Taux d'occupation

Le taux d'occupation τ est une grandeur sans dimension. Le taux d'occupation est mesuré à l'aide de capteurs (boucles magnétiques) enfouis dans la chaussée et sensibles aux variations du champ magnétique produites par le passage des masses métalliques des véhicules. Il est défini par le rapport du temps durant lequel la boucle est occupée divisé par une durée de référence. Lorsque la durée de référence est Δt, c'est-à-dire la durée utilisée pour le calcul du débit $q(x,t) = q(x, t - \frac{\Delta t}{2}, t + \frac{\Delta t}{2})$, on obtient la relation

$$\tau = (L + l)\rho \tag{3.7}$$

où L est la longueur moyenne des véhicules et l est la longueur de la boucle. Cette variable est souvent utilisée car les procédés de mesure du taux d'occupation sont moins complexes et moins coûteux que les procédés de mesure de la densité.

3.1.4 Vitesse du flot

La vitesse du flot est alors obtenue par la relation :

$$v(x,t) = \frac{q(x,t)}{\rho(x,t)}. \tag{3.8}$$

Ainsi, les données expérimentales observées sont le débit et le taux d'occupation. On obtient ensuite par les relations (3.7) et (3.8) la densité ρ et la vitesse v.

3.2 Problème d'assimilation de données

Dans cette partie, nous nous intéressons plus particulièrement à la mise en place du problème d'assimilation de données : nous supposerons connaître des valeurs de densité et de vitesse en des points particuliers de la portion de route et à des instants donnés. Ces valeurs peuvent être le résultat de mesures réelles observées. Nous voulons trouver les meilleures données initiales et les meilleures conditions aux bords permettant de calculer une solution continue du système de Aw-Rascle qui serait le plus en accord avec les valeurs observées de densité et de vitesse. Nous rappelons l'objectif sous-jacent : nous désirons obtenir des valeurs pour la densité et la vitesse à

chaque point du maillage du domaine spatial $[0; L]$ au temps final T. Ces valeurs calculées pourraient alors servir comme données initiales pour des simulations dans le but de prévoir l'évolution du trafic après l'instant T par exemple.

La densité et la vitesse des véhicules sont calculées à l'aide du système de Aw et Rascle. Nous faisons l'hypothèse qu'une condition aux bords à la fin de la section de route est donnée et connue précisément. Nous supposons ainsi que la condition aux bords en aval est donnée par la connaissance d'une fonction dépendant du temps représentant la vitesse à la fin de la route. Cela peut modéliser par exemple le fait que la vitesse des véhicules est imposée par la présence d'un feu tricolore.
Le problème est alors le suivant :

Problème d'optimisation 2 *(Assimilation de données).*
Etant donné un ensemble d'observations $(\rho_j^n, v_j^n)_{j,n}$ *en des points* $(x_j, t^n)_{j,n}$ *du domaine* $[0; L] \times [0; T]$, *trouver les conditions initiales* (ρ^i, v^i) *(l'indice* i *signifie "initial"), et éventuellement la condition aux bords en amont* $t \mapsto q(t)$ *telles que la solution* $U(x,t) = (\rho, \rho w)^t$ *du système de Aw-Rascle*

$$\begin{cases} \partial_t \rho + \partial_x(v\rho) = 0, \\ \partial_t(\rho w) + \partial_x(v\rho w) = 0, \\ (\rho v)(0,t) = q(t),\ v(L,t) = v_d(t), \\ \rho(x,0) = \rho_i(x),\ v(x,0) = v_i(x). \end{cases} \tag{3.9}$$

minimise une certaine fonction de coût J, *qui mesure la différence entre les valeurs calculées de la densité et de la vitesse en* $(x_j, t^n)_{j,n}$ *de cette solution et les valeurs observées* $(\rho_j^{obs}, v_j^{obs})_j$ *en ces mêmes points.*

Nous expliquerons plus tard comment ce problème peut être bien posé.

Nous avons mis en oeuvre deux stratégies différentes pour assimiler ces données observées afin de rechercher des conditions initiales et des conditions aux bords dans le cas d'un système hyperbolique.
La première stratégie est sans doute la plus naturelle. Nous considérons à la fois les conditions initiales (ρ^i, v^i) définies sur $[0; L]$ et la condition aux bords dépendant du temps q définie sur $[0; T]$ comme variables de contrôle. Le champ calculé (ρ, v) au temps final $t = T$ doit être une fonction continue de (ρ^i, v^i) et de q.
Dans la deuxième stratégie, nous décidons de considérer uniquement les conditions initiales (ρ^i, v^i) comme variables de contrôle. Il est évident que nous devons remplacer la condition en amont par une autre information pour espérer avoir à nouveau un problème bien posé. Ainsi nous étendons le domaine spatial en amont de notre section de route représentée par $[0, L]$ à un domaine $[-L', L]$. La longueur L' doit être assez grande pour fournir suffisamment d'informations. Dans cette stratégie, nous avons uniquement besoin de retrouver la condition initale (étendue) puisque la condition aux bords à $x = 0$ est intégrée dans la condition initiale. Cette seconde stratégie est applicable grâce à la nature hyperbolique du modèle EDP sous-jacent.
La figure 3.1 a) correspond à la première stratégie. Elle représente l'information provenant des

conditions initiales et des conditions en amont. La figure 3.1 b) correspond à la deuxième stratégie. Elle représente l'information qui provient uniquement de la condition initiale. La condition en amont devient inutile.

Chaque approche conduit à une mise en oeuvre numérique spécifique. Dans les deux parties qui

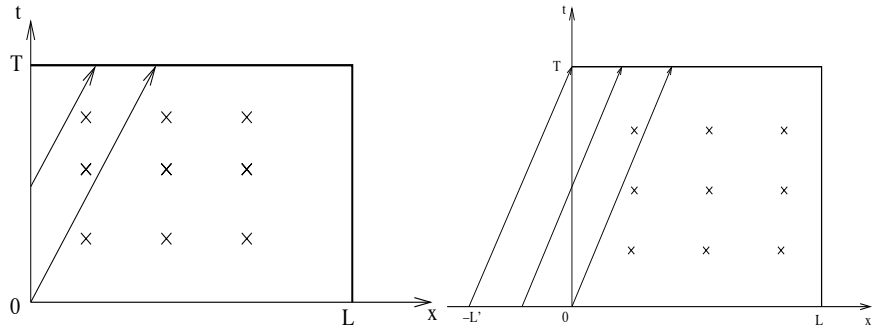

FIG. 3.1 – Les symboles "×" désignent les points de mesures en (x_j, t^n). a) Première stratégie : information provenant à la fois des conditions initiales sur $]0, L[$ et des conditions en amont sur $[0; T]$; b) Deuxième stratégie : information provenant uniquement des conditions initiales sur $]-L', L[$.

suivent nous présentons successivement les deux stratégies ainsi que des expériences numériques.

4

Procédé d'optimisation sur les conditions initiales et aux bords

Nous considérons le système homogène de Aw-Rascle sur le domaine $[0;L]\times[0;T]$ avec des conditions limites aux bords (CL) à chaque frontière :

$$\begin{cases} \partial_t\rho + \partial_x(v\rho) = 0, \\ \partial_t(\rho w) + \partial_x(v\rho w) = 0, \\ +CL: \ (\rho v)(0,t) = q(t),\ v(L,t) = v_d(t), \\ +CI: \ \rho(x,0) = \rho_i(x),\ v(x,0) = v_i(x). \end{cases} \tag{4.1}$$

Le système de Aw-Rascle est hyperbolique avec deux valeurs propres $\gamma_1 = v - c$ et $\gamma_2 = v$. Ainsi les conditions aux bords en amont données dans (4.1) imposent au flux d'être subsonique. Dans le cas contraire, une deuxième condition en amont serait nécessaire. De même, les conditions aux bords en aval imposent au flux d'être subsonique. Sinon, il n'y aurait pas de conditions en aval. Nous rappelons qu'avec le choix de la fonction P selon le modèle $\mathbf{M_0}$, (2.6) du chapitre 2, il suffit de choisir $v_{ref} = \mu\, v_m$, $\mu > 1$ pour que le flux reste subsonique (proposition 1). Nous nous placerons désormais dans ce cas. Ainsi, le problème (4.1) est bien posé.

4.1 Choix d'une fonctionnelle de coût

Nous désirons mesurer la différence entre les valeurs calculées par la résolution du système de Aw-Rascle et les valeurs observées. Nous introduisons ainsi la fonctionnelle de coût suivante

$$J^N = \frac{1}{2}\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\rho((x,t)_j) - \rho_j^{obs}}{\rho^c}\right)^2 + \frac{1}{2}\frac{1}{N}\sum_{j=1}^{N}\left(\frac{v((x,t)_j) - v_j^{obs}}{v^c}\right)^2. \tag{4.2}$$

Les densités ρ_j^{obs} et les vitesses v_j^{obs} désignent les quantités observées en N points $((x,t)_j)$ du domaine spatio-temporel. Les termes ρ_c et v_c sont des densités et des vitesses de référence choisies pour rendre les quantités sans dimension. Par exemple, nous pouvons prendre la moyenne des densités observées pour ρ_c et la moyenne des vitesses observées pour v_c. Nous voulons minimiser J^N en considérant que cette fonctionnelle dépend de variables d'optimisation qui sont d'une part

les données initiales $(\rho^i,\ v^i)$ et d'autre part les conditions aux bords en aval q.
Nous désirons prendre en compte les contraintes sur le signe des variables d'optimisation : les quantités ρ^i, v^i et q sont des fonctions positives. Nous écrivons alors ces fonctions comme les images des fonctions α, β et γ définies par

$$\rho^i = \rho_c\,\phi_\epsilon(\alpha),\quad v^i = v_c\,\phi_\epsilon(\beta),\quad q = q_c\,\phi_\epsilon(\gamma). \tag{4.3}$$

où ρ_c, v_c et q_c sont des quantités caractéristiques que nous prenons égales aux quantités de références utilisées dans l'expression de la fonctionnelle. La fonction ϕ_ϵ est un $\mathcal{C}^1$-difféomorphisme de $\mathbb{R}$ dans $]0;+\infty[$. La valeur de $\phi_\epsilon(x)$ doit être pratiquement égale à x lorsque $x > 0$ pour permettre à la contrainte de positivité d'être relaxée. Pour $x < 0$, $\phi_\epsilon(x)$ doit être proche de zéro. Comme exemple de telles fonctions, nous pouvons donner les fonctions appartenant à la famille à un paramètre

$$\phi_\epsilon(x) = \frac{x + \sqrt{\epsilon^2 + x^2}}{2}. \tag{4.4}$$

Il s'agit en fait d'une régularisation de la fonction $x \mapsto \max(0, x)$. Pour des raisons mathématiques, il est intéressant et parfois plus simple de travailler avec une version continue de la fonctionnelle. Une fois la méthode d'interpolation choisie (qui serait à définir), nous pouvons écrire la fonctionnelle sous la forme d'une intégrale sur le domaine $Q =]0, L[\times]0, T[$. Ainsi nous interpolons les vecteurs d'observation ρ^{obs} et v^{obs} afin d'obtenir deux fonctions d'observation $\mathcal{I}\rho^{obs}$ et $\mathcal{I}v^{obs}$ définies sur Q. La version continue de la fonctionnelle est ainsi :

$$\begin{aligned} J(\alpha,\beta,\gamma) = \frac{1}{2}\frac{1}{|Q|}\int_Q \left(\frac{\rho(\alpha,\beta,\gamma;x,t) - \mathcal{I}\rho^{obs}}{\rho^c}\right)^2 \mathrm{d}x\mathrm{d}t \\ + \frac{1}{2}\frac{1}{|Q|}\int_Q \left(\frac{v(\alpha,\beta,\gamma;x,t) - \mathcal{I}v^{obs}}{v^c}\right)^2 \mathrm{d}x\mathrm{d}t. \end{aligned} \tag{4.5}$$

En notant $X = (\alpha, \beta, \gamma)$ le vecteur des variables d'optimisation, le problème de minimisation que nous devons résoudre s'écrit :

Problème d'optimisation 3 *Trouver $\tilde{X}$ tel que*

$$J(\tilde{X}) = \min_{\substack{(\alpha,\beta)\in\mathcal{C}^1([0,L]),\ \gamma\in\mathcal{C}^1([0,T]),\\ (\alpha,\beta,\gamma)\ \textit{donné par}\ (4.3),\, U=(\rho,\rho w)\ \textit{solution de}\ (4.1)}} J(X). \tag{4.6}$$

4.2 Calcul du gradient de J

D'un point de vue numérique, le problème 3 de minimisation peut être résolu en utilisant des méthodes de gradient (gradient à pas optimal, gradient conjugué), des méthodes de quasi-Newton ou encore des méthodes de région de confiance ([1], [5] ou [6]). Il n'y a pas a priori de raison pour que la fonctionnelle (4.5) soit convexe par rapport à (α, β, γ). Ainsi J peut admettre éventuellement plusieurs minima locaux. Les méthodes numériques que nous utiliserons doivent être capables de repérer le minimum global parmi les minima locaux. Une recherche linéaire peut alors être greffée aux méthodes de gradient ou de quasi-Newton. Dans le cas des méthodes de régions de confiance, nous parlerons de recherche curviligne (voir [5] ou [6]). Nous présenterons

brièvement ces différentes méthodes dans la suite.
Dans tous les cas, il est nécessaire de calculer le gradient de la fonctionnelle J ou du moins une bonne représentation locale de ∇J. Dans les deux prochains paragraphes, nous expliquons la méthode d'état adjoint et d'équation adjointe que nous avons utilisée pour approcher le gradient de J ([10]).

Equation adjointe

Nous rappelons que ρ_i, v_i et q sont des fonctions des variables α, β et γ respectivement. La solution U est alors une fonction de ces trois variables. Nous notons $U(\alpha,\beta,\gamma)$ la solution du problème primal (PP)

$$(PP)\begin{cases} \partial_t U + \partial_x(f(U)) = 0, \\ \rho v(0,t) = q(t),\ v(L,t) = v_d(t), \\ U(x,0) = U_i(x), \end{cases} \tag{4.7}$$

où $f(U) = v\,U$ est le flux du fluide. Pour une perturbation donnée des variables de contrôle, la variation $\delta U = U(\alpha+\delta\alpha,\beta+\delta\beta,\gamma+\delta\gamma) - U(\alpha,\beta,\gamma)$ de la solution U par rapport à la variation des paramètres est solution au premier ordre du problème primal linéarisé (PPL)

$$(PPL)\begin{cases} \partial_t \delta U + \partial_x(A(U)\delta U) = 0, \\ \delta(\rho v)(0,t) = \delta q(t),\ \delta v(L,t) = 0, \\ \delta U(x,0) = \delta U_i(x). \end{cases} \tag{4.8}$$

Précisons que $\delta v(L,\cdot) = 0$ puisque $v(L,\cdot)$ ne dépend pas de α, β et γ. En effet, $v_d(t)$ est, par hypothèse, indépendant de U. La perturbation $(\delta\alpha,\delta\beta,\delta\gamma)$ conduit à une perturbation δJ de J :

$$\delta J = \frac{1}{|Q|}\int_Q \left(\frac{\rho-\rho^{obs}}{(\rho^c)^2}\right)\delta\rho\,\mathrm{d}x\,\mathrm{d}t + \frac{1}{|Q|}\int_Q \left(\frac{v-v^{obs}}{(v^c)^2}\right)\delta v\,\mathrm{d}x\,\mathrm{d}t. \tag{4.9}$$

Etant donné que

$$\delta v = \delta\left(\frac{\rho w}{\rho} - P(\rho)\right) = \frac{1}{\rho}\delta(\rho w) + \left(-\frac{\rho w}{\rho^2} - P'(\rho)\right)\delta\rho,$$

nous obtenons le lemme suivant.

Lemme 1 *La variation δJ est égale à*

$$\delta J = \frac{1}{|Q|}\int_Q \begin{pmatrix} \dfrac{\rho-\rho^{obs}}{(\rho^c)^2} + \dfrac{v-v^{obs}}{(v^c)^2}\left(-\dfrac{\rho w}{\rho^2} - P'(\rho)\right) \\ \dfrac{v-v^{obs}}{(v^c)^2}\dfrac{1}{\rho} \end{pmatrix} \cdot \delta U(x,t)\,dx\,dt. \tag{4.10}$$

D'autre part, nous avons

$$\delta J = (\nabla_\alpha J, \delta\alpha)_{L^2(0,L)} + (\nabla_\beta J, \delta\beta)_{L^2(0,L)} + (\nabla_\gamma J, \delta\gamma)_{L^2(0,T)}. \tag{4.11}$$

Nous désirons identifier (4.10) et (4.11), ainsi nous avons besoin d'exprimer δU en fonction de δX. Dans ce but, nous introduisons un problème adjoint, dérivé du problème primal linéarisé (4.8). La proposition que nous donnons ci-dessous donne le problème adjoint à considérer et

montre le lien entre l'expression de $\nabla_X J$ et la solution λ du problème adjoint.

Proposition 5 *Soit λ la solution du problème adjoint suivant*

$$\begin{cases} \partial_t\lambda + A(U)^t\partial_x\lambda = \dfrac{1}{|Q|}\begin{pmatrix} \dfrac{\rho-\rho^{obs}}{(\rho^c)^2} + \dfrac{v-v^{obs}}{(v^c)^2}\left(-\dfrac{\rho w}{\rho^2} - P'(\rho)\right) \\ \dfrac{v-v^{obs}}{(v^c)^2}\dfrac{1}{\rho} \end{pmatrix}, \\ \textit{Condition finale : } \lambda(x,T) = 0, \\ \textit{CL1 : } \lambda(L,t)\cdot\begin{pmatrix} 1 \\ w(L,t) + \rho P'(\rho)(L,t) \end{pmatrix} = 0, \\ \textit{CL2 : } \lambda_2(0,t) = 0. \end{cases} \tag{4.12}$$

Alors,

1. *la variation de J par rapport à δU_i et δq est donnée par*

$$\delta J = -\int_0^L \lambda(x,0)\cdot\delta U_i(x)\,dx - \int_0^T \lambda(0,t)\cdot\begin{pmatrix} 1 \\ 2w - P(\rho) \end{pmatrix}\delta q(t)\,dt. \tag{4.13}$$

2. *En notant $\lambda_1(x,t)$ et $\lambda_2(x,t)$ les deux composantes de $\lambda(x,t)$, le gradient de J par rapport à $X = (\alpha,\beta,\gamma)$ est donné par la formule suivante :*

$$\begin{cases} \nabla_\alpha J = -L\,(\lambda_1(\cdot,0) + \lambda_2(\cdot,0)(v + P(\rho) + \rho P'(\rho))(\cdot,0))\phi_\epsilon'(\alpha), \\ \nabla_\beta J = -L\,\lambda_2(\cdot,0)\rho(\cdot,0)\phi_\epsilon'(\beta), \\ \nabla_\gamma J = -T\,\lambda_1(0,\cdot)\phi_\epsilon'(\gamma). \end{cases} \tag{4.14}$$

La preuve est technique. Le lecteur interessé pourra se rapporter au chapitre 6.

4.3 Calcul numérique du gradient

D'un point de vue numérique, il est nécessaire de discrétiser à la fois la solution du problème adjoint et d'approcher l'intégrale de la fonctionnelle par des sommes finies. Nous expliquons dans la section suivante comment les équations du système adjoint sont discrétisées.

Approximation de la solution de l'équation adjointe

Tout d'abord, nous effectuons un changement de variable en temps dans (4.12) pour revenir à un problème avec condition initiale. Posons $\mu(x,t) = \lambda(x,T-t)$. Nous obtenons à partir de (4.12)

$$\begin{cases} \partial_t\mu(x,t) - A(U)^t(x,T-t)\partial_x\mu(x,t) = s(x,T-t), \\ \mu(x,0) = 0, \\ \mu(L,t)\cdot\begin{pmatrix} 1 \\ w(L,T-t) + \rho P'(\rho)(L,T-t) \end{pmatrix} = 0, \\ \mu_2(0,t) = 0 \end{cases} \tag{4.15}$$

avec

$$s(x,t) = -\frac{1}{|Q|}\begin{pmatrix} \dfrac{\rho-\rho^{obs}}{(\rho^c)^2} + \dfrac{v-v^{obs}}{(v^c)^2}\left(-\dfrac{\rho w}{\rho^2} - P'(\rho)\right) \\ \dfrac{v-v^{obs}}{(v^c)^2}\dfrac{1}{\rho} \end{pmatrix}. \tag{4.16}$$

Le système étant hyperbolique, nous allons utiliser de nouveau un schéma décentré. De plus, le système adjoint fait intervenir des équations linéaires. Nous choisissons alors un schéma linéaire implicite. Le fait que le schéma soit implicite permet d'obtenir un schéma numérique inconditionnellement stable pour le système adjoint. Ainsi, nous pouvons rendre les pas de temps plus grands pour un calcul plus rapide de l'état adjoint. Nous introduisons les notations suivantes :

$$B(U) = -A(U)^t, \quad B^k_{i+\frac{1}{2}} = B\left(\frac{U^k_i + U^k_{i+1}}{2}\right).$$

Pour la matrice B, les valeurs propres κ_1 et κ_2 sont les opposées des valeurs propres de A. Nous considérons le schéma suivant pour $1 \leq n \leq N$ et $2 \leq i \leq I-1$:

$$\begin{aligned} \frac{\mu^{n+1}_i - \mu^n_i}{\Delta t^n} + \frac{1}{h}\big(&B^{N-n+1,-}_{i+\frac{1}{2}}(\mu^{n+1}_{i+1} - \mu^{n+1}_i) \\ &+ B^{N-n+1,+}_{i-\frac{1}{2}}(\mu^{n+1}_i - \mu^{n+1}_{i-1})\big) = s^{N-n+2}_i. \end{aligned} \tag{4.17}$$

Nous rappelons que ce schéma est linéaire puisque les matrices $B_{i+\frac{1}{2}}$ ne dépendent pas de $(\mu^n_i)_i$. L'initialisation de μ est $\mu^1_i = 0$ pour tout i.

Comme nous l'avons vu, les conditions aux bords jouent un rôle majeur dans le calcul du gradient (équation (4.12)). Ainsi le traitement numérique des conditions aux bords doit-il faire l'objet d'une étude particulière.

Traitement numérique des conditions aux frontières

Conditions en amont pour $x = 0$

La connaissance de $\lambda_1(0,t)$ pour tout $t > 0$ est nécessaire pour calculer $\nabla_\gamma J$ d'après la proposition 5. Le but de cette partie est de calculer précisément $\lambda_1(0,t)$. Il est équivalent de calculer $\mu_1(0, T-t)$. Il s'agit ainsi d'exhiber une information sur les valeurs aux frontières qui n'est pas donnée par les conditions aux bords. Nous introduisons un état fantôme

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}^{n+1}_G$$

à gauche de l'état

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}^{n+1}_1$$

sur $[x_{-\frac{1}{2}}, x_{\frac{1}{2}}]$, $x_{-\frac{1}{2}} = 0$. Le système (4.15) impose $\mu_2(0,t) = 0$. Nous devons donc poser $(\mu_2)^{n+1}_G = 0$. Le fluide est subsonique : il y a une valeur propre négative et une valeur propre positive. Ainsi

nous pouvons calculer $(\mu_1)_G^{n+1}$ à l'aide des informations provenant de la droite (aval). Considérons maintenant le système

$$\partial_t \mu(x,t) + B^k_{-\frac{1}{2}} \partial_x \mu(x,t) = s(x, T-t) \tag{4.18}$$

où $B^k_{-\frac{1}{2}}$ est la matrice fixée et connue $B(\frac{U_G+U_1}{2}, t = T - t^{n+1})$. Les vecteurs propres à droite de $B^k_{-\frac{1}{2}}$ sont

$$(r^1)^k_{-\frac{1}{2}} = \begin{pmatrix} -w^k_{-\frac{1}{2}} \\ 1 \end{pmatrix}, \; (r^2)^k_{-\frac{1}{2}} = \begin{pmatrix} -w^k_{-\frac{1}{2}} - c^k_{-\frac{1}{2}} \\ 1 \end{pmatrix}$$

et ont pour valeurs propres associées

$$(\kappa_1)^k_{-\frac{1}{2}} = -v^k_{-\frac{1}{2}}, \quad (\kappa_2)^k_{-\frac{1}{2}} = -v^k_{-\frac{1}{2}} + c^k_{-\frac{1}{2}} \tag{4.19}$$

avec $(\kappa_1)^k_{-\frac{1}{2}} < 0$ et $(\kappa_2)^k_{-\frac{1}{2}} > 0$. Nous n'écrirons plus l'indice k dans la suite. Ainsi nous pouvons écrire $\mu(x,t) = (\mu_1, \mu_2)^T$ dans la base $((r^1)_{-\frac{1}{2}}, (r^2)_{-\frac{1}{2}})$

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \alpha_1 (r^1)_{-\frac{1}{2}} + \alpha_2 (r^2)_{-\frac{1}{2}}. \tag{4.20}$$

Nous diagonalisons ensuite l'équation (4.18) pour obtenir une équation où α_1 serait l'inconnue (l'équation concernant α_2 ne nous intéresse pas). Nous obtenons un système bien posé sur $[t^n, t^{n+1}]$

$$\begin{aligned} &\partial_t \alpha_1(x,t) + (\kappa_1)_{-\frac{1}{2}} \partial_x \alpha_1(x,t) = s_1(x, T-t) \\ &\alpha_1(x, t = t^n) = (\alpha_1)^n_G \text{ si } x < 0 \text{ et } \alpha_1(x, t = t^n) = (\alpha_1)^n_1 \text{ si } x > 0 \\ &\alpha_1(x = x_{\frac{1}{2}}, t) = (\alpha_1)^n_{\frac{1}{2}} \end{aligned} \tag{4.21}$$

où $s(x, T-t) = s_1(x, T-t)\,(r^1)_{-\frac{1}{2}} + s_2(x, T-t)\,(r^2)_{-\frac{1}{2}}$. La discrétisation du système (4.21) par un schéma explicite permet de calculer $(\alpha_1)_G$.
A l'aide de la condition $(\mu_2)_G = 0$, nous obtenons finalement $(\mu_1)_G^{n+1}$:

$$(\mu_1)_G^{n+1} = c_G^{n+1}\,(\alpha_1)_G^{n+1}$$

où $c_G^{n+1} = \rho_G^{n+1} P'^{n+1}_G$.

Nous traitons maintenant du problème des conditions en aval $x = L$ $(i = I)$.

Conditions en aval pour $x = L$

Le système (4.15) impose

$$\mu(L,t) \cdot \begin{pmatrix} 1 \\ w(L, T-t) + \rho P'(\rho)(L, T-t) \end{pmatrix} = 0.$$

Comme précédemment, nous introduisons un état fantôme

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}^{n+1}_G$$

à droite de l'état

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}_I^{n+1}$$

pour $[x_{I-\frac{1}{2}}, x_{I+\frac{1}{2}}]$, $x_{I-\frac{1}{2}} = L$. Considérons maintenant le système

$$\partial_t \mu(x,t) + B^k_{I-\frac{1}{2}} \partial_x \mu(x,t) = s(x, T-t) \tag{4.22}$$

où $B^k_{I-\frac{1}{2}}$ est la matrice fixée et connue $B(\frac{U_I+U_G}{2}, t = T - t^{n+1})$. Avec les mêmes notations que précédemment, nous écrivons $\mu(x,t) = (\mu_1, \mu_2)^T$ dans la base $((r^1)_{I-\frac{1}{2}}, (r^2)_{I-\frac{1}{2}})$

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \alpha_1 (r^1)_{I-\frac{1}{2}} + \alpha_2 (r^2)_{I-\frac{1}{2}}. \tag{4.23}$$

Nous diagonalisons l'équation (4.22). Dans le cas des conditions en aval, nous voulons avoir une équation où α_2 serait l'inconnue. Nous obtenons sur $[t^n, t^{n+1}]$

$$\begin{array}{l} \partial_t \alpha_2(x,t) + (\kappa_2)_{I-\frac{1}{2}} \partial_x \alpha_2(x,t) = s_2(x, T-t) \\ \alpha_2(x, t = t^n) = (\alpha_2)^n_I \text{ si } x \leq L \text{ et } \alpha_2(x, t = t^n) = (\alpha_2)^n_G \text{ si } x > L \\ \alpha_2(x = x_{I-\frac{3}{2}}, t) = (\alpha_2)^n_{I-\frac{3}{2}} \end{array} \tag{4.24}$$

La discrétisation du système (4.24) par un schéma explicite permet de calculer $(\alpha_2)^{n+1}_G$. Nous pouvons vérifier que la condition aval dans (4.15) impose $(\alpha_1)_G = 0$.
Finalement, nous obtenons :

$$(\mu_2)^{n+1}_G = (\alpha_2)^{n+1}_G$$

4.4 Algorithmes d'optimisation

Nous présentons brièvement dans cette partie plusieurs algorithmes d'optimisation que nous avons testés. Il s'agit des méthodes de gradient conjugué dans les cas quadratique et non quadratique, des méthodes de couplage dans ce dernier cas avec des méthodes de recherche linéaire, des méthodes de quasi-Newton et enfin de la méthode de région de confiance couplée avec la méthode de BFGS. Pour plus de détails, on pourra lire les livres [1], [6] et [5]. Le but de ces algorithmes est de minimiser une fonction de coût à valeurs réelles J de X, où X est un vecteur de $\mathbb{R}^N$.

Méthodes de gradient conjugué.

Rappelons l'algorithme de gradient conjugué dans le cas d'une fonctionnelle quadratique définie pour $X \in \mathbb{R}^N$ par $J(X) = a + b^T X + \frac{1}{2} X^T A X$, où a et b sont des vecteurs fixés de $\mathbb{R}^N$ et A est une matrice carrée de taille n.

Algorithme 1 *(Gradient conjugué) :*

- *Initialisation : on se donne X_0 et on pose $d_0 = \nabla J(X_0)$.*
- *Itération (k) :*

- $X_{k+1} = X_k - t_k d_k$ *avec* $t_k = \frac{d_k^T \nabla J(X_k)}{d_k^T A d_k}$,
- $d_{k+1} = \nabla J(X_{k+1}) - \beta_k d_k$ *avec* $\beta_k = \frac{d_k^T A \nabla J(X_{k+1})}{d_k^T A d_k}$.

Il existe plusieurs variantes pour l'expression de β_k. Le coefficient β_k pour la variante de Fletcher-Reeves s'écrit

$$\beta_k = \frac{\|\nabla J(X_{k+1})\|^2}{\|\nabla J(X_k)\|^2},$$

et pour la variante de Polak-Ribière,

$$\beta_k = \frac{(\nabla J(X_{k+1}) - \nabla J(X_k))^T \nabla J(X_{k+1})}{\|\nabla J(X_k)\|^2}.$$

La variante de Fletcher-Reeves ou celle de Polak-Ribière peuvent être appliquées lorsque la fonctionnelle n'est pas quadratique. Le seul changement par rapport à l'algorithme utilisé dans le cas quadratique est le calcul du terme t_k par une recherche linéaire et non plus par la formule précédente. En pratique, la méthode de Polak-Ribière est considérée comme meilleure que celle de Fletcher-Reeves.

Méthode de quasi-Newton

Avant de présenter les méthodes de quasi-Newton, nous rappelons l'algorithme de Newton. Lorsque J n'est pas quadratique, il s'agit avant tout d'approcher J près d'un point X_k par une fonction quadratique

$$\tilde{J}(X_k + \delta) = J(X_k) + \nabla J(X_k) \cdot \delta + \frac{1}{2}\delta^T J''(X_0)\delta \tag{4.25}$$

et de minimiser cette fonction en résolvant l'équation d'inconnue δ, $\nabla \tilde{J}(X_k + \delta) = 0$.
L'algorithme de Newton s'écrit :

Algorithme 2 *(Newton) :*
- *Initialisation : on se donne* X_0.
- *Itération* k *:*
  - $\delta_k = -J''(X_k)^{-1} \nabla J(X_k)$,
  - $X_{k+1} = X_k + \delta_k$.

En pratique, on ne cherche pas l'inverse de $J''(X_k)$ mais on calcule δ_k en résolvant le système linéaire $J''(X_k)\delta_k = -\nabla J(X_k)$.
Il est tout de même nécessaire de calculer le hessien de J. Les méthodes de quasi-Newton permettent d'éviter ce calcul. On remplace en effet le calcul $\delta_k = -J''(X_k)^{-1}\nabla J(X_k)$ par $\delta_k = t_k M_k \nabla J(X_k)$ où M_k est une matrice symétrique définie positive approchant $J''(X_k)^{-1}$ et calculée itérativement. Le coefficient positif t_k est calculé par une recherche linéaire.
Nous donnons ici la méthode B.F.G.S. (Broyden-Fletcher-Godfarb-Shanno) :

Algorithme 3 *(calcul de la matrice hessienne de J par B.F.G.S.) :*

- *Initialisation : soit M^0 une matrice symétrique définie positive.*
- *Itération (k) : X^{k+1} et X^k sont supposés connus.*
  - *Soit $\gamma^k = \nabla J(X^{k+1}) - \nabla J(X^k)$,*
  - *Soit $\delta^k = X^{k+1} - X^k$.*
  - *On calcule*

$$M^{k+1} = M^k + \left(1 + \frac{(\gamma^k)^T S^k \gamma^k}{(\delta^k)^T \gamma^k}\right) \frac{\delta^k (\delta^k)^T}{(\delta^k)^T \gamma^k} - \frac{\delta^k (\gamma^k)^T S^k + S^k \gamma^k (\delta^k)^T}{(\delta^k)^T \gamma^k}.$$

Il existe d'autres formules pour approcher le hessien de J que nous ne citerons pas ici. Nous allons présenter maintenant quelques méthodes de recherche linéaire.

Méthode de recherche linéaire

Nous avons vu que les méthodes de gradient conjugué pour une fonctionnelle non quadratique ou les méthodes de quasi Newton nécessitaient le calcul d'un paramètre t_k. Ce paramètre est obtenu par recherche linéaire le long de la direction de descente. Nous présentons essentiellement trois méthodes : la méthode d'Armijo, la méthode de Goldstein et Price et la méthode de Wolfe. Nous supposons qu'à l'itération k, la direction de descente d_k est calculée (par exemple par un des algorithmes précédents). Nous noterons $q(t) = J(X_k + td_k)$. Nous rappelons que nous savons calculer une approximation numérique de $q(t)$ et de $q'(t) = \nabla J(X_k + td_k)^T d_k$ pour chaque t, ces calculs pouvant être couteux. Il s'agit alors de trouver t qui minimise q (on minimise J dans la direction d_k) ou du moins un t convenable. Une recherche linéaire consiste alors à tester t pour savoir s'il est trop grand ou trop petit. Le schéma général d'une recherche linéaire est donné par l'algorithme 4.

Algorithme 4 *(Recherche linéaire) :*

- *Initialisation : soit t donné. On pose $t_g = 0$ et $t_d = 0$ (par exemple)*
- *On teste t.*
  - *Si t satisfait le test, alors on garde t. L'algorithme est fini.*
  - *Si t est trop grand, on pose $t_d = t$.*
  - *Si t est trop petit, on pose $t_g = t$.*
- *Si $t_d = 0$ alors on calcule un nouveau $t > t_g$ (extrapolation),*
- *sinon on calcule un nouveau $t \in]t_g; t_d[$.*

Les méthodes de recherche linéaire diffèrent entre elles par les tests utilisés. Après avoir choisi $0 < m_1 < m_2 < 1$, la recherche linéaire de Wolfe peut s'écrire :

- Si $q(t) \leq q(0) + m_1 t q'(0)$ et $q'(t) \geq m_2 q'(0)$ alors t satisfait le test.
- Si $q(t) > q(0) + m_1 t q'(0)$ alors t est trop grand $(t_d = t)$.
- Si $q(t) \leq q(0) + m_1 t q'(0)$ et $q'(t) < m_2 q'(0)$ alors t est trop petit $(t_g = t)$.

