THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *05/07/2021* **par :**

Mathias Dus

Problèmes de stabilisation au bord pour des systèmes d'équations aux dérivées partielles hyperboliques en dimension un d'espace

JURY

Emmanuel Trélat	Sorbonne Université	Rapporteur
Julie Valein	Institut Elie Cartan de Lorraine	Rapportrice
Lucie Baudouin	LAAS - CNRS	Examinatrice
Sylvain Ervedoza	Université de Bordeaux	Examinateur
Christophe Prieur	Université Grenobles Alpes	Examinateur
Franck Boyer	Université Toulouse 3	Directeur de thèse
Francesco Ferrante	Université Grenobles Alpes	Directeur de thèse

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse (UMR 5219)
Directeurs de Thèse :
Franck Boyer et *Francesco Ferrante*
Rapporteurs :
Emmanuel Trélat et *Julie Valein*

Remerciements

Mes premiers remerciements s'adressent tout naturellement à Franck Boyer et Francesco Ferrante pour leur expertise en théorie du contrôle, en automatique et en analyse numérique, bien sûr, mais aussi pour leur sympathie, leur humour et leur soutien dans les moments difficiles. Mais surtout, merci à ces tuteurs pour m'avoir éduqué et corrigé en faisant preuve d'une patience indéfectible mêlée à une grande exigence, et cela, toujours avec un regard bienveillant sur mon travail.

Je suis très reconnaissant et honoré que Emmanuel Trélat et Julie Valein aient accepté de lire et d'évaluer mon travail de recherche. Par ailleurs, je remercie chaleureusement Sylvain Ervedoza, Lucie Baudouin et Christophe Prieur d'avoir accepté de compléter mon jury.

Je tiens à remercier particulièrement Christophe Prieur qui, lors de mon stage de M2, m'a fait découvrir le sujet passionnant de l'automatique appliqué aux EDPs. Malgré la distance, ses corrections ainsi que nos discussions, ont contribué à faire le scientifique que je suis aujourd'hui.

Mes années au sein de l'IMT resteront marquées en moi, en partie grâce à la vie que ce laboratoire porte en lui. Ainsi, j'aimerais remercier les quelques personnes que j'ai pu rencontrer en ces murs.

Je commence par remercier quelques membres permanents du laboratoire. Tout d'abord, Marie Hélène Vignal, pour son accessibilité et ses précieux conseils, qui permettent aux jeunes chercheurs comme moi de comprendre les codes du milieu académique. Merci à Fanny Delebecque pour son implication et sa passion de l'enseignement, qui m'a permis de travailler dans les meilleurs conditions et ce, malgré la crise du COVID. Merci à Luca Amodei pour sa disponibilité et son sens de l'organisation, qui permettent aux chargés de TD de faire leur travail sereinement. Je voudrais particulièrement remercier Sylvain Ervedoza qui, avec son accessibilité et sa maîtrise impressionnante de la théorie du contrôle, m'a permis d'avancer à pas de géant dans l'un de mes travaux. Merci à Sepideh Mirrahimi de m'ouvrir à de nouveaux domaines grâce à son cours très intéressant sur les équations de Hamilton-Jacobi pour la biologie. Merci à Sophie Jan pour son travail remarquable au sein de l'AMIES, qui permet de faire mieux comprendre aux doctorants leur avenir professionnel au sein d'un organisme privé. Elle contribue à une véritable ouverture de la force de recherche toulousaine, pour l'inclure dans un monde concret et appliqué. Enfin, merci à Denis Matignon et Ghislain Haine pour m'avoir donné envie de faire une thèse et pour m'avoir donné le goût de la recherche scientifique.

L'administration universitaire, parfois lourde et absurde survit grâce à l'engagement de son personnel. Je tiens ici à remercier ces personnes dont le travail permet au laboratoire de fonctionner correctement. Tout d'abord Monique et Marie-Line pour toutes ces conversations durant ces trois années, qui vont beaucoup me manquer. Je salue également leur travail quotidien au service du laboratoire. Par ailleurs, je remercie grandement Nicole Lhermitte, Janani Chandran, Tamara Azaiez, Zohra Kallel pour leurs soutiens dans mes tâches administratives. Je tiens particulièrement à remercier Agnès Requis pour son écoute et ses conseils qui m'ont été très précieux et réconfortants.

Les doctorants sont le cœur battant du labo, une deuxième famille qui m'a permis de surmonter l'épreuve de la thèse.

Je tiens d'abord évidemment à remercier les éléments plus que prometteurs du bureau 302. Baptiste Fedele, mon grand frère (pas si grand) éphémère de thèse, avec qui j'ai noué des liens d'amitié forts. Sa patience, sa résilience, son dévouement et son humour en font une personne exceptionnelle. Merci à Joachim Crevat pour nos délires journaliers et nos fou rires quotidiens. Nos citations récurrentes de OSS 117 [1] ainsi que nos cris d'animaux sauvages (je m'excuse auprès de Jean Michel Roquejoffre pour le bruit) ont illuminé mes journées. Merci à Alexis Leculier

1. "Huuuuum, ça me sert à rien ..."

pour ta bonne humeur, ta bienveillance et surtout pour tes chaussons d'escalades après deux heures de session. On a vraiment bien rigolé avec nos blagues (parfois douteuses), nos interjections absurdes ("PARCOURT") et les quelques soirées arrosées qu'on a fait ensemble. Merci à Louis pour sa patience envers moi (et il en faut) et sa pédagogie pour m'expliquer les nombreux domaines mathématiques qui m'étaient alors inconnus. Ton travail fut une grande source d'inspiration et cela a même abouti à un article. Merci à la petite dernière Viviana, qui s'est très bien intégrée à l'élite du bureau 302. Je te souhaite une expérience de thèse, riche en rebondissements scientifiques avec des rencontres qui te permettront d'avancer loin.

Michelle, this is your paragraph ! On ne peut pas parler du bureau des légendes (302) sans parler de son personnage emblématique : Michelle Romanos. Un grand merci à toi pour ta bonne humeur, ton INCROYABLE énergie et ta force de caractère. Tes phrases mythiques ont illuminées la vie du bureau : "I am climbing the ladder of sexism", "Everybody wants to be Michelle", "Kapp Pralib"... Au delà de ça, tu as su m'écouter et me donner de très bons conseils quand j'en avais le plus besoin. Ce fut une très belle rencontre et j'espère pouvoir garder contact encore longtemps.

Je voudrais remercier les autres doctorants du bâtiment 1R3. Tout particulièrement, Corentin Kilque et sa clique, Perla Kfoury et Paola Allegrini. Votre bonne humeur et vos soirées permettent de rendre plus douce l'épreuve de la thèse. Merci à mon grand frère de thèse, Kuntal Bhandari pour ses conseils précieux. Merci à Joe Alhelou de supporter mes blagues et de partager avec moi, nos galères de thésard. Enfin, je voulais remercier le reste des doctorants du bâtiment 1R3 : Anthony Mur, François L'official, Alain Blaustein, Susely Figueroa...

L'IMT est vaste et au delà du bâtiment 1R3 "maths appli", j'aimerais rendre hommage au reste de l'institut qui travaille sur des sujets obscures et maléfiques.

J'aimerais remercier les doctorants stat/proba en commençant par Léaticia Colombani et Eva Lawrence, qui se sont consacrées à l'intérêt des doctorants au sein des instances du laboratoire et de celles de l'école doctorale. Merci à Pierre Gach, camarade de prépa, avec qui j'ai partagé mes galères. Merci à Maylis Varvenne, Fabien Montégut, Mickael Albertus, Clément Benesse, Clément Steiner ainsi que les petits nouveaux : Luca de Lara, Virgile Foy, Javi González Delgado, Alberto González Sanz...

J'adresse un hommage à l'équipe des Picards et je commence par Florian Bertuol, toujours disponible pour m'expliquer la folie des mathématiques pures. Passionné et toujours à l'écoute, je suis profondément convaincu que tu brilleras dans l'enseignement. Son dévouement envers les doctorants et les élèves de la faculté est remarquable et je te souhaite le meilleur pour la suite. Merci à Jean Marc Hok pour son courage, sa gentillesse et son implication dans le monde pédagogique. Je salue également Dominique, Dimitri, Elena et Sara.

Mes amis ont également été d'un très grand soutien. Il y a bien sûr Rémi Adon, grand poète et informaticien à ses heures perdues. Il y a aussi Laurent François, mon développeur binôme de Supaéro, homme à tout faire au grand coeur. On n'oubliera pas toute l'équipe de ma ville natale, Agen : Benjamin, Thomas, Victor, Vincent, Nicolas, Emmanuel, Antoine...

J'aimerais pour achever mes remerciements partager mes dernières pensées avec ma famille qui fut d'une aide immense.

Mes sentiments très affectueux s'adressent tout d'abord à ma maman pour son soutien inconditionnel. Quels que soient mes choix, tu as toujours été un roc, un appui et je sais que je peux toujours compter sur toi.

Dans cette thèse, il y a eu un avant et un après qui coïncide avec la rencontre de mon amour véritable, Maeva. Tu as été d'une grande patience, d'un grand soutien et tu m'as donné la force d'aller jusqu'au bout. Je salue également toute ma belle famille, Marc et Fabienne.

Résumé

Dans cette thèse, nous étudions le problème de stabilisation au bord de systèmes généraux d'équations aux dérivées partielles hyperboliques. Plus précisément, l'étude se focalise sur des systèmes où le transport est uniquement scalaire et où le sens de propagation de l'information est fixé. En outre, le contrôle choisi sera la plupart du temps sous la forme d'une loi de retour d'état (ou feedback) linéaire que l'on perturbera éventuellement par l'effet d'une saturation. Le travail est séparé en deux parties bien distinctes ; l'une se concentre sur des méthodes de Lyapunov, tandis que l'autre va plutôt utiliser des techniques propres au linéaire.

Pour la première partie, deux travaux principaux sont présentés. Dans un premier temps, nous ne considérons que des équations de transport linéaires à vitesses positives et cherchons à stabiliser exponentiellement le système dans L^∞ grâce à un feedback linéaire saturé. La méthode consiste à utiliser des techniques classiques de Lyapunov afin d'exhiber un bassin d'attraction et d'en donner une estimation fine. On généralise ensuite ce travail dans un cadre BV pour les systèmes de lois de conservation scalaires couplées au bord. Secondement, un système de lois de conservation scalaires à vitesses positives est discrétisé en utilisant un schéma à limiteur de pente. En s'inspirant des méthodes issues du cadre continu, une fonctionnelle de Lyapunov discrète est étudiée pour prouver la stabilisation exponentielle BV par feedback linéaire de la solution discrète.

Pour la seconde partie, deux études sont également exposées mais cette fois-ci, dans un cadre totalement linéaire. D'une part, il s'agit d'établir la possibilité de construire un feedback issu d'un placement de pôles pour stabiliser exponentiellement des edps hyperboliques linéaires avec couplage au bord et dans le domaine. D'autre part, nous développons une théorie du backstepping discrétisé pour stabiliser en temps fini un schéma numérique modélisant un système 2×2 avec couplage au bord et au sein du domaine.

Mots-Clés

Stabilisation, équations de transport, Backstepping, saturation, schémas numériques

Abstract

In this thesis, we study the problem of boundary stabilization of general hyperbolic systems of partial differential equations. More precisely, the analysis focuses on systems where the transport term is scalar and for which the information propagates in a fixed direction. In addition, the chosen control is most of the time a state feedback law for which a saturation is possibly applied. The work is divided into two distinct parts, one focusing on Lyapunov techniques while the other one uses the linearity of the problem.

In the first part of the thesis, two main works are presented. In the first one, only linear transport equations with positive velocities are considered. The main goal is to design a saturated linear feedback in order to stabilize exponentially the open-loop system in L^∞. The method consists of using classical Lyapunov techniques to exhibit a basin of attraction for which a fine estimate is given. We also extend this work to nonlinear scalar conservation laws in a BV framework.
In the other work, thanks to a slope limiter scheme, a system of scalar conservation laws is discretized. Inspired by "continuous" Lyapunov methods, a discrete Lyapunov functional is studied to prove the exponential BV stabilization of the discrete solution using a linear feedback.

In the second part of the thesis, two works are exposed as well, this time in a full linear framework. On the one hand, we study systems of linear transport equations of arbitrary dimension, coupled on the domain and at the boundary. Designing a controller from a pole placement algorithm, the exponential stabilization is proved in L^2. On the other hand, we develop a numerical Backstepping theory in order to stabilize in finite time a numerical scheme modeling a 2×2 linear system with in domain and boundary couplings.

Keywords

Stabilization, transport equations, Backstepping, saturation, numerical schemes

TABLE DES MATIÈRES

CHAPITRE 1

INTRODUCTION

1.1 Notions de contrôle

Pour introduire certaines notions de contrôle, considérons une simple équation différentielle ordinaire (EDO) :

$$\begin{cases} \dot{X} = & a(X) \\ X(0) = & X_0 \in \mathbb{R}^q \end{cases} \tag{1.1}$$

où $q \in \mathbb{N}$, $a : \mathbb{R}^q \to \mathbb{R}^q$ et X_0 est une donnée initiale choisie dans $\mathbb{R}^q$. En théorie du contrôle on va chercher à modifier la dynamique (1.1) de X pour que celle-ci corresponde à une trajectoire ou à une cible voulue. Ainsi, on introduit un opérateur b d'actuation et un contrôle $u(t) \in \mathbb{R}^m$ pour contrôler (1.1).

$$\begin{cases} \dot{X} = & a(X) + b(u(t)) \\ X(0) = & X_0 \in \mathbb{R}^q \end{cases} \tag{1.2}$$

La fonction $b : \mathbb{R}^m \to \mathbb{R}^q$ permet de modéliser l'action que le contrôle $u(t)$ peut avoir sur l'état X. On peut distinguer trois branches principales dans la théorie du contrôle :

— La contrôlabilité où on prend un temps final T et une cible X_1 et on regarde s'il est possible de trouver un contrôle tel que $X(T) = X_1$. Pour ce genre de problème, le contrôle dépend uniquement de l'état initial X_0.

— Le contrôle optimal où on impose un critère à minimiser et le problème sera de trouver un contrôle u qui minimise ce critère.

— Enfin, la stabilisation qui est la branche qui va le plus nous intéresser dans cette thèse. Lorsque l'on cherche à stabiliser un système, on considère un état X_{eq} d'équilibre du système (1.1) vérifiant $a(X_{eq}) = 0$ et on essaye de faire converger la solution $X(t)$ vers cet état. Pour ce faire, on impose une certaine manière d'écrire le contrôle. Celui-ci doit dépendre de l'état $X(t)$ à l'instant t :

$$u(t) = k(X(t)) \tag{1.3}$$

où k est une fonction bien choisie telle que :

$$\lim_{x \to \infty} X(t) = X_{eq}.$$

Il semblerait a priori qu'imposer la forme de contrôle (1.3) soit assez limitatif, cependant il s'avère que ce choix est très robuste vis à vis des perturbations extérieures. En effet, le contrôle dépendant de l'état à l'instant t, celui-ci peut s'adapter à l'éventualité d'une perturbation de la dynamique du système. De part cette robustesse, les méthodes de stabilisation par feedback sont très utilisées dans l'industrie.

Pour ce qui concerne les méthodes de contrôlabilité, le contrôle ne dépendant que de la condition initiale X_0, il ne pourra pas s'adapter à une éventuelle perturbation et l'état cible X_1 sera le plus souvent non atteint. On parle de manque de robustesse de la méthode par rapport à une perturbation. C'est d'autant plus flagrant lorsque le système étudié est non linéaire et chaotique comme celui du double pendule par exemple.

Ainsi et avant de continuer, il est nécessaire d'introduire le vocabulaire des automaticiens [45, Chapitre 1], ce dernier étant nécessaire afin de continuer notre présentation.

Definition 1. *On appellera X l'état du système. La variable $u(t)$ correspond au contrôle. Le système sans contrôle $u(t) = 0$ est ce qu'on appelle le système en boucle ouverte alors que celui avec contrôle de type $u(t) = k(X(t))$ (1.2) est le système en boucle fermé. On appellera un contrôle de type $u(t) = k(X(t))$ un contrôle par retour d'état ou un feedback.*

Ces notations proviennent de la vision du système sous forme de schémas blocs que l'on rappelle ici pour la boucle ouverte.

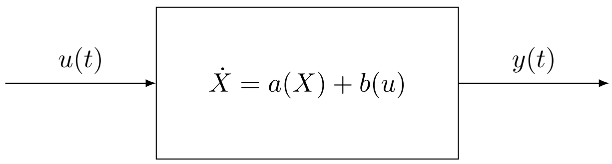


FIGURE 1.1 – Un schéma bloc en boucle ouverte

La variable $y(t) \in \mathbb{R}^p$ correspond aux observations données par les capteurs qui nous informent sur l'état X. En automatique linéaire, on a $y = CX$ où $C \in M_{pq}(\mathbb{R})$. Dans les cas les plus généraux, C n'est pas inversible et on n'a qu'une observation partielle de l'état.

Dans ce travail, on fera l'hypothèse que les capteurs permettent de mesurer l'état X avec une infinie précision, pour ne se focaliser uniquement que sur l'effet du contrôle. On prendra donc $y = X$. En boucle fermée, on aura le schéma bloc suivant :

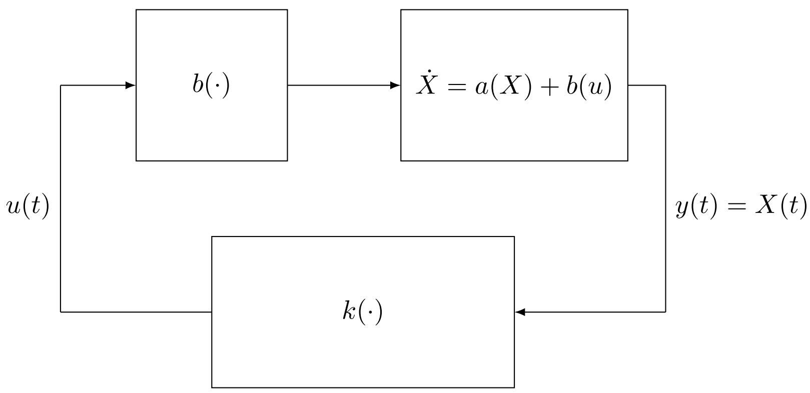


FIGURE 1.2 – Un schéma bloc en boucle fermée

Pour les notions de stabilité, on introduit les définitions suivantes en supposant que 0 est l'unique état d'équilibre du système ($a(0) = 0$) [29, Theorem 10.8] :

Definition 2.
— *Le point d'équilibre 0 ou le système* (1.2) *est dit stable (en 0) si pour tout* $\varepsilon > 0$, *il existe un* $\eta > 0$ *tel que si* $|X_0| \leq \eta$ *alors pour tout* $t \geq 0$, $|X(t)| \leq \varepsilon$.
— *Le point d'équilibre 0 ou le système* (1.2) *est globalement asymptotiquement stable si pour toute condition initiale* X_0,

$$\lim_{t \to +\infty} |X(t)| = 0.$$

— *Le point d'équilibre 0 ou le système* (1.2) *est globalement exponentiellement stable s'il existe* $\nu, C > 0$ *tels que pour toute condition initiale* X_0,

$$\forall t \geq 0,\ |X(t)| \leq Ce^{-\nu t}|X_0|.$$

— *Le point d'équilibre 0 ou le système* (1.2) *est dit localement asymptotiquement stable s'il existe* $\eta > 0$ *tel que pour toute condition initiale* X_0 *vérifiant* $|X_0| \leq \eta$*, la solution* $X(t)$ *tend vers 0. On définira de la même manière "localement exponentiellement stable".*

Ce vocabulaire est très classique chez les automaticiens pour caractériser la stabilité du système étudié. Il sera utilisé tout au long de la thèse.

1.2 Les équations de transport

Il ne s'agit pas d'approfondir la stabilisation de systèmes de dimension finie (EDO) qu'étudient les automaticiens, mais plutôt de se concentrer sur un type particulier de systèmes où l'état évolue dans des espaces de dimension infinie : les Équations aux Dérivées Partielles (EDP) hyperboliques. Plus particulièrement, on introduit une variable 1D d'espace x que l'on va pour l'instant supposer dans $\mathbb{R}$. Notre état sera noté R qui est une fonction de t et de x à valeur dans $\mathbb{R}^q$.

Ainsi l'état R évolue dans un espace de dimension infinie qui sera le plus souvent un espace de Lebesgue $L^p(\mathbb{R};\mathbb{R}^q)$. Dans cette thèse, l'état $R(t,x)$ en boucle ouverte vérifiera une équation de transport du type :

$$\partial_t R + \partial_x[f(R)] = S(R).$$

Cette equation permet de modéliser l'ensemble des phénomènes de transport. On peut l'écrire de la manière suivante si l'on suppose que le flux f et la solution R sont suffisamment réguliers :

$$\partial_t R + f'(R)\partial_x R = S(R). \tag{1.4}$$

On supposera également que $f'(R) \in M_q(\mathbb{R})$ est diagonalisable dans $\mathbb{R}$ quelque soit $R \in \mathbb{R}^q$ et que les valeurs propres de $f'(R)$ ne s'annulent pas.

1.2.1 La forme caractéristique

On introduit ici l'hypothèse d'écriture sous forme caractéristique [16, Partie 1.1] :

Hypothèse 1. *Il existe un changement de variable régulier* $U = \psi(R)$ *et une fonction* $\tilde{\Lambda} : \mathbb{R}^q \to D_q(\mathbb{R})$ *à valeur dans les matrices diagonales carrées de taille q tels que :*

$$\forall R \in \mathbb{R}^q,\ \ \psi(R)'f'(R) = \tilde{\Lambda}(R)\psi(R)'. \tag{1.5}$$

La variable U *se notera invariant de Riemann et l'on pourra réécrire* (1.4) *sous la forme caractéristique :*

$$\partial_t U + \tilde{\Lambda}(\psi^{-1}(U))\partial_x U = C(U).$$

Par la suite et pour ne pas introduire des notations superflues, on prendra $U \leftarrow R$ et $C(U) \leftarrow S(R)$. La forme caractéristique s'écrira :

$$\partial_t R + \Lambda(R)\partial_x R = S(R) \tag{1.6}$$

avec $\Lambda(R) = \tilde{\Lambda}(\psi^{-1}(R))$.

Il n'est pas évident que l'Hypothèse 1 soit satisfaite. Pour la prouver il faut et il suffit d'expliciter un changement de variable $U = \psi(R)$ tel que (1.5) soit vérifiée. Cela revient à trouver les solutions à des EDOs d'ordre un. Il a été montré dans [78, p.34-35] que cette équation peut toujours être résolue au moins localement pour $q \leq 2$. A contrario, pour des systèmes de plus grandes dimensions, il y a des cas où un tel changement de variable n'existe pas [36, p. 240]. Heureusement, ce sont des cas qui sortent du cadre de cette thèse.

1.2.2 Le cas particulier des équations scalaires

Un cas particulier va nous intéresser : il s'agit de la situation où $q = 1$. Ici, le flux f sera une fonction de $\mathbb{R}$ dans $\mathbb{R}$. Ce cadre comprend notamment les équations de transport linéaires où $f(R)$ est un opérateur de la forme λR avec λ un réel. Ce genre de modèle donne des modélisations très simples pour de nombreux phénomènes physiques. On pourra toujours écrire le système sous la forme suivante :

$$\partial_t R + \partial_x [V(R)R] = S(R)$$

où $V : \mathbb{R} \to \mathbb{R}$. Ci-dessous nous présentons une liste non exhaustive [16, Chapter 1] de modèles faisant intervenir des lois de conservation scalaires :

— Le modèle de Kac-Goldstein [61]-[69] à deux équations, pour la chimiotaxie qui modélise le mouvement de cellules selon le milieu chimique dans lequel elles se trouvent. On aura deux lois de conservation scalaires qui interagissent entre elles par un terme d'ordre zéro :

$$\left\{ \begin{array}{lcl} \partial_t \rho^+ + \gamma \partial_x \rho^+ + \mu(\rho^+, \rho^-)(\rho^- - \rho^+) & = & 0 \\ \partial_t \rho^- - \gamma \partial_x \rho^- + \mu(\rho^+, \rho^-)(\rho^+ - \rho^-) & = & 0 \end{array} \right. \tag{1.7}$$

Ici ρ^+ désigne la densité de cellules se déplaçant vers la droite ($\gamma > 0$) alors que ρ^- correspond aux cellules se déplaçant vers la gauche. Le terme μ est un terme d'échange entre les deux types de cellules. Cette équation est un bel exemple de système d'EDPs hyperboliques semilinéaires où le transport est linéaire alors que le terme d'ordre zero ne l'est pas.

— Les modèles de trafic routier à une route où la variable $R = \rho$ représente la densité de véhicules et $V(\rho)$ la vitesse associée. Le modèle de Garavello et Picolli [55, Chapter 3] donne une vitesse linéairement décroissante en fonction de la densité :

$$V(\rho) = V_{\max} \left(1 - \frac{\rho}{\rho_{\max}} \right).$$

On retrouve ainsi le modèle de Lighthill et Whitham [83] pour le trafic de véhicules routiers.

— Un modèle très simple de fluide à deux phases comme l'eau et l'huile, est donné par une loi de conservations scalaire. Ici $R = \rho$ correspond à la saturation d'eau. Le cas $\rho = 0$ équivaut à de l'huile pure tandis que $\rho = 1$ équivaut à de l'eau pure. Dans ce cas, on prendra une vitesse caractéristique de la forme [23] :

$$V(\rho) = \frac{\rho^2}{\rho^2 + (1-\rho)^2}.$$

— Enfin, en dernier exemple, on peut citer les modèles scalaires décrivant les mouvements d'eau au sein de canaux. Ici la variable d'état $R = H$ représente la hauteur d'eau au sein du canal. Afin de présenter le modèle, on rappelle les equations de Saint-Venant :

$$\left\{ \begin{array}{lcl} \partial_t H + \partial_x (HV) & = & 0 \\ \partial_t V + \partial_x (\frac{V^2}{2} + gH) + (C\frac{V^2}{H} - gS_b) & = & 0. \end{array} \right. \tag{1.8}$$

Ici, on a un système hyperbolique avec deux variables ($q = 2$) H et V, qui n'est donc pas scalaire. Le paramètre physique g est la constante de gravité, C le coefficient de viscosité et S_b correspond à l'inclinaison verticale du canal. On considère le régime physique suivant proposé dans [16, p. 44] :

$$C\frac{V^2}{H} - gS_b = 0$$

qui désigne un régime où la friction compense la gravité. Dans ce cas, on peut exprimer V en fonction de H et $k := \frac{S_b}{C}$ pour obtenir :

$$\partial_t H + \partial_x \left(kH\sqrt{\frac{H}{g}} \right) = 0$$

qui est une équation scalaire.

Pour les modèles qui présentent un couplage dans le terme de flux ($q > 1$), le problème d'existence et d'unicité de la solution est très difficile et n'est, à ce jour, pas complètement résolu. Souvent, il faut faire des hypothèses de petitesse sur la condition initiale pour avoir l'existence d'une solution. Les modèles scalaires sont, quant à eux, très accommodants puisqu'ils permettent d'avoir des solutions sans faire cette hypothèse de petitesse de la condition initiale. Cela nous permettra de faire des estimations fines de bassin d'attraction que l'on abordera en chapitres 2 et 3.

1.2.3 Les systèmes de lois de conservation scalaire

Considérons toujours le cas scalaire $q = 1$. Dans cette thèse, on se placera non pas dans la droite réelle entière, mais dans un domaine borné : le segment $[0, 1]$. On considère ainsi le système suivant :

$$\partial_t R + \lambda(R)\partial_x R = S(R).$$

où on suppose que la vitesse de propagation possède un signe constant :

Hypothèse 2.

$$\forall R \in \mathbb{R},\ \lambda(R) > 0 \ \textit{ou} \ \lambda(R) < 0.$$

Si $\lambda(R) > 0$, alors on doit spécifier la valeur de R en $x = 0$. Si au contraire $\lambda(R) < 0$, il faudra le faire en $x = 1$. On notera R_{in} cette valeur. Quant à R_{out}, cette notation désignera l'information qui sort du domaine. Admettons que l'on veuille modéliser un réseau avec $d \in \mathbb{N}$ equations scalaires qui interagissent entre elles par le bord et par un terme d'ordre zéro. Dans ce cas, on écrira le système de la manière suivante :

$$\forall 1 \leq i \leq d, \begin{cases} \partial_t R_i + \lambda_i(R_i)\partial_x R_i &= S_i(R_1, \cdots, R_q) \\ R_{in,i} &= h_i(R_{out}) + u_i(t) \end{cases}$$

où h_i est une fonction qui couple les equations au bord alors que $u_i(t)$ modélise un éventuel input extérieur. On peut également généraliser pour $q > 1$. Dans ce cadre là on se placera souvent en transport linéaire :

$$\forall 1 \leq i \leq d, \begin{cases} \partial_t R_i + \Lambda_i\partial_x R_i &= S_i(R_1, \cdots, R_q) \\ R_{in,i} &= h_i(R_{out}) + u_i(t) \end{cases}$$

où ici R_i, u_i sont des vecteurs de taille q. On peut par exemple imaginer un réseau de systèmes de Kac-Goldstein (1.7) où $q = 2$.

Remarque 1. Si on sort du cas scalaire et que l'on considère un réseau $d > 1$ de système $q > 1$ d'EDPs hyperboliques non linéaires, il va falloir admettre une hypothèse importante sur la matrice des vitesses $\Lambda(R)$:

Hypothèse 3. *Il existe* $1 \leq q_0 \leq q$ *tel que :*

$$\forall R,\ \lambda_1(R) \leq \lambda_2(R) \leq \cdots \leq \lambda_{q_0}(R) < 0 < \lambda_{q_0+1}(R) \leq \cdots \leq \lambda_q(R).$$

Cette hypothèse permet de fixer le sens du transport et ainsi (R_{in}, R_{out}) est bien défini. Pour illustrer cela on peut prendre le système de Saint Venant (1.8). On obtient alors :

$$R = (H, V),\ f(H, V) = (HV, \frac{V^2}{2} + gH)$$

et donc :

$$\Lambda(H, V) = \begin{pmatrix} V & H \\ g & V \end{pmatrix}.$$

Après de simples calculs, on obtient les valeurs propres de $\Lambda(H, V)$ ainsi que les invariants de Riemann correspondants :

$$\lambda_1 = V - \sqrt{gH} \text{ et } \lambda_2 = V + \sqrt{gH}$$

$$U_1 = V - 2\sqrt{gH} \text{ et } U_2 = V + 2\sqrt{gH}.$$

L'écoulement est dit fluvial si les deux vitesses caractéristiques sont de signe opposé. Cela est vrai si et seulement si le nombre de Froude local vérifie :

$$\forall t, x,\ Fr(t, x) := \frac{|V(t, x)|}{\sqrt{gH(t, x)}} < 1. \tag{1.9}$$

L'Hypothèse 3 pour le système de Saint-Venant n'est pas toujours vérifiée. En effet, il faut imposer en plus la condition de Froude (1.9), qui restreint l'état du système. Si (1.9) est satisfaite, on aura :

$$\lambda_1 < 0 \text{ et } \lambda_2 > 0.$$

Dans ce cas, $R_{in} = (U_1(t, 1), U_2(t, 0))$ et $R_{out} = (U_1(t, 0), U_2(t, 1))$. Lorsque le nombre de Froude vaut localement 1 au bord, les choses se compliquent puisque l'on ne pourra plus définir si ce dernier est entrant ou sortant.

Dans cette thèse, on ne considérera pas des systèmes tels que $q > 1$, mais on se concentrera sur des systèmes un peu plus simples, moins fidèles à la physique afin d'introduire des techniques nouvelles de stabilisation. Par la suite, on prendra donc $q = 1$.

1.2.4 Stabilisation

Il existe deux moyens de stabiliser une EDP ou un système d'équations de transport scalaires.

— Le premier moyen est d'appliquer un contrôle au sein du domaine. On définira le support du contrôle que l'on notera $\omega_i \subset [0, 1]$ et on aura un système en boucle fermée de type :

$$\forall 1 \leq i \leq d,\ \partial_t R_i + \partial_x[f_i(R_i)] = S(R_1, \cdots, R_d) + 1_{\omega_i} g(u(t)).$$

où 1_{ω_i} est la fonction indicatrice de l'ensemble ω_i. Cette méthode bien que très intéressante ne sera alors pas traitée dans cette thèse.

— On se focalisera plutôt sur un contrôle $u_i(t)$ stabilisant au bord :

$$\forall 1 \leq i \leq d, \begin{cases} \partial_t R_i + \lambda_i(R_i)\partial_x R_i &= S_i(R_1, \cdots, R_q) \\ R_{i,in} &= h_i(R_{out}) + b_i(u_i(t)). \end{cases} \tag{1.10}$$

Ensuite, on fera l'hypothèse que :

$$\forall 1 \leq i \leq d,\ S_i(0, \cdots, 0) = 0$$

pour que l'état $R_i = 0$ soit un équilibre de (1.10). Ainsi, on cherchera des contrôles de la forme d'un retour d'état $u_i = k_i(R_1, \cdots, R_n)$, où $R_1, \cdots, R_n$ est un élément de L^p, tels que en boucle fermée :

$$\lim_{t \to \infty} \|R_i(t, \cdot)\|_{L^p([0,1])} = 0.$$

On parlera donc de stabilisation autour de l'état nul de systèmes de lois de conservation scalaires.

Pour la suite et jusqu'à la fin de cette thèse, on écrira notre réseau d'équations sous la forme vectorielle suivante :

$$\partial_t R + \partial_x[f(R)] = S(R)$$

où $R(t, x) \in \mathbb{R}^d$ et f est un flux scalaire dans le sens où $f_i(R) = f_i(R_i)$ et $S : \mathbb{R}^d \to \mathbb{R}^d$.

1.3 Stabilisation d'équations scalaires sans couplage au sein du domaine

Dans cette section, on va négliger le terme source $S(R)$. Ainsi, par un simple changement de variable en espace, on peut se ramener à un système d'équations en boucle ouverte à vitesses positives :

$$\begin{cases} \partial_t R + \partial_x[f(R)] &= 0 \\ R(t, 0) &= h(R(t, 1)). \end{cases}$$

Dans cette thèse, l'opérateur h sera linéaire et on aura $h(R) = H_{BO}R$, où $H_{BO} \in M_d(\mathbb{R})$ est une matrice modélisant les échanges au bord. Le système en boucle fermée sera de la forme :

$$\begin{cases} \partial_t R + \partial_x[f(R)] &= 0 \\ R(t, 0) &= H_{BO}R(t, 1) + Bu(t) \end{cases}$$

avec $B \in M_{dm}$ la matrice d'actuation.

1.3.1 La commande par retour linéaire local

Dans cette partie, on prendra un contrôle de la forme $u(t) = KR(t, 1)$, où $K \in M_{md}$. Ainsi, le schéma bloc sera le suivant :

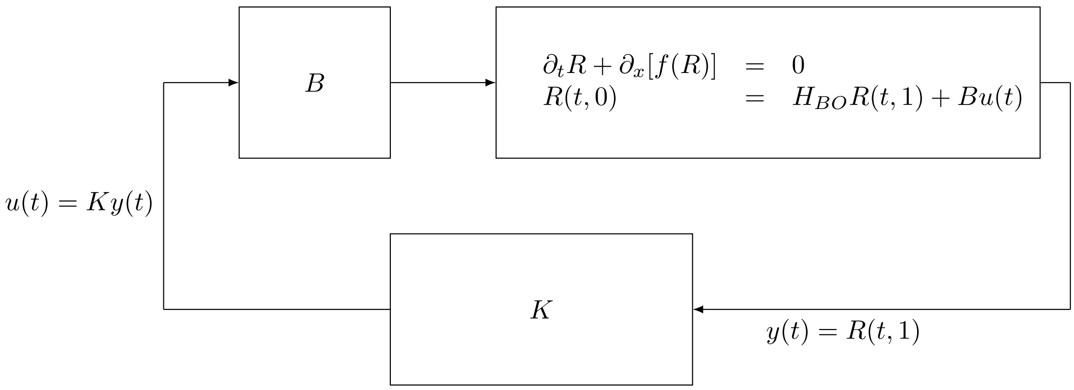


FIGURE 1.3 – Le schéma bloc en boucle fermée

On notera la matrice de retour au bord par $H := H_{BO} + BK$. Dans ce cadre, la littérature est assez riche et l'on peut caractériser la stabilité en fonction de H.

— Lorsque le flux est linéaire : $f(R) = \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_d)$ avec $\lambda_1, \cdots, \lambda_d > 0$, des conditions nécessaires de stabilité ont été prouvées dans de nombreux espaces fonctionnels. On citera par exemple :

  1. Les espaces de Sobolev de type $W^{m,p}([0,1])$ pour $m \in \mathbb{N}$ et $p \in [1, +\infty]$.
  2. Les espaces $C^m([0,1])$ and $(m \in \mathbb{N})$.
  3. L'espace $BV([0,1])$.

  En effet, dans [62, Theorem 3.5 p. 275], les auteurs prouvent que 0 est globalement exponentiellement stable dans les espaces 1.,2.,3. définis ci-dessus si et seulement si il existe $\delta > 0$ tel que

$$\left\{z \in \mathbb{C} \mid \det(I_d - \mathrm{diag}(e^{-z/\lambda_1}, \cdots, e^{-z/\lambda_d})H) = 0\right\} \subset \{z \in \mathbb{C} \mid Re(z) < -\delta\}. \quad (1.11)$$

  Cependant, le critère (1.11) n'est pas robuste par rapport à Λ. En fait, lorsque l'on fixe un couple (H, Λ) vérifiant (1.11), il est impossible de garantir que ce même critère soit vérifié pour $\tilde{\Lambda}$ avec $\tilde{\Lambda}$ arbitrairement proche de Λ [62, p. 285].
  Dans le même livre, Silkowski [62, Theorem 6.1 p. 286] démontre que pour tous les espaces de Banach 1.,2.,3. listés ci-dessus, 0 est globalement exponentiellement stable et que cette stabilité est robuste vis à vis de Λ si et seulement si

$$\rho_0(H) := \max\left\{\rho(\mathrm{diag}(e^{i\theta_1}, \cdots, e^{i\theta_d})H) \mid \theta_i \in \mathbb{R}\right\} < 1 \quad (1.12)$$

  où ρ désigne le rayon spectral. Cette condition est plus forte que (1.11) et pendant des années, la littérature a tenté de généraliser ces résultats à des flux non linéaires et vectoriels.

— Quand le flux est non linéaire et vectoriel, seules des conditions suffisantes de stabilité sont données et la plupart du temps la stabilité n'est prouvée que localement.

  1. Pour $C^m([0,1])$ avec $m \in \mathbb{N}^*$, une condition suffisante [15, 37, 95, 81] est :

$$\rho_\infty(H) := \inf_{\Delta \in D_d^+(\mathbb{R})} |\Delta H \Delta^{-1}|_\infty < 1 \quad (1.13)$$

où $|\cdot|_\infty$ est la norme infinie usuelle des matrices. Dans [37, 95, 81], la stabilité a été prouvée pour $m = 1$, mais les arguments permettent tout aussi bien de le démontrer pour $m > 0$.

2. Pour les espaces de Sobolev $W^{m,p}([0,1])$ une condition suffisante de stabilité s'écrit :

$$\rho_p(H) := \inf_{\Delta \in D_d^+(\mathbb{R})} |\Delta H \Delta^{-1}|_p < 1 \tag{1.14}$$

où $|\cdot|_p$ est la norme p usuelle des matrices.
Le cas $p = 2$ a été traité dans [30] et le cas général $p \geq 1$ dans [33]. Aussi, dans [30, 33], la stabilité fut prouvée pour $m = 2$, mais les arguments permettent également de le démontrer pour $m > 2$.

3. Pour $BV([0,1])$, peu de résultats sont connus. A notre connaissance, seul [32] étudie ce cas. Les auteurs y considèrent un système de lois de conservation $q = 2$ et donnent des conditions suffisantes sur H pour garantir une stabilité locale dans BV.

Dans cette thèse, on considère une situation intermédiaire à mi-chemin entre d'une part, le cas linéaire, et d'autre part celui très complexe des flux non linéaires vectoriels. En effet, il s'agit du cas des flux scalaires qui possède un intérêt théorique certain pour la stabilisation de systèmes d'EDPs hyperboliques plus complexes, tels que celui de Saint-Venant.

1.3.2 La commande par retour linéaire saturé

Maintenant, on décide de saturer notre commande stabilisante en introduisant une fonction de saturation $\sigma : \mathbb{R} \to \mathbb{R}$ tel qu'il existe $\sigma_s > 0$ (le niveau de saturation) tel que :

$$\begin{cases} \sigma(x) = x & \text{si } |x| \leq \sigma_s \\ \sigma(x) = \text{sign}(x)\sigma_s & \text{sinon.} \end{cases}$$

On définit ainsi la saturation par composante sur un vecteur de taille m par :

$$\forall R \in \mathbb{R}^m, \; [\sigma(R)]_i = \sigma(R_i).$$

Le schéma bloc correspondant est légèrement modifié et on obtient :

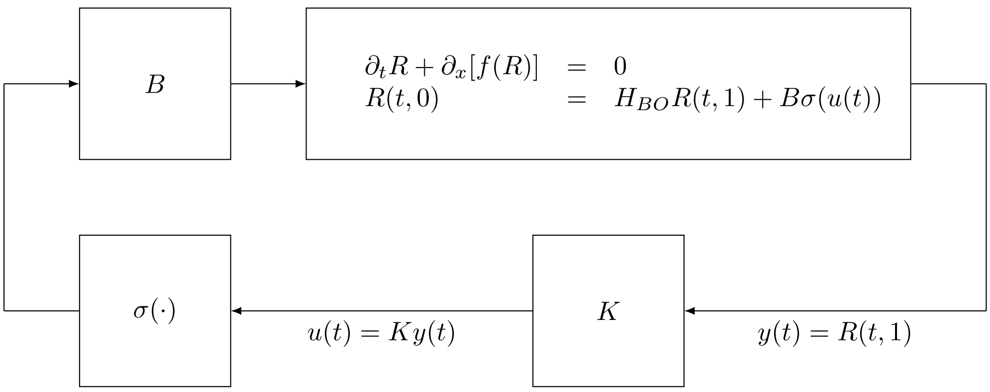


FIGURE 1.4 – Le schéma bloc en boucle fermée avec saturation

L'étude de la stabilisation avec saturation intéresse de plus en plus la communauté du contrôle, en raison des besoins en précision croissants pour la modélisation des actionneurs. Les actionneurs physiques ne peuvent pas fournir une amplitude infinie et il arrive qu'ils saturent, rendant ainsi les modèles de contrôle linéaire inadaptés. Pour éviter de rentrer dans la zone de non-linéarité, les ingénieurs font souvent le choix de prendre des actionneurs suffisamment puissants pour éviter la saturation. Cependant, surdimensionner un actionneur est loin d'être optimal en terme de masse et de coût d'opération. On pourra citer l'exemple de l'aéronautique où le moindre kilogramme ajouté est important. De plus, dans certaines situations exceptionnelles, les actionneurs peuvent être amenés à saturer et l'on se doit de prédire ce qu'il se passera pour éviter des circonstances dangereuses.

Négliger la saturation des actionneurs peut même mener à des comportements non désirés, voire à des situations catastrophiques. Dans [24] ou [102, Exemple 1.1, 1.2], les auteurs prennent un système de dimension finie et montrent avec des exemples que sous certaines conditions initiales, le système peut devenir instable, et ce même si le système non saturé est globalement stable. Ce constat nous fait entrevoir l'existence de bassins d'attraction, qui constituent l'ensemble des conditions initiales telles que la solution associée n'explose pas en norme en temps infini. A des fins d'illustration, nous allons reprendre l'exemple [102, Exemple 1.1] où on étudie l'EDO :

$$\dot{X} = H_{BO}X + B\sigma(KX)$$

avec :

$$H_{BO} := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, B := \begin{pmatrix} 0 \\ -1 \end{pmatrix}, K := \begin{pmatrix} 13 & 7 \end{pmatrix}.$$

Les valeurs propres de $H = H_{BO} + BK$ sont -3 et -4 donc le système non saturé est globalement stable. Si on sature la commande avec un niveau de saturation $\sigma_s = 5$, on aura un bassin d'attraction. Pour le représenter, on réalise une simulation numérique (schéma d'Euler explicite). Sur une fenêtre de temps $T = 10s$ on évalue le taux de convergence exponentiel de la norme euclidienne de la solution $\sqrt{X_1^2(t) + X_2^2(t)}$. Si ce taux est positif, alors la solution diverge. A contrario, s'il est négatif, alors la solution converge exponentiellement vite vers 0. On obtient ainsi la Figure 1.5 :

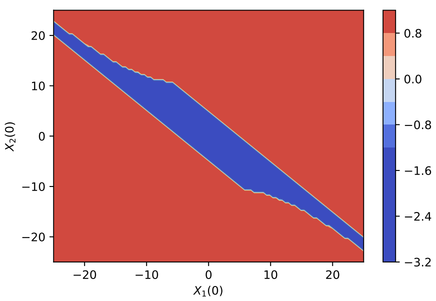


FIGURE 1.5 – Le bassin d'attraction dans notre exemple

On voit très clairement apparaître un bassin d'attraction en bleu où le système est stable. En revanche, dans la zone rouge la saturation est trop importante et le système est instable. Ce simple exemple rend palpable ce phénomène de perte de stabilité globale que l'on retrouve aussi lors du passage aux EDPs.

Pour les équations de transport, peu de travaux considèrent l'effet de la saturation. On pourra quand même citer [93] dans lequel le cas particulier de l'équation des ondes $z_{tt} = z_{xx}$ est étudié. Les auteurs prouvent une stabilité locale exponentielle dans H^1 dans le cas d'un contrôle au bord, puis d'un contrôle dans le domaine. On a également [85] où un résultat de stabilité globale sur l'équation de KdV stabilisé par un contrôle distribué et saturé, est démontré.

1.3.3 Méthode d'estimation du bassin d'attraction

Dans ce paragraphe, nous allons donner un résumé des techniques utilisées pour estimer le bassin d'attraction. Considérons un exemple pour nous aider à développer les idées. Considérons deux équations de transport linéaires couplées au bord telles que :

$$\left\{\begin{array}{rcl} \partial_t R_1 + \lambda_1 \partial_x R_1 & = & 0 \\ \partial_t R_2 + \lambda_2 \partial_x R_2 & = & 0 \\ R(t,0) & = & HR(t,1) + B\sigma(KR(t,1)) \end{array}\right. \tag{1.15}$$

où $\lambda_1, \lambda_2 > 0$ et $H, B, K \in M_2(\mathbb{R})$. Le système non saturé est le suivant :

$$\left\{\begin{array}{rcl} \partial_t R_1 + \lambda_1 \partial_x R_1 & = & 0 \\ \partial_t R_2 + \lambda_2 \partial_x R_2 & = & 0 \\ R(t,0) & = & (H+BK)R(t,1) \end{array}\right. \tag{1.16}$$

Afin de prouver des résultats de stabilité, on introduit ici la notion de fonctionnelle de Lyapunov stricte.

Définition 1. *Soit $p < \infty$, on dit que la fonctionnelle $V_p : L^p([0,1];\mathbb{R}^2) \to \mathbb{R}^+$ est une fonctionnelle de Lyapunov stricte pour le système* (1.16) *par rapport à la norme $L^p([0,1])$ si il existe $C, \gamma > 0$ tels que :*

$$\forall R \in L^p([0,1];\mathbb{R}^2),\ \frac{1}{C}\|R\|_{L^p([0,1])} \le V_p(R) \le C\|R\|_{L^p([0,1])},$$

et si de plus, pout toute solution classique R de (1.16) *on a :*

$$\forall t \ge 0,\ \frac{d}{dt}V_p(R(t,\cdot)) \le -\gamma V_p(R(t,\cdot)).$$

Tout l'enjeu sera de trouver une telle fonctionnelle de Lyapunov stricte pour le système sans saturation (1.16). Il en existe d'ailleurs une très classique pour (1.16). Il s'agit d'une norme L^p avec un poids exponentiel de la forme :

$$V_p(R) = \left(\int_0^1 \left(\frac{P_1^p}{\lambda_1}|R_1|^p + \frac{P_2^p}{\lambda_2}|R_2|^p\right) e^{-p\gamma x}dx\right)^{1/p}$$

où $P \in D_d^+(\mathbb{R})$ est à choisir astucieusement. On va essayer de le démontrer ici :

$$\begin{array}{rcl} \frac{d}{dt}V_p(R)^p & = & p\displaystyle\int_0^1 \left(\frac{P_1^p}{\lambda_1}R_1\partial_t R_1|R_1|^{p-2} + \frac{P_2^p}{\lambda_2}R_2\partial_t R_2|R_2|^{p-2}\right)e^{-p\gamma x}dx \\ & = & -p\displaystyle\int_0^1 \left(P_1^p R_1\partial_x R_1|R_1|^{p-2} + P_2^p R_2\partial_x R_2|R_2|^{p-2}\right)e^{-p\gamma x}dx \\ & = & -[(P_1^p R_1^p + P_2^p R_2^p)e^{-p\gamma x}]_0^1 - p\gamma\displaystyle\int_0^1 (P_1^p|R_1|^p + P_2^p|R_2|^p)e^{-p\gamma x}dx \\ & \le & -p\gamma\min(\lambda_1,\lambda_2)V_p^p(R) - (e^{-p\gamma}|PR(t,1)|_p^p - |P(H+BK)R(t,1)|_p^p). \end{array}$$

Maintenant il faut estimer le terme de bord. Pour cela, on introduit une hypothèse sur les matrices :

$$\rho_\infty(H) = \inf_{P\in D_d^+(\mathbb{R})} |P(H+BK)P^{-1}|_\infty < 1$$

et on prend $P \in D_d^+(\mathbb{R})$ tel que :

$$|P(H+BK)P^{-1}|_\infty < 1.$$

On estime le terme d'énergie sortant :

$$|PR(t,1)|_p^p \geq |PR(t,1)|_\infty^p$$

et le terme d'énergie rentrant :

$$|P(H+BK)R(t,1)|_p^p \leq d|P(H+BK)P^{-1}|_\infty^p |PR(t,1)|_\infty^p.$$

On obtient l'estimée du terme de bord suivante :

$$\begin{aligned} e^{-p\gamma}|PR(t,1)|_p^p - |P(H+BK)R(t,1)|_p^p \geq & \quad (e^{-p\gamma} - d|P(H+BK)P^{-1}|_\infty^p)|PR(t,1)|_\infty^p \\ \geq & \quad 0 \end{aligned}$$

où on a pris $\gamma > 0$ tel que $|P(H+BK)P^{-1}|_\infty < e^{-\gamma}$ et p suffisamment grand. Ainsi, on obtient :

$$p\frac{dV_p}{dt}V_p^{p-1} \leq -p\gamma \min(\lambda_1, \lambda_2)V_p^p,$$

ce qui donne :

$$\frac{dV_p}{dt} \leq -\gamma \min(\lambda_1, \lambda_2)V_p.$$

Il est à noter que cette estimée est uniforme en p et donc on a démontré le lemme de stabilité globale suivant dans L^∞ :

Lemme 1. *[15] Si la condition*

$$\rho_\infty(H+BK) < 1$$

est vérifiée, alors le système linéaire (1.16) *est globalement exponentiellement stable dans* $L^\infty([0,1])$.

Pour estimer le bassin d'attraction lorsqu'on sature le contrôle, il faut se baser sur des estimées de "deadzone". En supposant que le contrôle non saturé soit stabilisant ie $\rho_\infty(H+BK) < 1$ et en le comparant avec son homologue saturé, on obtient le lemme suivant qui sera démontré en chapitre 2.

Lemme 2. *Soit* $R \in \mathbb{R}^2, P \in D_d^+(\mathbb{R})$ *tels que* $|P(H+BK)P^{-1}| < e^{-\gamma}$. *Si*

$$|PR|_\infty \leq \frac{|PBP^{-1}|_\infty P_{\min}\sigma_s}{\big||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty|PKP^{-1}|_\infty - e^{-\gamma}\big|}, \tag{1.17}$$

alors

$$|P(HR + B\sigma(KR))|_\infty \leq e^{-\gamma}|PR|_\infty. \tag{1.18}$$

Ensuite en faisant une analyse de Lyapunov similaire au cas non saturé et en utilisant (1.17), on peut trouver un contour de Lyapunov de type $\sup_{x\in[0,1]}|PR(x)| \leq \textit{constante}$ où l'on aura une stabilité exponentielle. On a ainsi estimé le bassin d'attraction dans L^∞.

Remarque 2. Cette méthode ne permet pas de recouvrir la totalité du bassin d'attraction. En effet, nous utilisons des normes de L^p à poids pour nos fonctionnelles de Lyapunov. Or, les contours d'une norme sont forcément des ensembles convexes. Si l'on regarde bien la figure 1.5, on voit que le bassin d'attraction est loin d'être convexe. Avec nos méthodes, on ne pourra donc obtenir qu'au maximum le plus grand convexe inclus dans le bassin d'attraction.

1.3.4 Contributions

Les contributions de cette thèse pour l'étude du contrôle feedback saturé pour des réseaux d'EDPs de transport scalaires sont décrites ci-dessous :

- On va d'abord estimer un bassin d'attraction dans L^∞ pour un système (d quelconque) d'équations de transport linéaires avec couplage au bord. Ce sera l'objet du chapitre 2.
- On fera de même dans le cadre des lois de conservations scalaires, en utilisant des techniques de suivi de front d'ondes (chapitre 3).

La principale approche utilisée sera très similaire à la technique présentée ci-dessus. Dans un cadre numérique, on essayera d'adapter la démarche de Lyapunov présentée en paragraphe 1.3.3 pour prouver des résultats de stabilité. Cela fera l'objet de la prochaine section.

1.4 Méthode de Lyapunov discrète

Dans cette partie, nous nous concentrerons sur l'approximation numérique de systèmes de transport les plus simples de type :

$$\left\{\begin{array}{rcl} \partial_t R + \partial_x[f(R)] & = & 0 \\ R(t,0) & = & HR(t,1) \end{array}\right. \tag{1.19}$$

où $R \in \mathbb{R}^d$, $H \in M_d(\mathbb{R})$ et f est un flux scalaire dans le sens où $f_i(R) = f_i(R_i)$ $(i \in \{1, \cdots, d\})$ et à vitesse de propagation positive *ie* $f_i' > 0$.

1.4.1 Méthodes de volumes finis

Pour commencer, on va rappeler la méthode des volumes finis. Elle consiste à discrétiser notre variable d'espace $x \in [0,1]$ en un nombre fini $N \in \mathbb{N}$ de cellules. On fera de même pour la variable de temps t. Pour cela, on introduit deux paramètres de discrétisation $dx, dt > 0$ supposés petits et

$$\forall n \in \mathbb{N},\ 1 \leq j \leq N,\ \left\{\begin{array}{rcl} x_j & := & (j-1/2)dx \\ t^n & := & ndt \end{array}\right.,\ C_j := (x_j - dx/2, x_j + dx/2).$$

La méthode des volumes finis consiste à donner une approximation constante par morceaux $R_{\Delta x}$ de la solution de (1.19), aux temps t^n et sur les cellules C_j. On notera $(R_j^n)_{n,j}$ les valeurs de $R_{\Delta x}(t^n, \cdot)$ sur les cellules C_j. L'indice n est l'indice de temps et j l'indice d'espace. Ensuite, nous intégrons (1.19) sur une cellule d'espace-temps $[ndt, (n+1)dt] \times C_j$ et réalisons un bilan de flux :

$$\int_{C_j} R((n+1)dt,\cdot)dx-\int_{C_j} R(ndt,\cdot)dx+\int_{ndt}^{(n+1)dt} f(R(t,x_j+dx/2))dt-\int_{ndt}^{(n+1)dt} f(R(t,x_j-dx/2))dt = 0.$$

Afin de déterminer $(R_j^n)_{n,j}$, on donne une approximation des différents termes de la précédente équation.