Nous avons représenté un exemple de fonction q ainsi que les différents cas sur la figure 4.1 : les ensembles G_1 et G_2 désignent les t trop grands, les ensembles P_1 et P_2 désignent les t trop petits et les ensembles S_1 et S_2 désignent les t satisfaisant la règle de Wolfe. La règle de Wolfe nécessite

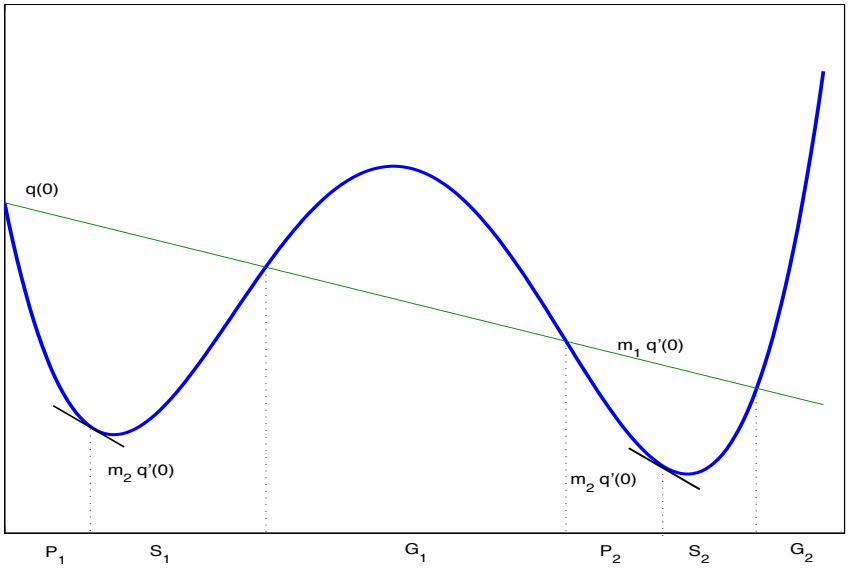

FIG. 4.1 – Règle de Wolfe : les ensembles G_1 et G_2 désignent les t trop grands, les ensembles P_1 et P_2 désignent les t trop petits et les ensembles S_1 et S_2 désignent les t satisfaisant la règle de Wolfe.

la connaissance de $q'(t)$ c'est-à-dire la connaissance du gradient de J au point $X_k + td_k$. Afin d'éviter ce calcul, on peut utiliser la règle de Goldstein et Price ou la règle d'Armijo.
La règle de Goldstein et Price remplace le test impliquant $q'(t)$ par un test concernant $\frac{q(t)-q(0)}{t}$.

- Si $q(t) \leq q(0) + m_1 t q'(0)$ et $\frac{q(t)-q(0)}{t} \geq m_2 q'(0)$ alors t satisfait le test.
- Si $q(t) > q(0) + m_1 t q'(0)$ alors t est trop grand ($t_d = t$).
- Si $\frac{q(t)-q(0)}{t} < m_2 q'(0)$ alors t est trop petit ($t_g = t$).

La règle d'Armijo teste uniquement le fait que t est trop grand mais jamais que t est trop petit :

- Si $q(t) \leq q(0) + m_1 t q'(0)$ alors t satisfait le test.
- Si $q(t) > q(0) + m_1 t q'(0)$ alors t est trop grand ($t_d = t$).

Ainsi cette règle est simple à programmer mais ne permet pas d'extrapoler t dans le cas où ce paramètre serait trop petit.

Méthode de région de confiance

Comme dans les méthodes de quasi-Newton, on approche J près d'un point X_k par une fonction quadratique.

$$\tilde{J}_k(\delta) = J(X_k) + \nabla J(X_k) \cdot \delta + \delta^T M \delta. \tag{4.26}$$

La matrice M est une approximation du hessien de J. On peut l'obtenir par exemple par la formule de BFGS.
Ainsi nous voulons minimiser J près de X_k en minimisant $\tilde{J}_k$. Cependant, l'approximation de

$J(X_k+\delta)$ par $\tilde{J}_k$ est a priori valable uniquement pour $\|\delta\| \leq \rho$ (près de X_k). Le paramètre ρ est appelé rayon de confiance. Il détermine la région dite région de confiance où l'on considère que le modèle quadratique est convenable.
Nous avons alors à résoudre le problème d'optimisation avec contrainte :
trouver

$$\min_{\|\delta\|\leq\rho} \tilde{J}_k(\delta). \tag{4.27}$$

On recherche ensuite les solutions de (4.27) sous la forme

$$\delta(\mu) = -(M+\mu I)^{-1}\nabla J(X_k)$$

(I désigne la matrice identité) où $\mu \geq 0$ doit vérifier $|\delta(\mu)| \leq \rho$ et $\mu(|\delta(\mu)| - \rho) = 0$. Voici l'algorithme de région de confiance à l'itération k.

Algorithme 5 *(Région de confiance) :*
- *Initialisation : $l = 0$, on donne ρ_0, on calcule M_k par la formule de BFGS (par exemple)*
- *Iteration (l) :*
  - *On calcule μ_l en résolvant par la méthode de Newton*

$$|(M_k+\mu I)^{-1}\nabla J(X_k)| = \rho_l$$

  - *On calcule $\delta_l = \delta(\mu_l)$*
  - *Si $J(X_k+\delta_l) \leq J(X_k) - m(J(X_k) - \tilde{J}_k(\delta_l))$, alors $X_{k+1} = X_k + \delta_l$. L'algorithme de région de confiance est terminé. On passe à l'itération $k+1$.*
  - *Sinon, $\rho_{l+1} = \frac{1}{2}\rho_l$. On passe à l'itération $l+1$.*

Le test sur δ_l utilise la règle d'Armijo appliquée à la fonction $\tilde{J}_k$ (cependant, on ne peut plus parler de recherche linéaire, il s'agit plutôt de recherche curviligne sur la trajectoire $\{X_k+\delta\}_{\delta\geq 0}$).

4.5 Résumé de l'algorithme d'optimisation

A l'aide de la connaissance du gradient, nous pouvons calculer de manière itérative les paramètres $X = (\alpha, \beta, \gamma)$ par une des méthodes rappelées précédemment (gradient conjugué, méthode de quasi-Newton, méthode par région de confiance). Actuellement, la méthode de région de confiance fait partie des algorithmes de minimisation les plus performants. Nous résumons l'algorithme complet de minimisation utilisé pour le problème de trafic routier :

Algorithme 6 *(Algorithme de minimisation) :*
- *Initialisation : α^0, β^0, γ^0 ;*
- *Itération (k) : soient $\alpha^{(k)}$, $\beta^{(k)}$, et $\gamma^{(k)}$ donnés,*
  - *On calcule $(\rho^i)^{(k)}$, $(v^i)^{(k)}$, $q^{(k)}$;*
  - *On calcule la solution approchée $U^{(k)}$ du problème de Aw-Rascle (2.4) en utilisant le schéma (2.12) ;*
  - *On calcule $J^{(k)}$ à l'aide de la formule (4.5) ;*

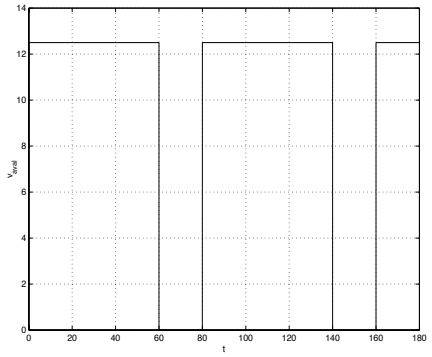

FIG. 4.2 – Vitesse choisie comme condition en aval

- *On calcule la solution approchée $\lambda^{(k)}$ du problème adjoint 4.12 en utilisant le schéma numérique 4.17 ;*
- *On calcule ∇J^k à l'aide de la formule* (4.14) *;*
- *On calcule finalement α^{k+1}, β^{k+1}, γ^{k+1} par l'algorithme 5 de région de confiance couplé avec l'algorithme 3 de BFGS.*

4.6 Validation numérique et expérience

Pour notre expérience numérique, nous choisissons comme condition en amont en $x = L$ une vitesse en créneau pour les véhicules. Cette vitesse est représentée sur la figure 4.2. Dans la figure 4.3, nous avons représenté les conditions initiales et les conditions aux bords observées, ainsi que celles choisies pour l'initialisation de l'optimisation (itération 1) et celles calculées (itération 75).

Nous observons quelques différences entre d'une part les conditions aux bords et les conditions initiales observées et d'autre part les conditions initiales et les conditions aux bords calculées, mais globalement, nous pouvons dire que les conditions U_0 et q sont calculées de manière satisfaisante, d'autant plus que ce qui nous intéresse dans un tel problème est davantage la valeur de U au temps final. En effet, nous rappelons que nous voulons utiliser ces valeurs pour faire des prévisions sur U après le temps T. Or les valeurs calculées de U au temps $t = T$ et celles observées sont identiques à l'oeil nu. Enfin, nous représentons sur la figure 4.4 les valeurs de $\log_{10} J$ et de $\log_{10} \|\nabla J\|$ au fil des itérations. La norme choisie pour le gradient est la norme $L^2(0, L) \times L^2(0, L) \times L^2(0, T)$.

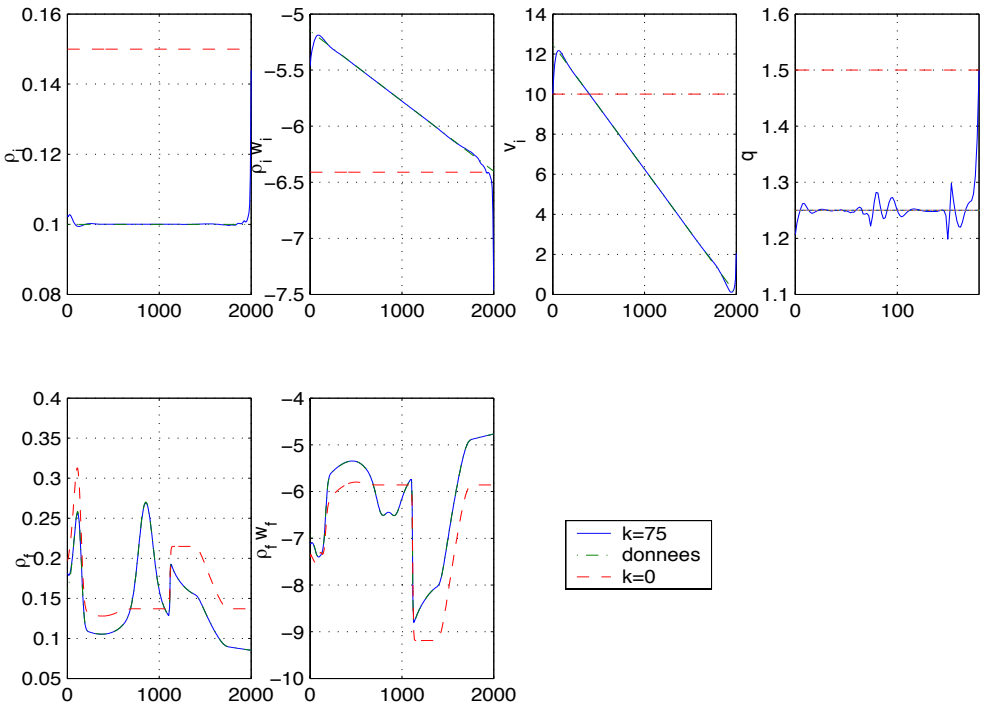

FIG. 4.3 – En haut : conditions initiales et conditions aux bords ; en bas : valeurs finales de ρ et ρw à $T = 180$. Nous avons représenté les valeurs observées (données), les valeurs initialisant l'algorithme d'optimisation ($k = 0$) et les valeurs calculées à l'itération $k = 75$ ($k = 75$).

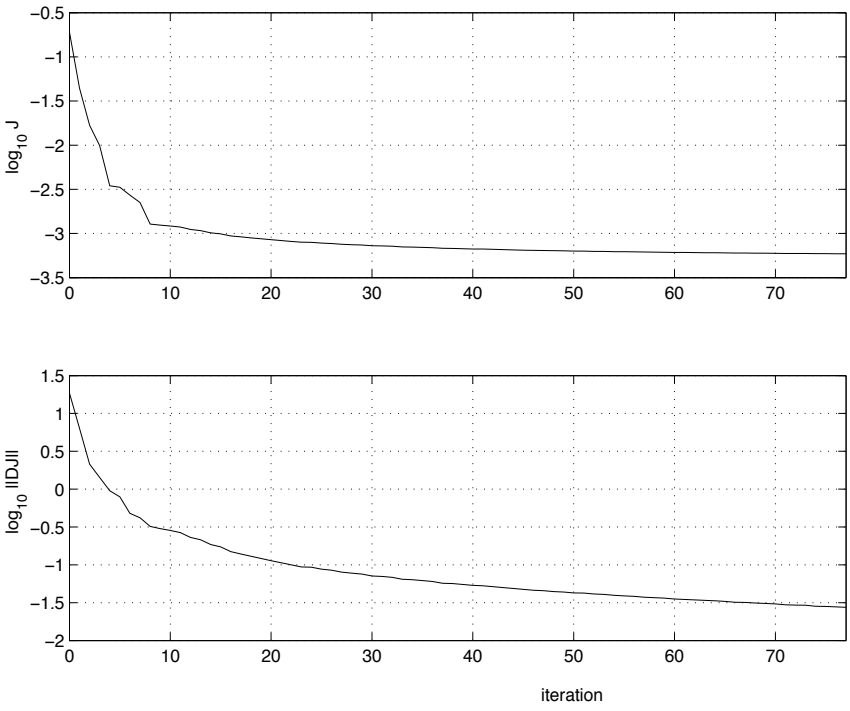

FIG. 4.4 – Valeurs de $\log_{10} J$ et de $\log_{10} \|\nabla J\|$.

5

Optimisation sur les conditions initiales uniquement.

5.1 Présentation de la seconde stratégie.

Nous considérons une nouvelle situation. Les observations sont données dans $[0;L]\times[0;T]$ et la vitesse en $x=L$ est encore connue. Cependant, nous ne voulons plus rechercher les conditions aux bords en $x=0$. La recherche simultanée des conditions initiales sur la section de route $[0;L]$ et de la condition en amont est maintenant remplacée par une recherche des conditions initiales sur une section plus étendue $[-L';L]$. La valeur de L' est choisie pour que la condition en amont en $x=-L'$ n'ait pas d'influence sur la solution sur $[0;L]\times[0;T]$. En effet, pour le système de Aw-Rascle, nous avons vu que l'information qui va de la gauche vers la droite (dans le sens de circulation des voitures) ne peut pas voyager plus rapidement que les véhicules eux-mêmes. Si nous étudions le système sur une durée de T secondes, nous devons prendre L' telle que $L'>T\times v_{max}$ où v_{max} est la vitesse maximale des véhicules. Le problème d'optimisation concerne maintenant le système suivant sur $[-L';L]\times[0;T]$:

$$\begin{cases} \partial_t\rho+\partial_x(v\rho)=0, \\ \partial_t(\rho w)+\partial_x(v\rho w)=0, \\ (\rho v)(-L',t)=q(t), v(L,t)=v_d(t), \\ \rho(x,0)=\rho_i(x), v(x,0)=v_i(x) \end{cases} \tag{5.1}$$

où q est une fonction positive fixée dont la valeur n'a pas d'importance. Par exemple, nous posons $q=0$. Nous voulons trouver les bonnes conditions initiales sur $[-L';L]$ à partir des observations obtenues uniquement sur $[0;L]$. Par conséquent, la fonctionnelle est définie par la même formule que dans le paragraphe précédent, mais elle ne dépend plus que des deux fonctions α et β définies sur $[-L';L]$:

$$J(\alpha,\beta)=\frac{1}{2}\frac{1}{|Q|}\int_Q\left(\frac{\rho(x,t)-\rho^{obs}(x,t)}{\rho^c}\right)^2+\left(\frac{v(x,t)-v^{obs}(x,t)}{v^c}\right)^2\mathrm{d}x\mathrm{d}t \tag{5.2}$$

Les variables d'optimisation sont maintenant $X=(\alpha,\beta)$ et le problème de minimisation que nous devons résoudre s'écrit :

Problème d'optimisation 4 *Trouver $\tilde{U}$ solution de (5.1) telle que*

$$J(\tilde{U}) = \min_{\substack{(\alpha,\beta)\in\mathcal{C}^1([-L',L]),\\ (\alpha,\beta) \text{ donnés par } (4.3),\, U=(\rho,\rho w) \text{ solution de } (4.1)}} J(\rho(X), v(X)). \tag{5.3}$$

L'ensemble Q est à nouveau $[0, L] \times [0, T]$. Nous expliquons dans la suite uniquement les points qui diffèrent de la première stratégie.

5.2 Détermination du gradient de J.

Nous obtenons un système adjoint pour λ de la même façon que dans le cas précédent. La seule différence est que nous intégrons cette équation sur$[-L'; L] \times [0; T]$ et non plus sur $[0; L] \times [0; T]$. Cependant, la fonctionnnelle définie par (5.2) est toujours obtenue en intégrant sur $[0; L] \times [0; T]$. Nous obtenons alors comme nouveau problème adjoint :

$$\begin{cases} \partial_t\lambda + A(U)^T\partial_x\lambda = s(x,t), \\ \lambda(x,T) = 0, \\ \lambda(L,t)\cdot \begin{pmatrix} 1 \\ w(L,t) + \rho P'(\rho)(L,t) \end{pmatrix} = 0, \\ \lambda_2(-L',t) = 0, \end{cases} \tag{5.4}$$

où

$$s(x,t) = \begin{cases} \frac{1}{|Q|} \begin{pmatrix} \frac{\rho-\rho^o}{(\rho^c)^2} + \frac{v-v^o}{(v^c)^2}(-\frac{\rho w}{\rho^2} - P'(\rho)) \\ \frac{v-v^o}{(v^c)^2}\frac{1}{\rho} \end{pmatrix} \text{ si } x \in [0; L], \\ 0 \text{ sinon.} \end{cases} \tag{5.5}$$

Nous rappelons que $|Q| = L \times T$. Le gradient de J est donné par

$$\begin{cases} \nabla_\alpha J = -(L + L')\left(\lambda_1(\cdot, 0) + (v + P(\rho) + \rho P'(\rho))(\cdot, 0)\right)\phi'_\epsilon(\alpha), \\ \nabla_\beta J = -(L + L')\,\lambda_2(\cdot, 0)\rho(\cdot, 0)\phi'_\epsilon(\beta). \end{cases} \tag{5.6}$$

5.3 Résultats numériques.

Nous désirons déduire les conditions initiales à partir des observations. La figure 5.1 représente la condition initiale pour la solution observée, les conditions initiales que nous avons choisies pour initialiser l'algorithme d'optimisation et les conditions initiales que nous avons calculées au bout de 100 itérations.

Nous remarquons que les conditions initiales sont calculées précisément sur $[0; L]$ mais ne le sont pas sur tout l'intervalle $[-L'; L]$. En effet, les informations qui sont trop éloignées de la zone d'observations (section $[0; L]$) n'ont pas forcément le temps d'arriver et d'influencer les quantités intervenant dans J (souvenons-nous que nous avons choisi $L = v_{\max}T$ où $v_{\max}$ est la vitesse maximale des voitures, or tous les véhicules ne roulent pas forcément à cette allure). De

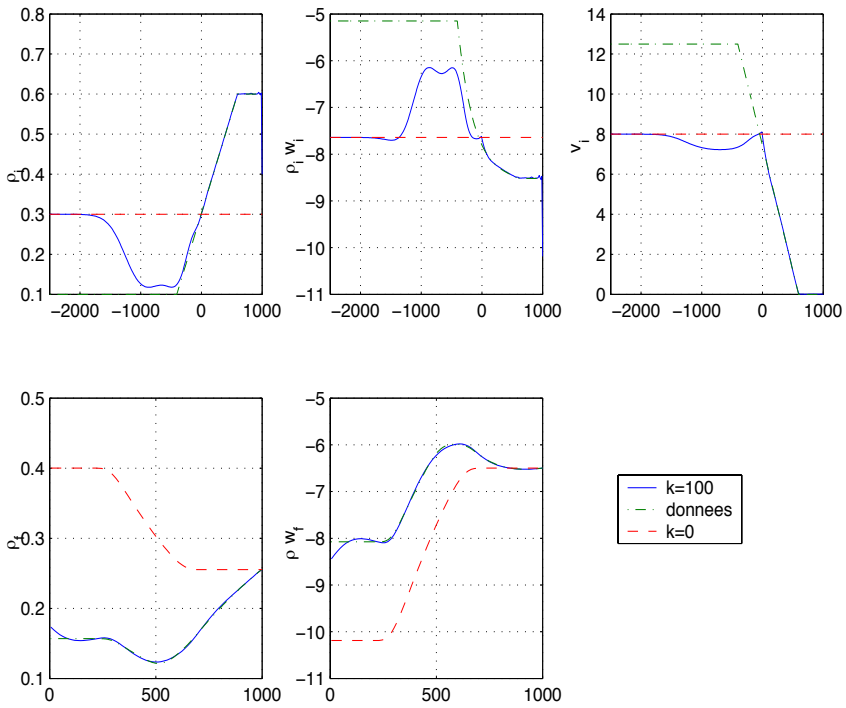

FIG. 5.1 – En haut : conditions initiales ; en bas : valeurs finales de ρ et ρw à $T = 180$. Nous avons représenté les valeurs observées (données), les valeurs initialisant l'algorithme d'optimisation ($k = 0$) et les valeurs calculées à l'itération $k = 100$ ($k = 100$).

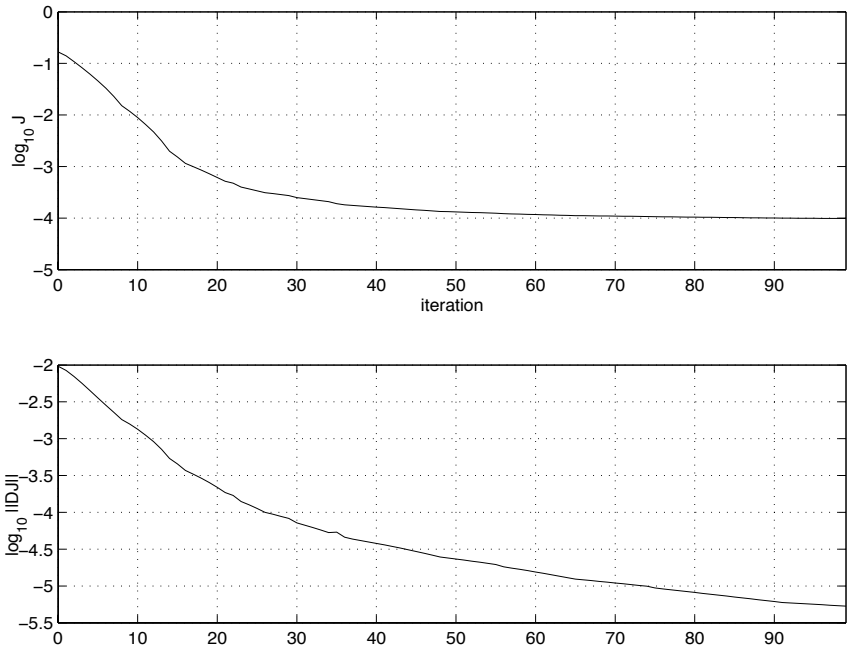

FIG. 5.2 – Valeurs de J et norme $L^2(-L',L) \times L^2(-L',L)$ du gradient de J.

plus, la densité et la vitesse au temps final $t = T$ (début de la prévision) sont très légèrement différentes des observations. Enfin, nous donnons avec la figure 5.2 les courbes représentant $\log_{10} J$ et $\log_{10} \|\nabla J\|$ en fonction du nombre d'itérations.

6

Preuve de la proposition 5

Nous donnons ici la preuve de la proposition 5. Elle permet d'expliquer la construction du problème adjoint.

Obtention du problème adjoint

Soit λ une fonction de classe $\mathcal{C}^1$ définie sur Q à valeurs dans $\mathbb{R}^2$. Nous notons (λ_1, λ_2) les deux composantes de λ. En multipliant le système linéarisé d'EDP (4.8) par λ, nous obtenons

$$\int_Q \lambda \cdot \partial_t \delta U + \lambda \cdot \partial_x (A(U)\delta U)\,\mathrm{d}x\mathrm{d}t = 0. \tag{6.1}$$

Nous intégrons par partie le membre de gauche pour obtenir une expression uniquement en fonction des variables primitives d'optimisation δU_i et δq :

$$\int_Q \partial_t \lambda \cdot \delta U + \partial_x \lambda \cdot A(U)\delta U\,\mathrm{d}x\mathrm{d}t = \int_0^L [\lambda \cdot \delta U]_0^T\,\mathrm{d}x + \int_0^T [\lambda \cdot A(U)\delta U]_0^L\,\mathrm{d}t. \tag{6.2}$$

Ainsi

$$\begin{aligned}\int_Q \partial_t \lambda \cdot \delta U + A(U)^T \partial_x \lambda \cdot \delta U\,\mathrm{d}x\mathrm{d}t& \\ = \int_0^L \lambda(x,T)\cdot\delta U(x,T) - \lambda(x,0)\cdot\delta U_i(x)\,\mathrm{d}x& \\ + \int_0^T \lambda(L,t)\cdot A\delta U(L,t) - \lambda(0,t)\cdot A\delta U(0,t)\,\mathrm{d}t.&\end{aligned} \tag{6.3}$$

Puisque nous avons au premier ordre

$$\delta w(L,t) = \delta v(L,t) + \delta P(\rho(L,t)) = P'(\rho(L,t))\delta\rho(L,t),$$

nous pouvons écrire $\delta U(L,T)$ en fonction de $\delta\rho(L,t)$:

$$\delta U(L,t) = \begin{pmatrix} \delta\rho(L,t) \\ \delta(\rho w(L,t)) \end{pmatrix} = \begin{pmatrix} 1 \\ w(L,t) + \rho P'(\rho)(L,t) \end{pmatrix} \delta\rho(L,t). \tag{6.4}$$

De même, nous obtenons $\delta U(0,t)$ en fonction de $\delta\rho(0,t)$ et de $\delta q(t)$:

$$\delta(\rho w)(0,t) = \delta(\rho v + \rho P)(0,t) = \delta q(t) + (P + \rho P'(\rho))\delta\rho(0,t). \tag{6.5}$$

Ainsi

$$\delta U(0,t) = \begin{pmatrix} 1 \\ P(\rho) + \rho P'(\rho) \end{pmatrix} \delta\rho(0,t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta q(t). \tag{6.6}$$

Nous écrivons maintenant (6.3) sous la forme

$$\begin{aligned} &\int_Q \partial_t\lambda \cdot \delta U + A(U)^T \partial_x\lambda \cdot \delta U \,\mathrm{d}x\mathrm{d}t \\ &\quad = \int_0^L \lambda(x,T)\cdot\delta U(x,T) - \lambda(x,0)\cdot\delta U_i(x)\,\mathrm{d}x \\ &\qquad + \int_0^T \lambda(L,t)\cdot A \begin{pmatrix} 1 \\ w(L,t) + \rho P'(\rho)(L,t) \end{pmatrix} \delta\rho(L,t) \\ &\qquad - \lambda(0,t)\cdot A\left(\begin{pmatrix} 1 \\ P(\rho) + \rho P'(\rho) \end{pmatrix} \delta\rho(0,t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta q(t) \right) \mathrm{d}t. \end{aligned} \tag{6.7}$$

Nous allons chercher λ comme solution d'une équation aux dérivées partielles avec des conditions finales et des conditions aux bords. Nous avons pour objectif de remplacer le membre de gauche de (6.7) par δJ à l'aide du lemme 1 et d'écrire le membre de droite uniquement en fonction de δU_i et δq. Les conditions finales et les conditions aux bords du problème adjoint seront ainsi choisies pour éliminer les termes indésirables. Avec ces contraintes, le système que nous devons résoudre s'écrit donc

$$\begin{cases} \partial_t\lambda + A(U)^T\partial_x\lambda = \frac{1}{|Q|} \begin{pmatrix} \frac{\rho-\rho^o}{(\rho^c)^2} + \frac{v-v^o}{(v^c)^2}(-\frac{\rho w}{\rho^2} - P'(\rho)) \\ \frac{v-v^o}{(v^c)^2}\frac{1}{\rho} \end{pmatrix}, \\ \lambda(x,T) = 0, \\ \lambda(L,t)\cdot A \begin{pmatrix} 1 \\ w(L,t) + \rho P'(\rho)(L,t) \end{pmatrix} = 0, \\ \lambda(0,t)\cdot A \begin{pmatrix} 1 \\ P(\rho) + \rho P'(\rho) \end{pmatrix} = 0. \end{cases} \tag{6.8}$$

Or

$$\begin{pmatrix} 1 \\ w(L,t) + \rho P'(\rho)(L,t) \end{pmatrix} \tag{6.9}$$

est un vecteur propre de A. De plus, nous trouvons que

$$A \begin{pmatrix} 1 \\ P(\rho) + \rho P'(\rho) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \gamma_1\gamma_2. \tag{6.10}$$

(Nous rappelons que γ_i sont les valeurs propres de A). Finalement le système (6.8) devient

$$\left\{\begin{array}{l} \partial_t\lambda + A(U)^T\partial_x\lambda = \frac{1}{|Q|}\left(\begin{array}{c} \frac{\rho-\rho^o}{(\rho^c)^2} + \frac{v-v^o}{(v^c)^2}(-\frac{\rho w}{\rho^2} - P'(\rho)) \\ \frac{v-v^o}{(v^c)^2}\frac{1}{\rho} \end{array}\right), \\ \lambda(x,T) = 0, \\ \lambda(L,t)\cdot\left(\begin{array}{c} 1 \\ w(L,t) + \rho P'(\rho)(L,t) \end{array}\right) = 0, \\ \lambda_2(0,t) = 0, \end{array}\right. \tag{6.11}$$

qui est le système adjoint du problème primal.

Détermination du gradient de J

Ainsi nous pouvons écrire grâce à (4.10), (6.7) et (6.11) l'expression de δJ en fonction de δU_i et de δq :

$$\delta J = -\int_0^L \lambda(x,0)\cdot\delta U_i(x)\,\mathrm{d}x - \int_0^T \lambda(0,t)\cdot\left(\begin{array}{c} 1 \\ 2w - P(\rho) \end{array}\right)\delta q(t)\,\mathrm{d}t. \tag{6.12}$$

A présent, nous allons déterminer le gradient de J par rapport aux variables de contrôle X. A partir de (4.13), nous obtenons

$$\delta U = \left(\begin{array}{c} \delta\rho \\ \delta(\rho w) \end{array}\right) = \left(\begin{array}{c} \delta\rho \\ \delta(\rho v + \rho P(\rho)) \end{array}\right) = \left(\begin{array}{c} 1 \\ w + \rho P'(\rho) \end{array}\right)\delta\rho + \left(\begin{array}{c} 0 \\ \rho \end{array}\right)\delta v. \tag{6.13}$$

Ainsi nous pouvons écrire δU_i et δq en fonction de $\delta\alpha$, $\delta\beta$ et $\delta\gamma$:

$$\begin{array}{l} \delta U_i = \left(\begin{array}{c} 1 \\ w + \rho P'(\rho) \end{array}\right)\phi_\epsilon'(\alpha)\delta\alpha + \left(\begin{array}{c} 0 \\ \rho \end{array}\right)\phi_\epsilon'(\beta)\delta\beta, \\ \delta q = \phi_\epsilon'(\gamma)\delta\gamma. \end{array} \tag{6.14}$$

Finalement nous obtenons :

$$\begin{aligned} \delta J = -\int_0^L \lambda(x,0)\cdot\left(\left(\begin{array}{c} 1 \\ v + P(\rho) + \rho P'(\rho) \end{array}\right)\phi_\epsilon'(\alpha)\delta\alpha + \left(\begin{array}{c} 0 \\ \rho \end{array}\right)\phi_\epsilon'(\beta)\delta\beta\right)\mathrm{d}x \\ - \int_0^T \lambda(0,t)\cdot\left(\begin{array}{c} 1 \\ 2w - P(\rho) \end{array}\right)\phi_\epsilon'(\gamma)\delta\gamma\,\mathrm{d}t. \end{aligned}$$

Nous rappelons que

$$\begin{aligned} \delta J = &< \nabla_\alpha J, \delta\alpha >_{L^2([0,L])} + < \nabla_\beta J, \delta\beta >_{L^2([0,L])} \\ &+ < \nabla_\gamma J, \delta\gamma >_{L^2([0,T])} \end{aligned} \tag{6.15}$$

($<>$ est le produit scalaire) et λ est solution de (4.12). Ainsi nous avons

$$\begin{array}{l} < \nabla_\alpha J, \delta\alpha > = -\int_0^L (\lambda_1(x,0) + \lambda_2(x,0)(v + P(\rho) + \rho P'(\rho))(x,0))\phi_\epsilon'(\alpha)\delta\alpha\,\mathrm{d}x, \\ < \nabla_\beta J, \delta\beta > = -\int_0^L \lambda_2(x,0)\rho(x,0)\phi_\epsilon'(\beta)\delta\beta\,\mathrm{d}x, \\ < \nabla_\gamma J, \delta\gamma > = -\int_0^T \lambda_1(0,t)\phi_\epsilon'(\gamma)\delta\gamma\,\mathrm{d}t \end{array}$$

(souvenons-nous que $\lambda_2(0,t) = 0$). Nous venons de trouver la formule attendue.

Bibliographie

[1] Allaire, G. *Analyse numérique et optimisation.* Les Editions de l'Ecole Polytéchnique, 2005.

[2] Aw, A., Klar, A., Materne, T., and Rascle, M. Derivation of continuum traffic flow models from microscopic follow-the-leader models. *SIAM J. Appl. Math. 63*, 1 (2000), 259–278.

[3] Aw, A., and Rascle, M. Resurrection of second order models of traffic flow ? *SIAM J. Appl. Math. 60*, 3 (2000), 916–938.

[4] Barcelo, J., Casas, J., Ferrer, J., and Garcia, D. Modelling advanced transport telematic applications with microscopic simulators : the case with aimsun2. *Traffic and Mobility, Simulation, Economics, Environment* (1999), 205–224.

[5] Bonnans, J., Gilbert, J., Lemarechal, C., and Sagastizabal, C. *Optimisation numérique.* Springer Verlag, 1997.

[6] Culioli, J. *Introduction à l'optimisation.* Ellipses, 1994.

[7] Daganzo, C. Requiem for second-order fluid approximations of traffic flow. *Transp. Res.-B 29B, No.4* (1995), 277–286.

[8] Daley, R. *Atmospheric Data Analysis.* Cambridge University Pres, 1991.

[9] Dimet, F. L., Navon, I., and Daescu, D. Second-order information in data assimialtion. *American Meteorological Society 130* (2002), 629–648.

[10] Dimet, F. L., and Talagrand, O. Variational algorithms for analysis and assimilation of meterological observations : theoritical aspects. *Tellus 38A* (1986), 97–110.

[11] Dubois, F., and Lefloch, P. Boundary conditions for nonlinear hyperbolic systems of conservation laws. *2nd International Conference on Hyperbolic Problem, Aachen* (juin 1988).

[12] Ghidaglia, J., Kumbaro, A., and Coq, G. L. Une méthode "volumes finis" à flux caractéristiques pour la résolution numérique des systèmes hyperboliques de lois de conservation. *C.R. Acad. Sci. Paris Sér I Math. 322*, 10 (1996), 981–988.

[13] Harten, A. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics. 49* (mars 1983), 357–393.

[14] Hoogendorn. State-of-the-art of vehicular traffic flow modelling. *Proceedings of the institution of Mechanical Engineers, Journal of Systems and Control Engineering 215*, 14 (2001), 283–303.

[15] Jaisson, P., and DeVuyst, F. Data assimilation for fluid traffic flow models and optimization algorithms. *IFIP, Turin, Italie* (juillet 2005).

[16] Klar, A., Kuhne, D., and Wegener, R. Mathematical models for vehicular traffic. *Surveys Math. Indust. 6* (1996), 215–339.

[17] Lighthill, M., and Whitham, J. On kinematic waves. i : Flow movement in long rivers. ii : A theory of traffic flow on long crowded roads. *Proc. Royal Soc. Edinburgh A229* (1955), 281–345.

[18] Luong, B., Blum, J., and Verron, J. A variational method for the resolution of a data assimilation problem in oceanograpy. *Inverse Problem 14* (1998), 979–997.

[19] Payne, H. Models of freeway traffic and control. *Simul. Council Proc. 28* (1971), 51–61.

[20] Prigogine, I. *A Boltzmann-like Approach to the Statistical Theory of Traffic Flow, in Theory of Traffic Flow.* Hermann, R., 1961.

[21] Prigogine, I., and Hermann, R. Kinetic theory of vehicular traffic. *American Elsevier* (1971).

[22] Richards, P. Shock waves on the highway. *Operations research 4*, 1 (1956), 42–51.

[23] Roe, P. Approximate riemann solvers, parameter vector and difference scheme. *Journal of Computational Physics 43* (1981), 357–372.

[24] Whitham, G. *Linear and Non Linear Waves.* Pure and Applied Mathematics. Wiley Interscience, 1974.

Troisième partie

Schéma hybride

Introduction

Dans les deux parties précédentes, nous avons vu que la modélisation du flux d'informations et la modélisation du trafic routier font intervenir des systèmes de lois de conservation pouvant s'écrire sous la forme

$$\partial_t u + \partial_x f(u) = 0,\ x \in I,\ t \geq 0,$$
$$u(x,0) = u^0(x),$$

où I est un intervalle de $\mathbb{R}$ (nous ne parlons pas ici des conditions aux bords). Nous avons alors utilisé le schéma numérique de Roe [15] pour obtenir une solution approchée. Cependant, nous nous sommes heurtés à une difficulté récurrente dans la résolution de ces problèmes en exhibant une solution non physique.
Dans cette partie, nous cherchons un nouveau type de schémas hybrides avec paramètre. L'ajustement de ce paramètre doit permettre d'éviter l'apparition de solutions non physiques tout en assurant un degré de précision acceptable.
Ainsi, nous nous intéressons aux systèmes de loi de conservation de façon générale :

$$\begin{aligned} &\partial_t u + \partial_x f(u) = 0,\ x \in \mathbb{R},\ t \geq 0, \\ &u(x,0) = u^0(x). \end{aligned} \tag{1}$$

Nous supposons que u appartient à un ouvert Ω de $\mathbb{R}^p$. Le flux physique $f : \Omega \to \mathbb{R}^p$ est supposé être localement lipschitzien. Afin d'obtenir un problème bien posé, nous supposons que le système est hyperbolique, ce qui signifie que pour tout état u dans Ω, la matrice jacobienne $A(u)$ du flux :

$$A(u) = \partial_u f(u)$$

est diagonalisable dans $\mathbb{R}$. Par souci de simplicité, nous supposons de plus que le système est strictement hyperbolique, c'est-à-dire que pour tout $u \in \Omega$, chaque valeur propre $\alpha_k(u)$ est de multiplicité un [6]. Dans toute la suite, nous écrirons les valeurs propres dans l'ordre croissant

$$\alpha_1(u) < \alpha_2(u) < ... < \alpha_p(u).$$

Nous désirons trouver un schéma numérique qui permette d'obtenir la solution physique. Nous désirons de plus que ce schéma conserve la propriété TVD et ne soit pas trop diffusif.
La notion de solution physique est la suivante. Dans toute la suite, nous supposons que le système hyperbolique possède une paire entropie-flux $(S(u), F(u))$ telle que l'entropie $S : \Omega \to \mathbb{R}$ soit une fonction strictement convexe. Nous supposons de plus que S est de classe $\mathcal{C}^2$ et F de classe $\mathcal{C}^1$. Ce couple doit satisfaire la relation de compatibilité [6], [17]

$$\partial_u S(u) \partial_u f(u) = \partial_u F(u). \tag{2}$$

D'après cette relation, les solutions régulières de (1) satisfont la loi de conservation scalaire supplémentaire

$$\partial_t S(u) + \partial_x F(u) = 0. \tag{3}$$

De manière générale, on demande que les solutions faibles de (1) satisfassent l'inégalité au sens des distributions

$$\partial_t S(u) + \partial_x F(u) \leq 0. \tag{4}$$

pour toute paire entropie-flux d'entropie. Sous certaines hypothèses, cette condition permet d'obtenir l'unicité de la solution de (1.1) voir par exemple [17],[6] . Cela permet de plus de sélectionner la solution faible physique parmi toutes les solutions faibles possibles. Cette solution est alors appelée solution entropique. Parmi les nombreux articles s'intéressant au sujet des solutions entropiques, citons les travaux de Jin et Xin qui proposent un schéma de relaxation [9], les travaux de LeFloch et Rohde [14], LeFloch, Mercier et Rohde [13] ou encore de Tadmor [21], [22].

Nous proposons le schéma suivant :

$$u_j^{n+1} = u_j^n - \lambda \left(\phi(u_j^n, u_{j+1}^n, \lambda, \theta_{j+\frac{1}{2}}) - \phi(u_{j-1}^n, u_j^n, \lambda, \theta_{j-\frac{1}{2}}) \right) \tag{5}$$

où le flux numérique dépend d'un paramètre réel θ :

$$\phi(u, v, \lambda, \theta) = \frac{f(u) + f(v)}{2} - \frac{1}{2} \lambda^\theta |A(u, v)|^{1+\theta} (v - u). \tag{6}$$

Le paramètre θ est déterminé en deux étapes. La première étape est une phase prédictrice où nous cherchons le paramètre θ permettant d'obtenir un schéma TVD qui soit le moins diffusif possible. La deuxième étape est une phase correctrice : dans les cellules où la dissipation d'entropie est positive, nous déterminons θ afin de rendre la dissipation négative. Le choix de l'approximation numérique de $\partial_t S(u) + \partial_x F(u)$ est important pour la qualité des résultats.

Le plan de cette partie est le suivant. Dans le premier chapitre, nous présentons le schéma hybride : définition du flux numérique hybride et équation équivalente. Dans le deuxième chapitre, nous déterminons les conditions sur θ pour obtenir un schéma TVD, d'ordre 2 en espace et 1 en temps dans le cas de loi de conservation scalaire non linéaire, puis nous étendons heuristiquement le schéma au cas des systèmes. Dans le troisième chapitre, nous présentons des expériences numériques afin de valider notre schéma. Enfin, dans le quatrième chapitre, nous expliquons comment corriger le paramètre θ afin de rendre le schéma entropique.

Un article sur cette partie est en cours de rédaction [3].

Mots-clés : lois de conservation, schémas numériques, entropie.

1

Présentation et analyse numérique du schéma hybride

1.1 Construction du schéma hybride

Nous allons d'abord rappeler les flux numériques classiques de Lax-Wendroff, Lax-Friedrichs, Lax-Friedrichs modifié, Roe et Rusanov. Puis nous allons mettre en évidence qu'il est possible d'interpoler trois de ces cinq flux numériques par le biais d'un seul paramètre réel θ, ce qui définira notre famille de flux hybrides.

1.1.1 Définition du flux numérique hybride

Nous nous intéressons à des schémas numériques pour l'approximation de solutions du système hyperbolique

$$\partial_t u + \partial_x f(u) = 0,\ x \in \mathbb{R},\ t \geq 0. \tag{1.1}$$

Nous considérons un maillage spatial uniforme avec un pas d'espace constant égal à h et notons par $x_j = jh$ les points du maillage. Nous posons $x_{j+\frac{1}{2}} = (j+\frac{1}{2})h$. Le pas de temps Δt^n est variable et nous posons $t^{n+1} = t^n + \Delta t^n$. Au temps discret t^n, les suites (u_j^n), $j \in \mathbb{Z}$ fournissent une approximation de la solution continue u :

$$u_j^n \approx \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t^n)\,dx.$$

Dans la suite, λ désigne le quotient du pas de temps par le pas d'espace, la dépendance de λ par rapport à n étant sous-entendue :

$$\lambda = \frac{\Delta t^n}{h}.$$

Les solutions de (1.1) sont approchées à l'aide de schémas numériques conservatifs :

$$u_j^{n+1} = u_j^n - \lambda\left(\phi(u_j^n, u_{j+1}^n, \lambda) - \phi(u_{j-1}^n, u_j^n, \lambda)\right) \tag{1.2}$$

où le flux numérique $\phi : \Omega \times \Omega \to \mathbb{R}^p$ est une fonction lipschitzienne sur $\Omega \times \Omega$. Pour certains flux numériques, il est possible d'avoir des inégalités d'entropie discrètes où apparaissent des flux

d'entropie numériques :

$$S(u_j^{n+1}) - S(u_j^n) + \lambda \left(\psi_{j+1/2}^n - \psi_{j-1/2}^n \right) \leq 0. \tag{1.3}$$

Le flux d'entropie numérique $\psi_{j+1/2}^n = \psi(u_j^n, u_{j+1}^n)$ est supposé être régulier et satisfaire la propriété de consistance $\psi(u,u) = F(u)$ [11]. Les schémas numériques qui satisfont une telle inégalité sont appelé schémas entropiques.

Différents schémas numériques conservatifs ont été proposés.

Schéma de Lax-Wendroff

Citons le schéma de Lax-Wendroff [12] dont le flux numérique s'écrit

$$\phi(u,v,\lambda) = \frac{f(u)+f(v)}{2} - \frac{1}{2}\lambda A(u,v)(f(v)-f(u)) \tag{1.4}$$

où $A(u,v)$ est une matrice continue en (u,v) qui a la propriété de consistance $A(u,u) = A(u)$. Le schéma de Lax-wendroff est le seul schéma à trois points du second ordre en temps et en espace. Il est stable pour des solutions régulières sous la condition de Courant-Friedrichs-Lewy (condition CFL) avec un nombre de Courant inférieur à un demi. Malheureusement, ce schéma est dispersif et produit de fortes oscillations pour les ondes de chocs. De plus ce schéma viole la propriété d'entropie au niveau discret et peut ainsi capturer des discontinuités non physiques telles que des chocs de raréfaction.

Schéma de Lax-Friedrichs

Le schéma de Lax-Friedrichs [4],[10] a un flux numérique qui s'écrit

$$\phi(u,v,\lambda) = \frac{f(u)+f(v)}{2} - \frac{1}{2\lambda}(v-u). \tag{1.5}$$

On peut montrer que le schéma de Lax-Friedrichs est stable sous une condition CFL avec un nombre de Courant inférieur à un. Malheureusement, il est uniquement du premier ordre et présente un caractère fortement dissipatif qui le rend peu approprié pour des simulations numériques pratiques.

Schéma de Lax-Friedrichs modifié

Le flux numérique du schéma de Lax-Friedrichs modifié [20] est

$$\phi(u,v,\lambda) = \frac{f(u)+f(v)}{2} - \frac{1}{4\lambda}(v-u). \tag{1.6}$$

On peut montrer que le schéma de Lax-Friedrichs modifié satisfait une inégalité d'entropie discrète pour toutes les paires entropie-flux. Cependant, ce schéma a le même inconvénient que le schéma de Lax-Friedrichs car il est très dissipatif.

Schéma de Roe

Un autre schéma couramment utilisé est le schéma de Roe [15] dont le flux numérique s'écrit

$$\phi(u,v) = \frac{f(u)+f(v)}{2} - \frac{1}{2}|\bar{A}(u,v)|(v-u). \tag{1.7}$$

Nous rappelons que la matrice de Roe $\bar{A}(u,v)$ doit satisfaire trois conditions

$$\forall u \in \Omega,\ \bar{A}(u,u) = \partial_u f(u),$$

$$\forall (u,v) \in \Omega^2,\ \bar{A}(u,v) \text{ est diagonalisable,}$$

$$\forall (u,v) \in \Omega^2,\ \bar{A}(u,v)(v-u) = f(v) - f(u).$$

Une telle linéarisation de Roe existe si le système hyperbolique possède une paire entropie-flux d'entropie [8]. Le schéma de Roe est stable sous la condition d'un nombre de Courant inférieur à un. Ce schéma est relativement précis en particulier pour les discontinuités se déplaçant lentement. Mais ce schéma viole aussi la propriété d'entropie. En particulier, les chocs stationnaires non physiques sont stables et des chocs de raréfactions peuvent apparaître à travers des détentes soniques. Harten a proposé une modification dans le calcul du flux qui permet d'obtenir un schéma entropique à partir du schéma de Roe [7].

Schéma de Rusanov

Enfin, nous présentons le schéma de Rusanov [16] dont le flux numérique s'écrit

$$\phi(u,v,\lambda) = \frac{f(u)+f(v)}{2} - \frac{1}{2}\alpha_{max}(v-u). \tag{1.8}$$

où α_{max} est le rayon spectral de la matrice $A(u,v)$.

Ces schémas ont à la fois leurs avantages et leurs inconvénients. Nous introduisons un schéma numérique hybride qui permet de passer de l'un à l'autre des schémas suivant les cas. Tout d'abord remarquons que le schéma de Lax-Wendroff, le schéma de Roe et le schéma de Lax-Friedrichs peuvent s'écrire sous la forme générale :

$$u_j^{n+1} = u_j^n - \lambda\left(\phi(u_j^n, u_{j+1}^n, \lambda, \theta) - \phi(u_{j-1}^n, u_j^n, \lambda, \theta)\right) \tag{1.9}$$

où le flux numérique dépend d'un paramètre réel θ :

$$\phi(u,v,\lambda,\theta) = \frac{f(u)+f(v)}{2} - \frac{1}{2}\lambda^\theta |A(u,v)|^{1+\theta}(v-u). \tag{1.10}$$

Lorsque la matrice $A(u,v)$ est une matrice de Roe, nous retrouvons le flux de Lax-Wendroff pour $\theta = 1$, le flux de Roe pour $\theta = 0$ et le flux de Lax-Friedrichs pour $\theta = -1$. Nous pouvons interpréter θ comme un paramètre de viscosité numérique. Lorsque $\theta = 1$, la viscosité numérique est faible et lorsque $\theta = -1$, la viscosité est la plus grande.
Nous voulons tirer partie des avantages complémentaires de ces schémas en utilisant une variable adaptative θ, a priori dans $[-1,1]$. Ainsi dans toute la suite, θ est une fonction de l'état u. Au niveau discret, nous indicerons la variable θ comme une fonction classique de l'espace et du temps par $\theta_{j+\frac{1}{2}}^n$.

Schéma hybride

Nous considérons ainsi le schéma conservatif suivant :

$$u_j^{n+1} = u_j^n - \lambda \left(\phi(u_j^n, u_{j+1}^n, \lambda, \theta_{j+\frac{1}{2}}^n) - \phi(u_{j-1}^n, u_j^n, \lambda, \theta_{j-\frac{1}{2}}^n) \right) \tag{1.11}$$

avec le flux numérique

$$\phi(u, v, \lambda, \theta) = \frac{f(u) + f(v)}{2} - \frac{1}{2}\lambda^\theta \left| A(u, v) \right|^{1+\theta} (v - u), \tag{1.12}$$

$$\theta_{j+\frac{1}{2}}^n = \theta(u_{j-1}^n, u_j^n, u_{j+1}^n, u_{j+2}^n, \lambda) \in \mathbb{R}. \tag{1.13}$$

Le schéma n'est donc plus un schéma à trois points puisque le flux numérique dépend de $(u_{j-1}^n, u_j^n, u_{j+1}^n, u_{j+2}^n)$ par le biais de θ . L'intérêt d'introduire la variable θ est clair : nous pouvons espérer trouver un algorithme qui donne la valeur 1 à θ dans les régions où la solution est régulière afin d'obtenir une précision au second ordre. D'un autre côté, dans une région où la solution est discontinue ou bien lorsque le gradient est très important, l'algorithme doit utiliser un schéma stable du premier ordre faiblement diffusif, c'est-à-dire utiliser une valeur de θ proche de zéro. Si cela est nécessaire, l'algorithme peut produire une diffusion numérique importante avec des valeurs de θ proche de -1. La possibilité pour θ de devenir inférieur à zéro peut être exploitée pour produire plus de dissipation numérique que le schéma de Roe, en particulier pour des ondes de raréfaction aux points soniques, lorsque le schéma de Roe créé en général des chocs d'entropie non physiques. En d'autres termes, nous pouvons espérer obtenir un schéma entropique.

1.1.2 Variantes du schéma hybride

Nous pouvons choisir différentes variantes du schéma hybride présenté dans le paragraphe précédent. Nous en citons quelques exemples.

Schéma hybride LW-upwind-LF

Remarquons que notre schéma ne nécessite pas forcément le calcul d'une matrice de Roe. Par exemple, nous pouvons utiliser la matrice $A(u, v) = \partial_u f(\frac{u+v}{2})$. Dans ce cas, bien sûr, le schéma hybride pour $\theta = 0$ n'est plus le schéma de Roe. L'avantage est que le schéma est plus facilement utilisable, seule une diagonalisation de $\partial_u f(u)$ étant nécessaire.

Schéma hybride LW-VFFC-LF

Une autre façon d'éviter le calcul d'une matrice de Roe est d'écrire le schéma hybride sous la forme d'un schéma VFFC ([5]). En effet, nous pouvons remarquer que le flux s'écrit

$$\phi(u, v, \lambda, \theta) = \frac{f(u) + f(v)}{2} - \frac{1}{2}\text{sign}(A)\lambda^\theta \left| A(u, v) \right|^\theta (f(v) - f(u))$$

pour $\theta \geq 0$ dans le cas où A est une matrice de Roe. Ainsi nous pouvons écrire

$$\phi(u, v, \lambda, \theta) = \frac{f(u) + f(v)}{2} - \frac{1}{2}\lambda^\theta B(f(v) - f(u))$$

où $B = \text{sign}(A) \left| A(u, v) \right|^\theta$, où A peut être une matrice différente d'une matrice de Roe, telle que $A(u, u) = \partial_u f(u)$. Précisons que ces schémas ne sont pas entropiques.

LW-Roe-Rusanov

Nous pouvons introduire avec cette méthode plusieurs types de schémas hybrides. Par exemple, nous pouvons prendre dans le schéma (1.11) le flux numérique

$$\phi(u,v,\lambda,\theta) = \frac{f(u)+f(v)}{2} - \frac{1}{2}\lambda^{\theta^+}\,\alpha_{max}^{\theta^++1}|\frac{1}{\alpha_{max}}A(u,v)|^{1+\theta}(v-u) \tag{1.14}$$

où $\theta^+ = \max(0,\theta)$. Pour ce schéma hybride, nous obtenons toujours les schémas de Lax-Wendroff et de Roe pour $\theta = 1$ et $\theta = 0$ respectivement et le schéma de Rusanov pour $\theta = -1$.

LW-Roe-LFM ou LW-upwind-LFM

De même, nous pouvons obtenir un schéma hybride pour lequel le cas $\theta = 1$ correspond au schéma de Lax-Wendroff et le cas $\theta = -1$ au schéma de Lax-Friedrichs modifié :

$$\phi(u,v,\lambda,\theta) = \frac{f(u)+f(v)}{2} - \frac{1}{2}\lambda^{\theta}2^{\theta_-}\,|A(u,v)|^{1+\theta}(v-u)$$

où $\theta_- = \min(0,\,\theta)$. Ce type de schéma hybride permet d'obtenir le schéma de Lax-Friedrichs modifié pour $\theta = -1$ qui est entropique. Nous utiliserons ce type de schéma dans nos applications numériques.

1.1.3 Forme vectorielle du schéma hybride

La technique d'hybridation présentée ci-dessus introduit un paramètre scalaire. Il est naturel d'introduire une technique d'hybridation plus sophistiquée en introduisant un paramètre à valeurs vectorielles $\vec{\theta} = (\theta_k)_{k=1,\ldots,p}$. Pour cela, nous supposons que la matrice $A(u,v)$ se diagonalise sous la forme $D(u,v) = diag(\alpha_k(u,v))$ suivant la matrice de vecteurs propres à droite $R(u,v)$:

$$A(u,v) = R(u,v)D(u,v)R(u,v)^{-1}.$$

Dans ce qui suit, nous utiliserons la notation suivante pour un exposant "vectoriel" de matrice défini par :

$$|A(u,v)|^{\vec{\theta}} = R(u,v)\,diag(|\alpha_k(u,v)|^{\theta_k})R(u,v)^{-1}. \tag{1.15}$$

Dans ce cas, le flux numérique s'écrit

$$\phi(u,v,\lambda,\vec{\theta}) = \frac{f(u)+f(v)}{2} - \frac{1}{2}\,(\lambda\mathbf{I})^{\vec{\theta}}\,|A(u,v)|^{\mathbf{e}+\vec{\theta}}(v-u) \tag{1.16}$$

avec $\mathbf{e} = (1,\ldots,1)$; la matrice $\mathbf{I}$ est la matrice identité dans $\mathbb{R}^p$.

Bien sûr, nous pouvons étendre les variantes du schéma hybride présentées ci-dessus au cas vectoriel.

1.1.4 Equations équivalentes pour les formes scalaire et vectorielle

Nous voulons étudier l'influence de θ sur le schéma numérique, en particulier en ce qui concerne la diffusion numérique. Nous donnons dans ce paragraphe les équations équivalentes associées respectivement à la forme scalaire et à la forme vectorielle du schéma hybride.

Nous écrivons ainsi le développement de Taylor au voisinage du point (x_j, t^n) au second ordre dans (1.11), (1.12). Nous obtenons le système d'EDP que le schéma numérique résout avec une précision du second ordre en temps et en espace. Pour simplifier les calculs, nous supposons que θ est une fonction localement constante près du point (x_j, t^n). Un calcul simple montre que l'équation équivalente de la forme scalaire de l'hybridation est

$$\partial_t u + \partial_x f(u) + \frac{\Delta t^n}{2}\partial_x[A^2(u)\partial_x u] - \frac{1}{2}(\Delta t^n)^\theta h^{1-\theta}\partial_x\left[|A(u)|^{1+\theta}\partial_x u\right] = 0. \tag{1.17}$$

Nous pouvons écrire de façon équivalente

$$\partial_t u + \partial_x f(u) - \frac{\Delta t^n}{2}\partial_x\left[\left((\lambda)^{\theta-1}|A(u)|^{\theta+1} - A^2(u)\right)\partial_x u\right] = 0.$$

La condition pour avoir une matrice de diffusion positive est exprimée à l'aide des valeurs propres α_k de $A(u)$:

$$(\lambda|\alpha_k|)^{\theta-1} - 1 \geq 0 \quad \forall k \in \{1, ..., p\}.$$

Pour $\theta \leq 1$, nous retrouvons les conditions classiques de Courant-Friedrichs-Lewy concernant le pas de temps Δt^n :

$$\lambda|\alpha_k| \leq 1 \quad \forall k. \tag{1.18}$$

Nous pouvons remarquer que pour $\theta = 1$, la matrice de diffusion est exactement zéro ; le schéma numérique correspondant est en effet le schéma de Lax-Wendroff. Dans le cas contraire, l'intensité de la matrice de diffusion est gouvernée par ses valeurs propres. Ces valeurs propres sont

$$\left((\lambda|\alpha_k|)^{\theta-1} - 1\right)(\alpha_k)^2 \text{ si } \alpha_k \neq 0 \text{ et } 0 \text{ sinon.}$$

Pour $\lambda|\alpha^k| > 0$, la fonction

$$q : \theta \mapsto (\alpha_k)^2((\lambda|\alpha_k|)^{\theta-1} - 1)$$

est de classe $\mathcal{C}^\infty$ sur $[-1, +\infty]$. De plus,

$$q'(\theta) = (\alpha_k)^2 \ln(\lambda|\alpha_k|)(\lambda|\alpha_k|)^{\theta-1} < 0$$

pour tout θ, sous la condition CFL (1.18). Cela signifie en particulier que l'intensité de la matrice de diffusion est une fonction décroissante du paramètre θ. De plus, nous passons du schéma de Lax-Wendroff au schéma de Lax-Friedrichs de façon très régulière.

Remarque 2 *Pour $\theta \geq 1$, la fonction f est négative et la limite de f lorsque θ tend vers $+\infty$ vaut $-(\alpha_k)^2$. Ainsi pour des valeurs de θ supérieures à 1 la matrice $\lambda^\theta|A(u)|^{\theta+1} - A^2(u)$ est une matrice d'antidiffusion dont l'intensité reste bornée lorsque θ tend vers $+\infty$.*

Les calculs effectués pour la forme scalaire de l'hybridation peuvent être étendus sans difficulté à la forme vectorielle avec des conclusions analogues. L'équation équivalente est alors :

$$\partial_t u + \partial_x f(u) - \frac{\Delta t^n}{2}\partial_x\left[\left((\lambda)^{\vec{\theta}-e}|A(u)|^{e+\vec{\theta}} - A^2(u)\right)\partial_x u\right] = 0$$

avec les conventions précisées ci-dessus.

2

Détermination des fonctions θ

2.1 Cas de l'équation de transport linéaire scalaire

Avant de nous attaquer au problème plus général et plus difficile d'une équation de loi de conservation scalaire non linéaire, nous nous intéressons au cas du transport linéaire. Nous désirons trouver les fonctions θ qui permettent d'obtenir un schéma TVD. Nous considérons donc l'équation

$$\partial_t u + a\partial_x u = 0, \tag{2.1}$$

où a est une constante. Pour fixer les idées, nous pouvons supposer par exemple que $a > 0$. Notre schéma hybride dans le cas scalaire s'écrit simplement :

$$u_j^{n+1} = u_j^n - \lambda \left(\phi_{j+\frac{1}{2}} - \phi_{j-\frac{1}{2}} \right), \tag{2.2}$$

avec le flux numérique

$$\phi_{j+\frac{1}{2}} = a\frac{u_j^n + u_{j+1}^n}{2} - \frac{1}{2}\lambda^{\theta^n_{j+\frac{1}{2}}} a^{1+\theta^n_{j+\frac{1}{2}}} \left(u_{j+1}^n - u_j^n \right). \tag{2.3}$$

2.1.1 Analyse TVD par le critère d'Harten et stabilité du schéma hybride

Nous rappelons les définitions suivantes (avec les notations précédentes).

Définition 1 *Un schéma*

$$u_j^{n+1} = u_j^n - \lambda(\phi_{j+\frac{1}{2}} - \phi_{j-\frac{1}{2}})$$

est dit TVD (Total Variation Diminishing) lorsque

$$\forall u^0,\, TV(u^1) \leq TV(u^0)$$

où $TV(u) = \sum_{j\in\mathbb{Z}} |u_{j+1} - u_j|$.

Définition 2 *Un schéma*

$$u_j^{n+1} = u_j^n - \lambda(\phi_{j+\frac{1}{2}} - \phi_{j-\frac{1}{2}})$$

est dit L^∞-stable lorsqu'il existe une constante K indépendante de Δt et h telle que

$$\forall u^0,\, \forall n,\, \max_{j\in\mathbb{Z}} |u_j^n| \leq K \max_{j\in\mathbb{Z}} |u_j^0|.$$

Posons

$$\nu = a\lambda, \quad \Delta u_{j+\frac{1}{2}} = u^n_{j+1} - u^n_j.$$

En suivant les idées d'Harten [7] et de Sweby [19], nous écrivons le schéma numérique sous la forme

$$u^{n+1}_j = u^n_j - \frac{\nu}{2}\left(1 - \nu^{\theta^n_{j+\frac{1}{2}}}\right)\Delta u_{j+\frac{1}{2}} + \left(1 + \nu^{\theta^n_{j-\frac{1}{2}}}\right)\Delta u_{j-\frac{1}{2}}. \tag{2.4}$$

Nous voulons déterminer les fonctions θ telles que le schéma hybride possède la propriété TVD. Nous posons

$$r_j = \frac{\Delta u_{j-\frac{1}{2}}}{\Delta u_{j+\frac{1}{2}}}.$$

Ainsi le schéma numérique peut s'écrire (dans le cas $a > 0$)

$$u^{n+1}_j = u^n_j - \frac{\nu}{2}\left(\frac{1 - \nu^{\theta^n_{j+\frac{1}{2}}}}{r_j} + (1 + \nu^{\theta^n_{j-\frac{1}{2}}})\right)\Delta u_{j-\frac{1}{2}} \tag{2.5}$$

ce qui donne la forme incrémentale

$$u^{n+1}_j = u^n_j + C^n_{j+\frac{1}{2}}\Delta u_{j+\frac{1}{2}} - D^n_{j-\frac{1}{2}}\Delta u_{j-\frac{1}{2}} \tag{2.6}$$

de coefficients

$$\begin{cases} C^n_{j+\frac{1}{2}} = 0, \\ D^n_{j-\frac{1}{2}} = \dfrac{\nu}{2}\left(\dfrac{1 - \nu^{\theta^n_{j+\frac{1}{2}}}}{r_j} + (1 + \nu^{\theta^n_{j-\frac{1}{2}}})\right). \end{cases}$$

La proposition suivante donne une condition suffisante pour avoir un schéma TVD [7].

Proposition 6 *Un schéma écrit sous une forme incrémentale*

$$u^{n+1}_j = u^n_j + C^n_{j+\frac{1}{2}}\Delta u_{j+\frac{1}{2}} - D^n_{j-\frac{1}{2}}\Delta u_{j-\frac{1}{2}} \tag{2.7}$$

est TVD sous la condition suffisante : pour tout n et j,

$$C^n_{j+\frac{1}{2}} \geq 0, \ D^n_{j+\frac{1}{2}} \geq 0$$

et

$$C^n_{j+\frac{1}{2}} + D^n_{j+\frac{1}{2}} \leq 1.$$

A l'aide de la forme incrémentale, nous pouvons montrer qu'un schéma est L^∞-stable :

Proposition 7 *Un schéma écrit sous une forme incrémentale*

$$u^{n+1}_j = u^n_j + C^n_{j+\frac{1}{2}}\Delta u_{j+\frac{1}{2}} - D^n_{j-\frac{1}{2}}\Delta u_{j-\frac{1}{2}} \tag{2.8}$$

est L^∞-stable sous la condition suffisante : pour tout n et j,

$$C^n_{j+\frac{1}{2}} \geq 0, \ D^n_{j+\frac{1}{2}} \geq 0$$

et

$$C^n_{j+\frac{1}{2}} + D^n_{j-\frac{1}{2}} \leq 1.$$

Dans notre cas, les critères pour montrer qu'un schéma possède la propriété TVD et la stabilité L^∞ deviennent simplement $0 \leq D^n_{j+\frac{1}{2}} \leq 1$. Nous cherchons donc $\theta^n_{j+\frac{1}{2}}$ qui permet de satisfaire ces inégalités.
Nous supposons dans toute la suite que la condition CFL $\nu < 1$ est satisfaite. Nous cherchons alors les fonctions θ telles que pour tous r, r',

$$0 \leq \frac{\nu}{2}\left(\frac{1-\nu^{\theta(r,\nu)}}{r} + 1 + \nu^{\theta(r',\nu)}\right) \leq 1, \tag{2.9}$$

c'est-à-dire

$$-1 \leq \frac{1-\nu^{\theta(r,\nu)}}{r} + \nu^{\theta(r',\nu)} \leq \frac{2}{\nu} - 1 \quad \forall r, r'. \tag{2.10}$$

Considérons en premier l'inégalité gauche de (2.10).

Inégalité $-1 \leq \frac{1-\nu^{\theta(r,\nu)}}{r} + \nu^{\theta(r',\nu)}$

Nous considérons un paramètre intermédiaire α dans $[0;1]$. Nous cherchons alors les conditions sur θ telles que

$$\begin{cases} \alpha \leq \nu^{\theta(r',\nu)} \\ -1-\alpha \leq \frac{1-\nu^{\theta(r,\nu)}}{r} \end{cases} \tag{2.11}$$

La première inégalité $\alpha \leq \nu^{\theta(r',\nu)}$ est vérifiée si et seulement si $\theta(r',\nu) \leq \frac{\ln\alpha}{\ln\nu}$. Pour la seconde inégalité $-1-\alpha \leq \frac{1-\nu^{\theta(r,\nu)}}{r}$, nous distinguons les deux cas $r \geq 0$ et $r < 0$.
Pour le premier cas $r \geq 0$, l'inégalité est vérifiée si et seulement si $\theta(r,\nu) \geq \frac{\ln(1+(1+\alpha)r)}{\ln\nu}$. Pour le second cas $r < 0$, si $r \leq \frac{-1}{1+\alpha}$, l'inégalité est toujours vérifiée. Si $\frac{-1}{1+\alpha} < r < 0$, l'inégalité est vérifiée ssi $\theta(r,\nu) \leq \frac{\ln(1+(1+\alpha)r)}{\ln\nu}$. Nous remarquons que si $r \in [\frac{-1}{1+\alpha}; \frac{-1+\alpha}{1+\alpha}]$, alors $\frac{\ln(1+(1+\alpha)r)}{\ln\nu} \geq \frac{\ln\alpha}{\ln\nu}$. Ces conditions sont résumées dans le tableau ci-dessous.

r	$-\infty$	$\frac{-1+\alpha}{1+\alpha}$	0	$+\infty$
Conditions sur θ	$\theta \leq \frac{\ln\alpha}{\ln\nu}$	$\theta \leq \frac{\ln(1+(1+\alpha)r)}{\ln\nu}$	$\frac{\ln(1+(1+\alpha)r)}{\ln\nu} \leq \theta \leq \frac{\ln\alpha}{\ln\nu}$	

Pour le cas $\alpha = 0$, nous obtenons les mêmes résultats en remplaçant $\ln\alpha$ par $-\infty$.
Nous considérons maintenant l'inégalité de droite dans (2.10).

Inégalité $\frac{1-\nu^{\theta(r,\nu)}}{r} + \nu^{\theta(r',\nu)} \leq \frac{2}{\nu} - 1$

Nous considérons un réel β dans $[0;1]$. Nous cherchons alors les conditions sur θ telles que

$$\begin{cases} \nu^{\theta(r',\nu)} \leq \beta(\frac{2}{\nu} - 1), \\ \frac{1-\nu^{\theta(r,\nu)}}{r} \leq (1-\beta)(\frac{2}{\nu} - 1). \end{cases} \tag{2.12}$$

La première inégalité $\nu^{\theta(r',\nu)} \leq \beta(\frac{2}{\nu} - 1)$ est vérifiée si et seulement si $\theta(r',\nu) \geq \frac{\ln\beta(\frac{2}{\nu}-1)}{\ln\nu}$. Pour la seconde inégalité de (2.12), nous distinguons les deux cas $r \geq 0$ et $r < 0$. Pour le premier cas $r < 0$, l'inégalité est vérifiée *ssi* $\theta(r,\nu) \geq \frac{\ln(1-(\frac{2}{\nu}-1)(1-\beta)r)}{\ln\nu}$. Pour le second cas $r \geq 0$, si $0 \leq r < \frac{1}{(\frac{2}{\nu}-1)(1-\beta)}$, alors la condition est $\theta(r,\nu) \leq \frac{\ln(1-(\frac{2}{\nu}-1)(1-\beta)r)}{\ln\nu}$. Si $r > \frac{1}{(\frac{2}{\nu}-1)(1-\beta)}$, alors

l'inégalité est automatiquement vérifiée.
Revenons au cas $r \in [0; \frac{1}{(\frac{2}{\nu}-1)(1-\beta)}]$. Pour $r' = r$, l'équation (2.12) entraîne que β doit satisfaire

$$\frac{\ln \beta(\frac{2}{\nu}-1)}{\ln \nu} \leq \theta(r,\nu) \leq \frac{\ln(1-(\frac{2}{\nu}-1)(1-\beta)r)}{\ln \nu}.$$

Cette condition est équivalente à

$$\beta \geq \frac{1}{\frac{2}{\nu}-1}. \tag{2.13}$$

Dans ce cas, les deux conditions : $\theta(r,\nu) \geq \frac{\ln(1-(\frac{2}{\nu}-1)(1-\beta)r)}{\ln \nu}$ et $\theta(r,\nu) \geq \frac{\ln \beta(\frac{2}{\nu}-1)}{\ln \nu}$ peuvent être résumées par $\theta(r,\nu) \geq \frac{\ln \beta(\frac{2}{\nu}-1)}{\ln \nu}$ sur $]-\infty; \frac{1-\beta(\frac{2}{\nu}-1)}{(1-\beta)(\frac{2}{\nu}-1)}]$ et $\theta(r,\nu) \geq \frac{\ln(1-(\frac{2}{\nu}-1)(1-\beta)r)}{\ln \nu}$ sur $[\frac{1-\beta(\frac{2}{\nu}-1)}{(1-\beta)(\frac{2}{\nu}-1)}; 0]$. Ces résultats sont résumés dans le tableau ci-dessous :

r	$-\infty$	$\frac{1-\beta(\frac{2}{\nu}-1)}{(1-\beta)(\frac{2}{\nu}-1)}$	0
θ	$\theta \geq \frac{\ln \beta(\frac{2}{\nu}-1)}{\ln \nu}$		$\theta \geq \frac{\ln(1-(\frac{2}{\nu}-1)(1-\beta)r}{\ln \nu}$

r	0	$\frac{1}{(\frac{2}{\nu}-1)(1-\beta)}$	$+\infty$
θ	$\frac{\ln \beta(\frac{2}{\nu}-1)}{\ln \nu} \leq \theta \leq \frac{\ln(1-(\frac{2}{\nu}-1)(1-\beta)r)}{\ln \nu}$		$\theta \geq \frac{\ln \beta(\frac{2}{\nu}-1)}{\ln \nu}$

2.1.2 Zone TVD et dissipation numérique minimale

Comme nous l'indique notre étude précédente, plusieurs types de régions TVD existent. Elles dépendent à la fois des deux paramètres α et β. Nous désirons identifier une région qui permet d'avoir un schéma TVD avec une dissipation minimale. Nous devons ainsi trouver une région TVD où θ prend les valeurs les plus grandes possibles. Les conditions sur α ne permettent pas la détermination d'une région TVD que nous pourrions qualifier de meilleure qu'une autre. Par souci de simplification des conditions sur θ, nous choisissons $\alpha = 0$.
Pour choisir β, nous fixons un réel $r > 0$. Nous voulons que la borne supérieure $\frac{\ln(1-(\frac{2}{\nu}-1)(1-\beta)r)}{\ln \nu}$ soit la plus grande possible en fonction de β. Or cette fonction de β (à r fixé) est décroissante, ainsi il est suffisant de poser $\beta = \frac{1}{\frac{2}{\nu}-1}$. Dans ce cas, les nouvelles conditions provenant de β pour θ sont :

r	$-\infty$	0	$\frac{1}{2(\frac{1}{\nu}-1)}$	$+\infty$
θ	$\theta \geq 0$	$0 \leq \theta \leq \frac{\ln(1-2(\frac{1}{\nu}-1)r)}{\ln \nu}$	$\theta \geq 0$	

En regroupant les conditions provenant de α et β, nous obtenons finalement les conditions suivantes sur θ que nous résumons dans la proposition ci-dessous.