— Dans toute cette thèse, les termes temporels seront approximés par :

$$\begin{cases} \int_{C_j} R((n+1)dt,\cdot)dx \approx & R_j^{n+1}dx \\ \int_{C_j} R(ndt,\cdot)dx \approx & R_j^n dx. \end{cases}$$

— Pour les termes spatiaux et sans se soucier pour l'instant du bord, on prendra une approximation à trois points de la forme :

$$\begin{cases} \int_{ndt}^{(n+1)dt} f(R(t,x_j+dx/2))dt \approx & \tilde{f}(R_{j-1}^n, R_j^n, R_{j+1}^n)dt \\ \int_{ndt}^{(n+1)dt} f(R(t,x_j-dx/2))dt \approx & \tilde{f}(R_{j-2}^n, R_{j-1}^n, R_j^n)dt \end{cases}$$

où $\tilde{f}$ est une approximation du flux à choisir judicieusement.

On pourra ainsi (sans se soucier du bord) définir $(R_j^n)_{n,j}$ par récurrence sur l'indice de temps n :

$$\frac{R_j^{n+1}-R_j^n}{dt} + \frac{\tilde{f}(R_{j-1}^n, R_j^n, R_{j+1}^n) - \tilde{f}(R_{j-2}^n, R_{j-1}^n, R_j^n)}{dx} = 0. \tag{1.20}$$

On dit alors que l'équation (1.20) est un schéma numérique. On peut citer plusieurs propriétés classiques et importantes d'un schéma numérique. Pour cela supposons que l'on ait accès à une solution R aussi régulière que l'on veut au système d'EDPs (1.19). On pose

$$\begin{aligned}\mathcal{C}_j^n := & \frac{R((n+1)dt,x_j)-R(ndt,x_j)}{dt} \\ & + \frac{\tilde{f}(R(ndt,x_{j-1}),R(ndt,x_j),R(ndt,x_{j+1})) - \tilde{f}(R(ndt,x_{j-2}),R(ndt,x_{j-1}),R(ndt,x_j))}{dx}\end{aligned}$$

que l'on appellera l'erreur de consistance du schéma.

Définition 2. *On dit que le schéma est consistant d'ordre $p_t \in \mathbb{N}$ en temps et p_x en espace si pour tout $T>0$, il existe $C>0$ indépendant de dx, dt tel que :*

$$\forall ndt < T, j,\ |\mathcal{C}_j^n| \leq C(dx^{p_x} + dt^{p_t}).$$

Pour donner un exemple prenons le cas d'une simple equation de transport linéaire à vitesse unité, approximée par un schéma très simple de type :

$$\frac{R_j^{n+1}-R_j^n}{dt} + \frac{R_j^n - R_{j-1}^n}{dx} = 0.$$

Ensuite, on pose R aussi régulière que l'on veut et solution de l'équation de transport $\partial_t R + \partial_x R = 0$. On a ainsi avec un développement de Taylor :

$$\frac{R_j^{n+1}-R_j^n}{dt} = \partial_t R(ndt,x_j) + \frac{dt}{2}\partial_{tt}R(ndt,x_j) + o(dt)$$

et

$$\frac{R_j^n - R_{j-1}^n}{dx} = \partial_x R(ndt, x_j) + \frac{dx}{2}\partial_{xx} R(ndt, x_j) + o(dx).$$

En sommant, on obtient $\mathcal{C}_n^j = \frac{dt}{2}\partial_{tt} R(ndt, x_j) + \frac{dx}{2}\partial_{xx} R(ndt, x_j) + o(dt, dx)$ et ainsi le schéma est d'ordre de consistance égal à un en temps et un en espace. Une autre définition très importante introduit la notion de stabilité du schéma.

Définition 3. *On dit que le schéma est stable pour la norme L^p si pour tout $T > 0$, il existe $C > 0$ indépendant des paramètres de la discrétisation, telle que :*

$$\forall R_0 \in L^p,\ \forall ndt < T,\ \sum_j |R_j^n|^p dx \leq C \|R^0\|_{L^p}^p.$$

Finalement, la notion de schéma convergent est donnée ici.

Définition 4. *On dit que le schéma est convergent dans un espace de Banach E de norme $\|\cdot\|_E$ s'il existe $R \in E$ tel que :*

$$\lim_{dt, dx \to 0} \|R_{\Delta x} - R\|_E = 0.$$

En général, on dira que le schéma est convergent s'il existe une sous-suite convergente dans E de $R_{\Delta x}$. Parfois, il s'agira uniquement d'une convergence faible et on dira quand même que le schéma est convergent.

1.4.2 Le choix de la fonction de flux

Il existe un grand nombre de fonctions $\tilde{f}$ pour approximer le flux aux interfaces de C_j. Dans cette thèse, nous nous concentrerons sur deux types de fonction $\tilde{f}$:

— Comme on a posé l'hypothèse que les vitesses de propagation étaient positives, on peut supposer que le flux à droite de C_j est sortant et ne s'exprime qu'avec R_j^n :

$$\tilde{f}(R_{j-1}^n, R_j^n, R_{j+1}^n) = f(R_j^n).$$

Ce schéma bien connu se nommera alors schéma "upwind", cas particulier du schéma de Godunov [60]. Sous une condition de CFL :

$$\frac{dt \max_{1 \leq i \leq d, R} f_i'(R_i)}{dx} < 1,$$

on peut montrer qu'il est stable et convergent. Son principal inconvénient est qu'il est très diffusif, ce qui régularise fortement la solution. Dans un contexte hyperbolique dans lequel on peut se trouver en présence de discontinuités, le schéma upwind donne des solutions assez éloignées des celles continues lorsque la discrétisation est grossière. De plus, c'est un schéma d'ordre de consistance égale à un en temps et en espace uniquement.

— Pour pallier ce problème, on introduit une fonction de flux un peu plus générale de type [108] [79] :

$$\tilde{f}(R_{j-1}^n, R_j^n, R_{j+1}^n) = f(R_j^n + \tilde{R}_j^n)$$

où

$$\tilde{R}_j^n = \phi\left(\frac{R_j^n - R_{j-1}^n}{dx}, \frac{R_{j+1}^n - R_j^n}{dx}\right)\frac{dx}{2}.$$

Le schéma associé se nomme schéma à limiteur de pente. Lorsque la solution est assez régulière, on prendra $\tilde{R}_j^n = \frac{R_{j+1}^n - R_j^n}{dx}\frac{dx}{2}$ pour retrouver localement un schéma de type centré d'ordre deux. Or il est bien connu que le schéma centré est instable notamment autour des extremums locaux et des discontinuités de la solution. C'est pourquoi on introduit le limiteur de pente ϕ de la forme :

$$\phi(u, v) = \phi_r(u/v)v. \tag{1.21}$$

Pour expliquer la construction de ce schéma, il faut distinguer deux cas. Si l'on se trouve loin d'un extremum local et que la solution est régulière *ie* $u/v \approx 1$ alors $\phi_r(u/v) \approx 1$ et on retrouve le schéma d'ordre deux centré. Si au contraire, la solution admet localement un extremum local, ou bien est discontinue, alors on prendra $\phi_r(u/v) \approx 0$ pour retrouver le schéma upwind qui possède de bonnes propriétés de stabilité. Ainsi, on obtient un schéma globalement moins diffusif et plus précis, tout en conservant de la stabilité.

1.4.3 Les conditions aux bords

Dans cette thèse, nous travaillons dans un domaine borné $x \in [0, 1]$. Cela pose un problème dans (1.20) puisque lorsque $j = N$ et $j = 0$, R_{N+1}^n et R_0^n n'ont pas été définis. Pour pallier ce problème, on impose un schéma de type upwind à droite du domaine :

$$\frac{R_N^{n+1} - R_N^n}{dt} + \frac{f(R_N^n) - f(R_{N-1}^n)}{dx} = 0,$$

et on introduit une cellule fantôme en $x_0 := -dx/2$ et on définit :

$$R_0^n = HR_N^n.$$

1.4.4 Contributions

D'une part, la contribution du chapitre 4 sera de trouver une fonctionnelle de Lyapunov stricte dans l'espace fonctionnel $BV([0, 1])$ pour le système d'équations vérifié par $R_{\Delta x}$ dans le cadre d'un schéma à limiteur de pente. D'autre part, il s'agira d'analyser des résultats numériques montrant l'influence du choix du limiteur ϕ sur la dynamique énergétique de la solution.

1.5 Les méthodes linéaires

Dans cette section, nous allons nous concentrer plutôt sur l'espace $L^2([0, 1])$. Les EDPs seront linéaires avec une interaction au sein du domaine. On prendra $d = 2$ équations de transport avec une vitesse positive et une vitesse négative :

$$\begin{cases} \partial_t \bar{R} + \lambda_+(x)\partial_x \hat{R} &= \bar{M}_{11}(x)\bar{R} + \bar{M}_{12}(x)\bar{S} \\ \partial_t \bar{S} - \lambda_-(x)\partial_x \bar{S} &= \bar{M}_{21}(x)\bar{R} + \bar{M}_{22}(x)\bar{S} \\ \bar{R}(t, 0) &= \bar{u}(t) \\ \bar{S}(t, 1) &= \bar{h}\bar{R}(t, 1). \end{cases}$$

On retrouve ce type de modèle lorsque l'on linéarise des systèmes d'EDPs physiques type Saint-Venant. Par un changement de variable [16, p 176], on peut éliminer les termes diagonaux dans le membre de droite et en ces nouvelles variables, le système prend la forme suivante :

$$\begin{cases} \partial_t R + \lambda_+(x)\partial_x R &= M_{12}(x)S \\ \partial_t S - \lambda_-(x)\partial_x S &= M_{21}(x)R \\ R(t,0) &= u(t) \\ S(t,1) &= hR(t,1). \end{cases}$$

Finalement, pour simplifier notre étude on ne gardera que des paramètres constants :

$$\begin{cases} \partial_t R + \lambda_+\partial_x R &= M_{12}S \\ \partial_t S - \lambda_-\partial_x S &= M_{21}R \\ R(t,0) &= u(t) \\ S(t,1) &= hR(t,1). \end{cases} \tag{1.22}$$

Le système (1.22) sans condition au bord peut s'écrire sous forme matricielle :

$$\partial_t U_t + \Lambda\partial_x U = MU$$

où :

$$U(t,x) = \begin{pmatrix} R(t,x) \\ S(t,x) \end{pmatrix}, \Lambda = \begin{pmatrix} \lambda_+ & 0 \\ 0 & -\lambda_- \end{pmatrix}, M = \begin{pmatrix} 0 & M_{12} \\ M_{21} & 0 \end{pmatrix}.$$

Pour stabiliser ce système, on pourrait envisager les méthodes de Lyapunov des sections précédentes mais il a été démontré dans [13], [64] que si les termes de couplages M étaient trop importants dans un sens à définir, alors on ne peut pas trouver de fonctionnelle de Lyapunov quadratique dans L^2 pour le système (1.22). De plus, une analyse spectrale [16, Proposition 5.2] montre que pour n'importe quel contrôle de la forme $u(t) = fR(t,0)$ (f réel) le système sera toujours instable dans L^2. Il faudra donc radicalement changer de paradigme pour stabiliser un tel système. On rappelle le théorème prouvé dans [13] :

Théorème 1. *Le système* (1.22) *admet une fonctionnelle quadratique de Lyapunov dans* L^2 *de la forme :*

$$V(R,S) = \sqrt{\int_0^1 q_R(x)R^2(x) + q_S(x)S^2(x) + q_{RS}(x)R(x)S(x)dx},$$

si et seulement si la solution de :

$$\begin{cases} \eta'(x) = & \left|\frac{M_{12}}{\lambda_+} + \frac{M_{21}}{\lambda_-}\eta^2\right| \\ \nu(0) = & 0 \end{cases}$$

est définie sur $[0,1]$.

Dans cette section, on va donc mettre de côté les méthodes de Lyapunov, et l'on va plutôt utiliser des techniques permettant de stabiliser (1.22) qu'elles que soient les valeurs de M. On profitera fortement de la linéarité du système.

1.5.1 Stabilisation par placement de pôles

Dans ce cadre, on va considérer le système (1.22) comme une équation d'évolution dans l'espace de dimension finie qu'est $L^2([0,1])$:

$$\frac{d}{dt}U = \mathcal{A}U + \mathcal{B}u(t)$$

où :

- L'opérateur $\mathcal{A}$ est l'opérateur linéaire non borné tel que :

$$\begin{cases} D(\mathcal{A}) = & \left\{U \in L^2([0,1];\mathbb{R}^2) \mid U' \in L^2([0,1];\mathbb{R}^2), R(0) = 0, S(1) = hR(1)\right\} \\ \mathcal{A}U = & -\Lambda U' + MU. \end{cases}$$

  C'est l'opérateur représentant la boucle ouverte.
- L'opérateur $\mathcal{B}$ est l'opérateur de contrôle :

$$\mathcal{B}u := (\sqrt{\lambda_1}u, 0)\delta(x)$$

  où δ est la distribution de Dirac usuelle.

En chapitre 5 et dans un cadre plus général que $d = 2$ équations, on emploiera le Spectral Mapping Theorem qui n'est pas vrai pour tous les opérateurs non bornés mais qui l'est pour notre opérateur de boucle ouverte $\mathcal{A}$. Ce théorème est prouvé dans [82] et on le rappelle ici :

Théorème 2. *Le spectre de l'opérateur $\mathcal{A}$ et celui son semigroupe associé sont reliés par l'égalité suivante :*

$$\sigma(e^{\mathcal{A}t})\backslash\{0\} = \overline{e^{\sigma(A)t}}\backslash\{0\}.$$

Plus particulièrement, dans [82] on montre que le spectre de l'opérateur de la boucle ouverte est uniquement discret avec un nombre fini de valeurs propres instables. Ainsi, afin de stabiliser notre système, on projettera l'état R sur l'espace propre instable de dimension finie, projection que l'on stabilisera à l'aide des techniques de la dimension finie. On citera le théorème de placement de pôles :

Théorème 3. *[29, Section 10.1] Soit (A, B) $(A \in M_d(\mathbb{R}), B \in M_{dm}(\mathbb{R}))$ un couple de matrices contrôlables alors pour tout d-uplet de valeurs propres $(\mu_i)_i$, il existe une matrice $K \in M_{md}(\mathbb{R})$ telle que le spectre de $A + BK$ soit exactement $(\mu_i)_i$.*

Ainsi on pourra trouver un actionneur de dimension finie qui stabilisera (1.22). On montrera cela de manière rigoureuse en chapitre 5 pour un cas plus général que seulement deux équations de transport.

Remarque 3. Pour synthétiser ce genre de contrôle, on aura besoin de connaître la projection de l'état sur son espace instable. Cela oblige à connaître l'état sur la totalité de l'espace $x \in [0, 1]$ et pas seulement sur le bord. On touche ici à une limitation pratique de la méthode. On a besoin de plus d'informations sur l'état qu'uniquement celle du bord. En réalité, on ne va pas observer directement l'état dans sa totalité, mais plutôt construire ce que l'on appelle un observateur qui, à partir des informations aux bords, donne une bonne approximation en temps réel de l'état sur $[0, 1]$. On peut citer [25, 54, 58, 39, 110] qui étudient ce genre de problématique.

1.5.2 La méthode de backstepping pour la stabilisation en temps fini

La méthode de backstepping est présentée ici. Elle sera appliquée pour obtenir une stabilisation en temps fini dont la définition est donnée ici :

Définition 5. *Un contrôle par retour d'état stabilise en temps fini $T > 0$ le système* (1.22) *dans* $L^2([0,1])$ *si pour toute condition initiale* $(R^0, S^0) \in L^2([0,1])^2$,

$$\forall t \geq T, R(t,\cdot) = 0 \text{ et } S(t,\cdot) = 0$$

dans $L^2([0,1])$.

La méthode de backstepping a d'abord été construite pour les EDOs non linéaires (le livre [72] en est une bonne introduction), où l'idée principale est de trouver une transformation bijective, changeant notre système en un système plus facile à stabiliser. La littérature est vaste pour étendre cette technique aux EDPs. On citera [100, 73, 99] et notamment le livre [74] qui est une très bonne introduction au domaine.

On utilise la stratégie de backstepping afin de synthétiser un contrôle. L'idée est de construire une transformation inversible de L^2 vers L^2. On prendra la transformée du deuxième ordre de Volterra qui permet de passer à un système que l'on qualifiera de système cible. Cette transformée est décrite ci-dessous :

$$U^\star(t,x) = U(t,x) - \int_x^1 P(x,\xi)U(t,\xi)d\xi \tag{1.23}$$

où P prend ses valeurs dans $M_{2,2}(\mathbb{R})$ et est définie sur le triangle $\{(x,\xi) \mid 0 \leq x \leq \xi \leq 1\}$. On exprime la dérivée en temps de $U^\star$:

$$\begin{aligned} \partial_t U^\star(t,x) &= \partial_t U(t,x) - \int_x^1 P(x,\xi)\partial_t U(t,\xi)d\xi \\ &= \partial_t U(t,x) - \int_x^1 P(x,\xi)(-\Lambda\partial_\xi U + MU)(t,\xi)d\xi \\ &= \partial_t U(t,x) + P(x,1)\Lambda U(t,1) - P(x,x)\Lambda U(t,x) \\ &\quad - \int_x^1 (\partial_\xi P\Lambda + PM)(x,\xi)U(t,\xi)d\xi. \end{aligned}$$

Pour la dérivée en espace :

$$\partial_x U^\star(t,x) = \partial_x U(t,x) + P(x,x)U(t,x) - \int_x^1 P_x(x,\xi)U(t,\xi)d\xi.$$

En rassemblant les résultats précédents, on obtient le système d'EDPs vérifié par $U^\star$:

$$\begin{aligned} \partial_t U^\star(t,x) + \Lambda\partial_x U^\star(t,x) = & \int_x^1 (-\Lambda\partial_x P - \partial_\xi P\Lambda - PM)(x,\xi)U(t,\xi)d\xi \\ & +(M + \Lambda P(x,x) - P(x,x)\Lambda)U(t,x) \\ & +P(x,1)\Lambda U(t,1). \end{aligned}$$

On choisit maintenant P comme l'unique solution du système :

$$\begin{cases} \partial_\xi P\Lambda + \Lambda\partial_x P + PM &= 0 \\ M + \Lambda P(x,x) - P(x,x)\Lambda &= 0 \\ P(x,1)\Lambda(1,h)^T &= 0 \end{cases} \tag{1.24}$$

qui est bien posé d'après [34]. Ainsi, le système vérifié par $U^\star$ est plus simple :

$$\partial_t U^\star + \Lambda \partial_x U^\star = 0$$

et on a supprimé le terme d'ordre zero. Il faut désormais donner les conditions aux bords pour l'état cible $U^\star$. D'après (1.23), on a :

$$\begin{cases} U_1^\star(t,0) &= u(t) - \int_0^1 p_{11}(0,\xi)U_1(\xi) + p_{12}(0,\xi)U_2(\xi)d\xi \\ U_2^\star(t,1) &= U_2(t,1) = hU_1(t,1) = hR_1^\star(t,1). \end{cases}$$

Le contrôle $u(t)$ est donc choisi tel que :

$$U_1^\star(t,0) = u(t) - \int_0^1 p_{11}(0,\xi)U_1(\xi) + p_{12}(0,\xi)U_2(\xi)d\xi = 0$$

et le système cible s'écrit de la manière suivante :

$$\begin{cases} \partial_t U^\star + \Lambda \partial_x U^\star &= 0 \\ U_1^\star(t,0) &= 0 \\ U_2^\star(t,1) &= hU_2^\star(t,1). \end{cases} \tag{1.25}$$

Après un temps $t \geq 1/\lambda_+ + 1/\lambda_-$, l'état solution de (1.25) est nul et comme la transformée de Volterra est bijective, l'état primal l'est aussi.

Cette méthode est bien connue dans un contexte que l'on qualifiera de continu. Dans un cadre numérique où l'on discrétise les équations, la théorie du backstepping est inexistante. Cela est d'autant plus dommageable que les actionneurs utilisés en industrie sont numériques, c'est à dire que le contrôle est exprimé dans un espace de dimension finie. C'est sur ce problème que nous nous pencherons dans la dernière partie de l'introduction.

1.6 La méthode de backstepping discretisé

On considère toujours (1.22) mais à la différence qu'ici nous nous placerons dans une situation numérique. La méthode de backstepping étant principalement linéaire, il est impératif de choisir un schéma linéaire. On prendra pour cela le schéma le plus simple, le schéma upwind.

1.6.1 Le schéma

Soient N, α deux entiers et $dx_- := 1/N$ et $dx_+ := 1/\alpha N$. On introduit également le pas de temps $dt > 0$ vérifiant une condition de CFL classique :

$$\begin{cases} \nu_+ := & \frac{\lambda_+ dt}{dx_+} \leq 1 \\ \nu_- := & \frac{\lambda_- dt}{dx_-} \leq 1. \end{cases} \tag{1.26}$$

Dans cette section, on aura une grille de temps et deux grilles d'espace :

$$\begin{cases} \forall n \in \mathbb{N}, t^n := ndt & & \\ \forall 1 \leq j_c \leq N, x_{j_c}^c := (j_c - 1/2)dx_- & \text{and} & C_{j_c}^c := (x_{j_c}^c - dx_-/2, x_{j_c}^c + dx_-/2) \\ \forall 1 \leq j_f \leq \alpha N, x_{j_f}^f := (j_f - 1/2)dx_+ & \text{and} & C_{j_f}^f := (x_{j_f}^f - dx_+/2, x_{j_f}^f + dx_+/2). \end{cases}$$

L'approximation numérique (R^n, S^n) (n est l'indice temporel) est constante par morceaux sur les cellules $C_{j_f}^f$ (respectivement $C_{j_c}^c$). Le schéma est donné ci-dessous :

— La condition initiale se calcule comme suit : $R^0_{j_f} = \frac{1}{dx_+}\int_{C^f_{j_f}} R^0(x)dx$ et $S^0_{j_c} = \frac{1}{dx_-}\int_{C^c_{j_c}} S^0(x)dx$.

— Si on suppose que (R^n, S^n) est donné, on calcule (R^{n+1}, S^{n+1}) en utilisant les formules suivantes :

$$\left\{\begin{array}{lcl} R^{n+1} & = & R^n + dt(-\lambda_+\partial_x^+ R^n + M_{12}\Pi_{f\leftarrow c}S^n + Bu^n) \\ S^{n+1} & = & S^n + dt(\lambda_-\partial_x^- S^n + M_{21}\Pi_{c\leftarrow f}R^n + B_2R^n) \end{array}\right. \tag{1.27}$$

Pour décrire les opérateurs dans (1.27), nous aurons besoin de la définition suivante afin de passer de la grille fine à la grille grossière et vice versa :

Définition 6. *Pour tout* $1 \le i_c \le N$,

$$N_f(i_c) := \left\{1 \le i_f \le \alpha N \mid x^f_{i_f} \in C^c_{i_c}\right\}.$$

De plus, $1 \le N_c(i_f) \le N$ *est l'unique indice tel que* $C^f_{i_f} \subset C^c_{N_c(i_f)}$.

Les opérateurs dans la définition du schéma (1.27) sont décrits ci-dessous :

1. On introduit l'opérateur de transport à vitesse positive $\partial_x^+ \in M_{\alpha N,\alpha N}(\mathbb{R})$:

$$\partial_x^+ := \begin{pmatrix} 1/dx_+ & 0 & \cdots & \cdots & 0 \\ -1/dx_+ & 1/dx_+ & \cdots & \cdots & \vdots \\ 0 & -1/dx_+ & 1/dx_+ & \cdots & \vdots \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & -1/dx_+ & 1/dx_+ \end{pmatrix}.$$

Pour celui à vitesse négative, on a $\partial_x^- \in M_{N,N}(\mathbb{R})$:

$$\partial_x^- := \begin{pmatrix} -1/dx_- & 1/dx_- & \cdots & \cdots & 0 \\ 0 & -1/dx_- & 1/dx_- & \cdots & \vdots \\ 0 & 0 & -1/dx_- & 1/dx_- & \vdots \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & -1/dx_- \end{pmatrix}.$$

2. On introduit l'opérateur de projection de la grille grossière vers la grille fine $\Pi_{f\leftarrow c} \in M_{\alpha N,N}(\mathbb{R})$. Afin de la définir, on prend une cellule grossière indexée par $1 \le i_c \le N$ et un vecteur grossier $S \in \mathbb{R}^N$:

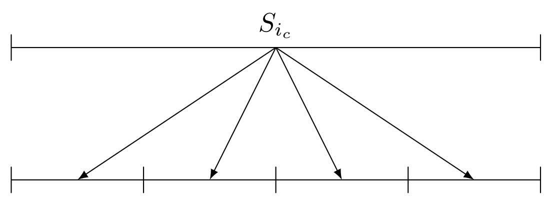


FIGURE 1.6 – La définition de $\Pi_{f\leftarrow c}$

Le vecteur fin $\Pi_{f\leftarrow c}S \in \mathbb{R}^{\alpha N}$ est construit en copiant la valeur de S dans la cellule grossière correspondante *i.e.* :

$$\forall j_f \in N_f(i_c), (\Pi_{f\leftarrow c}S)_{j_f} = S_{i_c}.$$

Pour la projection de la grille fine vers la grille grossière, on définit l'opérateur $\Pi_{c\leftarrow f} \in M_{N,\alpha N}(\mathbb{R})$. Pour ce faire, on considère un vecteur fin R et une cellule grossière indexée par $1 \leq i_c \leq N$

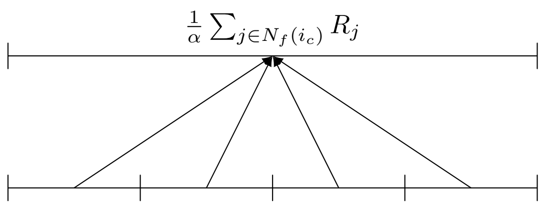


FIGURE 1.7 – La définition de $\Pi_{c\leftarrow f}$

Le réel $(\Pi_{c\leftarrow f}R)_{i_c}$ est calculé en prenant la moyenne arithmétique de R sur l'ensemble des cellules fines correspondant à la cellule grossière i_c. Ainsi, on a la propriété de dualité suivante :

$$\langle R, \Pi_{f\leftarrow c}S\rangle_f dx_+ = \langle \Pi_{c\leftarrow f}R, S\rangle_c dx_- \tag{1.28}$$

avec $\langle\cdot,\cdot\rangle_f, \langle\cdot,\cdot\rangle_c$ sont les produits scalaires canoniques respectifs de $\mathbb{R}^{\alpha N}$ et $\mathbb{R}^N$. Et donc :

$$\Pi_{f\leftarrow c} = \alpha \Pi_{c\leftarrow f}^T. \tag{1.29}$$

3. L'opérateur de contrôle discrétisé $B \in M_{\alpha N,1}$ est donné ci-dessous :

$$B := \begin{pmatrix} \lambda_+/dx_+ \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

4. L'opérateur d'interaction au bord $B_2 \in M_{N,\alpha N}(\mathbb{R})$ est défini tel que :

$$B_{2,N,\alpha N} := \frac{h\lambda_-}{dx_-}$$

et nul pour toutes les autres composantes.

1.6.2 La méthode de Backstepping discrétisée

Dans cette section, il s'agit d'appliquer la méthode de backstepping à notre système discrétisé (1.27). On introduit pour cela la transformation de Volterra du deuxième ordre :

$$\begin{pmatrix} R^\star \\ S^\star \end{pmatrix} = \mathcal{T}\begin{pmatrix} R \\ S \end{pmatrix} \iff \left\{ \begin{array}{lcl} R^\star & = & R - P_{11}Rdx_+ - P_{12}Sdx_- \\ S^\star & = & S \end{array} \right.$$

où la structure de P_{11}, P_{12} est triangulaire supérieure dans un certain sens que l'on définira en chapitre 6. Tout l'enjeu de la méthode de backstepping sera de trouver un couple (P_{11}, P_{12}) et un contrôle de type feedback $u(t)$ tel que le système cible soit de la forme :

$$\left\{\begin{array}{rcl} \dfrac{R^{\star n+1} - R^{\star n}}{dt} + \lambda_+ \partial_x^+ R_n^\star & = & 0 \\ \dfrac{S^{\star n+1} - S^{\star n}}{dt} - \lambda_- \partial_x^- S^{\star n} & = & (B_2 + M_{21}\Pi_{c\leftarrow f})(R^{\star n} + L_{11}R^{\star n}dx_+ + L_{12}S^{\star n}dx_-) \end{array}\right. \tag{1.30}$$

où L_{11}, L_{12} sont aussi triangulaires supérieures dans un certain sens.

Remarque 4. L'équivalent de (1.30) en continu est :

$$\left\{\begin{array}{lcl} \partial_t R^\star(t,x) + \lambda_+ \partial_x R^\star(t,x) & = & 0 \\ \partial_t S^\star(t,x) - \lambda_- \partial_x S^\star(t,x) & = & M_{21}(R^\star(t,x) + \int_x^1 (L_{11}(x,\xi)R^\star(t,\xi) + L_{12}(x,\xi)S^\star(t,\xi))d\xi) \\ R^\star(t,0) & = & 0 \\ S^\star(t,1) & = & hR^\star(t,1). \end{array}\right.$$

Cela ne ressemble pas vraiment au système cible (1.25) plus simple. Cependant, on peut montrer que les deux systèmes sont équivalents, dans le sens où il existe une transformation bijective continue entre les deux systèmes.

La dernière étape de la méthode de backstepping sera de montrer que la solution du système est nulle lorsque le temps d'extinction $T = \frac{1}{\lambda_+} + \frac{1}{\lambda_-}$ est dépassé.

1.6.3 Contributions

Dans le chapitre 6, nous suivrons rigoureusement la méthode de backstepping discrétisé présentée ci-dessus en faisant attention à deux points :

- La transformation de backstepping et son inverse n'explosent pas (dans un certain sens) lorsque le paramètre de discrétisation tend vers zéro. Sans cela, il est impossible de garantir la stabilité du système primal même si l'on établit la stabilité du système cible, puisqu'en effet les constantes de continuité de la transformée peuvent exploser.
- On ne pourra pas prouver une extinction stricte dans le sens où après un temps $t = \frac{1}{\lambda_+} + \frac{1}{\lambda_-}$ la solution numérique avec contrôle au bord ne sera rigoureusement pas nulle. Cela provient du fait que l'on travaille avec un système linéaire de dimension finie. Dès lors, on obtiendra plutôt un résultat de type :

  Théorème 4. *Pour $T > \frac{1}{\lambda_+} + \frac{1}{\lambda_-}$ and $1 < p < \infty$, il existe une constante C indépendante de la discrétisation telle que pour tout n tel que $\frac{1+dt^{1/8}}{1-C\sqrt{dt}}\left(\frac{1}{\lambda_+} + \frac{1}{\lambda_-}\right) \leq ndt \leq T$ et toute condition initiale $(R^0, S^0) \in L^\infty([0,1])^2$,*

  $$\|R^n\|_{L^p([0,1])} + \|S^n\|_{L^p([0,1])} \leq C(\|R^0\|_{L^\infty([0,1])} + \|S^0\|_{L^\infty([0,1])})dx_-.$$

1.7 Organisation du manuscrit

Ce manuscrit est écrit selon le modèle de compilation d'articles. Il traitera dans l'ordre présenté ci-dessous les sujets présentés en introduction :

— Le chapitre 2 correspond à l'étude d'un système d'EDPs de transport scalaires linéaires avec une loi de feedback au bord saturé. Il s'agit de donner une estimation du bassin d'attraction. L'approche est basée sur une étude de semigroupes non linéaires.

— Le chapitre 3 se focalise sur les lois de conservation scalaires avec retour au bord saturé. Le cadre sera L^∞/BV.

— Le chapitre 4 sera consacré aux schémas à limiteur de pente pour donner une approximation numérique des systèmes de lois de conservation scalaires décrits dans le chapitre précédent. On prouvera une stabilité de Lyapunov dans un cadre BV lorsque le retour est linéaire.

— Le chapitre 5 s'inscrit dans un cadre totalement linéaire. On y étudie la stabilisation de systèmes d'équations de transport linéaires couplées au bord et dans le domaine. L'idée principale se base sur un placement de pôles que l'on justifiera rigoureusement.

— Le chapitre 6 se focalisera sur le backstepping discrétisé. On y établi la stabilisation en temps fini d'un système de deux équations de transport linéaires à vitesses de signe opposé. Le couplage se fait au bord et dans le domaine.

— Enfin, en chapitre 7, nous exposons nos conclusions et perspectives.

Cette thèse a donné lieu a deux publications, deux articles soumis et un en préparation :

— M. Dus, F. Ferrante and C. Prieur, **On L^∞ stabilization of semilinear hyperbolic systems by saturated boundary control**, *ESAIM : Control, Optim. Cal. Var.*, vol. 26 (23) (2020).

— M. Dus, **Bv exponential stability for systems of scalar conservation laws using saturated controls**, *SIAM journal on Control and Optimization*, (2021).

— M. Dus, **Exponential stability of a general slope limiter scheme for scalar conservation laws subject to a dissipative boundary condition**, *Submitted*, https ://hal.archives-ouvertes.fr/hal-03116551, (2021).

— M. Dus, F. Ferrante and C. Prieur, **Spectral stabilization of linear transport equations with boundary and in-domain couplings**, *Submitted*, (2021).

— M. Dus, **The discretized backstepping : an application to a general system of linear hyperbolic partial differential equations**, *in preparation*, (2021).

CHAPITRE 2

SUR LA STABILISATION L^∞ DES SYSTÈMES HYPERBOLIQUES SEMILINÉAIRES PAR UN CONTRÔLE AU BORD SATURÉ

Le contenu de ce chapitre correspond à un article écrit avec F. Ferrante et C. Prieur. Il a donné lieu à la publication suivante :

M. Dus, F. Ferrante and C. Prieur, **On L^∞ stabilization of semilinear hyperbolic systems by saturated boundary control**, *ESAIM : Control, Optim. Cal. Var.*, vol. 26 (23) (2020).

Abstract.

Ce chapitre considère un système diagonal d'équations aux dérivées partielles hyperboliques semilinéaires à vitesses positives. La condition au bord est sous la forme d'un feedback saturé. Grâce à une approche semigroupe non-linéaire, on y établit la stabilisation locale exponentielle des solutions dans $L^\infty([0,1])$. De plus, une estimation du bassin d'attraction y est explicité.

La suite de ce chapitre est écrite en anglais.

Sommaire

2.1 Introduction

2.1.1 Literature Review

In this chapter, we are interested in the effect of saturation on systems of semilinear transport equations coupled at the boundary. To the best of our knowledge, the first work analyzing the effect of saturation in infinite-dimensional systems is [98]. In particular, in [98] the author focuses on the case of compact and bounded control operators, with an a priori constraint. The results in [76] suggest the use of an observability condition for the analysis of systems modeled by PDEs controlled via closed-loop saturating controllers. In particular, the contraction semigroup obtained from the saturating closed-loop system is compared with the corresponding saturation-free semigroup.

The literature on unsaturated linear boundary control for semilinear hyperbolic systems is rich ; see, e.g., [16] or [74], just to mention a few. However, when input saturation comes into play, the inherent nonlinear nature of the problem renders the analysis much harder. As such, only a few papers focused on saturated boundary control of systems modeled via PDEs. For example, in [80], the authors prove that two-dimensional quasilinear hyperbolic systems with opposite velocities are stabilizable with bounded C^1 boundary control inputs. Nevertheless, the method mainly relies on characteristics and does not seem to be generalizable to system with a larger number of PDEs. The results in [93] are tailored to the wave equation $z_{tt} = z_{xx}$ subject to a nonlinear saturated boundary condition ; which is a special second-order hyperbolic PDE. In particular, inspired by [10], the authors in [93] rely on the theory of nonlinear semigroups to prove well-posedness and global H^1 exponential stability for the wave equation, in the presence of distributed or boundary saturated controllers. The main idea consists of using a sector bounded approach inspired by the literature of finite dimensional systems [102], to ensure exponential decay of an H^1-Lyapunov functional.

Here, we directly consider the following system of semilinear hyperbolic PDEs of arbitrary dimension $d \in \mathbb{N}$

$$\begin{cases} \partial_t R + \Lambda \partial_x R = & g \circ R \\ R(t,0) = & HR(t,1) + Bu(t) \\ R(0,\cdot) = & R_0 \end{cases}$$

where Λ is a diagonal positive definite matrix, H and B are $d \times d$ real matrices, and $g \in C^1(\mathbb{R}^d)$ is a globally Lipschitz function, with Lipschitz constant L_g, such that $g(0) = 0$. Moreover, g is diagonal in the sense that for all $R \in \mathbb{R}^d$, $g_i(R) = g_i(R_i)$.

The open-loop system may turn out to be unstable if the matrix H is too "large" (in the sense of a certain norm). The source term g has also its impact on the stability. According to its form, it could make the open-loop system more or less stable.

In [15], authors found a sufficient condition on matrix $K \in M_d(\mathbb{R})$ such that the linear control $u(t) = KR(t,1)$ ensures C^1 exponential stability of the following "unsaturated" closed-loop system :

$$\begin{cases} \partial_t R + \Lambda \partial_x R = & 0 \\ R(\cdot,0) = & (H+BK)R(\cdot,1) \\ R(0,\cdot) = & R_0 \in C^1([0,1]). \end{cases} \tag{2.1}$$

In particular, recalling that for all matrices $M \in M_d(\mathbb{R})$, $|M|_\infty = \max_{i=1..d} \sum_{j=1}^d |M_{i,j}|$, [15, Theorem 3.3] established that if :

$$\rho_\infty(H+BK) := \inf_{P \in D_d^+(\mathbb{R})} |P(H+BK)P^{-1}|_\infty < 1$$

then the unsaturated system (2.1) is C^1 exponentially stable for the canonical norm of $C^1([0,1])$. Note that [15, Theorem 3.3] was proven for small initial data and for quasilinear systems.

2.1.2 Definition of the system and contribution

Here, it is assumed that there exists a matrix gain K such that $\rho_\infty(H+BK) < 1$ and we will study the L^∞ stability of the saturated closed-loop system :

$$\begin{cases} \partial_t R + \Lambda \partial_x R = & g \circ R \\ R(0,.) = & HR(1,.) + B\sigma(KR(1,.)) \\ R(.,0) = & R_0 \in L^\infty([0,1]) \end{cases} \tag{2.2}$$

with σ defined as a saturation by component *i.e.* there exists a $\sigma_s > 0$ such that for all $i \in [\![1,d]\!],\ x \in \mathbb{R}$,

$$\begin{cases} \sigma_i(x) = x & \text{if } |x| \leq \sigma_s \\ \sigma_i(x) = \text{sign}(x)\sigma_s & \text{otherwise.} \end{cases}$$

Remark 1. *For simplicity, we take σ such that the value of the saturation level σ_s is identical for each component. This is not a restriction with respect to the general case*

$$\begin{cases} \sigma_i(x) = x & \textit{if } |x| \leq \sigma_{s,i} \\ \sigma_i(x) = \textit{sign}(x)\sigma_{s,i} & \textit{otherwise} \end{cases}$$

where $(\sigma_{s,i})_{i\in[\![1,d]\!]} \in (\mathbb{R}^+)^d$.

The initial data R_0 being in $L^\infty([0,1])$, solutions of (2.2) has to be understood in a weak sense. The main contribution of this chapter is to answer the two following problems :

Problem 1. *Define the sense of a weak solution to system (2.2) and prove a well-posedness theorem.*

Problem 2. *Prove the L^∞ local exponential stability of this system with an estimation of the region of attraction.*

Problem 1 will be solved using a smooth approximation of the system (2.2) coming from a smoothed sequence of saturations. Convergence of semigroups allows to define weak solutions to (2.2) and prove the well-posedness. Problem 2 will be tackled using an approximation of the $L^\infty([0,1])$ norm by $L^p([0,1])$ norms ($p \in \mathbb{N}$).

The rest of this chapter is organized as follows. In Section 2.2, all main results are formulated by two theorems ; the first one states the well-posedness and the other, the exponential stability. In the same section, an estimation of the region of attraction is given. In Section 2.3, the estimated region of attraction for systems taken from the literature is compared with the region of non-saturation. Some concluding remarks and further orientations are given in Section 2.4.

2.2 Main results

In this section, results for well-posedness and exponential stability are stated.

2.2.1 Problem 1

To properly define a weak solution to system (2.2), we need to give a precise sense to the trace of this solution on the lines $s \mapsto (t = s, x = 0)$ and $s \mapsto (t = s, x = 1)$. To do so, smoothed solutions subject to smoothed saturations are used. Such smoothed saturations approximate σ in the sense of Definition 3.

Definition 3. *$(\sigma_n)_n$ is a smooth approximation of σ if it is in $C^1(\mathbb{R})$ and converges uniformly to σ on $\mathbb{R}$.*

Remark 2. *An example of smoothed saturation $(\sigma_n)_n$ (with n an integer) approximation of σ is defined by :*

$$\begin{cases} \sigma'_{n,i}(x) = & 1 \text{ if } x \in [-\sigma_s, \sigma_s] \\ \sigma'_{n,i}(x) = & \frac{1}{2} + \frac{\cos(n(x-\sigma_s))}{2} \quad \text{if } x \in [\sigma_s, \sigma_s + \pi/n] \\ \sigma'_{n,i}(x) = & \frac{1}{2} + \frac{\cos(n(x+\sigma_s))}{2} \quad \text{if } x \in [-\sigma_s, -\sigma_s - \pi/n] \\ \sigma'_{n,i}(x) = & 0 \text{ otherwise} \end{cases}$$

and σ_n which is a primitive of σ'_n, is chosen as :

$$\begin{cases} \sigma_{n,i}(x) = & x \text{ if } x \in [-\sigma_s, \sigma_s] \\ \sigma_{n,i}(x) = & \frac{x+\sigma_s}{2} + \frac{\sin(n(x-\sigma_s))}{2n} \quad \text{if } x \in [\sigma_s, \sigma_s + \pi/n] \\ \sigma_{n,i}(x) = & \frac{x-\sigma_s}{2} + \frac{\sin(n(x+\sigma_s))}{2n} \quad \text{if } x \in [-\sigma_s, -\sigma_s - \pi/n] \\ \sigma_{n,i}(x) = & \sigma_s + \frac{\pi}{2n} \text{ if } x > \sigma_s + \frac{\pi}{n} \\ \sigma_{n,i}(x) = & -\sigma_s - \frac{\pi}{2n} \text{ if } x < -\sigma_s - \frac{\pi}{n}. \end{cases}$$

It is easy to show that $(\sigma_n)_n$ tends uniformly towards σ on $\mathbb{R}$. We represent both sequences $(\sigma_n)_n$ and $(\sigma'_n)_n$ in Figure 2.1.

Remark 3. *In this chapter and for every approximation of σ, the maximum of σ_n will be denoted $\sigma_{s,n}$ for all integers n.*

Now, taking a smooth approximation $(\sigma_n)_n$ of σ and an integer n, we define another system "smoother" than (2.2). The system subject to the saturation σ_n is defined by :

$$\begin{cases} \partial_t R_n + \Lambda \partial_x R_n = & g \circ R_n \\ R_n(\cdot, 0) = & H R_n(\cdot, 1) + B\sigma_n(K R_n(\cdot, 1)) \\ R_n(0, \cdot) = & R_{0,n} \in H^2([0,1]). \end{cases} \tag{2.3}$$

Thanks to Theorem 3 given in Appendix, we will show that if the initial data is $H^2([0,1])$ and satisfies compatibility conditions of order 1 (2.4) then the previous system of PDEs has a unique solution in $C^0([0,T], H^1([0,1])) \cap C^1([0,T], L^2([0,1]))$ for any $T > 0$. Hence, traces of this unique solution $R_n(0,.)$ and $R_n(1,.)$ are well-defined on the almost everywhere sense.

Remark 4. *Note that compatibility conditions of order 1 depends on the chosen σ_n. They are expressed as follows*

$$\begin{cases} R_0(0) = & H R_0(1) + B\sigma_n(K R_0(1)) \\ R'_0(0) = & \Lambda^{-1}\Big([H + B\sigma'_n(R_0(1))](\Lambda R'_0(1) - g(R_0(1))) + g(R_0(0))\Big). \end{cases} \tag{2.4}$$

The definition of weak solutions is given here.

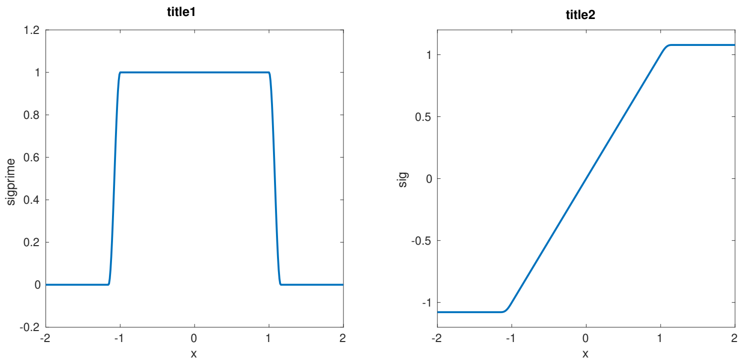


FIGURE 2.1 – Functions σ_n' and σ_n for $n = 20$ and $\sigma_s = 1$

Definition 4. *For all $T > 0$, $R \in C^0([0,T], L^2([0,1]))$ is a weak solution to the problem :*

$$\begin{cases} \partial_t R + \Lambda \partial_x R = & g \circ R \\ R(\cdot, 0) = & HR(\cdot, 1) + B\sigma(KR(\cdot, 1)) \\ R(0, \cdot) = & R_0 \in L^2([0,1]), \end{cases} \tag{2.5}$$

if there exists a sequence $(\sigma_n)_n$, smooth approximation of σ, and a sequence $(R_{0,n})_n$ in $H^2([0,1])$ satisfying the compatibility conditions of order 1 (2.4) *(which depend on the saturation σ_n chosen) tending towards R_0 in $L^2([0,1])$, such that the sequence of solutions $(R_n)_n$ to (2.3) converges towards R in $C^0([0,T], L^2([0,1]))$.*

Remark 5. *This definition is different from the common definition of a weak L^2 solution [16, Definition A.3]. The adjoint (for the usual $L^2([0,T])$ scalar product) of the boundary operator $f \mapsto Hf + B\sigma(Kf)$ may not exist. As a consequence, it is impossible to define a boundary condition on test functions. Therefore, we cannot use the common notion of weak solutions. In [12], authors proved the well-posedness of quasilinear scalar problems subject to L^∞ boundary conditions. They used the method of vanishing viscosity to prove the existence and the uniqueness of the weak solution. The method consists of using a regularized system with additional viscosity and pass to the limit in the weak formulation of the PDE considered. Here, we use the same idea : a regularized system is considered and by a passage to the limit, the weak solution is defined.*

It turns out that this problem is well-posed in the sense of Hadamard :

Theorem 1 (Well-Posedness)**.** *There exists a unique weak solution to problem (2.5). Moreover, the flow operator defined by :*

$$U_T : \begin{cases} L^2([0,1]) & \to & C^0([0,T], L^2([0,1])) \\ R_0 & \mapsto & R \end{cases}$$

is continuous for all $T > 0$.

This theorem is proven in Appendix A.

Remark 6. Theorem 1 holds for whatever σ bounded and continuous such that there exists a smooth approximation of σ in the sense of Definition 3. Moreover, Theorem 1 is also valid for a nondiagonal (but Lipschitz) source term.

2.2.2 Problem 2

For exponential stability, we introduce the Lipschitz constants $L_{g,i}$ such that

$$\forall R \in \mathbb{R}^d,\ i \in [\![1,d]\!],\ |g_i(R_i)| \le L_{g,i}|R_i|. \tag{2.6}$$

Hence, for all integers i, the scalar function $R \mapsto g_i(R)$ is a scalar function from $\mathbb{R}$ to $\mathbb{R}$ which is $L_{g,i}$ Lipschitz. Then defining

$$\forall f \in L^\infty([0,1]),\ V(f) := \max_{i\in[\![1,d]\!]} |P_i f_i e^{-\mu x}|_{L^\infty([0,1])} \tag{2.7}$$

where $P = \mathrm{diag}(P_i)$ is selected such that $|P(H+BK)P^{-1}|_\infty \le 1$ (possible because $\rho_\infty(H+BK) < 1$ is assumed all along this article). One gets the following result :

Theorem 2 (Exponential Stability)**.** *Suppose $\rho_\infty(H+BK) < 1$. For all $P \in D_d^+(\mathbb{R})$ satisfying $|P(H+BK)P^{-1}| < 1$, all positive $\mu < -\log(|P(H+BK)P^{-1}|_\infty)$ and for all initial data $R_0 \in L^\infty([0,1])$, if*

$$\mu\lambda_{\min} - L_{g,\max} \ge 0$$

where $L_{g,\max} := \max_{i\in[\![1,d]\!]} L_{g,i}$,

and

$$V(R_0) < e^{-\mu}\frac{|PBP^{-1}|_\infty \sigma_s P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu}|}, \tag{2.8}$$

then, the weak solution to (2.5) verifies :

$$\forall t \ge 0,\ V(R(t,\cdot)) \le e^{-(\mu\lambda_{\min} - L_{g,\max})t} V(R_0).$$

This theorem is proven in Appendix B.

Remark 7. *As*

$$\forall R \in L^\infty([0,1]),\ \min_i \{P_i\}\, e^{-\mu} \|R\|_{L^\infty([0,1])} \le V(R) \le \max_i \{P_i\}\, \|R\|_{L^\infty([0,1])},$$

Theorem 2 gives the classical local exponential stability of system (2.5) *with respect to the usual $L^\infty([0,1])$ norm.*

Remark 8. *Concerning the estimation of the region of attraction, we will see in the proof of Theorem 2 (Remark 11) that such an estimation (2.8) comes from a sector bounded condition imposed on the dead-zone function ; the difference between the linear and the saturated control.*

2.3 Numerical example

In this section, we analyze a typical example of diagonal semilinear systems taken from [92] ; an article considering the same kind of systems plus a disturbance ; discarded here for our purposes. Matrices are defined as :

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{pmatrix},\ H = \begin{pmatrix} 0 & 1.1 \\ 1 & 0 \end{pmatrix},\ B = I_2.$$

In [92], authors consider a system of transport PDEs with positive velocities and without source term. They give a method to find a gain matrix K such that the equivalent linear system subject to the control $R(t,0) = (H + BK)R(t,1)$ be L^2 exponentially stable and robust. Three gain matrices K were compared with an increasing rate (or at least an estimation of this rate) of exponential decay for the L^2 norm :

$$K_1 = \begin{pmatrix} 0 & -0.1050 \\ -0.1045 & 0 \end{pmatrix}$$

$$K_2 = \begin{pmatrix} 0 & -0.4777 \\ -0.4651 & 0 \end{pmatrix}.$$

The last gain is taken from [52] :

$$K_3 = \begin{pmatrix} 0 & -0.7 \\ -1 & 0 \end{pmatrix}.$$

For all these matrix gains, we evaluate the region of attraction thanks to the estimation given in Theorem 2 and evaluate if it is larger than the domain where the saturation does not apply.

We take $\mu = 0$ in (2.8) and approximate the stability region of Theorem 2 by :

$$V(R_0) < \frac{|PBP^{-1}|_\infty \sigma_s P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - 1|} \tag{2.9}$$

equivalent to :

$$\forall i \in [\![1,d]\!],\ |R_{0,i}|_{L^\infty([0,1])} < |R_{0,i}|_{L^\infty,\lim}$$

where $|R_{0,i}|_{L^\infty,\lim} := \frac{1}{P_i} \frac{|PBP^{-1}|_\infty \sigma_s P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - 1|}$.

Taking $P \in D_d^+(\mathbb{R})$ minimizing $|P(H+BK)P^{-1}|_\infty$ and a saturation such that $\sigma_s = 1$, one gets the numerical results from Table 2.1 for the estimation of the region of attraction :

K	K_1	K_2	K_3
$\rho_\infty(H+BK)$	0.95	0.58	$\simeq 0$
$\lvert R_{0,1}\rvert_{L^\infty,\lim}$	17.5	11.0	$\simeq 0$
$\lvert R_{0,2}\rvert_{L^\infty,\lim}$	16.6	10.1	$\simeq 0$
$\lvert (KR_0)_1\rvert_{\lim}$	1.84	5.61	$\simeq 0$
$\lvert (KR_0)_2\rvert_{\lim}$	1.93	5.91	$\simeq 0$

Table 2.1 – The estimation of the region of attraction

where $|(KR_0)_1|_{\lim} := \sup\left\{|(KR)_1| \mid |R_i| < |R_{0,i}|_{L^\infty,\lim},\ \forall i \in [\![1,2]\!]\right\}$ and $|(KR_0)_2|_{\lim} := \sup\left\{|(KR)_2| \mid |R_i| < |R_{0,i}|_{L^\infty,\lim},\ \forall i \in [\![1,2]\!]\right\}$.

We added a row giving the values of $\rho_\infty(H+BK)$ as it gives an estimation of the rate of convergence $\mu\lambda_{\min} - L_{g,\max}$ of the L^∞ norm of the solution R. This can be seen from the condition "$\mu < -\log(|P(H+BK)P^{-1}|_\infty)$" of Theorem 2.

Note that for K_1 and K_2, $|(KR_0)_1|_{\lim}$ and $|(KR_0)_2|_{\lim}$ are both larger than the saturation $\sigma_s = 1$ and hence, the estimated region of attraction is larger than the linear unsaturated region. Then, we also remark that there exists a balance between the rate of convergence of the saturated system estimated by $\rho_\infty(H + BK)$ and the region of attraction. Keep in mind that the smaller $\rho_\infty(H + BK)$, the larger the estimation of the rate of exponential convergence of the saturated system. From this and results presented in Table 2.1, if one wants a vast region of attraction, the estimation of the rate of convergence will not be important. On the contrary, if one wants a strong rate of convergence, then the region of attraction will be limited.

Another comment has to be made on the case $K = K_3$. Here the matrix gain K_3 is chosen such that $\rho_\infty(H+BK) \simeq 0$ which means that the system is exponentially stable with a very large rate of convergence. From Table 2.1, the estimation of the region of attraction gives bad results. This is mainly because $|PB^+K^+P^{-1}|_\infty \simeq +\infty$. This last analysis tends to confirm the link between the estimation of the rate of convergence and the estimation of the region of attraction underlined earlier.

2.4 Conclusion

The well-posedness and the local L^∞ exponential stability of a wide class of diagonal semi-linear systems was established. The PDEs under consideration resulted from a transport with constant velocities coupled with a nonlinear source term and a nonlinear boundary condition. The saturated control was applied at the boundary in order to stabilize the open-loop system. The well-posedness was tackled using nonlinear semigroup techniques. The stability has been proven using convergence of semigroups and Lyapunov theory.

This work let some questions open. The case of mixed positive and negative velocities is not treated; the method of [31] which differentiates the Lyapunov functional for components with positive and negative velocities seems to be a promising idea. The case of space varying velocities (with constant sign) is also interesting and already solved in[16, Chapter 3.5] for unsaturated systems. The generalization to non-diagonal source terms would be an important improvement. The article [64], where a specific space dependent Lyapunov functional is introduced, would be a good starting point to tackle the problem. Finally, the L^∞ (or even L^p) stability for systems of nonlinear d scalar conservation laws remains an open question even for unsaturated controllers. In [19], a feedback control was found for a single scalar conservation law whose flux is either convex or concave. Additionally, authors of [35] study the stabilization of a nonlocal one-dimensional conservation law. Starting from a linearized system, they find a sufficient condition for stability and adapt the proof to the full nonlinear PDE. However, to our knowledge, nothing seems to be generalizable to conservation laws of arbitrary dimension.

2.A Proof of Theorem 1

Let $X = L^2([0, 1])$ be the base space; the scalar product on X was introduced by [31] and is defined by :

$$\forall u, v \in X,\ (u, v) := \int_0^1 u^T \overline{v} e^{\nu(x-1)} dx. \tag{2.10}$$

2.A.1 Existence and uniqueness of solution with a smoothed saturation

Take an arbitrary smooth approximation $(\sigma_n)_n$ of σ in the sense of Definition 3. For all integers n, we define the operator A_n by :

$$\begin{cases} A_n R = & -\Lambda R' \\ D(A_n) = & \left\{R \in H^1([0,1]); R(0) = HR(1) + B\sigma_n(KR(1))\right\}. \end{cases}$$

Moreover, the operator G can be defined as follows :

$$\begin{cases} GR = & g \circ R \\ D(G) = & L^2([0,1]) \end{cases}$$

The following theorem states the well-posedness for the closed-loop system whose control is smoothly saturated.

Theorem 3. *There exists $\zeta > 0$ dependent on σ_s, H, B and K such that for all integers n, the operator $A_n + G$ is ζ dissipative. Moreover, $A_n + G$ generates a semigroup T_n of type ζ and for all $R_{0,n} \in D(A_n)$, $T_n(.)R_{0,n}$ is the $C^0([0,T], L^2([0,1]))$ solution to the Cauchy problem :*

$$\begin{cases} \partial_t R_n & = -\Lambda \partial_x R_n + g \circ R_n \\ R_n(t=0) & = R_{0,n} \\ R_n(t) & \in D(A_n) \end{cases} \tag{2.11}$$

where $\partial_t R_n$ is defined as the Fréchet derivative with respect to t in the L^2 space :

$$\left\|\frac{R_n(t+dt,\cdot) - R_n(t,\cdot)}{dt} - \partial_t R_n(t,\cdot)\right\|_{L^2([0,1])} \to_{dt \to 0} 0$$

Finally, if $R_{0,n}$ is $H^2([0,1])$ and satisfies compatibility conditions of order 1 (2.4) *then the solution R_n belongs to $C^1([0,T], L^2([0,1])) \cap C^0([0,T], H^1([0,1]))$.*

Remark 9. *A definition of the ζ dissipativity can be found in [86]. Moreover, in the following proof, we do not use the form of σ_n but the fact that it is bounded in C^1 (n fixed). Hence, conclusions of Theorem 3 are valid for whatever bounded $\sigma_n \in C^1(\mathbb{R})$.*

Let n be an integer. To get the conclusions of Theorem 3, we need to prove some properties on the operators $A_n + G$. They are listed below :

— $A_n + G$ is ζ dissipative with ζ independent on n.

— It satisfies the range condition : $\mathrm{Rg}(I - \rho(A_n + G)) \supset D(A_n + G)$ for all positive ρ sufficiently small.

— $A_n + G$ is closed

Having done that, we use [86, Theorem 5.12] to prove that $A_n + G$ generates the expected semigroup.

Proof. Let n be an integer fixed all along the proof. Constants may depend on n but, in the following, this dependence is skipped when it is useless.

(1) $\mathbf{A_n + G}$ is ζ dissipative :

Let u and v in $D(A_n + G) = D(A_n)$. Recalling the definition of the scalar product (2.10), one has

$$((A_n+G)u-(A_n+G)v, u-v) = -\int_0^1 (u'-v')^T\Lambda(\overline{u-v})e^{\nu(x-1)}dx+\int_0^1 (g\circ u-g\circ v)^T(\overline{u-v})e^{\nu(x-1)}dx \tag{2.12}$$

By an integration by parts, one has :

$$\begin{aligned}((A_n + G)u - (A_n + G)v, u - v) = &-[(u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}]_0^1 + \int_0^1 (u - v)^T\Lambda(\overline{u'-v'})e^{\nu(x-1)}dx \\ &+\nu\int_0^1 (u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}dx + \int_0^1 (g \circ u - g \circ v)^T(\overline{u-v})e^{\nu(x-1)}dx.\end{aligned}$$

From (2.12), one gets :

$$\begin{aligned}((A_n + G)u - (A_n + G)v, u - v) = &-[(u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}]_0^1 - \overline{((A_n + G)u - (A_n + G)v, u - v)} \\ &+\nu\int_0^1 (u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}dx + 2\mathrm{Re}\int_0^1 (g \circ u - g \circ v)^T(\overline{u-v})e^{\nu(x-1)}dx.\end{aligned}$$

It implies necessarily that :

$$\begin{aligned}2\mathrm{Re}((A_n + G)u - (A_n + G)v, u - v) = &- [(u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}]_0^1 \\ &+ \nu\int_0^1 (u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}dx \\ &+ 2\mathrm{Re}\int_0^1 (g \circ u - g \circ v)^T(\overline{u-v})e^{\nu(x-1)}dx\end{aligned}$$

and taking the real part in last equation :

$$\begin{aligned}2\mathrm{Re}((A_n + G)u - (A_n + G)v, u - v) = -\mathrm{Re}\left([(u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}]_0^1\right) \\ + \nu\ \mathrm{Re}\int_0^1 (u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}dx + 2\mathrm{Re}\int_0^1 (g \circ u - g \circ v)^T(\overline{u-v})e^{\nu(x-1)}dx.\end{aligned} \tag{2.13}$$

Using the fact that velocities are bounded from above,

$$\begin{aligned}2\mathrm{Re}((A_n + G)u - (A_n + G)v, u - v) \leq &- \mathrm{Re}\left([(u - v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}]_0^1\right) + \nu\lambda_{\max}\mathrm{Re}(u - v, u - v) \\ &+ 2\mathrm{Re}\int_0^1 (g \circ u - g \circ v)^T(\overline{u-v})e^{\nu(x-1)}dx.\end{aligned}$$

Using the fact that g is Lipschitz, there exists a constant $\varsigma > 0$ depending on L_g such that :

$$2\mathrm{Re}((A_n+G)u-(A_n+G)v, u-v) \leq -\mathrm{Re}[(u-v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}]_0^1+(\nu\lambda_{\max}+\varsigma)\mathrm{Re}(u-v, u-v). \tag{2.14}$$

To simplify the notation, $u(1)$ and $v(1)$ will be denoted, respectively, by u_1 and v_1 in following computations. Boundary terms can be rewritten as follows :

$$\begin{aligned}\mathrm{Re}\left([(u-v)^T\Lambda(\overline{u-v})e^{\nu(x-1)}]_0^1\right) &= \mathrm{Re}(u_1-v_1)^T\Lambda\mathrm{Re}(u_1-v_1)\\ &- e^{-\nu}\mathrm{Re}[H(u_1-v_1)+B(\sigma_n(Ku_1)-\sigma_n(Kv_1))]^T\Lambda\mathrm{Re}[H(u_1-v_1)+B(\sigma_n(Ku_1)-\sigma_n(Kv_1))]\\ &+\mathrm{Im}(u_1-v_1)^T\Lambda\mathrm{Im}(u_1-v_1)\\ &- e^{-\nu}\mathrm{Im}[H(u_1-v_1)+B(\sigma_n(Ku_1)-\sigma_n(Kv_1))]^T\Lambda\mathrm{Im}[H(u_1-v_1)+B(\sigma_n(Ku_1)-\sigma_n(Kv_1))].\end{aligned} \tag{2.15}$$

For $\nu=\nu_n$ large enough compared to the norm of H, B and $\sigma_{s,n}$, we deduce that

$$\mathrm{Re}\left([(u-v)^T\Lambda(\overline{u-v})e^{\nu_n(x-1)}]_0^1\right) \geq \frac{1}{2}(\mathrm{Re}(u_1-v_1)^T\Lambda\mathrm{Re}(u_1-v_1)+\mathrm{Im}(u_1-v_1)^T\Lambda\mathrm{Im}(u_1-v_1)) \geq 0$$

which implies with (2.14) that :

$$\begin{aligned}2\mathrm{Re}\left((A_n+G-\frac{\lambda_{\max}\nu_n+\varsigma}{2}I)u-(A_n+G-\frac{\lambda_{\max}\nu_n+\varsigma}{2}I)v, u-v\right)\\ \leq -\mathrm{Re}[(u-v)^T\Lambda(\overline{u-v})e^{\nu_n(x-1)}]_0^1 \leq 0\end{aligned} \tag{2.16}$$

and therefore A_n+G is $\frac{\lambda_{\max}\nu_n+\varsigma}{2}$ dissipative.

Recall that, ν_n was taken large enough compared to the norms of H, B and $\sigma_{s,n}$. As $(\sigma_{s,n})_n$ is a bounded sequence, we can take a sequence $(\nu_n)_n$ bounded from above by a positive real ν. As A_n+G is $\frac{\nu_n\lambda_{\max}+\varsigma}{2}$ dissipative and $\frac{\nu_n\lambda_{\max}+\varsigma}{2} \leq \frac{\nu\lambda_{\max}+\varsigma}{2}$, A_n+G is $\frac{\nu\lambda_{\max}+\varsigma}{2}=\zeta$ dissipative with ζ independent on n.

(2) $\mathbf{A_n+G}$ satisfies the range condition :

Let us now prove the following range condition :

$$\exists\rho_{sup}>0;\ \forall\rho\in(0,\rho_{sup}),\ \mathrm{Rg}(I-\rho(A_n+G))\supset D(A_n). \tag{2.17}$$

It is equivalent to prove that for all v in $D(A_n)$, there exists an element u in $D(A_n)$ such that :

$$\begin{cases} u+\rho\Lambda u'-\rho g(u)= & v\\ u(0)= & Hu(1)+B\sigma_n(Ku(1))\end{cases}$$

This property is the most difficult to prove. It consists of proving the existence of a solution to a nonlinear ODE with a nonlinear boundary condition. To prove the existence, first, we will deal with the nonlinear boundary condition and then, using a fixed point theorem, the nonlinear source term will be taken into account ; the method being inspired from [93] and [105].

(2.1) Taking into account the nonlinear boundary condition

Let us now prove the following range condition :

$$\forall (v_1, v_2) \in D(A_n + G) = D(A_n),\ v_1 + \rho G v_2 \in \mathrm{Rg}(I - \rho A_n). \tag{2.18}$$

To do so, take v_1, v_2 both in $C^0([0,1])$. To prove assertion (2.18), we have to find an element u in $D(A_n)$ solution of :

$$\begin{cases} u + \rho\Lambda u' = & v_1 + \rho(g \circ v_2) \\ u(0) = & Hu(1) + B\sigma_n(Ku(1)) \end{cases}$$

equivalent to :

$$\begin{cases} u' + \frac{\Lambda^{-1}}{\rho} u = & \Lambda^{-1}\big(\frac{v_1}{\rho} + g \circ v_2\big) \\ u(0) = & Hu(1) + B\sigma_n(Ku(1)). \end{cases} \tag{2.19}$$

We define $\mathcal{T} : C^0([0,1]) \to C^1([0,1])$

$$\begin{aligned} \mathcal{T} : C^0([0,1]) & \to C^1([0,1]) \\ y \quad & \mapsto \text{ u solution of the following system} \end{aligned}$$

$$\begin{cases} u' + \frac{\Lambda^{-1}}{\rho} u = & \Lambda^{-1}\big(\frac{v_1}{\rho} + g \circ v_2\big) \\ u(0) = & Hu(1) + B\sigma_n(Ky(1)). \end{cases} \tag{2.20}$$

If we prove that $\mathcal{T}$ is well-defined and admits a fixed point in $D(A_n)$, then assertion (2.18) is proven.

For all y in $C^0([0,1])$, solutions u to the ODE in (2.20) are $C^1([0,1])$ (because $v_1, g(v_2) \in C^0([0,1])$) and can be expressed as follow :

$$\forall x \in [0,1],\ u(x) = e^{-\frac{\Lambda^{-1}x}{\rho}} Z(y) + \int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)} \times \Lambda^{-1}\left(\frac{v_1}{\rho} + g \circ v_2\right) ds$$

where $Z(y)$ is a constant of $\mathbb{R}^d$. Thus, for all y in $C^0([0,1])$, u is a solution to system (2.20) if and only if $Z(y)$ satisfies the following equation :

$$Z(y) = He^{-\frac{\Lambda^{-1}}{\rho}} Z(y) + H\int_0^1 e^{-\frac{\Lambda^{-1}}{\rho}(1-s)} \times \Lambda^{-1}\left(\frac{v_1}{\rho} + g \circ v_2\right) ds + B\sigma_n(Ky(1))$$

equivalent to :

$$(I_d - He^{-\frac{\Lambda^{-1}}{\rho}})Z(y) = H\int_0^1 e^{-\frac{\Lambda^{-1}}{\rho}(1-s)} \times \Lambda^{-1}\left(\frac{v_1}{\rho} + g \circ v_2\right) ds + B\sigma_n(Ky(1))$$

and for ρ sufficiently small, say $\rho < \theta(H, \Lambda)$ with $\theta(H, \Lambda) > 0$, one can invert the last relation :

$$\forall y \in C^0([0,1]),\ Z(y) = (I_d - He^{-\frac{\Lambda^{-1}}{\rho}})^{-1}\left(H\int_0^1 e^{-\frac{\Lambda^{-1}}{\rho}(1-s)} \times \Lambda^{-1}\left(\frac{v_1}{\rho} + g \circ v_2\right) ds + B\sigma_n(Ky(1))\right). \tag{2.21}$$

Last equation ensures that for $\rho < \theta(H, \Lambda)$, $\mathcal{T}$ is well defined and

$$\forall x \in [0,1],\ \mathcal{T}(y)(x) = e^{-\frac{\Lambda^{-1}x}{\rho}} Z(y) + \int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)} \times \Lambda^{-1}\left(\frac{v_1}{\rho} + g \circ v_2\right) ds$$

with $Z(y)$ defined in (2.21).

The operator $\mathcal{T}$ being well-defined, we can focus on the fixed point argument. As σ_n is bounded, $Z(y)$ is bounded when y scans $C^0([0,1])$. Hence, $\mathcal{T}(C^0([0,1]))$ is a bounded set of $C^1([0,1])$ and there exists a set $K = \{w \in C^1([0,1]); \|w\|_{C^0([0,1])} \leq M \text{ and } \|w'\|_{C^0([0,1])} \leq M\}$ where M is a constant such that $\mathcal{T}(C^0([0,1])) \subset K$. As K is bounded in $C^1([0,1])$, K is compact in $C^0([0,1])$ by Ascoli-Arzela's theorem. Moreover, $\mathcal{T}$ is continuous and K is closed and convex allowing to use Schauder fixed point theorem to conclude that :

$$\exists u \in K : \mathcal{T}(u) = u.$$

Hence, $u \in D(A_n) = D(A_n + G)$ and the assertion (2.18) is proven.

(2.2) Taking into account the source term

Let v be in $D(A_n + G) = D(A_n)$, $\rho \leq \theta(H, \Lambda)$ and $\mathcal{H} : C^0([0,1]) \mapsto C^0([0,1])$ be such that for all w in $C^0([0,1])$, $\mathcal{H}(w)$ is solution of :

$$\begin{cases} u' + \frac{\Lambda^{-1}}{\rho} u = & \Lambda^{-1}(\frac{v}{\rho} + g \circ w) \\ u(0) = & Hu(1) + B\sigma_n(Ku(1)). \end{cases}$$

By assertion (2.18), $\mathcal{H}$ is well defined (take $v_1 \leftarrow v$ and $v_2 \leftarrow w$). We will prove that $\mathcal{H}$ has a fixed point in $D(A_n)$ which implies that the range condition (2.17) is verified. To do so, we will prove that there exists a ball of C^0 ; B_r of radius r such that B_r is invariant under $\mathcal{H}$ and $\mathcal{H}(B_r)$ precompact in $C^0([0,1])$.

(2.2.1) There exists a ball of $C^0([0,1])$, $\mathbf{B_r}$, invariant under $\mathcal{H}$

Let w be in $C^0([0,1])$ and let recall the definition of the usual $C^0([0,1])$ norm :

$$\|w\|_{C^0([0,1])} := \max_{x\in[0,1]} |w(x)| = \max_{x\in[0,1]} \sqrt{\sum_{i=1}^{d} w_i(x)^2}. \tag{2.22}$$

The continuous function $\mathcal{H}(w)$ can be expressed as :

$$\forall w \in C^0([0,1]),\ \forall x \in [0,1],\ \mathcal{H}(w)(x) = e^{-\frac{\Lambda^{-1}x}{\rho}} Z(w) + \int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)} \times \Lambda^{-1}\left(\frac{v}{\rho} + g \circ w\right) ds \tag{2.23}$$

where :

$$\forall w \in C^0([0,1]),\ Z(w) = (I_d - He^{-\frac{\Lambda^{-1}}{\rho}})^{-1}\Big(H \int_0^1 e^{-\frac{\Lambda^{-1}}{\rho}(1-s)} \times \Lambda^{-1}\big(\frac{v}{\rho} + g \circ w\big)ds + B\sigma_n(K\mathcal{H}(w)(1))\Big). \tag{2.24}$$

Fixing an w in $C^0([0,1])$, we focus on a term present in both expressions of $Z(w)$ and of $\mathcal{H}(w)$:

$$\forall x \in [0,1],\ \int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)} \times \frac{\Lambda^{-1}}{\rho} v\, ds = v(x) - e^{-\frac{\Lambda^{-1}}{\rho}x} v(0) - \int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)} v'\, ds.$$

Hence,

$$\forall x\in[0,1],\ \left|\int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)}\times\frac{\Lambda^{-1}}{\rho}v\ ds\right| = \sqrt{\sum_{i=1}^d\left(\int_0^x e^{-\frac{\lambda_i^{-1}}{\rho}(x-s)}\times\frac{\lambda_i^{-1}}{\rho}v_i\ ds\right)^2}$$

$$=\sqrt{\sum_{i=1}^d\left(v_i(x)-e^{-\frac{\lambda_i^{-1}}{\rho}x}v_i(0)-\int_0^x e^{-\frac{\lambda_i^{-1}}{\rho}(x-s)}v_i'\ ds\right)^2}$$

$$\le\sqrt{\sum_{i=1}^d v_i(x)^2}+\sqrt{\sum_{i=1}^d e^{-\frac{2\lambda_i^{-1}}{\rho}x}v_i^2(0)}$$

$$+\sqrt{\sum_{i=1}^d\left(\int_0^x e^{-\frac{\lambda_i^{-1}}{\rho}(x-s)}v_i'\ ds\right)^2}$$

and because $e^{-\frac{\Lambda^{-1}}{\rho}(x-s)}\le I_d$ when $s\le x$:

$$\forall x\in[0,1],\ \left|\int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)}\times\frac{\Lambda^{-1}}{\rho}v\ ds\right|\le 2\|v\|_{C^0([0,1])}+\|v\|_{H^1([0,1])}=:C(v). \tag{2.25}$$

Remark that $C(v)$ is independent on ρ, w.

The second term to study is :

$$\forall x\in[0,1],\ \left|\int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)}\times\Lambda^{-1}(g\circ w)\ ds\right|^2 = \sum_{i=1}^d\left(\int_0^x e^{-\frac{\lambda_i^{-1}}{\rho}(x-s)}\times\lambda_i^{-1}(g_i\circ w)\ ds\right)^2$$

$$\le\sum_{i=1}^d|g_i\circ w|^2_{L^2([0,1])}\frac{\rho}{2\lambda_i}(1-e^{-2\frac{\lambda_i^{-1}}{\rho}x})$$

$$\le\sum_{i=1}^d\frac{\rho}{2\lambda_i}|g_i\circ w|^2_{L^2([0,1])}\le\frac{\rho}{2\min_i\lambda_i}\|g\circ w\|^2_{C^0([0,1])}$$

where we have used Cauchy-Schwarz inequality to get first inequality.

Using the fact that $g(0)=0$ and g Lipschitz as a function from $\mathbb{R}^d\mapsto\mathbb{R}^d$ for the canonical $\mathbb{R}^d$ norm, one gets :

$$\forall w\in C^0([0,1]),\|g\circ w\|_{C^0([0,1])}\le\max_{x\in[0,1]}|(g\circ w)(x))-g(0)|+|g(0)|\le L_g\max_{x\in[0,1]}|w(x)|.$$

Hence,

$$\forall w\in C^0([0,1]),\|g\circ w\|_{C^0([0,1])}\le L_g\|w\|_{C^0([0,1])}$$

and as a consequence :

$$\forall w\in C^0([0,1]),x\in[0,1],\ \left|\int_0^x e^{-\frac{\Lambda^{-1}}{\rho}(x-s)}\times\Lambda^{-1}(g\circ w)\ ds\right|\le\sqrt{\frac{\rho}{2\min_i\lambda_i}}L_g\|w\|_{C^0([0,1])}. \tag{2.26}$$

Finally, using the boundedness of σ_n, one gets the existence of a constant C_3 (only dependent on σ_s, H, B) such that $\forall \rho > 0$:

$$|(I_d - He^{-\frac{\Lambda^{-1}}{\rho}})^{-1} B\sigma_n(K\mathcal{H}(w)(1))| \leq C_3. \tag{2.27}$$

Injecting inequalities (2.25), (2.26) and (2.27) in the definition of $Z(w)$ (2.24), one gets the existence of $C_1(v), C_2, C_3$ independent on ρ, w such that :

$$\forall w \in C^0([0,1]), |Z(w)| \leq C_1(v) + C_2\sqrt{\rho}\|w\|_{C^0([0,1])} + C_3. \tag{2.28}$$

Putting all together inequalities (2.25), (2.26) and (2.28) in the expression of $\mathcal{H}(w)$ (2.23), there exists three positive constants $\tilde{C_1}(v)$, $\tilde{C_2}$, $\tilde{C_3}$ independent on ρ, w such that :

$$\forall w \in C^0([0,1]), \|\mathcal{H}(w)\|_{C^0([0,1])} \leq \tilde{C_1}(v) + \tilde{C_2}\sqrt{\rho}\|w\|_{C^0([0,1])} + \tilde{C_3}. \tag{2.29}$$

Taking $\rho < \frac{1}{\tilde{C_2}^2}$ (note that this bound is independent on v), it is possible to choose a strict positive r such that $\tilde{C_1}(v) + \tilde{C_2}\sqrt{\rho}r + \tilde{C_3} \leq r$ and :

$$\forall w \in B_r, \ \|\mathcal{H}(w)\|_{C^0([0,1])} \leq r.$$

which is no more than B_r is invariant under $\mathcal{H}$.

(2.2.2) The set $\mathcal{H}(B_r)$ is precompact in $\mathbf{C^0}([\mathbf{0},\mathbf{1}])$

The following claim allows to get compactness :

Claim 1. *The set $\mathcal{H}(B_r)$ is uniformly bounded in $C^1([0,1])$.*

Proof of Claim 1. Let $w \in B_r$. Denoting $u = \mathcal{H}(w)$, we have

$$u' = -\frac{\Lambda^{-1}}{\rho}u + \Lambda^{-1}\Big(\frac{v}{\rho} + g \circ w\Big).$$

As u, v, w are all continuous, u' is also continuous. By previous section, $u \in B_r$ and as a consequence

$$\|u'\|_{C^0([0,1])} \leq C$$

where C may depend on $(r, \rho, \lambda, L_g, v)$ but not on w. This ends the proof of Claim 1. □

To conclude, using Ascoli-Arzela's theorem, $\mathcal{H}(B_r)$ is relatively compact in $C^0([0,1])$ and by Schauder fixed point theorem, $\mathcal{H}$ admits a fixed point which ends the proof of the range condition (2.17).

(3) $\mathbf{A_n} + \mathbf{G}$ is a closed operator :

To prove this, we first prove that A_n is closed. Then, using the continuity of G, we conclude on the closedness of the operator $A_n + G$.

(3.1) $\mathbf{A_n}$ is a closed operator :

Take a sequence $(u_k)_k$ of elements of $D(A_n)$ such that $\lim_{k\to\infty} u_k =: u$ in $L^2([0,1])$ and $\lim_{k\to\infty} A_n u_k =: \tilde{u}$ in $L^2([0,1])$. We have to show that u belongs to $D(A_n)$ and that $\tilde{u} = A_n u$. Let us define for all integers k, $\tilde{u}_k := A_n u_k \to_{L^2([0,1])} \tilde{u}$ which can also be written as :

$$\forall k \in \mathbb{N},\ u_k' = -\Lambda^{-1}\tilde{u}_k \to_{L^2([0,1])} -\Lambda^{-1}\tilde{u}. \tag{2.30}$$

As a consequence,

$$\forall k \in \mathbb{N}, \forall x \in [0,1], u_k(x) = u_k(0) - \int_0^x \Lambda^{-1}\tilde{u}_k(s)ds = Hu_k(1) + B\sigma_n(Ku_k(1)) - \int_0^x \Lambda^{-1}\tilde{u}_k(s)ds.$$

Let x be in $[0,1]$, we have $|\int_0^x \Lambda^{-1}(\tilde{u}_k(s) - \tilde{u}(s))ds| \leq C \times \|\tilde{u}_k - \tilde{u}\|_{L^2([0,1])}$ which tends to zero as k tends towards infinity. Hence,

$$\forall x \in [0,1],\ \int_0^x \Lambda^{-1}\tilde{u}_k(s)ds \xrightarrow[k\to\infty]{} \int_0^x \Lambda^{-1}\tilde{u}(s)ds. \tag{2.31}$$

Then, as $H^1([0,1]) \subset C^0([0,1])$ continuously :

$$\forall k, m \in \mathbb{N},\ |u_k(1) - u_m(1)| \leq C \times \|u_k - u_m\|_{H^1([0,1])}$$

where C is the constant of the continuous injection $H^1([0,1]) \subset C^0([0,1])$.

Moreover, by (2.30), $\|u_k - u_m\|^2_{H^1([0,1])} = \|u_k - u_m\|^2_{L^2([0,1])} + \|\Lambda^{-1}\tilde{u}_k - \Lambda^{-1}\tilde{u}_m\|^2_{L^2([0,1])}$ which tends to zero as (k,m) tends towards infinity. Hence, $(u_k(1))_k$ is Cauchy and tends towards a real u_1. From this and the pointwise convergence stated in (2.31), $Hu_k(1) + B\sigma_n(Ku_k(1)) - \int_0^x \Lambda^{-1}\tilde{u}_k(s)ds$ converges pointwise towards a function denoted w (which can be different from u because the convergence of w is just pointwise) and :

$$\forall x \in [0,1],\ w(x) = Hu_1 + B\sigma_n(Ku_1) - \int_0^x \Lambda^{-1}\tilde{u}(s)ds. \tag{2.32}$$

Obviously, w is in $H^1([0,1])$ and $w_x = \Lambda^{-1}\tilde{u}$. The convergence of $(u_k)_k$ towards w is also in $L^2([0,1])$:

$$\begin{aligned}\int_0^1 (w(x) - u_k(x))^2 dx = &\int_0^1 \Big\{H(u_1 - u_k(1)) + B\sigma_n(Ku_1) - B\sigma_n(Ku_k(1)) \\ &+ \int_0^x \Lambda^{-1}(\tilde{u}(s) - \tilde{u}_k(s))ds\Big\}^2 dx \\ \leq &2 \times (H(u_1 - u_k(1)) + B\sigma_n(Ku_1) - B\sigma_n(Ku_k(1)))^2 \\ &+ 2 \times \int_0^1 \left(\int_0^x \Lambda^{-1}(\tilde{u}(s) - \tilde{u}_k(s))ds\right)^2 dx \\ \leq &C \times (|u_k(1) - u_1|^2 + \|\tilde{u}_k - \tilde{u}\|^2_{L^2([0,1])})\end{aligned}$$

where we have used the identity $(a+b)^2 \leq 2a^2 + 2b^2$ to get first inequality and C is a constant depending on $H, B, \sigma_{s,n}$ and Λ.

The right-hand side of last equation tending towards zero as k tends to infinity, we have proven that $(u_k)_k$ tends towards w in $L^2([0,1])$ and by the uniqueness of the limit $w = u$ in $L^2([0,1])$. Note that as w is continuous, we can take $w = u$ in the sense of $C^0([0,1])$. Moreover, as $u_k(1)$ tends towards $w(1) = u(1)$ (because $(u_k)_k$ tends towards w pointwise) and towards u_1 (by definition of u_1) at the same time, we have that $u(1) = u_1$. Injecting this last equality in (2.32), we have :

$$\forall x \in [0,1],\ u(x) = Hu(1) + B\sigma_n(Ku(1)) - \int_0^x \Lambda^{-1}\tilde{u}(s)ds.$$

Thus, $u \in D(A_n)$ and $\tilde{u} = A_n u$.

(3.2) $\mathbf{A_n + G}$ is a closed operator

As A_n is closed by previous paragraph and G is continuous as an operator from $L^2([0,1])$ into $L^2([0,1])$ (because g is Lipschitz), $A_n + G$ is closed.

(4) First conclusions on the proof of Theorem 3

As $A_n + G$ satisfies the range condition and is ζ dissipative ; by [86, Theorem 5.12], $A_n + G$ generates a unique semigroup T_n of type ζ. By Remark 2 p. 148 (and Theorem 4.10 (ii)) of the same book, the additional facts that $A_n + G$ is closed and $L^2([0,1])$ is reflexive ; if $R_{0,n} \in D(A_n)$ then $t \mapsto T_n(t)R_{0,n}$ is the unique solution of the Cauchy problem :

$$\left\{\begin{array}{rl} \partial_t R_n = & (A_n + G)R_n \\ R_n(t=0) = & R_{0,n} \\ R_n(t) \in & D(A_n) \end{array}\right.$$

in the sense that for almost every $t \geq 0$, $T_n(t)R_{0,n}$ is in $D(A_n)$, time-Fréchet differentiable in $L^2([0,1])$ and verifies the system presented just above.

As T_n is a semigroup of type ζ in $L^2([0,1])$, we have :

$$t \mapsto T_n(t)R_{0,n} \in C^0([0,T], L^2([0,1])).$$

(5) Regularity of the solution

Here we use the C^1 regularity assumptions on g and σ_n. The augmented system satisfied by $U_n := (R_n, \partial_x R_n)$ writes

$$\left\{\begin{array}{rl} \partial_t U_n + \mathrm{diag}(\Lambda,\Lambda)\partial_x U_n = & g_2 \circ U \\ R_n(0) = & HR_n(1) + B\sigma_n(KR_n(1)) \\ \partial_x R_n(0) = & \Lambda^{-1}\Big([H + B\sigma_n'(R_n(1))](\Lambda\partial_x R_n(1) - g(R_n(1))) + g(R_n(0))\Big) \end{array}\right. \tag{2.33}$$

where

$$\left\{\begin{array}{lll} g_2 : \mathbb{R}^{2d} & \mapsto & \mathbb{R}^{2d} \\ U := (U_1, U_2) & \to & (g(U_1), g'(U_1)U_2) \end{array}\right.$$

In substance, this augmented system is very similar to system (2.11). The transport part is identical and the source term is Lipschitz (g' is bounded by assumption). For the boundary condition of $\partial_x R_n$, it is linear in the variable $\partial_x R_n$. Knowing this, we can easily adapt the reasoning used for the system (2.11) for the variable R_n alone to the system (2.33), the unbounded operator being now defined as

$$\begin{cases} D(A_{2,n}) = \left\{ (U_1, U_2) \in H^1([0,1])^2 \quad \middle| \quad \begin{aligned} U_1(0) = &\ HU_1(1) + B\sigma_n(KU_1(1)), \\ U_2(0) = &\ \Lambda^{-1}\Big([H + B\sigma_n'(U_1(1))](\Lambda U_2(1) \\ &\ -g(U_1(1))) + g(U_1(0))\Big) \end{aligned} \right\} \\ A_{2,n}(U_1, U_2) = -(\Lambda U_1', \Lambda U_2') \end{cases}$$

and the source term

$$\begin{cases} D(G_2) = & (L^2([0,1]))^2 \\ G_2(U_1, U_2) = & g_2(U_1, U_2). \end{cases}$$

Recall that to apply the theory of nonlinear semigroups from [86], we need to prove the three following statements ; $A_{2,n} + G_2$ is ζ dissipative for some $\zeta \in \mathbb{R}$, satisfies the range condition and is closed. As these proofs are very similar to what was done in the parts (1) (2) and (3) for the operator $A_n + G$, we will just give a sketch of the proof and insist on crucial hypothesis.

— For ζ dissipativity, it is a comparison between boundary terms. The fact that σ_n' is bounded (n fixed) is the key hypothesis to show this ζ dissipativity.

— For range condition, we need to solve $(I - \rho(A_{2,n} + G_2))(u_1, u_2) = (v_1, v_2)$ where (u_1, u_2) is the unknown, $(v_1, v_2) \in D(A_{2,n})$ and ρ belonging to $[0, \rho_0]$ with ρ_0 independent on (v_1, v_2) to determine. We find u_1 using the fact that $A_n + G$ satisfies the range condition. For u_2, it suffices to remark that the ODE to solve is linear in u_2 (even for the boundary condition) with a bounded linear source term (g' is bounded as g is Lispchitz).

— Finally, for the closedness one can proceed using same techniques as in previous sections and the closedness of $A_n + G$.

Hence $(R_n, \partial_x R_n)$ is $C^0([0,T], L^2([0,1]))$. As a consequence, $R_n \in C^0([0,T], H^1([0,1]))$ and as $\partial_t R_n = g(R_n) - \Lambda \partial_x R_n$, we also have $\partial_t R_n \in C^0([0,T], L^2([0,1]))$. To conclude, $R_n \in C^0([0,T], H^1([0,1])) \cap C^1([0,T], L^2([0,1]))$.

All points of Theorem 3 are now proven. □

2.A.2 Convergence of solutions with smoothed saturations in $L^2([0,1])$

We define the operator A by :

$$\begin{cases} AR & = -\Lambda R' \\ D(A) & = \left\{ R \in H^1([0,1]); R(0) = HR(1) + B\sigma(KR(1)) \right\}. \end{cases}$$

Note that $\overline{D(A)} = L^2([0,1])$ because $C_c^\infty((0,1)) \subseteq D(A)$ and $C_c^\infty((0,1))$ is a dense subset of $L^2([0,1])$ by [22, Corollary 4.23] [1]. The following lemma will be useful to prove the convergence of semigroups $(T_n)_n$.

1. In fact, it is shown that $C_c^\infty(\Omega)$ is dense in $L^2(\Omega)$ for Ω open of $\mathbb{R}^N$ ($N \in \mathbb{N}$)) but the proof can be easily adapted to our context.

Lemma 1. *It holds $A \subset \lim_{n\to\infty} A_n$ which means that every element of the graph of A is the limit of a sequence $\{(x_n, A_n x_n)\}_n$ (in $L^2([0,1]) \times L^2([0,1])$) where $x_n \in D(A_n)$.*

Proof. Let $R \in D(A)$. Let us define the sequence $(R_n)_n$ in $H^1([0,1])$ by :

$$\forall n \in \mathbb{N}, \forall x \in [0,1],\ R_n(x) := (1-x^2)B[\sigma_n(KR(1)) - \sigma(KR(1))] + R(x).$$

For all integers n, we have $R_n(0) = HR(1) + B\sigma_n(KR(1))$ (use the fact that $R \in D(A)$) and $R_n(1) = R(1)$. Hence, for all integers n, $R_n \in D(A_n)$ and by the pointwise convergence of $(\sigma_n)_n$, we have $R_n \to_{H^1([0,1])} R$ which is equivalent to $(R_n, A_n R_n) \to_{L^2([0,1])\times L^2([0,1])} (R, AR)$.

□

The convergence of semigroups $(T_n)_n$ is given by the following theorem :

Theorem 4. *The operator $A + G$ generates a semigroup T of type ζ. Moreover, for all initial data R_0 in $\overline{D(A)} = L^2([0,1])$ and sequence $(R_{0,n})_n$ such that $R_{0,n} \in D(A_n)$ for all integers n and converging to R_0 in $L^2([0,1])$:*

$$\forall T > 0,\ \lim_{n\to\infty} T_n(\cdot)R_{0,n} = T(\cdot)R_0 \in C^0([0,T], L^2([0,1])). \tag{2.34}$$

Proof. After Lemma 3, all the $A_n + G$ are ζ dissipative. Moreover, $A + G$ satisfies the range condition, the proof being identical to the proof of "$A_n + G$ satisfies the range condition (2.17)" (because the argument only uses the boundedness and the continuity of the saturation operator). Hence, it satisfies the distance condition :

$$\forall Q \in \overline{D(A)},\ \liminf_{\rho\to 0^+} \rho^{-1} d(\mathrm{Rg}(I - \rho(A+G)), Q) = 0. \tag{2.35}$$

Finally, $A \subset \lim_{n\to\infty} A_n$ by previous lemma. As a consequence, by the continuity of G as an operator from $L^2([0,1])$ to $L^2([0,1])$, $A + G \subset \lim_{n\to\infty} A_n + G$.

Take $R_0 \in \overline{D(A)}$ and a sequence $(R_{0,n})_n$ of elements from $D(A_n)$ for all n converging to R_0 in $L^2([0,1])$. From [86, Theorem 6.8], $A + G$ generates a semigroup T and

$$\lim_{n\to+\infty} T_n(t)R_{0,n} = T(t)R_0 \text{ in } L^2([0,1])$$

for all time $t \geq 0$. Moreover, the equality above holds uniformly on every bounded interval of $[0, \infty)$ which proves Theorem 4. □

2.A.3 Conclusion of the proof of Theorem 1

(1) Existence

Let R_0 be the initial data in $\overline{D(A)} = L^2([0,1])$. Take a smooth approximation $(\sigma_n)_n$ of σ and $(R_{0,n})_n$ a sequence of $C^\infty((0,1))$ satisfying compatibility conditions of order 1 converging towards R_0 in $L^2([0,1])$. Note that there exists at least one sequence $(R_{0,n})_n$ of such initial data. Indeed, $C_c^\infty((0,1))$ is dense in $L^2([0,1])$ and functions of $C_c^\infty((0,1))$ obviously satisfy compatibility conditions of order 1.

By Theorem 3, for all integers n, the operator $A_n + G$ generates a semigroup T_n and $R_n := T_n(.)R_0$ is the unique solution to (2.3).

After Theorem 4, $(R_n)_n$ converges in $C^0([0,T], L^2([0,1]))$ towards an element $R \in C^0([0,T], L^2([0,1]))$. R satisfies all requirements for being a weak solution to system (2.5).

(2) Uniqueness

The limit R neither depends on the smooth approximation $(\sigma_n)_n$ nor on the sequence of initial data $(R_{0,n})_n$ chosen. To prove this, we take an arbitrary sequence of smooth approximations $(\tilde{\sigma}_n)_n$ and define the corresponding sequence of operators $(\tilde{A}_n)_n$. Then, take an arbitrary sequence $(\tilde{R}_{0,n})_n$ of elements of $D(\tilde{A}_n) \cap H^2([0,1])$ converging to R_0 in $L^2([0,1])$. For all integers n, we define $\tilde{R}_n$ as the solution of :

$$\begin{cases} \partial_t R_n + \Lambda \partial_x R_n = & g \circ R_n \\ R_n(\cdot, 0) = & H R_n(\cdot, 1) + B\tilde{\sigma}_n(K R_n(\cdot, 1)) \\ R_n(0, \cdot) = & \tilde{R}_{0,n}. \end{cases}$$

We note $\tilde{R}$, the limit in $C^0([0,T], L^2([0,1]))$ of $(\tilde{R}_n)_n$. We define $(\overline{\sigma}_n)_n$ as equal to σ_n if n is even and equal to $\tilde{\sigma}_n$ otherwise. In the same way, we define $(\overline{R}_{0,n})_n$ as equal to $R_{0,n}$ if n is even and equal to $\tilde{R}_{0,n}$ otherwise. Obviously, $(\overline{\sigma}_n)_n$ is a smooth approximation of σ and $(\overline{R}_{0,n})_n$ is a sequence of H^2 function satisfying compatibility conditions of order 1 and converging to R_0 in $L^2([0,1])$. Hence, we can apply what was done before to prove that there exists a limit $\overline{R}$ in $C^0([0,T], L^2([0,1]))$ of the sequence $(\overline{R_n})_n$. As a consequence $R = \overline{R} = \tilde{R}$ and the limit is unique.

(3) Continuity of $\mathbf{U_T}$

Let $R_0, \tilde{R}_0$ be two initial data in $L^2([0,1])$ and $R, \tilde{R}$ the associated weak solutions. Take an arbitrary smooth approximation $(\sigma_n)_n$ of σ and sequences of approximation $(R_{0,n})_n$ and $(\tilde{R}_{0,n})_n$ of R_0 and $\tilde{R}_0$ respectively. Using the ζ dissipativity of the $A_n + G$ proven in Theorem 3 with ζ independent on n, one gets :

$$\forall n \in \mathbb{N}, \forall 0 \leq t \leq T, \ \|T_n(t)R_{0,n} - T_n(t)\tilde{R}_{0,n}\|_{L^2([0,1])} \leq e^{\zeta t}\|R_{0,n} - \tilde{R}_{0,n}\|_{L^2([0,1])}.$$

Taking the limit for all time $t \leq T$ (see Theorem 4) :

$$\forall 0 \leq t \leq T, \ \|T(t)R_0 - T(t)\tilde{R}_0\|_{L^2([0,1])} \leq e^{\zeta t}\|R_0 - \tilde{R}_0\|_{L^2([0,1])}$$

which implies that :

$$\forall 0 \leq t \leq T, \ \|T(t)R_0 - T(t)\tilde{R}_0\|_{L^2([0,1])} \leq e^{\zeta T}\|R_0 - \tilde{R}_0\|_{L^2([0,1])}.$$

It shows that U_T is Lipschitz and hence continuous. This concludes the proof of Theorem 1.

2.B Proof of Theorem 2

Take a $P \in D_d^+(\mathbb{R})$ satisfying $|P(H + BK)P^{-1}|_\infty < 1$ and a positive $\mu < -\log(|P(H + BK)P^{-1}|_\infty)$. Take the smooth approximation $(\sigma_n)_n$ of σ from Remark 2. Notation $A_n, D(A_n)$... are the same as in Appendix A.1.

Let R_0 be an initial data in $\overline{D(A)} = L^2([0,1])$ (for the moment we do not suppose that R_0 is in $L^\infty([0,1])$) . As $C_c^\infty((0,1))$ is dense in $L^2([0,1])$ (see [22]) and as $C_c^\infty((0,1)) \subset D(A_n)$ for all integers n, one can construct a sequence $(R_{0,n})_n$ belonging to $C_c^\infty((0,1)) \subset D(A_n)$ converging to R_0 in $L^2([0,1])$. In what follows, we will consider such a sequence of initial data.

2.B.1 Local L^∞ stability of $C^1([0,T], L^2([0,1])) \cap C^0([0,T], H^1([0,1]))$ solutions

All along section B.1, n is a **fixed** integer.

R_n is the solution to the smooth problem :

$$\begin{cases} \partial_t R_n + \Lambda \partial_x R_n = & g \circ R_n \\ R_n(\cdot, 0) = & H R_n(\cdot, 1) + B\sigma_n(K R_n(\cdot, 1)) \\ R_n(0, \cdot) = & R_{n,0} \in D(A_n) \cap H^2([0,1]). \end{cases}$$

As the initial data is $H^2([0,1])$, satisfies compatibility conditions of order 1, the solution belongs to $C^1([0,T], L^2([0,1])) \cap C^0([0,T], H^1([0,1]))$ for all $T > 0$ by Theorem 3.

Recall the definition of $V(.)$,

$$\forall f \in L^\infty([0,1]),\ V(f) := \max_{i \in [\![1,d]\!]} |P_i f_i e^{-\mu x}|_{L^\infty([0,1])} \tag{2.36}$$

where $\mathrm{diag}_{i \in [\![1,d]\!]}(P_i) = P$.

For all integers p, the functional V_{2p} is defined by :

$$\forall f \in L^\infty([0,1]),\ V_{2p}(f) := \left(\sum_{i=1}^{d} \int_0^1 q_{p,i}^{2p} f_i^{2p} e^{-2p\mu x} dx \right)^{1/2p}$$

where $Q_p = \mathrm{diag}_{i \in [\![1,d]\!]}(q_{p,i})$ will be chosen wisely later.

Claim 2. *We can differentiate $V_{2p}(R_n(t,\cdot))$ with respect to time and*

$$\forall t \geq 0,\ \frac{dV_{2p}^{2p}(R_n(t,\cdot))}{dt} = 2p \sum_{i=1}^{d} \int_0^1 q_{p,i}^{2p} \partial_t R_{n,i} R_{n,i}^{2p-1} e^{-2p\mu x} dx. \tag{2.37}$$

Proof of Claim 2. Take $t \geq 0$, $dt > 0$ and $T > t$.

$$\left| \frac{V_{2p}^{2p}(R_n(t+dt,\cdot)) - V_{2p}^{2p}(R_n(t,\cdot))}{dt} - 2p \sum_{i=1}^{d} \int_0^1 q_{p,i}^{2p} \partial_t R_{n,i} R_{n,i}^{2p-1} e^{-2p\mu x} dx \right|$$
$$\leq \sum_{i=1}^{d} \int_0^1 q_{p,i}^{2p} \left| \frac{R_{n,i}^{2p}(t+dt,x) - R_{n,i}^{2p}(t,x)}{dt} - 2p \partial_t R_{n,i}(t,x) R_{n,i}^{2p-1}(t,x) \right| e^{-2p\mu x} dx.$$

Thus it is sufficient to prove that at the limit when dt tends to zero

$$\forall i \in [\![1,d]\!],\ \left| \frac{R_{n,i}^{2p}(t+dt,\cdot) - R_{n,i}^{2p}(t,\cdot)}{dt} - 2p \partial_t R_{n,i}(t,\cdot) R_{n,i}^{2p-1}(t,\cdot) \right| \to_{L^1([0,1])} 0.$$

Let i be in $[\![1,d]\!]$,

$$\int_0^1 \left|\frac{R_{n,i}^{2p}(t+dt,\cdot)-R_{n,i}^{2p}(t,\cdot)}{dt}-2p\partial_t R_{n,i}(t,\cdot)R_{n,i}^{2p-1}(t,\cdot)\right|dx$$
$$=\int_0^1 \left|\frac{R_{n,i}(t+dt,\cdot)-R_{n,i}(t,\cdot)}{dt}(R_{n,i}(t+dt,\cdot)^{2p-1}+\cdots+R_{n,i}(t,\cdot)^{2p-1})-2p\partial_t R_{n,i}(t,\cdot)R_{n,i}^{2p-1}(t,\cdot)\right|dx$$
$$\leq \|\frac{R_{n,i}(t+dt,\cdot)-R_{n,i}(t,\cdot)}{dt}-\partial_t R_{n,i}\|_{L^2([0,1])}\times\sqrt{\int_0^1 |R_{n,i}(t+dt,\cdot)^{2p-1}+\cdots+R_{n,i}(t,\cdot)^{2p-1}|^2\,dx}$$
$$+\int_0^1 |\partial_t R_{n,i}|\left|R_{n,i}(t+dt,\cdot)^{2p-1}+\cdots+R_{n,i}(t,\cdot)^{2p-1}-2pR_{n,i}(t,\cdot)^{2p-1}\right|dx$$

The term $\left|\left|\frac{R_{n,i}(t+dt,\cdot)-R_{n,i}(t,\cdot)}{dt}-\partial_t R_n\right|\right|_{L^2([0,1])}$ converges towards zero because $R_{n,i}$ is in $C^1([0,T],L^2([0,1]))$ while the term $\int_0^1 \left|R_{n,i}(t+dt,\cdot)^{2p-1}+\cdots+R_{n,i}(t,\cdot)^{2p-1}\right|^2 dx$ is bounded because $R_{n,i}\in C^0([0,T],H^1([0,1]))\subseteq C^0([0,T]\times[0,1])$. Finally, the term $\int_0^1 \left|\partial_t R_{n,i}\right|\left|R_{n,i}(t+dt,\cdot)^{2p-1}+\cdots+R_{n,i}(t,\cdot)^{2p-1}-2pR_{n,i}(t,\cdot)^{2p-1}\right| dx$ is bounded by $\|\partial_t R_{n,i}\|_{L^2([0,1])}\|R_{n,i}(t+dt,\cdot)^{2p-1}+\cdots+R_{n,i}(t,\cdot)^{2p-1}-2pR_{n,i}(t,\cdot)^{2p-1}\|_{L^2([0,1])}$. The term $\|\partial_t R_{n,i}\|_{L^2([0,1])}$ is bounded as R_n is in $C^1([0,T],L^2([0,1]))$ whereas the other term converges towards zero by the dominated convergence theorem. Therefore, Claim 2 is proven. □

Next, as R_n verifies $\partial_t R_n+\Lambda\partial_x R_n=g\circ R_n$, one gets :

$$\frac{dV_{2p}^{2p}(R_n(t,\cdot))}{dt}=2p\sum_{i=1}^d\int_0^1 q_{p,i}^{2p}\left(-\lambda_i\partial_x R_{n,i}+g_i\circ R\right)R_{n,i}^{2p-1}e^{-2p\mu x}dx. \tag{2.38}$$

There are two terms in (2.38) whose origins are different :

— the transport term :

$$W_{2p}(R_n(t,\cdot)):=-2p\sum_{i=1}^d\int_0^1 q_{p,i}^{2p}\lambda_i\partial_x R_{n,i}R_{n,i}^{2p-1}e^{-2p\mu x}dx \tag{2.39}$$

— and the source term :

$$Z_{2p}(R_n(t,\cdot)):=2p\sum_{i=1}^d\int_0^1 q_{p,i}^{2p}(g_i\circ R)R_{n,i}^{2p-1}e^{-2p\mu x}dx \tag{2.40}$$

so that :

$$\frac{dV_{2p}^{2p}(R_n(t,\cdot))}{dt}=W_{2p}(R_n(t,\cdot))+Z_{2p}(R_n(t,\cdot)). \tag{2.41}$$

Define :

$$\forall t\geq 0,\ \xi_n(t):=PR_n(t,1). \tag{2.42}$$

In the rest of the proof, we will sometimes denote $\xi_n(t)$ as ξ_n for readability.

Study of the term transport term W_{2p}

Lemma 2. *For all* $\mu < -\log(|P(H+BK)P^{-1}|_\infty)$ *and* $\varepsilon > 0$ *there exists* $\tilde{p} = \tilde{p}(\mu, H, B, K, P, \varepsilon)$ *such that for all initial data* $R_{0,n} \in H^2([0,1])$ *satisfying compatibility conditions of order 1, the solution* $R_n = T_n(.)R_{0,n}$ *to (2.3) verifies the following assertion.*

If at time $\bar{t} \geq 0$,

$$|\xi_n(\bar{t})|_{\max} < \frac{|PBP^{-1}|_\infty \sigma_{s,n} P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu-\varepsilon}|}. \tag{2.43}$$

Then :

$$\forall p > \tilde{p},\ W_{2p}(R_n(\bar{t},\cdot)) \leq -2p\mu\lambda_{\min} V_{2p}^{2p}(R_n(\bar{t},\cdot)). \tag{2.44}$$

Remark 10. The positive real ε is introduced to prevent an eventual dependence of $\tilde{p}$ with respect to $\bar{t}$. Indeed, further in the chapter, we integrate (2.44) with respect to time. This would not be possible if $\tilde{p} = \tilde{p}(\bar{t})$.

Proof. In this proof we will sometimes drop the time dependence notation $\xi(\bar{t})$ or $R_n(\bar{t})$ for readability.

Using an integration by parts in (2.39), we compute :

$$W_{2p}(R_n(\bar{t},\cdot)) = -\sum_{i=1}^{d} q_{p,i}^{2p}\lambda_i \left[R_{n,i}^{2p} e^{-2p\mu x}\right]_0^1 - 2p\mu \sum_{i=1}^{d} \int_0^1 q_{p,i}^{2p}\lambda_i R_{n,i}^{2p} e^{-2p\mu x} dx$$

and therefore

$$W_{2p}(R_n(\bar{t},\cdot)) \leq -W_{2p,1}(R_n(\bar{t},\cdot)) - 2p\mu\lambda_{\min} V_{2p}^{2p}(R_n(\bar{t},\cdot)) \tag{2.45}$$

where

$$W_{2p,1}(R_n(\bar{t},\cdot)) := \sum_{i=1}^{d} q_{p,i}^{2p}\lambda_i \left[R_{n,i}^{2p} e^{-2p\mu x}\right]_0^1.$$

Using the fact that $R_n(\bar{t},\cdot)$ is in $D(A_n)$, one gets :

$$W_{2p,1}(R_n(\bar{t},\cdot)) = \sum_{i=1}^{d} q_{p,i}^{2p}\lambda_i R_{n,i}^{2p}(\bar{t},1) e^{-2p\mu} - q_{p,i}^{2p}\lambda_i \left\{HR_n(\bar{t},1) + B\sigma_n(KR_n(\bar{t},1))\right\}_i^{2p}.$$

We have,

$$W_{2p,1}(R_n(t,\cdot)) = W_{2p,11} + W_{2p,12} \tag{2.46}$$

with :

$$W_{2p,11} := \sum_{i=1}^{d} q_{p,i}^{2p}\lambda_i R_{n,i}^{2p}(\bar{t},1) e^{-2p\mu} \tag{2.47}$$

and

$$W_{2p,12} := -\sum_{i=1}^{d} q_{p,i}^{2p} \lambda_i \left\{ HR_n(\bar{t},1) + B(\sigma_n(KR_n(\bar{t},1))) \right\}_i^{2p}. \tag{2.48}$$

Note that if $\xi_n = 0$, $W_{2p,1}$ is zero and the conclusion of Lemma 2 holds. In what follows, we suppose that $\xi_n \neq 0$.

Study of $\mathbf{W_{2p,11}}$

For what follows, we choose Q_p such that :

$$\boxed{\forall i \in [\![1,d]\!], \forall p \in \mathbb{N},\ q_{p,i} = \lambda_i^{-1/2p} P_i} \tag{2.49}$$

Inspired by [16, p. 123], we get the following estimates for $W_{2p,11}$

$$\begin{array}{rcl} W_{2p,11} & = & \sum_{i=1}^{d} q_{p,i}^{2p} \lambda_i R_{n,i}^{2p}(\bar{t},1) e^{-2p\mu} \\ & \geq & |\xi_n|_{\max}^{2p} e^{-2p\mu} \end{array} \tag{2.50}$$

where we have used the equations $q_{p,i}^{2p}\lambda_i = P_i^{2p}$, $|\xi_n|_{\max}^{2p} \leq \sum_{i=1}^{d} |\xi_i|^{2p}$ and (2.42).