Proposition 8 *Soit une fonction θ satisfaisant les conditions*

r	$-\infty$	-1	0	$\frac{1}{2(\frac{1}{\nu}-1)}$	$+\infty$
θ	$\theta \geq 0$	$0 \leq \theta \leq \frac{\ln(1+r)}{\ln \nu}$	$0 \leq \theta \leq \frac{\ln(1-2(\frac{1}{\nu}-1)r)}{\ln \nu}$	$\theta \geq 0$	

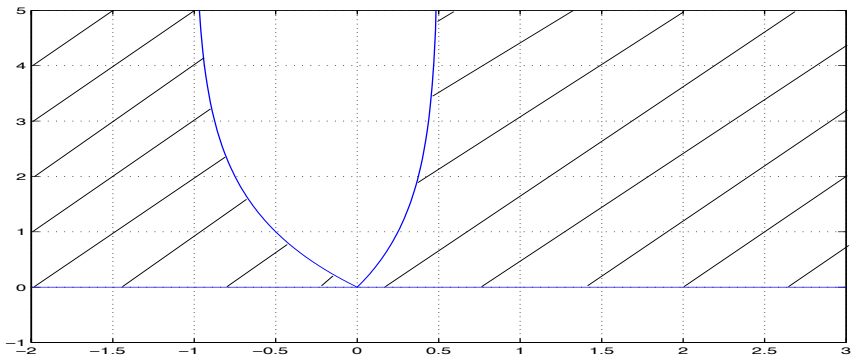

FIG. 2.1 – Région TVD (zone hachurée) dans le cas $\alpha = 0$ et $\beta = \frac{1}{\frac{2}{\nu}-1}$, avec en abscisse r. La condition CFL est ici $\nu = 0.5$.

alors le schéma hybride (2.2)-(2.3) *dans le cas linéaire scalaire possède la propriété TVD et la stabilité* L^∞.

Remarquons que la proposition 8 donne un exemple parmi d'autres de conditions suffisantes sur la fonction θ. La zone où doit se situer le paramètre θ en fonction de r est représentée sur la figure 2.1

2.2 Cas de l'équation scalaire non linéaire

Nous considérons maintenant l'équation de loi de conservation scalaire non linéaire :

$$\partial_t u + \partial_x f(u) = 0. \tag{2.14}$$

Nous avons le schéma suivant :

$$u_j^{n+1} = u_j^n - \lambda(\phi_{j+\frac{1}{2}}^n - \phi_{j-\frac{1}{2}}^n) \tag{2.15}$$

où

$$\phi_{j+\frac{1}{2}} = \frac{f(u_{j+1}^n) + f(u_j^n)}{2} - \frac{1}{2}\lambda^{\theta_{j+\frac{1}{2}}^n} |a_{j+\frac{1}{2}}|^{1+\theta_{j+\frac{1}{2}}^n} (u_{j+1}^n - u_j^n) \tag{2.16}$$

est le flux numérique et où $a_{j+\frac{1}{2}} = a(u_j^n, u_{j+1}^n)$ et $(u,v) \mapsto a(u,v)$ est la fonction définie par

$$a(u,v) = \begin{cases} f'(u) \text{ si } u = v \\ \frac{f(v)-f(u)}{v-u} \text{ sinon.} \end{cases}$$

Nous rappelons que $\Delta u_{j+\frac{1}{2}} = u_{j+1}^n - u_j^n$. Nous considérons qu'il existe une constante μ (ne dépendant ni de j ni de n) telle que la condition $|\frac{dt}{dx} f'(u)| \leq \mu$ est satisfaite. Afin de trouver la zone TVD, nous nous inspirons de l'étude des schémas avec limiteurs de flux ([19],[6]). Comme dans le cas linéaire, nous allons écrire le schéma sous une forme incrémentale.

2.2.1 Forme incrémentale du schéma

Ainsi le schéma (2.15)-(2.16) peut s'écrire :

$$u_j^{n+1} = u_j^n - \lambda \left(\frac{f_{j+1}^n + f_j^n}{2} - \frac{1}{2}\lambda^{\theta_{j+\frac{1}{2}}} |a_{j+\frac{1}{2}}|^{1+\theta^n_{j+\frac{1}{2}}} \Delta u^n_{j+\frac{1}{2}} - \left(\frac{f_j^n + f_{j-1}^n}{2} - \frac{1}{2}\lambda^{\theta_{j-\frac{1}{2}}} |a_{j-\frac{1}{2}}|^{1+\theta^n_{j-\frac{1}{2}}} \Delta u^n_{j-\frac{1}{2}} \right) \right). \tag{2.17}$$

Or

$$\frac{f_{j+1}^n + f_j^n}{2} - \frac{f_j^n + f_{j-1}^n}{2} = \frac{f_{j+1}^n - f_j^n}{2} + \frac{f_j^n - f_{j-1}^n}{2},$$

et nous avons supposé que $a(u,v)(v-u) = f(v) - f(u)$. Ainsi,

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2} \left((a_{j+\frac{1}{2}} - \lambda^{\theta_{j+\frac{1}{2}}} |a_{j+\frac{1}{2}}|^{1+\theta^n_{j+\frac{1}{2}}}) \Delta u^n_{j+\frac{1}{2}} + (a_{j-\frac{1}{2}} + \lambda^{1+\theta_{j-\frac{1}{2}}} |a_{j-\frac{1}{2}}|^{1+\theta^n_{j-\frac{1}{2}}}) \Delta u^n_{j-\frac{1}{2}} \right). \tag{2.18}$$

Nous réécrivons l'expression précédente sous la forme :

$$u_j^{n+1} = u_j^n - \frac{1}{2} \Big((\lambda a_{j+\frac{1}{2}} - |\lambda a_{j+\frac{1}{2}}|^{1+\theta^n_{j+\frac{1}{2}}}) \Delta u^n_{j+\frac{1}{2}} + (\lambda a_{j-\frac{1}{2}} + |\lambda a_{j-\frac{1}{2}}|^{1+\theta^n_{j-\frac{1}{2}}}) \Delta u^n_{j-\frac{1}{2}} \Big).$$

Nous définissons a^+ la partie positive de a et a^- la partie négative de a. Ainsi $a = a^+ + a^-$ et $|a| = a^+ - a^-$. Nous avons alors

$$\begin{aligned} u_j^{n+1} = u_j^n - \frac{1}{2} \Big(& \lambda a^+_{j+\frac{1}{2}} (1 - |\lambda a_{j+\frac{1}{2}}|^{\theta^n_{j+\frac{1}{2}}}) \Delta u^n_{j+\frac{1}{2}} \\ & + \lambda a^-_{j+\frac{1}{2}} (1 + |\lambda a_{j+\frac{1}{2}}|^{\theta^n_{j+\frac{1}{2}}}) \Delta u^n_{j+\frac{1}{2}} \\ & + \lambda a^+_{j-\frac{1}{2}} (1 + |\lambda a_{j-\frac{1}{2}}|)^{\theta^n_{j-\frac{1}{2}}}) \Delta u^n_{j-\frac{1}{2}} \\ & + \lambda a^-_{j-\frac{1}{2}} (1 - |\lambda a_{j-\frac{1}{2}}|)^{\theta^n_{j-\frac{1}{2}}}) \Delta u^n_{j-\frac{1}{2}}) \Big) \end{aligned}$$

Nous supposons désormais que ($a^+_{j-\frac{1}{2}} \neq 0$ ou $a^+_{j+\frac{1}{2}} = 0$) c'est-à-dire qu'il n'est pas possible d'avoir à la fois : ($a^+_{j-\frac{1}{2}} = 0$ et $a^+_{j+\frac{1}{2}} \neq 0$). Cette hypothèse est en particulier vérifiée si le cellule j ne se trouve pas au voisinage d'un point sonique. Pour simplifier, nous supposons donc que nous ne sommes pas au voisinage d'un point sonique.
Comme dans le cas linéaire, nous posons $\nu_j = \lambda a_j$, $\nu_j^+ = \lambda a_j^+$ et $\nu_j^- = \lambda a_j^-$. Dans ce cas,

$$\begin{aligned} u_j^{n+1} = u_j^n & - \frac{\nu^+_{j-\frac{1}{2}}}{2} \left((1 - |\nu_{j+\frac{1}{2}}|^{\theta^n_{j+\frac{1}{2}}}) \frac{\nu^+_{j+\frac{1}{2}} \Delta u^n_{j+\frac{1}{2}}}{\nu^+_{j-\frac{1}{2}} \Delta u^n_{j-\frac{1}{2}}} + (1 + |\nu_{j-\frac{1}{2}}|^{\theta^n_{j-\frac{1}{2}}}) \right) \Delta u^n_{j-\frac{1}{2}} \\ & + \frac{-\nu^-_{j+\frac{1}{2}}}{2} \left((1 + |\nu_{j+\frac{1}{2}}|^{\theta^n_{j+\frac{1}{2}}}) \quad + (1 - |\nu_{j-\frac{1}{2}}|^{\theta^n_{j-\frac{1}{2}}})) \frac{\nu^-_{j-\frac{1}{2}} \Delta u^n_{j-\frac{1}{2}}}{\nu^-_{j+\frac{1}{2}} \Delta u^n_{j+\frac{1}{2}}} \right) \Delta u^n_{j+\frac{1}{2}} \end{aligned}$$

avec les conventions $\frac{\nu^+_{j+\frac{1}{2}}}{\nu^+_{j-\frac{1}{2}}} = 0$ et $\frac{\nu^-_{j-\frac{1}{2}}}{\nu^-_{j+\frac{1}{2}}} = 0$ dès que le numérateur s'annule. Nous obtenons alors la proposition :

Proposition 9 *En absence de point sonique, le schéma* (2.15) *peut s'écrire sous la forme incrémentale*

$$u_j^{n+1} = u_j^n + C_{j+\frac{1}{2}} \Delta u^n_{j+\frac{1}{2}} - D_{j-\frac{1}{2}} \Delta u^n_{j-\frac{1}{2}}, \tag{2.19}$$

avec comme coefficients

$$D_{j-\frac{1}{2}} = \frac{\nu^+_{j-\frac{1}{2}}}{2} \left((1 - |\nu_{j+\frac{1}{2}}|^{\theta^n_{j+\frac{1}{2}}}) \frac{\nu^+_{j+\frac{1}{2}} \Delta u^n_{j+\frac{1}{2}}}{\nu^+_{j-\frac{1}{2}} \Delta u^n_{j-\frac{1}{2}}} + (1 + |\nu_{j-\frac{1}{2}}|^{\theta^n_{j-\frac{1}{2}}}) \right),$$

$$C_{j+\frac{1}{2}} = \frac{-\nu^-_{j+\frac{1}{2}}}{2} \left((1 + |\nu_{j+\frac{1}{2}}|^{\theta^n_{j+\frac{1}{2}}}) + (1 - |\nu_{j-\frac{1}{2}}|^{\theta^n_{j-\frac{1}{2}}}) \frac{\nu^-_{j-\frac{1}{2}} \Delta u^n_{j-\frac{1}{2}}}{\nu^-_{j+\frac{1}{2}} \Delta u^n_{j+\frac{1}{2}}} \right)$$

où $\nu_{j+\frac{1}{2}} = \lambda a_{j+\frac{1}{2}}$ *et* $\nu_{j-\frac{1}{2}} = \lambda a_{j-\frac{1}{2}}$.

Nous posons alors

$$r_j^+ = \frac{\nu^+_{j-\frac{1}{2}} \Delta u^n_{j-\frac{1}{2}}}{\nu^+_{j+\frac{1}{2}} \Delta u^n_{j+\frac{1}{2}}}, \tag{2.20}$$

$$r_j^- = \frac{\nu^-_{j+\frac{1}{2}} \Delta u^n_{j+\frac{1}{2}}}{\nu^-_{j-\frac{1}{2}} \Delta u^n_{j-\frac{1}{2}}}. \tag{2.21}$$

Nous décidons de définir l'exposant $\theta_{j+\frac{1}{2}}$ par

$$\theta_{j+\frac{1}{2}} = \begin{cases} \theta(r_j^+, |\nu_{j+\frac{1}{2}}|) & \text{si} \quad \nu_{j+\frac{1}{2}} \geq 0, \\ \theta(r_{j+1}^-, |\nu_{j+\frac{1}{2}}|) & \text{si} \quad \nu_{j+\frac{1}{2}} < 0 \end{cases} \tag{2.22}$$

où θ est une fonction sur laquelle nous allons imposer des conditions pour que le schéma soit TVD et L^∞-stable par les critères rappelés dans les propositions 6 et 7. Nous avons alors

$$D_{j+\frac{1}{2}} = \frac{\nu^+_{j+\frac{1}{2}}}{2} \left(\frac{1 - |\nu_{j+\frac{3}{2}}|^{\theta(r^+_{j+1})}}{r^+_{j+1}} + 1 + |\nu_{j+\frac{1}{2}}|^{\theta(r_j^+)} \right), \tag{2.23}$$

$$C_{j+\frac{1}{2}} = \frac{-\nu^-_{j+\frac{1}{2}}}{2} \left(1 + |\nu_{j+\frac{1}{2}}|^{\theta(r^-_{j+1})} + \frac{1 - |\nu_{j-\frac{1}{2}}|^{\theta(r_j^-)}}{r_j^-} \right). \tag{2.24}$$

2.2.2 Détermination de la région TVD et L^∞-stable

Nous voulons trouver une fonction $\theta(r, \nu)$ telle que $C_{j+\frac{1}{2}}$ et $D_{j+\frac{1}{2}}$ satisfassent les conditions :

$$D_{j+\frac{1}{2}} \geq 0, \tag{2.25}$$

$$C_{j+\frac{1}{2}} \geq 0, \tag{2.26}$$

$$C_{j+\frac{1}{2}} + D_{j+\frac{1}{2}} \leq 1, \tag{2.27}$$

pour la région TVD et la condition supplémentaire

$$C_{j+\frac{1}{2}} + D_{j-\frac{1}{2}} \leq 1 \tag{2.28}$$

pour que le schéma soit $L^\infty-$stable.
Suffisamment loin d'un point sonique (il suffit juste d'avoir le même signe pour $a_{j+\frac{1}{2}}$, $a_{j-\frac{1}{2}}$, $a_{j+\frac{3}{2}}$), nous avons en particulier soit ($D_{j+\frac{1}{2}} = 0$ et $D_{j-\frac{1}{2}} = 0$) soit ($C_{j+\frac{1}{2}} = 0$). Nous donnons dans la proposition suivante un exemple de conditions suffisantes pour que le schéma possède la propriété TVD et soit L^∞-stable.

Proposition 10 *Soit μ le nombre CFL. Soit une fonction $\theta(r,\nu)$ satisfaisant les conditions suivantes :*

r	$-\infty$	-1	0
θ	$0 \leq \theta$	$0 \leq \theta \leq \frac{\ln(1+r)}{\ln \nu}$	

r	0	$\frac{1}{2(\frac{1}{\mu}-1)}$	$+\infty$
θ	$0 \leq \theta \leq \frac{\ln(1-2(\frac{1}{\mu}-1)r}{\ln \nu}$	$0 \leq \theta$	

alors le schéma hybride (2.15)-(2.16) *avec les notations* (2.20)-(2.21) *et la définition de $\theta_{j+\frac{1}{2}}$ donnée par* (2.22) *possède la propriété TVD et la L^∞-stabilité.*

Preuve. Nous cherchons des conditions suffisantes sur la fonction θ pour obtenir un schéma TVD et L^∞-stable :

Conditions provenant de l'inégalité (2.25)

Nous supposons que $a_{j+\frac{1}{2}} > 0$ et donc $a_{j-\frac{1}{2}} > 0$ et $a_{j+\frac{3}{2}} > 0$. Ainsi loin d'un point sonique, $C_{j+\frac{1}{2}} = 0$. Pour simplifier l'écriture, posons $\nu' = \nu^+_{j+\frac{1}{2}}$ et $\nu'' = |\nu_{j+\frac{3}{2}}|$. Ainsi

$$D_{j+\frac{1}{2}} = \frac{\nu'}{2}\left(1 + \nu'^{\theta(r',\nu')} + \frac{1-\nu''^{\theta(r'',\nu'')}}{r''}\right). \tag{2.29}$$

L'inégalité (2.25) devient

$$\nu'(1 + \nu'^{\theta'} + \frac{1-\nu''^{\theta''}}{r''}) \geq 0. \tag{2.30}$$

Comme $\nu'^{\theta'} \geq 0$, il est suffisant (mais non nécessaire, voir l'étude du cas linéaire en fonction du paramètre α) d'avoir

$$\frac{1-\nu''^{\theta''}}{r''} \geq -1. \tag{2.31}$$

Cette inégalité est équivalente à

$$\begin{cases} \theta(r'',\nu'') \geq \frac{\ln(1+r'')}{\ln \nu''}, \text{ si } r'' \geq 0, \\ \theta(r'',\nu'') \leq \frac{\ln(1+r'')}{\ln \nu''}, \text{ si } -1 \leq r'' < 0, \end{cases} \tag{2.32}$$

(si $r'' < -1$, l'inégalité est toujours vérifiée).

Conditions provenant de l'inégalité (2.26)

De même en considérant que $a_{j+\frac{1}{2}} < 0$, nous avons loin d'un point sonique, $D_{j+\frac{1}{2}} = 0$. Posons $\nu = |\nu_{j-\frac{1}{2}}|$ et $\nu' = -\nu^-_{j+\frac{1}{2}}$.
Ainsi

$$C_{j+\frac{1}{2}} = \frac{\nu'}{2}\left(1 + \nu'^{\theta'} + \frac{1-\nu^\theta}{r}\right). \tag{2.33}$$

Nous n'obtenons aucune condition supplémentaire sur la fonction θ.

Conditions provenant de l'inégalité (2.27)

Par exemple, nous supposons que $a^-_{j+\frac{1}{2}} = 0$. Ainsi l'inégalité (2.27) est équivalente à

$$D_{j+\frac{1}{2}} \leq 1 \tag{2.34}$$

c'est-à-dire

$$\frac{\nu'}{2}\left(\frac{1-\nu''^{\theta(r'',\nu'')}}{r''} + 1 + \nu'^{\theta(r',\nu')}\right) \leq 1 \tag{2.35}$$

d'où

$$\frac{1-\nu''^{\theta(r'',\nu'')}}{r''} + \nu'^{\theta(r',\nu')} \leq \frac{2}{\nu'} - 1. \tag{2.36}$$

Comme dans le cas linéaire, nous considérons un paramètre réel intermédiaire β dans $[0;1]$. Nous cherchons alors les conditions sur θ telles que

$$\begin{cases} \nu'^{\theta(r',\nu')} \leq \beta(\frac{2}{\nu'} - 1), \\ \frac{1-\nu''^{\theta(r'',\nu'')}}{r''} \leq (1-\beta)(\frac{2}{\nu'} - 1). \end{cases} \tag{2.37}$$

La première inégalité $\nu'^{\theta(r',\nu')} \leq \beta(\frac{2}{\nu'} - 1)$ est équivalente à $\theta(r',\nu') \geq \frac{\ln\beta(\frac{2}{\nu'}-1)}{\ln\nu'}$. Pour la seconde inégalité $\frac{1-\nu''^{\theta(r'',\nu'')}}{r''} \leq (1-\beta)(\frac{2}{\nu'} - 1)$: nous avons supposé que $\nu' \leq \mu$ ainsi $\frac{2}{\nu'} - 1 \geq \frac{2}{\mu} - 1$. Il est alors suffisant que

$$\frac{1-\nu''^{\theta(r'',\nu'')}}{r''} \leq (1-\beta)(\frac{2}{\mu} - 1). \tag{2.38}$$

Nous distinguons les deux cas $r'' \geq 0$ et $r'' < 0$.

- Premier cas $r'' < 0$: nous obtenons $\theta(r'',\nu'') \geq \frac{\ln(1-(\frac{2}{\mu}-1)(1-\beta)r'')}{\ln\nu''}$.
- Second cas $r \geq 0$: si $0 < r < \frac{1}{(\frac{2}{\nu}-1)(1-\beta)}$, alors la condition est $\theta(r'',\nu'') \leq \frac{\ln(1-(\frac{2}{\mu}-1)(1-\beta)r'')}{\ln\nu''}$.
  Si $r'' > \frac{1}{(\frac{2}{\nu''}-1)(1-\beta)}$, alors l'inégalité est vérifiée.

Si nous supposons maintenant que $\nu^+_{j+\frac{1}{2}} = 0$, alors nous pouvons facilement prouver qu'il n'y a pas de nouvelles conditions. Comme dans le cas linéaire, nous déterminons quelles conditions sont redondantes.

Résumé des conditions sur θ

Après calculs, nous pouvons résumer les résultats dans les tableaux suivants. Le premier tableau donne les conditions sur la borne inférieure de θ.

r	$-\infty$	$\frac{1-\beta(\frac{2}{\nu}-1)}{(1-\beta)(\frac{2}{\nu}-1)}$	0
θ	$\theta \geq \frac{\ln\beta(\frac{2}{\nu}-1)}{\ln\nu}$		$\theta \geq \frac{\ln(1-(\frac{2}{\mu}-1)(1-\beta)r}{\ln\nu}$

r	0	$\beta(\frac{2}{\nu}-1)-1$	$+\infty$
θ	$\theta \geq \frac{\ln(1+r)}{\ln\nu}$		$\theta \geq \frac{\ln\beta(\frac{2}{\nu}-1)}{\ln\nu}$

Le tableau suivant donne les conditions sur la borne supérieure de θ. Nous remarquons que la position relative de $\frac{1-\beta(\frac{2}{\nu}-1)}{(1-\beta)(\frac{2}{\nu}-1)}$ et -1 dépend de ν et μ, mais n'a pas d'importance pour la définition de θ.

r	$-\infty$	-1	0
θ		$\theta \leq \frac{\ln(1+r)}{\ln\nu}$	

r	0	$\frac{1}{(\frac{2}{\mu}-1)(1-\beta)}$	$+\infty$
θ	$\theta \leq \frac{\ln(1-(\frac{2}{\mu}-1)(1-\beta)r}{\ln\nu}$		

Afin que de telles fonctions θ existent, le paramètre β doit satisfaire

$$\frac{\ln\beta(\frac{2}{\nu}-1)}{\ln\nu} \leq \theta(r,\nu) \leq \frac{\ln(1-(\frac{2}{\mu}-1)(1-\beta)r)}{\ln\nu}, \tag{2.39}$$

si r appartient à $[0;\frac{1}{(\frac{2}{\nu}-1)(1-\beta)}]$ et

$$\frac{\ln(1+r)}{\ln\nu} \leq \theta(r,\nu) \leq \frac{\ln(1-(\frac{2}{\mu}-1)(1-\beta)r)}{\ln\nu}, \tag{2.40}$$

si r appartient à $[-1;0]$. La seconde condition est toujours vérifiée. La première condition est équivalente à

$$\beta \geq \frac{1}{\frac{2}{\nu}-1}. \tag{2.41}$$

Mais cette condition doit être satisfaite pour tout ν tel que $0 \leq \nu \leq \mu$. Ainsi cette condition devient

$$\beta \geq \frac{1}{\frac{2}{\mu}-1}. \tag{2.42}$$

Comme dans le cas linéaire, nous pouvons prendre $\beta = \frac{1}{\frac{2}{\mu}-1}$. Nous avons ainsi prouvé la proposition 10.

2.3 Conditions sur θ pour un schéma d'ordre deux en espace (cas scalaire non linéaire)

Dans ce paragraphe, $(u,v) \mapsto a(u,v)$ est une fonction telle que $a(u,u) = f'(u)$. Nous avons ainsi $(\partial_u a)(u,u) = (\partial_v a)(u,u) = \frac{1}{2}f''(u)$. Nous désirons déterminer les conditions sur θ pour obtenir un schéma d'ordre le plus élévé possible. Nous considérons alors une solution u suffisamment régulière de l'équation (1.1). Nous effectuons un développement de Taylor au voisinage d'un point (x_j, t^n).

Notations. Nous utilisons les notations habituelles, $u_j^n \simeq u(x_j,t^n)$ et $u_j^{n+1} \simeq u(x_j,t^{n+1})$. Les dérivées de θ par rapport à ν et à r sont notées respectivement θ_ν et θ_r. Pour simplifier l'écriture, nous omettons l'exposant n lorsque nous considérons l'instant t^n ainsi que l'indice j lorsque nous considérons les valeurs prises en x_j. Par exemple, nous notons $f(u_j^n)$ par f et $f'(u_j^n)$ par f_u.
Nous posons

$$R_j^n = u_j^{n+1} - u_j^n + \lambda(\phi_{j+\frac{1}{2}} - \phi_{j-\frac{1}{2}}). \tag{2.43}$$

Nous rappelons que le flux numérique de notre schéma est donné par :

$$\phi_{j+\frac{1}{2}} = \frac{f(u_{j+1}^n) + f(u_j^n)}{2} - \frac{1}{2}\lambda^{\theta_{j+\frac{1}{2}}^n} |a_{j+\frac{1}{2}}|^{1+\theta_{j+\frac{1}{2}}^n} (u_{j+1}^n - u_j^n).$$

Pour fixer les idées, nous supposons que $a_{j+\frac{1}{2}}$ et $a_{j-\frac{1}{2}}$ sont des valeurs positives. Nous verrons que le raisonnement est le même si l'une ou les deux quantités sont négatives. Ainsi,

$$\begin{aligned}
\theta_{j+\frac{1}{2}} &= \theta(r_j^+, \nu_{j+\frac{1}{2}}),\\
\theta_{j-\frac{1}{2}} &= \theta(r_{j-1}^+, \nu_{j-\frac{1}{2}}),
\end{aligned}$$

et

$$r_j^+ = \frac{a_{j-\frac{1}{2}}^+ \Delta u_{j-\frac{1}{2}}^n}{a_{j+\frac{1}{2}}^+ \Delta u_{j+\frac{1}{2}}^n}.$$

Nous avons la proposition suivante.

Proposition 11 *Soit θ une fonction sur $\mathbb{R} \times [0;1[$ vérifiant pour tout réel ν, les conditions*

$$\begin{aligned}
&\theta(1,\nu) = 1,\\
&\theta_r(1,\nu) = 0,\\
&\theta_\nu(1,\nu) = 0.
\end{aligned}$$

Alors le schéma hybride (2.15)-(2.16) *avec les notations* (2.20)-(2.21) *et la définition de $\theta_{j+\frac{1}{2}}$ donnée par* (2.22) *est du second ordre en espace et du second ordre en temps.*

Preuve : L'erreur de troncature est

$$\begin{aligned}
R_j^n = u_j^{n+1} - u_j^n + \lambda\Bigg(&\frac{f_{j+1}^n - f_{j-1}^n}{2}\\
&- \frac{1}{2}(\lambda^{\theta_{j+\frac{1}{2}}} |a_{j+\frac{1}{2}}^n|^{1+\theta_{j+\frac{1}{2}}} \Delta u_{j+\frac{1}{2}}^n - \lambda^{\theta_{j-\frac{1}{2}}} |a_{j-\frac{1}{2}}^n|^{1+\theta_{j-\frac{1}{2}}} \Delta u_{j-\frac{1}{2}}^n)\Bigg).
\end{aligned} \tag{2.44}$$

Nous avons

$$\begin{aligned}
u_j^{n+1} - u_j^n &= u_t\, dt + \frac{u_{tt}}{2}dt^2 + o(dt^2),\\
\Delta u_{j+\frac{1}{2}}^n &= u_{j+1}^n - u_j^n = u_x dx + \frac{u_{xx}}{2}dx^2 + o(dx^2),\\
\Delta u_{j-\frac{1}{2}}^n &= u_j^n - u_{j-1}^n = u_x dx - \frac{u_{xx}}{2}dx^2 + o(dx^2),
\end{aligned}$$

et

$$\begin{aligned}
a_{j+\frac{1}{2}}^+ &= a_{j+\frac{1}{2}} = a + a_u u_x dx + o(dx),\\
a_{j-\frac{1}{2}}^+ &= a_{j-\frac{1}{2}} = a - a_u u_x dx + o(dx).
\end{aligned}$$

Donc après calculs, nous obtenons :

$$r_j^+ = 1 - \left(\frac{u_{xx}}{u_x} + 2\frac{a_u}{a}u_x\right)dx + o(dx),$$

d'où

$$\theta_{j+\frac{1}{2}} = \theta(1,\lambda a) + (-\theta_r(1,\lambda a)(\frac{u_{xx}}{u_x} + 2\frac{a_u}{a}u_x) + \theta_\nu(1,\lambda a)\lambda a_u u_x)dx + o(dx) \tag{2.45}$$

et

$$\begin{aligned}
\lambda^{\theta_{j+\frac{1}{2}}}|a_{j+\frac{1}{2}}|^{1+\theta_{j+\frac{1}{2}}} = \lambda^{\theta(1,\lambda a)}a^{1+\theta(1,\lambda a)}(1 - \ln(\lambda a)(\theta_r(1,\lambda a)(\frac{u_{xx}}{u_x} + 2\frac{a_u}{a}u_x)\\
-\theta_\nu(1,\lambda a)\lambda a_u u_x + (1+\theta(1,\lambda a))\frac{a_u}{a}u_x)dx) + o(dx).
\end{aligned}$$

Nous pouvons calculer de même $\theta_{j-\frac{1}{2}}$:

$$r_{j-1}^+ = 1 - (\frac{1}{2}\frac{u_{xx}}{u_x} + 2\frac{a_u}{a}u_x)dx + o(dx),$$

d'où

$$\theta_{j-\frac{1}{2}} = \theta(1,\lambda a) + (-\theta_r(1,\lambda a)(\frac{1}{2}\frac{u_{xx}}{u_x} + 2\frac{a_u}{a}u_x) - \theta_\nu(1,\lambda a)\lambda a_u u_x)dx + o(dx) \tag{2.46}$$

et

$$\begin{aligned}
\lambda^{\theta_{j-\frac{1}{2}}}|a_{j-\frac{1}{2}}|^{1+\theta_{j-\frac{1}{2}}} = \lambda^{\theta(1,\lambda a)}a^{1+\theta(1,\lambda a)}(1 - \ln(\lambda a)(\theta_r(1,\lambda a)(\frac{1}{2}\frac{u_{xx}}{u_x} + 2\frac{a_u}{a}u_x)\\
+\theta_\nu(1,\lambda a)\lambda a_u u_x) - (1+\theta(1,\lambda a))\frac{a_u}{a}u_x)dx) + o(dx).
\end{aligned}$$

Nous avons

$$\begin{aligned}
f_{j+1}^n &= f + a u_x dx + \frac{1}{2}(a u_{xx} + a_u u_x^2)dx^2 + o(dx^2)\\
f_{j-1}^n &= f - a u_x dx + \frac{1}{2}(a u_{xx} + a_u u_x^2)dx^2 + o(dx^2)
\end{aligned}$$

d'où

$$\frac{f_{j+1}^n - f_{j-1}^n}{2} = a u_x dx + o(dx^2).$$

En regroupant tous les résultats, nous obtenons

$$\begin{aligned} R = u_t dt + \frac{u_{tt}}{2} dt^2 + a u_x dt - \frac{1}{2} dt^{\theta(1,\lambda a)+1} dx^{-\theta(1,\lambda a)} a^{1+\theta(1,\lambda a)} (\ln(\lambda a)(\theta_r(1,\lambda a)\frac{-u_{xx}}{2} + 2\theta_\nu(1,\lambda a)\lambda a_u u_x) \\ -\frac{1}{2} dt^{\theta(1,\lambda a)+1} dx^{1-\theta(1,\lambda a)} a^{1+\theta(1,\lambda a)} (2\frac{a_u}{a} u_x^2 (1+\theta(1,\lambda a)) + u_{xx}) \\ +o(dx^2) + o(dt^2). \end{aligned} \tag{2.47}$$

Or si u est une solution régulière de

$$\partial_t u + \partial_x f(u) = 0 \tag{2.48}$$

alors on a

$$\begin{cases} u_{xt} + f''(u) u_x^2 + f'(u) u_{xx} = 0 \\ u_{tt} + f''(u) u_t u_x + f'(u) u_{xt} = 0. \end{cases} \tag{2.49}$$

En multipliant la première équation par $-f'(u)$ et en additionnant les deux équations membre à membre, nous obtenons, grâce à la relation $u_t = -f'(u)u_x$:

$$u_{tt} - (f'(u))^2 u_{xx} - 2f''(u) f'(u) u_x^2 = 0. \tag{2.50}$$

Nous avons par hypothèse $a = f'(u)$ et $a_u = \frac{1}{2} f''(u)$ d'où

$$u_{tt} - a^2 u_{xx} - 4 a_u a u_x^2 = 0. \tag{2.51}$$

D'où, en imposant pour tout réel ν,

$$\theta(1,\nu) = 1, \tag{2.52}$$

$$\theta_r(1,\nu) = 0, \tag{2.53}$$

$$\theta_\nu(1,\nu) = 0, \tag{2.54}$$

nous obtenons

$$R = o(dx^2) + o(dt^2). \tag{2.55}$$

Enfin, le fait que $a_{j+\frac{1}{2}}$ et $a_{j-\frac{1}{2}}$ soient positifs intervient uniquement dans le développement limité à l'ordre 1 de θ. Si nous imposons $\theta_r(1,\lambda a) = 0$ et $\theta_\nu(1,\lambda a) = 0$, les signes de $a_{j+\frac{1}{2}}$ et $a_{j-\frac{1}{2}}$ n'interviennent pas dans le raisonnement. Ainsi, le schéma est du second ordre en espace et du second ordre en temps.
Nous avons prouvé la proposition 11.