Hence :

$$\frac{W_{2p,11}}{|\xi_n|_{\max}^{2p}} \geq e^{-2p\mu} \geq 0. \tag{2.51}$$

Study of $\mathbf{W_{2p,12}}$:

The term $W_{2p,12}$ is deeply related to the effect of saturation. We recall the definition of $W_{2p,12}$ in (2.48) :

$$W_{2p,12} = -\sum_{i=1}^{d} \left\{ HR_n(\bar{t},1) + B(\sigma_n(KR_n(\bar{t},1))) \right\}_i^{2p})$$

equivalent to :

$$W_{2p,12} = -\sum_{i=1}^{d} P_i^{2p} \upsilon_i^{2p} \tag{2.52}$$

with :

$$\upsilon = HR_n(\bar{t},1) + B\sigma_n(KR_n(\bar{t},1)).$$

The following claim allows to neglect the term $W_{2p,12}$ relatively to $W_{2p,11}$ at the limit when p tends to infinity :

Claim 3. *Under the conditions of Lemma 2 and if $\xi_n \neq 0$, then there exists a $\tilde{p} = \tilde{p}(\mu, H, B, K, P, \varepsilon)$ such that :*

$$\forall p > \tilde{p},\ |W_{2p,12}| \leq \frac{e^{-2p\mu}|\xi_n|_{\max}^{2p}}{2}. \tag{2.53}$$

Proof of Claim 3. Let i be in $[\![1,d]\!]$. Denoting $R_{n,j}(\bar{t},1)$ by $R_{n,j}$ ($j\in[\![1,d]\!]$) and Sat_i the set

$$\mathrm{Sat}_i=\{j\in[\![1,d]\!]\mid |(KR_n)_j|>\sigma_{s,n} \text{ and } B_{i,j}\neq 0\},$$

one has

$$\begin{aligned}
P_i|\upsilon_i| =& P_i\left|\sum_{j=1}^{d}(H+BK)_{i,j}R_{n,j}+\sum_{j=1}^{d}B_{i,j}\left(\sigma_{n,j}([KR_n]_j)-[KR_n]_j\right)\right|\\
=& P_i\left|\sum_{j=1}^{d}(H+BK)_{i,j}R_{n,j}+\sum_{j\in\mathrm{Sat}_i}B_{i,j}\left(\sigma_{n,j}([KR_n]_j)-[KR_n]_j\right)\right|\\
\leq& P_i\sum_{j=1}^{d}|(H+BK)_{i,j}R_{n,j}|+P_i\sum_{j\in\mathrm{Sat}_i}\left|B_{i,j}\left(\sigma_{n,j}([KR_n]_j)-\sum_{k=1}^{d}K_{j,k}R_{n,k}\right)\right|\\
\leq& P_i\sum_{j=1}^{d}|(H+BK)_{i,j}R_{n,j}|+P_i\sum_{j\in\mathrm{Sat}_i}|B_{i,j}|\left(|\sum_{k=1}^{d}K_{j,k}R_{n,k}|-\sigma_{s,n}\right)\\
\leq& |P(H+BK)P^{-1}|_\infty|\xi_n|_{\max}+|PBP^{-1}|_\infty(|PKP^{-1}|_\infty|\xi_n|_{\max}-\sigma_{s,n}P_{\min})
\end{aligned}$$

If Sat_i is empty *ie* if the saturation does not act on the ith coordinate, then $\upsilon_i=\chi_i$ and

$$P_i|\upsilon_i| \;\leq |P(H+BK)P^{-1}|_\infty|\xi_n|_{\max}. \tag{2.54}$$

Otherwise, Sat_i is non empty and

$$P_i|\upsilon_i| \leq |P(H+BK)P^{-1}|_\infty|\xi_n|_{\max}+|PBP^{-1}|_\infty(|PKP^{-1}|_\infty|\xi_n|_{\max}-\sigma_{s,n}P_{\min}). \tag{2.55}$$

Moreover, by assumption $|P(H+BK)P^{-1}|_\infty<e^{-\mu}$ and so, there exist $\alpha<1$ such that :

$$|P(H+BK)P^{-1}|_\infty=\alpha e^{-\mu}.$$

Additionally, by (2.43),

$$|P(H+BK)P^{-1}|_\infty|\xi_n|_{\max}+|PBP^{-1}|_\infty(|PKP^{-1}|_\infty|\xi_n|_{\max}-\sigma_{s,n}P_{\min})<e^{-\mu-\varepsilon}|\xi_n|_{\max}.$$

Injecting the last two equations in (2.54)-(2.55),

$$\forall i\in[\![1,d]\!],\quad P_i|\upsilon_i| \quad\leq\quad \max(\alpha e^{-\mu},e^{-\mu-\varepsilon})|\xi_n|_{\max}. \tag{2.56}$$

Finally, from (2.56) and (2.52), we have

$$\forall p\in\mathbb{N},\ \frac{W_{2p,12}}{|\xi_n|_{\max}^{2p}}\leq 2d\times\max\{\alpha^{2p},e^{-2p\varepsilon}\}e^{-2p\mu}.$$

Hence, there exists an integer $\tilde{p}$ (depending on (ε,α) but not on $|\xi_n|_{\max}$) such that

$$\forall p\geq\tilde{p},\ \left|\frac{W_{2p,12}}{|\xi_n|^{2p}}\right|\leq\frac{e^{-2p\mu}}{2}.$$

This ends the proof of Claim 3. $\square$

As a consequence, by (2.51), (2.53) and for $p > \tilde{p}$, $W_{2p,1} = W_{2p,11} + W_{2p,22} \geq \frac{|\xi_n|_{\max}^{2p} e^{-2p\mu}}{2} \geq 0$. Injecting last statement in (2.45), one gets :

$$W_{2p}(R_n(\bar{t},\cdot)) \leq -2p\mu\lambda_{\min} V_{2p}^{2p}(R_n(\bar{t},\cdot)).$$

This finishes the proof of Lemma 2. □

Remark 11. *In the proof of the above result, we enforce $W_{2p,12} =_{p\to+\infty} o(|\xi_n|_{\max}^{2p} e^{-2p\mu})$. This condition recalls a local sector bounded condition[102, Section 1.7.2]. Indeed, the term $W_{2p,12}$ is induced by the difference between the linear control law and the its saturated version, i.e., this term arises from a deadzone nonlinearity. The term on the right-hand side of* (2.57)*, $e^{-\mu}|\xi_n|_{\max}$, represents the state of the system. Thus formally speaking, the condition*

$$W_{2p,12} =_{p\to+\infty} o(|\xi_n|_{\max}^{2p} e^{-2p\mu}), \tag{2.57}$$

is somewhat equivalent to a regional sector condition[102, Section 1.7.2].

Remark 12. *Conditions (2.43) are less restrictive when the saturation $\sigma_{s,n}$ is weaker and when the exponential decay rate μ decreases.*

Analysis of the term Z_{2p}

Lemma 3. *For all integers p and for all time $t \geq 0$,*

$$|Z_{2p}(R_n(t,\cdot))| \leq 2pL_{g,\max} V_{2p}^{2p}(R_n(t,\cdot)).$$

Proof. Recall the definition of $Z_{2p}(R_n(t,\cdot))$ (2.40) :

$$Z_{2p}(R_n(t,\cdot)) = 2p \sum_{i=1}^{d} \int_0^1 q_{p,i}^{2p}(g_i \circ R_n) R_{n,i}^{2p-1} e^{-2p\mu x} dx.$$

Using the hypothesis on g in (2.6) :

$$|Z_{2p}(R_n(t,\cdot))| \leq 2p \sum_{i=1}^{d} \int_0^1 q_{p,i}^{2p} L_{g,i} R_{n,i}^{2p} e^{-2p\mu x} dx \leq 2pL_{g,\max} V_{2p}^{2p}(R_n(t,\cdot)).$$

□

Conclusion on the L^∞ stability of regular solutions

Lemma 4. *Under conditions of Lemma 2, the solution $R_n = T_n(.)R_{0,n}$ to problem (2.3) satisfies the following statement.*

If at a time $\bar{t} \geq 0$, condition (2.43) is satisfied then :

$$\frac{dV_{2p}(R_n(\bar{t},\cdot))}{dt} \leq -(\mu\lambda_{\min} - L_{g,\max}) V_{2p}(R_n(\bar{t},\cdot))$$

for all $p > \tilde{p}$ where $\tilde{p}$ defined in Lemma 2.

Proof. Let $p > \tilde{p}$. By (2.41), we have :

$$\frac{dV_{2p}^{2p}}{dt} = 2p\frac{dV_{2p}}{dt}V_{2p}^{2p-1} = W_{2p} + Z_{2p}. \tag{2.58}$$

After Lemmas 2 and 3,

$$W_{2p}(R_n(\bar{t},\cdot)) \leq -2p\mu\lambda_{\min}V_{2p}^{2p}(R_n(\bar{t},\cdot))$$

and

$$|Z_{2p}(R_n(\bar{t},\cdot))| \leq 2pL_{g,\max}V_{2p}^{2p}(R_n(\bar{t},\cdot)).$$

Summing previous inequalities and dividing by $2pV_{2p}^{2p-1}$ in (2.58), one gets :

$$\frac{dV_{2p}}{dt} \leq -(\mu\lambda_{\min} - L_{g,\max})V_{2p}.$$

□

Before going further into the proof, we need the following lemma which may be useful in future works :

Lemma 5.

$$\forall R \in L^\infty([0,1]), V_{2p}(R) \to_{p\to+\infty} V(R).$$

Moreover, the convergence is uniform on all bounded sets of $H^1([0,1])$.

Proof. Let R be in $L^\infty([0,1])$

$$\forall p \in \mathbb{N},\ V_{2p}(R) = \left(\sum_{i=1}^{d}\int_0^1 P_i^{2p}\lambda_i R_i^{2p}e^{-2p\mu x}dx\right)^{1/2p}.$$

As a consequence,

$$\lambda_{\min}^{1/2p}\|x \mapsto PR(x)e^{-\mu x}\|_{L^{2p}([0,1])} \leq V_{2p}(R) \leq \lambda_{\max}^{1/2p}\|x \mapsto PR(x)e^{-\mu x}\|_{L^{2p}([0,1])}.$$

By a classic result of analysis $\|Q\|_{L^{2p}([0,1])}$ converges towards $\|Q\|_{L^\infty([0,1])}$ for all $Q \in L^\infty([0,1])$. Hence, passing to the limit in last inequalities :

$$\|x \mapsto PR(x)e^{-\mu x}\|_{L^\infty([0,1])} \leq \lim_{p\to+\infty} V_{2p}(R) \leq \|x \mapsto PR(x)e^{-\mu x}\|_{L^\infty([0,1])}$$

and :

$$\lim_{p\to+\infty} V_{2p}(R) = \|x \mapsto PR(x)e^{-\mu x}\|_{L^\infty([0,1])} = V(R).$$

We also need to prove the uniform convergence for all bounded sets of $H^1([0,1])$. We will prove it for the case of scalar functions ; the case of vector valued function being similar.

Take an $r > 0$. Let B_r be the ball of radius r in $H^1([0,1])$. Define for $\omega > 0$ and f in B_r, $S_{f,\omega} := \{x \in [0,1];\ |f(x)| \geq |f|_{L^\infty([0,1])} - \omega\}$. For all $\omega > 0$ and f in B_r, there exists an x in $S_{f,\omega/2}$.

For all y in $[0,1]$, we have :

$$f(y) = f(x) + \int_x^y f'(z)dz.$$

Using Cauchy-Schwartz inequality for the L^2 canonical scalar product, one gets

$$|f(y)| \geq |f(x)| - r\sqrt{|x-y|}.$$

As x is in $S_{\omega/2}$:

$$|f(y)| \geq |f|_{L^\infty([0,1])} - \frac{\omega}{2} - r\sqrt{|x-y|}$$

Thus, for all y in $[0,1]$ such that $\sqrt{|x-y|} \leq \frac{\omega}{2r}$:

$$|f(y)| \geq |f|_{L^\infty([0,1])} - \omega$$

which implies that :

$$\forall y \in [0,1];\ \sqrt{|x-y|} \leq \frac{\omega}{2r} \implies y \in S_{f,\omega}.$$

As a consequence :

$$\forall f \in B_r,\ \forall 0 < \omega < 2r,\ \mu(S_{f,\omega}) \geq \frac{\omega^2}{4r^2} \tag{2.59}$$

where μ designates the usual Lebesgue measure.

Now using the definition of $S_{f,\omega}$ and (2.59), one has for all $f \in B_r, \omega > 0, p \in \mathbb{N}$:

$$\begin{aligned}\left(\tfrac{\omega}{2r}\right)^{1/p}(|f|_{L^\infty([0,1])} - \omega) \leq \quad \mu(S_{f,\omega})^{1/2p}(|f|_{L^\infty([0,1])} - \omega) &= \left(\int_{S_{f,\omega}} (|f|_{L^\infty([0,1])} - \omega)^{2p} dx\right)^{1/2p} \\ &\leq |f|_{L^{2p}([0,1])}.\end{aligned}$$

Moreover,

$$\forall f \in L^\infty([0,1]),\ |f|_{L^p([0,1])} \leq |f|_{L^\infty([0,1])}.$$

Hence :

$$\forall f \in B_r,\ \forall \omega > 0, \forall p \in \mathbb{N},\ \left(\frac{\omega}{2r}\right)^{1/p}(|f|_{L^\infty([0,1])} - \omega) \leq |f|_{L^{2p}([0,1])} \leq |f|_{L^\infty([0,1])}$$

which implies :

$$\forall f \in B_r,\ \forall \omega > 0, \forall p \in \mathbb{N},\ 0 \leq \left| |f|_{L^{2p}([0,1])} - |f|_{L^\infty([0,1])} \right| \leq \left| \left(\frac{\omega}{2r}\right)^{1/p}(|f|_{L^\infty([0,1])} - \omega) - |f|_{L^\infty([0,1])} \right|.$$

Now take an $0 < \varepsilon < 1$, f in B_r and pose $\omega = \frac{\varepsilon}{2}$;

$$\begin{aligned}\left|\left(\frac{\omega}{2r}\right)^{1/p}(|f|_{L^\infty([0,1])}-\omega)-|f|_{L^\infty([0,1])}\right| &\leq \left|\left(\frac{\omega}{2r}\right)^{1/p}-1\right|\left||f|_{L^\infty([0,1])}-\omega\right| \\ &+ \left||f|_{L^\infty([0,1])}-\omega-|f|_{L^\infty([0,1])}\right| \\ &\leq \left|\left(\frac{\omega}{2r}\right)^{1/p}-1\right|(r+\omega)+\omega \\ &= \left|\left(\frac{\varepsilon}{4r}\right)^{1/p}-1\right|(r+\frac{\varepsilon}{2})+\frac{\varepsilon}{2} \\ &\leq \left|\left(\frac{\varepsilon}{4r}\right)^{1/p}-1\right|(r+1)+\frac{\varepsilon}{2}.\end{aligned}$$

There exists a $P(r,\varepsilon)$ (independent on f) such that for $p > P(r,\varepsilon)$:

$$\left|\left(\frac{\varepsilon}{4r}\right)^{1/p}-1\right|(r+1) \leq \frac{\varepsilon}{2}$$

so that :

$$\forall f \in B_r,\ \left||f|_{L^{2p}([0,1])}-|f|_{L^\infty([0,1])}\right| \leq \varepsilon.$$

This finishes the proof of Lemma 5. □

Lemma 6. *For all initial data $R_{0,n} \in H^2([0,1])$ satisfying compatibility conditions of order 1 and $\mu < -\log(|P(H+BK)P^{-1}|_\infty)$, if*

$$\begin{cases} \mu\lambda_{\min} - L_{g,\max} & \geq & 0 \\ V(R_{0,n}) & < & e^{-\mu}\dfrac{|PBP^{-1}|_\infty \sigma_{s,n} P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu}|}\end{cases}$$

then, the solution $R_n = T_n(.)R_{0,n}$ to (2.3) verifies :

$$\forall t \geq 0,\ V(R_n(t,\cdot)) \leq e^{-(\mu\lambda_{\min}-L_{g,\max})t}V(R_{0,n}).$$

Proof. Using a continuity argument, we can take a $\varepsilon > 0$ such that

$$V(R_{0,n}) < e^{-\mu}\frac{|PBP^{-1}|_\infty \sigma_{s,n} P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu-\varepsilon}|}. \tag{2.60}$$

Define for all integers p,

$$T_p := \sup\left\{T \geq 0;\ \forall t \in [0,T],\ V_{2p}(R_n(t,\cdot)) \leq e^{-(\mu\lambda_{\min}-L_{g,\max})t}V_{2p}(R_{0,n})\right\}.$$

(1) $(T_p)_p$ is not a bounded sequence

We will prove that the sequence $(T_p)_p$ is not bounded by contradiction. Suppose $(T_p)_p$ bounded and take a subsequence still denoted $(T_p)_p$ converging towards a limit denoted T_∞.

(1.1) T_∞ is strictly positive

For all integers i in $[\![1,d]\!]$

$$\xi_{n,i}(0)e^{-\mu} = P_i R_{n,i}(1,0)e^{-\mu} \leq V(R_n(.,0)) = V(R_{0,n}).$$

By (2.60),

$$|\xi_n|_{\max}(0) < \frac{|PBP^{-1}|_\infty \sigma_{s,n} P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu-\varepsilon}|}.$$

By continuity of $t \mapsto \xi_n(t) = PR_n(t,1)$, there exists a $dt_0 > 0$ independent on p such that :

$$\forall t \in [0, dt_0],\ |\xi_n|_{\max}(t) < \frac{|PBP^{-1}|_\infty \sigma_{s,n} P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu-\varepsilon}|}.$$

Applying Lemma 4 and noting that condition (2.43) is satisfied for $\bar{t} \leftarrow t \in [0, dt_0]$, there exists a $\tilde{p}(\mu, H, B, K, \Lambda, \varepsilon) \in \mathbb{N}$ (independent on time) :

$$\forall p > \tilde{p},\ \forall t \in [0, dt_0],\ \frac{dV_{2p}(R_n(t,\cdot))}{dt} \leq -(\mu\lambda_{\min} - L_{g,\max})V_{2p}(R_n(t,\cdot)).$$

Integrating on $[0, dt_0]$, it holds

$$\forall p > \tilde{p}, \forall t \in [0, dt_0],\ V_{2p}(R_n(t,\cdot)) \leq e^{-(\mu\lambda_{\min} - L_{g,\max})t} V_{2p}(R_{0,n}).$$

This allows to assert that :

$$\forall p > \tilde{p},\ T_p \geq dt_0$$

and passing to the limit,

$$T_\infty \geq dt_0 > 0.$$

(1.2) Proof of the contradiction

As $R_n \in C^0([0,1] \times [0,T])$ for all $T > 0$ and by definition of T_p :

$$\forall p \in \mathbb{N},\ V_{2p}(R_n(T_p,\cdot)) = e^{-(\mu\lambda_{\min} - L_{g,\max})T_p} V_{2p}(R_{0,n}).$$

As $(R_n(T_p))_p$ is a bounded sequence of $H^1([0,1])$ (because R_n is in $C^0([0,T], H^1([0,1]))$ for all $T > 0$), we can use the uniform convergence proven in Lemma 5 to pass to the limit as p tends towards infinity in last equation :

$$V(R_n(T_\infty,\cdot)) = e^{-(\mu\lambda_{\min} - L_{g,\max})T_\infty} V(R_{0,n}) \leq V(R_{0,n}).$$

As a consequence, for all integers i in $[\![1,d]\!]$

$$\xi_{n,i}(T_\infty)e^{-\mu} = P_i R_{n,i}(T_\infty,1)e^{-\mu} \leq V(R_n(T_\infty,\cdot)) \leq V(R_{0,n})$$

and by (2.60), we have :

$$|\xi_n|_{\max}(T_\infty) < \frac{|PBP^{-1}|_\infty \sigma_{s,n} P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu-\varepsilon}|}.$$

As $t \mapsto \xi_n(t) = PR_n(t,1)$ is a continuous function (remember that R_n is in $C^0([0,1]\times[0,T])$ for all $T>0$), there exists a $0 < dt < T_\infty$ independent on p (but dependent on n) such that :

$$\forall t \in [T_\infty - dt, T_\infty + dt],\ |\xi_n|_{\max}(t) < \frac{|PBP^{-1}|_\infty \sigma_{s,n} P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu-\varepsilon}|}.$$

Applying Lemma 4 and noting that condition (2.43) is satisfied for $\bar{t} \leftarrow t \in [T_\infty - dt, T_\infty + dt]$, there exists $\tilde{p}(\mu, H, B, K, \Lambda, \varepsilon) \in \mathbb{N}$ such that :

$$\forall p > \tilde{p},\ \forall t \in [T_\infty - dt, T_\infty + dt],\ \frac{dV_{2p}(R_n(t,\cdot))}{dt} \leq -(\mu\lambda_{\min} - L_{g,\max}) V_{2p}(R_n(t,\cdot))$$

which is no more than :

$$\forall p > \tilde{p},\ \forall t \in [T_\infty - dt, T_\infty + dt],\ \frac{d[V_{2p}(R_n(t,\cdot)) e^{(\mu\lambda_{\min} - L_{g,\max})t}]}{dt} \leq 0.$$

Integrating last statement on $[T_\infty - dt, t]$ for $T_\infty - dt \leq t \leq T_\infty + dt$:

$$\forall p > \tilde{p},\ \forall t \in [T_\infty - dt, T_\infty + dt],\ V_{2p}(R_n(t,\cdot)) \leq e^{-(\mu\lambda_{\min} - L_{g,\max})(t+dt-T_\infty)} V_{2p}(R_n(T_\infty - dt, \cdot)) \tag{2.61}$$

Moreover, as $T_\infty - dt < T_\infty$ and $\lim_{p\to\infty} T_p = T_\infty$, there exists a $P(dt)$ such that for all $p > P(dt)$, $T_\infty - dt < T_p$. It implies that :

$$\forall p > P(dt),\ \forall t \leq T_\infty - dt,\ V_{2p}(R_n(t,\cdot)) \leq e^{-(\mu\lambda_{\min} - L_{g,\max})t} V_{2p}(R_{0,n}).$$

Using last inequality and (2.61), one has :

$$\forall p > \max(P(dt), \tilde{p}), \forall t \leq T_\infty + dt,\ V_{2p}(R_n(t,\cdot)) \leq e^{-(\mu\lambda_{\min} - L_{g,\max})t} V_{2p}(R_{0,n}).$$

As a consequence and by the definition of T_p :

$$\forall p > \max(P(dt), \tilde{p}),\ T_p \geq T_\infty + dt$$

passing to the limit as p tends towards infinity, one gets :

$$T_\infty \geq T_\infty + dt$$

implying that $dt \leq 0$ which is a contradiction.

(2) Proof of exponential convergence

Hence, $(T_p)_p$ is not a bounded sequence and there exists a non decreasing subsequence $(T_{\phi(p)})_p$ of $(T_p)_p$ such that :

$$\lim_{p\to\infty} T_{\phi(p)} = +\infty.$$

Take an arbitrary time $t \geq 0$, there exists a P such that for $p > P$ then $T_{\phi(p)} \geq t$. It implies that :

$$\forall p > P,\ V_{2\phi(p)}(R_n(t,\cdot)) \leq e^{-(\mu\lambda_{\min} - L_{g,\max})t} V_{2\phi(p)}(R_{0,n})$$

Passing to the limit as p tends towards infinity, we get :

$$V(R_n(t,\cdot)) \leq e^{-(\mu\lambda_{\min} - L_{g,\max})t} V(R_{0,n})$$

and Lemma 6 is proven. □

2.B.2 Local $L^\infty([0,1])$ stability for weak solutions

Finally, we prove Theorem 2. The integer n is not fixed anymore and we will pass to the limit in the exponential stability of the sequence $(R_n)_n$.

Lemma 7. *For all $R_0 \in L^\infty([0,1])$, there exists a sequence $(R_{0,n})_n \in C_c^2((0,1))$ converging to R_0 in $L^2([0,1])$ such that*

$$\forall n \in \mathbb{N},\ V(R_{0,n}) \leq V(R_0).$$

Proof. By [22, Corollary 4.23], for all $R \in L^\infty([0,1])$ there exists a sequence of elements of $C_c^2((0,1))$, $(R_n)_n$, such that :

$$\forall n \in \mathbb{N},\ \|R_n\|_{L^\infty([0,1])} \leq \|R\|_{L^\infty([0,1])}$$

and

$$R_n \rightarrow_{n\rightarrow+\infty} R$$

where the convergence holds in $L^2([0,1])$.

Now take R_0 in $L^\infty([0,1])$, $V(R_0) = \|x \mapsto PR_0(x)e^{-\mu x}\|_{L^\infty([0,1])}$. Applying the Corollary stated above to $x \mapsto PR_0(x)e^{-\mu x}$, there exists a subsequence $(S_{0,n})_n$ of elements of $C_c^2((0,1))$ such that :

$$\forall n \in \mathbb{N},\ \|S_{0,n}\|_{L^\infty([0,1])} \leq V(R_0)$$

and

$$S_{0,n} \rightarrow_{n\rightarrow+\infty} \left[x \mapsto PR_0(x)e^{-\mu x}\right]$$

in $L^2([0,1])$.

Defining, for all integers n, $R_{0,n} = x \mapsto P^{-1}S_{0,n}(x)e^{\mu x}$, one gets :

$$\forall n \in \mathbb{N},\ V(R_{0,n}) \leq V(R_0)$$

and

$$R_{0,n} \rightarrow_{n\rightarrow+\infty} R_0$$

in $L^2([0,1])$. This concludes the proof of Lemma 7.

□

Take the sequence $(R_{0,n})_n \in D(A_n)$ given by preceding Lemma 7 converging to R_0 in $L^2([0,1])$. We denote also $(R_n)_n = (T_n(.)R_{0,n})_n$.

As for all integers n, $V(R_{0,n}) \leq V(R_0)$, $\sigma_{s,n} \geq \sigma_s$ and (2.8) :

$$V(R_{0,n}) < e^{-\mu} \frac{|PBP^{-1}|_\infty \sigma_{s,n} P_{\min}}{||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\mu}|},$$

we can apply Lemma 6 for all integers n,

$$\forall n \in \mathbb{N},\ \forall t \geq 0,\ V(R_n(t,\cdot)) \leq e^{-(\mu \lambda_{\min} - L_{g,\max})t} V(R_{0,n}). \tag{2.62}$$

Now let $t \geq 0$ and $p \in \mathbb{N}$. By Fatou's lemma and the fact that $(R_n(t,\cdot))_n$ converges up to a subsequence towards $R(t,\cdot)$ in the almost everywhere sense (because it converges in $L^2([0,1])$ by Theorem 4) :

$$V_{2p}(R(t,\cdot)) \leq \liminf_{n \to +\infty} V_{2p}(R_n(t,\cdot)).$$

Using the fact that for all $R \in L^\infty([0,1])$, $V_{2p}(R) \leq d^{1/2p} \max_i \{\lambda_i^{-1/2p}\} V(R)$ (Remember (2.49) and the definition of V and V_{2p}), we have :

$$\begin{aligned} V_{2p}(R(t,\cdot)) &\leq d^{1/2p} \max\nolimits_i \{\lambda_i^{-1/2p}\} \liminf\nolimits_{n \to +\infty} V(R_n(t,\cdot)) \\ &\leq d^{1/2p} \max\nolimits_i \{\lambda_i^{-1/2p}\} e^{-(\mu \lambda_{\min} - L_{g,\max})t} \liminf\nolimits_{n \to +\infty} V(R_{0,n}) \\ &\leq d^{1/2p} \max\nolimits_i \{\lambda_i^{-1/2p}\} e^{-(\mu \lambda_{\min} - L_{g,\max})t} V(R_0) \end{aligned}$$

where we have used the fact that $V(R_{0,n}) \leq V(R_0)$ by construction.

Passing to the limit when p goes to infinity, one gets :

$$V(R(t,\cdot)) \leq e^{-(\mu \lambda_{\min} - L_{g,\max})t} V(R_0).$$

Theorem 2 is proven.

CHAPITRE 3

STABILITÉ EXPONENTIELLE BV POUR LES SYSTÈMES DE LOIS DE CONSERVATION SCALAIRES PAR CONTRÔLE AU BORD SATURÉ

Le contenu de ce chapitre correspond à un article. Il a donné lieu à la publication suivante :

M. Dus, **Bv exponential stability for systems of scalar conservation laws using saturated controls**, *SIAM journal on Control and Optimization*, (2021).

Abstract.

Dans ce chapitre, on ajoute une nonlinéarité supplémentaire qui portera sur le flux. On aura donc un système de lois de conservation scalaires à vitesses positives, couplées au bord par une loi de feedback d'abord linéaire puis saturée. Le cadre de travail sera $BV([0,1])$ pour le retour linéaire. En revanche et toujours dans BV, un exemple sera présenté pour montrer qu'il est impossible de stabiliser notre système au delà de la zone de linéarité lorsque le retour est saturé. Ainsi, on se focalisera sur un autre espace plus adapté, $L^\infty([0,1])$, pour lequel on démontrera l'existence d'un bassin d'attraction avec estimation explicite de celui-ci. L'approche utilise principalement la technique de suivi de front (wavefront tracking) décrite dans [20].

La suite de ce chapitre est écrite en anglais.

Sommaire

3.1 Introduction

In this chapter, we are interested in systems of $d \in \mathbb{N}$ scalar conservation laws with strictly positive characteristic velocities. The system under consideration is of the form :

$$\forall i \in [\![1, d]\!], \begin{cases} \partial_t R_i + \partial_x f_i(R_i) & = & 0 \\ R_i(t, 0) & = & g_i(R(t, 1)) \\ R_i(0, x) & = & R_{0,i}(x) \end{cases} \tag{3.1}$$

where $R_i : \mathbb{R}^+ \times [0, 1] \mapsto \mathbb{R}$, $f_i : \mathbb{R} \mapsto \mathbb{R}$ and $g_i : \mathbb{R}^d \mapsto \mathbb{R}$. For coherence, all characteristic velocities are positive and consequently, the boundary condition in (3.1) is adapted. More specifically, we are interested in the stabilization of (3.1) using feedback control laws at the boundary. The problem is equivalent to find sufficient conditions on g such that for any initial data R_0, the solution to (3.1) converges exponentially fast towards zero in the sense that

$$\forall t \geq 0, \ \|R(t, .)\|_X \leq C e^{-\gamma t} \|R_0\|_X \tag{3.2}$$

where $C, \gamma > 0$ are constants independent on t and $\| \cdot \|_X$ is a norm on a Banach space X.

3.1.1 An example

One can consider the basic scalar model for open channel [16, p.44] :

$$\partial_t R + \partial_x \left(k R \sqrt{R} \right) = 0 \tag{3.3}$$

where $R > 0$ is the height of water in the channel, k is a coefficient calculated from the viscous friction, the vertical slope of the channel and the gravity. This simplified model corresponds to a regime where the friction is compensated by the gravity. Written in the flow rate variable $Q = kR^{3/2} > 0$ [1], (3.3) writes :

$$\partial_t Q + \frac{9}{8} k^{2/3} \partial_x \left(Q^{4/3} \right) = 0. \tag{3.4}$$

Now imagine that seven channels are linked as depicted in Figure 3.1 ;

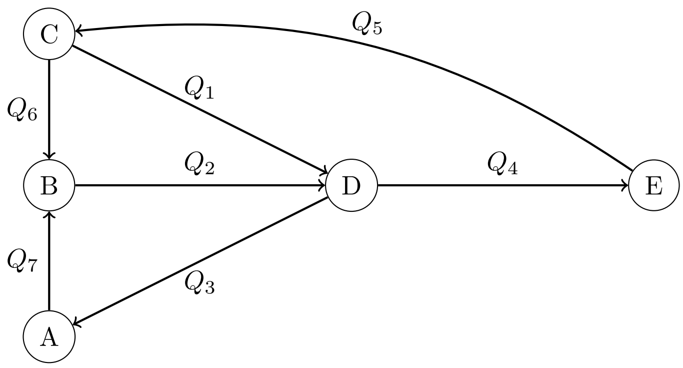


FIGURE 3.1 – An example of system

1. Here the change of variable is done when the solution is regular. With discontinuous solutions, equations (3.3) and (3.4) may not be equivalent.

Flow rates $(Q_1, \cdots, Q_7)$ are functions of space $x \in [0,1]$ and time $t \geq 0$ and are all subject (3.4). To hope for a well-posed problem, it is necessary to define $(Q_1(t,0), \cdots, Q_7(t,0))$. In this paper, it is given by transfer operators associated to nodes A, B, C, D, E. For example, at node B, it is physically relevant to impose that the flow rate in channel 2 is the sum of flow rates from channels 6 and 7 :

$$Q_2(t,0) = Q_6(t,1) + Q_7(t,1) := g_B(Q_6(t,1), Q_7(t,1)).$$

For completeness, paragraph [16, Section 4.2] presents another example of such scalar models coupled by the boundary.

More generally, the family of systems we study constitutes a simplified model for more realistic systems. In [16, Chapter 1], typical examples of hyperbolic PDEs with feedback boundary conditions are cited ; the telegrapher equations for electrical lines, the shallow water (Saint-Venant) equations for open channels [65], the isothermal Euler equations for gas flow in pipelines or even the Aw-Rascle equations [9] for road traffic. It should be noted that in previous examples, there is often an in-domain coupling which is not present in our model. Moreover, fluxes are not scalar which render the analysis far more complicated. In fact, this paper focuses on a simplified version of those systems in order to introduce techniques helping in the complex study of general systems of conservation and balance laws from [16, Chapter 1].

Additionally, the stabilization of similar systems with non-local terms receive more and more attention. We can cite [18] where the authors add a nonlocal zeroth order term to be stabilized. In this article, uncertainties on parameters and on the state of the system are allowed and an adaptive command built from an observer is designed. In [35], authors propose a spectral analysis to stabilize a scalar linear transport equation with a non-local velocity. The control is localized at the boundary. Then, by a Lyapunov analysis they prove a local stability result for the nonlinear version of the system.

3.1.2 Linear feedback

For the case where $g = H \in M_d(\mathbb{R})$ is a linear operator, the literature is quite rich.

When the flux is linear

It can be written as $f(R) = \Lambda R = \text{diag}(\lambda_1, \cdots, \lambda_d)R$ with $\lambda_1, \cdots, \lambda_d > 0$ and the problem of stabilization can be treated for the following classical functional spaces X :

a. Sobolev spaces $W^{m,p}([0,1])$ for $m \in \mathbb{N}$ and $p \in [1, +\infty]$.

b. Spaces $C^m([0,1])$ with $(m \in \mathbb{N})$.

c. $BV([0,1])$.

A literature review was given in Section 1.3.1 describing sufficient conditions on H ensuring exponential stability. Some works are also available when there is an additional source term coupling the equations in the domain. One can cite papers [64, 65, 37, 17, 42] where Lyapunov methods allow to prove exponential stability for the linearized Saint-Venant system.

When the flux is nonlinear for general hyperbolic systems

For some years, many results came out in the case of nonlinear and non scalar fluxes. Only sufficient conditions of stability are given and most of the time this stability is only proved to be local. Also in this case, a literature review is given in Section 1.3.1 where special attention was given to [32] where authors study the BV stabilization with a linear feedback ($g(R) = H$), of a general 2×2 hyperbolic system.

In this chapter, we also place ourselves in a BV context and find a sufficient condition on H to ensure a global BV stability. Contrary to [32], non scalar fluxes are discarded here. In this case, solutions are only proved to exist for small initial data. This is why, we rather consider scalar decentralized fluxes (see Section 1.3) for which solutions exist for any initial data in BV. This hypothesis on the flux is all the more important that when we will study saturated feedback laws, the basin of attraction will be estimated. This would not be possible with solutions defined only for small initial data.

3.1.3 Saturated control law

Let us introduce a matrix $H \in M_d(\mathbb{R})$ potentially unstable in the sense that $\rho_\infty(H) > 1$. Then, it is assumed that there exist matrices $B, K \in M_d(\mathbb{R})$ such that $\rho_\infty(H+BK) < 1$. Finally, the following stabilization problem is considered :

$$\begin{cases} R_t + \partial_x f(R) &= 0 \\ R(t,0) &= HR(t,1) + Bu(t) \\ R(0,x) &= R_0(x). \end{cases} \tag{3.5}$$

If $u(t) := KR(t,1)$, the control is a linear feedback and as $\rho_\infty(H + BK) < 1$, it will be proven in Section 3.3.1, that the solution to (3.5) converges exponentially fast towards zero in BV.

Now suppose that the control is saturated imposing $u(t) := \sigma(KR(t,1))$ with σ defined as a saturation by component *ie* there exists a $\sigma_s > 0$ such that :

$$\forall i \in [\![1,d]\!],\ x \in \mathbb{R},\ \begin{cases} \sigma_i(x) = x & \text{if } |x| \le \sigma_s \\ \sigma_i(x) = \text{sign}(x)\sigma_s & \text{otherwise.} \end{cases}$$

It is natural to ask ourselves if this property of stability is conserved through the saturation. In fact, a basin of attraction exists and an estimation of it is given in Theorem (1) which is the main result of this paper.

Additionally, it is argued that in a BV context, it is not possible to get a basin of attraction bigger than the region of linearity. We rather prove an L^∞ local stability result with an estimation of the basin of attraction. Then, the exponential decay of the BV norm is shown, for solutions whose initial data belongs to the L^∞ basin of attraction.

3.1.4 Scalar conservation laws

The feedback laws being presented, we can now focus on the partial differential equation in itself. The flux f verifies the following Hypothesis 1 of regularity :

Hypothesis 1. *For all $i \in [\![1,d]\!]$, $f_i \in C^1(\mathbb{R})$ and there exist $\alpha_i, \beta_i > 0$;*

$$\forall i \in [\![1,d]\!],\ \alpha_i \le f_i' \le \beta_i.$$

Such hypothesis allows to define the maximal and the minimal velocity :

$$\begin{cases} c_{\max} &:= \max_{i \in [\![1,d]\!]} \beta_i \\ c_{\min} &:= \min_{i \in [\![1,d]\!]} \alpha_i. \end{cases} \tag{3.6}$$

The aim of this section is to give a very short introduction to scalar conservation laws without giving any proof (see [20] for more details).

The set of functions with bounded variations

It is well-known that the space BV is well-adapted for conservation laws (see [20] for instance). This is why, we give the definition and main properties of such a space here :

Definition 5. *Let $R : [0,1] \mapsto \mathbb{R}^d$ be a vector valued function. We say that R has bounded variations if*

$$\forall n \in \mathbb{N},\ \forall\ x_1 < ... < x_n \in [0,1],\ \sum_{i=1}^{n-1} |R(x_{i+1}) - R(x_i)| < \infty$$

where $|\cdot|$ is the canonical euclidean norm.

We denote $TV_{[0,1]}(R) = \sup_{n,\ (x_1,...,x_n)} \left\{ \sum_{i=1}^{n-1} |R(x_{i+1}) - R(x_i)| \right\}$ the total variation of R. $BV([0,1])$ is the space of vector valued functions with bounded variations and it is a Banach space when $BV([0,1])$ is embedded with the norm $\|.\|_{BV([0,1])}$ defined as

$$\forall R \in BV([0,1]),\ \|R\|_{BV([0,1])} = TV_{[0,1]}(R) + \|R\|_{L^1([0,1])}. \tag{3.7}$$

The reason why we consider this space is because any function with bounded variations has a left and a right limit at each point x of $[0,1]$. Hence, it is easy to define the trace operator and impose a boundary condition. Moreover, $BV([0,1])$ has a very interesting property of compactness which will be very useful when we will pass to the limit in the Lyapunov analysis of approximating solutions. These properties are summed up in a lemma and a theorem :

Lemma 8. *Let $R : [0,1] \mapsto \mathbb{R}^d$ with bounded variations. Then for all $x \in (0,1)$, the left and right limit*

$$R(x^-) = \lim_{y \to x^-} R(y), \quad R(x^+) = \lim_{y \to x^+} R(y)$$

exist.

Moreover, $R(0^+)$ and $R(1^-)$ are also well defined and R has at most countably many point of discontinuities.

Proof. This is an adaptation of [20, Lemma 2.1]. □

Defining the value of R at each jump by $R(x) = R(x^+)$, we can say that R is right continuous in the L^1 equivalence class. The following theorem is from Helly and states the compactness of $BV([0,1])$ in $L^1_{loc}(\mathbb{R}^+, L^1([0,1]))$.

Theorem 5. *[20, Theorem 2.4] Let $(R_\nu)_\nu$ be a sequence of functions from $\mathbb{R}^+ \times [0,1]$ into $\mathbb{R}^d$ such that there exist constants C, M and L satisfying*

$$\forall \nu > 1,\ \forall x \in [0,1],\ \forall t \geq 0,\ TV_{[0,1]}(R_\nu(t,.)) \leq C,\ |R_\nu(t,x)| \leq M, \tag{3.8}$$

and

$$\forall 0 \leq t, s \leq T,\ \|R_\nu(t,.) - R_\nu(s,.)\|_{L^1([0,1])} \leq L|t-s|. \tag{3.9}$$

Then there exists a subsequence $(R_\mu)_\mu$ converging strongly towards a certain R in $L^1_{loc}(\mathbb{R}^+, L^1([0,1]))$ and this limit satisfies (3.8)-(3.9) with R_ν replaced by R.

Entropy

The concept of entropy is primordial in order to guaranty uniqueness of solutions to conservation laws. This is why we recall some basic definitions in this section.

If one considers the conservation law $\partial_t R + \partial_x f(R) = 0$ in the usual weak sense :

$$\forall \phi \in C_c^1((0,T)\times(0,1);\mathbb{R}^d), \int_0^T \int_0^1 (\partial_t \phi R + \partial_x \phi f(R)) = 0,$$

it is commonly known that this PDE (associated with fixed boundary and initial conditions) can have several weak solutions (see Example 4.3 from [20]). In order to restrain the set of solutions, an entropy functional was introduced ([36], [77]) and is defined as follows :

Definition 6. *A continuously differentiable convex function $\eta : \mathbb{R}^d \mapsto \mathbb{R}$ is called an entropy for the conservation law $\partial_t R + \partial_x f(R) = 0$ with entropy flux $q : \mathbb{R}^d \mapsto \mathbb{R}$, if*

$$\forall R \in \mathbb{R}^d,\ D\eta(R)\cdot Df(R) = Dq(R).$$

For scalar conservation laws of the form $\partial_t u + \partial_x f_1(u) = 0$, every convex function is an entropy and the usual choice is $\eta(u) := |u-k|$ with flux $q(u) := (f_1(u) - f_1(k))\mathrm{sign}(u-k)$ where k is an arbitrary real. Knowing this, we introduce the notion of entropy solution to (3.1).

Definition 7. *Under Hypothesis 1, we say that $R \in L^\infty_{loc}(\mathbb{R}^+, BV([0,1]))$ is an entropy solution on $[0,T]$ to the system*

$$\begin{cases} \partial_t R + \partial_x f(R) = 0 \\ R(.,0) = g(R(.,1)) \\ R(0,.) = R_0 \in BV([0,1]), \end{cases} \tag{3.10}$$

if :

—

$$\forall k \in \mathbb{R}^d,\ \sum_{i=1}^d \int_0^T \int_0^1 \left\{ |R_i - k_i|\partial_t\phi + (f_i(R_i) - f_i(k_i))sign(R_i - k_i)\partial_x\phi \right\} dxdt \geq 0 \tag{3.11}$$

for all $\phi \geq 0$ and $\phi \in C_c^1((0,T)\times(0,1);\mathbb{R})$.

— *$R(0,.) = R_0$ in the almost everywhere sense.*

— *$R(.,0^+) = g(R(.,1^-))$ in the almost everywhere sense.*

Remark 13. Here the entropy functional and its flux are defined for all k in $\mathbb{R}^d$ by

$$\forall R \in \mathbb{R}^d,\ \eta_k(R) = \sum_{i=1}^d |R_i - k_i|,\ q_k(R) = \sum_{i=1}^d (f_i(R_i) - f_i(k_i))\mathrm{sign}(R_i - k_i). \tag{3.12}$$

Moreover, equation (3.11) can be rewritten as

$$\partial_t \eta_k(R) + \partial_x q_k(R) \leq 0$$

in a weak sense. Hence entropy solutions are the solutions of (3.1) which make the entropy η decrease.

3.1.5 The contribution

Now that all the notions have been introduced, we can be more specific concerning the main contributions of this paper :

- State and prove a well-posedness result of (3.1) in a BV context.
  To help us in the task, we use front tracking techniques from DiPerna [43] and Bressan [20] to get an entropy solution in the domain considered. To deal with the boundary condition, the article [32] is the reference work. One could use results from [32] for which well-posedness is proven for system of 2×2 equations. Here, the proof is simpler and adapted to the context of scalar equations.
- State and prove a global exponential stability result for linear feedback laws.
  To our knowledge, no global stabilization result holds for feedback laws of the form $R(t,0) = HR(t,1)$ in a BV entropy context. The article [90] proposes also a feedback law of the form $R(t,0) = g(\|R(t,.)\|_{L^1})$. However, in physical systems the L^1 norm of the solution is not always accessible by observations. Additionally, the article [32] which considers a 2×2 system of conservation laws gives only a local stabilization result for an entropy solution.
- The key result of this paper is the statement and the proof of a local exponential stability result for saturated feedback laws. We will see that this is not possible in a BV context. To our knowledge, only [49] has studied this kind of saturated feedback laws in an L^∞ context and for the case of constant characteristic velocities.

3.1.6 Outline

In Section 3.2, we present and prove an approximation and a well-posedness result for the entropy BV solution to (3.1). The technique of front tracking are mainly used. Then in Section 3.3, a sufficient condition for global BV stability is given in the case of a linear feedback. Additionally, we give a sufficient condition for the local L^∞ stability in the case of a saturated feedback with an estimation of the basin of attraction. Finally, Section 3.5 is devoted to concluding remarks and perspectives.

3.2 Well-posedness and approximation results

This section is devoted to the well-posedness of (3.10). Additionally, we prove the existence of a suitable approximation by piecewise constant functions of the solution to (3.10). This sequence of approximation is crucial for the stability analysis.

3.2.1 Piecewise constant entropy solutions

Piecewise constant functions play an important role in the theory of BV solutions to conservation laws. Let us recall the definition of what a piecewise constant function is in our context.

Definition 8. *An element R of $L^\infty_{loc}(\mathbb{R}^+, BV([0,1]))$ is piecewise constant if for all $T > 0$, R viewed as a function defined on $[0,T] \times [0,1]$ is constant on a finite number of polyhedra. The edges of such polyhedra are called the fronts of R. Additionally, the absolute value of the jump across the front is called the intensity of the front.*

In this paper, the concept of approximating sequence of piecewise constant functions (PCF) is used in the proof of stability and well-posedness.

Definition 9. *$(R_\nu)_\nu$ is an approximating sequence of PCFs of an entropy solution R to* (3.10) *if :*

- *For $\nu > 1$ fixed, R_ν is piecewise constant in the sense of Definition 8 and takes its values in $2^{-(n+1)\nu}\mathbb{Z}$ on strips*

$$\{(x,t) \mid 0 \leq x \leq 1,\ \max\{(x+n-1)/c_{\max}, 0\} \leq t \leq (x+n)/c_{\max}\}$$

  for $n \in \mathbb{N}$. The velocities of fronts are all bounded from below by $c_{\min}$ and from above by $c_{\max}$ (see (3.6) *for the definition of $c_{\min}$ and $c_{\max}$).*
- *For $\nu > 1$ fixed, no more than one front at a time can interact with the right boundary.*
- *For $\nu > 1$ fixed, if at a time $t \geq 0$ several fronts interact, the sum of intensities of outgoing fronts is inferior to the sum of intensities of ingoing fronts.*
- *The sequence $(R_\nu(0,.))_\nu$ converges toward R_0 in $BV([0,1])$.*
- *The approximated boundary condition is verified :*

$$\forall n \in \mathbb{N},\ \forall t\ s.t\ \frac{n}{c_{\max}} \leq t \leq \frac{n+1}{c_{\max}},\ R_\nu(t,0^+) = g_{(n+2)\nu}(R_\nu(t,1^-)) \tag{3.13}$$

  where :

$$\forall R \in \mathbb{R}^d,\ \forall \nu > 1,\ g_\nu(R) = 2^{-\nu}\big(E(2^\nu g(R))\big). \tag{3.14}$$

- $\forall t \geq 0,\ \Delta t > 0,$

$$TV_{[0,1]}(R(t,.)) \leq \limsup_{\nu\to+\infty} \sup_{s\in[t,t+\Delta t]} TV_{[0,1]}(R_\nu(s,.))$$

  and

$$\|R(t,.)\|_{L^\infty([0,1])} \leq \limsup_{\nu\to+\infty} \sup_{s\in[t,t+\Delta t]} \|R_\nu(s,.)\|_{L^\infty([0,1])}. \tag{3.15}$$

3.2.2 The result of well-posedness and approximation

Now we give the first result of this paper :

Theorem 6. *Under Hypothesis 1 and for all $R_0 \in BV([0,1])$, $g \in Lip(\mathbb{R}^d, \mathbb{R}^d)$, there exists a unique entropy solution $R \in L^\infty_{loc}(\mathbb{R}^+, BV([0,1]))$ to* (3.10)*. Moreover, there exists an approximating sequence of PCF $(R_\nu)_\nu$ of the entropy solution R.*

Proof. A proof is given in Appendix A for the existence and Appendix B for the uniqueness. □

3.3 Lyapunov analysis

Before going into the stability analysis, the functional TV_H defined on the space BV, is intoduced. For all matrices H in $M_d(\mathbb{R})$, it is defined as follows :

$$\forall R \in BV([0,1]),\ TV_H(R) = TV_{[0,1]}(R) + |HR(1^-) - R(0^+)|, \tag{3.16}$$

where $R(1^-)$ and $R(0^+)$ has to be understood as the left and right limits of the function R at $x = 1$ and $x = 0$.

Moreover, the Hypothesis 2 is imposed :

Hypothesis 2. *The feedback matrix H verifies :*

$$\rho_\infty(H) < 1.$$

Remark 14. By [32, Remark 1.4],

$$\forall M \in M_d(\mathbb{R}),\ \rho_\infty(M) = \rho_1(M^T) = \rho_1(M) = \rho_\infty(M^T).$$

The following lemma ensures the equivalence between TV_H and $\|\cdot\|_{BV([0,1])}$.

Lemma 9. *Assume Hypothesis 2. The functional TV_H defined in* (3.16) *is a norm on $BV([0,1])$ equivalent to the norm $\|\cdot\|_{BV([0,1])}$ defined in* (3.7)*. Moreover, there exists a constant $C > 0$ such that*

$$\forall R \in BV([0,1]),\ \|R\|_{L^\infty([0,1])} \le C\ TV_H(R). \tag{3.17}$$

Proof. We first prove the following claim :

$$\forall R \in \mathbb{R}^d,\ |R| \le C|R - HR|. \tag{3.18}$$

Let $P \in D_d^+(\mathbb{R})$ such that

$$|PHP^{-1}|_\infty < 1.$$

The map $\|\cdot\|_\infty : \begin{cases} M_d(\mathbb{R}) & \to & \mathbb{R}^+ \\ M & \mapsto & |PMP^{-1}|_\infty \end{cases}$ defines an algebra norm on $M_d(\mathbb{R})$ and $\|H\|_\infty < 1$. Hence, $I_d - H$ is invertible, which gives (3.18) with $C := |(I-H)^{-1}|$.

$$\begin{aligned} TV_H(R) &= TV_{[0,1]}(R) + |HR(1^-) - R(0^+)| \\ &\le TV_{[0,1]}(R) + |HR(1^-) - HR(0^+)| + |HR(0^+) - R(0^+)| \\ &\le TV_{[0,1]}(R) + |H||R(1^-) - R(0^+)| + |H - I_d||R(0^+)| \\ &\le (1+|H|)TV_{[0,1]}(R) + |H - I_d||R(0^+)|. \end{aligned}$$

Take $x \in [0,1]$, by the triangle inequality,

$$\begin{aligned} TV_H(R) &\le (1+|H|)TV_{[0,1]}(R) + |H - I_d||R(0^+) - R(x)| + |H - I_d||R(x)| \\ &\le (1+|H|+|H-I_d|)TV_{[0,1]}(R) + |H - I_d||R(x)|. \end{aligned}$$

Integrating with respect to x on $[0,1]$, one obtains :

$$\begin{aligned} TV_H(R) &\le (1+|H|+|H-I_d|)TV(R) + |H - I_d|\|R\|_{L^1([0,1])} \\ &= C\|R\|_{BV([0,1])}. \end{aligned}$$

where $C = 1 + |H| + |H - I_d|$.

To get the converse inequality, we remark that by (3.18),

$$|R(1^-)| \le C|HR(1^-) - R(1^-)|.$$

As a consequence,

$$\begin{array}{rcl}\|R\|_{BV([0,1])} & = & TV_{[0,1]}(R)+\|R\|_{L^1([0,1])} \\ & \leq & TV_{[0,1]}(R)+|R(1^-)|+\|R-R(1^-)\|_{L^1([0,1])} \\ & \leq & 2TV_{[0,1]}(R)+C|HR(1^-)-R(1^-)| \\ & \leq & 2TV_{[0,1]}(R)+C|HR(1^-)-R(0^+)|+C|R(0^+)-R(1^-)| \\ & \leq & (2+C)TV_{[0,1]}(R)+C|HR(1^-)-R(0^+)|\end{array}$$

and both norms are equivalent. Concerning the L^∞ estimate (3.17), take a couple $(x,y)\in[0,1]^2$ and using again the triangle inequality

$$|R(x)|\leq|R(x)-R(y)|+|R(y)|\leq TV_{[0,1]}(R)+|R(y)|.$$

Integrating with respect to y on $[0,1]$, one gets

$$|R(x)|\leq TV_{[0,1]}(R)+\|R\|_{L^1([0,1])}=\|R\|_{BV([0,1])}.$$

And as this is true for all x in $[0,1]$,

$$\|R\|_{L^\infty([0,1])}\leq\|R\|_{BV([0,1])}.$$

The equivalence between the norms $\|\cdot\|_{BV([0,1])}$ and TV_H proved earlier allows to deduce (3.17). □

3.3.1 Lyapunov analysis for the unsaturated system

In this section, we consider the following system

$$\left\{\begin{array}{lcl}\partial_t R+\partial_x f(R) & = & 0 \\ R(.,0) & = & HR(.,1) \\ R(0,.) & = & R_0\in BV([0,1])\end{array}\right. \tag{3.19}$$

where the feedback operator g presented in the introduction is replaced by a matrix $H\in M_d(\mathbb{R})$.

The main theorem of this section is presented here :

Theorem 7. *Under Hypothesis 2 and if* $0<\gamma<-\log(\rho_\infty(H))$*, then the unique entropy solution of* (3.19) *satisfies*

$$\forall t\geq 0,\ \|R\|_{BV([0,1])}\leq Ce^{-\gamma c_{\min}t}\|R_0\|_{BV([0,1])}$$

where $C>0$ *is a constant which does not depend on* R_0 *and* t*.*

A candidate Lyapunov functional first introduced by Glimm [59] and then by Coron et al [32] applies well to piecewise constant functions and is defined by :

Definition 10. *Let* R *be a piecewise constant function on* $[0,1]$ *and taking its values in* $\mathbb{R}^d$*. Take* $i\in[\![1,d]\!]$ *:*

- *We denote* $x_{i,1}<x_{i,2}<\cdots<x_{i,n_i}$ *the discontinuities of* R_i *(*n_i *being the number of discontinuities).*
- *For all* $j\in[\![1,n_i]\!]$, $r^l_{i,j}$, $r^r_{i,j}$ *designate the respective left and right state of* R_i *around* $x_{i,j}$*.*

The Lyapunov functional $\mathcal{L}$ evaluated at R writes

$$\mathcal{L}(R) = \sum_{i=1}^{d} P_i \sum_{j=1}^{n_i} |r^r_{i,j} - r^l_{i,j}| e^{-\gamma x_{i,j}} + \sum_{i=1}^{d} P_i |[HR]_i(1^-) - R_i(0^+)| \tag{3.20}$$

where $\gamma > 0$ and $P = \mathrm{diag}\{P_i, i \in [\![1,d]\!]\} \in D_d^+(\mathbb{R})$ will be selected later.

Remark 15. Obviously, there exists a constant $C(H,P,\gamma) > 1$ such that for all R piecewise constant :

$$\frac{\mathcal{L}(R)}{C(H,P,\gamma)} \leq TV_H(R) \leq C(H,P,\gamma)\mathcal{L}(R). \tag{3.21}$$

Remark 16. In our case, the boundary terms in (3.20) are not zero since the boundary condition is approximated by (3.13).

Theorem 7 is proved using a piecewise approximation of the solution for which the exponential decay of the Lyapunov functional $\mathcal{L}$ is established. As a last step, we pass to the limit.

Proof. We consider $(R_\nu)_\nu$ an approximating sequence of PCFs of the entropy solution R in the sense of Definition 9. Such a sequence exists by Theorem 6. The following lemma asserts the exponential stability of the approximation :

Lemma 10. *If $0 < \gamma < -\log(\rho_\infty(H))$. Then, for all $P \in D_d^+(\mathbb{R}^d)$ such that $|P^{-1}H^T P|_\infty < e^{-\gamma}$, there exists $\tilde{\nu}(P,H,\gamma)$ such that*

$$\forall \nu > \tilde{\nu},\ \forall t \geq 0,\ \mathcal{L}(R_\nu) \leq e^{-\gamma c_{\min} t}\mathcal{L}(R_{0,\nu}) + \frac{E(c_{\max}t)+1}{2^\nu}\sum_{i=1}^{d} P_i. \tag{3.22}$$

Proof. Fix $\nu > 1$, $P \in D_d^+(\mathbb{R}^d)$ such that $|P^{-1}H^T P|_\infty < e^{-\gamma}$ and time $0 \leq t \leq 1/c_{\max}$.

Three cases are to be considered :

— (Case 1) If at time t there is no interaction between two fronts nor between a front and the boundary, then $\mathcal{L}(R_\nu)$ is differentiable and because the boundary term is constant locally around t one gets :

$$\begin{aligned}\frac{d\mathcal{L}(R_\nu(t,.))}{dt} &= -\gamma \sum_{i=1}^{d} P_i \sum_{j=1}^{n_i} \frac{dx_{i,j}}{dt} |r^r_{i,j} - r^l_{i,j}| e^{-\gamma x_{i,j}} \\ &\leq -\gamma c_{\min} \sum_{i=1}^{d} P_i \sum_{j=1}^{n_i} |r^r_{i,j} - r^l_{i,j}| e^{-\gamma x_{i,j}}.\end{aligned}$$

Here, we used the fact that for all integers $i \in [\![1,d]\!]$, characteristic velocities $\frac{dx_{i,j}}{dt}$ are bounded from below by $c_{\min} > 0$. Finally, by the definition of $\mathcal{L}(R_\nu(t,.)))$,

$$\begin{aligned}\frac{d\mathcal{L}(R_\nu(t,.))}{dt} \leq\ & -\gamma c_{\min}\mathcal{L}(R_\nu(t,.))) \\ & +\gamma c_{\min} \textstyle\sum_{i=1}^{d} P_i |[HR_\nu]_i(t,1^-) - R_{\nu,i}(t,0^+)| \\ \leq\ & -\gamma c_{\min}\mathcal{L}(R_\nu(t,.))) + \frac{\gamma c_{\min}}{2^\nu}\sum_{i=1}^{d} P_i\end{aligned} \tag{3.23}$$

where we used (3.13) with g replaced by H to get last equation.

— (Case 2) When a front interaction happens, the total variation is non increasing by construction and as a consequence

$$\mathcal{L}(R_\nu(t^+,.)) - \mathcal{L}(R_\nu(t^-,.)) \leq 0.$$

Here we used the third point of Definition 9.

— (Case 3) When an interaction of a front with the boundary happens, computations are a bit more difficult. Suppose that such a front is of type $i \in [\![1,d]\!]$ and has $(R_{i,l}, R_{i,r})$ as respective left and right state (see Figure 3.2). We note its intensity by $I_i := |R_{i,l} - R_{i,r}|$. Note that as R_ν takes its values in $2^{-\nu}\mathbb{Z}$ on the triangle $\{(x,t) \mid 0 < t < x/c_{\max}\}$:

$$I_i \geq 2^{-\nu}. \tag{3.24}$$

Moreover, recall that simultaneous interactions of fronts with the boundary are forbidden by construction. Using the approximate boundary condition (3.13) with g replaced by the linear operator H, it holds

$$\begin{aligned}\mathcal{L}(R_\nu(t^+,.)) - \mathcal{L}(R_\nu(t^-,.)) \leq & \sum_{j=1}^{d} P_j |H_{j,i}(R_{i,r} - R_{i,l})| - e^{-\gamma} I_i P_i \\ & + 2^{-2\nu+2} \textstyle\sum_{j=1}^{d} P_j.\end{aligned} \tag{3.25}$$

The second term on the right-hand side of (3.25) corresponds to the leaving front (which is of type i). The first term results from the entering fronts at the left boundary. Note that an entering front of type $j \in [\![1,d]\!]$ may rather be a fan of fronts (see Figure 3.2). This is not problematic because the sum of the intensities of the fronts composing the fan is equal to the difference of extremal states of the fan by construction (see Appendix A for details). The last term in (3.25) corresponds to the approximation of the boundary condition (3.13).

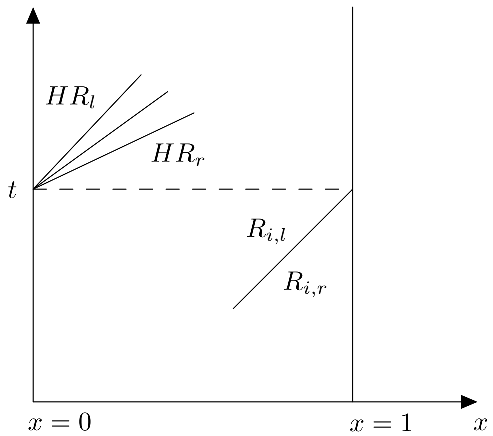


FIGURE 3.2 – Case 3

Then, using the definition of $|\cdot|_\infty$ and (3.24), one gets :

$$\begin{aligned}\mathcal{L}(R_\nu(t^+,.)) - \mathcal{L}(R_\nu(t^-,.)) \leq & \ \Big(\sum_{j=1}^d \frac{P_j}{P_i}|H_{j,i}| - e^{-\gamma}\Big)P_i I_i \\ & +2^{-\nu+2}\textstyle\sum_{j=1}^d P_j I_i \\ \leq & \ \Big(|P^{-1}H^T P|_\infty + 2^{-\nu+2}\sum_{j=1}^d P_j/P_i \\ & -e^{-\gamma}\Big)P_i I_i.\end{aligned}$$

Remark 17. Here we see why the approximated boundary condition (3.13) is essential. Thanks to it, the error term $2^{-2\nu+2}\sum_{j=1}^d P_j$ coming from the approximation of g by g_ν can be bounded by the intensity $I_i \geq 2^{-\nu}$ of the front hitting the right boundary.

As $|P^{-1}H^T P|_\infty - e^{-\gamma} < 0$ by assumption, we can take ν sufficiently large say $\nu \geq \tilde{\nu}(P,H,\gamma)$ such that

$$\mathcal{L}(R_\nu(t^+,.)) - \mathcal{L}(R_\nu(t^-,.)) \leq 0$$

(Case 2) and (Case 3) can occur only a finite number of times on finite time intervals because R_ν is piecewise constant in the sense of Definition 8. Consequently, one can integrate (3.23) with respect to time to get :

$$\forall 0 \leq t \leq 1/c_{\max},\ \mathcal{L}(R_\nu(t,.)) \leq e^{-\gamma c_{\min} t}\mathcal{L}(R_{0,\nu}) + \frac{1}{2^\nu}\sum_{i=1}^d P_i.$$

For time $n/c_{\max} \leq t \leq (n+1)/c_{\max}$ where n is an integer, one easily proves by induction that :

$$\forall n/c_{\max} \leq t \leq (n+1)/c_{\max}\ \mathcal{L}(R_\nu(t,.)) \leq e^{-\gamma c_{\min} t}\mathcal{L}(R_{0,\nu}) + \frac{n+1}{2^\nu}\sum_{i=1}^d P_i.$$

This ends the proof of Lemma 10. □

Now, we conclude on the proof of Theorem 7 taking $t \geq 0$ fixed. By (3.22) and (3.21), there exists a constant $C > 0$ such that

$$\forall \nu > 0,\ TV_H(R_\nu(t,.)) \leq C\Big(e^{-\gamma c_{\min} t}TV_H(R_{0,\nu}) + \frac{E(c_{\max}t)+1}{2^\nu}\sum_{i=1}^d P_i\Big).$$

Using the equivalence between the norm TV_H and the norm $\|\cdot\|_{BV([0,1])}$,

$$\forall \nu > 0,\ \|R_\nu(t,.)\|_{BV([0,1])} \leq C\Big(e^{-\gamma c_{\min} t}\|R_{0,\nu}\|_{BV([0,1])} + \frac{E(c_{\max}t)+1}{2^\nu}\sum_{i=1}^d P_i\Big) \qquad (3.26)$$

where the constant $C > 0$ may have changed.

As $(R_\nu)_\nu$ is an approximating sequence of PCFs of R, one has :

$$\begin{cases} \lim\limits_{\nu\to\infty} R_\nu(0,.) & = \quad R_0 \in BV([0,1]) \\ \forall \tau \geq 0,\ d\tau > 0,\ TV_{[0,1]}(R(\tau,.)) & \leq \quad \limsup\limits_{\nu\to\infty}\ \sup\limits_{s\in[\tau,\tau+d\tau]} TV_{[0,1]}(R_\nu(s,.)). \end{cases}$$

Moreover, by Remark A.4,

$$\forall \tau \geq 0, \ \lim_{\nu \to \infty} \|R_\nu(\tau, .) - R(\tau, .)\|_{L^1([0,1])} = 0.$$

We have for all $dt > 0$,

$$\begin{array}{rcl} \|R(t,.)\|_{BV([0,1])} & \leq & \limsup\limits_{\nu\to\infty}\Big(\sup\limits_{s\in[t,t+dt]} TV_{[0,1]}(R_\nu(s,.)) + \|R_\nu(t,.)\|_{L^1([0,1])}\Big) \\ & \leq & \limsup\limits_{\nu\to\infty} \sup\limits_{s\in[t,t+dt]} \Big(TV_{[0,1]}(R_\nu(s,.)) + \|R_\nu(s,.)\|_{L^1([0,1])}\Big) \\ & = & \limsup\limits_{\nu\to\infty} \sup\limits_{s\in[t,t+dt]} \|R_\nu(s,.)\|_{BV([0,1])} \\ & \leq & C\limsup\limits_{\nu\to\infty} \Big(e^{-\gamma c_{\min} t}\|R_{0,\nu}\|_{BV([0,1])} + \frac{E(c_{\max}t)+1}{2^\nu}\sum_{i=1}^d P_i\Big) \\ & = & Ce^{-\gamma c_{\min} t}\|R_0\|_{BV([0,1])} \end{array}$$

where (3.26) has been used to get the fourth equation.

This finishes the proof of Theorem 7. □

3.3.2 Stability analysis for the saturated system

In this section, we consider the following system :

$$\begin{cases} \partial_t R + \partial_x f(R) & = & 0 \\ R(.,0) & = & [H \cdot + B\sigma(K\cdot)]R(.,1) \\ R(0,.) & = & R_0 \in BV([0,1]). \end{cases} \tag{3.27}$$

The deadzone function is defined by :

$$\forall R \in \mathbb{R}^d, \ \phi(R) = \sigma(R) - R \tag{3.28}$$

and Hypothesis 3 :

Hypothesis 3. *The matrices H, B, K are chosen such that :*

$$\rho_\infty(H + BK) < 1.$$

Here the main result is different since we prove local exponential stability (Proposition 1). It is not possible to study directly the problem of BV stability because of the lack of contractivity of the saturation σ.

Remark 18. *In Figure 3.3, we represent the boundary operator $H \cdot + B\sigma(K\cdot)$ for $d = 1$, $H = 2$, $B = 1$, $K = -1.5$ and $\sigma_s = 2$. Except for the zone of linearity, the boundary operator is only 2-Lipschitz. As a consequence, it is possible to construct a front whose left/right states are arbitrary close to the zone of linearity and whose intensity increases after a passage through the feedback operator. This is why it is not possible to get a basin of attraction in BV norm larger than the zone of linearity. We rather prove the L^∞ local stability with a basin of attraction in L^∞.*

This section is devoted to the proof of the following proposition and theorem (the definition of $\|\cdot\|_{\infty,P}$ is given in the section notation) :

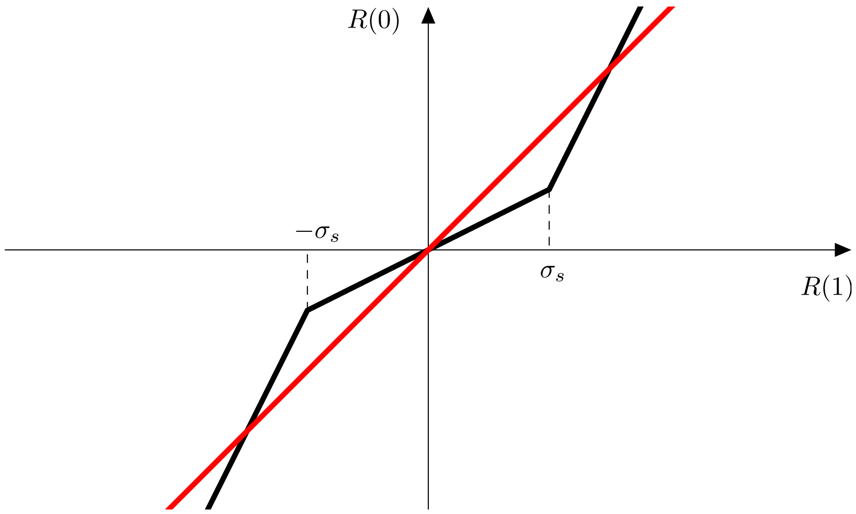


FIGURE 3.3 – The feedback operator (black line) compared with the graph of the function $R(0) = R(1)$ (red line)

Proposition 1. *Under Hypothesis 3, if* $0 < \gamma < -\log(\rho_\infty(H+BK))$*. Then, for all* $P \in D_d^+(\mathbb{R}^d)$ *such that* $|P(H+BK)P^{-1}|_\infty \leq e^{-\gamma}$*, there exists a constant* C *depending on* (H, B, K, P, γ) *such that if* $R_0 \in BV([0,1])$ *and if :*

$$\|R_0\|_{\infty,P} < \frac{|PBP^{-1}|_\infty P_{\min}\sigma_s}{\left||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty|PKP^{-1}|_\infty - e^{-\gamma}\right|}. \tag{3.29}$$

Then, the unique entropy solution $R \in L^\infty_{loc}(\mathbb{R}^+, BV([0,1]))$ *of* (3.27) *satisfies,*

$$\forall t \geq 0, \|R(t,.)\|_{L^\infty([0,1])} \leq Ce^{-\gamma c_{\min} t}\|R_0\|_{L^\infty([0,1])} \tag{3.30}$$

where C *depends on the parameters of the problem but not* R_0*.*

For cases where $\rho_\infty(H) > 1$, the denominator in (3.29) is not zero :

Remark 19. If $\rho_\infty(H) > e^{-\gamma}$, then we claim that for all $P \in D_d^+(\mathbb{R})$:

$$|P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty|PKP^{-1}|_\infty - e^{-\gamma} > 0.$$

Proof of the claim of Remark 19. Let P be in $D_d^+(\mathbb{R})$. As $\rho_\infty(H) > e^{-\gamma}$,

$$|PHP^{-1}|_\infty > e^{-\gamma}.$$

This gives by the triangle inequality :

$$|P(H+BK)P^{-1}|_\infty + |PBKP^{-1}|_\infty > e^{-\gamma}.$$

Finally, by the fact that

$$\forall A, B \in M_d(\mathbb{R}),\ |PABP^{-1}|_\infty \leq |PAP^{-1}|_\infty|PBP^{-1}|_\infty,$$

we have :

$$|P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty|PBP^{-1}|_\infty > e^{-\gamma}$$

and the claim is proved. □

The following theorem is a consequence of Proposition 1 and constitutes a BV exponential stability result.

Theorem 8. *Under the conditions of Proposition 1,*

$$\forall t \geq 0,\ \|R(t,.)\|_{BV([0,1])} \leq Ce^{-\gamma c_{\min} t}\|R_0\|_{BV([0,1])}$$

where C depends on the parameters of the problem but not R_0.

Let us assume for the time being Proposition 1 and prove Theorem 8 :

Proof of Theorem 8. Equation (3.30) implies that at a certain time denoted $t^\star$ depending on $\|R_0\|_{L^\infty([0,1])}$, the solution enters in the zone of linearity and stays in it. Then, Theorem 7 implies :

$$\forall t \geq t^\star,\ \|R(t,.)\|_{BV([0,1])} \leq Ce^{-\gamma c_{\min}(t-t^\star)}\|R(t^\star,.)\|_{BV([0,1])} \tag{3.31}$$

where C depends on $H, B, K, P, \gamma, \sigma_s$.

Then, for $t \leq t^\star$, one can prove using the same techniques from Section 3.1 that :

$$\forall 0 \leq t \leq t^\star,\ \|R(t,.)\|_{BV([0,1])} \leq e^{\nu t}\|R_0\|_{BV([0,1])} \tag{3.32}$$

where $\nu > 0$ is a constant depending on $c_{\max}, \gamma$ and a Lipschitz constant of the feedback operator $H + B\sigma(K)$. From (3.32) and (3.31), one gets :

$$\forall t \geq 0,\ \|R(t,.)\|_{BV([0,1])} \leq Ce^{-\gamma c_{\min} t}\|R_0\|_{BV([0,1])}$$

where C depends on the parameters of the problem and on $\|R_0\|_{L^\infty([0,1])}$. As the bound (3.29) holds, we can conclude that C does not depend on $\|R_0\|_{L^\infty([0,1])}$ and the corollary is proved. □

The following lemma is useful for the proof of Proposition 1.

Lemma 11. *Let $R \in \mathbb{R}^d$ be such that :*

$$|PR|_\infty \leq \frac{|PBP^{-1}|_\infty P_{\min}\sigma_s}{\big||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty|PKP^{-1}|_\infty - e^{-\gamma}\big|}. \tag{3.33}$$

Then,

$$|P(HR + B\sigma(KR))|_\infty \leq e^{-\gamma}|PR|_\infty.$$

Proof. Let i be in $[\![1,d]\!]$. If $sat_i(R) := \{j \in [\![1,d]\!],\ |\ |[KR]_j| > \sigma_s \text{ and } B_{i,j} \neq 0\}$ is empty, then :

$$\begin{aligned} P_i|HR + B\sigma(KR)|_i &= P_i|(H+BK)R)|_i \\ &\leq |P(H+BK)P^{-1}|_\infty|PR|_\infty \\ &\leq e^{-\gamma}|PR|_\infty. \end{aligned}$$

If the set $\mathrm{sat}_i(R)$ is not empty, then :

$$\begin{array}{rl} P_i|HR+B\sigma(KR)|_i = & P_i|(H+BK)R+B\phi(KR))|_i \\ \leq & \sum_{j=1}^d P_i|(H+BK)_{i,j}R_j| \\ & +\sum_{j\in sat_i(R)}^d P_i|B_{i,j}|(|[KR]_j|-\sigma_s) \\ \leq & \sum_{j=1}^d P_i|(H+BK)_{i,j}\frac{1}{P_j}P_jR_j| \\ & +\sum_{j\in sat_i(R)}^d P_i|B_{i,j}|\frac{P_j}{P_j}(|[KR]_j|-\sigma_s) \\ \leq & |P(H+BK)P^{-1}|_\infty|PR|_\infty \\ & +|PBP^{-1}|_\infty(|PKP^{-1}|_\infty|PR|_\infty - P_{\min}\sigma_s) \\ \leq & e^{-\gamma}|PR|_\infty \end{array}$$

where we have used the hypothesis (3.33) to get the last inequality. □

Now the focus is on the proof of Proposition 1.

Proof of Proposition 1. Take $P \in D_d^+(\mathbb{R})$ such that $|P(H+BK)P^{-1}|_\infty < e^{-\gamma}$ and $R_0 \in BV([0,1])$ satisfying (3.29). We consider $(R_\nu)_\nu$ an approximating sequence of PCFs of the entropy solution R in the sense of Definition 9. Such a sequence exists because of Theorem 6. Then, we analyze the exponential damping of R_ν for a fixed $\nu > 1$. As $(R_{0,\nu})_\nu$ converges towards R_0 in $BV([0,1])$, it holds for ν sufficiently large :

$$\|R_{0,\nu}\|_{\infty,P} \leq \frac{|PBP^{-1}|_\infty P_{\min}\sigma_s}{\left||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty|PKP^{-1}|_\infty - e^{-\gamma}\right|} \tag{3.34}$$

We first recall the definition of $c_{\min}, c_{\max}$ the respective minimum and maximum velocity, in (3.6). Let $t \leq 1/c_{\min}$ and $x > c_{\max}t$ be in $[0,1]$. Constructing the light cone enclosed by line with slopes $1/c_{\min}$ and $1/c_{\max}$ and passing through (t,x), the following estimate is obtained :

$$|PR_\nu(t,x)|_\infty \leq \|R_{0,\nu}\|_{\infty,P} \tag{3.35}$$

The argument of the light cone can be justified by the fact that the L^∞ norm does not increase by fronts interaction (see Appendix Appendix A.2.3) and because fronts velocities belongs to $[c_{\min}, c_{\max}]$.

When $x \leq c_{\max}t$, constructing the light cone enclosed by lines with slopes $1/c_{\min}$ and $1/c_{\max}$ and passing through (t,x), one gets :

$$|PR_\nu(t,x)|_\infty \leq \max\{\|R_{0,\nu}\|_{\infty,P}, \sup_{t\in[0,1/c_{\min}]} |PR_\nu(t,0)|\}.$$

The boundary condition gives :

$$|PR_\nu(t,x)|_\infty \leq \max\{\|R_{0,\nu}\|_{\infty,P}, \sup_{t\in[0,1/c_{\min}]} |P[H\cdot + B\sigma(K\cdot)]R_\nu(t,1)|\}.$$

By (3.35) applied to $x=1$ and (3.34), hypothesis of Lemma 11 are verified and consequently :

$$|PR_\nu(t,x)|_\infty \leq \max\{\|R_{0,\nu}\|_{\infty,P}, e^{-\gamma}\|R_{0,\nu}\|_{\infty,P}\} \leq \|R_{0,\nu}\|_{\infty,P}.$$

Next we proceed by induction on intervals of the form $t \in [n/c_{\min}, (n+1)/c_{\min}]$ with $n \in \mathbb{N}$. Suppose that :

$$\forall t \in [n/c_{\min}, (n+1)/c_{\min}],\ \|R_\nu(t,\cdot)\|_{\infty,P} \leq e^{-\gamma n}\|R_{0,\nu}\|_{\infty,P}.$$

Let $(n+1)/c_{\min} \leq t \leq (n+2)/c_{\min}$ and x be in $[0,1]$. Constructing the light cone enclosed by lines with slopes $1/c_{\min}$ and $1/c_{\max}$ and passing through (t,x), one gets the existence of a $t^\star \in [n/c_{\min}, (n+2)/c_{\min}]$ such that :

$$|PR_\nu(t,x)|_\infty \leq |PR_\nu(t^\star,0)|_\infty \leq |P[H\cdot + B\sigma(K\cdot)]R_\nu(t^\star,1)|. \tag{3.36}$$

Using same reasoning as in the case $n = 0$, it can be proved that :

$$\|R_\nu(t^\star,\cdot)\|_{\infty,P} \leq \|R_\nu(n/c_{\min},\cdot)\|_{\infty,P}.$$

Hence, by the hypothesis of induction :

$$|PR_\nu(t^\star,1)|_\infty \leq \|R_\nu(t^\star,\cdot)\|_{\infty,P} \leq e^{-\gamma n}\|R_{0,\nu}\|_{\infty,P} \leq \|R_{0,\nu}\|_{\infty,P}. \tag{3.37}$$

As a consequence, by (3.34) :

$$|PR_\nu(t^\star,1)|_\infty \leq \frac{|PBP^{-1}|_\infty P_{\min}\sigma_s}{\left||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\gamma}\right|}$$

Thus, we can use Lemma 11 in (3.36) to get :

$$\begin{array}{rcl} |PR_\nu(t,x)|_\infty & \leq & e^{-\gamma}|PR_\nu(t^\star,1)|_\infty \\ & \leq & e^{-\gamma}\|R_\nu(t^\star,\cdot)\|_{\infty,P}. \end{array}$$

Hence by the induction hypothesis,

$$\|R_\nu(t,\cdot)\|_{\infty,P} \leq e^{-\gamma}\|R_\nu(t^\star,\cdot)\|_{\infty,P} \leq e^{-\gamma(n+1)}\|R_{0,\nu}\|_{\infty,P}$$

where (3.37) has been used. To conclude, we have :

$$\forall t \geq 0,\ \|R_\nu(t,\cdot)\|_{\infty,P} \leq e^{-\gamma(c_{\min}t-1)}\|R_{0,\nu}\|_{\infty,P}.$$

It remains to prove the exponential decay for the solution R. It suffices to use property (3.15) and to take a sequence of initial data piecewise constant such that :

$$\forall \nu > 1,\ \|R_{0,\nu}\|_{\infty,P} \leq \|R_0\|_{\infty,P}.$$

Owing this, one passes to the limit as ν goes to infinity to get :

$$\forall t \geq 0,\ \|R(t,\cdot)\|_{\infty,P} \leq e^{-\gamma(c_{\min}t-1)}\|R_0\|_{\infty,P}.$$

This ends the proof of Proposition 1. □

3.4 Numerical results

Here, we study a numerical example with saturation and show the relevance of the estimation of the region of attraction (3.29).

3.4.1 Relevance of the estimation of the basin of attraction

In this section, an example of system of scalar conservation laws is analyzed for $d = 2$ with saturated feedback control law with $\sigma_s = 1$. Matrices are defined as follows.

$$H = \begin{pmatrix} 0 & 1.1 \\ 1 & 0 \end{pmatrix}, \; B = I_2, \; K = \begin{pmatrix} 0 & -0.1050 \\ -0.1045 & 0 \end{pmatrix}.$$

We take a nonlinear flux $f(R) = \Lambda R + 0.2(\arctan(R_1), \arctan(R_2))$ with

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{pmatrix}.$$

The open-loop system can be represented by the graph given in Figure 3.4 :

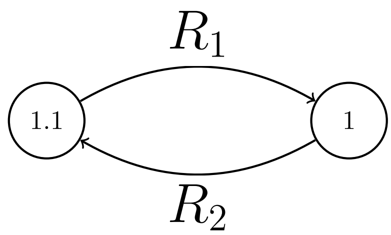


FIGURE 3.4 – The open-loop system

We recall the estimation of the basin of attraction for $\gamma > 0$ and $P \in D_d^+(\mathbb{R})$:

$$\|R_0\|_{\infty,P} \leq \frac{|PBP^{-1}|_\infty P_{\min}\sigma_s}{\left||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - e^{-\gamma}\right|}. \tag{3.38}$$

Using an optimization routine from python, we calculate $P \in D_d^+(\mathbb{R})$ such that $|P(H+BK)P^{-1}|_\infty$ is minimal. The code gives :

$$P = \begin{pmatrix} 0.974 & 0 \\ 0 & 1.026 \end{pmatrix}.$$

To estimate the largest region of attraction, we take $\gamma = 0$ in (3.38) which gives the following criteria of stability :

$$\|R_0\|_{\infty,P} \leq \frac{|PBP^{-1}|_\infty P_{\min}\sigma_s}{\left||P(H+BK)P^{-1}|_\infty + |PBP^{-1}|_\infty |PKP^{-1}|_\infty - 1\right|}. \tag{3.39}$$

3.4.2 Numerical simulations

Still keeping the matrices from previous section, we take a certain range of initial data R_0 constant on $[0,1]$ belonging to the estimated region of attraction and simulate the behavior of the solution. For example, one can take R_0 constant with value in $(-40,40)^2$ and look if the solution does not blow up at infinite time in L^∞ norm. We briefly describe the scheme used. The space step is $dx = 1/N$ ($N \in \mathbb{N}^*$) and the time step $dt > 0$ such that the following CFL condition holds :

$$c_{max}\frac{dt}{dx} \leq 1 - \xi \tag{3.40}$$

with

$$0 < \xi < 1.$$

For computation, we take $dt = 10^{-2}$ and $\frac{dt}{dx} = 0.4$. Doing so, the space-time mesh is given by :

$$\forall n \in \mathbb{N},\ 1 \leq j \leq N,\ \left\{ \begin{array}{lll} x_j & := & (j - 1/2)dx \\ t^n & := & ndt. \end{array} \right.$$

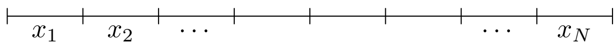


FIGURE 3.5 – The space grid

The scheme is a finite volume one given by the minmod slope limiter method [79]. It is of the form :

$$\frac{R_j^{n+1} - R_j^n}{dt} + \frac{f_{j+1/2}^n - f_{j-1/2}^n}{dx} = 0$$

where for all $n \geq 0$, $1 \leq i \leq 2$:

$$\begin{array}{rl} \forall 2 \leq j \leq N-1,\ f_{i,j+1/2}^n = & f_i\left(R_{i,j}^n + minmod(\frac{R_{i,j}^n - R_{i,j-1}^n}{dx}, \frac{R_{i,j+1}^n - R_{i,j}^n}{dx})\frac{dx}{2}\right) \\ f_{i,N+1/2}^n = & f_i\left(R_{i,N}^n\right) \\ f_{i,N-1/2}^n = & f_i\left(R_{i,N-1}^n\right) \\ f_{i,3/2}^n = & f_i\left(R_{i,1}^n\right) \\ f_{i,1/2}^n = & f_i\left(\left[(H + B\sigma(\cdot))R_N^n\right]_i\right). \end{array}$$

The minmod function is defined below for all $a, b \in \mathbb{R}$:

$$minmod(a,b) := \left\{ \begin{array}{ll} 0 & \text{if } ab \leq 0 \\ a & \text{if } ab \geq 0 \text{ and } |a| \leq |b| \\ b & \text{otherwise.} \end{array} \right.$$

One can cite [48] for the study of stability of such numerical system subject to linear boundary conditions.

In Figure 3.6, contours correspond to the rate of exponential decay wrt L^∞ norm of the numerical solution for a time window of 50 seconds. If it is negative, the solution decays exponentially in norm. If it is positive, we have exponential divergence. The orange square is the estimated region of attraction while the blue one encloses the zone where saturation does not occur at $t = 0$.

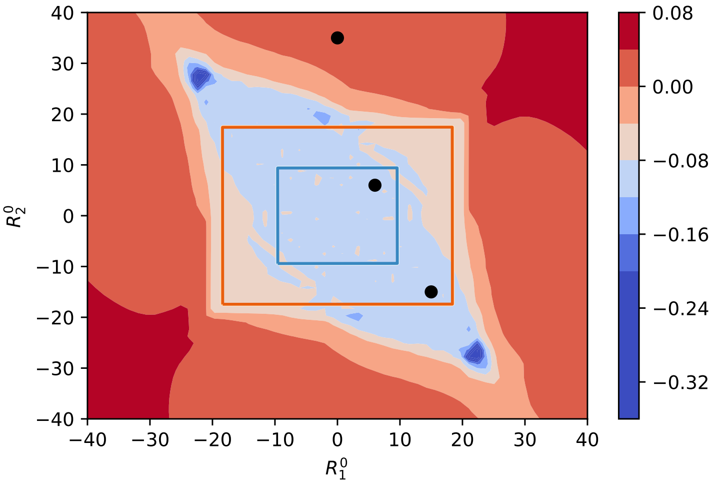
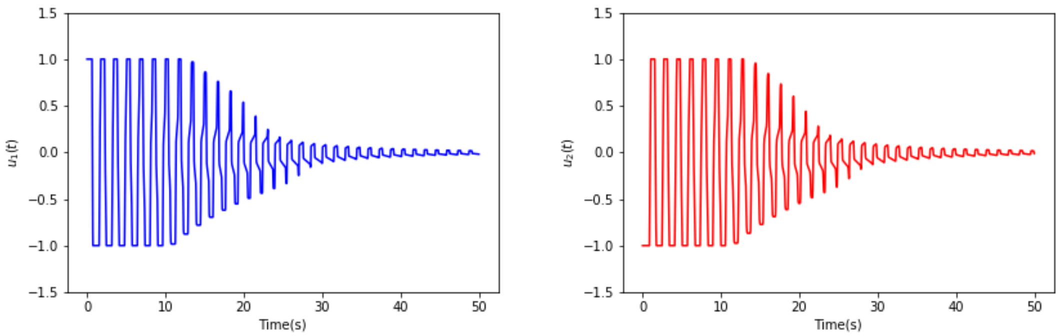


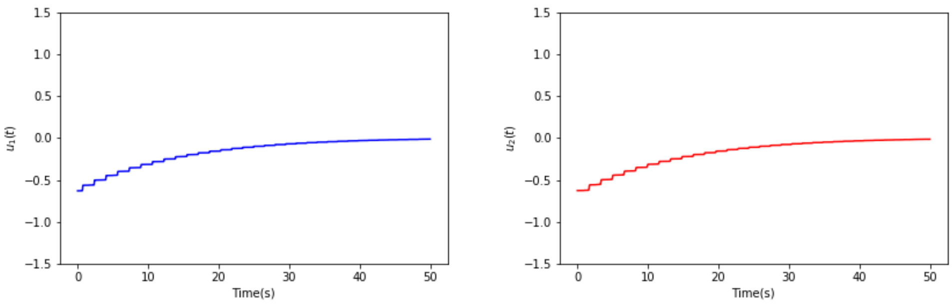
FIGURE 3.6 – The basin of attraction

We also pick three initial data R_0 in different regions of Figure 3.6 and observe the dynamic of the solution. For example, one can take $R_0(x) = (15, -15)$ on $[0, 1]$. The black dots in Figure 3.6 correspond to these initial data. The values of controls are plotted in Figure 3.7-3.9 where $u_1(t) = \sigma(KR(t, 1))_1$ and $u_2(t) = \sigma(KR(t, 1))_2$. Concerning Figure 3.7, we observe that some saturation occurs from $t = 0$ until time $t \approx 13$, then the solution enters in the zone of linearity. The Figure 3.8 represents a case where the system stays in the zone of linearity whereas in Figure 3.9, the initial data is out of the basin of attraction.



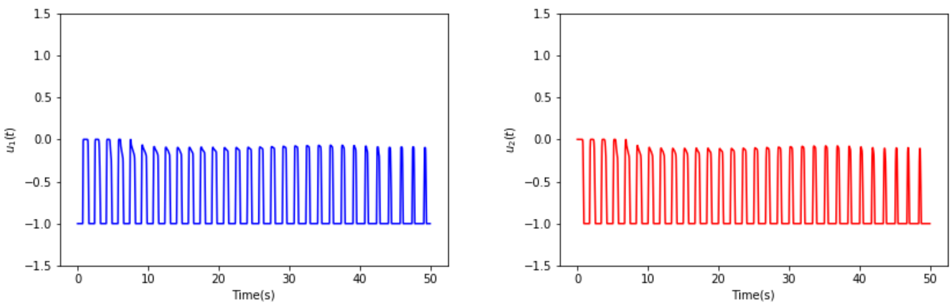
(a) The control $u_1(t)$ (b) The control $u_2(t)$

FIGURE 3.7 – The case $R_0(x) = (15, -15)$



(a) The control $u_1(t)$ (b) The control $u_2(t)$

FIGURE 3.8 – The case $R_0(x) = (6, 6)$



(a) The control $u_1(t)$ (b) The control $u_2(t)$

FIGURE 3.9 – The case $R_0(x) = (0, 35)$

3.5 Conclusion

The well-posedness for a wide class of systems of scalar conservation laws with boundary unsaturated and saturated feedback laws was established. The ρ_∞ criteria was established in the BV context for linear feedback laws. Then, for saturated feedback laws, we proved with an example that estimating a basin of attraction in BV was not relevant. We rather gave an estimation of the basin of attraction in L^∞ and deduce the exponential decay of the BV norm of solutions whose initial data belongs to this basin of attraction.

Some questions remain open. The estimation (3.29) may not be optimal. Moreover, a method of maximizing the basin of attraction where the matrix K is the variable of optimization is not given in this article. This is not an easy task since criterion (3.29) is not convex with respect to K. Finally, the other big gap to bridge is the stabilization of general systems of conservation laws the main difficulty coming from the well-posedness. The initial-boundary value problem for hyperbolic systems of conservation laws is indeed a very delicate matter, even when no characteristic speed vanishes. We refer to [97, 3, 27, 44, 32] and the references therein.

3.A Existence of a solution

All this section is dedicated to the proof of the existence result of Theorem 6.

3.A.1 The approximated problem

Let $\nu \geq 1$, i in $[\![1,d]\!]$ and define $f_{\nu,i}$ the piecewise affine approximation of f_i coinciding with f_i at all $2^{-\nu}j$ nodes ($j \in \mathbb{Z}$) by :

$$f_{\nu,i}(s) = \frac{s-2^{-\nu}j}{2^{-\nu}} f_i(2^{-\nu}(j+1)) + \frac{2^{-\nu}(j+1)-s}{2^{-\nu}} f_i(2^{-\nu}j) \text{ for } s \in [2^{-\nu}j, 2^{-\nu}(j+1)].$$

The sequence $(f_\nu)_\nu$ is introduced in order to construct a piecewise constant entropy solution. The following lemma gives its main properties :

Lemma 12. *For all $T > 0$, there exists a constant $C(g,T)$ such that for all $\nu > 1$ and $R_{0,\nu}$ piecewise constant taking its values in $2^{-\nu}\mathbb{Z}$, there exists R_ν piecewise constant in the sense of Definition 8 verifying the following assertions :*

— *The approximated boundary condition* (3.13) *is verified.*

— *Two fronts cannot interact simultaneously with the right boundary.*

— $\forall k \in \mathbb{R}^d, \phi \in C^1_c((0,T)\times(0,1);\mathbb{R})$ *:*

$$\int_0^T \int_0^1 \eta_k(R_\nu)\partial_t\phi + q_k(R_\nu)\partial_x\phi dxdt \geq -C(g,T)\frac{TV(R_{0,\nu})}{\nu}\|\phi\|_{L^\infty(\mathbb{R}^+\times[0,1])}. \tag{3.41}$$

where :

$$\begin{cases} \eta_k(R_\nu(t,x)) &= \displaystyle\sum_{i=1}^d |R_{\nu,i}(t,x) - k_i| \\ q_k(R_\nu(t,x)) &= \displaystyle\sum_{i=1}^d |f_i(R_{\nu,i}(t,x)) - f_i(k_i)|. \end{cases} \tag{3.42}$$

— *The following bounds hold :*

$$\forall t \leq T,\ TV_{[0,1]}(R_\nu(t,.)) \leq C(g,T)TV_{[0,1]}(R_{0,\nu}).$$

$$\forall t \leq T,\ \|R_\nu(t,.)\|_{L^\infty([0,1])} \leq C(g,T)\|R_{0,\nu}(t,.)\|_{L^\infty([0,1])}. \tag{3.43}$$

Proof. See Appendix 3.A.2 □

It is relatively easy to construct piecewise constant functions that make the entropy decrease. The main rules of construction are presented in the following lemma.

Lemma 13. *[Characterization of entropy piecewise constant functions]*

A piecewise constant function R in the sense of Definition 8 verifies the condition of entropy decay (3.11) *if and only if for all integers i in $[\![1,d]\!]$ and all fronts $\gamma(t)$ of R_i,*

— *The Rankine-Hugoniot condition holds for R_i^l the left state and R_i^r the right state :*

$$\dot{\gamma}(t)[R_i^r(t) - R_i^l(t)] = f_i(R_i^r(t)) - f_i(R_i^l(t)). \tag{3.44}$$

— *If $R_i^l(t) < R_i^r(t)$ then :*

$$\forall \alpha \in [0,1], f_i(\alpha R_i^r(t) + (1-\alpha)R_i^l(t)) \geq \alpha f_i(R_i^r(t)) + (1-\alpha) f_i(R_i^l(t)). \tag{3.45}$$

— *If $R_i^l(t) > R_i^r(t)$ then :*

$$\forall \alpha \in [0,1], f_i(\alpha R_i^r(t) + (1-\alpha)R_i^l(t)) \leq \alpha f_i(R_i^r(t)) + (1-\alpha) f_i(R_i^l(t)). \tag{3.46}$$

Proof. This corresponds to [20, Theorem 4.4]. □

We call conditions (3.45)-(3.46), the entropy decay conditions of fronts ; it selects values before and after the front such that the entropy of the solution decreases with time. If a front verifies such conditions, we say that the front is **entropic**.

3.A.2 Proof of Lemma 12

Now we prove Lemma 12 constructing step by step a piecewise constant solution. We begin by solving a Riemann problem to get a solution near $t = 0$.

The Riemann problem

Let i be an integer of $[\![1, d]\!]$, $\nu > 1$ and $R_i^l, R_i^r \in \mathbb{R}$ be two states. We recall techniques from [20, pp .108-113] to solve the Riemann problem associated to (R_i^l, R_i^r) when taking f_ν as flux. There are two cases to consider :

— If $R_i^l < R_i^r$. Then, we consider f_i^* the largest convex function inferior to $f_{\nu,i}$ on $[R_i^r, R_i^l]$. Denote also $w_0 := R_i^l < w_1 < w_2 < \cdots < w_n := R_i^r$ the states where $f_i^{*,\prime}$ jumps. We give an example for $n = 2$ on Figure 3.10 :

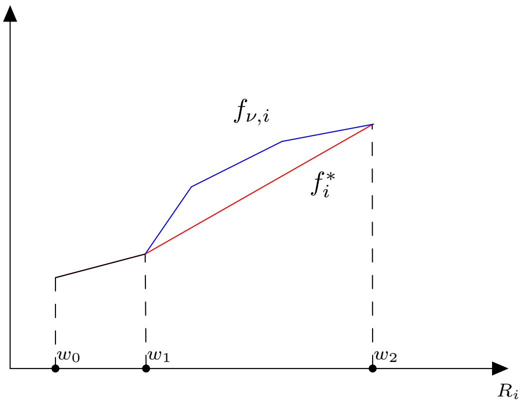


FIGURE 3.10 – The case $R_i^l < R_i^r$

Introducing the speeds

$$\lambda_l = \frac{f_{\nu,i}(w_l) - f_{\nu,i}(w_{l-1})}{w_l - w_{l-1}}, \ l \in [\![1, n]\!], \tag{3.47}$$

we define the solution to the Riemann problem as :

$$R_{\nu,i}(t,x) = \begin{cases} w_0 \text{ if } x < t\lambda_1 \\ w_l \text{ if } t\lambda_l < x < t\lambda_{l+1}, \ l \in [\![1, n-1]\!] \\ w_n \text{ if } x > t\lambda_n. \end{cases} \tag{3.48}$$

This solution is entropic because it is piecewise constant, all fronts are entropic (3.45) and satisfy the Rankine-Hugoniot condition (3.44).

— If $R_i^l > R_i^r$. Then, we consider f_i^* the smallest concave function larger than $f_{\nu,i}$. Denote also $w_0 := R_i^l > w_1 > w_2 > \cdots > w_n := R_i^r$ the states where $f_i^{*,\prime}$ jumps. We give an example for $n = 2$ on the Figure 3.11.

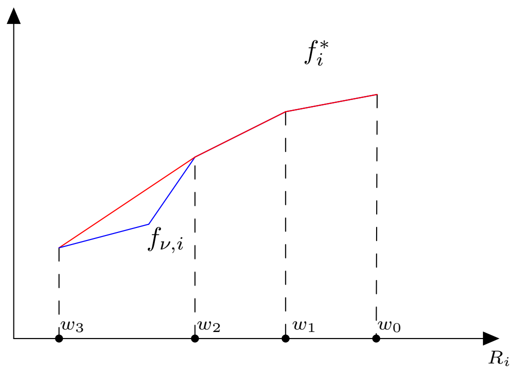


FIGURE 3.11 – The case $R_i^l > R_i^r$

Defining velocities $(\lambda_l)_{l \in [\![1,q]\!]}$ as in (3.47), we define the local solution also as :

$$R_{\nu,i}(t,x) = \begin{cases} w_0 \text{ if } x < t\lambda_1 \\ w_l \text{ if } t\lambda_l < x < t\lambda_{l+1}, \ l \in [\![1, n-1]\!] \\ w_n \text{ if } x > t\lambda_n. \end{cases} \tag{3.49}$$

Local in time solution

Take a fixed $\nu > 1$. Let us define what we will call the limit line $t \mapsto c_{\max} t$ with maximal speed. Thanks to the Riemann solver defined in the previous section, we can find an entropy solution until a front interaction happens. The corresponding picture is given in Figure 3.12.

Dealing with shock interactions

We recall the method described in [20, pp. 111-112]. Two cases have to be considered :

— (Case 1) All the incoming jumps have the same sign. Suppose they are all positive and let us denote $w_0 < w_1 < ... < w_n$ ($n \in \mathbb{N}$) the consecutive "incoming" states. As all incoming fronts are entropic, we have :

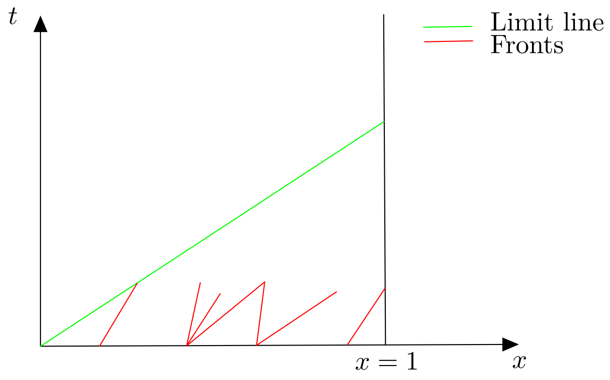


FIGURE 3.12 – The local in time solution

$$\forall \alpha \in [0,1],\ f(\alpha w_i + (1-\alpha)w_{i+1}) \geq \alpha f(w_i) + (1-\alpha) f(w_{i+1}). \tag{3.50}$$

The fact that we have converging fronts gives that the function h built from lines passing through points $(w_i, f(w_i))_{i\in[\![0,n]\!]}$ is concave. Moreover, by (3.50) :

$$\forall w \in [w_0, w_n],\ h(w) \leq f(w). \tag{3.51}$$

Hence, by the concavity of h and (3.51) :

$$\forall \alpha \in [0,1],\ f(\alpha w_0 + (1-\alpha)w_n) \geq h(\alpha w_0 + (1-\alpha)w_n) \geq \alpha f(w_0) + (1-\alpha) f(w_n).$$

Thus, it is possible to link the extremal (w_0 and w_n) states by a unique entropic front whose jump intensity is strictly equal to the sum of the intensities of incoming jumps. Hence, in this case the total variation is conserved and so is the L^∞ norm.

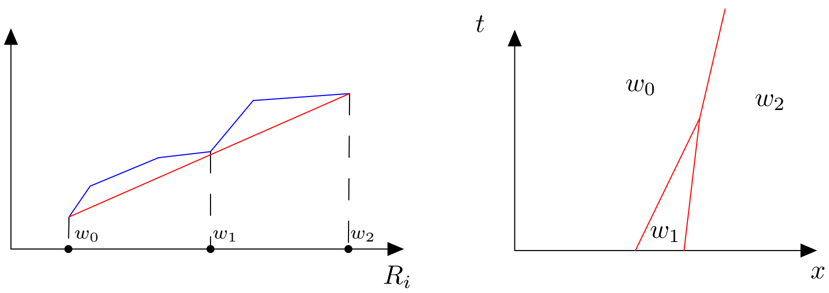


FIGURE 3.13 – All jumps have the same sign

— (Case 2) Not all jumps have the same sign. Let us denote $w_0, w_1, ..., w_n$ ($n \in \mathbb{N}$) the consecutive “incoming” states. It is possible to link the extremal (w_0 and w_n) states using fronts whose jumps have the same sign. To do so, it suffices to solve a Riemann problem between extremal states w_0 and w_n as in Section 3.A.2. Moreover, by the triangle inequality, the total variation decreases at least by $2 \times 2^{-\nu}$. Concerning the L^∞ norm, it is conserved. This is because the $f_{\nu,i}$s are non decreasing.

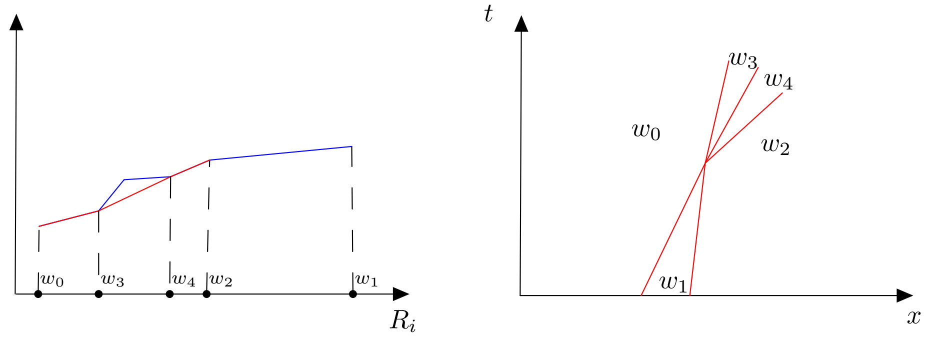


Figure 3.14 – Not all jumps have the same sign

(Case 1) creates a unique front and let the total variation unchanged whereas (Case 2) can create several fronts but the total variation decreases by at least by $2 \times 2^{-\nu}$. Consequently, (Case 2) can happen only a finite number of times and the number of fronts remains bounded as time evolves. As a consequence, it is possible to construct an entropy piecewise constant approximate solution under the limit line verifying

$$\begin{cases} \forall 0 \leq t \leq 1/c_{\max},\ TV_{[c_{\max}t,1]}(R_{\nu,i}(t,.)) & \leq & TV_{[0,1]}(R_{0,\nu,i}) \\ \forall x \in [0,1],\ TV_{[0,x/c_{\max}]}(R_{\nu,i}(.,x)) & \leq & TV_{[0,1]}(R_{0,\nu,i}) \end{cases} \tag{3.52}$$

The corresponding picture is given in Figure 3.15 :

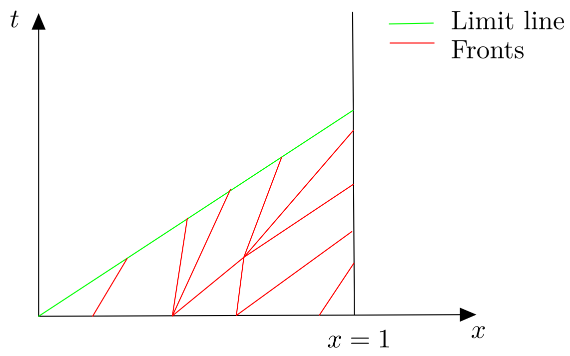


Figure 3.15 – The solution under the limit line

Remark 20. If $n \geq 2$ ($n \in \mathbb{N}$) fronts interacts exactly at $x = 1$ for some time $t > 0$. Then, we modify a bit the velocity of $n - 1$ fronts to prevent this situation. Taking one of such fronts, we denote λ and $\tilde{\lambda}$ the respective former and new velocities. We can choose them such that $|\lambda - \tilde{\lambda}| \leq \frac{1}{\nu}$.

Finishing the construction

To construct locally the solution above the limit line, we impose the boundary condition

$$\forall 0 \leq t \leq 1/c_{\max},\ R_\nu(t, 0^+) = g_{2\nu}(R_\nu(t, 1^-))$$

where we recall that :

$$\forall R \in \mathbb{R}^d, \nu > 1, i \in [\![1, d]\!],\ g_{\nu,i}(R) = 2^{-\nu}\big(E(2^\nu g_i(R))\big).$$

Then, to construct a local solution, we solve the different Riemann problems as in section 3.A.2 this time using the approximated flux $f_{2\nu}$. More precisely if at a time t, $R_i(t^-, 0^+) \neq R_i(t^+, 0^+)$, we solve the Riemann problem with $R_i(t^+, 0^+)$ as left state and $R_i(t^-, 0^+)$ as right state. Hence we are able to get a solution locally above the limit line taking its values in $2^{-2\nu}\mathbb{Z}$.

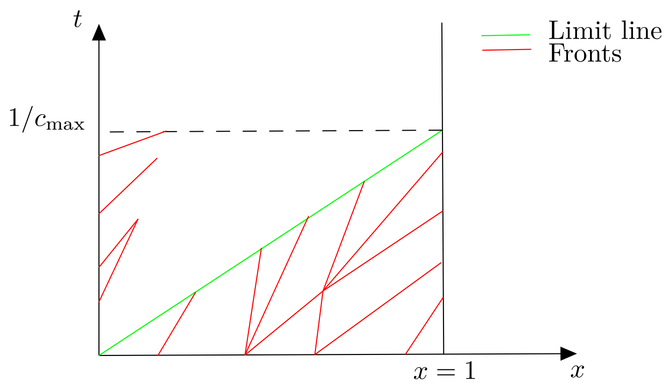


FIGURE 3.16 – The local solution above the limit line

Finally, we extend fronts coming from the zone under the limit line and deal with front interactions as in section 3.A.2 this time using the approximated flux $f_{2\nu}$. The final picture is given in Figure 3.17. This is very important to remark that the picture under the limit line cannot be modified by fronts coming from the left boundary. This is because the limit line has maximal velocity.