Remarque 3 *Nous verrons dans le paragraphe suivant que le fait que les conditions* (2.52)-(2.54) *soient requises pour tout ν n'apportent pas réellement de difficultés pour trouver de "bonnes fonctions" θ.*

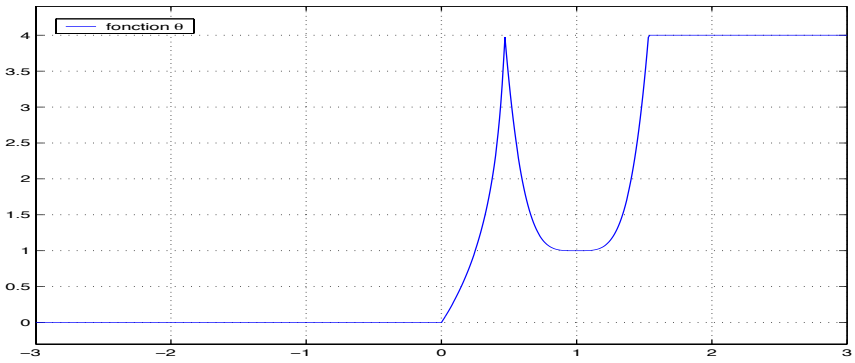

FIG. 2.2 – Fonction θ représentée dans le cas $\mu = 0.5$, $p = 4$, $n = 4$.

2.4 Exemples de fonctions θ

Nous donnons une famille à deux paramètres de fonctions θ qui permettent d'obtenir un schéma hybride TVD, du second ordre en espace et du second ordre en temps. Remarquons que, à ν fixé, la courbe représentative de la fonction θ doit passer par le point de coordonnées $(1, 1)$ pour obtenir le second ordre en espace. Ce point doit appartenir à la zone TVD. Si nous considérons la zone TVD donnée par les conditions de la proposition 10, la famille de fonctions données ci-dessous convient pour un nombre CFL μ inférieur à $\frac{2}{3}$:

$$\theta_{p,n}(r,\nu) = \begin{cases} 0 \text{ si } r \leq 0, \\ \frac{\ln(1-\gamma r)}{\ln(\nu)} \text{ si } 0 < r \leq \frac{(1-\nu^p)}{\gamma}, \\ \frac{p-1}{\left(\frac{1-\nu^p}{\gamma-1}\right)^n}(r-1)^n + 1 \text{ si } \frac{(1-\nu^p)}{\gamma} \leq r < 2 - \frac{1-\nu^p}{\gamma}, \\ p \text{ si } r \geq 2 - \frac{1-\nu^p}{\gamma}. \end{cases} \tag{2.56}$$

où $\gamma = 2(\frac{1}{\mu} - 1)$. Nous remarquons que les conditions (2.52)-(2.54) sont vérifiées. Nous avons représenté sur la figure 2.2 la courbe représentative de $\theta_{4,4}$ dans le cas $p = 4$, $n = 4$, $\mu = \nu = 0.5$.

2.5 Extension heuristique au cas du système

Nous avons ici un approche heuristique pour étendre l'étude du cas scalaire au cas des systèmes. Nous rappelons les notations vectorielles du paragraphe 1.1.3. Nous supposons que la matrice $A(u,v)$ se diagonalise sous la forme $\Lambda(u,v) = \text{diag}(\alpha_k(u,v))$ suivant la matrice des vecteurs propres à droite $R(u,v)$:

$$A(u,v) = R(u,v)D(u,v)R(u,v)^{-1}.$$

La notation pour un exposant "vectoriel" de matrice est définie par :

$$|A(u,v)|^{\vec{\theta}} = R(u,v)\, diag(|\alpha_k(u,v)|^{\theta_k})R(u,v)^{-1}. \tag{2.57}$$

Le schéma est alors

$$u_j^{n+1} = u_j^n - \lambda\Big(\phi(u_j^n, u_{j+1}^n, \lambda, \overrightarrow{\theta}^n_{j+\frac{1}{2}}) - \phi(u_{j-1}^n, u_j^n, \lambda, \overrightarrow{\theta}^n_{j-\frac{1}{2}})\Big), \tag{2.58}$$

avec

$$\phi(u,v,\lambda,\vec{\theta}) = \frac{f(u)+f(v)}{2} - \frac{1}{2}\left(\lambda \mathbf{I}\right)^{\vec{\theta}} |A(u,v)|^{\mathbf{e}+\vec{\theta}}(v-u), \tag{2.59}$$

et $\mathbf{e} = (1,...,1)$; la matrice $\mathbf{I}$ est la matrice identité dans $\mathbb{R}^p$. Le paramètre θ est alors déterminé par

$$\theta_{k,j+\frac{1}{2}} = \begin{cases} \theta(r^+_{k,j}, |\nu_{j+\frac{1}{2}}|) \text{ si } \nu_{k,j+\frac{1}{2}} \geq 0, \\ \theta(r^-_{k,j+1}, |\nu_{j+\frac{1}{2}}|) \text{ si } \nu_{k,j+\frac{1}{2}} < 0, \end{cases} \tag{2.60}$$

où

$$\nu_{k,j+\frac{1}{2}} = \frac{dt}{dx}\alpha_{k,j+\frac{1}{2}}, \tag{2.61}$$

et

$$r^+_{k,j} = \frac{\alpha^+_{k,j-\frac{1}{2}}(R^{-1}(u^n_j - u^n_{j-1}))_k}{\alpha^+_{k,j+\frac{1}{2}}(R^{-1}(u^n_{j+1} - u^n_j))_k},$$
$$r^-_{k,j} = \frac{\alpha^-_{k,j+\frac{1}{2}}(R^{-1}(u^n_{j+1} - u^n_j))_k}{\alpha^-_{k,j-\frac{1}{2}}(R^{-1}(u^n_j - u^n_{j-1}))_k}.$$

Nous décidons de conserver la fonction θ déterminée dans le cas scalaire.

3

Expériences numériques

Nous désirons valider le schéma hybride par différentes expériences numériques. Nous utilisons la fonction $\theta_{4,4}$ représentée dans la figure 2.2. Cette fonction est définie par

$$\theta_{4,4}(r,\nu) = \begin{cases} 0 \text{ si } r \leq 0, \\ \frac{\ln(1-\gamma r)}{\ln(\nu)} \text{ si } 0 < r \leq \frac{(1-\nu^4)}{\gamma}, \\ \frac{3}{\left(\frac{1-\nu^4}{\gamma-1}\right)^4}(r-1)^4 + 1 \text{ si } \frac{(1-\nu^4)}{\gamma} \leq r < 2 - \frac{1-\nu^4}{\gamma}, \\ 4 \text{ si } r \geq 2 - \frac{1-\nu^4}{\gamma}. \end{cases} \tag{3.1}$$

Nous rappelons que $\gamma = 2(\frac{1}{\mu} - 1)$ et μ est le nombre CFL.

3.1 Cas scalaire

3.1.1 Equation de transport linéaire pour la fonction sinus

Nous considérons ici l'équation de transport linéaire avec une condition initiale régulière :

$$\partial_t u + \partial_x u = 0,\ x \in]0;1[,\ t > 0, \tag{3.2}$$

avec comme condition initiale

$$u(x,0) = \sin(\pi x + \frac{\pi}{4}),\ x \in]0;1[,$$

et comme condition aux bords

$$u(0,t) = \sin(-\pi t + \frac{\pi}{4}),\ t > 0.$$

La solution exacte de ce problème est :

$$u(x,t) = \sin(\pi(x-t) + \frac{\pi}{4}).$$

Nous désirons comparer la précision et l'ordre en espace du schéma hybride à deux autres schémas, le schéma de Roe qui est TVD mais diffusif et le schéma de Lax-Wendroff qui n'est pas TVD

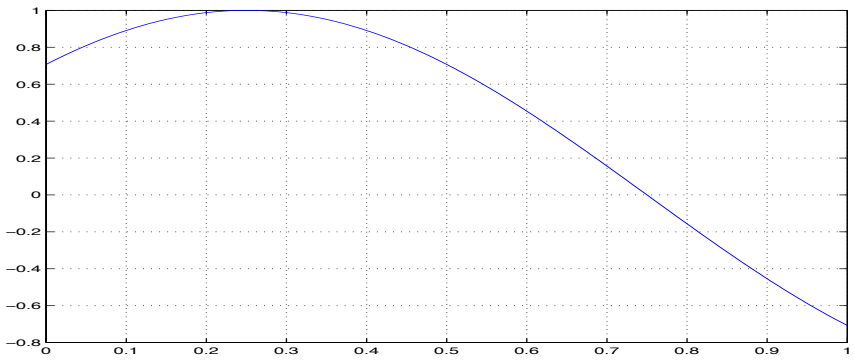

FIG. 3.1 – Condition initiale pour l'équation de transport linéaire avec donnée initiale régulière.

mais est du second ordre.
Le maillage utilisé pour la simulation est un maillage uniforme en espace. Le nombre de Courant choisi est $\mu = 0.5$. Nous considérons des maillages de tailles différentes pour observer l'ordre des trois schémas. Le nombre de points du maillage varie de 200 à 1600 points. La donnée initiale est représentée dans la figure 3.1. Les figures 3.2, 3.3 et 3.4 représentent les courbes de la solution calculée au temps final par les schémas de Roe, Lax-Wendroff et hybride respectivement. Enfin, la figure 3.5 représente sur un même graphique les trois courbes d'erreur relative (en norme L^1) en fonction du pas d'espace en échelle logarithmique. La pente de la courbe représentative de $\log_{10} err$ en fonction de $\log_{10} h$ est 1.00 pour le schéma de Roe. Elle est de 1.99 pour le schéma de Lax-Wendroff. Enfin, la pente est de 2.01 pour le schéma hybride. Nous vérifions ainsi que le schéma hybride est d'ordre 2 dans le cas linéaire.

3.1.2 Equation de transport linéaire pour une fonction continue en chapeau

Nous considérons ici l'équation de transport linéaire avec une condition initiale continue :

$$u_0(x) = \begin{cases} 0 \text{ si } x \in]0;\frac{1}{4}[\\ x - \frac{1}{4} \text{ si } x \in]\frac{1}{4};\frac{1}{2}[, \\ -x + \frac{3}{4} \text{ si } x \in]\frac{1}{2};\frac{3}{4}[, \\ 0 \text{ si } x \in]\frac{3}{4};1[. \end{cases}$$

La solution exacte est

$$u(x,t) = u_0(x-t).$$

La donnée initiale est représentée sur la figure 3.6. Le schéma de Roe reste là encore diffusif (voir figure 3.7) : la pointe n'est pas bien capturée par le schéma et les raréfactions sont lissées. La pente de la courbe d'erreur en norme L^1 en échelle logarithmique est 0.88. Le schéma de Lax-Wendroff présente des oscillations (faibles) au pied de la raréfaction. La pointe est mieux capturée (figure 3.8). La pente de la courbe d'erreur est 1.16. Le schéma hybride permet de calculer une solution qui se superpose presque à la solution exacte. La pente de la courbe d'erreur est 1.38 environ. L'erreur pour le schéma hybride est inférieure à l'erreur commise par le schéma de Lax-Wendroff (figure 3.10 et figure 3.11).

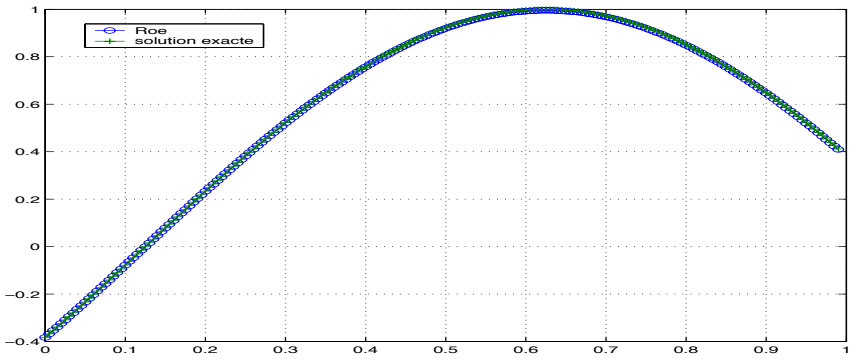

FIG. 3.2 – Schéma de Roe pour l'équation de transport linéaire avec donnée initiale régulière ; $\mu = 0.5$; $T = 0.375$; 200 points.

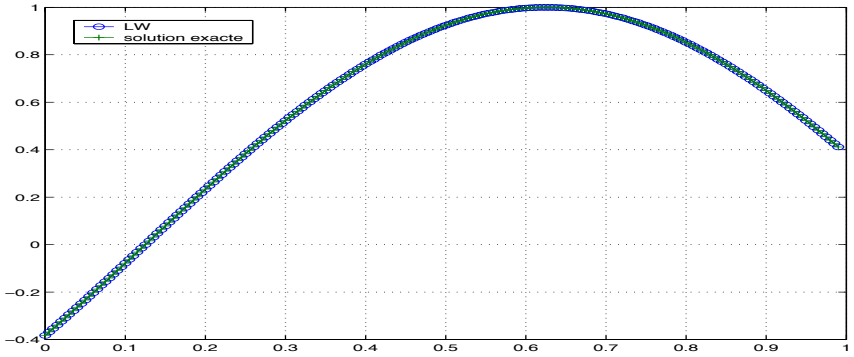

FIG. 3.3 – Schéma de Lax-Wendroff pour l'équation de transport linéaire avec donnée initiale régulière ; $\mu = 0.5$; $T = 0.375$; 200 points.

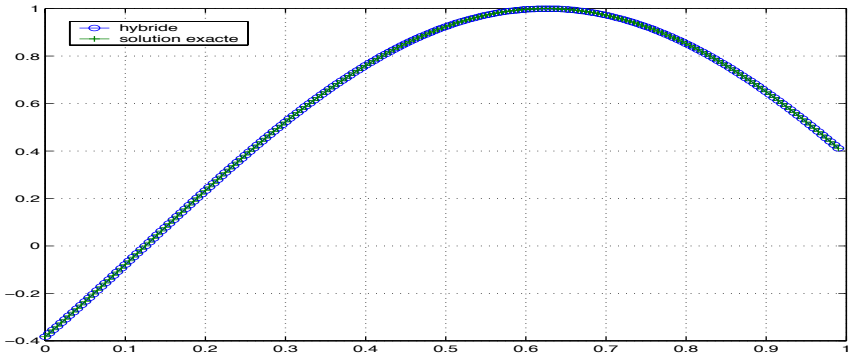

FIG. 3.4 – Schéma hybride pour l'équation de transport linéaire avec donnée initiale régulière ; $\mu = 0.5$; $T = 0.375$; 200 points.

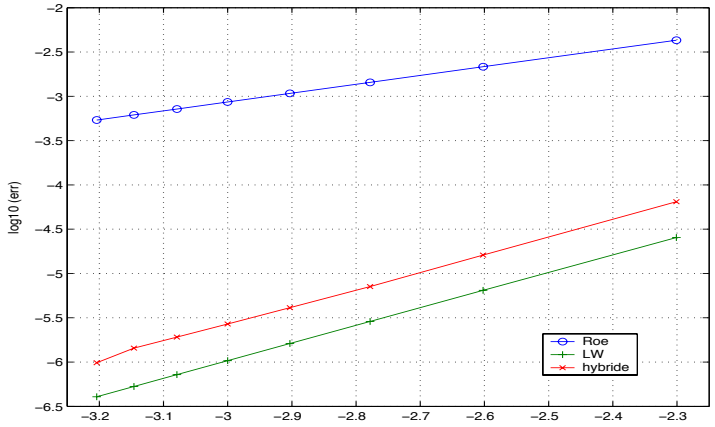

FIG. 3.5 – Comparaison de l'erreur pour les schémas de Roe, de Lax-Wendroff et hybride pour l'équation de transport linéaire avec donnée initiale régulière ; $\mu = 0.5$; $T = 0.375$: courbe $\log_{10} err$ en fonction de $\log_{10} h$. Pentes : 1.00 pour le schéma de Roe, 1.99 pour le schéma de Lax Wendroff, 2.01 pour le schéma hybride.

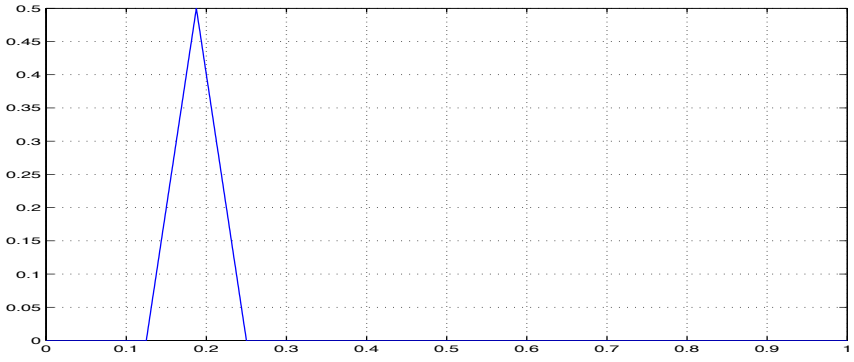

FIG. 3.6 – Condition initiale pour l'équation de transport linéaire avec donnée continue.

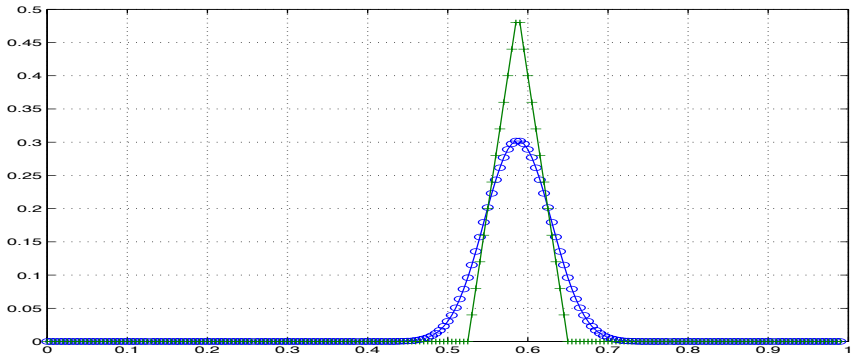

FIG. 3.7 – Schéma de Roe pour l'équation de transport linéaire avec une donnée initiale continue : $\nu = 0.5$; $T = 0.4$; 200 points.

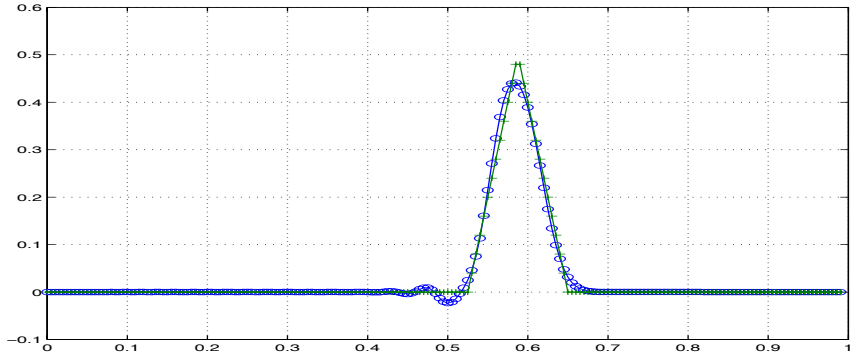

FIG. 3.8 – Schéma de Lax Wendroff pour l'équation de transport linéaire avec une donnée initiale continue : $\nu = 0.5$; $T = 0.4$; 200 points.

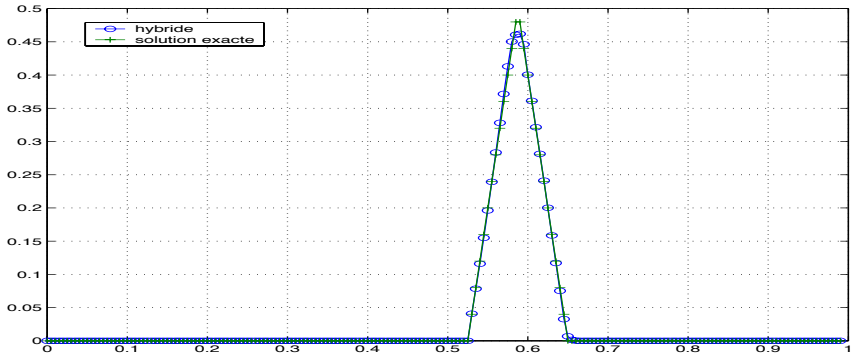

FIG. 3.9 – Schéma hybride pour l'équation de transport linéaire avec une donnée initiale continue : $\nu = 0.5$; $T = 0.4$; 200 points.

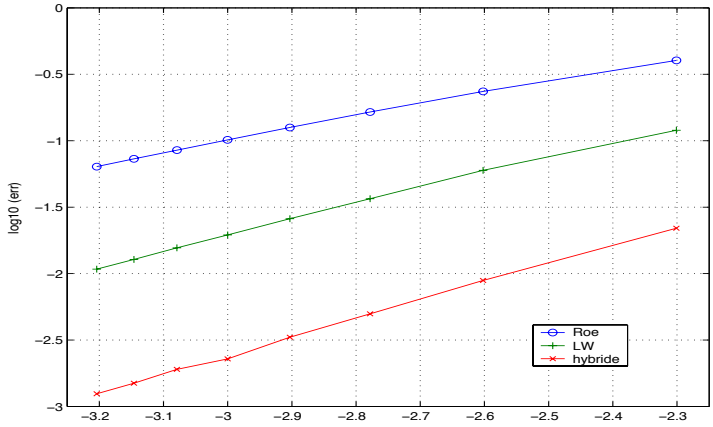

FIG. 3.10 – Comparaison de l'erreur pour les schémas de Roe, de Lax-Wendroff et hybride pour l'équation de transport linéaire avec donnée initiale continue ; $\mu = 0.5$; $T = 0.4$: courbe $\log_{10} err$ en fonction de $\log_{10} h$(norme L^1). Pentes : 0.88 pour le schéma de Roe, 1.16 pour le schéma de Lax Wendroff, 1.38 pour le schéma hybride.

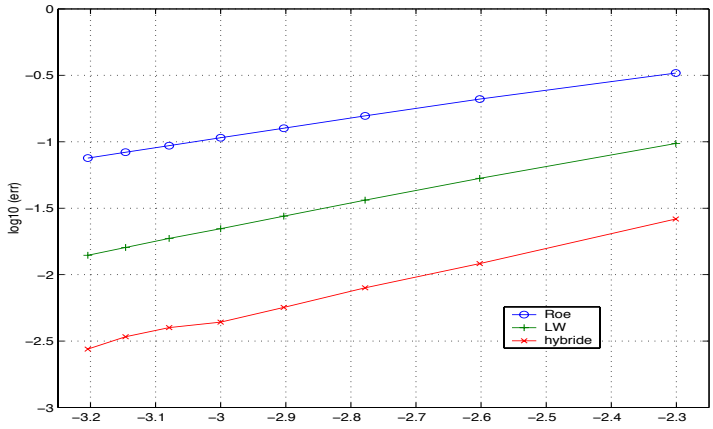

FIG. 3.11 – Comparaison de l'erreur pour les schémas de Roe, de Lax-Wendroff et hybride pour l'équation de transport linéaire avec donnée initiale continue ; $\mu = 0.5$; $T = 0.4$: courbe $\log_{10} err$ en fonction de $\log_{10} h$ (norme L^2). Pentes : 0.71 pour le schéma de Roe, 0.93 pour le schéma de Lax Wendroff, 1.08 le schéma hybride.

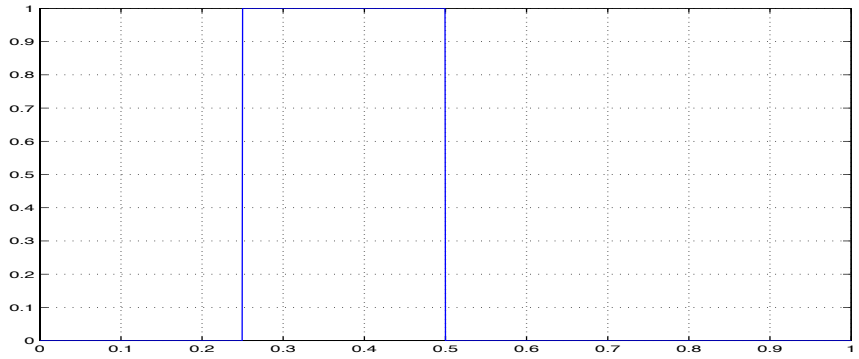

FIG. 3.12 – Condition initiale pour l'équation de transport linéaire avec donnée initiale constante par morceaux.

3.1.3 Equation de transport linéaire pour une fonction constante par morceaux

Nous considérons à nouveau l'équation (3.2). La donnée initiale est maintenant discontinue :

$$u(x,0) = \begin{cases} 0 \text{ si } x \in]0;\frac{1}{4}[, \\ 1 \text{ si } x \in]\frac{1}{4};\frac{1}{2}[, \\ 0 \text{ si } x \in]\frac{1}{2};1[. \end{cases}$$
$$u(0,t) = 0,\ t > 0$$

La solution exacte est alors

$$u(x,t) = u(x-t,0).$$

La donnée intiale est représentée sur la figure 3.12. Nous comparons la précision des schémas de Roe, de Lax-Wendroff et hybride pour une donnée initiale en forme de créneau. La fonction θ est la fonction déjà utilisée précédemment. Les résultats sont présentés dans les figures 3.13, 3.14 et 3.15. Le choc est bien capturé avec le schéma hybride, tandis que nous observons qu'il ne l'est pas du tout avec le schéma de Roe. De plus, le schéma hybride est bien TVD ce qui n'est pas le cas du schéma de Lax-Wendroff, qui présente de fortes oscillations. La pente d'erreur en norme L^1 du schéma de Roe est 0.50, celle du schéma de Lax-Wendroff est 0.60 et celle du schéma hybride est 0.96 (figure 3.16). En norme L^2, figure 3.17, nous retrouvons que la pente d'erreur du schéma de Roe est 0.25 (convergence en $h^{\frac{1}{4}}$), celle du schéma de Lax-Wendroff est 0.32 et celle du schéma hybride est 0.43 (pente de la droite des moindres carrés).

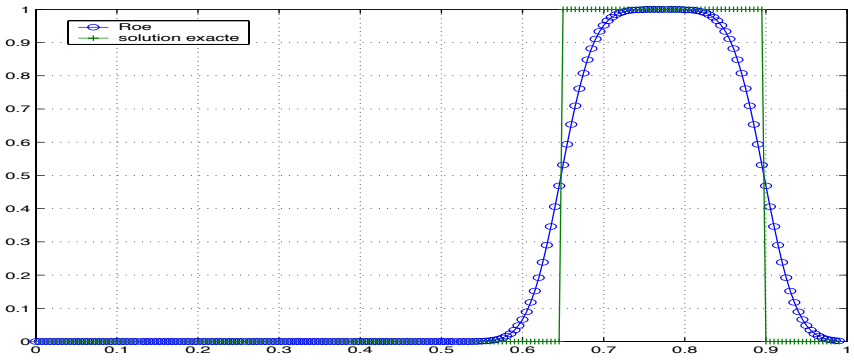

FIG. 3.13 – Schéma de Roe pour l'équation de transport linéaire avec donnée initiale discontinue, constante par morceaux ; $\mu = 0.5$; $T = 0.4$; 200 points.

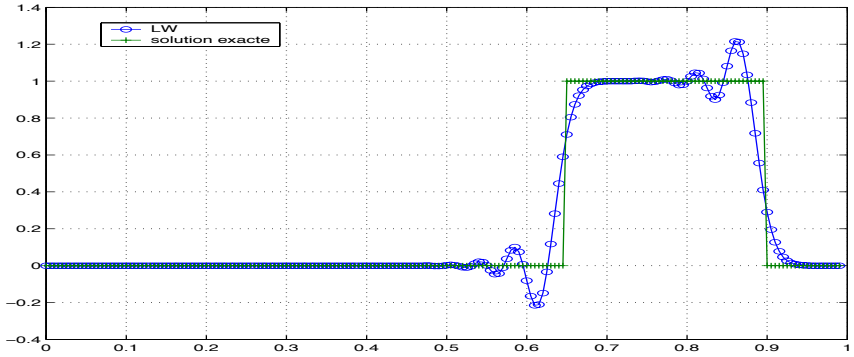

FIG. 3.14 – Schéma de Lax-Wendroff pour l'équation de transport linéaire avec donnée initiale discontinue, constante par morceaux ; $\mu = 0.5$; $T = 0.4$; 200 points.

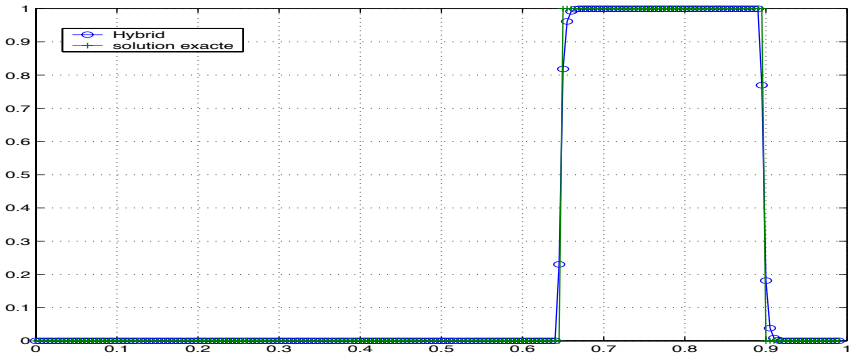

FIG. 3.15 – Schéma hybride pour l'équation de transport linéaire avec donnée initiale discontinue, constante par morceaux ; $\mu = 0.5$; $T = 0.4$; 200 points.

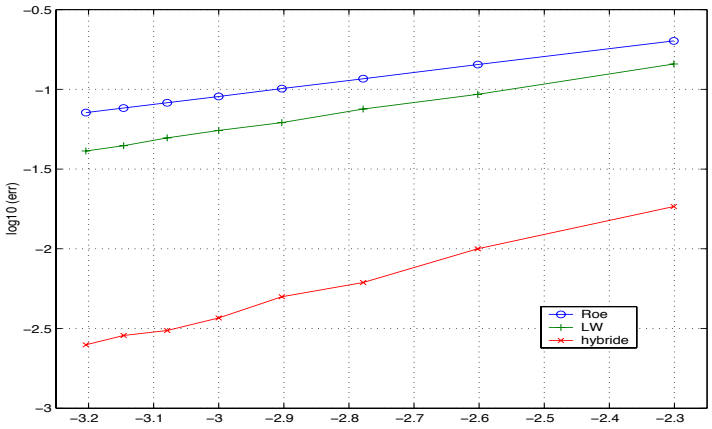

FIG. 3.16 – Comparaison de l'erreur pour les schémas de Roe, de Lax-Wendroff et hybride pour l'équation de transport linéaire avec donnée initiale discontinue, constante par morceaux (normes L^1) ; $\mu = 0.5$; $T = 0.4$: courbe $\log_{10} err$ en fonction de $\log_{10} h$. Pentes : 0.50 pour le schéma de Roe, 0.60 pour le schéma de Lax Wendroff, 0.96 pour le schéma hybride.

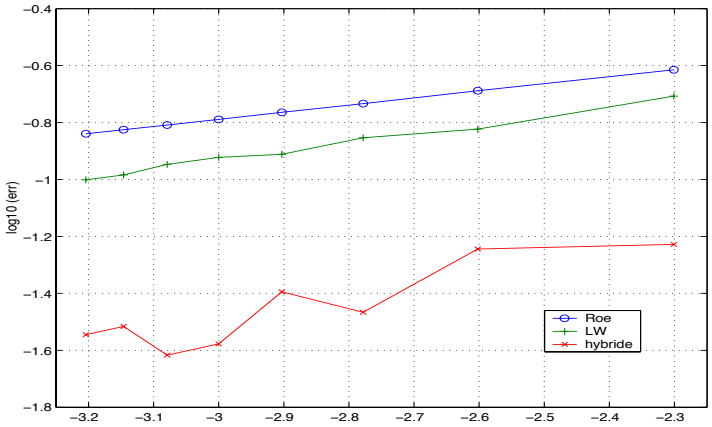

FIG. 3.17 – Comparaison de l'erreur pour les schémas de Roe, de Lax-Wendroff et hybride pour l'équation de transport linéaire avec donnée initiale discontinue, constante par morceaux (normes L^2) ; $\mu = 0.5$; $T = 0.4$: courbe $\log_{10} err$ en fonction de $\log_{10} h$. Pentes : 0.25 pour le schéma de Roe, 0.32 pour le schéma de Lax Wendroff, 0.43 pour le schéma hybride (on peut noter les fluctuations d'erreurs dans le cas hybride dues au faible nombre de points dans la discontinuité et à ses fluctuations).

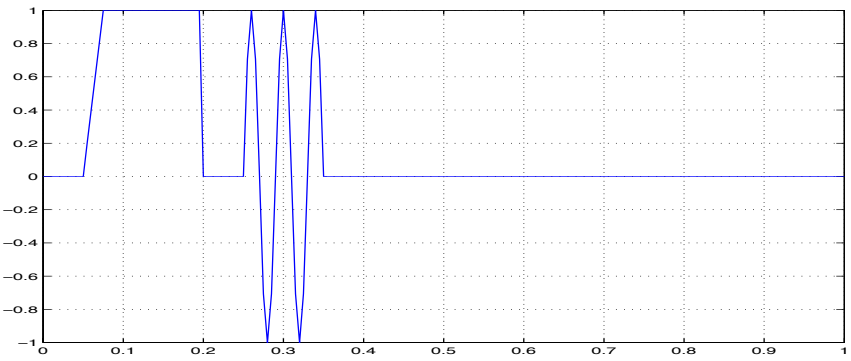

FIG. 3.18 – Donnée initiale pour l'équation de transport linéaire avec raréfaction, choc et perturbations, $\mu = 0.5$.

3.1.4 Equation de transport linéaire avec raréfaction, choc et perturbations

Nous étudions maintenant l'équation (3.2) avec comme donnée initiale :

$$u(x,0) = \begin{cases} 0 \text{ si } x \in]0;0.05[, \\ 2(x-0.05)/(0.15-0.1) \text{ si } x \in]0.05;0.075[, \\ 1 \text{ si } x \in]0.075;0.2[, \\ 0 \text{ si } x \in]0.2;0.25[, \\ \sin(5\pi(x-0.25)/(0.35-0.25)) \text{ si } x \in]0.25;0.35[, \\ 0 \text{ si } x \in]0.35;1[. \end{cases}$$

et comme condition aux bords

$$u(0,t) = 0,\ t > 0.$$

Ainsi la donnée initiale présente une onde de raréfaction suivie d'un choc puis de perturbations (voir figure 3.18). La solution exacte est

$$u(x,t) = u(x-t,0).$$

Le schéma de Roe est très diffusif. Le choc n'apparaît plus et le profil des perturbations a été grandement modifié (figure 3.19). Le schéma de Lax-Wendroff conserve le profil des perturbations, mais des oscillations apparaissent (figure 3.20). Nous observons par contre que la solution calculée par le schéma hybride se superpose parfaitement à la solution exacte. Ce schéma conserve le choc et le profil des perturbations (figure 3.21). Les tests présentés sont effectués avec 1600 points mais le schéma hybride donne déja de très bons résultats avec un maillage de 400 points (figure 3.25), ce qui n'est pas du tout le cas des autres schémas (figure 3.23, figure 3.24).