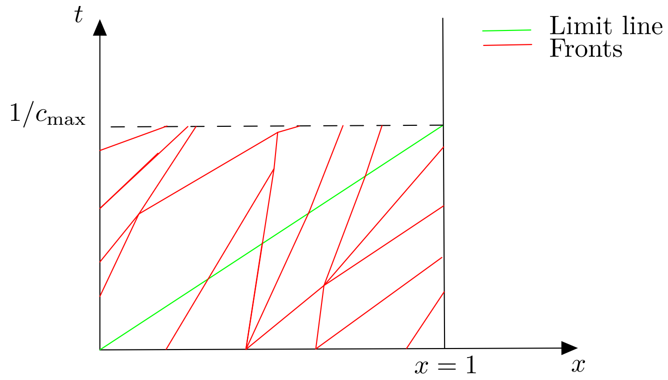


FIGURE 3.17 – The solution

Concerning the total variation, the way we dealt with front interactions prevents the total variation from increasing when we compare the total strength of ongoing fronts with the one of outgoing fronts. As a consequence,

$$\begin{cases} \forall 0 \leq t \leq 1/c_{\max},\ TV_{[0,c_{\max}t]}(R_{\nu,i}(t,.)) \leq & TV_{[0,1]}(R_{0,\nu,i}) \\ & +TV_{[0,1/c_{\max}]}(R_{\nu,i}(.,0^+)) \\ \forall x \in [0,1],\ TV_{[x/c_{\max},1/c_{\max}]}(R_{\nu,i}(.,x)) \leq & TV_{[0,1]}(R_{0,\nu,i}) \\ & +TV_{[0,1/c_{\max}]}(R_{\nu,i}(.,0^+)) \end{cases} \tag{3.53}$$

Conclusion

All previous steps can be repeated on intervals $[k/c_{\max}, (k+1)/c_{\max}]$ and a solution defined for all time is built. Now let $T > 0$. There are several points to verify :

- (Boundary condition) The approximated boundary condition (3.13) is satisfied by construction.
- (Boundary interactions) Two fronts cannot interact simultaneously at the right boundary by construction.
- (Estimate on the total variation). Using (3.52), (3.53) and the fact that R_ν satisfies the approximated boundary condition (3.13), one can deduce that there exists a constant $C(g,T)$ (depending on the Lipschitz constant of g and T) such that

$$\forall 0 \leq t \leq T,\ TV_{[0,1]}(R_\nu(t,.)) \leq C(g,T)TV_{[0,1]}(R_{0,\nu}(t,.)). \tag{3.54}$$

- (Estimate on the entropy) Take a positive test function $\phi \in C_c^1((0,T)\times(0,1))$, $T > 0$ and $k \in \mathbb{R}^d$. Then, by integration by parts, one obtains :

$$\int_0^T \int_0^1 [\eta_k(R_\nu)\partial_t\phi + q_k(R_\nu)\partial_x\phi dxdt = \sum_\alpha [\dot{\gamma_\alpha}(\eta_k(R^r_{\nu,\alpha}) - \eta_k(R^l_{\nu,\alpha})) - (q_k(R^r_{\nu,\alpha}) - q_k(R^l_{\nu,\alpha}))]\phi(t,\gamma_\alpha)$$

where α runs over the discontinuities $t \to (t, \gamma_\alpha(t))$ of R_ν.
We denote P the set of physical fronts *ie* the fronts for which the velocity has not been modified. NP designates the complement of P. As fronts of P are entropic by construction, we have :

$$
\begin{aligned}
&\int_0^T \int_0^1 \eta_k(R_\nu)\partial_t\phi_i + q_k(R_\nu)\partial_x\phi_i dxdt \\
&\geq 0 + \sum_{\alpha\in NP} [\dot{\gamma_\alpha}(\eta_k(R^r_{\nu,\alpha})\eta_k(R^l_{\nu,\alpha})) - (q_k(R^r_{\nu,\alpha}) - q_k(R^l_{\nu,\alpha}))]\phi(t,\gamma_\alpha) \\
&= \sum_{\alpha\in NP} [\tilde{\lambda_\alpha}(\eta_k(R^r_{\nu,\alpha}) - \eta_k(R^l_{\nu,\alpha})) - (q_k(R^r_{\nu,\alpha}) - q_k(R^l_{\nu,\alpha}))]\phi(t,\gamma_\alpha)
\end{aligned}
$$

where we replaced the notation $\dot{\gamma}$ by $\tilde{\lambda}$ to emphasize the fact that it corresponds to a modified velocity (see Remark 20) ; the unmodified "entropic" velocity being denoted λ. Hence,

$$
\begin{aligned}
&\int_0^T \int_0^1 \eta_k(R_\nu)\partial_t\phi + q_k(R_\nu)\partial_x\phi dxdt \\
&\geq \sum_{\alpha\in NP} [\tilde{\lambda_\alpha}(\eta_k(R^r_{\nu,\alpha}) - \eta_k(R^l_{\nu,\alpha})) - (q_k(R^r_{\nu,\alpha}) - q_k(R^l_{\nu,\alpha}))]\phi(t,\gamma_\alpha) \\
&= \sum_{\alpha\in NP} [\lambda_\alpha(\eta_k(R^r_{\nu,\alpha}) - \eta_k(R^l_{\nu,\alpha})) - (q_k(R^r_{\nu,\alpha}) - q_k(R^l_{\nu,\alpha})) \\
&+(\tilde{\lambda_\alpha} - \lambda_\alpha)(\eta_k(R^r_{\nu,\alpha}) - \eta_k(R^l_{\nu,\alpha}))]\phi(t,\gamma_\alpha) \\
&\geq 0 - \tfrac{TV(R_\nu(t,.))}{\nu}\|\phi\|_{L^\infty(\mathbb{R}^+\times[0,1])} \\
&\geq -C(g,T)\tfrac{TV(R_{0,\nu})}{\nu}\|\phi\|_{L^\infty(\mathbb{R}^+\times[0,1])}
\end{aligned}
$$

where we have used chronologically :

- The fact that an unmodified velocity corresponds to an entropy front
- The equation $|\lambda - \tilde{\lambda}| \leq \frac{1}{\nu}$ from Remark 20
- The fact that η_k is 1-Lipschitz
- The estimate (3.54) proven before.

— (L^∞ estimate). Remark that when we solved a Riemann problem, the L^∞ norm did not increase. This is mainly because we are dealing with non decreasing fluxes. The only way for the L^∞ norm to increase is through the boundary condition. As a consequence, the estimate (3.43) holds.

This finishes the proof of Lemma 12.

3.A.3 End of the proof of the existence result

To conclude on the existence, we will use Lemma 12 and Helly's Theorem 5. There are several points to prove :

— (Entropy decay) Take $T > 0$, $R_0 \in BV([0,1])$ and a sequence $(R_{0,\nu})_\nu$ of piecewise constant functions converging to R_0 in BV (such a sequence exists by [20, Lemma 2.2]). For all $\nu > 1$, we denote $(R_\nu)_\nu$ the sequence of piecewise constant functions of Lemma 12.

By Lemma 12, there exists a $C(g,T) > 0$ such that

$$\forall 0 \leq t \leq T,\ TV_{[0,1]}(R_\nu(t,.)) \leq C(g,T)TV_{[0,1]}(R_{0,\nu}).$$

As $\lim_{\nu\to\infty} R_{0,\nu} = R_0 \in BV([0,1])$,

$$\forall 0 \le t \le T,\ TV_{[0,1]}(R_\nu(t,.)) \le C(g,T,R_0). \tag{3.55}$$

Next by (3.43) and the fact that the L^∞ norm of the elements of $(R_{0,\nu})_\nu$ are bounded, we have

$$\forall 0 \le t \le T,\ \|R_\nu(t,.)\|_{L^\infty([0,1])} \le C(g,T,R_0). \tag{3.56}$$

Finally for all $0 \le s,t \le T$ and by the finiteness of the speed of propagation :

$$\begin{aligned} \|R_\nu(t,.) - R_\nu(s,.)\|_{L^1([0,1])} &\le c_{\max}(t-s)\max_{u\in[s,t]} TV_{[0,1]}(R_\nu(u,.)) \\ &\le c_{\max}(t-s)C(g,T,R_0) \end{aligned} \tag{3.57}$$

where we have used (3.55).

By Helly's Theorem (Theorem 5), there exists a subsequence of $(R_\nu)_\nu$ still denoted $(R_\nu)_\nu$ converging in $L^1_{loc}(\mathbb{R}^+, L^1([0,1]))$ to an element $R \in L^\infty_{loc}(\mathbb{R}^+, BV([0,1]))$. Moreover,

$$\forall 0 \le s,t \le T,\ \|R(t,.) - R(s,.)\|_{L^1([0,1])} \le c_{\max}(t-s)C(g,T,R_0). \tag{3.58}$$

As $(f_\nu)_\nu$ converges uniformly towards f on bounded intervals, we can pass to the limit in (3.41) to get (3.11).

— (Initial condition). Let $\varepsilon > 0$ and $s > 0$

$$\begin{aligned} \|R(0,.) - R_\nu(0,.)\|_{L^1([0,1])} \le\ & \|R(0,.) - R(s,.)\|_{L^1([0,1])} \\ & + \|R(s,.) - R_\nu(s,.)\|_{L^1([0,1])} \\ & + \|R_\nu(s,.) - R_\nu(0,.)\|_{L^1([0,1])} \\ \le\ & 2C(g,R_0)s + \|R(s,.) - R_\nu(s,.)\|_{L^1([0,1])}. \end{aligned}$$

where we have used (3.58).

Integrating with respect to s on an interval $[0,t]$ for $0 \le t \le 1/c_{\max}$, one gets

$$\begin{aligned} \|R(0,.) - R_\nu(0,.)\|_{L^1([0,1])} \le\ & C(g,R_0)t \\ & + \frac{1}{t}\int_0^t \|R(s,.) - R_\nu(s,.)\|_{L^1([0,1])} ds \\ \le\ & C(g,R_0)t \\ & + \frac{1}{t}\int_0^{1/c_{\max}} \|R(s,.) - R_\nu(s,.)\|_{L^1([0,1])} ds \end{aligned}$$

Taking $t = \frac{\varepsilon}{2C(g,R_0)}$ and ν sufficiently large such that $\int_0^{1/c_{\max}} \|R(s,.) - R_\nu(s,.)\|_{L^1([0,1])} ds \le \frac{\varepsilon^2}{4C(g,R_0)}$, one finally obtains :

$$\|R(0,.) - R_\nu(0,.)\|_{L^1([0,1])} \le \varepsilon.$$

By the fact that $(R_{0,\nu})_\nu$ converges towards R_0 in $L^1([0,1])$, we deduce that $R(0,.) = R_0$ in a L^1 sense and $R(0,.) = R_0$ almost everywhere.

Remark 21. We can repeat the same procedure for any $t \ge 0$ and

$$\forall t \ge 0,\ \lim_{\nu\to\infty} \|R_\nu(t,.) - R(t,.)\|_{L^1([0,1])} = 0.$$

— (Boundary condition). For the boundary condition, it suffices to consider the variable x as a time variable.

Using (3.52), (3.53) and the approximated boundary condition (3.13), one can easily prove that

$$\forall x \in [0,1],\ TV_{[0,1/c_{\max}]}(R_\nu(.,x)) \leq C(g) TV_{[0,1]}(R_{0,\nu}(.)).$$

As $(R_{0,\nu})_\nu$ is bounded in BV,

$$\forall x \in [0,1],\ TV_{[0,1/c_{\max}]}(R(.,x)) \leq C(g, R_0). \tag{3.59}$$

Additionally, with (3.56) we get the L^∞ estimate

$$\forall \nu > 1,\ \forall x \in [0,1],\ \|R_\nu(.,x)\|_{L^\infty([0,1/c_{\max}])} \leq C(g, R_0). \tag{3.60}$$

Finally, using (3.59) and recalling the definition $c_{\min} := \min_i \alpha_i$ of front velocities, we have for $0 \leq x, y \leq 1$:

$$\forall \nu > 1,\ \|R_\nu(.,x) - R_\nu(.,y)\|_{L^1([0,1/c_{\max}])} \leq \frac{|x-y|}{c_{\min}} C(g, R_0). \tag{3.61}$$

By Helly's Theorem (Theorem 5), $(R_\nu)_\nu$ tends towards R in $L^\infty_{loc}([0,1], L^1([0,1/c_{\max}]))$. Using a similar argument as in the previous item of the proof, one shows that $R(.,0^+) = g(R(.,1^-))$ in the almost everywhere sense on $[0, 1/c_{\max}]$. We repeat the argument to get the same conclusion for all time.

Hence, R is a solution of (3.10) in the sense of Definition 7. It remains to prove that $(R_\nu)_\nu$ is an approximating sequence of PCFs of the entropy solution R in the sense of Definition 9. By construction, $(R_\nu)_\nu$ satisfies the first five points of Definition 9. It remains only to prove the bound :

$$\forall t \geq 0,\ \delta t > 0,\ TV_{[0,1]}(R(t,.)) \leq \limsup_{\nu\to\infty} \sup_{s\in[t,t+\Delta t]} TV_{[0,1]}(R_\nu(s,.)). \tag{3.62}$$

This is a consequence of Helly's Theorem. Indeed, take $t \geq 0$, $\Delta t > 0$ and $n \in \mathbb{N}^*$. Instead of applying Helly's Theorem on an interval of the form $[0,T]$ for the sequence $(R_\nu)_{\nu>1}$, we apply it on the interval $[t, t+\Delta t]$ for the sequence $(R_\nu)_{\nu>n}$.

As

$$\forall s \in [t, t+\Delta t],\ \forall \nu > n,\ TV_{[0,1]}(R_\nu(s,.)) \leq \sup_{\substack{u\in[t,t+\Delta t]\\ \upsilon>n}} TV_{[0,1]}(R_\upsilon(u,.)),$$

we deduce by Helly's Theorem that

$$\forall s \in [t, t+\Delta t],\ TV_{[0,1]}(R(s,.)) \leq \sup_{\substack{u\in[t,t+\Delta t]\\ \upsilon>n}} TV_{[0,1]}(R_\upsilon(u,.)).$$

Passing to the limit as n goes to infinity gives (3.62). To get the estimate (3.15), the proof is similar. The existence part of Theorem 6 is proven.

3.B Uniqueness

We will adapt the method of doubling variables of Kruzhkov to our boundary value problem. Let u, v be two entropy solutions of (3.10) with their respective initial data u_0, v_0.

We will first show the uniqueness on the triangle T_1 :

$$T_1 := \{(t,x) \mid c_{\max}t \leq x \leq 1,\ 0 \leq t \leq 1/c_{\max}\}$$

To do so, let $0 < t \leq 1/c_{\max}$ and define the domain Ω_t by :

$$\Omega_t := \{(s,x);\ 0 \leq s \leq t,\ c_{\max}s \leq x \leq 1\}\,.$$

We give a graphical representation of Ω_t in Figure 3.18.

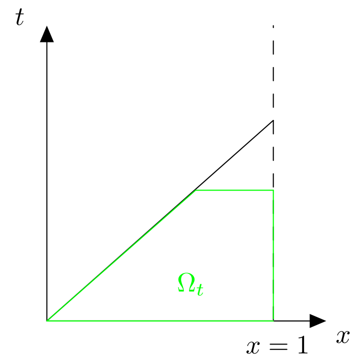


FIGURE 3.18 – The domain Ω_t

Formally, as u in entropy on Ω_t, we have for all $k \in \mathbb{R}^d$

$$\begin{aligned} 0 \quad \geq & \quad \int\int_{\Omega_t} \partial_t \eta_k(u) + \partial_x q_k(u) dx dt \\ = & \quad \int_{1-c_{\max}t}^{1} \eta_k(u) dx - \int_0^1 \eta_k(u_0) dx \\ & + \int_0^t c_{\max}\eta_k(u(\tau, c_{\max}\tau)) - q_k(u(\tau, c_{\max}\tau)) d\tau + \int_0^t q_k(u(1,s)) ds. \end{aligned}$$

The third term is positive because $c_{\max}$ is superior to all the Lipschitz constants of the f_is. The last term is positive since all the f_is are non decreasing. Hence,

$$\int_{1-c_{\max}t}^{1} \eta_k(u) dx \leq \int_0^1 \eta_k(u_0) dx. \tag{3.63}$$

It is equivalent to :

$$\forall k \in \mathbb{R}^d,\ \sum_{i=1}^{d} \int_{1-c_{\max}t}^{1} |u_i(t,x) - k_i| dx \leq \sum_{i=1}^{d} \int_0^1 |u_{0,i}(x) - k_i| dx.$$

Kruzhkov's doubling variable method allows to replace the k_i by the v_i to give :

$$\sum_{i=1}^{d}\int_{1-c_{\max}t}^{1}|u_i(t,x)-v_i(t,x)|dx \leq \sum_{i=1}^{d}\int_{0}^{1}|u_{0,i}(x)-v_{0,i}(x)|dx.$$

As a consequence, the solution is unique on the triangle T_1.

Remark 22. Rigorous justifications of previous computations can be found in the proof of [20, Theorem 6.2].

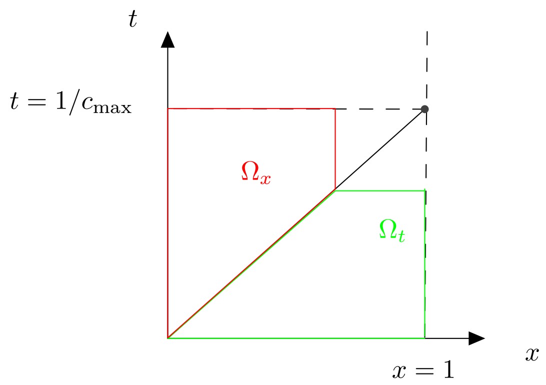


FIGURE 3.19 – The domain Ω_x

Now let x be in $]0,1[$, we apply the same strategy to the set

$$\Omega_x := \{(s,y);\quad 0 \leq y \leq x,\ y/c_{\max} \leq s \leq 1/c_{\max}\}$$

represented in Figure 3.19.

Integrating $\partial_t \eta_k(u) + \partial_x q_k(u) \leq 0$ in Ω_x, one obtains : $\forall x \in [0,1], k \in \mathbb{R}^d$,

$$\int_{1/c_{\max}-x/c_{\max}}^{1/c_{\max}} q_k(u(t,x))dt \leq \int_{0}^{1/c_{\max}} q_k(u(t,0))dt + \int_{0}^{x/c_{\max}} c_{\max}\eta_k(u(t,c_{\max}t)) - q_k(u(t,c_{\max}t))dt.$$

This is equivalent to : $\forall x \in [0,1],\ k \in \mathbb{R}^d$,

$$\begin{aligned}&\sum_{i=1}^{d}\int_{1/c_{\max}-x/c_{\max}}^{1/c_{\max}} (f_i(u_i)-f_i(k_i))\text{sign}(u_i-k_i)dt\\ &\leq \sum_{i=1}^{d}\int_{0}^{1/c_{\max}} (f_i(u_i(t,0))-f_i(k_i))\text{sign}(u_i(t,0)-k_i)dt\\ &+\int_{0}^{x/c_{\max}} c_{\max}\eta_k(u(t,c_{\max}t)) - q_k(u(t,c_{\max}t))dt.\end{aligned}$$

As all the f_i are non decreasing ($q_k \geq 0$) and all $c_{\max}$ Lipschitz : $\forall x \in [0,1],\ k \in \mathbb{R}^d$,

$$\sum_{i=1}^{d}\int_{1/c_{\max}-x/c_{\max}}^{1/c_{\max}}|f_i(u_i)-f_i(k_i)|dt \leq c_{\max}\sum_{i=1}^{d}\int_{0}^{1/c_{\max}}|u_i(t,0)-k_i|dt + c_{\max}\int_0^{x/c_{\max}}\eta_k(u(t,c_{\max}t))dt.$$

Following Kruzhkov's method, the k_is can be replaced by the v_is and for all x in $[0,1]$:

$$\begin{aligned}\sum_{i=1}^{d}\int_{1/c_{\max}-x/c_{\max}}^{1/c_{\max}}|f_i(u_i)-f_i(v_i)|dt &\leq c_{\max}\sum_{i=1}^{d}\int_{0}^{1/c_{\max}}|u_i(t,0)-v_i(t,0)|dt \\ &+ c_{\max}\int_0^{x/c_{\max}}|u(t,c_{\max}t)-v(t,c_{\max}t)|dt.\end{aligned}$$

As u, v satisfy the boundary condition on $[0, 1/c_{\max}]$, there exists a constant $C(g, c_{\max})$ depending on the Lipschitz constant of g such that for all x in $[0,1]$:

$$\sum_{i=1}^{d}\int_{1/c_{\max}-x/c_{\max}}^{1/c_{\max}}|f_i(u_i)-f_i(v_i)|dt \leq C(g,c_{\max})\sum_{i=1}^{d}\int_{0}^{1/c_{\max}}|u_i(t,1)-v_i(t,1)|dt + c_{\max}\int_0^{x/c_{\max}}|u(t,c_{\max}t)-v(t,c_{\max}t)|dt.$$

If $u_0 = v_0$, we have seen that u and v coincide on T_1. This implies that if $u_0 = v_0$, u and v coincide on the segment $\{1\} \times [0, 1/c_{\max}]$ and on the line $(x = c_{max}t, t)$ for $t \leq 1/c_{\max}$. As a consequence,

$$\forall x \in [0,1],\ \sum_{i=1}^{d}\int_{1/c_{\max}-x/c_{\max}}^{1/c_{\max}}|f_i(u_i)-f_i(v_i)|dt = 0.$$

By the monoticity of the f_is, u and v coincide on the triangle T_2 defined by

$$T_2 := \{(t,x)\ |0 \leq x \leq 1,\ x/c_{\max} \leq t \leq 1/c_{\max}\}.$$

To conclude, u and v coincide for $t \leq 1/c_{\max}$ and repeating this argument, we can prove the uniqueness for all time. This finishes the proof of the uniqueness.

CHAPITRE 4

SUR LA STABILITÉ EXPONENTIELLE DES SCHÉMAS À LIMITEUR DE FLUX POUR LES LOIS DE CONSERVATION SCALAIRES SOUMISES À UNE CONDITION AU BORD DISSIPATIVE

Le contenu de ce chapitre correspond à un article soumis chez Mathematics of Control, Signals, and Systems (MCSS).

Abstract.

Dans ce chapitre, on étudie toujours le même type de système de lois de conservations couplées au bord par un feedback linéaire. En gardant le cadre $BV([0,1])$, on discrétise les équations en utilisant les bonnes propriétés des schémas à limiteur de pente. D'une part, un résultat de stabilisation exponentielle est démontré pour la version numérique du système, la méthode consistant à utiliser le formalisme de Harten pour l'inclure dans une étude de Lyapunov discrète. D'autre part, une analyse comparative sera exposée pour montrer l'influence du choix du schéma sur la stabilisation du système numérique en jeu.

La suite de ce chapitre est écrite en anglais.

Sommaire

4.1 Introduction

In this chapter, we are also interested in the BV exponential stability of systems of scalar conservation laws using boundary feedback laws. The system under consideration is a set of d nonlinear scalar conservation laws coupled at the boundary by a square matrix H of size d :

$$\forall 1 \leq i \leq d, \begin{cases} \partial_t R_i + \partial_x [f_i(R_i)] & = & 0 \\ R_i(t,0) & = & [HR(t,1)]_i \\ R_i(0,x) & = & R_i^0(x) \end{cases} \tag{4.1}$$

where $R : \mathbb{R}^+ \times [0,1] \mapsto \mathbb{R}^d$ $(d \in \mathbb{N}^*)$, $f_i : \mathbb{R} \mapsto \mathbb{R}$. For coherence, it is assumed that all characteristic velocities are positive and consequently, the boundary condition in (4.1) is adapted. The case where the sign of characteristic velocities is not fixed is out of the scope of this chapter. This corresponds to a problem of traffic junction treated in [7, 6] just to mention a few.

The criterion ensuring exponential stability which is relevant in this chapter is given in Chapter 3 corresponding to the article [47] :

$$\rho_\infty(H) = \inf_{\Delta \in D_d^+(\mathbb{R})} |\Delta H \Delta^{-1}|_\infty < 1. \tag{4.2}$$

Owing (4.2), we proved in Chapter 3 that the exponential stability of (4.1) in the space $BV([0,1])$, holds. In this work, we also consider the Banach space $BV([0,1])$ and focus on the exponential stability of the corresponding numerical solutions.

4.1.1 The numerical problem

In this chapter, the focus is on finite volume approximations of system (4.1) :

$$\frac{R_j^{n+1} - R_j^n}{dt} + \frac{f_{j+1/2}^n - f_{j-1/2}^n}{dx} = 0$$

where n, j are respectively the time and space index. More precisely, we find sufficient conditions on H such that the discretized version of (4.1) is exponentially stable.

The upwind flux $f_{j+1/2}^n = f(R_j^n)$ is known for its simplicity and its good properties. It is consistent and under a classic hypothesis of CFL, it is Total Variation Decreasing (TVD) and monotone [53] for scalar equations. These characteristics allow to prove easily that when the parameter of discretization tends towards zero, the numerical solution tends towards the unique entropy solution of the problem. For an introduction on the notion of entropy solution, we refer to [53] or Section 3.1.4. However, it is over diffusive and only first order accurate. In order to correct such behavior, one can use the additional precision of a second order scheme taking :

$$f_{j+1/2}^n = f(R_j^n + \tilde{R}_j^n)$$

where for example $\tilde{R}_j^n = \dfrac{R_{j+1}^n - R_j^n}{dx}\dfrac{dx}{2}$. However, a second order scheme cannot be TVD by Godunov Theorem and have oscillatory behaviors around discontinuities. Moreover, it can also be unstable as it is the case for the example cited just above. This is why one introduces a slope limiter ϕ [79] whose role is to damp the effect of the second order flux around discontinuities. The slope limiter scheme is defined taking :

$$\tilde{R}_j := \phi\left(\frac{R_j^n - R_{j-1}^n}{dx}, \frac{R_{j+1}^n - R_j^n}{dx}\right)\frac{dx}{2}$$

where ϕ is called the slope limiter that will be defined carefully later. The main contribution of the chapter is to find sufficient condition on ϕ and H to prove the exponential $BV([0,1])$ stability of the numerical solution built from the slope limiter scheme giving an explicit formula for the dissipation rate.

There are several results for the stabilization of discretized 1D wave equation. A very instructive survey from Trélat [106] gives an insight of techniques for the stabilization of discretized hyperbolic PDEs. One can also cite [104] where the authors prove the boundary stabilization of a semi-discretized 1D wave equation. Adding a numerical viscosity term damping high frequencies, they prove a uniform exponential stability decay of the energy uniformly with respect to the parameter of discretization. The article [103] deals with the case of an in-domain damping.

The main concern is to obtain a uniform observability inequality whose constant is independent on the parameter of discretization. This is not an easy task since the observability of the continuous system does not necessarily imply the same for the fully discrete or the semi discrete system. The survey [112] summarizes most of the main results about this question. The main problem comes from high frequency modes and their group velocity. The group velocity of two consecutive high frequency modes tends to zero as the discretization parameter goes to zero which prevent from using Ingham's inequality. This makes the observation (at a finite fixed time) very difficult as the discretization get finer [84]. This is why it is primordial to add a non physical numerical viscosity in order to damp these high frequencies.

For completeness, we cite [50] where the author proves the uniform exponential decay for another semi-discretized scheme simulating an in-domain damped wave equation. The case of semilinear wave equation is treated in [1]. The method to deduce exponential stabilization for the full discretized case is given in [51].

However to the author's knowledge, few result are known about the stabilization of discretized transport equations of the form (4.1). The method presented here is very different from observation inequalities techniques. We rather use discrete Lyapunov functions to prove exponential decay of our solution inspired from [57, 11]. This is presented in Section 2.

4.1.2 Problem setting

For wellposedness and scheme convergence results, we consider a more general boundary condition :

$$\begin{cases} \partial_t R + \partial_x[f(R)] & = & 0 \\ R(t,0) & = & g(R(t,1)) \\ R(0,x) & = & R^0(x) \in BV([0,1]). \end{cases} \tag{4.3}$$

where the following hypothesis on f and g hold :

— f is diagonal in the sense that :

$$\forall 1 \le i \le d,\ R \in \mathbb{R}^d,\ f_i(R) = f_i(R_i).$$

Moreover, all the f_i seen as functions from $\mathbb{R}$ to $\mathbb{R}$ are non decreasing. Doing so, the boundary condition of (4.3) has sense.

— There exist two positive constants $0 < v_{\min} < v_{\max}$ such that for all $1 \le i \le d$,

$$\forall (u,v) \in \mathbb{R}^2,\ v_{\min}|u-v| \le |f_i(u) - f_i(v)| \le v_{\max}|u-v|. \tag{4.4}$$

— The boundary feedback function $g : \mathbb{R}^d \mapsto \mathbb{R}^d$ is Lipschitz with Lipschitz constant $L_g > 0$. Moreover, we impose $g(0) = 0$.

Remark 23. The components of a solution R to (4.3) interact with each others only at the boundary.

In this chapter, the framework is $BV([0,1])$. This space is embedded with its classical norm $||\cdot||_{BV([0,1])}$ defined by

$$\forall R \in BV([0,1]), ||R||_{BV([0,1])} = TV_{[0,1]}(R) + ||R||_{L^1([0,1])} \tag{4.5}$$

where :

$$TV(R) = \sup_{n,\ (x_1,\ldots,x_n)} \left\{ \sum_{i=1}^{n-1} |R(x_{i+1}) - R(x_i)| \right\}.$$

The reason why the BV space is considered is because any function with bounded variations has a left limit and a right limit at each point x of $[0,1]$. Hence, it is easy to define the trace operator and impose a boundary condition. Moreover, $BV([0,1])$ has a very interesting property of compactness which will be very useful when we will pass to the limit when the parameter of discretization tends towards zero. These properties are given in Lemma 8 and Theorem 5 (Helly's theorem) from Chapter 3.

When the initial data is supposed to be L^∞ only, an entropic solution exists and is unique [75]. Moreover, it is possible to define the trace in a certain sense [89, 109]. However, the method we use to prove exponential stability crucially uses the BV hypothesis and this is why the L^∞ framework is out of the scope of this chapter.

4.1.3 Outline

The chapter is organized as follows. In Section 4.2, we precisely define the slope limiter scheme and prove the convergence in $BV([0,1])$ of the corresponding numerical solution. As a by-product, a well-posedness result for system (4.3) is proved. In Section 4.3, the framework is restricted to linear boundary feedback operator $g(R) = HR$. Then, the exponential stability of the scheme in $BV([0,1])$ is proved giving an explicit formula for the dissipation rate. Passing to the limit, we give another proof of the exponential stability of the continuous system which was already stated in [47]. In Section 4.4, simulations are given to illustrate results from Section 4.3. A study of saturated controls is also given. Finally, the conclusions and perspectives are exposed in the last part of this work.

4.2 Well-posedness and numerical approximation results

In this section, we introduce rigorously the slope limiter scheme corresponding to (4.3) and prove the convergence of such a scheme. Firstly, let us consider a space step $dx = 1/N$ ($N \in \mathbb{N}^*$) and a time step $dt > 0$ such that the following CFL condition holds (recall the definition of $v_{\max}$ in (4.4)) :

$$v_{max}\nu \leq 1 - \xi \tag{4.6}$$

with

$$\nu := \frac{dt}{dx}$$

and where ξ is a real number such that :

$$0 < \xi < 1.$$

Doing so, the space-time mesh is given by :

$$\forall n \in \mathbb{N},\ 1 \leq j \leq N, \left\{ \begin{array}{lll} x_j & := & (j-1/2)dx \\ t^n & := & ndt \end{array} \right. ,\ C_j := (x_j - dx/2, x_j + dx/2).$$

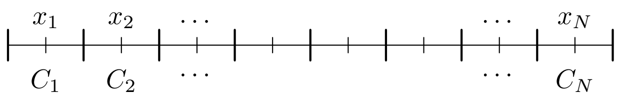


FIGURE 4.1 – The space grid

All along this chapter, the superscript n is exclusively allocated to designate time indices. Whereas, subscripts i and j are the indices corresponding respectively to the vector component and space grid. The numerical approximation $R^0_{\Delta x}$ of $R^0 \in BV([0,1])$ is given below

$$\forall 1 \leq j \leq N,\ x \in C_j,\ R^0_{\Delta x}(x) = \frac{1}{dx} \int_{C_j} R^0(x)dx. \tag{4.7}$$

In our context, the definition of a four point finite volume scheme is given below :

Definition 11. *A four-point finite volume scheme is defined a follows :*

— *The scheme is initialized defining* $(R^0_j)_{1 \leq j \leq N}$ *as in* (4.7).

— *Then the numerical solution is built by induction. For* $n \in \mathbb{N}$, $(R^n_j)_{1 \leq j \leq N}$ *is supposed to be given. We define* $(R^{n+1}_j)_{1 \leq j \leq N}$ *using the following relation*

$$\forall\ 2 \leq j \leq N-1,\ \frac{R^{n+1}_j - R^n_j}{dt} + \frac{f^n_{j+1/2} - f^n_{j-1/2}}{dx} = 0 \tag{4.8}$$

where $f^n_{j+1/2}$ *is the approximation of the flux at the right interface of cell* j *:*

$$\forall 1 \leq j \leq N-1,\ f^n_{j+1/2} = \tilde{f}(R^n_{j-1}, R^n_j, R^n_{j+1}). \tag{4.9}$$

The function $\tilde{f}$ *is called the three point approximation of the flux* f. *The choice of such a function will be given later with careful justifications.*

— *At the boundary the scheme degenerates into the upwind one :*

$$\left\{ \begin{array}{l} \dfrac{R^{n+1}_N - R^n_N}{dt} + \dfrac{f(R^n_N) - f(R^n_{N-1})}{dx} = 0 \\ \dfrac{R^{n+1}_1 - R^n_1}{dt} + \dfrac{f(R^n_1) - f(R^n_0)}{dx} = 0 \\ R^n_0 := g(R^n_N). \end{array} \right. \tag{4.10}$$

There are many possibilities for the choice of the function $\tilde{f}$. For example, one can take :

$$\tilde{f}(R^n_{j-1}, R^n_j, R^n_{j+1}) = f(R^n_j).$$

It corresponds to the classic upwind scheme when the transport velocity is positive. Recalling what was said in the introduction, the first main advantage of such scheme is that it is TVD and consistent allowing convergence towards a weak solution to (4.3). The other advantage is that it is monotone forcing the scheme to converge towards the unique entropy solution of (4.3). However, it is known for its diffusivity giving poor results in simulation. This is why one introduces the slope limiter flux [108] [79] :

$$\tilde{f}(R^n_{j-1}, R^n_j, R^n_{j+1}) = f(R^n_j + \tilde{R}^n_j) \tag{4.11}$$

where the term $\tilde{R}^n_j$ is given by a piecewise linear approximation of the solution :

$$\forall 1 \leq i \leq d,\ \tilde{R}^n_{i,j} = \phi\left(\frac{R^n_{i,j} - R^n_{i,j-1}}{dx}, \frac{R^n_{i,j+1} - R^n_{i,j}}{dx}\right)\frac{dx}{2}$$

where ϕ is assumed to be of the form :

$$\phi(u,v) = \begin{cases} 0 & \text{if } v = 0 \\ \phi_r(u/v)\psi_{dx}(v) & \text{otherwise} \end{cases} \tag{4.12}$$

such that there exists $0 \leq \beta < 1/2$ such that :

$$|\psi_{dx}(v)| \leq v \text{ and } |\psi_{dx}(v)| \leq dx^{-\beta} \tag{4.13}$$

and $\phi_r : \mathbb{R} \to \mathbb{R}$ and for all $r \leq 0,\ \phi_r(r) = 0$.

The natural choice for ψ_{dx} is :

$$\psi_{dx}(v) = \begin{cases} v & \text{if } |v| \leq dx^{-\beta} \\ sign(v)dx^{-\beta} & \text{otherwise.} \end{cases}$$

Hence, if $R^n_{i,j+1} - R^n_{i,j} \neq 0$, then :

$$\tilde{R}^n_{i,j} := \phi_r\left(r^n_{i,j+1/2}\right)\psi_{dx}\left(\frac{R^n_{i,j+1} - R^n_{i,j}}{dx}\right)\frac{dx}{2}.$$

The variable $r^n_{i,j+1/2} \in \mathbb{R}$ is defined by :

$$\forall 1 \leq i \leq d,\ r^n_{i,j+1/2} := \frac{R^n_{i,j} - R^n_{i,j-1}}{R^n_{i,j+1} - R^n_{i,j}}.$$

Remark 24. When the flux is linear, slope limiter schemes and flux limiter schemes coincide [79, p. 114].

The slope limiter ϕ is designed such that when the solution is regular i.e $r_{j+1/2} \approx 1$, the flux $\tilde{f}$ is closed to a second order flux. On the contrary, when the solution is not regular ($r_{j+1/2} \gg 1$) or admits a local extrema ($r_{j+1/2} \leq 0$), the second order scheme gives poor oscillatory results. In this case, we prefer to give more weight on the upwind TVD scheme imposing $\phi \approx 0$. We define the following coefficient for all $1 \leq i \leq d,\ 1 \leq j \leq N-1$, which will be useful all along this chapter :

$$
\begin{aligned}
a^n_{i,j-1} &:= \nu \frac{f_i(R^n_{i,j}+\tilde{R}^n_{i,j}) - f_i(R^n_{i,j-1}+\tilde{R}^n_{i,j-1})}{R^n_{i,j}+\tilde{R}^n_{i,j}-R^n_{i,j-1}-\tilde{R}^n_{i,j-1}} \\
&\times \Big(1 + \frac{\phi_r(r^n_{i,j+1/2})}{2r^n_{i,j+1/2}} \frac{\psi_{dx}(\frac{R^n_{i,j+1}-R^n_{i,j}}{dx})}{\frac{R^n_{i,j+1}-R^n_{i,j}}{dx}} - \frac{\phi_r(r^n_{i,j-1/2})}{2} \frac{\psi_{dx}(\frac{R^n_{i,j}-R^n_{i,j-1}}{dx})}{\frac{R^n_{i,j}-R^n_{i,j-1}}{dx}}\Big).
\end{aligned} \tag{4.14}
$$

For the pathologic cases where :

— $R^n_{i,j} - R^n_{i,j-1} = 0$ (*ie* $r^n_{i,j+1/2} = 0$ and $r^n_{i,j-1/2}$ ill-defined), then $a^n_{i,j-1} := 0$.

— $r^n_{i,j+1/2}$ is ill-defined, the coefficient $\frac{\phi_r(r^n_{i,j+1/2})}{2r^n_{i,j+1/2}} \frac{\psi_{dx}(\frac{R^n_{i,j+1}-R^n_{i,j}}{dx})}{\frac{R^n_{i,j+1}-R^n_{i,j}}{dx}}$ is replaced by zero in (4.14).

— $R^n_{i,j}+\tilde{R}^n_{i,j}-R^n_{i,j-1}-\tilde{R}^n_{i,j-1} = 0$, then the coefficient $\frac{f_i(R^n_{i,j}+\tilde{R}^n_{i,j})-f_i(R^n_{i,j-1}+\tilde{R}^n_{i,j-1})}{R^n_{i,j}+\tilde{R}^n_{i,j}-R^n_{i,j-1}-\tilde{R}^n_{i,j-1}}$ is replaced by $f'_i(R^n_{i,j}+\tilde{R}^n_{i,j})$ in (4.14).

For $j = N$, we set :

$$
a^n_{i,N-1} := \nu \frac{f^i(R^n_{i,N}) - f^i(R^n_{i,N-1})}{R^n_{i,N} - R^n_{i,N-1}}.
$$

if $R^n_{i,N} - R^n_{i,N-1} \neq 0$ and $a^n_{i,N-1} := 0$ otherwise.

The diagonal matrix with $a^n_{i,j}$ as entries is denoted $A^n_j := \text{diag}\{a^n_{i,j} \mid 1 \le i \le d\}$. Hence, the scheme can be reformulated under an upwind form :

$$
R^{n+1}_j = R^n_j - A^n_{j-1}(R^n_j - R^n_{j-1}).
$$

In order to ensure that our scheme is TVD, we impose the following condition of the slope limiter ϕ :

Hypothesis 4. *There exists* $0 \le v_{num} \le v_{\min}$ *such that :*

$$
0 \le \phi_r(r) \le \min\left\{2\left(\frac{1}{\nu v_{\max}} - 1\right) r, 2\left(1 - \frac{v_{num}}{v_{\min}}\right)\right\}
$$

This hypothesis is fundamental to ensure Harten's condition [63] :

Lemma 14. *If Hypothesis 4 is satisfied, then Harten's condition is satisfied :*

$$
\nu v_{num} \le a^i_{n,j} \le 1. \tag{4.15}
$$

Proof. Looking at (4.14), one has :

$$
\begin{aligned}
a^n_{i,j} &\le \nu v_{\max}\left(1 + \frac{\phi_r(r^n_{i,j+1/2})}{2r^n_{i,j+1/2}}\right), \\
a^n_{i,j} &\ge \nu v_{\min}\left(1 - \frac{\phi_r(r^n_{i,j-1/2})}{2}\right).
\end{aligned}
$$

By Hypothesis 4 :

$$
\begin{aligned}
\nu v_{\max}\left(1 + \frac{\phi_r(r^n_{i,j+1/2})}{2r^n_{i,j+1/2}}\right) &\le 1, \\
\nu v_{\min}\left(1 - \frac{\phi_r(r^n_{i,j-1/2})}{2}\right) &\ge \nu v_{num}
\end{aligned}
$$

which immediately gives the result of the Lemma.

□

Before going into the main result of this section, we introduce a new notation. Let $R_{\Delta x}$ be the piecewise constant function equal to R_j^n on each cell $[ndt, (n+1)dt] \times C_j$. This numerical approximation helps us proving the existence of a solution to (4.3). Meanwhile, Kruzhkov theory allows to prove the uniqueness. To prove the following Theorem 9, techniques are quite classical and plenty of works deal with the convergence of slope limiter schemes for scalar conservation laws in several dimensions and when the domain is the whole real line [26] [71] [88]. In our case, this is a bit more complex since we have to take into account the boundary condition.

Theorem 9. *System* (4.3) *admits a unique entropy solution* $R \in L^\infty_{loc}(\mathbb{R}^+, BV([0,1])) \cap Lip_{loc}(\mathbb{R}^+, L^1([0,1]))$ *in the sense that for all* $T > 0$ *:*

— *The entropy decay estimate holds :*

$$\forall k \in \mathbb{R}^d,\ \sum_{i=1}^{d} \int_0^T \int_0^1 \left\{ |R_i - k_i| \partial_t \varphi_i + (f_i(R_i) - f_i(k_i)) sign(R_i - k_i) \partial_x \varphi_i \right\} dxdt \geq 0 \quad (4.16)$$

for all $\varphi_i \geq 0$ *and* $\varphi_i \in C^1_c(]0, T[\times]0,1[)$.

— $R(0,.) = R^0$ *in the almost everywhere sense.*

— $R(.,0^+) = g(R(.,1^-))$ *in the almost everywhere sense.*

Moreover, the following convergence properties hold

$$\forall t \geq 0, dt > 0,\ TV_{[0,1]}(R(t,.)) \leq \limsup_{N \to +\infty} \sup_{s \in [t, t+dt]} TV_{[0,1]}(R_{\Delta x}(s,.)). \quad (4.17)$$

$$\forall t \geq 0,\ \lim_{N \to \infty} ||R(t,.) - R_{\Delta x}(t,.)||_{L^1([0,1])} = 0. \quad (4.18)$$

Proof. The proof is given in Appendix 4.A. □

4.3 Exponential BV stability

In this section, we focus on a particular case of system of (4.3).

$$\begin{cases} \partial_t R + \partial_x [f(R)] & = & 0 \\ R(t,0) & = & HR(t,1) \\ R(0,x) & = & R^0(x) \in BV([0,1]). \end{cases} \quad (4.19)$$

where $H \in M_d(\mathbb{R})$. We introduce the following stability hypothesis.

Hypothesis 5. *The feedback matrix* H *satisfies :*

$$\rho_1(H) < 1.$$

Remark 25. By [32, Remark 1.4]

$$\forall M \in M_d(\mathbb{R}),\ \rho_1(H) = \rho_\infty(H)$$

so that Hypothesis 5 can also be written as :

$$\rho_\infty(H) < 1.$$

The following Lyapunov functional is proposed for functions which are constant on the space mesh.

Definition 12. *Let N be in $\mathbb{N}^*$, $\gamma > 0$ and $P \in D_d^+(\mathbb{R})$. For any function R piecewise constant on cells C_j taking its values in $\mathbb{R}^d$, the BV Lyapunov functional $\mathcal{L}$ is given by*

$$\mathcal{L}(R) = \sum_{i=1}^{d} P_i \sum_{j=0}^{N-1} |R_{i,j+1} - R_{i,j}| e^{-\gamma x_j}$$

where

$$x_0 := -dx/2, R_0 := HR_N.$$

Next lemma ensures the equivalence between $\mathcal{L}$ and $||\cdot||_{BV([0,1])}$ defined in (4.5).

Lemma 15. *Suppose Hypothesis 5. For all $P \in D_d^+(\mathbb{R}^+)$ such that $|PHP^{-1}|_1 \leq 1$ and for all $\gamma > 0$, there exists a constant $C(P, \gamma, H) > 1$ such that for all $N \in \mathbb{N}^*$ and all functions R piecewise constant on cells $(C_j)_j$:*

$$\frac{\mathcal{L}(R)}{C(P,\gamma,H)} \leq ||R||_{BV([0,1])} \leq C(P,\gamma,H)\mathcal{L}(R). \tag{4.20}$$

Proof. We define the extension $\tilde{R}$ of R at $C_0 := (-dx, 0)$ setting $\tilde{R}(x) = HR(1)$ on C_0. In [47, Lemma 2], it is proved that :

$$\frac{\mathcal{L}(R)}{C(P,\gamma,H)} \leq TV(\tilde{R}) + \|\tilde{R}\|_{L^1([0,1])} \leq C(P,\gamma,H)\mathcal{L}(R). \tag{4.21}$$

Additionally, as $\tilde{R}$ is an extension on the left boundary of R and by the definition of the BV norm, it holds :

$$\|R\|_{BV([0,1])} \leq TV(\tilde{R}) + \|\tilde{R}\|_{L^1([0,1])} = \|R\|_{BV([0,1])} + |R_1 - \tilde{R}_0| + |\tilde{R}_0| dx.$$

As $\tilde{R}_0 = HR_N$ by definition and as $\|\cdot\|_{L^\infty([0,1])} \leq C\|\cdot\|_{BV([0,1])}$ ($C > 0$ is a constant depending on the parameters of the problem only), one has :

$$\|R\|_{BV([0,1])} \leq TV(\tilde{R}) + \|\tilde{R}\|_{L^1([0,1])} \leq C(H)\|R\|_{BV([0,1])}.$$

Hence, using (4.21), it holds :

$$\frac{\mathcal{L}(R)}{C(P,\gamma,H)} \leq ||R||_{BV([0,1])} \leq C(P,\gamma,H)\mathcal{L}(R).$$

where we may have changed the constant $C(P, \gamma, H)$. The lemma is proved. □

The reason why we introduce a discrete Lyapunov functional is because we cannot deduce the exponential stability of the numerical solution from the stability of the "continuous" one proven in [47]. Indeed, the convergence of the numerical solution towards the continuous solution is only proven to be in $L^1_{loc}(\mathbb{R}^+; L^1([0,1]))$ in general [53, Theorem 5.2]. This is not enough to prove an exponential stability result even in $L^1([0,1])$. This is why, we need to study the stability property of the scheme in itself introducing a discrete Lyapunov functional. The main result of this chapter is given in the following theorem whose proof is the object of the section.

Theorem 10. *Under Hypothesis 4-5 and for all $\gamma > 0, 0 \le c < v_{num}$, $P \in D_d^+(\mathbb{R}^d)$ such that $|PHP^{-1}|_1 < e^{-\gamma}$, there exists ε such that if $\gamma dt < \varepsilon$, the discrete Lyapunov estimate holds :*

$$\forall n \ge 0, \ \mathcal{L}(R^n) \quad \le \quad e^{-c\gamma n dt}\mathcal{L}(R^0).$$

To understand the consequences of Theorem 10, we consider a limiter ϕ_r satisfying Hypothesis 4 with $v_{num} = \alpha v_{\min}$ and $\alpha \in [0,1]$. Then, a Sweby diagram [101] is drawn in Figure 4.2 where the colored region represents the zone corresponding to Hypothesis 4. Theorem 10 states that for such ϕ_r the dissipation rate of the numerical solution is bounded from below by $\alpha v_{\min}(= v_{num})$. Remark that the best estimation $v_{num} = v_{\min} \iff \alpha = 1$ is obtained for only one scheme $\phi_r = 0$ which is the upwind scheme.

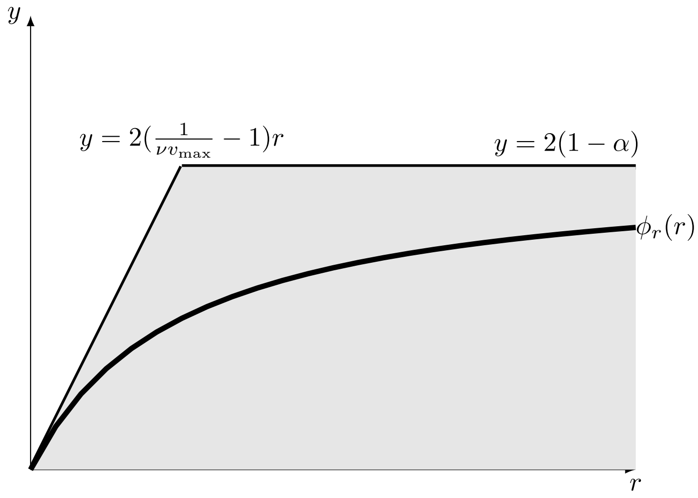


FIGURE 4.2 – The Sweby diagram

Proof. Let N, n be positive integers. By the definition of $\mathcal{L}$, we have :

$$\mathcal{L}(R^{n+1}) = \sum_{i=1}^{d} P_i \sum_{j=0}^{N-1} |R^{n+1}_{i,j+1} - R^{n+1}_{i,j}| e^{-\gamma x_j}.$$

The ith component of the Lyapunov functional $\mathcal{L}$ is given here :

$$\mathcal{L}_i(R^{n+1}) \quad := \quad \sum_{j=0}^{N-1} |R^{n+1}_{i,j+1} - R^{n+1}_{i,j}| e^{-\gamma x_j}.$$

For $1 \le i \le d$ and $1 \le j \le N-1$, we estimate the intensity of the discontinuity $R^{n+1}_{i,j+1} - R^{n+1}_{i,j}$.

$$R^{n+1}_{i,j+1} - R^{n+1}_{i,j} = (1 - a^n_{i,j})(R^n_{i,j+1} - R^n_{i,j}) + a^n_{i,j-1}(R^n_{i,j} - R^n_{i,j-1})$$

where a^n_j was defined in (4.14). Because of the Harten's condition (4.15) we have for $1 \le i \le d$ and $1 \le j \le N-1$:

$$|R^{n+1}_{i,j+1} - R^{n+1}_{i,j}| \le (1 - a^n_{i,j})|R^n_{i,j+1} - R^n_{i,j}| + a^n_{i,j-1}|R^n_{i,j} - R^n_{i,j-1}|. \tag{4.22}$$

Multiplying by $e^{-\gamma x_j}$ and summing over $1 \le j \le N-1$, we get

$$\begin{aligned}\textstyle\sum_{j=1}^{N-1} |R^{n+1}_{i,j+1} - R^{n+1}_{i,j}|e^{-\gamma x_j} \le \; & \sum_{j=1}^{N-2}(1 - a^n_{i,j}(1 - e^{-\gamma dx}))|R^n_{i,j+1} - R^n_{i,j}|e^{-\gamma x_j} \\ & +(1 - a^n_{i,N-1})|R^n_{i,N} - R^n_{i,N-1}|e^{-\gamma x_{N-1}} \\ & +a^n_{i,0}|R^n_{i,1} - R^n_{i,0}|e^{-\gamma dx/2}.\end{aligned}$$

Coming back to the Lyapunov functional $\mathcal{L}_i$, it holds :

$$\begin{aligned}\mathcal{L}_i(R^{n+1}) \le \; & \sum_{j=1}^{N-2}(1 - a^n_{i,j}(1 - e^{-\gamma dx}))|R^n_{i,j+1} - R^n_{i,j}|e^{-\gamma x_j} \\ & +(1 - a^n_{i,N-1})|R^n_{i,N} - R^n_{i,N-1}|e^{-\gamma x_{N-1}} \\ & +a^n_{i,0}|R^n_{i,1} - R^n_{i,0}|e^{-\gamma dx/2} \\ & +|R^{n+1}_{i,1} - R^{n+1}_{i,0}|e^{\gamma dx/2}.\end{aligned}$$

Now, we estimate $|R^{n+1}_{i,1} - R^{n+1}_{i,0}|$:

$$\begin{aligned}R^{n+1}_{i,1} - R^{n+1}_{i,0} &= R^n_{i,1} + a^n_{i,0}(R^n_{i,0} - R^n_{i,1}) - R^{n+1}_{i,0} \\ &= (R^n_{i,1} - R^n_{i,0})(1 - a^n_{i,0}) + R^n_{i,0} - R^{n+1}_{i,0}.\end{aligned} \tag{4.23}$$

Taking the absolute value and by triangle inequality, one obtains :

$$|R^{n+1}_{i,1} - R^{n+1}_{i,0}| \le (1 - a^n_{i,0})|R^n_{i,1} - R^n_{i,0}| + |R^n_{i,0} - R^{n+1}_{i,0}|.$$

This gives :

$$\begin{aligned}\mathcal{L}_i(R^{n+1}) \le \; & \sum_{j=1}^{N-2}(1 - a^n_{i,j}(1 - e^{-\gamma dx}))|R^n_{i,j+1} - R^n_{i,j}|e^{-\gamma x_j} \\ & +(1 - a^n_{i,N-1})|R^n_{i,N} - R^n_{i,N-1}|e^{-\gamma x_{N-1}} \\ & +(1 - a^n_{i,0}(1 - e^{-\gamma dx}))|R^n_{i,1} - R^n_{i,0}|e^{\gamma dx/2} \\ & +|R^{n+1}_{i,0} - R^n_{i,0}|e^{\gamma dx/2}\end{aligned}$$

which is no more than :

$$\begin{aligned}\mathcal{L}_i(R^{n+1}) \le \; & \sum_{j=0}^{N-2}(1 - a^n_{i,j}(1 - e^{-\gamma dx}))|R^n_{i,j+1} - R^n_{i,j}|e^{-\gamma x_j} \\ & +(1 - a^n_{i,N-1})|R^n_{i,N} - R^n_{i,N-1}|e^{-\gamma x_{N-1}} \\ & +|R^{n+1}_{i,0} - R^n_{i,0}|e^{\gamma dx/2}.\end{aligned}$$

Using the fact that there exists ε such that for $\gamma dx \leq \varepsilon$, $1-e^{-\gamma dx} \geq \frac{c}{v_{num}}\gamma dx$ (because $\frac{c}{v_{num}} < 1$ by assumption), we get :

$$\begin{aligned}\mathcal{L}_i(R^{n+1}) \leq \quad & \sum_{j=0}^{N-2}(1-a_{i,j}^n\frac{c\gamma dx}{v_{num}})|R_{i,j+1}^n - R_{i,j}^n|e^{-\gamma x_j} \\ & +(1-a_{i,N-1}^n)|R_{i,N}^n - R_{i,N-1}^n|e^{-\gamma x_{N-1}} \\ & +|R_{i,0}^{n+1} - R_{i,0}^n|e^{\gamma dx/2}.\end{aligned}$$

Then we add and subtract $(1-a_{i,N-1}^n\frac{c\gamma dx}{v_{num}})|R_{i,N}^n - R_{i,N-1}^n|e^{-\gamma(N-1)dx}$ in the first and second line respectively to get :

$$\begin{aligned}\mathcal{L}_i(R^{n+1}) \leq \quad & \sum_{j=0}^{N-1}(1-a_{i,j}^n\frac{c\gamma dx}{v_{num}})|R_{i,j+1}^n - R_{i,j}^n|e^{-\gamma x_j} \\ & -a_{i,N-1}^n(1-\frac{c\gamma dx}{v_{num}})|R_{i,N}^n - R_{i,N-1}^n|e^{-\gamma x_{N-1}} \\ & +|R_{i,0}^{n+1} - R_{i,0}^n|e^{\gamma dx/2}.\end{aligned}$$

Finally, using the fact that $a_{N-1}^n(R_N^n - R_{N-1}^n) = -(R_N^{n+1} - R_N^n)$ and the fact that $1-a_{i,j}^n\frac{c\gamma dx}{v_{num}} \leq 1-c\gamma dt$ (see (4.15)), the estimate of $\mathcal{L}_i$ writes :

$$\begin{aligned}\mathcal{L}_i(R^{n+1}) \leq \quad & (1-c\gamma dt)\mathcal{L}_i(R^n) \\ & -(1-\frac{c\gamma dx}{v_{num}})|R_{i,N}^{n+1} - R_{i,N}^n|e^{-\gamma x_{N-1}} \\ & +|R_{i,0}^{n+1} - R_{i,0}^n|e^{\gamma dx/2}.\end{aligned}$$

Using the discrete boundary condition (4.10), one changes the last boundary term :

$$\begin{aligned}\mathcal{L}_i(R^{n+1}) \leq \quad & (1-c\gamma dt)\mathcal{L}_i(R^n) \\ & -(1-\frac{c\gamma dx}{v_{num}})|R_{i,N}^{n+1} - R_{i,N}^n|e^{-\gamma x_{N-1}} \\ & +|[HR_N^{n+1}]_i - [HR_N^n]_i|e^{\gamma dx/2}.\end{aligned}$$

Multiplying by P_i and summing over all the $1 \leq i \leq d$, an estimate on $\mathcal{L}$ writes :

$$\begin{aligned}\mathcal{L}(R^{n+1}) \leq \quad & (1-c\gamma dt)\mathcal{L}(R^n) \\ & -\sum_{i=1}^{d} P_i|R_{i,N}^{n+1} - R_{i,N}^n|(1-\frac{c\gamma dx}{v_{num}})e^{-\gamma x_{N-1}} \\ & +\sum_{i=1}^{d} P_i|[HR_N^{n+1}]_i - [HR_N^n]_i|e^{\gamma dx/2}.\end{aligned}$$

Using the definition of $|\cdot|_1$, the boundary terms can be compared with each others :

$$\begin{aligned}\mathcal{L}(R^{n+1}) \leq \quad & (1-c\gamma dt)\mathcal{L}(R^n) \\ & +(|PHP^{-1}|_1 e^{\gamma dx/2} - (1-\frac{c\gamma dx}{v_{num}})e^{-\gamma x_{N-1}})\sum_{i=1}^{d} P_i|R_{i,N}^{n+1} - R_{i,N}^n|.\end{aligned}$$

As $P \in D_d^+(\mathbb{R})$ is such that $|PHP^{-1}|_1 < e^{-\gamma}$, we get for γdx sufficiently small (that is to say $\gamma dx < \varepsilon(c, v_{num}, H)$) :

$$\mathcal{L}(R^{n+1}) \quad \leq \quad \Big(1 - c\gamma dt\Big)\mathcal{L}(R^n).$$

Summing with respect to time, one obtains :

$$\mathcal{L}(R^{n+1}) \quad \leq \quad \Big(1 - c\gamma dt\Big)^{n+1}\mathcal{L}(R^0).$$

Using the fact that $1 - x \leq e^{-x}$, the result of Lemma 10 is proved.

□

Remark 26. The rate of exponential convergence γc given by Theorem 10 is weaker than the one given in [47] for the continuous setting $\gamma v_{\min}$. This is mainly due to the choice of the limiter which guaranties only (4.15). To recover the rate of convergence from [47], one should impose :

$$\nu v_{min} \leq a^i_{n,j} \leq 1$$

which is true for the upwind scheme but not for other limiters.

The following Theorem was already proved in [47]. The main difference here is that the proof comes from the convergence of our finite volume scheme.

Theorem 11. *Under Hypothesis 5 and for all $P \in D_d^+(\mathbb{R}), \gamma > 0$ such that $|PHP^{-1}|_1 < e^{-\gamma}$, there exists a constant $C(P,\gamma,H)$ depending only on P,γ,H such that the entropy solution to* (4.19) *verifies*

$$\forall t \geq 0,\ ||R(t,\cdot)||_{BV([0,1])} \leq C(P,\gamma,H)\exp\Big(-\gamma v_{\min} t\Big)||R^0||_{BV([0,1])}.$$

Proof. For this proof, we take $\phi \equiv 0$ which corresponds to the upwind scheme. As a consequence, it is possible to take $v_{num} = v_{\min}$ in Hypothesis 4. Then, it suffices to pass to the limit as N goes to infinity in Lemma 10 using estimates (4.18), (4.17) and the equivalence between norms $\mathcal{L}, ||\cdot||_{BV([0,1])}$ (see (4.20)).

□

4.4 Simulations

In this section, we illustrate numerically that slope limiter schemes are clearly less dissipative than the upwind scheme and capture better the behavior of the continuous solution. For computations, we consider the case where $d = 4$, $\frac{dt}{dx} = 0.4$ and the flux is given by :

$$\forall 1 \leq i \leq 4,\ x \in \mathbb{R},\ f_i(x) = 0.5x + \frac{0.5}{2^{i-1}}\arctan(x).$$

In such a case, min and max velocities are $v_{\min} = 0.5$ and $v_{\max} = 1$. Concerning the limiter, it is chosen as :

$$\phi_r(r) = \max\{\min\{r,1\},0\},\ \psi_{dx}(v) = \begin{cases} v & \text{if } |v| \leq dx^{-\beta} \\ sign(v)dx^{-\beta} & \text{otherwise} \end{cases}$$

with $\beta = 0.49$. This limiter is called a minmod limiter.

4.4.1 Dissipativity of the scheme

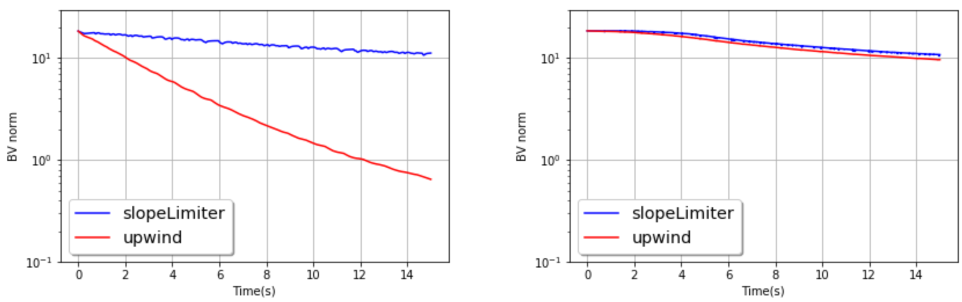
If the focus is on the dissipativity of the scheme alone, a relevant boundary condition is :

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

In such a case, the boundary condition does not dissipate energy. For a smooth initial data :

$$R_1(t = 0, .) = \cdots = R_4(t = 0, .) = \cos(2\pi x),$$

we compare the $BV([0, 1])$ norm of the solution given by the upwind scheme with the one given by the limiter one. It is expected that the limiter scheme is less dissipative than the upwind one when the solution is regular.



(a) $dt = 10^{-2}$ (b) $dt = 5 \times 10^{-4}$

FIGURE 4.3 – The BV norm of the solution for the non dissipative case

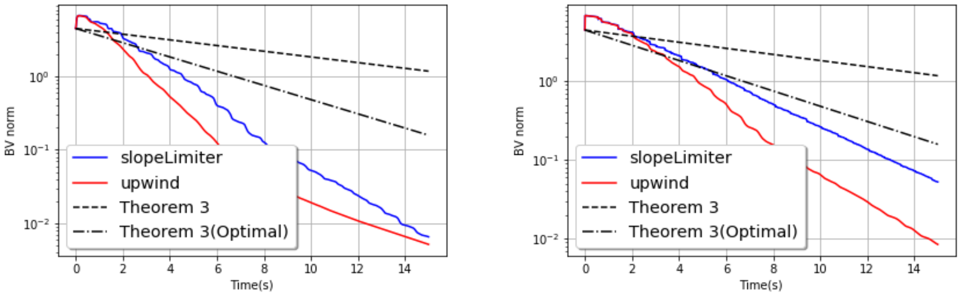
One sees in Figure 4.3a-4.3b that this is indeed the case. Obviously when the grid is fine enough, the dissipation created by the upwind scheme is negligible and results for both schemes are similar. Next, we can do the same analysis imposing this time, a dissipative feedback boundary condition :

$$H = \begin{pmatrix} 0.5 & 0.1 & 0 & 0.1 \\ 0.1 & 0.5 & 0.1 & 0 \\ 0 & 0.1 & 0.5 & 0.1 \\ 0.1 & 0 & 0.1 & 0.5 \end{pmatrix}.$$

An optimization routine gives $\rho_1(H) = 0.7$. Moreover, a less regular initial data is examined :

$$\begin{cases} R_1^0 = & x \geq 0.5 \\ R_2^0 = & -(x \leq 0.5) + (x \geq 0.5) \\ R_3^0 = & x \geq 0.5 \\ R_4^0 = & -(x \leq 0.5) + (x \geq 0.5). \end{cases}$$

One obtains Figures 4.4a-4.4b for the evolution of the BV norm :



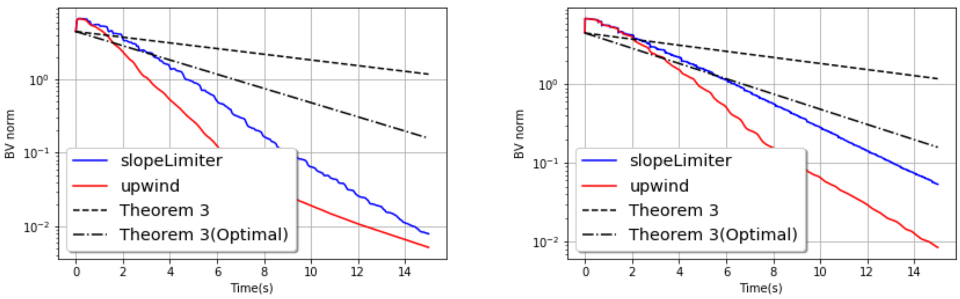
(a) $dt = 10^{-2}$ (b) $dt = 5 \times 10^{-4}$

FIGURE 4.4 – The BV norm of the solution for the dissipative case

The black dashed line corresponds to the estimation of the rate of convergence γv_{num} ($v_{num} = 0.5 v_{\min}, \gamma = -\log(0.7)$ here) given by Theorem 10. We see this is not optimal since the rate of convergence is underestimated. Then, when the solution is small enough, it should be wise to take $v_{\min} = \min_i f_i'(0)$ in Theorem 10 in order to estimate the rate of convergence when the solution is small (in $L^\infty([0,1])$). This rate is represented by the dashed dotted line for the upwind scheme ($v_{num} = v_{\min}$) in figures 4.4a-4.5b. This estimation fits better with the numerical experiments. With these results, one sees that Theorem 10 gives only a lower bound on the exponential convergence rate and not a precise estimation of it. Another remark to make is that the limiter scheme captures better the dissipation rate, the upwind scheme dissipating too much when the grid is coarse.

If we change the limiter, we have similar numerical results. For example, one can take the Van Leer limiter :

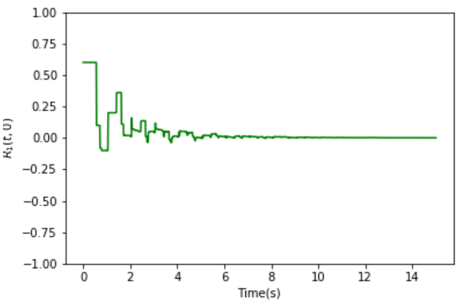
$$\phi_r(r) = \frac{r + |r|}{1 + |r|}.$$



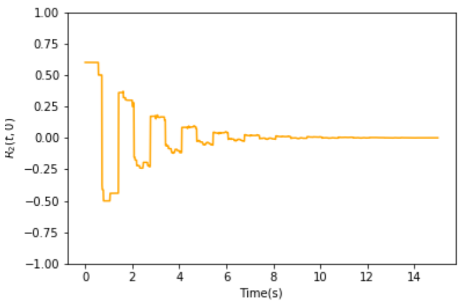
(a) $dt = 10^{-2}$ (b) $dt = 5 \times 10^{-4}$

FIGURE 4.5 – The Van-Leer case

For completeness, we plot a part of the left trace of the solution for $t < 15$ ($dt = 5 \times 10^{-4}$).



(a) $R_1(t,0)$



(b) $R_2(t,0)$

FIGURE 4.6 – Trace of the solution

One clearly sees that the solution is rapidly damped when time evolves.

4.4.2 Saturated control

In this section, we analyze an example of system of scalar conservation laws for $d = 2$ with saturated feedback control law. The same example as [47] is considered :

$$\begin{cases} \partial_t R + \partial_x[f(R)] & = & 0 \\ R(t,0) & = & HR(t,1) + Bu(t) \\ R(0,x) & = & R^0(x) \in BV([0,1]) \end{cases} \tag{4.24}$$

where :

$$u(t) = \sigma(KR(t,1)).$$

The operator σ is the saturation by component with level of saturation equal to one :

$$\forall 1 \leq i \leq 2,\ x \in \mathbb{R}, \begin{cases} \sigma_i(x) = x & \text{if } |x| \leq 1 \\ \sigma_i(x) = \text{sign}(x) & \text{otherwise.} \end{cases}$$

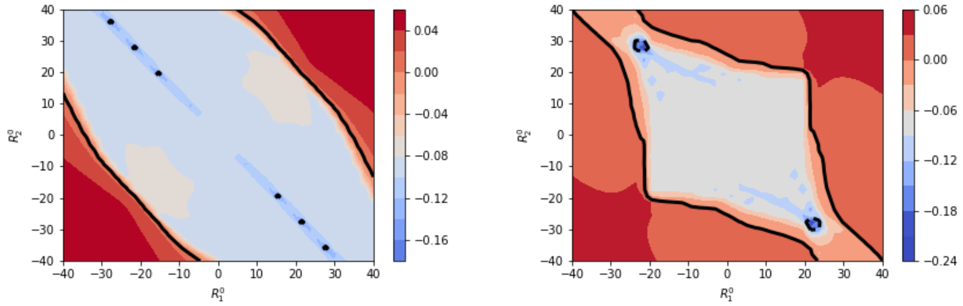
Matrices are defined as follows.

$$H = \begin{pmatrix} 0 & 1.1 \\ 1 & 0 \end{pmatrix},\ B = I_2,\ K = \begin{pmatrix} 0 & -0.1050 \\ -0.1045 & 0 \end{pmatrix}.$$

As nonlinear flux, one takes $f(R) = \Lambda R + 0.2(\arctan(R_1), \arctan(R_2))$ with

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{pmatrix}.$$

For the scheme parameters, it is imposed that $dt = 10^{-2}$ and $\nu = 0.4$. Then, we take R^0 constant with value in $(-40, 40)^2$ and look if the solution does not blow up at infinite time in BV norm. In this way, it is possible to estimate the projection of the basin of attraction onto the space of constant initial data. Results are given below for the upwind and the minmod slope limiter scheme :



(a) Upwind

(b) Minmod limiter

FIGURE 4.7 – Comparison of basins of attraction ($dt = 10^{-2}$)

The x-axis corresponds to the value of the first component of the initial data $R_1^0(x) = R_1^0 \in \mathbb{R}$ whereas the y-axis corresponds to the value of the second component of the initial data $R_2^0(x) = R_2^0 \in \mathbb{R}$. Contours correspond to the rate of exponential decay of the numerical solution for a time window of 50 seconds. If it is negative, the solution decays exponentially fast in BV norm. If it is positive, we have exponential divergence. The black thick contour corresponds to a dissipation rate equal to zero. Here, the same figure is plotted with a finer discretization $dt = 10^{-3}$ for the minmod limiter scheme.

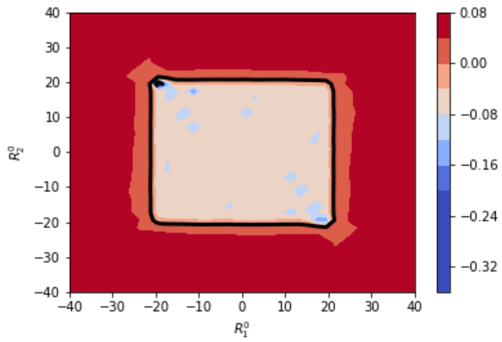


FIGURE 4.8 – Basin of attraction for $dt = 10^{-3}$

We see that when the grid is coarse, the upwind scheme overestimates the basin of attraction whereas it is not the case for the minmod slope limiter scheme.

4.5 Conclusion

The exponential BV stability of general slope limiter scheme for system (4.19) has been established with an estimate of the dissipation rate. The main idea was to use Harten's formalism. Then, using new discrete Lyapunov techniques, we concluded on the exponential stability of the numerical solution in the spirit of the continuous case. The dissipation rate is estimated from below uniformly with respect to the discretization. Finally, the convergence of the scheme towards the entropy solution was established using the boundedness of ψ_{dx}.

There are some questions which remain open. The case of time dependent flux may not change the philosophy of the proof of exponential stability. However if we add a source term coupling the equations inside the domain the analysis is far from being obvious. Then the case where there are positive and negative characteristic velocities is not treated but this is only a question of change of variable. Finally combining both difficulties of velocities with different signs and the one of source terms, we know that a Lyapunov function does not exist in the continuous case [14]. It does not seem extravagant to suppose that such Lyapunov function does not exist in the discrete case. Backstepping techniques were designed to solve this difficulty when the flux is linear but it is not obvious that such methods can be directly applied to the corresponding numerical system.

4.A Proof of Theorem 9

The uniqueness can be proven by classical Kruzhkov techniques [47, Appendix B]. In this section, we give more attention on the existence result and the convergence of the scheme. Let N be a positive integer and denote $R_{\Delta x}$ the numerical approximation given in Definition 11. We recall that we defined R_0^n as :

$$R_0^n = g(R_N^n).$$

4.A.1 BV estimates

The three following lemmas allow to get BV estimates on $R_{\Delta x}$.

Lemma 16. *[Space to space BV estimate]*
Let $0 \leq j_0 < j_1 \leq N-1$ be integers, then for all integers n, m such that $m \leq j_0$:

$$\sum_{j=j_0}^{j_1} |R_{j+1}^n - R_j^n| \leq \sum_{j=j_0-m}^{j_1} |R_{j+1}^{n-m} - R_j^{n-m}|. \tag{4.25}$$

Moreover, the following L^∞ estimate holds :

$$\max_{j_0 \leq j \leq j_1} |R_j^n| \leq \max_{j_0-m \leq j \leq j_1} |R_j^{n-m}|. \tag{4.26}$$

We give a graphical interpretation of previous lemma. It corresponds to a bound on the space BV norm of the red line in Figure 4.9 by the space BV norm of the blue line.

Proof. We will show this by induction on m. For $m = 0$, the property is trivially verified. For $m = 1$, we have :

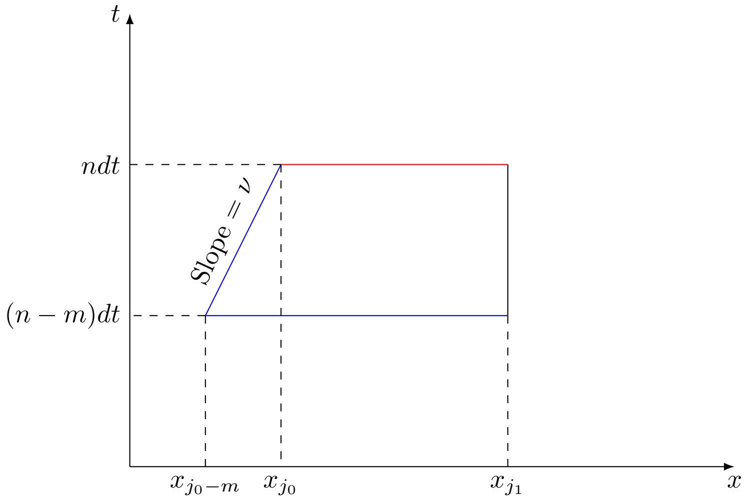


FIGURE 4.9 – Space to space BV estimate.

$$\begin{aligned}
\sum_{j=j_0}^{j_1} |R_{j+1}^n - R_j^n| \leq & \sum_{j=j_0}^{j_1} |I_d - A_j^{n-1}||R_{j+1}^{n-1} - R_j^{n-1}| \\
& + \sum_{j=j_0}^{j_1} |A_{j-1}^{n-1}||R_j^{n-1} - R_{j-1}^{n-1}| \\
= & \sum_{j=j_0}^{j_1} |I_d - A_j^{n-1}||R_{j+1}^{n-1} - R_j^{n-1}| \\
& + \sum_{j=j_0-1}^{j_1-1} |A_j^{n-1}||R_{j+1}^{n-1} - R_j^{n-1}| \\
\leq & \sum_{j=j_0-1}^{j_1} |R_{j+1}^{n-1} - R_j^{n-1}|
\end{aligned}$$

where we have used (4.22) to get the first inequality and the fact that the A_j^ns verify $0 \leq A_j^n < I_d$ for last inequality. The case $m > 1$ is easily proven by induction. For the L^∞ estimate (4.26) for $m = 1$:

$$\begin{aligned}
\max_{j_0 \leq j \leq j_1} |R_j^n| &= \max_{j_0 \leq j \leq j_1} |R_j^{n-1} + A_{j-1}^n (R_{j-1}^{n-1} - R_j^{n-1})| \\
&\leq \max_{j_0 \leq j \leq j_1} |(I_d - A_{j-1}^n) R_j^{n-1} + A_{j-1}^n R_{j-1}^{n-1}| \\
&\leq \max_{j_0-1 \leq j \leq j_1} \{|R_j^{n-1}|\}.
\end{aligned}$$

Again, the case for $m > 1$ is easily proven by induction. This ends the proof of Lemma 16. □

Lemma 17. *[Time to space BV estimate]*
Let $1 \leq \bar{j} \leq N$, n, m be integers such that $\bar{j} - 1 - m > 0$

$$\sum_{l=n-m}^{n} |R^{l+1}_{\bar{j}} - R^{u}_{\bar{j}}| \leq \sum_{j=\bar{j}-1-m}^{\bar{j}-1} |R^{n-m}_{j+1} - R^{n-m}_{j}|. \tag{4.27}$$

Moreover, the following L^∞ estimate holds :

$$\max_{n-m\leq l\leq n+1} |R^{l}_{\bar{j}}| \leq \max_{\bar{j}-1-m\leq j\leq \bar{j}} |R^{n-m}_{j}|. \tag{4.28}$$

Similarly to Lemma 16, we give a graphical interpretation of previous lemma. The time BV norm of the red line in Figure 4.10 is bounded by the space BV norm of the blue line.

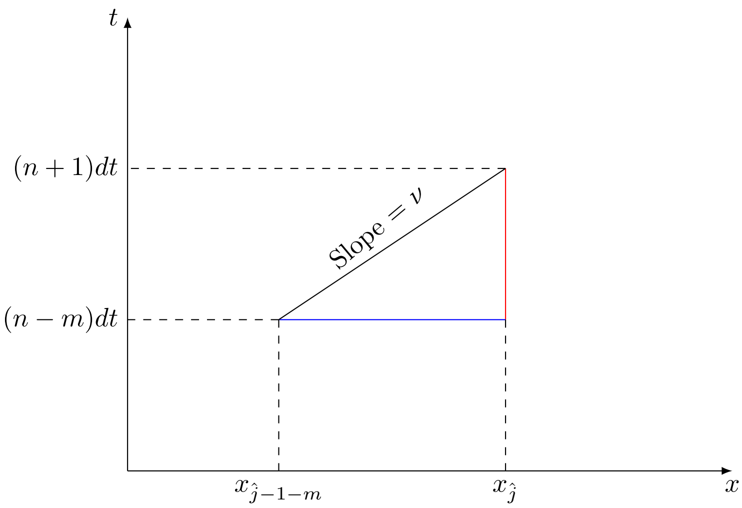


FIGURE 4.10 – Time to space BV estimate.

Proof. We will proceed by induction on m. For $m = 0$, the following TV estimate holds :

$$\begin{aligned} |R^{n+1}_{\bar{j}} - R^{n}_{\bar{j}}| &\leq |A^{n}_{\bar{j}-1}(R^{n}_{\bar{j}-1} - R^{n}_{\bar{j}})| \\ &\leq |R^{n}_{\bar{j}-1} - R^{n}_{\bar{j}}|. \end{aligned}$$

which is the required result. For the L^∞ estimate (4.28) :

$$\begin{aligned} |R^{n+1}_{\bar{j}}| &\leq |R^{n}_{\bar{j}} + A^{n}_{\bar{j}-1}(R^{n}_{\bar{j}-1} - R^{n}_{\bar{j}})| \\ &\leq |(I_d - A^{n}_{\bar{j}-1})R^{n}_{\bar{j}} + A^{n}_{\bar{j}-1}R^{n}_{\bar{j}-1}| \\ &\leq \max\{|R^{n}_{\bar{j}-1}|, |R^{n}_{\bar{j}}|\} \end{aligned}$$

which is the required result.

Suppose that (4.27) is true for $m - 1 \geq 1$. Then we have,

$$\begin{aligned} \sum_{l=n-m}^{n} |R^{l+1}_{\bar{j}} - R^{l}_{\bar{j}}| \leq\ & \textstyle\sum_{j=\bar{j}-1-m+1}^{\bar{j}-1} |R^{n-m+1}_{j+1} - R^{n-m+1}_{j}| \\ & + |R^{n-m+1}_{\bar{j}} - R^{n-m}_{\bar{j}}|. \end{aligned}$$

Using the definition of the scheme, one gets

$$\begin{aligned}\sum_{l=n-m}^{n} |R_{\bar{j}}^{l+1} - R_{\bar{j}}^{l}| \leq & \sum_{j=\bar{j}-1-m+1}^{\bar{j}-1} |I_d - A_j^{n-m}||R_{j+1}^{n-m} - R_j^{n-m}| \\ & + \sum_{j=\bar{j}-1-m+1}^{\bar{j}-1} |A_{j-1}^{n-m}||R_j^{n-m} - R_{n-m,j-1}| \\ & + |A_{\bar{j}-1}^{n-m}||R_{\bar{j}}^{n-m} - R_{\bar{j}-1}^{n-m}|.\end{aligned}$$

Rearranging all the terms and taking into account that fact that the $a_{i,j}^n$s verify $0 \leq a_{i,j}^n \leq 1$, we get

$$\sum_{l=n-m}^{n} |R_{\bar{j}}^{l+1} - R_{\bar{j}}^{l}| \quad \leq \quad \textstyle\sum_{j=\bar{j}-1-m}^{\bar{j}-1} |R_{j+1}^{n-m} - R_j^{n-m}|$$

which is what was required.

For the L^∞ estimate, the hypothesis of induction gives :

$$\max_{n-(m-1)\leq l \leq n+1} |R_{\bar{j}}^{l}| \quad \leq \quad \max_{\bar{j}-1-(m-1)\leq j \leq \bar{j}} |R_j^{n-m+1}|.$$

The space to space L^∞ estimate (4.26) gives :

$$\max_{n-(m-1)\leq l \leq n+1} |R_{\bar{j}}^{l}| \quad \leq \quad \max_{\bar{j}-m-1\leq j \leq \bar{j}} |R_j^{n-m}|.$$

This finishes the proof of the Lemma.

□

Lemma 18. *[Space to time/space BV estimate]*

Let $0 \leq j_0 < j_1 \leq N-1$ be integers, then for all integers m, n

$$\sum_{j=j_0}^{j_1} |R_{j+1}^n - R_j^n| \leq \sum_{j=j_0}^{j_1} |R_{j+1}^{n-m} - R_j^{n-m}| + \sum_{l=n-m}^{n-1} |R_{j_0}^{l+1} - R_{j_0}^{l}|. \tag{4.29}$$

Moreover, the following L^∞ estimate holds :

$$\max_{j_0\leq j\leq j_1+1} |R_{\bar{j}}^n| \leq \max_{n-m\leq l\leq n} |R_{j_0}^l| + \max_{j_0\leq j\leq j_1+1} |R_j^{n-m}|. \tag{4.30}$$

Similarly to Lemma 16, we give a graphical interpretation of previous Lemma. The space total variation of the red line in Figure 4.11 is bounded by the time/space total variation of the blue line.

Proof. We proceed by induction this time on m. If $m = 1$ and $j_0 \geq 1$, then

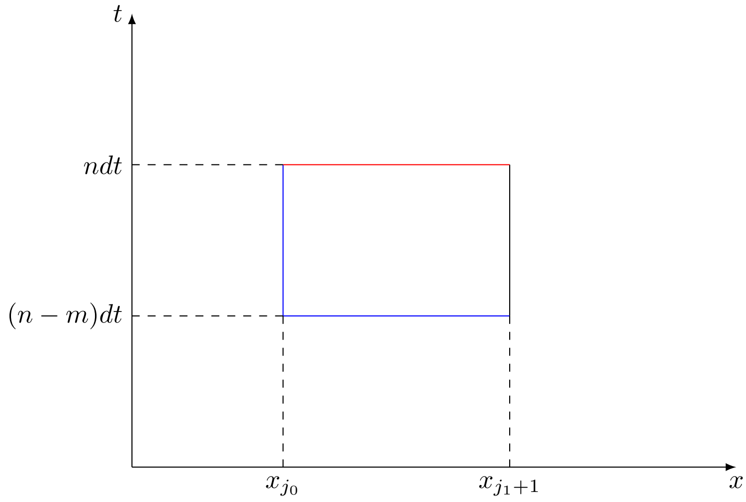


FIGURE 4.11 – Space to time/space BV estimate.

$$\begin{aligned}
\sum_{j=j_0}^{j_1} |R_{j+1}^n - R_j^n| \leq & \sum_{j=j_0}^{j_1} |I_d - A_j^{n-1}||R_{j+1}^{n-1} - R_j^{n-1}| + |A_{j-1}^{n-1}||R_j^{n-1} - R_{j-1}^{n-1}| \\
= & |A_{j_0-1}^{n-1}(R_{j_0}^{n-1} - R_{j_0-1}^{n-1})| \\
& + \sum_{j=j_0}^{j_1-1} |R_{j+1}^{n-1} - R_j^{n-1}| + |I_d - A_{j_1}^{n-1}||R_{j_1+1}^{n-1} - R_{j_1}^{n-1}| \\
= & |R_{j_0}^n - R_{j_0}^{n-1}| \\
& + \sum_{j=j_0}^{j_1-1} |R_{j+1}^{n-1} - R_j^{n-1}| + |I_d - A_{j_1}^{n-1}||R_{j_1+1}^{n-1} - R_{j_1}^{n-1}| \\
\leq & |R_{j_0}^n - R_{j_0}^{n-1}| \\
& + \sum_{j=j_0}^{j_1} |R_{j+1}^{n-1} - R_j^{n-1}|
\end{aligned}$$

which gives the required result. By induction, it is easy to get the result for any m such that $n-m \geq 0$. Now we consider the case where $j_0 = 0$ and $m = 1$. Using the definition of the scheme, we have :

$$\begin{aligned}
\sum_{j=0}^{j_1} |R_{j+1}^n - R_j^n| \leq & \sum_{j=1}^{j_1} |I_d - A_j^{n-1}||R_{j+1}^{n-1} - R_j^{n-1}| + |A_{j-1}^{n-1}||R_j^{n-1} - R_{j-1}^{n-1}| \\
& + |R_1^n - R_0^n|.
\end{aligned}$$

Moreover from (4.23), one deduces that

$$|R_1^n - R_0^n| \quad \leq \quad |R_0^n - R_0^{n-1}| + |I_d - A_0^{n-1}||R_1^{n-1} - R_0^{n-1}|.$$

This gives

$$\begin{aligned}\sum_{j=0}^{j_1}|R_{j+1}^n - R_j^n| \leq & \sum_{j=1}^{j_1}|I_d - A_j^{n-1}||R_{j+1}^{n-1} - R_j^{n-1}| + |A_{j-1}^{n-1}||R_j^{n-1} - R_{j-1}^{n-1}| \\ & + |R_0^n - R_0^{n-1}| + |I_d - A_0^{n-1}||R_1^{n-1} - R_0^{n-1}| \\ \leq & \textstyle\sum_{j=0}^{j_1}|R_{j+1}^{n-1} - R_j^{n-1}| + |R_0^n - R_0^{n-1}|.\end{aligned}$$

which gives the required result. Again, we use the induction to deduce the result for any m such that $n - m \geq 0$. The L^∞ estimate (4.30) can be proven with similar techniques and is left to the reader. □

4.A.2 The L^∞ estimate

Now, define for all $ndt < \frac{1}{2(1-\xi)v_{\max}} := \tau$, the index $j_L(n)$ such that $2(1-\xi)v_{\max}(n-1)dt \leq j_L(n)dx < 2(1-\xi)v_{\max}ndt$. From now on, we suppose that the time index n verifies $ndt < \tau$. Estimate (4.26) gives

$$\max_{\substack{ndt<\tau \\ j_L(n)+1\leq j\leq N}} |R_j^n| \leq \max_{0\leq j\leq N} |R_j^0| \tag{4.31}$$

whereas estimate (4.30) gives

$$\max_{\substack{ndt<\tau \\ 0\leq j\leq j_L(n)}} |R_j^n| \leq \max_{ndt<\tau} |R_0^n| + \max_{0\leq j\leq N} |R_j^0|.$$

Then, the Lipschitz character of the boundary condition allows to transform last inequality

$$\max_{\substack{ndt<\tau \\ 0\leq j\leq j_L(n)}} |R_j^n| \leq C(L_g) \max_{ndt<\tau} |R_N^n| + \max_{0\leq j\leq N} |R_j^0|$$

where $C(L_g)$ is a constant depending only on L_g. Finally, injecting estimate (4.28) in last equation, one gets :

$$\max_{\substack{ndt<\tau \\ 0\leq j\leq j_L(n)}} |R_j^n| \leq (1 + C(L_g)) \max_{0\leq j\leq N} |R_j^0|. \tag{4.32}$$

Gathering (4.32) and (4.31), we get :

$$\|R_{\Delta x}\|_{L^\infty([0,\tau]\times[0,1])} \leq (1 + C(L_g))\|R_{\Delta x}^0\|_{L^\infty([0,1])}.$$

Repeating this process, we get the following L^∞ estimate for any $T > 0$:

$$\|R_{\Delta x}\|_{L^\infty([0,T]\times[0,1])} \leq C\|R_{\Delta x}^0\|_{L^\infty([0,1])}.$$

where C depends only on T, L_g, ξ and $v_{\max}$. By the fact that $(R_{\Delta x}^0)_N$ is bounded in L^∞ (recall that R^0 is in $BV([0,1])$), we have

$$\|R_{\Delta x}\|_{L^\infty([0,T]\times[0,1])} \leq C\|R^0\|_{L^\infty([0,1])} \tag{4.33}$$

where C depends only on T and L_g.

4.A.3 The TV estimate

We have,

$$TV(R^n_\bullet) = TV_{[0,j_L]}(R^n_\bullet) + TV_{[j_L,N]}(R^n_\bullet). \tag{4.34}$$

By Lemma 18, the first term in (4.34) can be bounded :

$$TV_{[0,j_L]}(R^n_\bullet) \leq TV_{[0,n]}(R^\bullet_0) + TV(R^0_\bullet).$$

Moreover, by the Lipschitz character of the boundary condition, we get

$$TV_{[0,j_L]}(R^n_\bullet) \leq L_g TV_{[0,n]}(R^\bullet_N) + TV(R^0_\bullet).$$

Finally, using Lemma 17, we get

$$TV_{[0,j_L]}(R^n_\bullet) \leq (1+L_g)TV(R^0_\bullet). \tag{4.35}$$

With Lemma 16, we get :

$$TV_{[j_L,N]}(R^n_\bullet) \leq TV(R^0_\bullet). \tag{4.36}$$

and with (4.35), one gets :

$$\forall ndt < \tau,\ TV(R^n_\bullet) \leq C(L_g,T)TV(R^0_\bullet).$$

Repeating this process, one gets the following bound in TV for $ndt < T$ with $T > 0$:

$$TV(R^n_\bullet) \leq C(L_g,T)TV(R^0_\bullet)$$

where $C(L_g,T)$ is a constant that depends solely on L_g and T. By the fact that $(R^0_\cdot)_N$ converges towards R^0 in $BV([0,1])$, we have

$$\forall ndt \leq T,\ TV(R^n_\bullet) \leq C(L_g,T,R^0). \tag{4.37}$$

4.A.4 The $L^1([0,1])$ continuity estimate

For all integers $n > 1$,

$$\begin{aligned}
\sum_{j=0}^{N} |R^{n+1}_j - R^n_j|dx =& \sum_{j=1}^{N} |A^n_{j-1}(R^n_j - R^n_{j-1})|dx + |R^{n+1}_0 - R^n_0|dx \\
\leq& \sum_{j=1}^{N} |A^n_{j-1}||R^n_j - R^n_{j-1}|dx + L_g|R^{n+1}_N - R^n_N|dx \\
=& \sum_{j=0}^{N-1} |A^n_j||R^n_{j+1} - R^n_j|dx + L_g|A^n_{N-1}(R^n_N - R^n_{N-1})|dx \\
\leq& (1+L_g)TV(R^n_\bullet)dx.
\end{aligned}$$

By the triangle inequality, one gets for $n, m > 0$

$$\sum_{j=0}^{N} |R^{n+m}_j - R^n_j|dx \quad \leq \quad dx \sum_{l=n}^{n+m-1} (1+L_g)TV(R^l_\bullet).$$

And by the boundedness of the total variation (4.37), we get

$$\sum_{j=0}^{N} |R_j^{n+m} - R_j^n| dx \quad \leq \quad C(L_g, (n+m)dt, R^0) m dx.$$

which is no more than :

$$\|R_{\Delta x}((n+m)dt, \cdot) - R_{\Delta x}(ndt, \cdot)\|_{L^1([0,1])} \quad \leq \quad C(L_g, f, (n+m)dt, R^0) m dt. \tag{4.38}$$

4.A.5 The $L^1_{loc}(\mathbb{R}^+)$ continuity estimate

We define for $T > 0$, $n_T := E(T/dt)$. For all integers $0 \leq j \leq N-1$, we have the following estimate :

$$\begin{aligned} \sum_{n=0}^{n_T} |R_{j+1}^n - R_j^n| dt &= \sum_{n=0}^{n_T} dt |(a_j^{n+1})^{-1}(R_{j+1}^{n+1} - R_{j+1}^n)| \\ &\leq Cdt \sum_{n=0}^{n_T} |R_{j+1}^{n+1} - R_{j+1}^n|. \end{aligned}$$

where we have used the fact that the $a_{i,j}^n$s are bounded from below because of the Harten's condition (4.15). By the triangular inequality, one gets for $j_0 < j_1$

$$\sum_{n=0}^{n_T} |R_{j_1}^n - R_{j_0}^n| dt \quad \leq \quad Cdt \sum_{j=j_0}^{j_1} TV_{[1, n_T+1]}(R_{j+1}^{\bullet}). \tag{4.39}$$

Similarly to proof of the boundedness of the space total variation (see (4.37)), we can prove the corresponding result for the time total variation. That is to say :

$$\forall T > 0, 0 \leq j \leq N, \ TV_{[0, n_T]}(R_j^{\bullet}) \leq C(T, R^0, L_g).$$

Injecting this in (4.39), one gets :

$$\|R_{\Delta x}(\cdot, j_1 dx) - R_{\Delta x}(\cdot, j_0 dx)\|_{L^1([0,T])} \leq C(T, R^0, L_g, f)(j_1 - j_0) dx. \tag{4.40}$$

4.A.6 Conclusion

By Helly's Theorem, the sequence $(R_{\Delta x})_N$ converges in $L^1_{loc}(\mathbb{R}^+ \times [0,1])$ to a function $R \in L^1_{loc}(\mathbb{R}^+, BV([0,1])) \cap Lip_{loc}(\mathbb{R}^+, L^1([0,1]))$. It remains to prove that R is an entropy solution to (4.3).

Let $k \in \mathbb{R}^d$ and define $q_k(u) := f(u \top k) - f(k \perp u)$. The proof of the following result crucially depends on (4.13) and is given in a more general context in [26]. We give it here for completeness.

Lemma 19. *[Discrete entropy estimate]*
For all $k \in \mathbb{R}^d$, $N \in \mathbb{N}^*$, we have for all $n \in \mathbb{N}, 1 \leq j \leq N$:

$$\tfrac{1}{dt}(|R_j^{n+1} - k_i| - |R_j^n - k_i|) + \tfrac{1}{dx}(q_k(R_j^n + \tilde{R_j^n}) - q_k(R_{j-1}^n + \tilde{R_{j-1}^n})) \quad \leq \quad Cdx^{1-2\beta}. \tag{4.41}$$

Proof. Let R_p^n, R_m^n be defined as :

$$\left\{\begin{array}{ll} R_{p,j}^n := & R_j^n - 2\nu(f(R_j^n + \tilde{R}_j^n) - f(R_j^n)) \\ R_{m,j}^n := & R_j^n + 2\nu(f(R_{j-1}^n + \tilde{R}_{j-1}^n) - f(R_j^n)) \end{array}\right.$$

Let us also define :

$$\left\{\begin{array}{ll} E_{p,j}^n := & \eta(R_{p,j}^n) - \eta(R_j^n) + 2\nu(q_k(R_j^n + \tilde{R}_j^n) - q_k(R_j^n)) \\ E_{m,j}^n := & \eta(R_{m,j}^n) - \eta(R_j^n) - 2\nu(q_k(R_{j-1}^n + \tilde{R}_{j-1}^n) - q_k(R_j^n)). \end{array}\right.$$

We now estimate $E_{p,j}^n$. To do so, we define :

$$\begin{array}{ll} a(\tau) := & R_i^n + \tau\tilde{R}_i^n \\ c(\tau) := & R_i^n - 2\nu(f(a(\tau)) - f(R_i^n)). \end{array}$$

With these definitions, we can interpret the terms of $E_{p,j}^n$:

$$\begin{array}{rl} \eta(R_{p,j}^n) - \eta(R_j^n) = & \eta(c(1)) - \eta(c(0)) \\ = & \displaystyle\int_0^1 \frac{d}{d\tau}\eta(c(\tau))d\tau \\ = & \displaystyle -2\nu\int_0^1 \eta'(c(\tau))f'(a(\tau))\tilde{R}_i^n d\tau. \end{array}$$

For the entropy flux :

$$\begin{array}{rl} q_k(R_j^n + \tilde{R}_j^n) - q_k(R_j^n) = & q_k(a(1)) - q_k(a(0)) \\ = & \displaystyle\int_0^1 \frac{d}{d\tau}q_k(a(\tau))d\tau \\ = & \displaystyle\int_0^1 \eta'(a(\tau))f'(a(\tau))\tilde{R}_i^n d\tau. \end{array}$$

Hence, we have :

$$\begin{array}{rl} E_{p,j}^n = & \displaystyle 2\nu\int_0^1 f'(a(\tau))\tilde{R}_i^n(\eta'(a(\tau)) - \eta'(c(\tau)))d\tau \\ = & 2\nu\int_0^1 f'(a(\tau))\tilde{R}_i^n\eta''(\zeta_\tau)(a(\tau)) - c(\tau))d\tau \end{array}$$

where $\zeta_\tau \in]\min(a(\tau), c(\tau)), \max(a(\tau), c(\tau))[$. As :

$$|a(\tau) - c(\tau)| = |\tau\tilde{R}_j^n + 2\nu(f(R_j^n + \tau\tilde{R}_j^n) - f(R_j^n))|$$

The following estimate holds :

$$\begin{array}{rl} E_{p,j}^n \leq & 2\nu\|\eta''\|_{L^\infty(\mathbb{R})}v_{\max}|\tilde{R}_j^n|^2(1 + 2\nu v_{\max}) \\ \leq & Cdx^{2-2\beta}. \end{array} \tag{4.42}$$

By a similar proof, we have the same estimate for $E_{m,j}^n$. Finally, as

$$R_j^{n+1} = \frac{1}{2}(R_{p,j}^n + R_{m,j}^n).$$

Hence, by the convexity of the entropy :

$$\begin{array}{rl} \eta(R_j^{n+1}) \leq & \frac{1}{2}(\eta(R_{p,j}^n) + \eta(R_{m,j}^n)) \\ = & \eta(R_j^n) - \nu(q_k(R_j^n + \tilde{R}_j^n) - q_k(R_{j-1}^n + \tilde{R}_{j-1}^n)) + \frac{E_{p,j}^n + E_{m,j}^n}{2} \end{array}$$

which immediately give the result with (4.42) and the corresponding estimate for $E^n_{m,j}$.

□

Then we take $\varphi \in C_c(]0,T[\times]0,1[)$. Multiplying (4.41) by $I^n_j(\varphi) := \displaystyle\int_{t^n}^{t^{n+1}} \int_{x_j-dx/2}^{x_j+dx/2} \varphi(t,x)dtdx$ and summing over n, j, one gets :

$$\sum_{n=0}^{n_T}\sum_{j=1}^{N} \frac{\eta(R_j^{n+1})-\eta(R_j^n)}{dt} I^n_j(\varphi) + \sum_{n=0}^{n_T}\sum_{j=1}^{N} \frac{q_k(R_j^n+\tilde{R}_j^n - q_k(R^n_{j-1}+\tilde{R}^n_{j-1}))}{dx} I^n_j(\varphi) \leq C dx^{1-2\beta}.$$

A numerical integration by parts (owing the fact that φ is zero at the boundary of $[0,T]\times[0,1]$), we get :

$$\sum_{n=0}^{n_T}\sum_{j=1}^{N} \eta(R_j^n)\frac{I_j^{n-1}(\varphi)-I_j^n(\varphi)}{dt} + \sum_{n=0}^{n_T}\sum_{j=1}^{N} q_k(R_j^n+\tilde{R}_j^n)\frac{I_j^n(\varphi)-I_{j+1}^n(\varphi)}{dx} \leq C dx^{1-2\beta}.$$

Passing to the limit owing that $\beta < 1/2$, one gets

$$\int_0^T \int_0^1 \eta_k(R)\partial_t\varphi + q_k(R)\partial_x\varphi dxdt \geq 0.$$

Boundary conditions and convergence estimates

To finish the proof of the existence of a solution to (4.3), it remains to show that the initial and boundary conditions are verified in the almost everywhere sense. Let $\varepsilon > 0$ and $0 < s < 1$

$$\begin{aligned} ||R(0,.)-R_{\Delta x}(0,.)||_{L^1([0,1])} &\leq ||R(0,.)-R(s,.)||_{L^1([0,1])} + ||R(s,.)-R_{\Delta x}(s,.)||_{L^1([0,1])} \\ &\quad + ||R_{\Delta x}(s,.)-R_{\Delta x}(0,.)||_{L^1([0,1])} \\ &\leq 2C(g,R^0)s + ||R(s,.)-R_{\Delta x}(s,.)||_{L^1([0,1])}. \end{aligned}$$

where we have used (4.38) to get the last inequality and the fact that $R \in Lip_{loc}(\mathbb{R}^+, L^1([0,1]))$. Integrating with respect to s on an interval $[0,t]$ for $0 \leq t \leq 1$, one gets

$$\begin{aligned} ||R(0,.)-R_{\Delta x}(0,.)||_{L^1([0,1])} &\leq C(g,R^0)t + \frac{1}{t}\int_0^t ||R(s,.)-R_{\Delta x}(s,.)||_{L^1([0,1])}ds \\ &\leq C(g,R^0)t + \frac{1}{t}\int_0^1 ||R(s,.)-R_{\Delta x}(s,.)||_{L^1([0,1])}ds \end{aligned}$$

Taking $t = \frac{\varepsilon}{2C(g,R^0)}$ and N sufficiently large such that $\displaystyle\int_0^1 ||R(s,.)-R(s,.)||_{L^1([0,1])}ds \leq \frac{\varepsilon^2}{4C(g,R^0)}$, one finally obtains :

$$||R(0,.)-R_{\Delta x}(0,.)||_{L^1([0,1])} \leq \varepsilon.$$

By the fact that $R^0_{\Delta x}$ converges towards R^0 in $L^1([0,1])$, we deduce that $R(0,.) = R^0$ in a L^1 sense and $R(0,.) = R^0$ almost everywhere.

Remark 27. We can repeat the same procedure for any $t \geq 0$ and prove (4.18).

Remark 28. We can use the same technique to prove that the boundary condition $R(t,0) = g(R(t,1))$ is satisfied in the almost everywhere sense. In this case, the continuity estimate (4.40) is the key tool (see [47, p.24] for details).

The final step is to prove the convergence estimate (4.17). We give the sketch of the proof (see [47, p.24-25] for more details). In fact (4.17) is a consequence of Helly's Theorem applied to the sequence $(R_{\Delta x})_{N \geq \tilde{N}}$ where $\tilde{N} > 1$, on a time interval $[t, t+dt]$ ($dt > 0$). Then it suffices to take the limit when $\tilde{N}$ goes to infinity and when dt goes to zero.

CHAPITRE 5

STABILISATION PAR MÉTHODE SPECTRALE DE SYSTÈMES GÉNÉRAUX D'ÉQUATIONS DE TRANSPORT AVEC COUPLAGES AU SEIN DU DOMAINE

Le contenu de ce chapitre correspond à un article soumis chez Compte Rendus de l'Académie des Sciences (Mathématiques).

Abstract.

Ce chapitre se place dans un cadre linéaire où on rajoute un terme de couplage au sein du domaine. Contrairement aux chapitres précédents, les méthodes de Lyapunov sont ici inefficaces puisqu'il a été prouvé dans [13, 64] que si le terme de couplage est trop fort dans un certain sens alors on ne peut pas trouver une fonctionnelle de Lyapunov simple. Ainsi, on préférera utiliser une méthode spectrale de placement de pôles pour utiliser pleinement le caractère linéaire du système.

La suite de ce chapitre est écrite en anglais.

Sommaire

5.1 Introduction

5.1.1 Literature review

In the work, we investigate boundary stabilization of a class of linear first-order hyperbolic systems of PDEs on a finite space domain $x \in [0,1]$. Such systems are predominant in modeling traffic flow [9], heat exchangers [111], open channel flow [16, Chapter 1.4] or multiphase flow [38, 46, 40]. The couplings between states traveling in opposite directions, both in-domain and at the boundaries, may induce instability leading to undesirable behaviors. For example, oscillatory two-phase flow regimes occurring on oil and gas production systems directly result, in some cases, from these mechanisms [40]. The dynamics of most of these industrial systems are described by nonlinear transport equations. If we linearize systems presented before, one obtains a system of equations of the form :

$$\begin{cases} \partial_t R + \Lambda \partial_x R &= MR \\ R_1(t,0) &= u(t) \\ R_2(t,1) &= HR_1(t,1) \end{cases} \tag{5.1}$$

where $R = (R_1, R_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and :

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & -\Lambda_2 \end{pmatrix}, \ M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}.$$

Matrices have dimensions $\Lambda_1 \in D^+_{d_1}(\mathbb{R})$, $\Lambda_2 \in D^+_{d_2}(\mathbb{R})$, $M_{12} \in M_{d_1 d_2}(\mathbb{R})$, $M_{21} \in M_{d_2 d_1}(\mathbb{R})$, $M_{11} \in M_{d_1 d_1}(\mathbb{R})$ and $M_{22} \in M_{d_2 d_2}(\mathbb{R})$. It should be noticed that in most of the cases presented in the first paragraph, the linearized system is not homogeneous in the sense that matrices depend on the space variable.

To exponentially stabilize system (5.1), feedback controls $u(t)$ depending on the boundary values $R_2(t,0)$ were designed in the literature. Lyapunov techniques allow to establish exponential stabilization in Sobolev or C^p spaces when term M is supposed to be small. Applications to linearized Saint Venant systems are given in [64, 65, 37, 17, 42].

However when the in-domain coupling term M is too large, simple quadratic Lyapunov function may not be found [13, 64]. Moreover, spectral analysis shows that when the entries of M exceed a certain amplitude, the system is unstable for any control of the form $u(t) = FR(t,1)$ ($F \in M_{d_1 d_2}(\mathbb{R})$) [16, Proposition 5.2]. Note that in [16, Proposition 5.2] this was proven only for $d_1 = d_2 = 1$.

To overcome this problem, one can relax the assumption of a control depending only on the value of the state at the boundary. Doing so, it is possible to construct a full-state feedback depending on the value of R on all the domain $[0,1]$. In this work, we consider an integral feedback of the form $u(t) = \int_0^1 k(x)R(x)dx$ where k is a kernel to be defined. As a consequence, to use the proposed method, one needs to measure the state R on all the domain, which is impossible in industrial applications (see [38] for example). Some works [25, 54, 58, 39, 110] solve this difficulty designing a boundary observer of the state R. Here for simplicity, it is assumed that the observation of the state is complete in order to focus only on the effect of the control.

The core of the method is based on spectral theory applied to a well-behaved open-loop operator for which the spectrum is reduced to its point spectrum. Our analysis is greatly inspired from works [82, 96, 87] where the authors study a vast class of linear hyperbolic systems with in domain and boundary couplings. In order to explain the great lines of the proof of exponential stabilization, it is needed to define the problem in a semigroup form. This is the object of the next two sections.

5.1.2 Preliminaries

Let $\mathcal{H}$ be an Hilbert space. The scalar product on $\mathcal{H}$ is denoted $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and the associated norm is given by $\| \cdot \|_{\mathcal{H}}$.

The following formalism is taken from the book [107, Chapter 2]. Here, concepts are introduced without proof to ease the presentation. For more details, we refer to [107, Chapter 2].

Generator of a strongly continuous semigroup

The notion of semigroup is fundamental in this chapter. Its definition is given here :

Definition 13. *A family $(T_t)_{t\geq 0}$ of operators in $L(\mathcal{H})$ is a strongly continuous semigroup on $\mathcal{H}$ if :*

— $T_0 = I_{\mathcal{H}}$.

— $\forall t, \tau \geq 0,\ T_{t+\tau} = T_t T_\tau$ *(the semigroup property)*

— $\forall z \in \mathcal{H},\ \lim_{t\to 0} T_t z = z$ *(the strong continuity property)*

The generator of the semigroup $(T_t)_{t\geq 0}$ is defined as follows :

Definition 14. *Let D be the subset of $\mathcal{H}$ such that :*

$$D := \left\{ R \in \mathcal{H} \mid \lim_{t\to 0^+} \frac{T_t R - R}{t} \text{ exists in } \mathcal{H} \right\}.$$

Then, the operator $\mathcal{A} : D \to \mathcal{H}$ is defined such that :

$$\mathcal{A}R = \lim_{t\to 0^+} \frac{T_t R - R}{t},\ \forall R \in D.$$

This operator is called the generator of $(T_t)_{t\geq 0}$ and for the rest of the chapter, we use the very classic notation $T_t \leftarrow e^{\mathcal{A}t}, D \leftarrow D(\mathcal{A})$. It is also said that $\mathcal{A}$ generates the semigroup $e^{\mathcal{A}t}$.

The Lumer-Phillips Theorem states necessary and sufficient conditions on an unbounded operator $(\mathcal{A}, D(\mathcal{A}))$ to generate a strongly continuous semigroup.

Theorem 12. *[Lumer-Phillips] Let $\mathcal{A}$ be an unbounded operator defined on $D(\mathcal{A}) \subset \mathcal{H}$. The operator $\mathcal{A}$ generates a strongly continuous semigroup $e^{\mathcal{A}t}$ if and only if :*

— *$D(\mathcal{A})$ dense in $\mathcal{H}$.*

— *$\mathcal{A}$ is closed.*

— $\forall R \in D(\mathcal{A}),\ \langle R, \mathcal{A}R \rangle_{\mathcal{H}} \leq \zeta \|R\|^2_{\mathcal{H}}$ *where $\zeta \in \mathbb{R}$. This property is called the ζ dissipativity of $\mathcal{A}$.*

— *The resolvent set of $\mathcal{A}$*

$$\rho(\mathcal{A}) := \left\{ \lambda \in \mathbb{C} \mid \lambda I - \mathcal{A} \text{ is invertible and } (\lambda I - \mathcal{A})^{-1} \in \mathcal{L}(\mathcal{H}) \right\}$$

is not empty.