3.1.5 Equation de Burgers avec une donnée initiale régulière

Nous considérons dans ce paragraphe l'équation de Burgers :

$$\partial_t u + \partial_x(\tfrac{u^2}{2}) = 0,\ x \in]0;1[,\ t > 0,$$
$$u(x,0) = u_0(x).$$

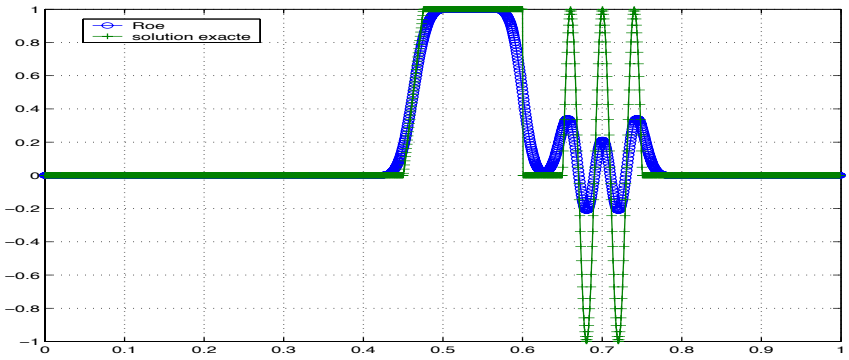

FIG. 3.19 – Schéma de Roe pour l'équation de transport linéaire avec raréfaction, choc et perturbations ; $\mu = 0.5$; $T = 0.4$; 1600 points.

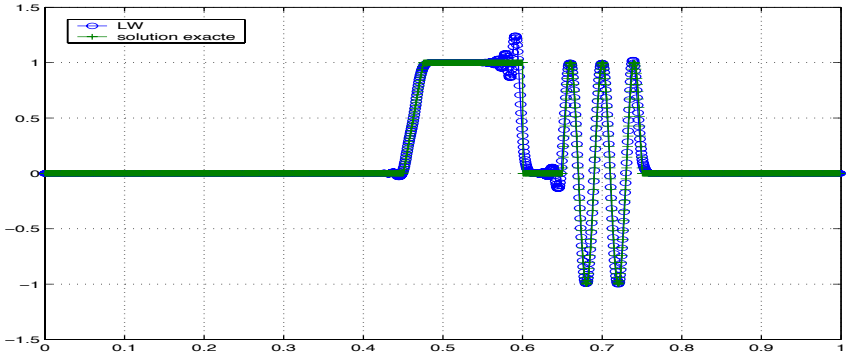

FIG. 3.20 – Schéma de Lax-Wendroff pour l'équation de transport linéaire avec raréfaction, choc et perturbations ; $\mu = 0.5$; $T = 0.4$; 1600 points.

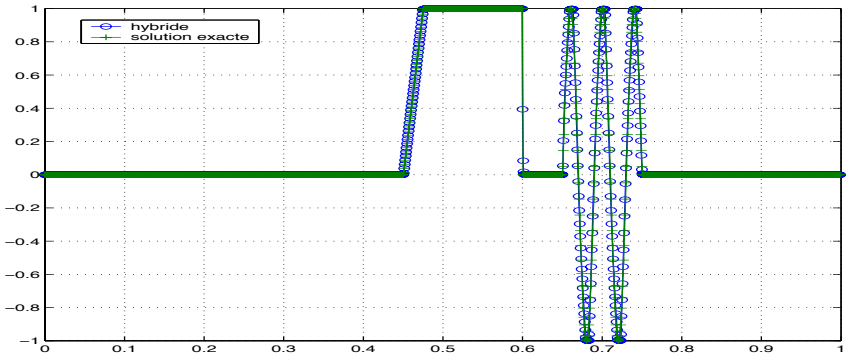

FIG. 3.21 – Schéma hybride pour l'équation de transport linéaire avec raréfaction, choc et perturbations ; $\mu = 0.5$; $T = 0.4$; 1600 points.

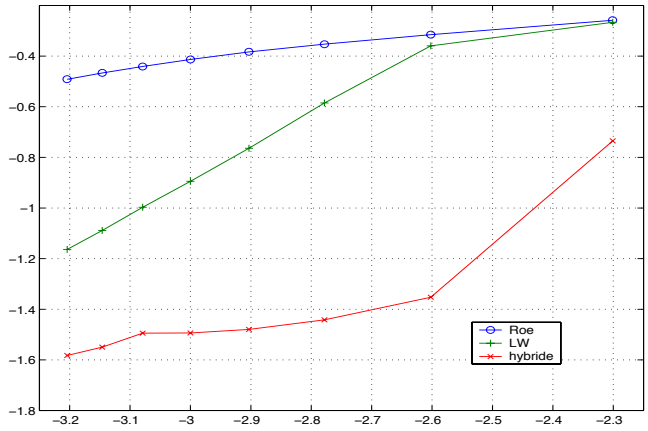

FIG. 3.22 – Comparaison des erreurs pour les schémas de Roe, Lax-Wendroff et hybride pour l'équation de transport linéaire avec raréfaction, choc et perturbations : $\log_{10} err$ en fonction de $\log_{10} h$ (norme L^1) ; $\mu = 0.5$; $T = 0.4$. Pentes : 0.29 pour le schéma de Roe, 1.33 pour le schéma de Lax Wendroff, 0.34 pour le schéma hybride (nous n'avons pas considéré l'erreur obtenue pour 200 points).

Dans notre premier test, la donnée initiale est une fonction régulière :

$$u(x,0) = \sin(\pi x + \frac{\pi}{4}),\ x \in]0;1[.$$

et la condition aux bords est

$$u(0,t) = \sin(-\pi t + \frac{\pi}{4}),\ t > 0$$

Le maillage est uniforme avec 200 points et le nombre de Courant est 0.5. Les schémas de Roe et hybride (figure 3.26 et 3.28 respectivement) donnent les mêmes résultats sauf pour le début de la raréfaction qui est plus nette pour le schéma hybride. Le schéma de Lax-Wendroff présente une nouvelle fois des oscillations (figure 3.27).

3.1.6 Equation de Burgers avec une donnée initiale en créneau.

Nous nous intéressons une nouvelle fois à l'équation de Burgers. Dans ce second test, la donnée initiale est maintenant :

$$u(x,0) = \begin{cases} 0 \text{ si } x \in]0;\frac{1}{4}[, \\ 1 \text{ si } x \in]\frac{1}{4};\frac{1}{2}[, \\ 0 \text{ si } x \in]\frac{1}{2};1[. \end{cases}$$

Le maillage est uniforme avec 200 points et le nombre de Courant est 0.5. Le schéma de Lax-Wendroff ne convient pas du tout (figure 3.30). Les schémas de Roe et hybride sont représentés figures 3.29 et 3.31 respectivement. Le schéma hybride capture mieux le choc.

3.1.7 Comparaison de différentes fonctions $\theta_{n,p}$ dans le cas linéaire

Nous pouvons choisir des fonctions pour le paramètre $\theta_{n,p}$ différentes de celle utilisée dans les tests numériques. Nous testons dans cette partie plusieurs fonctions θ pour voir comment le

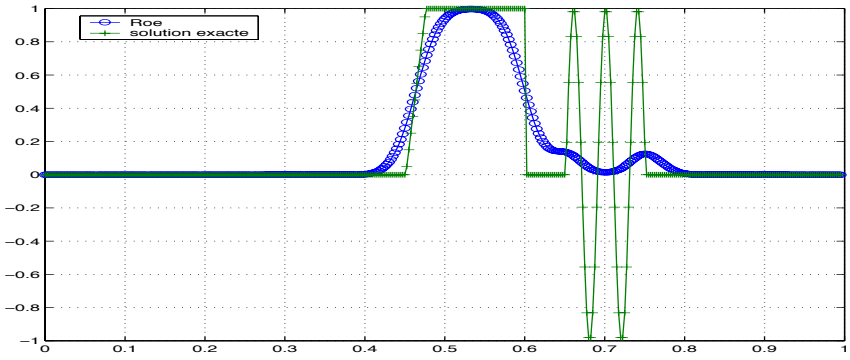

FIG. 3.23 – Schéma de Roe pour l'équation de transport linéaire avec raréfaction, choc et perturbations : $\log_{10} err$ en fonction de $\log_{10} h$; $\mu = 0.5$; $T = 0.4$; ; 400 points.

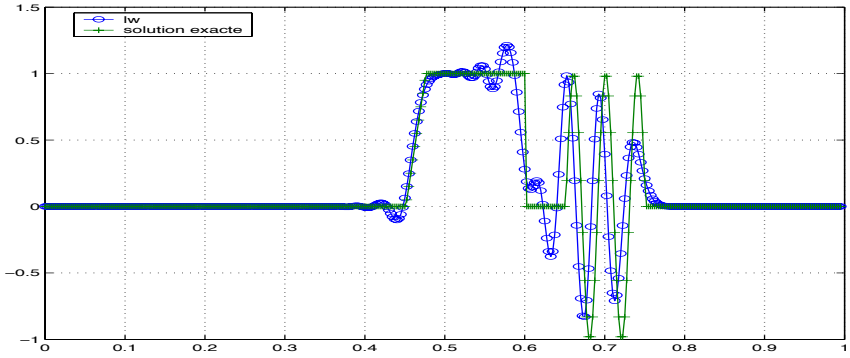

FIG. 3.24 – Schéma de Lax-Wendroff pour l'équation de transport linéaire avec raréfaction, choc et perturbations ; $\mu = 0.5$; $T = 0.4$; ; 400 points.

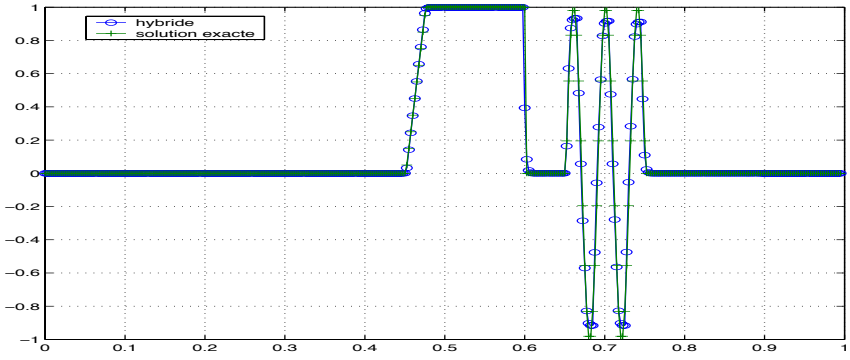

FIG. 3.25 – Schéma hybride pour l'équation de transport linéaire avec raréfaction, choc et perturbations ; $\mu = 0.5$; $T = 0.4$; 400 points.

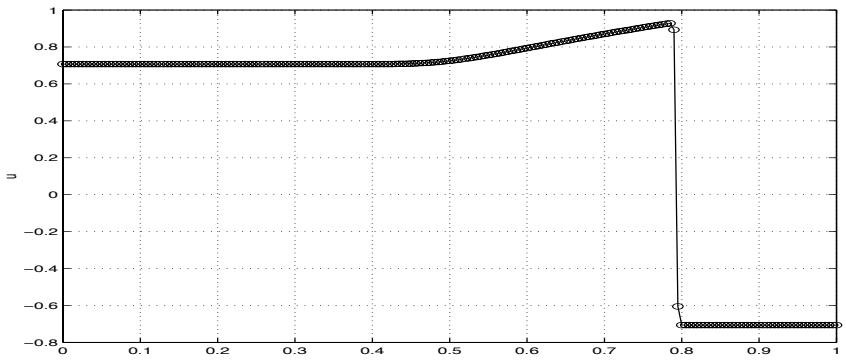

FIG. 3.26 – Schéma de Roe pour l'équation de Burgers avec une donnée initiale régulière ; $\mu = 0.5$; $T = 0.7$.

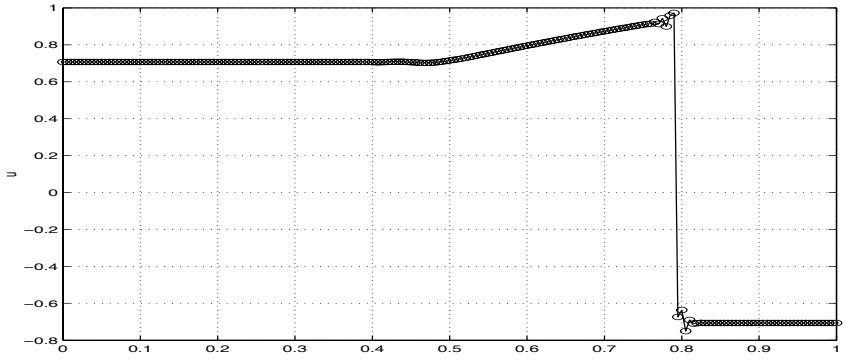

FIG. 3.27 – Schéma de Lax Wendroff pour l'équation de Burgers avec une donnée initiale régulière ; $\mu = 0.5$; $T = 0.7$.

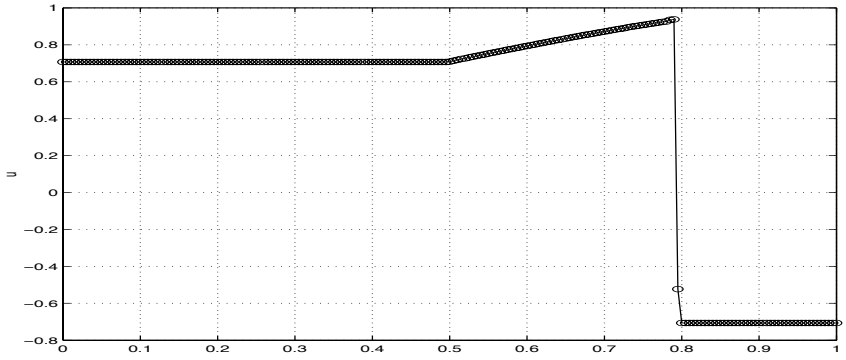

FIG. 3.28 – Schéma hybride pour l'équation de Burgers avec une donnée initiale régulière ; $\mu = 0.5$; $T = 0.7$.

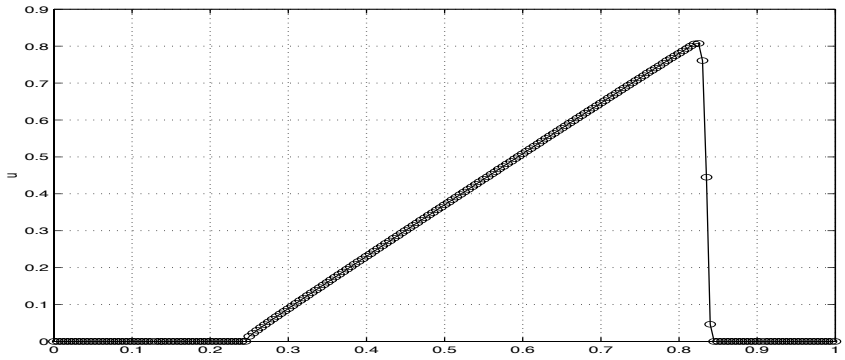

FIG. 3.29 – Schéma de Roe pour l'équation de Burger avec une donnée initiale en créneau ; $\mu = 0.5$; $T = 0.7$.

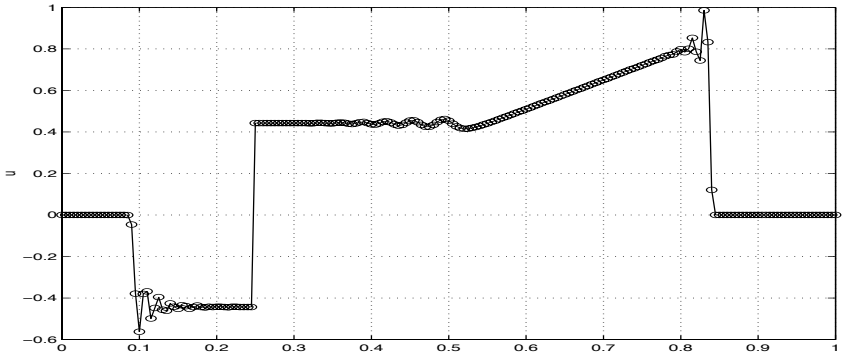

FIG. 3.30 – Schéma de Lax-Wendroff pour l'équation de Burger avec une donnée initiale en créneau ; $\mu = 0.5$; $T = 0.7$.

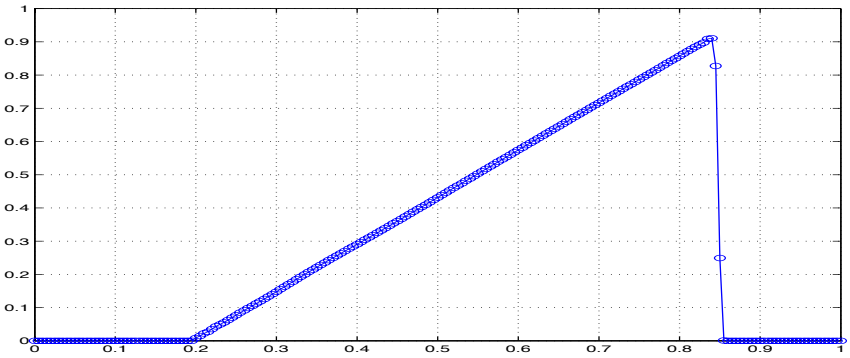

FIG. 3.31 – Schéma hybride pour l'équation de Burger avec une donnée initiale en créneau ; $\mu = 0.5$; $T = 0.7$.

schéma se comporte dans le cas de l'équation de transport linéaire avec une donnée initiale en créneau.

D'après notre étude, la courbe doit se situer dans la zone TVD pour éliminer les oscillations. Dans un premier test, nous prenons une fonction dont la courbe se situe juste au-dessus des courbes limitant la zone TVD (nous avons multiplié par 1.001 puis 1.01 la fonction $\theta_{4,4}$ définie par (2.56) avec $p = 4$ et $n = 4$). Le test de la figure 3.32 montre que la variation totale (VT) au cours du temps augmente. De plus, de légères oscillations parasites sont visibles. Cela confirme l'analyse numérique concernant la zone TVD.

Le second test concerne les fonctions $\theta(r, \nu)$ définie par (2.56) mais avec différents p. Nous avons représenté sur la figure 3.33 les cas $p = 1$, $p = 4$ et $p = 16$. Les erreurs relatives en normes L^1 sont 0.23 pour $p = 1$, 0.04 pour $p = 4$ et 0.11 pour $p = 16$. Le cas $p = 4$ semble donner le meilleur résultat. Le comportement du schéma varie peu lorsque $p > 16$. Cela s'explique par l'étude de l'équation équivalente (voir la remarque 2 du paragraphe 1.1.4)

Le troisième test concerne la puissance n de la fonction qui intervient au voisinage de 1 dans l'expression (2.56) de θ en gardant $p = 4$. Nous avons représenté sur la figure 3.34 les cas $n = 2$, $n = 4$ et $n = 8$. Les erreurs relatives en normes L^1 sont 0.13 pour $n = 2$, 0.04 pour $n = 4$ et 0.08 pour $n = 8$. Le cas $n = 4$ semble donner le meilleur résultat.

Enfin, nous désirons tester les fonctions θ qui sont différentes de 0 pour $r < 0$. La fonction testée ici est définie par

$$\theta^m(r,\nu) = \begin{cases} 8 \text{ si } r \leq \nu^8 - 1, \\ \frac{\ln(1+r)}{\ln(\nu)} \text{ si } \nu^8 - 1 \leq r \leq 0, \\ \frac{\ln(1-\gamma r)}{\ln(\nu)} \text{ si } 0 < r \leq \frac{(1-\nu^4)}{\gamma}, \\ \frac{3}{\left(\frac{1-\nu^4}{\gamma-1}\right)^4}(r-1)^n + 1 \text{ si } \frac{(1-\nu^4)}{\gamma} \leq r < 2 - \frac{1-\nu^4}{\gamma}, \\ 4 \text{ si } r \geq 2 - \frac{1-\nu^4}{\gamma}. \end{cases} \tag{3.3}$$

Cette fonction modifiée θ^m satisfait les conditions suffisantes pour avoir un schéma TVD (voir la preuve de la proposition 10). La différence n'est visible que lors d'un changement de pente ($r < 0$). Nous testons donc cette fonction pour une condition initiale en chapeau. La différence est faible mais visible en zoomant près du sommet comme nous l'avons fait sur la figure 3.35. L'erreur relative en norme L^1 est 0.009 pour la fonction $\theta_{4,4}$ et 0.018 pour la fonction θ^m qui ne s'annule pas pour $r < 0$.

Ainsi la fonction $\theta_{4,4}$ utilisée dans les tests numériques semble la plus satisfaisante.

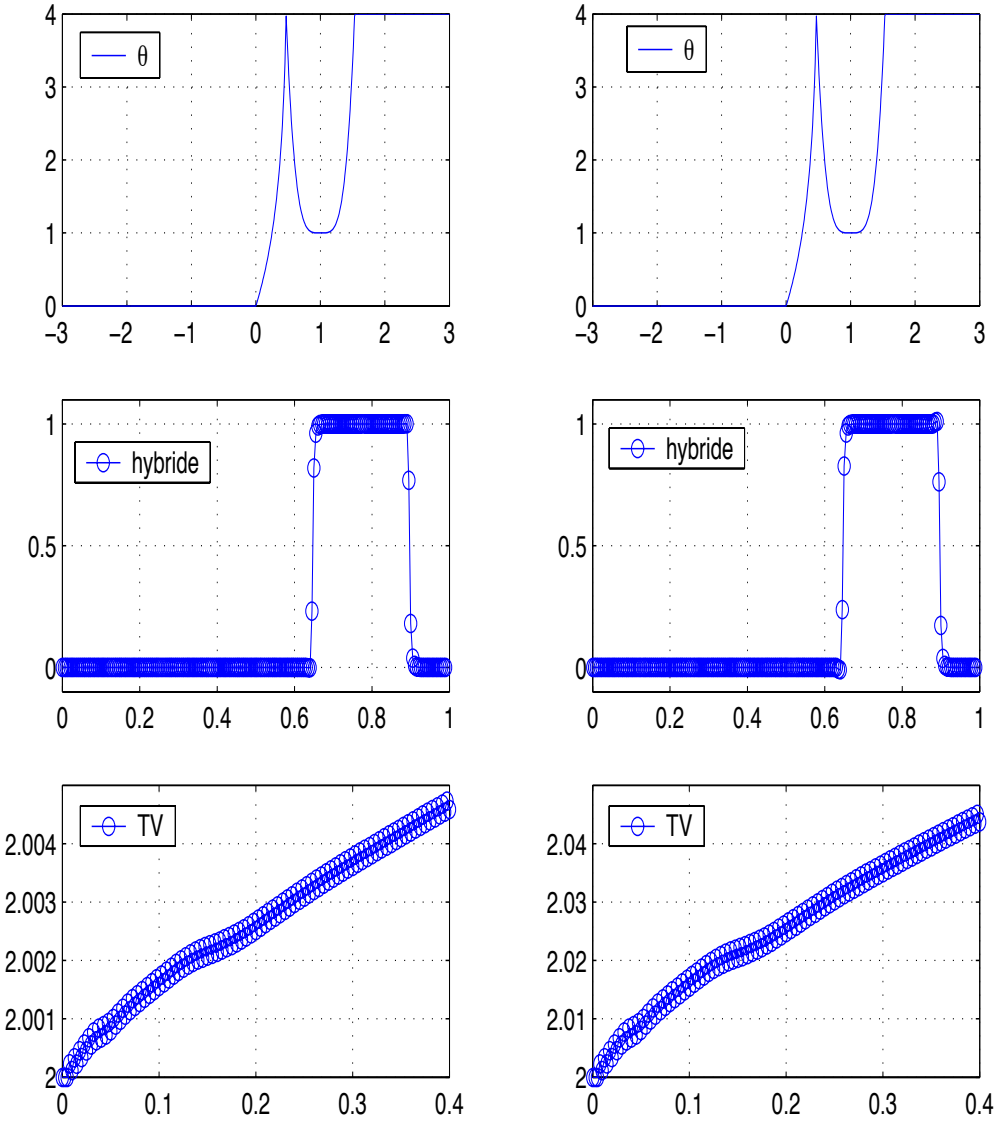

FIG. 3.32 – Fonction θ, solution du schéma hybride avec une donnée initiale en créneau dans le cas linéaire et variation totale (VT) en fonction du temps ; $\nu = 0.5$; $T = 0.4$; 200 points : a) cas $\theta = 0.001 \times \theta_{4,4}$; b) cas $\theta_{4,4} = 0.01 \times \theta_0$. Nous remarquons que le schéma hybride n'est pas TVD pour ces fonction θ. Dans le deuxième cas, on peut voir de faibles oscillations parasites (qui existent dans le premier cas, mais qui ne sont pas visibles à l'oeil nu).

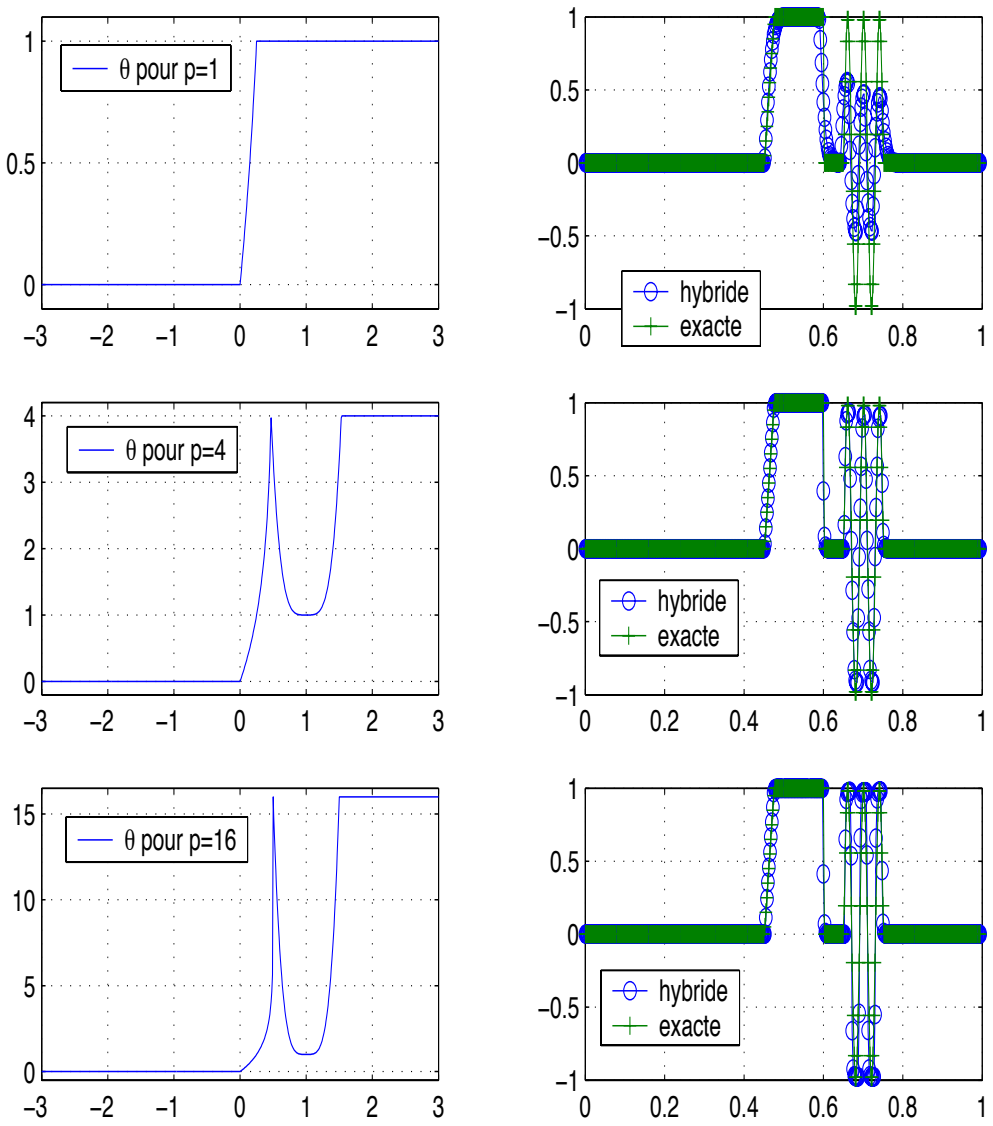

FIG. 3.33 – Fonction θ, solution du schéma hybride avec une donnée initiale avec raréfaction, choc et perturbations dans le cas linéaire ; $\mu = 0.5$; $T = 0.4$; 400 points. A gauche : fonction θ dans les cas $p = 1$, $p = 4$ et $p = 16$; à droite : solution exacte et solution numérique obtenue par le schéma hybride correspondantes. Le cas $p = 1$ n'est pas satisfaisant. Le cas $p = 16$ paraît mieux capturer les sommets des oscillations mais l'erreur entre la solution exacte et la solution numérique est plus grande que pour le cas $p = 4$.

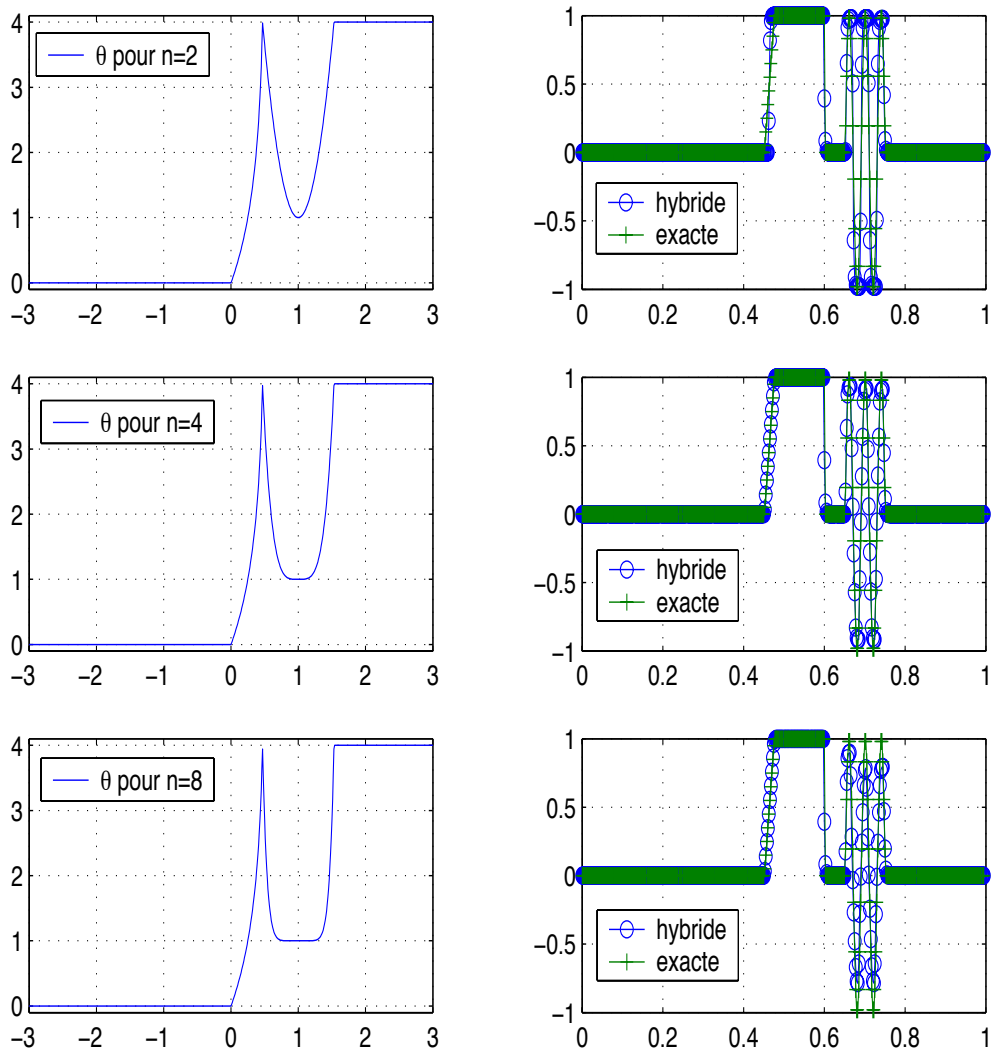

FIG. 3.34 – Fonction θ, solution du schéma hybride avec une donnée initiale avec raréfaction, choc et perturbations dans le cas linéaire ; $\mu = 0.5$; $T = 0.4$; 400 points. A gauche : fonction θ dans les cas $n = 2$, $n = 4$, $n = 8$ (et $p = 4$) ; à droite : solution exacte et solution numérique obtenue par le schéma hybride correspondantes. Le cas $n = 2$ paraît mieux capturer les sommets des oscillations mais ne donne pas une bonne approximation de la raréfaction. Le cas $n = 4$ semble être un bon compromis.

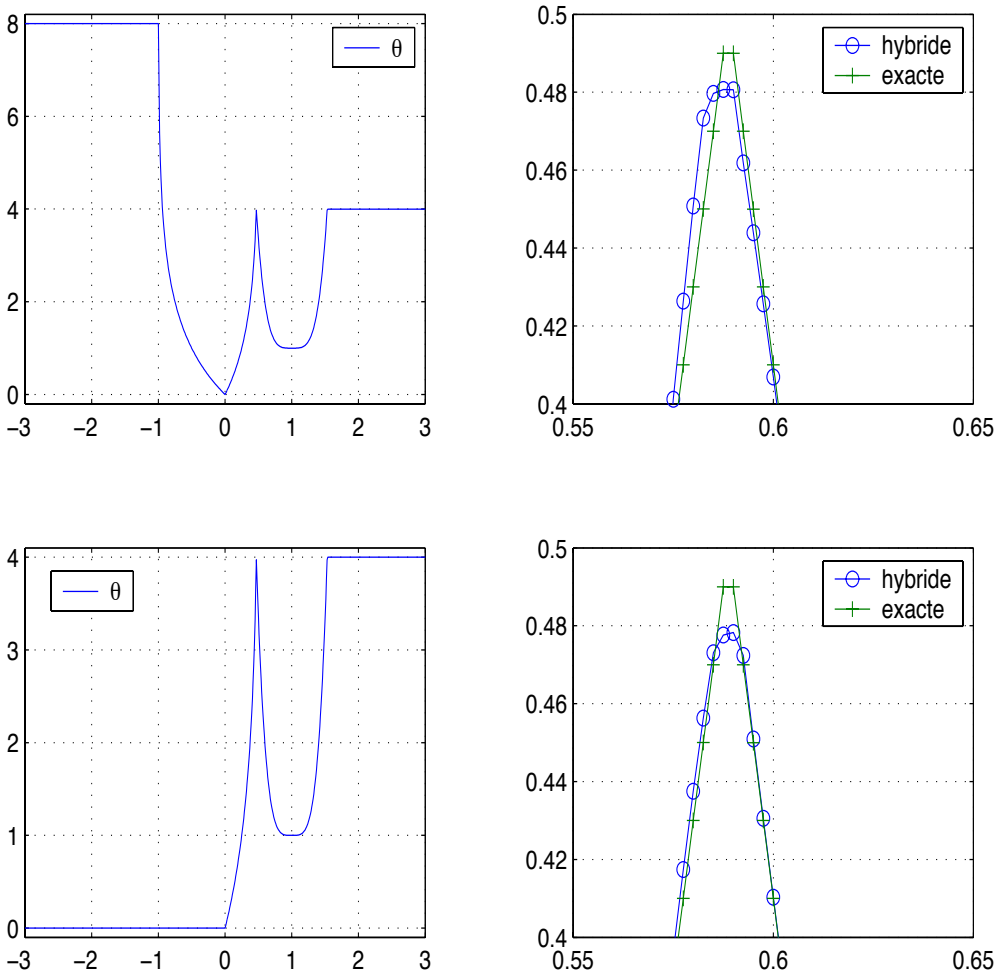

FIG. 3.35 – Fonction θ, solution du schéma hybride avec une donnée initiale en chapeau dans le cas linéaire ; $\mu = 0.5$; $T = 0.4$; 400 points. A gauche : fonction θ ($n = 4$ et $p = 4$) ; à droite : solutions "zoomées" exacte et calculée par le schéma hybride correspondantes.

3.2 Cas des systèmes en 1D

3.2.1 Problème du tube à choc pour les équations d'Euler compressibles

Nous considérons maintenant les équations d'Euler compressibles en une dimension d'espace. Ces équations s'écrivent sous la forme conservative

$$\partial_t u + \partial_x f(u) = 0,\ x \in]0;1[,\ t > 0, \tag{3.4}$$

où $u = (\rho, \rho v, \rho E)$ est le vecteur des variables conservatives, $f(u) = (\rho v, \rho v^2 + p, (\rho E + p)v)$ est le flux, ρ est la densité, v la vitesse, p la pression, ρE l'énergie volumique. L'hypothèse des gaz parfaits fournit l'équation d'état

$$p = (\gamma - 1)\left(\rho E - \frac{1}{2}\rho v^2\right). \tag{3.5}$$

Nous choisissons pour nos tests numériques la valeur 1.4 pour γ. Dans le premier test, nous considérons le cas test de Sod [18]. Il s'agit d'une donnée initiale constante par morceaux avec un état à droite $u_R = (\rho_R, \rho_R v_R, \rho_R E_R)$ et un état à gauche $u_L = (\rho_L, \rho_L v_L, \rho_L E_L)$ d'une interface située au point d'abscisse $x = 0.5$:

$$\begin{array}{ll} \rho_R = 1, & \rho_R = 0.125, \\ v_L = 0, & v_R = 0, \\ p_L = 1, & p_R = 0.1. \end{array} \tag{3.6}$$

La solution exacte est constituée d'une 1-raréfaction, d'une 2-discontinuité de contact et d'un 3-choc. Nous comparons les schémas de Roe et hybride (figure 3.36 et figure 3.37) avec un maillage uniforme de 200 points et un nombre de Courant de 0.5. Le choc est bien mieux capturé pour le schéma hybride. Remarquons que le schéma de Lax-Wendroff n'aboutit pas (explosion du calcul).

3.2.2 Onde de raréfaction pour les équations d'Euler compressibles

Nous considérons toujours les équations d'Euler. Les conditions initiales sont maintenant $u_L = (\rho_L, \rho_L v_L, \rho_L E_L)$ et $u_R = (\rho_R, \rho_R v_R, \rho_R E_R)$ avec une interface à $x = 0.5$:

$$\begin{array}{ll} \rho_L = 5, & \rho_R = 0.125, \\ v_L = 0, & v_R = 0, \\ p_L = 5, & p_R = 0.1. \end{array} \tag{3.7}$$

Le maillage est uniforme avec 200 points, le nombre de Courant est $\nu = 0.45$. Les solutions calculées par le schéma de Roe et le schéma hybride sont représentées au temps $T = 0.18$ dans les figures 3.38 et 3.39. Le schéma hybride capture mieux le choc. Cependant le schéma de Roe et le schéma hybride présentent tous les deux un choc entropique. Nous savons que le schéma de Roe n'est pas un schéma entropique. Le schéma hybride ne l'est pas non plus. Cette apparition d'un choc va motiver une analyse pour tenter de rendre le schéma entropique dans la partie suivante. Avant de nous intéresser à la dissipation d'entropie, nous terminons par l'extension du schéma hybride aux systèmes en 2D.

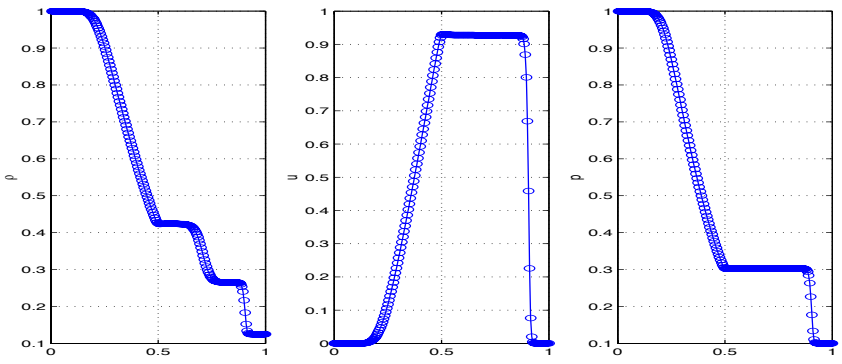

FIG. 3.36 – Schéma de Roe pour les équations d'Euler : tube à choc, 200 points $\nu = 0.5$; $T = 0.23$.

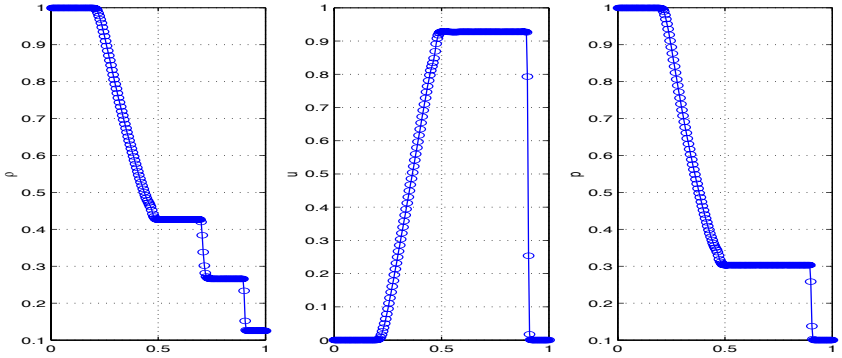

FIG. 3.37 – Schéma hybride pour les équations d'Euler : tube à choc, 200 points, $\nu = 0.5$; $T = 0.23$.

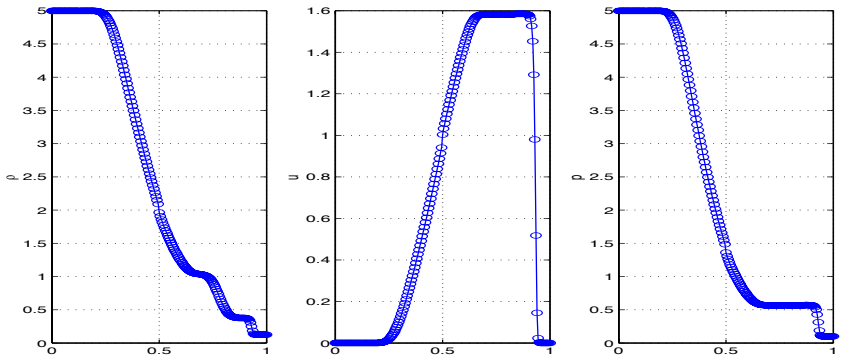

FIG. 3.38 – Schéma de Roe pour les équations d'Euler avec une onde de raréfaction ; 200 points $\nu = 0.45$; $T = 0.18$.

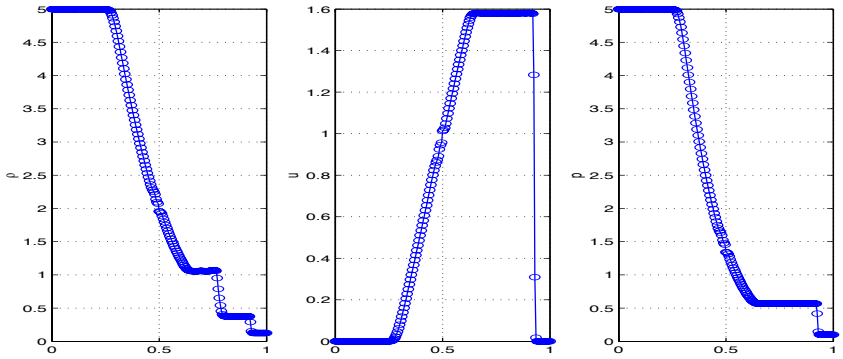

FIG. 3.39 – Schéma hybride pour les équations d'Euler avec une onde de raréfaction ; 200 points $\nu = 0.45$; $T = 0.18$.

3.3 Cas des systèmes en 2d pour les équations d'Euler

3.3.1 Problème d'un choc par réflexion

Nous considérons les équations d'Euler compressibles en deux dimensions d'espace :

$$\partial_t u + \partial_x f(u) + \partial_y g(u) = 0,\ x \in]0; 4.12829[,\ y \in]0; 1[\, t > 0, \tag{3.8}$$

où $u = (\rho, \rho v, \rho w, \rho E)$ est le vecteur des variables conservatives, $f(u) = (\rho v, \rho v^2 + p, \rho vw, (\rho E + p)v)$ et $g(u) = (\rho w, \rho vw, \rho w^2 + p, (\rho E + p)w)$ sont les flux, v et w sont les composantes de la vitesse en x et en y respectivement. L'hypothèse des gaz parfaits fournit l'équation d'état

$$p = (\gamma - 1)\left(\rho E - \frac{1}{2}\rho(v^2 + w^2)\right). \tag{3.9}$$

avec $\gamma = 1.4$. Nous considérons le problème de la réflexion d'un choc oblique sur le côté inférieur d'un domaine rectangulaire défini par $0 \leq x \leq 4.12829$ et $0 \leq y \leq 1$ [1]. La donnée initiale est u_0 donnée par

$$\rho = 1,\ u = 2.9,\ v = 0 \text{ et } p = \frac{1}{1.4}. \tag{3.10}$$

Les conditions aux bords sont :

- pour $x = 0$ (côté gauche du domaine), toutes les variables sont fixées et prennent la même valeur qu'au temps initial ;
- pour $x = 4.12829$ (côté droit du domaine), aucune variable n'est imposée ;
- pour $y = 1$ (bord supérieur), les variables sont fixées par

$$\rho = 1.7,\ v = 2.619,\ w = -0.506,\ p = 1.528; \tag{3.11}$$

- pour $y = 0$, les conditions sont celles d'une réflexion. Numériquement, nous rajoutons une ligne d'états fantômes sous le bord inférieur. Dans chaque cellule, la densité ρ, la composante v de la vitesse en x et la pression p sont identiques aux quantités des cellules directement au-dessus tandis que la composante w de la vitesse en y est l'opposé de la quantité des cellules directement au-dessus.

La solution exacte est representée sur la figure 3.40. Le maillage utilisé est un maillage rectangulaire uniforme avec $80 \times 25 = 2000$ cellules. Nous avons représenté les lignes de niveaux pour la densité, la pression, les composantes v puis w de la vitesse. Le schéma de Roe est très diffusif (figure 3.41). La solution du schéma de Lax-Wendroff est très éloignée de la solution exacte et présente de fortes oscillations (figure 3.42). Le schéma hybride par contre donne des résultats satisfaisants (figure 3.43).

3.3.2 Problème de la marche ascendante

Dans ce cas test, les équations sont toujours les équations d'Euler compressibles en 2D. Il s'agit d'un écoulement dans un domaine rectangulaire. Nous considérons maintenant le problème de la marche ascendante.Le domaine est rectangulaire : $0 \leq x \leq 3$ et $0 \leq y \leq 1$. La marche est un rectangle défini par : $0.6 \leq x \leq 3$ et $0 \leq y \leq 0.2$. La condition initiale est u_0 donnée par

$$\rho = 1.4,\ v = 3,\ w = 0 \text{ et } p = 1. \tag{3.12}$$

Les conditions aux bords sont :

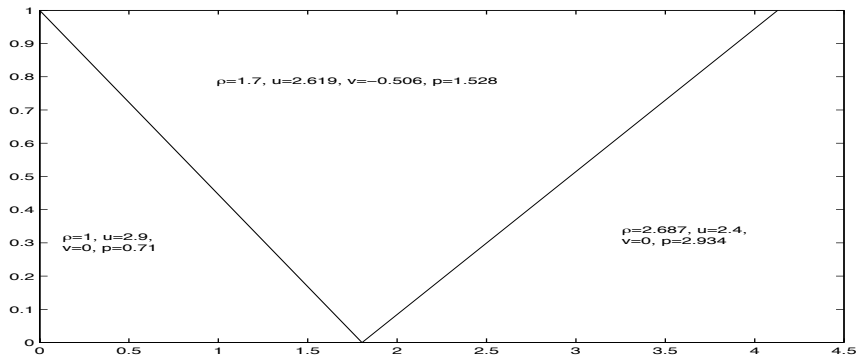

FIG. 3.40 – Solution exacte pour le problème d'un choc avec réflexion en 2D.

- pour $x = 0$ (côté gauche du domaine), toutes les variables sont fixées et prennent la même valeur qu'au temps initial ;
- pour $x = 3$ (côté droit du domaine), aucune variable n'est imposée (nous ne considérons pas le domaine défini par la marche) ;
- pour $x = 0.6$ et $0 \leq y \leq 0.2$, les conditions sont celles d'une réflexion par rapport à une paroi verticale ;
- pour $y = 1$ (bord supérieur), les conditions sont celles d'une réflexion par rapport à une paroi horizontale ;
- pour $y = 0$, $0 \leq x \leq 0.6$ ainsi que pour $y = 0.2$, $0.6 \leq x \leq 3$, les conditions sont celles d'une réflexion par rapport à une paroi horizontale.

Les réflexions sont traitées comme dans le cas test précédent. Le maillage utilisé est un maillage rectangulaire uniforme avec $120 \times 40 = 4800$ cellules. Nous avons représenté les lignes de niveaux pour la densité, la composante v puis w de la vitesse en x et en y respectivement et la pression.La solution du schéma hybride rend compte de la structure de la solution (figure 3.44). Cependant, nous voyons apparaître un choc entropique au-dessus de la marche.

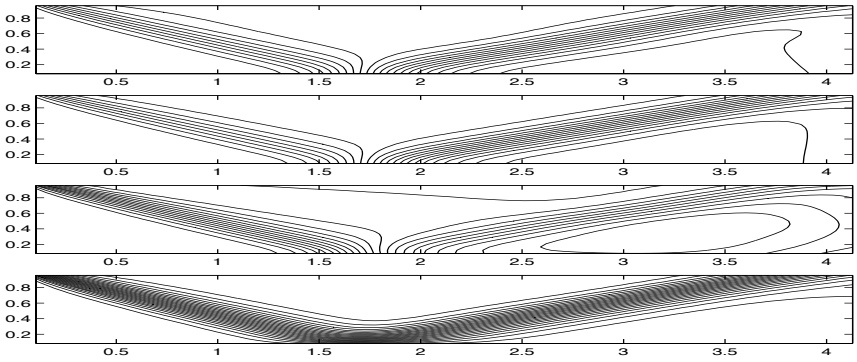

FIG. 3.41 – Schéma de Roe pour le problème d'un choc avec reflexion en 2D ; $\mu = 0.4$; $T = 1.5$.

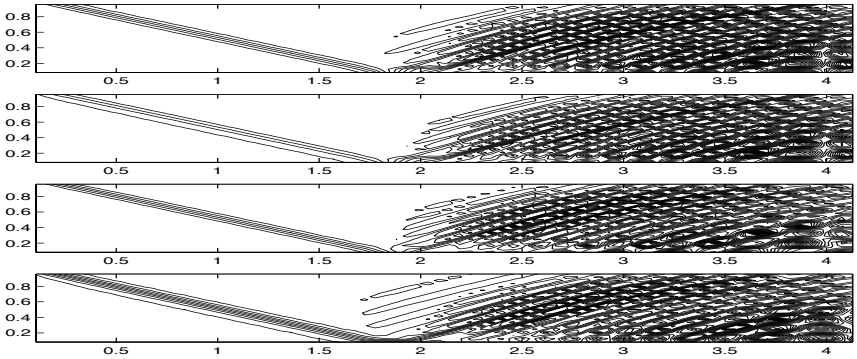

FIG. 3.42 – Schéma de Lax-Wendroff pour le problème d'un choc avec reflexion en 2D ; $\nu = 0.4$; $T = 1.5$.

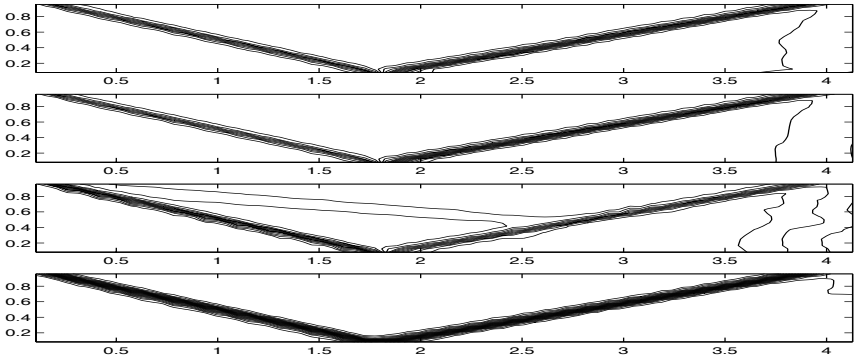

FIG. 3.43 – Schéma hybride pour le problème d'un choc avec reflexion en 2D ; $\nu = 0.4$; $T = 1.5$.

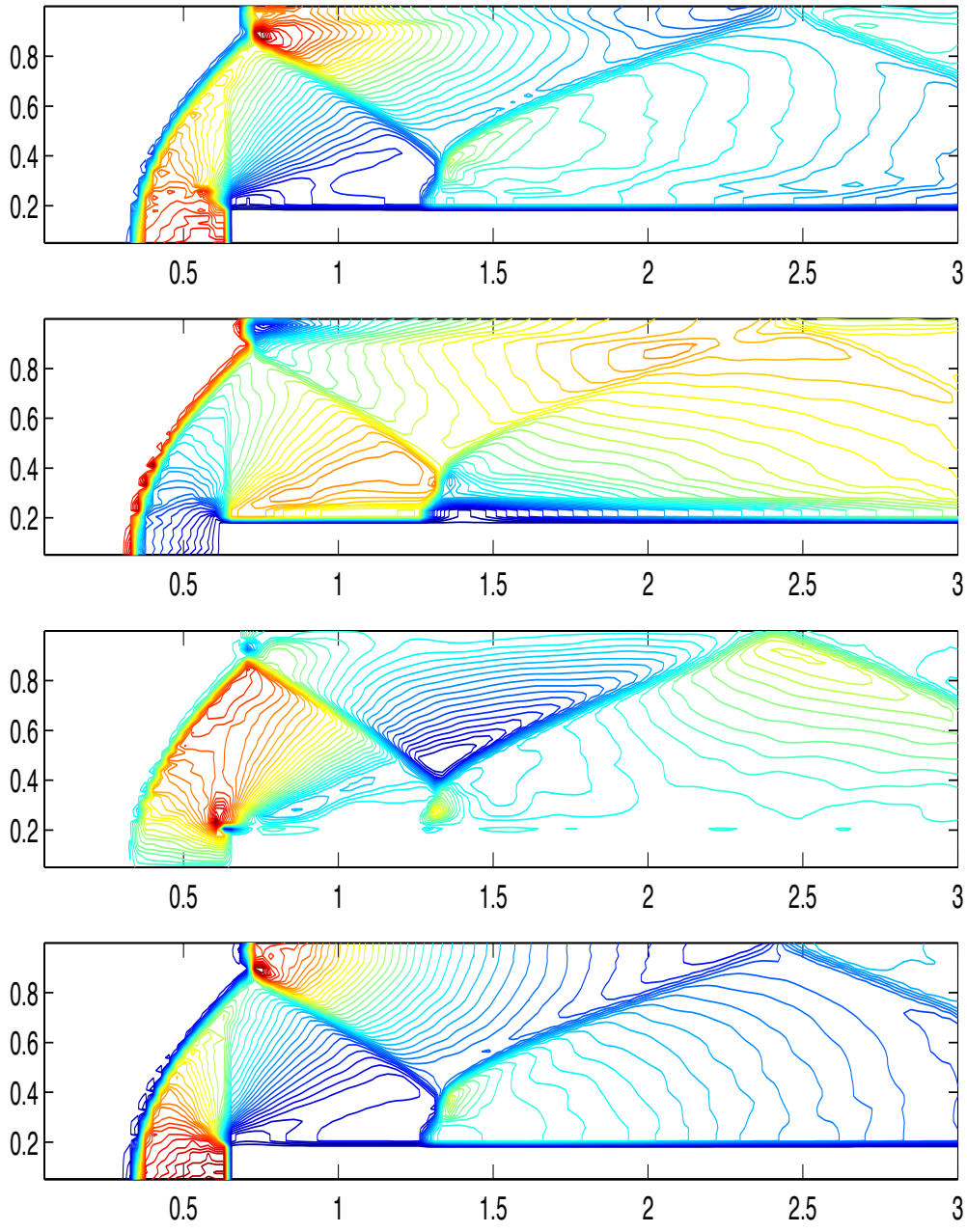

FIG. 3.44 – Schéma hybride pour la marche ascendante ; $\mu = 0.45$; $T = 4$.

4

Optimisation du paramètre θ de diffusion par approximation de la dissipation d'entropie numérique

4.1 Présentation du problème

Nous avons vu dans le chapitre précédent que le schéma hybride proposé permet d'obtenir de bons résultats notamment en ce qui concerne la capture des chocs. Cependant, dans certains cas (cas test des équations d'Euler en une dimension d'espace), notre schéma peut capturer des solutions non physiques et faire apparaître un choc d'entropie. Il est alors nécessaire de corriger le paramètre θ afin d'augmenter l'intensité de la matrice de diffusion. Le critère retenu pour la correction du paramètre θ concerne la production/dissipation d'entropie. Nous voulons en effet qu'au niveau discret, la solution calculée par le schéma satisfasse une inégalité d'entropie. Ainsi, dans chaque cellule où cette inégalité n'est pas vérifiée, le paramètre θ sera ajusté afin d'augmenter l'intensité de la matrice de diffusion, qui est une fonction décroissante de θ d'après notre analyse, jusqu'à ce que l'inégalité soit satisfaite. En premier lieu, il est nécessaire d'obtenir une bonne approximation de la production/dissipation d'entropie. Nous avons choisi d'utiliser la formule de Tadmor [20],[21],[22]. Ensuite, nous devons garantir l'existence d'une valeur de θ pour laquelle la solution numérique satisfait l'inégalité discrète d'entropie. Enfin, nous présentons un algorithme de type prédicteur/correcteur à partir du schéma hybride permettant de ne pas capturer les solutions non physiques.

4.2 Expression de la dissipation numérique

Nous considérons le système de lois de conservation

$$\partial_t u + \partial_x f(u) = 0,\ x \in \mathbb{R},\ t \geq 0. \tag{4.1}$$

Nous supposons qu'il existe une paire entropie-flux d'entropie (S, F) au sens donné par Lax. Ainsi

$$S'(u)f'(u) = F'(u), \tag{4.2}$$

et S est supposé strictement convexe. La solution faible entropique (ou physique) de (4.1) doit satisfaire l'inéquation aux dérivées partielles

$$\partial_t S(u) + \partial_x F(u) \leq 0,\ x \in \mathbb{R},\ t \geq 0, \tag{4.3}$$

au sens des distributions.
Il s'agit alors de trouver un schéma qui permette de satisfaire l'inéquation (4.3) numériquement dans chaque cellule. Il faut alors connaître numériquement la quantité $\partial_t S(u) + \partial_x F(u)$. Nous choisissons de considérer la formulation de la dissipation d'entropie numérique introduite par Tadmor que nous présentons ci-dessous (pour plus de détails, voir les articles de Tadmor [20, 21, 22]). Grâce à la stricte convexité de S, nous pouvons introduire les variables entropiques du système $\xi = \xi(u) = (S'(u))^T$. La dissipation d'entropie

$$\eta = \partial_t S(u) + \partial_x F(u), \tag{4.4}$$

est calculée à la multiplication par Δt^n près, à l'aide de la formule suivante :

$$\eta_j^{n+1} = S(u_j^{n+1}) - S(u_j^n) + \lambda(\psi(u_j^n, u_{j+1}^n, \lambda, \theta) - \psi(u_{j-1}^n, u_j^n, \lambda, \theta)). \tag{4.5}$$

Le flux numérique d'entropie proposé par Tadmor est donné par

$$\begin{aligned} \psi(u, v, \lambda, \theta) =& < \frac{\xi(u) + \xi(v)}{2}, \phi(u, v, \lambda, \theta) > \\ & -\frac{1}{2}\big(< \xi(u), f(u) > - F(u) \\ & \qquad + < \xi(v), f(v) > - F(v)\big) \end{aligned} \tag{4.6}$$

où $<,>$ désigne le produit scalaire dans $\mathbb{R}^3$ et ϕ est le flux numérique du schéma hybride. Nous pouvons vérifier facilement que le flux d'entropie numérique ψ est consistant avec le flux d'entropie F. Nous remarquons que la dépendance du flux d'entropie numérique ψ en θ provient de celle du flux numérique ϕ.

4.3 Existence d'un paramètre θ permettant d'obtenir une dissipation d'entropie négative

Dans ce paragraphe, nous présentons les travaux de Tadmor ([20, 21, 22]) sur la dissipation d'entropie. Le but est de montrer que la dissipation d'entropie numérique du schéma de Lax-Friedrichs modifié (LFm) est toujours négative.

4.3.1 Cas des systèmes semi-discrets

On pose

$$\begin{aligned} g(\xi) &= f(u(\xi)), \\ \varphi(\xi) &= < \xi; u(\xi) > - S(u(\xi), \\ \gamma(\xi) &= < \xi; g(\xi) > - F(u(\xi)). \end{aligned}$$

Ainsi

$$u(\xi) = \nabla_\xi \varphi(\xi),$$
$$g(\xi) = \nabla_\xi \gamma(\xi).$$

On s'intéresse dans un premier temps au schéma semi-discret (en espace) :

$$\mathrm{d}_t u_j(t) = -\frac{1}{\Delta x}(\phi_{j+\frac{1}{2}} - \phi_{j-\frac{1}{2}}). \tag{4.7}$$

On a alors

$$\mathrm{d}_t S(u_j(t)) = < \xi_j(t), \mathrm{d}_t u_j(t) > = < \xi_j(t), -\frac{1}{\Delta x}(\phi_{j+\frac{1}{2}} - \phi_{j-\frac{1}{2}}) > . \tag{4.8}$$

Ainsi, avec la définition (4.6) du flux d'entropie numérique et les notations $\Delta\xi_{j+\frac{1}{2}} = \xi^n_{j+1} - \xi^n_j$ et $\Delta\gamma_{j+\frac{1}{2}} = \gamma(\xi^n_{j+1}) - \gamma(\xi^n_j)$, on obtient

$$\begin{aligned} \mathrm{d}_t S(u_j(t)) + \frac{1}{\Delta x}(\psi_{j+\frac{1}{2}} - \psi_{j-\frac{1}{2}}) = \frac{1}{2\Delta x}(&< \Delta\xi_{j+\frac{1}{2}}, \phi_{j+\frac{1}{2}} > - \Delta\gamma_{j+\frac{1}{2}}) \\ &+ \frac{1}{2\Delta x}(< \Delta\xi_{j-\frac{1}{2}}, \phi_{j-\frac{1}{2}} > - \Delta\gamma_{j-\frac{1}{2}}). \end{aligned} \tag{4.9}$$

On obtient alors le théorème suivant [22].

Théorème 1 *Le schéma semi-discret* (4.7) *est entropiquement stable (respectivement conservatif) si* $< \Delta\xi_{j+\frac{1}{2}}, \Delta\gamma_{j+\frac{1}{2}} > \leq \Delta\gamma_{j+\frac{1}{2}}$ *(respectivement* $< \Delta\xi_{j+\frac{1}{2}}, \Delta\gamma_{j+\frac{1}{2}} > = \Delta\gamma_{j+\frac{1}{2}}$*).*

On considère les schémas pouvant s'écrire sous la forme visqueuse

$$\mathrm{d}_t u_j(t) = -\frac{1}{2\Delta x}(f(u_{j+1}) - f(u_{j-1})) + \frac{1}{2\Delta x}(Q_{j+\frac{1}{2}}\Delta\xi_{j+\frac{1}{2}} - Q_{j-\frac{1}{2}}\Delta\xi_{j-\frac{1}{2}}), \tag{4.10}$$

ce qui est le cas de notre schéma hybride. La matrice $Q_{j+\frac{1}{2}}$ est la matrice de viscosité. Ainsi

$$Q_{j+\frac{1}{2}}\Delta\xi_{j+\frac{1}{2}} = f(u_{j+1}) + f(u_j) - 2\phi_{j+\frac{1}{2}}. \tag{4.11}$$

On cherche un flux numérique ϕ pour lequel le schéma (4.7) est entropiquement conservatif. On a alors le théorème suivant [22].

Théorème 2 *Le schéma* (4.7) *est entropiquement conservatif pour le flux numérique (par rapport aux variables entropiques)*

$$\phi^*_{j+\frac{1}{2}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} g(\xi_{j+\frac{1}{2}}(s))\, ds. \tag{4.12}$$

où $\xi_{j+\frac{1}{2}}(s) = \frac{\xi_j + \xi_{j+1}}{2} + s\Delta\xi_{j+\frac{1}{2}}$.
La matrice de viscosité associée est alors

$$Q^*_{j+\frac{1}{2}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} 2sB(\xi_{j+\frac{1}{2}}(s))\, ds$$

avec $B(\xi) = g'(\xi)$.

Démonstration. Dans ce cas, on a

$$\begin{aligned} < \Delta\xi_{j+\frac{1}{2}}, \phi^*_{j+\frac{1}{2}} > &= \int_{-\frac{1}{2}}^{\frac{1}{2}} < \Delta\xi_{j+\frac{1}{2}}, g(\xi_{j+\frac{1}{2}}(s)) > \, ds \\ &= \int_{\xi_{-\frac{1}{2}}}^{\xi_{\frac{1}{2}}} < \mathrm{d}\xi, g(\xi) > \\ &= \int_{\xi_{-\frac{1}{2}}}^{\xi_{\frac{1}{2}}} < \mathrm{d}\xi, \mathrm{d}_\xi \gamma(\xi) > \\ &= \Delta\gamma_{j+\frac{1}{2}}. \end{aligned}$$

Après une intégration par partie,

$$g^*_{j+\frac{1}{2}} = \frac{f(u_j) + f(u_{j+1})}{2} - \int_{-\frac{1}{2}}^{\frac{1}{2}} sB(\xi_{j+\frac{1}{2}}(s))\, ds\, \Delta\xi_{j+\frac{1}{2}}.$$

Donc d'après l'expression de la matrice de viscosité Q, on obtient

$$Q^*_{j+\frac{1}{2}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} 2sB(\xi_{j+\frac{1}{2}}(s))\, ds,$$

ce qui termine la démonstration.
On obtient donc un critère pour montrer qu'un schéma est entropiquement stable :

Théorème 3 *Pour le schéma semi-discret* (4.7)*, la dissipation d'entropie est*

$$d_t S(\xi_j(t)) + \frac{1}{\Delta x}(\psi_{j+\frac{1}{2}} - \psi_{j-\frac{1}{2}}) = -\frac{1}{4\Delta x}(< \Delta\xi_{j+\frac{1}{2}}, D_{j+\frac{1}{2}}\Delta\xi_{j+\frac{1}{2}} > + < \Delta\xi_{j-\frac{1}{2}}, D_{j-\frac{1}{2}}\Delta\xi_{j-\frac{1}{2}} >)$$

où $D = Q - Q^*$.
Ainsi le schéma (4.7) *est entropiquement stable si* $< \Delta\xi_{j+\frac{1}{2}}, D_{j+\frac{1}{2}}\Delta\xi_{j+\frac{1}{2}} > \geq 0$.

Démonstration. D'après (4.11),

$$< \Delta\xi_{j+\frac{1}{2}}, \phi_{j+\frac{1}{2}} > -\Delta\psi_{j+\frac{1}{2}} = < \Delta\xi_{j+\frac{1}{2}}, \phi^*_{j+\frac{1}{2}} - \frac{1}{2} D_{j+\frac{1}{2}}\Delta\xi_{j+\frac{1}{2}} > -\Delta\psi_{j+\frac{1}{2}}.$$

Or par définition de ϕ^*,

$$< \Delta\xi_{j+\frac{1}{2}}, \phi^*_{j+\frac{1}{2}} - \Delta\psi_{j+\frac{1}{2}} = 0.$$

Donc

$$< \Delta\xi_{j+\frac{1}{2}}, \phi_{j+\frac{1}{2}} > -\Delta\psi_{j+\frac{1}{2}} = - < \Delta\xi_{j+\frac{1}{2}}, \frac{1}{2} D_{j+\frac{1}{2}}\Delta\xi_{j+\frac{1}{2}} > .$$

On conclue par (4.9).

4.3.2 Cas des schémas totalement discrétisés

Dissipation d'entropie numérique

Nous présentons le calcul de la dissipation d'entropie numérique pour un schéma totalement discrétisé présenté dans l'article [22]. On considère le schéma totalement discret (explicite)

$$u_j^{n+1} = u_j^n - \lambda(\phi^n_{j+\frac{1}{2}} - \phi^n_{j-\frac{1}{2}}).$$

On désire calculer

$$\eta = S_j^{n+1} - S_j^n + \lambda(\tilde{\Psi}_{j+\frac{1}{2}} - \tilde{\Psi}_{j-\frac{1}{2}}).$$

On a alors le théorème [22]

Théorème 4 *La dissipation d'entropie η est égale à*

$$\eta = -\lambda \mathcal{E}_j^{esp} + \mathcal{E}_j^{temps}$$

où

$$\mathcal{E}_j^{esp} = \frac{1}{4} < \Delta\xi_{j+\frac{1}{2}}, D_{j+\frac{1}{2}}\Delta\xi_{j+\frac{1}{2}} > + \frac{1}{4} < \Delta\xi_{j-\frac{1}{2}}, D_{j-\frac{1}{2}}\Delta\xi_{j-\frac{1}{2}} >$$

et

$$\mathcal{E}_j^{temps} = \int_{-\frac{1}{2}}^{\frac{1}{2}} (\frac{1}{2} + s) < \Delta\xi_j^{n+\frac{1}{2}}, H(\xi_j^{n+\frac{1}{2}}(s))\Delta\xi_j^{n+\frac{1}{2}} > ds,$$

où $H(\xi) = u_\xi(\xi)$ *et* $\xi_j^{n+\frac{1}{2}} = \xi_j^n + (s + \frac{1}{2})\Delta\xi_j^{n+\frac{1}{2}}$.

Démonstration On calcule de deux manières différentes la quantité

$$< \xi_j^n, u_j^{n+1} - u_j^n > .$$

D'une part,

$$< \xi_j^n, u_j^{n+1} - u_j^n > = < \xi_j^n, \phi_{j+\frac{1}{2}}^n - \phi_{j-\frac{1}{2}}^n > = \psi_{j+\frac{1}{2}} - \psi_{j-\frac{1}{2}} + \mathcal{E}_j^{esp}.$$

exactement comme dans le cas semi-discret. D'autre part, puisque

$$\xi_j^{n+\frac{1}{2}} = \xi_j^n + (s + \frac{1}{2})\Delta\xi_j^{n+\frac{1}{2}},$$

on obtient

$$S_j^{n+1} - S_j^n = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{d}{ds} S(u(\xi_j^{n+\frac{1}{2}})(s))\, ds = \int_{-\frac{1}{2}}^{\frac{1}{2}} < H(\xi_j^{n+\frac{1}{2}}(s))\Delta\xi_j^{n+\frac{1}{2}}, \xi_j^{n+\frac{1}{2}}(s) > \, ds$$

car $H = u_\xi$. D'où

$$S_j^{n+1} - S_j^n = \int_{-\frac{1}{2}}^{\frac{1}{2}} < H(\xi_j^{n+\frac{1}{2}}(s))\Delta\xi_j^{n+\frac{1}{2}}, \xi_j^n ds > + \mathcal{E}_j^{temps}.$$

Enfin,

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} < H(\xi_j^{n+\frac{1}{2}}(s))\Delta\xi_j^{n+\frac{1}{2}}, \xi_j^n ds > = < \xi_j^n, u_j^{n+1} - u_j^n > .$$

On obtient alors

$$< \xi_j^n, u_j^{n+1} - u_j^n > = S_j^{n+1} - S_j^n - \mathcal{E}_j^{temps}.$$

On termine la démonstration en égalant les deux expressions trouvées de $< \xi_j^n, u_j^{n+1} - u_j^n >$.

Dissipation d'entropie numérique pour le schéma de Lax-Friedrichs modifié

A notre connaissance, il n'existe pas de démonstration correcte permettant de dire que le schéma de Lax-Friedrichs modifié est entropique pour le flux de Tadmor, dans le cas général ou même linéaire scalaire. Nous donnons ici une démonstration dans le cas particulier d'une équation linéaire scalaire. Nous considérons l'équation d'advection linéaire :

$$\partial_t u + a\partial_x u = 0,\ a \in \mathbb{R},$$

dont nous voulons approcher la solution par le schéma de Lax-Friedrichs modifié :

$$u_j^{n+1} = u_j^n - \lambda(g_{j+\frac{1}{2}} - g_{j-\frac{1}{2}}),$$

où

$$g_{j+\frac{1}{2}} = \frac{f_j + f_{j+1}}{2} - \frac{1}{4\lambda}\Delta u_{j+\frac{1}{2}}.$$

Ainsi en réarrangeant les termes, nous pouvons écrire la quantité $\Delta u_j^{n+\frac{1}{2}}$ sous la forme :

$$\Delta u_j^{n+\frac{1}{2}} = -\frac{\lambda}{2}\left((a - \frac{1}{2\lambda})\Delta u_{j+\frac{1}{2}} + (a + \frac{1}{2\lambda})\Delta u_{j-\frac{1}{2}}\right).$$

Dans ce qui suit, nous noterons $\Delta_+ = \Delta u^n_{j+\frac{1}{2}}$ et $\Delta_- = \Delta u^n_{j-\frac{1}{2}}$ pour simplifier les expressions. D'après le théorème 4, nous avons

$$\begin{aligned}\mathcal{E}_j^{temps} &= \frac{1}{2}|\Delta u_j^{n+\frac{1}{2}}|^2 \\ &= \frac{1}{8}\left((\Delta_+ + \Delta_-)^2(a\lambda)^2 - (\Delta_+^2 - \Delta_-^2)(a\lambda) + \frac{1}{4}(\Delta_+ - \Delta_-)^2\right),\end{aligned}$$

et, puisque $D = Q - Q^* = \frac{1}{2\lambda}$ pour le schéma de Lax-Friedrichs modifié dans le cas linéaire, nous obtenons

$$\mathcal{E}_j^{esp} = \frac{1}{8\lambda}(\Delta_+^2 + \Delta_-^2). \tag{4.13}$$

Ainsi

$$\begin{aligned}\mathcal{E}_j &= \mathcal{E}_j^{temps} - \lambda\mathcal{E}_j^{esp} \\ &= \frac{1}{8}\left((\Delta_+ + \Delta_-)^2(a\lambda)^2 - (\Delta_+^2 - \Delta_-^2)(a\lambda) - \frac{3}{4}(\Delta_+^2 + \Delta_-^2) - \frac{1}{2}\Delta_+\Delta_-\right).\end{aligned}$$

Si $\Delta_+ + \Delta_- = 0$, alors $\mathcal{E}_j \leq 0$. Dans le cas contraire, nous obtenons un trinôme du second degré d'inconnue $a\lambda$ de discriminant :

$$\delta = 4(\Delta_+ + \Delta_-)^2(\Delta_+^2 + \Delta_-^2) \geq 0.$$

Les solutions sont alors

$$(a\lambda)_1 = \frac{(\Delta_+ - \Delta_-) + 2\sqrt{\Delta_+^2 + \Delta_-^2}}{2(\Delta_+ + \Delta_-)}, \tag{4.14}$$

$$(a\lambda)_2 = \frac{(\Delta_+ - \Delta_-) - 2\sqrt{\Delta_+^2 + \Delta_-^2}}{2(\Delta_+ + \Delta_-)}. \tag{4.15}$$

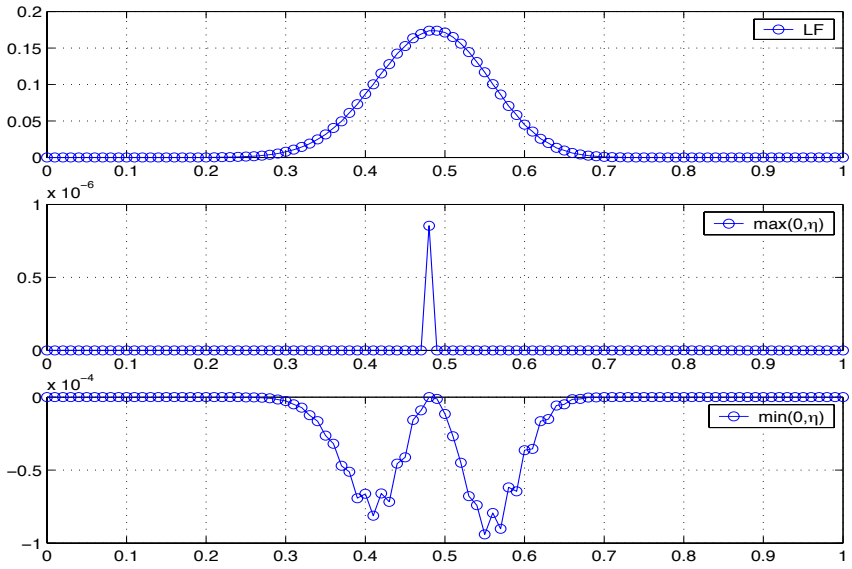

FIG. 4.1 – Schéma de Lax-Friedrichs pour l'équation de transport linéaire scalaire avec une condition initiale en chapeau ; 100 points, $\mu = 0.5$; $T = 0.3$. Il y a production d'entropie.

En passant aux coordonnées polaires $\Delta_+ = r\cos\theta$ et $\Delta_- = r\sin\theta$, nous obtenons deux fonctions de θ :

$$(a\lambda)_1 = \frac{\cos(\theta + \frac{\pi}{4}) + \sqrt{2}}{2\sin(\theta + \frac{\pi}{4})}, \tag{4.16}$$

$$(a\lambda)_2 = \frac{\cos(\theta + \frac{\pi}{4}) - \sqrt{2}}{2\sin(\theta + \frac{\pi}{4})}, \tag{4.17}$$

et nous pouvons facilement montrer en étudiant ces fonctions que l'une des solutions du trinôme est inférieure ou égale à $-\frac{1}{2}$ et l'autre solution est supérieure ou égale à $\frac{1}{2}$. Ainsi nous avons démontré que (au moins) sur l'intervalle $[-\frac{1}{2}; \frac{1}{2}]$, le trinôme du second degré d'inconnue $(a\lambda)$ est négatif. Nous avons donc démontré la proposition suivante.

Proposition 12 *Le schéma de Lax-Friedrichs modifié dans le cas d'une équation scalaire linéaire satisfait l'inégalité d'entropie discrète avec le flux d'entropie de Tadmor sous la condition CFL* $|a\lambda| \leq \frac{1}{2}$.

Remarque 4 *La même étude pour le schéma de Lax-Friedrichs montre que nous ne pouvons pas obtenir un principe d'entropie pour ce schéma avec le flux de Tadmor : la borne inférieure (supérieure respectivement) de la solution positive (respectivement négative) est égale à* 0. *Ainsi lorsque* $\Delta_+ + \Delta_-$ *tend vers* 0 *(par exemple, lorsque la solution numérique présente un "chapeau" symétrique),* $|a\lambda|$, *et donc* $\frac{dt}{dx}$ *doit tendre aussi vers* 0 *pour satisfaire l'inégalité d'entropie. La situation est illustrée dans la figure 4.1. La figure 4.2 permet au contraire de voir que la dissipation d'entropie est bien négative dans ce cas pour le schéma de Lax-Friedrichs modifié.*

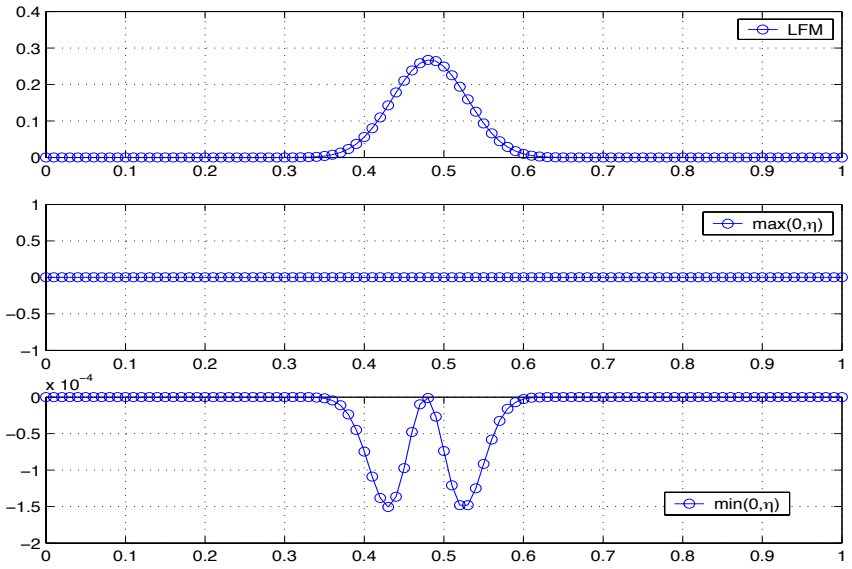

FIG. 4.2 – Schéma de Lax-Friedrichs modifié pour l'équation de transport linéaire scalaire avec une condition initiale en chapeau ; 100 points, $\mu = 0.5$; $T = 0.3$. Il n'y a pas de production d'entropie.

Nous avons ainsi une approximation numérique de $\partial_t S(u) + \partial_x F(u)$ grâce au flux de Tadmor. Nous allons maintenant utiliser ce flux d'entropie numérique pour corriger le paramètre θ du schéma hybide.

4.4 Détermination de θ dans chaque cellule

4.4.1 Méthode itérative

Nous présentons ici une détermination de $\theta_{j+\frac{1}{2}}$ par un procédé itératif basé sur le principe exposé par De Vuyst (voir [2]). Dans chaque cellule $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [t^n, t^{n+1}]$, nous considérons

$$u_j^{n+1} = u_j^n - \lambda(\phi(u_j^n, u_{j+1}^n, \lambda, \theta_{j+\frac{1}{2}}) - \phi(u_{j-1}^n, u_j^n, \lambda, \theta_{j-\frac{1}{2}})). \tag{4.18}$$

Pour la première étape, nous décidons de prendre la même valeur θ_j pour les deux paramètres $\theta_{j-\frac{1}{2}}$ et $\theta_{j+\frac{1}{2}}$. Ensuite nous cherchons θ_j qui rend la dissipation d'entropie numérique négative dans cette cellule par un procédé itératif. Dans chaque cellule, l'algorithme est ainsi :

Algorithme 7 *Algorithme itératif dans la cellule* $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [t^n, t^{n+1}]$.
- *Soit* $\Delta\theta > 0$ *un pas fixé. Soit* $q = 0$, $\theta_j^{n,q} = 1$ *(Lax Wendroff).*
- *Tant que* $\eta_j^{n+1}(\theta_j^{n,q}) > \delta$ *et* $\theta_j^{n,q} > -1$ *faire*
  - $\theta_j^{n,q+1} = \theta_j^{n,q} - \Delta\theta$
  - $q = q + 1$
  - *Calcul de* $\eta_j^{n+1}(\theta_j^{n,q})$

- *fin Tant que.*
- *Poser* $\theta_j^n = \theta_j^{n,q}$.

Ainsi nous avons à la fin de la première étape θ_j^n dans chaque cellule. Mais dans ce cas, le schéma hybride n'est pas conservatif puisque nous avons à chaque interface $j+\frac{1}{2}$ deux flux distincts $\phi(u_j, u_{j+1}, \theta_j)$ et $\phi(u_j, u_{j+1}, \theta_{j-1})$. Nous décidons alors dans la deuxième étape de prendre $\theta^n_{j+\frac{1}{2}} = \min(\theta^n_{j-1}, \theta^n_j)$ pour prendre en compte le fait que la diffusion dans le schéma hybride est une fonction décroissante de θ. A la fin de la deuxième étape, nous avons donc calculé le flux numérique hybride $\phi(u_j^n, u_{j+1}^n, \lambda, \theta^n_{j+\frac{1}{2}})$ pour tout j.

4.4.2 Procédé itératif avec prédicteur

Nous désirons greffer ce procédé itératif à la prédiction de θ effectuée à l'aide de la fonction $\theta(r,\nu)$ déterminée dans le paragraphe 2. Nous considérons alors avoir un prédicteur $\theta^n_{pred,j+\frac{1}{2}}$ pour $\theta^n_{j+\frac{1}{2}}$ déterminé à l'aide de la fonction $\theta(r,\nu)$ (ou par tout autre procédé). Cependant, les paramètres $\theta^n_{pred,j-\frac{1}{2}}$ et $\theta^n_{pred,j+\frac{1}{2}}$ ne rendent pas forcément la dissipation d'entropie négative dans la cellule j. Nous nous plaçons alors dans la cellule $[x^n_{j-\frac{1}{2}}, x^n_{j+\frac{1}{2}}]\times[t^n, t^{n+1}]$. Nous considérons dans $\mathbb{R}^2$ le segment $[A_jB_j]$ où A_j a pour coordonnées $(\theta^n_{pred,j-\frac{1}{2}}, \theta^n_{pred,j+\frac{1}{2}})$ et B_j a pour coordonnées $(-1,-1)$. Nous cherchons alors $(\theta^{n,+}_{j-\frac{1}{2}}, \theta^{n,-}_{j+\frac{1}{2}})$ qui rend la dissipation d'entropie $\eta_j^{n+1}(\theta^{n,+}_{j-\frac{1}{2}}, \theta^{n,-}_{j+\frac{1}{2}})$ négative. Dans chaque cellule, l'algorithme est :

Algorithme 8 *Algorithme itératif après une phase prédictrice*

- *Soit* $\Delta s > 0$ *un pas fixé. Soit* $q=0$, $s=\Delta s$, $\theta^{n,+,q}_{j-\frac{1}{2}} = \theta^n_{pred,j-\frac{1}{2}}$, $\theta^{n,-,q}_{j+\frac{1}{2}} = \theta^n_{pred,j+\frac{1}{2}}$.
- *Tant que* $\eta_j^{n+1}(\theta^{n,+,q}_{j-\frac{1}{2}}, \theta^{n,-,q}_{j+\frac{1}{2}}) > 0$ *et* $s < 1$ *faire*
  - $\theta^{n,+,q+1}_{j-\frac{1}{2}} = \theta^{n,+,q}_{j-\frac{1}{2}} + s(-1 - \theta^{n,+,q}_{j-\frac{1}{2}})$
  - $\theta^{n,-,q+1}_{j+\frac{1}{2}} = \theta^{n,-,q}_{j+\frac{1}{2}} + s(-1 - \theta^{n,-,q}_{j+\frac{1}{2}})$
  - $q = q+1$
  - $s = s + \Delta s$
  - *calcul de* $\eta_j^{n+1}(s) = \eta_j^{n+1}(\theta^{n,+,q}_{j-\frac{1}{2}}, \theta^{n,-,q}_{j+\frac{1}{2}})$
- *fin Tant que.*
- *Poser* $\theta^{n,+}_{j-\frac{1}{2}} = \theta^{n,+,q}_{j-\frac{1}{2}}$ *et* $\theta^{n,-}_{j+\frac{1}{2}} = \theta^{n,-,q}_{j+\frac{1}{2}}$.

Lorsque les deux valeurs $\theta^{n,-}_{j+\frac{1}{2}}$ et $\theta^{n,+}_{j+\frac{1}{2}}$ ont été calculées dans chaque cellule j, nous posons

$$\theta^n_{j+\frac{1}{2}} = \min(\theta^{n,-}_{j+\frac{1}{2}}, \theta^{n,+}_{j+\frac{1}{2}}) \tag{4.19}$$

et nous calculons $\phi(u_j^n, u_{j+1}^n, \lambda, \theta^n_{j+\frac{1}{2}})$.
Cet algorithme est directement applicable aux cas des systèmes sous sa forme vectorielle en remplaçant θ par $\vec{\theta}$. Cependant cette approche itérative peut être très coûteuse en temps de calcul. Comme le fait remarquer De Vuyst dans [2], nous pouvons utiliser une méthode de Newton pour trouver le paramètre s tel que $\eta_j^{n+1}(s) = 0$ (il faut s'assurer tout de même que du fait des approximations, nous obtenions de façon certaine une valeur approchée de s tel que $\eta_j^{n+1}(s) \leq 0$).

4.5 Récapitulatif de l'algorithme

L'algorithme complet peut s'écrire ainsi

Algorithme 9 *Algorithme complet.*
Tant que $t < T$
- *Pour chaque* j, *calculer* $\theta^n_{pred,j-\frac{1}{2}}$ *et* $\theta^n_{pred,j+\frac{1}{2}}$. *Fin pour.*
- *Pour chaque* j,
  - *poser* $s = 0$,
  - *calculer* $\phi^{n,+}_{j-\frac{1}{2}} = \phi(u^n_{j-1}, u^n_j, \theta^n_{pred,j-\frac{1}{2}})$ *et* $\phi^{n,-}_{j+\frac{1}{2}} = \phi(u^n_j, u^n_{j+1}, \theta^n_{pred,j+\frac{1}{2}})$,
  - *calculer* u^{n+1}_j,
  - *calculer* $\eta = \eta(u^{n+1}_j)$.
  - *Tant que* $\eta > \delta$,
    - *s=s+ds,*
    - *calcul de* $\theta^{n,+}_{j-\frac{1}{2}}(s) = \theta^n_{pred,j-\frac{1}{2}} + s(-1 - \theta^n_{pred,j-\frac{1}{2}})$,
    - *calcul de* $\theta^{n,-}_{j+\frac{1}{2}}(s) = \theta^n_{pred,j+\frac{1}{2}} + s(-1 - \theta^n_{pred,j+\frac{1}{2}})$,
    - *calcul de* $\phi^{n,+}_{j-\frac{1}{2}} = \phi(u^n_{j-1}, u^n_j, \theta^{n,+}_{j-\frac{1}{2}})$ *et* $\phi^{n,-}_{j+\frac{1}{2}} = \phi(u^n_j, u^n_{j+1}, \theta^{n,-}_{j+\frac{1}{2}})$,
    - *calcul de* u^{n+1}_j,
    - *calcul de* $\eta = \eta(u^{n+1}_j)$,
  - *fin tant que.*
- *Fin pour.*
- *Pour chaque* j,
  - *calcul de* $\theta^n_{j+\frac{1}{2}} = min(\theta^{n,-}_{j+\frac{1}{2}}, \theta^{n,+}_{j+\frac{1}{2}})$,
  - *calcul de* $\phi^n_{j-\frac{1}{2}} = \phi(u^n_{j-1}, u^n_j, \theta^n_{j-\frac{1}{2}})$ *et* $\phi^n_{j+\frac{1}{2}} = \phi(u^n_j, u^n_{j+1}, \theta^n_{j+\frac{1}{2}})$,
  - *calcul de* u^{n+1}_j,
- *fin pour.*

Fin tant que.

En réalité, il n'est pas nécessaire de recalculer u^{n+1}_j pour tous les indices j après le calibrage des paramètres θ avec l'entropie. En effet, seules les valeurs de la solution dans les cellules j où la dissipation d'entropie était strictement positive et leurs voisines directes $j+1$ et $j-1$ doivent être modifiées à cause de l'étape $min(\theta^{n,-}_{j+\frac{1}{2}}, \theta^{n,+}_{j+\frac{1}{2}})$. Les flux qui interviennent dans les autres cellules ne sont en effet pas modifiés.

4.6 Expériences numériques

4.6.1 Equation de Burger

Nous testons notre algorithme sur l'équation de Burger :

$$\partial_t u + \partial_x(\frac{u^2}{2}) = 0,\ x \in]0,1[,\ t > 0$$

avec comme condition initiale

$$u(x,0) = \begin{cases} 0 \text{ si } 0 < x \leq 0.2 \\ 1 \text{ si } 0.2 < x \leq 0.4 \\ 1 - 5(x - 0.4) \text{ si } 0.4 < x \leq 0.6 \\ 0 \text{ sinon,} \end{cases}$$

et comme condition aux bords

$$u(0,t) = 0,\ t > 0.$$

Nous représentons à gauche sur la figure 4.3 la solution numérique obtenue avec le schéma hybride sans correction d'entropie et à droite celle obtenue avec correction d'entropie à différents instants. Nous voyons que l'algorithme permet de rendre la dissipation d'entropie numérique négative, sans altérer la précision du résultat au niveau du choc.

4.6.2 Equation d'Euler 1d

Nous reprenons le test du paragraphe 3.2.2 pour lequel le schéma hybride fournissait une solution non entropique. Nous considérons les équations d'Euler. Nous rappelons les conditions initiales : $u_L = (\rho_L, \rho_L v_L, \rho_L E_L)$ et $u_R = (\rho_R, \rho_R v_R, \rho_R E_R)$ avec une interface à $x = 0.5$:

$$\begin{array}{ll} \rho_L = 5, & \rho_R = 0.125, \\ v_L = 0, & v_R = 0, \\ p_L = 5, & p_R = 0.1. \end{array} \tag{4.20}$$

Nous avons vu l'apparition d'un choc d'entropie à l'interface pour la solution numérique du schéma hybride (figure 3.39). En utilisant la méthode prédicteur-correcteur, nous arrivons à faire disparaître le choc d'entropie tout en gardant une bonne capture des chocs. La figure 4.4 montre les résultats pour un maillage de 200 points. Le flux d'entropie numérique utilisé est le flux de Tadmor, le schéma hybride est le schéma avec Lax-Friedrichs modifié pour $\theta = -1$, le nombre CFL est égal à 0.45.

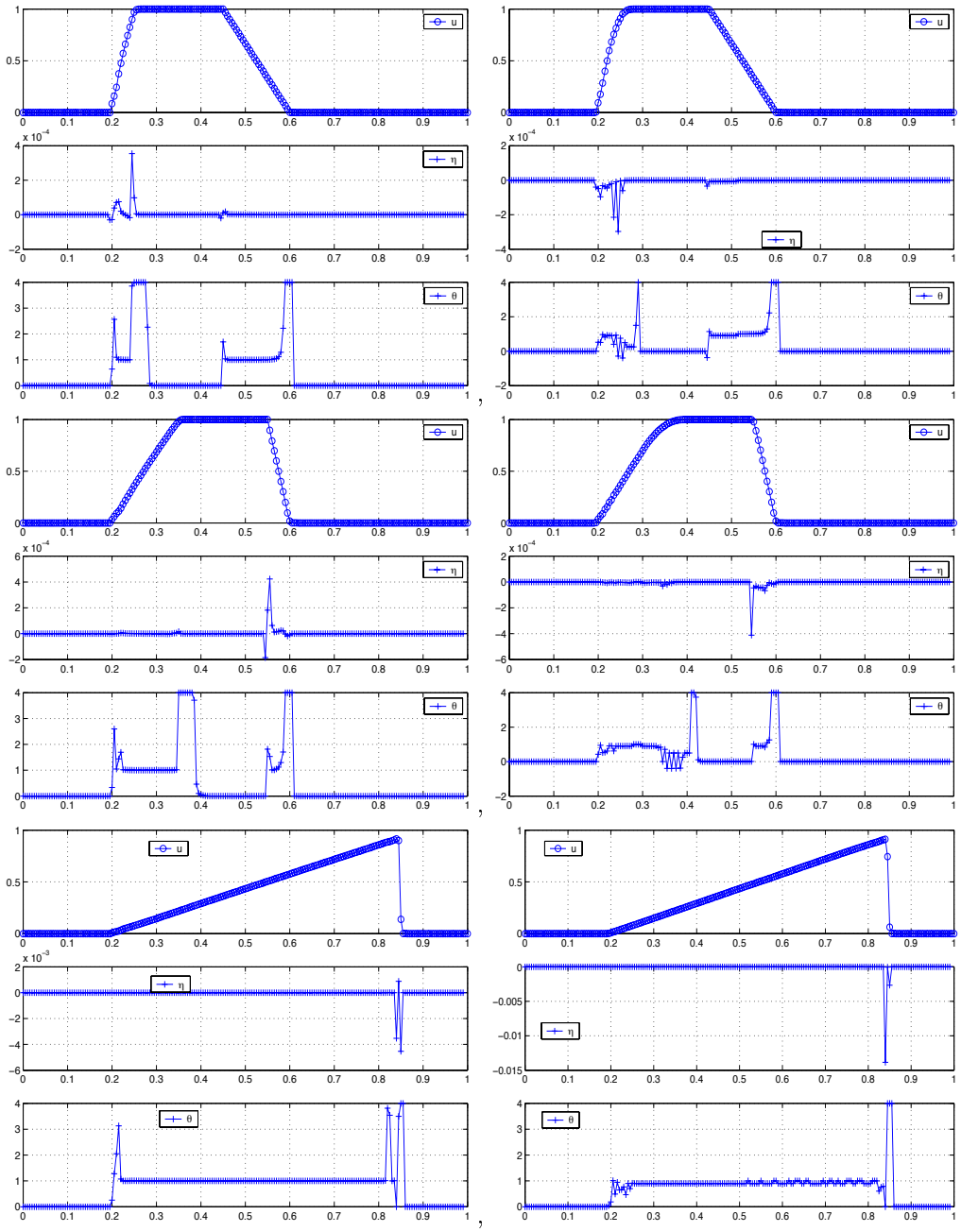

FIG. 4.3 – Schéma hybride pour l'équation de Burgers, $\nu = 0.5$, 200 points a) A gauche : schéma hybride sans correction d'entropie, $t = 0.05$, $t = 0.15$, $t = 0.7$; b) A droite : schéma hybride avec correction d'entropie (Lax-Friedrichs modifié avec flux d'entropie de Tadmor) $t = 0.05$, $t = 0.15$, $t = 0.7$.

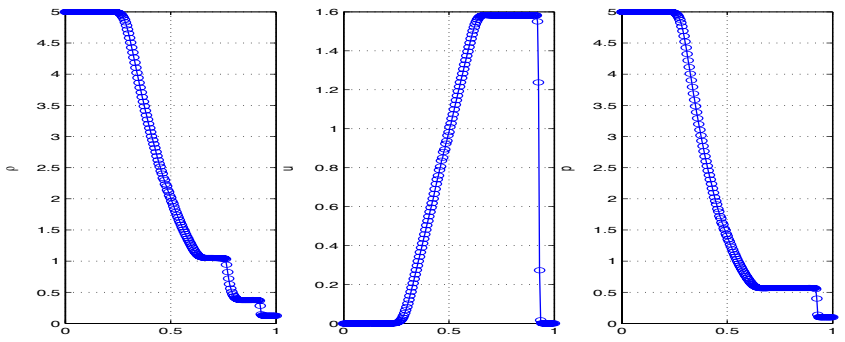

FIG. 4.4 – Schéma hybride (Lax-Friedrichs modifié avec flux d'entropie de Tadmor) pour les équations d'Euler avec une onde de raréfaction ; 200 points $\nu = 0.45$; $T = 0.18$.

Bibliographie

[1] Colella, P. Multidimensional upwind methods for hyperbolic conservation laws. *J. Comp. Phys. 87* (1990), 171–200.

[2] DeVuyst, F. Stable and accurate hybrid finite volume methods based on pure convexity arguments for hyperbolic systems of conservation law. *JCP 193* (2004), 426–468.

[3] DeVuyst, F., and Jaisson, P. A novel second order accurate hybrid numerical approach for the numerical solution of systems of conservation law. *En cours de rédaction*.

[4] Friedrichs, K. Symmetric hyperbolic linear differential equations. *Comm. Pure Appl. Math. 7* (1954), 345–392.

[5] Ghidaglia, J., Kumbaro, A., and Coq, G. L. Une méthode "volumes finis" à flux caractéristiques pour la résolution numérique des systèmes hyperboliques de lois de conservation. *C.R. Acad. Sci. Paris Sér I Math. 322*, 10 (1996), 981–988.

[6] Godlewski, E., and Raviart, P. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. New York, Springer, 1996.

[7] Harten, A. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics. 49* (mars 1983), 357–393.

[8] Harten, A. On the symmetric form of systems of conservation laws with entropy. *J. Comp. Phys. 49* (1983), 151–164.

[9] Jin, S., and Xin, Z. The relaxation schemes for systems of conservation laws in arbritary space dimensions. *Commun Pure Appl. Math. 45* (1995), 235–276.

[10] Lax, P. Weak solutions of non-linear hyperbolic equations and their numerical computations. *Comm. Pure Appl. Math. 7* (1954), 159–193.

[11] Lax, P. Hyperbolic systems of conservation laws and the mathematical theory of shock waves. *SIAM J. Numer. Anal. 43* (1972), 357–372.

[12] Lax, P., and Wendroff, B. Systems of conservation laws. *Commun. Pure Appl. Math. 13* (1960), 217–237.

[13] Lefloch, P., Mercier, J., and Rohde, C. Fully discrete, entropy conservative schemes of arbitrary order. *SIAM J. Numer. Anal. 40*, 5 (2002), 1968–1992.

[14] Lefloch, P., and Rohde, C. High-order schemes, entropy inequalities and nonclassical shocks. *SIAM J. Numer. Anal. 37*, 6 (2000), 2023–2060.

[15] Roe, P. Approximate riemann solvers, parameter vector and difference scheme. *Journal of Computational Physics 43* (1981), 357–372.

[16] Rusanov, V. Calculation of interaction of non-steady shock-waves with obstacles. *J. Comput. Math. Phys. 1* (1961), 267–279.

[17] Serre, D. *Système de Lois de Conservation I : hyperbolicité, entropies, ondes de choc.* Diderot Editeur, 1996.

[18] Sod, G. A survey of several finite difference methods for systems of non linear hyperbolic conservation laws. *J. Comp. Phys. 27* (1978), 1–31.

[19] Sweby, P. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM J. Numer. Anal. 21*, 5 (1984), 995–1011.

[20] Tadmor, E. Numerical viscosity and the entropy condition for conservative difference scheme. *Math. Comp. 43* (1984), 369–381.

[21] Tadmor, E. The numerical viscosity of entropy stable schemes for systems of conservation laws. *Acta Numerica 49*, 179 (1987), 91–103.

[22] Tadmor, E. Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependant problems. *Acta Numerica* (2004), 451–512.

Conclusion

Nous nous sommes intéressés dans cette thèse à différents problèmes d'optimisation faisant intervenir des systèmes hyperboliques d'EDP.

Flux d'informations. Nous avons modélisé les flux d'information pour des systèmes à tâches partagées à l'aide d'un modèle fluide d'EDP introduit par De Vuyst. Nous avons alors trouvé une méthode avec EDO pour calculer les temps de service des systèmes. Cela permettra de mieux comprendre le comportement des systèmes à tâches partagées comme certains serveurs web. Afin d'éprouver le modèle, nous avons proposé deux tests numériques sur la qualité de service et nous avons montré l'utilité du modèle.
Le calcul du temps de réponse est surtout important lorsque le système est congestionné. C'est justement dans ce cadre que la modélisation fluide par un système d'EDP-EDO est valable. Néanmoins, il serait intéressant dans un travail futur de prendre en compte les effets stochastiques et d'analyser la transition entre le domaine où la modélisation fluide est pertinente et le domaine où les effets stochastiques sont prépondérants.

Trafic routier. Ensuite, nous nous sommes intéressés au cas du trafic routier. Nous avons choisi le modèle du second ordre de Aw-Rascle. Il s'agissait d'utiliser ce modèle avec un procédé d'assimilation de données afin de prévoir les conditions de circulation futures. Le procédé d'assimilation de données est basé sur l'optimisation : nous avons retrouvé la solution numérique du système de Aw-Rascle qui minimise l'erreur par rapport aux observations. Nous avons présenté deux stratégies différentes afin de trouver cette solution minimisante. Dans la première stratégie, nous avons choisi comme variables d'optimisation les conditions initiales et les conditions aux bords. Dans la deuxième stratégie, nous avons uniquement recherché les conditions initiales, mais sur une portion de route plus étendue. Nous avons utilisé une méthode adjointe pour résoudre le problème d'optimisation.
Dans un travail futur, nous désirons utiliser cette méthode avec des données réelles. Notamment, il faudrait réfléchir à la méthode d'interpolation des données discrètes afin d'obtenir des données définies sur tout le domaine. Enfin, nous pourrions appliquer notre travail à des extensions du modèle de Aw-Rascle.

Schémas hybrides. Lors des tests numériques concernant le trafic routier, nous avons rencontré des solutions numériques non physiques. Nous nous sommes alors naturellement intéressés à l'analyse des schémas numériques permettant d'approcher les solutions physiques des systèmes hyperboliques d'EDP. Nous avons exhibé un schéma hybride dépendant d'un paramètre θ. Ce paramètre est ajusté afin d'obtenir un schéma possédant la propriété TVD (phase prédictrice de l'algorithme). Ensuite le calcul de la dissipation/production numérique d'entropie dans chaque cellule permet de corriger le paramètre afin d'augmenter l'intensité de la matrice de diffusion (phase correctrice de l'algorithme). Nous avons obtenu des résultats numériques convaincants.
Trois pistes de recherche future semblent prometteuses. Premièrement, nous avons utilisé le flux de Tadmor pour le calcul de la dissipation/production d'entropie. L'analyse complète permettant de montrer qu'un schéma donné (par exemple, le schéma de Lax-Friedrichs modifié) fournit une solution entropique est difficile. Il serait intéressant de trouver une expression numérique différente de la dissipation/production d'entropie plus adaptée à notre schéma. Deuxièmement, nous avons remarqué qu'au voisinage d'un point sonique, l'intensité de la matrice de diffusion était identique à celle du schéma de Roe (à l'exception de la valeur donnant le schéma de Lax-

Friedrichs modifié). Ainsi, nous ne pouvons pas espérer obtenir l'utilisation par notre algorithme d'un schéma moins diffusif que le schéma de Lax-Friedrichs modifié, au voisinage d'un point sonique. Il est donc nécessaire de trouver une autre méthode pour la phase correctrice de notre algorithme. Enfin, il serait très intéressant d'éviter une méthode itérative de la phase correctrice. La solution apportée à ces trois problèmes devrait permettre de rendre plus performant l'algorithme prédicteur-correcteur.

Résumé

Cette thèse concerne la modélisation par des EDP et la résolution numérique de problèmes d'optimisation pour les flux d'informations et pour le trafic routier. Nous proposons un nouveau schéma hybride. En premier, nous nous intéressons à une modélisation fluide proposée par De Vuyst, d'un système multi-tâches devant traiter plusieurs types de requêtes. Nous exhibons un système d'EDO permettant de calculer les temps de service. Nous résolvons numériquement deux problèmes de contrôle optimal pour garantir une qualité de service. Ensuite, nous traitons un problème d'assimilation de données en trafic routier et nous donnons un algorithme capable de prévoir les conditions futures de circulation sur une section de route. Le flux routier est modélisé par le système hyperbolique de Aw-Rascle. Nous devons minimiser une fonctionnelle dont les variables d'optimisation sont les conditions initiales et/ou les conditions aux bords. Nous obtenons la solution du système de Aw-Rascle par la méthode de Roe et nous calculons le gradient de la fonctionnelle par une méthode adjointe. Enfin, nous présentons un nouveau schéma hybride à un paramètre qui permet de calculer numériquement les solutions des systèmes hyperboliques. Ce schéma possède la propriété TVD. Il est du second ordre en espace et en temps. Après une première phase prédictrice, nous pouvons corriger le paramètre dans les cellules où il y a production d'entropie numérique. Nous obtenons ainsi un schéma qui capture la solution physique.

Mots-clés : EDP, flux d'information, loi de conservation, système à tâches partagées, trafic routier, assimilation de données, modèle de Aw-Rascle, schémas numériques, entropie, modélisation fluide.

Abstract

This thesis deals with PDE modeling and numerical resolution of optimisation problems for multithread system and traffic flow. We propose a new hybrid scheme. First, we are interesting by fluid models of a multithread/multitask system proposed by De Vuyst. We find 0DEs which are used for the computation of the service times. We numerically solve two problem of optimal control of quality of service (QoS) management. Then we deal with traffic data assimilation and algorithms able to predict the traffic flow on a road section. The traffic flow is modelized by the Aw-Rascle hyperbolic system. We have to minimize a functional whose optimization variables are initial condition and/or upstream boundary conditions. We use the Roe method to compute the solution of the traffic flow modelling system. Then we compute the gradient of the functional by an adjoint method. This gradient will be used to optimize the functional. Last, we propose a new hybrid scheme with one parameter which permit the scheme to have the TVD property and the space and time second order accuracy. After a first predictor step, we can correct the parameter in the cells where the entropy production is positive. Thus, the scheme can capture the physical solution.

Key-words : PDE, conservation law, multithread system, traffic flow, data assimilation, Aw-Rascle model, numerical scheme, entropy, fluid models.