Remark 29. The operator $(\lambda I - \mathcal{A})^{-1}$ appearing in the definition of $\rho(\mathcal{A})$ is denoted by $R(\lambda, \mathcal{A})$. Moreover, one can easily generalize the notion of spectrum to unbounded operators ; the spectrum of $\mathcal{A}$ is given by :

$$\sigma(\mathcal{A}) := \mathbb{C} \setminus \rho(\mathcal{A}).$$

The point spectrum of $\mathcal{A}$ is given as :

$$\sigma_p(\mathcal{A}) := \{\lambda \in \sigma(\mathcal{A}) \mid \lambda I - A \text{ is not injective}\}.$$

If $R_0 \in D(A)$, it is not difficult to prove that $T(t)R_0 \in D(\mathcal{A})$ for all time $t \geq 0$ and :

$$\frac{d}{dt} e^{\mathcal{A}t} R_0 = \mathcal{A} e^{\mathcal{A}t} R_0, \forall t \geq 0.$$

In other words, if we note $R(t) := e^{\mathcal{A}t} R_0$:

$$\frac{d}{dt} R(t) = \mathcal{A} R(t), \ \forall t \geq 0. \tag{5.2}$$

Hence, the semigroup representation can be very useful to treat PDEs when the initial data is in the domain of the operator considered.

The adjoint semigroup

When R_0 is not in $D(\mathcal{A})$, things become more difficult since we are not allowed to differentiate $e^{\mathcal{A}t} R_0$. In order to consider less regular solutions, we need to introduce duality. The adjoint semigroup is a key tool to understand this aspect.

Let us define the space :

$$D(\mathcal{A}^\star) := \left\{ \varphi \in \mathcal{H} \mid \sup_{R \in D(\mathcal{A}), R \neq 0} \frac{\langle \mathcal{A}R, \varphi \rangle_{\mathcal{H}}}{\|R\|_{\mathcal{H}}} < \infty \right\}$$

and the adjoint operator $A^\star$ is defined by the Riesz representation theorem as :

$$\langle \mathcal{A}R, \varphi \rangle_{\mathcal{H}} = \langle R, \mathcal{A}^\star \varphi \rangle_{\mathcal{H}}, \ \forall R \in D(\mathcal{A}), \varphi \in D(\mathcal{A}^\star).$$

Hence, when R is supposed to be in $\mathcal{H}$ only, it is tempting to still write $\langle \mathcal{A}R, \varphi \rangle = \langle R, \mathcal{A}^\star \varphi \rangle$ when $\varphi \in D(\mathcal{A}^\star)$ so that the problem (5.2) is written in the duality form :

$$\langle \frac{d}{dt} R(t), \varphi \rangle_{\mathcal{H}} = \langle R(t), A^\star \varphi \rangle_{\mathcal{H}}, \forall \varphi \in D(\mathcal{A}^\star).$$

Here φ correspond to the test functions of our problem. Moreover and by density, one can prove that $e^{\mathcal{A}t} R_0$ is a solution to previous equation. Hence, we found a solution to problem (5.2) when R_0 is in $\mathcal{H}$.

The embedding $\mathcal{H}_1^d \subset \mathcal{H} \subset \mathcal{H}_{-1}$

In this chapter, the control operator is singular in the sense that it does not belong to $\mathcal{L}(\mathbb{U}, \mathcal{H})$ ($\mathbb{U}$ is the control space). Hence, we cannot use the bracket $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ when the control operator is involved. Fortunately, one can generalize the notion of duality to a larger space than $\mathcal{H}$.

Let $\mathcal{H}_1^d$ be a normed dense subset of $\mathcal{H}$. For all $R \in \mathcal{H}$, the star norm of R writes :

$$\|R\|_\star := \sup_{\varphi \in \mathcal{H}_1^d, \ \|\varphi\|_{\mathcal{H}_1^d} \leq 1} \langle R, \varphi \rangle_{\mathcal{H}}.$$

Definition 15. *The space* $\mathcal{H}_{-1}$ *is the completion of* $\mathcal{H}$ *with respect to the norm* $\|\cdot\|_\star$ *and hence* $\mathcal{H}_1^d \subset \mathcal{H} \subset \mathcal{H}_{-1}$. *Moreover, for all* $R \in \mathcal{H}_{-1}$ *and* $(R_n)_n$ *such that* $\lim_{n\to\infty} \|R - R_n\|_\star = 0$, *we define the duality bracket as :*

$$\langle R, \varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} := \lim_{n\to\infty} \langle R_n, \varphi\rangle_{\mathcal{H}}.$$

The space $\mathcal{H}$ *which defines the scalar product is called the pivot space. The space* $\mathcal{H}_{-1}$ *is named as the dual of* $\mathcal{H}_1^d$ *with respect to the pivot space* $\mathcal{H}$.

The notion of duality is very important when one treats solution which does not necessarily pertain to $D(\mathcal{A})$.

Theorem 13. *Let* $\lambda \in \rho(\mathcal{A})$ *and define :*

$$\mathcal{H}_1^d := (D(\mathcal{A}^\star), \|(\bar{\lambda} I - \mathcal{A}^\star)\cdot\|_{\mathcal{H}})$$

and $\mathcal{H}_{-1}$ *be the completion of* $\mathcal{H}$ *with respect to the norm* $\|(\lambda I - \mathcal{A})^{-1}\cdot\|_{\mathcal{H}}$. *Spaces* $\mathcal{H}_1^d$ *and* $\mathcal{H}_{-1}$ *are independent on the choice of* λ.

Additionally, the space $\mathcal{H}_{-1}$ *is the dual of* $\mathcal{H}_1^d$ *with respect to the pivot space* $\mathcal{H}$. *Moreover, one can extend the operator* $\mathcal{A}$ *so that (keeping the same notation for the extension)* $\mathcal{A} \in \mathcal{L}(\mathcal{H}, \mathcal{H}_{-1})$ *and the associated semigroup* (e^{At}) *can be extended in* $\mathcal{L}(\mathcal{H}_{-1})$.

Remark 30. In the proof of last theorem, the extension of $\mathcal{A}$ is built from $\mathcal{A}^{\star\star}$.

As a consequence, if the operator of control B is in $\mathcal{L}(\mathbb{U}, \mathcal{H}_{-1})$, it is easy to define a solution to :

$$\frac{d}{dt} R = \mathcal{A}R + Bu$$

with $u \in L^2([0,T])$ $(T > 0)$. Using the duality bracket, one can write last equation as :

$$\langle \frac{d}{dt} R, \varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} = \langle R, \mathcal{A}^\star\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \langle Bu, \varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d},\ \forall \varphi \in D(\mathcal{A}^\star)$$

and the closed-loop problem has sense.

5.1.3 The abstract problem

Now that the semigroup framework has been recalled, we can apply it to our problem. Let $\mathcal{H} = L^2([0,1]; \mathbb{C}^d)$ be the base space embedded with the usual scalar product :

$$\langle f, g\rangle_{\mathcal{H}} := \sum_{i=1}^{d} \int_0^1 \overline{f_i(x)} g_i(x) dx.$$

The open-loop operator $\mathcal{A}$ is given by :

$$\begin{cases} D(\mathcal{A}) = & \{R \in \mathcal{H} \mid R' \in \mathcal{H}, R_1(0) = 0, R_2(1) = HR_1(1)\} \\ \mathcal{A}R = & -\Lambda R' + MR. \end{cases}$$

It is easy to check (left for the reader) that $\mathcal{A}$ is a closed densely defined operator that satisfies the hypothesis of Lumer-Phillips Theorem. Hence, it generates a strongly continuous semigroup denoted $(e^{\mathcal{A}t})_{t\geq 0}$ on $\mathcal{H}$. For its adjoint $\mathcal{A}^\star$, it can be shown that :

$$\begin{cases} D(\mathcal{A}^\star) = & \{R \in \mathcal{H} \mid R' \in \mathcal{H}, R_2(0) = 0, \Lambda_1 R_1(1) = H^T \Lambda_2 R_2(1)\} \\ \mathcal{A}^\star R = & \Lambda R' + M^T R. \end{cases}$$

The control space $L^2(\mathbb{R}_+, \mathbb{U}) := L^2(\mathbb{R}_+, \mathbb{R}^{d_1})$ is embedded with the canonical norm of $L^2(\mathbb{R}_+, \mathbb{R}^{d_1})$. To define the control operator, we introduce the space $\mathcal{H}_{-1}$ which is the dual of $D(A^\star)$ when we take $\mathcal{H}$ as pivot space, namely :

$$\mathcal{H}_{-1} := \overline{(D(A), \|R(\lambda, \mathcal{A}) \cdot \|_{\mathcal{H}})}.$$

where $\lambda \in \rho(\mathcal{A})$ is taken arbitrarily in the resolvant set of $\mathcal{A}$. Moreover, the primal of $\mathcal{H}_{-1}$ is denoted :

$$\mathcal{H}_1^d := (D(\mathcal{A}^\star), \|(\lambda I - A^\star) \cdot \|_{\mathcal{H}}).$$

The control operator $\mathcal{B} \in \mathcal{L}(\mathbb{U}, \mathcal{H}_{-1})$ writes :

$$\mathcal{B}u := (\sqrt{\Lambda_1} u, 0_{d_2})\delta(x) \in \mathcal{H}_{-1}$$

where δ is the usual Dirac delta distribution. Note that in this chapter, we will not make the difference between $\mathcal{A}, \mathcal{A}^\star$ and their canonical extension $(\mathcal{A})^{\star\star}, (\mathcal{A}^\star)^{\star\star}$ in $\mathcal{L}(\mathcal{H}, \mathcal{H}_{-1})$.

With this notation, the abstract evolution problem on $\mathcal{H}_{-1}$ is defined here :

$$\begin{cases} \frac{dR}{dt} = & \mathcal{A}R + \mathcal{B}u(t) \\ R(0) = & R_0. \end{cases} \tag{5.3}$$

The notation being introduced, one can present the main theorem of this work and the rest of the chapter is dedicated to its proof.

Theorem 14. *There exists a linear feedback control of the form $u := \mathcal{K}R$ with $\mathcal{K} \in \mathcal{L}(\mathcal{H})$ for which there exists $C, \delta > 0$ such that for all initial condition $R_0 \in \mathcal{H}$, the system* (5.3) *is well-posed and the corresponding unique solution $R \in C(\mathbb{R}^+, \mathcal{H}_{-1})$ to* (5.3) *verifies :*

$$\|R(t)\|_{\mathcal{H}_{-1}} \leq Ce^{-\delta t}\|R_0\|_{\mathcal{H}_{-1}}, \ \forall t \geq 0.$$

A sketch of the proof of this exponential stabilizability result is given below.

- First, we prove that $\mathcal{A}$ has only a discrete spectrum with a finite number of unstable eigenvalues.
- Then, the focus is on the unstable finite dimensional part $\mathcal{M}$ of the system using a projection on the unstable eigenspace. Proving a controlability result for the operator $(A_{|\mathcal{M}}, B_{|\mathcal{M}})$, it is possible to use the pole placement theorem to find a state feedback control stabilizing the unstable part.
- Finally, we prove that the whole closed-loop system is well-posed and exponentially stable. This is not immediate since the control synthesized from the finite dimensional unstable part can destabilize the remaining one.

The chapter is organized as follows :

Outline : In Section 5.2, we give the definition of solution, state and prove an admissibility condition for well-posedness. Section 5.3 is dedicated to the spectral study of the open-loop operator $\mathcal{A}$. Then, Section 5.4 gives a rigorous proof of Theorem 14. Next, the control constructed in Theorem 14 is saturated in a certain sense and we give a local stability result in Section 5.5. In the same part, numerical illustrations of the results previously presented, are exposed. Finally, as a conclusion, perspectives and open problems are stated.

5.2 Solution definition and admissibility condition

Let $u \in L^2(0,T,\mathbb{U})$ be given. The definition of a solution to (5.3) is proposed below :

Definition 16. *The function $R \in C^1([0,T],\mathcal{H}_{-1}) \cap C([0,T];\mathcal{H})$ is a solution to* (5.3) *if for all $\varphi \in \mathcal{H}_1^d$*

$$\langle R(t),\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} = \langle R_0, e^{\mathcal{A}^\star t}\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \int_0^t \langle u(s), \mathcal{B}^\star e^{\mathcal{A}^\star(t-s)}\varphi\rangle_{\mathbb{U},\mathbb{U}} ds, \ \forall 0 \leq t \leq T. \tag{5.4}$$

where $\mathbb{U}$ is identified with its dual.

The following lemma is an admissibility result allowing to prove that the solution has regularity in $\mathcal{H}$.

Lemma 20. *The admissibility property holds for all time $T \geq 0$:*

$$\int_0^T \|\mathcal{B}^\star e^{\mathcal{A}^\star(T-t)}\varphi\|_{\mathbb{U}}^2 dt \leq C^2\|\varphi\|_{\mathcal{H}}^2, \ \forall \varphi \in \mathcal{H}. \tag{5.5}$$

where C is a constant depending on the parameters of the problem.

Proof. Let $\varphi \in \mathcal{H}_1^d$; the case $\varphi \in \mathcal{H}$ is easily deduced by a density argument. Let $z(t) := e^{t\mathcal{A}^\star}\varphi$ the solution to :

$$\begin{cases} \partial_t z - \Lambda\partial_x z &= M^T z \\ z_2(0) = 0, & \Lambda_1 z_1(1) - H^T\Lambda_2 z_2(1) = 0 \\ z(t=0) &= \varphi. \end{cases}$$

We define the functional V by :

$$V(z) := \int_0^1 z_1^T z_1 e^{-\gamma x} + z_2^T z_2 e^{-\gamma(1-x)} dx = \int_0^1 z^T \Gamma z dx$$

where $\Gamma(x) = \mathrm{diag}(e^{-\gamma x} I_{d_1}, e^{-\gamma(1-x)} I_{d_2})$ will be chosen later. Using integration by parts, we get :

$$\begin{aligned} \frac{dV(z)}{dt} &= \int_0^1 \partial_t z^T \Gamma z + z^T \Gamma \partial_t z dx \\ &= \int_0^1 (\Lambda\partial_x z + M^T z)^T \Gamma z + z^T \Gamma(\Lambda\partial_x z + M^T z) dx \\ &= 2\int_0^1 z^T \Lambda\Gamma\partial_x z dx + \int_0^1 z^T(\Gamma M^T + M\Gamma) z dx \\ &= [z^T\Lambda\Gamma z]_0^1 - \int_0^1 z^T \Lambda \partial_x \Gamma z dx + \int_0^1 z^T(\Gamma M^T + M\Gamma) z dx. \end{aligned}$$

Using the fact that $z \in \mathcal{H}_1^d$ for all time, the boundary terms are estimated below :

$$\begin{aligned} [z^T\Lambda\Gamma z]_0^1 &= -z_1(0)^T\Lambda_1 z_1(0) + z_2(0)^T \Lambda_2 z_2(0) e^{-\gamma} \\ &\quad + z_1(1)^T \Lambda_1 z_1(1) e^{-\gamma} - z_2(1)^T\Lambda_2 z_2(1) \\ &= -z_1(0)^T\Lambda_1 z_1(0) \\ &\quad + z_2(1)^T \Lambda_2 H^T \Lambda_1^{-1} H^T \Lambda_2 z_2(1) e^{-\gamma} - z_2(1)^T \Lambda z_2(1). \end{aligned}$$

Thus, for $\gamma > 0$ large enough :

$$[z^T\Lambda\Gamma z]_0^1 \leq -\frac{1}{2} z_1(0)^T \Lambda_1 z_1(0) = -\frac{1}{2}\|B^\star e^{t\mathcal{A}^\star}\varphi\|_{\mathbb{U}}^2$$

Hence :

$$\frac{dV}{dt} \leq -\frac{1}{2}\|B^\star e^{t\mathcal{A}^\star}\varphi\|_{\mathbb{U}}^2 - \int_0^1 z^T \Lambda \partial_x \Gamma z dx + \int_0^1 z^T(\Gamma M^T + M\Gamma)z dx.$$

Using the fact that $\partial_x \Gamma \Lambda = -\gamma \Gamma |\Lambda|$, the following estimate on V holds :

$$\frac{dV}{dt} \leq -\frac{1}{2}\|B^\star e^{\mathcal{A}^\star t}\varphi\|_{\mathbb{U}}^2 + C_{m,\gamma}V$$

where $C_{m,\gamma}$ depends on M, γ. Integrating, one obtains :

$$V(t) - V(0)e^{C_{m,\gamma}t} \leq -\frac{1}{2}\int_0^t \|B^\star e^{\mathcal{A}^\star s}\varphi\|_{\mathbb{U}}^2 e^{C_{m,\gamma}(t-s)} ds$$

which immediately gives the existence of a constant C depending on t and the parameters of the problem such that :

$$C^2 V(z(0) = \varphi) \geq \int_0^T \|\mathcal{B}^\star e^{\mathcal{A}^\star (T-t)}\varphi\|_{\mathbb{U}}^2 dt$$

which is the required result since V is equivalent to the square norm on $\mathcal{H}$. The proof of the general case of $\varphi \in \mathcal{H}$ follows by density of $\mathcal{H}_1^d$ in $\mathcal{H}$. □

By [29, Theorem 2.37], Lemma 20 gives the following results.

Lemma 21. *If $u \in L^2(0, T, \mathbb{U})$ then there exists a unique solution to* (5.3) *in the sense of Definition 16.*

The aim of this chapter is to stabilize (5.3) showing the existence of an admissible operator (the notion of admissible operator will be defined latter) $\mathcal{K} : \mathcal{H}_{-1} \to \mathbb{U}$ such that :

$$\begin{cases} \frac{dR}{dt} &= (\mathcal{A} + \mathcal{B}\mathcal{K})R \\ R(0) &= R_0 \end{cases} \tag{5.6}$$

is well-posed and its solution verifies the bound :

$$\|R(t, \cdot)\|_{\mathcal{H}_{-1}} \leq Ce^{-\delta t}\|R_0\|_{\mathcal{H}_{-1}}, \forall t \geq 0$$

where $\delta, C > 0$ do not depend on R_0.

5.3 Spectral analysis of the open-loop problem

Before going into the proof of Theorem 14, we need some information the open-loop operator $\mathcal{A}$. More precisely, it is proved that $\mathcal{A}$ has a spectrum reduced to its point spectrum with a finite number of unstable eigenvalues.

Proposition 2. *The spectrum of the open-loop operator verifies :*

— $\sigma(\mathcal{A}) = \sigma_p(\mathcal{A})$.

— *There exists $r > 0$ such that :*

$$\sigma_p(\mathcal{A}) \subset \{z \in \mathbb{C} \mid \Re z < r\}.$$

— *The unstable part of the spectrum ; $\sigma_p(\mathcal{A}) \cap \{\lambda \in \mathbb{C} \mid \Re\lambda \geq 0\}$ has a finite cardinal.*

The proof will be the object of this section.

5.3.1 Structure of the spectrum

The first lemma states that the spectrum of $\mathcal{A}$ is in fact its point spectrum.

Lemma 22. *The spectra of $\mathcal{A}$ consists of isolated eigenvalues of finite geometric multiplicity ie $\sigma(\mathcal{A}) = \sigma_p(\mathcal{A})$.*

Proof. Let $\lambda \in \mathbb{C}$ and consider the unique solution to :

$$\begin{cases} -\Lambda R' + MR &= \lambda R \\ R_1(0) &= 0 \\ R_2(1) &= HR_1(0). \end{cases} \tag{5.7}$$

In particular, by defining :

$$T(x, y, \lambda) = e^{-\Lambda^{-1}(\lambda I_d - M)(x-y)}.$$

The solution to (5.7) is given by :

$$R(x) = T(x, 0, \lambda)(0, I_{d_2})^T v(0)$$

where $v(0) \in \mathbb{C}^{d_2}$. In order to satisfy the right border boundary condition, we need to impose :

$$(H, -I_{d_2})T(1, 0, \lambda)(0, I_{d_2})^T v(0) = 0_{d_1}.$$

Hence, denoting :

$$U(\lambda) := (H, -I_{d_2})T(1, 0, \lambda)(0, I_{d_2})^T,$$

the point spectrum of the system is given by the zeros of the following characteristic equation :

$$\det(U(\lambda)) = 0. \tag{5.8}$$

Moreover, for all $\lambda \in \sigma_p(\mathcal{A})$, the corresponding eigenspace is given by :

$$Eig(\mathcal{A}, \lambda) = \left\{ T(x, 0, \lambda)(0, I_{d_2})^T v(0) \mid v(0) \in \ker(U(\lambda)) \right\}.$$

The geometric multiplicity is less than d_2 (the dimension of $v(0)$) and for $\lambda \in \mathbb{C} \setminus \sigma_p(\mathcal{A})$:

$$R(\lambda, \mathcal{A})G = T(x, 0, \lambda)(0, I_{d_2})v(0) - \int_0^x T(x, y, \lambda)\Lambda^{-1}G dy, \ \forall G \in \mathcal{H} \tag{5.9}$$

with :

$$v(0) = U(\lambda)^{-1}(H, -I_{d_2}) \int_0^1 T(x, y, \lambda)\Lambda^{-1}G dy.$$

With (5.9), it is easy to see that $R(\lambda, \mathcal{A})$ is bounded in $\mathcal{L}(H)$ when $\lambda \notin \sigma_p(\mathcal{A})$ and hence $\sigma(\mathcal{A}) = \sigma_p(\mathcal{A})$. □

Remark 31. In the proof of the previous lemma, we have shown that :

$$R(\lambda, \mathcal{A}) = U(\lambda)^{-1}E(\lambda), \ \forall \lambda \in \rho(\mathcal{A}) \tag{5.10}$$

where $E(\lambda)$ is a $\mathcal{H}$ valued entire function of λ.

The expression (5.10) gives immediately that :

Lemma 23. *The algebraic multiplicity of $\lambda \in \sigma(\mathcal{A})$ is given by the multiplicity of the zeros of $\kappa(\lambda) := \det(U(\lambda))$.*

5.3.2 Analysis of the unstable part of the spectrum

In this section, we end the proof of Proposition 2. The result given next ensures that the characteristic equation of the spectrum (5.8) can be approximated by the same equation removing the effect of non-diagonal 0th order term coming from the matrix M for which the spectrum is known. To be clear, the following notation is introduced :

$$M_0 = \begin{pmatrix} M_{11} & 0 \\ 0 & M_{22} \end{pmatrix}$$

and $\mathcal{A}_0$ the operator in $\mathcal{H}$:

$$\begin{cases} D(\mathcal{A}_0) = & \{R \in \mathcal{H} \mid R' \in \mathcal{H}, R(0) = 0, R_2(1) = HR_1(1)\} \\ \mathcal{A}_0 R = & -\Lambda R' + M_0 R. \end{cases}$$

Obviously, we can prove Lemma 22 for operator $\mathcal{A}_0$ and exhibit a characteristic equation for $\mathcal{A}_0$:

$$\kappa_0(\lambda) := \det(U_0(\lambda)) = 0.$$

where :

$$U_0(\lambda) := (H, -I_{d_2}) e^{-\Lambda^{-1}(\lambda I_d - M_0)} (0, I_{d_2})^T. \tag{5.11}$$

By simple computations :

$$\det(U_0(\lambda)) = \det(e^{\Lambda_2^{-1}(\lambda - M_{22})})$$

which does not have zeros. This means that the operator $\mathcal{A}_0$ generates a semigroup and $\rho(\mathcal{A}_0) = \mathbb{C}$. We present the following result taken from [82] which states that spectrum of $\mathcal{A}_0$ and $\mathcal{A}$ are closed for large imaginary part :

Lemma 24. *Let $r > 0$ be such that $\mathbb{C}_r := \{z \in \mathbb{C} \mid |\Re z| \leq r\}$. We have the following :*

$$\lim_{|\Im\lambda| \to \infty} |U(\lambda) - U_0(\lambda)| = 0 \tag{5.12}$$

and the convergence is uniform on $\mathbb{C}_r$ for all $r > 0$.

Moreover, we have the following property which will be exploited for the complex functional κ_0 :

Lemma 25. *[82] Let f be an exponential polynomial of the form $f(\lambda) = \sum_{i=1}^{r} a_j e^{b_k \lambda}$ ($\lambda, a_j \in \mathbb{C}, b_j \in \mathbb{R}$). Let $Z := \{\lambda \in \mathbb{C} \mid f(\lambda) = 0\}$ denote the zero set of f. For all $\delta > 0, \alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$ there exists a constant $m(\delta, \alpha, \beta) > 0$ such that for all $\lambda \in \mathbb{C}$ satisfying $dist(\lambda, Z) > \delta$, $\alpha < \Re\lambda < \beta$, we have $|f(\lambda)| > m(\delta, \alpha, \beta)$.*

A direct consequence of previous lemma is that on all strips of the form $\alpha < \Re\lambda < \beta$, $\inf |\kappa_0(\lambda)| > 0$ (κ_0 does not admit zeros and is an exponential polynomial).

Corollary 1. *For all $\alpha < \beta$, we have*

$$\inf_{\lambda \in \mathbb{C},\ \alpha < \Re\lambda < \beta} |\kappa_0(\lambda)| > 0.$$

To conclude, Rouché's Theorem is recalled here :

Theorem 15. *[Rouché] Let $U \subset \mathbb{C}$ be an open connected set and f, g two meromorphic functions on U with finite number of zeros and poles. Let γ be a closed smooth curve in U that does not intersect the set of zeros of f or g and that forms the border ∂K of a compact set K. If*

$$|f(z) - g(z)| < |g(z)|, \ \forall z \in \gamma,$$

then :

$$Z_f - P_f = Z_g - P_g$$

where Z_f, Z_g designate the number of zeros of f and g in K and P_f, P_g designate the number of poles of f and g in K.

By applying Rouché's Theorem to $f = \kappa$ and $g = \kappa_0$ and using Lemma 24 and Corollary 1, one has that $\kappa(\lambda)$ has zeros located near the real axis. Owing this and the fact that $\kappa(\lambda)$ is an entire function, it has a finite number of zeros in the right-half plane. Combining this with Lemma 22, we easily conclude on the proof of Proposition 2.

5.4 Proof of Theorem 14

Let us denote $\mathcal{M}$ the finite dimensional unstable generalized eigenspace (Jordan blocks) of $\mathcal{A}$ and $\mathcal{M}'$ its topological complement. We denote by $\alpha := \dim(\mathcal{M})$. Let $P : \mathcal{H} \to \mathcal{M}$ be the projection onto $\mathcal{M}$ defined as [70, Theorem 6.17] :

$$P = -\frac{1}{2i\pi} \oint_{\Gamma} R(\lambda, \mathcal{A}) d\lambda$$

where Γ is any contour enclosing the unstable eigenvalues of $\mathcal{A}$ (this is possible because of the separation of unstable eigenvalues). As P and $\mathcal{A}$ commute, it is possible to decompose $\mathcal{A}$ on the topological sum $\mathcal{M} \bigoplus \mathcal{M}' = \mathcal{H}$ and the abstract stabilization problem becomes :

$$\begin{cases} P\dot{R} = & P\mathcal{A}PR + P\mathcal{B}u(t) \\ (I-P)\dot{R} = & (I-P)\mathcal{A}(I-P)R + (I-P)\mathcal{B}u(t). \end{cases} \tag{5.13}$$

5.4.1 Stabilization of the finite dimensional part

First, we stabilize the finite-dimensional part without considering the infinite-dimensional part taking $u(t)$ of the form $u(t) = KPR(t)$ where K is a matrix of dimension $d_1 \times \alpha$. Hence, we have to solve a finite-dimensional stabilization problem where the open-loop matrix is the restriction denoted $A_{\mathcal{M}}$ of A on $\mathcal{M}$ and the control matrix is $B_{\mathcal{M}} = P\mathcal{B} \in M_{\alpha d_1}(\mathbb{R})$.

Proposition 3. *The system $(A_{\mathcal{M}}, B_{\mathcal{M}})$ is controllable. Hence, there exists $K \in M_{d_1 \alpha}$ such that $A_{\mathcal{M}} + B_{\mathcal{M}} K$ is Hurwitz.*

Proof. We show the result by using the Fattorini-Hautus test (also known as the Popov–Belevitch–Hautus test). It is necessary to prove that :

$$\ker(\lambda I - A^{\star}_{\mathcal{M}}) \cap \ker B_{\mathcal{M}}{}^{\star} = \{0\}, \ \forall \lambda \in \mathbb{C} \tag{5.14}$$

which reduces the analysis to eigenspaces only (and not generalized eigenspaces). In order to prove (5.14), the eigenvalues of $\mathcal{A}^{\star}$ are calculated. This is equivalent to solve :

$$\begin{cases} \Lambda R' + M^T R = & \lambda R \\ R_1(1) = & \Lambda_1^{-1} H^T \Lambda_2 R_2(1) \\ R_2(0) = & 0. \end{cases} \tag{5.15}$$

To do so, we introduce the operator :

$$\tilde{T}(x, y, \lambda) := e^{\Lambda^{-1}(\lambda I_d - M^T)(x-y)}.$$

The solution to (5.15) is given by :

$$R(x) = \tilde{T}(x, 0, \lambda)(I_{d_1}, 0)^T v^\star(0)$$

where $v^\star(0) \in \mathbb{C}^{d_1}$. In order to satisfy the right-border boundary condition, it is needed to impose :

$$(-I_{d_1}, \Lambda_1^{-1} H^T \Lambda_2)\tilde{T}(1, 0, \lambda)(I_{d_1}, 0)^T v^\star(0) = 0_{d_1}.$$

Hence the spectrum of $\mathcal{A}^\star$ is given by the equation :

$$\det(\tilde{U}(\lambda)) := \det((-I_{d_1}, \Lambda_1^{-1} H^T \Lambda_2)\tilde{T}(1, 0, \lambda)(0, I_{d_2})^T) = 0, \tag{5.16}$$

where :

$$\tilde{U}(\lambda) = (-I_{d_1}, \Lambda_1^{-1} H^T \Lambda_2)\tilde{T}(1, 0, \lambda)(0, I_{d_2})^T.$$

Similarly to $\mathcal{A}$, the spectrum of $\mathcal{A}^\star$ corresponds to its point spectrum and for all $\lambda \in \sigma_p(\mathcal{A}^\star)$:

$$Eig(\mathcal{A}^\star, \lambda) = \left\{ \tilde{T}(x, 0, \lambda)(I_{d_1}, 0)^T v(0) \mid v(0) \in \ker(\tilde{U}(\lambda)) \right\}.$$

To conclude, it suffices to remark that for all $\lambda \in \sigma(\mathcal{A}^\star)$ and all $v^\star(0) \in \ker(\tilde{U}(\lambda))$:

$$B^\star \tilde{T}(x, 0, \lambda)(I_{d_1}, 0)^T v^\star(0) = 0 \iff v^\star(0) = 0.$$

This proves Fattorini's condition :

$$\ker(\lambda I - \mathcal{A}^\star) \cap \ker B^\star = \{0\}, \ \forall \lambda \in \mathbb{C}$$

which immediately implies (5.14). The finite-dimensional system $(A_\mathcal{M}, B_\mathcal{M})$ is controllable and we can apply a pole placement theorem to find a matrix gain K such that $A_\mathcal{M} + B_\mathcal{M} K$ is Hurwitz. □

5.4.2 Well-posedness of the closed-loop system

Now we take a gain matrix K stabilizing the finite-dimensional part of (5.13), define $\mathcal{K} := KP$ and system (5.6) split as follows :

$$\begin{cases} P\dot{R} = & (P\mathcal{A}P + P\mathcal{B}KP)R \\ (I-P)\dot{R} = & ((I-P)\mathcal{A}(I-P) + (I-P)\mathcal{B}KP)R. \end{cases} \tag{5.17}$$

Our notion of solution is given in the following definition :

Definition 17. *If* $R_0 \in \mathcal{H}$ *is the initial data considered and* $T > 0$*. The element* $R \in C^1([0,T], \mathcal{H}_{-1}) \cap C([0,T]; \mathcal{H})$ *is a solution to* (5.6) *if for all* $\varphi \in \mathcal{H}_1^d$

$$\langle R(t), \varphi \rangle_{\mathcal{H}_{-1}, \mathcal{H}_1^d} = \langle R_0, e^{\mathcal{A}^\star t} \varphi \rangle_{\mathcal{H}_{-1}, \mathcal{H}_1^d} + \int_0^t \langle KPR, \mathcal{B}^\star e^{\mathcal{A}^\star (t-s)} \varphi \rangle_{\mathbb{U}, \mathbb{U}} ds, \ \forall 0 \leq t \leq T. \tag{5.18}$$

where $\mathbb{U}$ *is identified with its dual.*

Proposition 4. *There exists a unique solution to* (5.6) *in the sense of Definition 17.*

Proof. Let $T > 0$. We use a Banach-Picard fixed-point theorem proving existence and uniqueness at the same time. Let us define $\mathcal{T} : C([0,T], \mathcal{H}) \to C([0,T], \mathcal{H})$ the application such that for all $\varphi \in \mathcal{H}$:

$$\langle (\mathcal{T}R)(t), \varphi \rangle_{\mathcal{H}} = \langle R_0, e^{\mathcal{A}^\star t} \varphi \rangle_{\mathcal{H}} + \int_0^t \langle KPR(s), \mathcal{B}^\star e^{\mathcal{A}^\star (t-s)} \varphi \rangle_{\mathbb{U}, \mathbb{U}} ds, \ \forall R \in C([0,T], \mathcal{H}), 0 \leq t \leq T.$$

Let $R, Q \in C([0,T], \mathcal{H})$, we have :

$$\langle (\mathcal{T}(R-Q))(t), \varphi \rangle_{\mathcal{H}} = \int_0^t \langle KP(R(s) - Q(s)), \mathcal{B}^\star e^{\mathcal{A}^\star (t-s)} \varphi \rangle_{\mathbb{U}, \mathbb{U}} ds, \ \forall 0 \leq t \leq T.$$

By Cauchy-Schwartz inequality and Lemma 20 :

$$\begin{aligned} \langle (\mathcal{T}(R-Q))(t), \varphi \rangle_{\mathcal{H}} &\leq \sqrt{\textstyle\int_0^t \|\mathcal{B}^\star e^{(t-s)\mathcal{A}^\star} \varphi\|_{\mathbb{U}}^2 ds} \times \sqrt{\textstyle\int_0^t \|KP(R(s) - Q(s))\|_{\mathbb{U}}^2 ds} \\ &\leq \sqrt{C\|\varphi\|_{\mathcal{H}}^2} \times \sqrt{t\|KP\|_{\mathcal{L}(\mathcal{H},\mathbb{U})}^2 \|R-Q\|_{C([0,T],\mathcal{H})}^2} \end{aligned}$$

where we have used the fact that KP is in $\mathcal{L}(\mathcal{H}, \mathbb{U})$. Indeed, P is a projection, hence $P \in \mathcal{L}(\mathcal{H})$ and K is a matrix. As a consequence,

$$\langle (\mathcal{T}(R-Q))(t), \varphi \rangle_{\mathcal{H}} \leq C\sqrt{T} \|KP\|_{\mathcal{L}(\mathcal{H},\mathbb{U})} \|R-Q\|_{C([0,T],\mathcal{H})} \|\varphi\|_{\mathcal{H}}, \ \forall 0 \leq t \leq T.$$

Taking T sufficiently small (uniformly with respect to R_0), it holds :

$$\langle (\mathcal{T}(R-Q))(t), \varphi \rangle_{\mathcal{H}} \leq \frac{\|R-Q\|_{C([0,T],\mathcal{H})}}{2} \|\varphi\|_{\mathcal{H}}, \ \forall 0 \leq t \leq T.$$

As a consequence,

$$\|\mathcal{T}(R-Q))(t)\|_{\mathcal{H}} \leq \frac{\|R-Q\|_{C([0,T],\mathcal{H})}}{2}, \ \forall 0 \leq t \leq T.$$

We can apply Banach-Picard theorem to assert the existence of a unique fixed point of $\mathcal{T}$ in $C([0,T], \mathcal{H})$ for T sufficiently small. By a bootstrap argument, we conclude on the existence and uniqueness in $C([0,T], \mathcal{H})$ for all $T > 0$. This unique solution is denoted by R and for $\varphi \in \mathcal{H}_1^d$:

$$\langle R(t), \varphi \rangle_{\mathcal{H}_{-1}, \mathcal{H}_1^d} = \langle R_0, e^{\mathcal{A}^\star t} \varphi \rangle_{\mathcal{H}_{-1}, \mathcal{H}_1^d} + \int_0^t \langle KPR(s), \mathcal{B}^\star e^{\mathcal{A}^\star (t-s)} \varphi \rangle_{\mathbb{U}, \mathbb{U}} ds, \ \forall 0 \leq t \leq T. \tag{5.19}$$

The equation (5.19) is equivalent to :

$$\langle R(t),\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} = \langle R_0, e^{\mathcal{A}^\star t}\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \int_0^t \langle \mathcal{B}KPR(s), e^{\mathcal{A}^\star (t-s)}\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} ds,\ \forall 0\leq t\leq T.$$

Owing the fact that $R\in C([0,T],\mathcal{H})$, $\mathcal{B}\in\mathcal{L}(\mathbb{U},\mathcal{H}_{-1})$ and $s\mapsto e^{\mathcal{A}^\star s}\varphi\in C^1([0,T],\mathcal{H}_1^d)$, one deduces that $R\in C^1([0,T],\mathcal{H}_{-1})$. Moreover :

$$\begin{aligned}\langle \frac{dR}{dt}(t),\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} &= \langle R_0, \mathcal{A}^\star e^{\mathcal{A}^\star t}\varphi\rangle_{\mathcal{H}} + \langle KPR(t), \mathcal{B}^\star\varphi\rangle_{\mathbb{U},\mathbb{U}}\\ &\quad - \int_0^t \langle KPR(s), \mathcal{B}^\star e^{\mathcal{A}^\star (t-s)}\mathcal{A}^\star\varphi\rangle_{\mathbb{U},\mathbb{U}} ds,\ \forall 0\leq t\leq T.\end{aligned}$$

All terms in last equation are convergent because of Lemma 20. Indeed,

$$\begin{aligned}\textstyle\int_0^t \langle KPR(s), \mathcal{B}^\star\mathcal{A}^\star e^{\mathcal{A}^\star (t-s)}\varphi\rangle_{\mathbb{U},\mathbb{U}} ds &\leq C||\mathcal{A}^\star\varphi||_{\mathcal{H}}\times ||KP||_{\mathcal{L}(\mathcal{H})}||R||_{C([0,T],\mathcal{H})}\\ &\leq C||\varphi||_{\mathcal{H}_1^d}\times ||KP||_{\mathcal{L}(\mathcal{H})}||R||_{C([0,T],\mathcal{H})}.\end{aligned}$$

This concludes the proof of Proposition 4. □

5.4.3 Conclusion on the stability of the whole system

The vector gain $K\in M_{1\alpha}(\mathbb{R})$ (α is the dimension of $\mathcal{M}$) is chosen such that the finite dimensional system exponentially converges to zero with rate $\tau>0$ fixed.

$$\|PR\|_{\mathcal{H}}\leq Ce^{-\tau t},\ \forall t\geq 0. \tag{5.20}$$

In order to conclude on the whole stability, we have to prove that the infinite-dimensional part in (5.17) is not destabilized by the control. To prove the stability of the infinite-dimensional part, it suffices to see the second equation of (5.17) as an inhomogeneous Cauchy problem on $\mathcal{M}'$:

$$(I-\dot{P})R = \mathcal{A}_{\mathcal{M}'}(I-P)R + (I-P)\mathcal{B}u(t)$$

where $u(t)$ verifies the following estimate :

$$|u|_{\mathbb{U}}\leq Ce^{-\tau t},\ \forall t\geq 0$$

and $\mathcal{A}_{\mathcal{M}'}$ is the restriction of $\mathcal{A}$ on $\mathcal{M}'$. Our choice of feedback matrix K gives that the spectrum of $\mathcal{A}_{\mathcal{M}'}$ is stable :

$$\sigma(\mathcal{A}_{\mathcal{M}'})\subset\{z\in\mathbb{C}\mid \Re z<-\tilde{\tau}\}$$

with $\tilde{\tau}<0$. To conclude, we need the following spectral mapping Theorem from [82] :

Theorem 16. *The spectral mapping theorem holds true :*

$$\sigma(e^{\mathcal{A}t})\setminus\{0\} = \overline{e^{\sigma(\mathcal{A})t}}\setminus\{0\},\ \forall t\geq 0. \tag{5.21}$$

Corollary 2. *Property* (5.21) *is also true for* $\mathcal{A}_{M'}$*.*

Proof. To characterize the spectrum of a semigroup when comparing it with its generator, we will need the following theorem :

Theorem 17. *Gearhart-Prüss Spectral Mapping Theorem [56, 94]*
Let $(e^{At})_{t\geq 0}$ be a C_0 semigroup generated by A in a Hilbert space. Then $e^{\lambda t} \in \rho(e^{At})$ iff

$$\left\{\lambda + i2\pi t^{-1} z \mid z \subset \mathbb{Z}\right\} \in \rho(A) \text{ and } \sup_{z\in\mathbb{Z}} \|R(\lambda + i2\pi t^{-1} z, A)\| < \infty.$$

We first show that $\sigma(e^{\mathcal{A}_{\mathcal{M}'} t}) \setminus \{0\} \subset \overline{e^{\sigma(\mathcal{A}_{\mathcal{M}'})t}} \setminus \{0\}$. Let $\lambda \in \mathbb{C}$ such that $e^{\lambda t} \notin \overline{e^{\sigma(\mathcal{A}_{\mathcal{M}'})t}}$. We have to show that $e^{\lambda t} \in \rho(e^{\mathcal{A}_{\mathcal{M}'} t})$. By the definition of λ, there exists $\delta > 0$ such that :

$$\bigcup_{z\in\mathbb{Z}} B(\lambda + 2i\pi z/t, \delta) \subset \rho(\mathcal{A}_{\mathcal{M}'}). \tag{5.22}$$

We need to prove that $R(\lambda+2i\pi z/t, \mathcal{A}_{\mathcal{M}'}) = R(\lambda+2i\pi z/t, \mathcal{A})(I-P)_{|\mathcal{M}'}$ [70, Theorem 6.17] is uniformly bounded with respect to $z \in \mathbb{Z}$. Note that for z large enough, $\lambda+2i\pi z/t \in \rho(\mathcal{A})$ because of Lemma 24 and Rouché theorem. As a consequence, it suffices to prove that $R(\lambda + 2i\pi z/t, \mathcal{A})$ is bounded when z goes to infinity. The following lemma will help us to conclude :

Lemma 26. *[82] Let $U \subset \rho(\mathcal{A})$ so that $\sup_{\lambda\in U} |\Re\lambda| < \infty$ and $\inf_{\lambda\in U} |\kappa_0(\lambda)| > 0$. Then, there exists $d > 0$ such that for $\lambda \in U$ and $|\Im\lambda| \geq d$:*

$$R(\lambda, \mathcal{A}) = R(\lambda, \mathcal{A}_0) + \mathcal{E}(\lambda, \mathcal{A})/\lambda$$

with $R(\lambda, \mathcal{A}_0)$ and $\mathcal{E}(\lambda, \mathcal{A})$ bounded on U. In particular, $R(\lambda, \mathcal{A})$ is bounded on U.

Because of Corollary 1, one has :

$$\inf_{\bigcup_{z\in\mathbb{Z}} B(\lambda+2i\pi z/t,\delta)} |\kappa_0(\lambda)| \neq 0. \tag{5.23}$$

We apply previous lemma to $U = \bigcup_{z\in\mathbb{Z}} B(\lambda + 2i\pi z/t, \delta)$ which gives immediately that $R(\lambda + 2i\pi z/t, \mathcal{A})$ is bounded when z goes to infinity. Hence,

$$\sigma(e^{\mathcal{A}_{\mathcal{M}'} t}) \setminus \{0\} \subset \overline{e^{\sigma(\mathcal{A}_{\mathcal{M}'})t}} \setminus \{0\}.$$

The reverse inclusion is classic for all closed densely defined operators. This finishes the proof of Corollary 2. □

Property (5.21) holds true for $\mathcal{A}_{\mathcal{M}'}$. Hence, the growth bound of $\mathcal{A}_{\mathcal{M}'}$ verifies :

$$\omega(\mathcal{A}_{\mathcal{M}'}) = \sup\{\Re\lambda \mid \lambda \in \sigma(\mathcal{A}_{\mathcal{M}'})\} =: -\tilde{\tau} < 0.$$

This allows to prove the following stability result :

Lemma 27. *There exists $C > 0$ such that for all $t \geq 0$:*

$$\|e^{\mathcal{A}_{\mathcal{M}'} t}(I - P)\|_{\mathcal{L}(\mathcal{H})} \leq Ce^{-\tilde{\tau}t} \tag{5.24}$$

and

$$\|e^{\mathcal{A}_{\mathcal{M}'} t}(I - P)\|_{\mathcal{L}(\mathcal{H}_{-1})} \leq Ce^{-\tilde{\tau}t}. \tag{5.25}$$

Proof. The first inequality is proven by Gelfand formula [107, Remark 2.2.16] and Corollary 2. For the second inequality, take $z \in \mathcal{H}_{-1}, \varphi \in \mathcal{H}_1^d$.

$$\begin{array}{rcl}\langle e^{\mathcal{A}_{\mathcal{M}'}t}(I-P)z,\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} &=& \langle z, e^{\mathcal{A}^\star t}(I-P^\star)\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d}\\ &\leq& \|z\|_{\mathcal{H}_{-1}}\|e^{\mathcal{A}^\star t}(I-P^\star)\varphi\|_{\mathcal{H}_1^d}\\ &=& \|z\|_{\mathcal{H}_{-1}}\|(\lambda I - A^\star)e^{\mathcal{A}^\star t}(I-P^\star)\varphi\|_{\mathcal{H}}\end{array}$$

As $e^{A^\star t}$ and $\lambda I - A^\star$ commutes and as $A^\star$ commutes with $P^\star$, one gets :

$$\begin{array}{rcl}\langle e^{\mathcal{A}_{\mathcal{M}'}t}(I-P)z,\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} &\leq& \|z\|_{\mathcal{H}_{-1}}\|e^{\mathcal{A}^\star t}(I-P^\star)(\lambda I-\mathcal{A}^\star)\varphi\|_{\mathcal{H}}\\ &\leq& \|z\|_{\mathcal{H}_{-1}}\|e^{\mathcal{A}^\star_{\mathcal{M}'}t}(I-P^\star)\|_{\mathcal{L}(\mathcal{H})}\|(\lambda I-\mathcal{A}^\star)\varphi\|_{\mathcal{H}}\\ &=& \|z\|_{\mathcal{H}_{-1}}\|e^{\mathcal{A}^\star_{\mathcal{M}'}t}(I-P^\star)\|_{\mathcal{L}(\mathcal{H})}\|\varphi\|_{\mathcal{H}_1^d}\end{array}$$

To conclude, we use the fact that the norm of the adjoint is also the norm of the anti-adjoint and by (5.24) :

$$\langle e^{\mathcal{A}_{\mathcal{M}'}t}z,\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} \leq \|z\|_{\mathcal{H}_{-1}}Ce^{-\tau t}\|\varphi\|_{\mathcal{H}_1^d}$$

which finishes the proof of the lemma. □

By the definition of solution (Definition 17), we have for all $\varphi \in \mathcal{H}_1^d$:

$$\langle R(t),\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} = \langle R_0, e^{\mathcal{A}^\star t}\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \int_0^t \langle KPR, \mathcal{B}^\star e^{\mathcal{A}^\star(t-s)}\varphi\rangle_{\mathbb{U},\mathbb{U}}ds.$$

The following decomposition is used to conclude on the exponential stability :

$$\langle R(t),\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} = \langle R(t),P^\star\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \langle R(t),(I-P^\star)\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d}$$

and we estimate both terms on the right hand side of last equation.

For the simplest one :

$$\begin{array}{rcl}\langle R(t),P^\star\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} &=& \langle PR(t),\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d}\\ &\leq& Ce^{-\tau t}\|\varphi\|_{\mathcal{H}_1^d}\end{array}$$

where we used (5.20) and the fact that $\mathcal{H}_1^d$ is continuously embedded in $\mathcal{H}$.

For the other estimate :

$$\begin{array}{rcl}\langle R(t),(I-P^\star)\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} &=& \langle R_0, e^{\mathcal{A}^\star t}(I-P^\star)\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \int_0^t\langle KPR,\mathcal{B}^\star e^{\mathcal{A}^\star(t-s)}(I-P^\star)\varphi\rangle_{\mathbb{U},\mathbb{U}}ds\\ &=& \langle e^{\mathcal{A}t}(I-P)R_0,\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \int_0^t\langle e^{\mathcal{A}(t-s)}(I-P)BKPR,\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d}ds\\ &=& \langle e^{\mathcal{A}_{\mathcal{M}'}t}(I-P)R_0,\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \int_0^t\langle e^{\mathcal{A}_{\mathcal{M}'}(t-s)}(I-P)BKPR,\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d}ds\\ &\leq& \Big(\|e^{\mathcal{A}_{\mathcal{M}'}t}(I-P)\|_{\mathcal{L}(\mathcal{H}_{-1})}\|R_0\|_{\mathcal{H}_{-1}}\\ && +\displaystyle\int_0^t \|B\|_{\mathcal{L}(\mathbb{U},\mathcal{H}_{-1})}\|K\|_{\mathcal{L}(\mathcal{H}_{-1},\mathbb{U})}\|PR(s)\|_{\mathcal{H}_{-1}}\|e^{\mathcal{A}_{\mathcal{M}'}(t-s)}(I-P)\|_{\mathcal{L}(\mathcal{H}_{-1})}dt\Big)\\ && \times\|\varphi\|_{\mathcal{H}_1^d}\end{array}$$

where we used the fact that P and $e^{s\mathcal{A}}$ commute.

With (5.25) and the fact that $\|PR(s)\|_{\mathcal{H}_{-1}} \leq e^{-\tau s}$:

$$\|(I-P)R(t)\|_{\mathcal{H}_{-1}} \leq e^{-\tilde{\tau}t}\|(I-P)R_0\|_{\mathcal{H}_{-1}} + C\int_0^t e^{-(t-s)\tilde{\tau}}e^{-\tau s}ds \tag{5.26}$$

where C depends on K and the parameters of the problem. Hence, the exponential stability in $\mathcal{H}_{-1}$ holds with rate (at least) $\min(\tilde{\tau}, \tau)$. This finishes the proof of Theorem 14.

Remark 32. If instead of (5.1), we focus on the stabilization of :

$$\left\{\begin{array}{lll} \partial_t R + \Lambda \partial_x R & = & MR \\ R_1(t,0) & = & FR_2(t,0) + u(t) \\ R_2(t,1) & = & HR_1(t,1) \end{array}\right. \tag{5.27}$$

where F is a real matrix of suitable dimension. Then, the characteristic equation of the system will be modified giving eigenvalues at high frequency. Indeed, in this case :

$$\kappa_0(\lambda) = \det\Big((H, -I_{d_2})e^{-\Lambda^{-1}(\lambda I_d - M_0)}(F, I_{d_2})^T\big) = \det(He^{-\Lambda_1^{-1}(\lambda I - M_{0,11})}F - e^{\Lambda_2^{-1}(\lambda I - M_{0,22})}\Big)$$

which may have zeros at $|\Im\lambda| \gg 1$. This has important consequences on the spectrum. Indeed, in such a case it is possible to have an infinite number of unstable eigenvalues ! To illustrate this, we give an example with $d_1 = d_2 = 1$:

$$\left\{\begin{array}{lll} \partial_t R + \Lambda \partial_x R & = & MR \\ R_1(t,0) & = & 2R_2(t,0) \\ R_2(t,1) & = & 2R_1(t,1) \end{array}\right.$$

with :

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, M = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Then, the characteristic equation $\kappa_0(\lambda) = 0$ becomes :

$$4e^{-(\lambda-1)} - e^{\lambda-1} = 0$$

for which solutions are explicit :

$$\forall k \in \mathbb{Z},\ \lambda_k = \log(2) + 1 + 2ik\pi$$

which immediately implies that the number of unstable poles is infinite.

5.5 Towards a nonlinear control

5.5.1 The saturation of the control

In this section, the following abstract problem is considered :

$$\left\{\begin{array}{lll} \frac{dR}{dt} & = & \mathcal{A}R + \mathcal{B}\sigma(\mathcal{K}R) \\ R(0) & = & R_0 \end{array}\right. \tag{5.28}$$

where $\sigma : \mathbb{U} \to \mathbb{U}$ is the usual saturation by components with $\sigma_s > 0$ the saturation level. If $\mathcal{K} := KP$ where P is the projection on the unstable space of $\mathcal{A}$ and K is a matrix, we can easily give the notion of solution :

Definition 18. *If $R_0 \in \mathcal{H}$ is the initial data considered. We say that $R \in C^1([0,T],\mathcal{H}_{-1}) \cap C([0,T];\mathcal{H})$ is a solution to* (5.6) *if for all $\varphi \in \mathcal{H}_1^d$*

$$\langle R(t),\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} = \langle R_0, e^{\mathcal{A}^\star t}\varphi\rangle_{\mathcal{H}_{-1},\mathcal{H}_1^d} + \int_0^t \langle \sigma(KPR), \mathcal{B}^\star e^{\mathcal{A}^\star (t-s)}\varphi\rangle_{\mathbb{U},\mathbb{U}} ds,\ \forall t \geq 0. \tag{5.29}$$

where $\mathbb{U}$ is identified with its dual.

Using a fixed point argument as in previous section, it is easy to prove that such a solution exists and is unique. As we have seen in the previous section, the stability of the finite- dimensional part implies the stability of the entire system. As a consequence, to estimate a basin of attraction, techniques from finite-dimension literature will be used. More precisely, the following theorem from [102, Theorem 3.1 p 125] is a key tool :

Theorem 18. *If there exist a symmetric definite positive matrix $W \in M_{\alpha\alpha}(\mathbb{R})$, $S > 0$, a matrix $Z \in M_{1\alpha}(\mathbb{R})$ such that :*

$$\begin{pmatrix} W(A_{\mathcal{M}} + (P\mathcal{B})K)^T + (A_{\mathcal{M}} + (P\mathcal{B})K)W & (P\mathcal{B})S - Z^T \\ S(P\mathcal{B})^T - Z & -2S \end{pmatrix} < 0 \tag{5.30}$$

and

$$\begin{pmatrix} W & WK^T - Z^T \\ KW - Z & \sigma_s^2 \end{pmatrix} \geq 0 \tag{5.31}$$

then, the cylinder $\mathcal{E}(W^{-1},1) := \left\{R \in \mathcal{H} \mid (PR)^T W^{-1}(PR) < 1\right\}$ is a region of stability for system (5.28)*.*

This theorem has a surprising consequence. The basin of attraction is expressed in the unstable finite-dimensional subspace of $\mathcal{A}$. As a consequence, if we suppose that we can find a solution W, S, Z for the matrix inequalities (5.30)-(5.31). Then, taking an initial data $R_0 \in \mathcal{H}$ in the cylinder $\mathcal{E}(W^{-1},1)$, one can take an arbitrarily large stable part $(I-P)R_0$ such that the corresponding solution is stable. Hence, the basin of attraction is not bounded in $\mathcal{H}$.

5.5.2 Numerical simulations

Linear feedback

In this section, an example illustrating the problem of stabilization, is discretized. The space and time steps are denoted respectively dx and dt. We denote $N := E(1/dx) \in \mathbb{N}$. For this example, the dimensions are $d_1 = d_2 = 1$ and :

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, M = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, H = 1.$$

The state vector $R \in M_{2N,1}(\mathbb{R})$ is now a vector such that $R(1 \cdots N), R(N+1 \cdots 2N)$ is the numerical version of R_1, R_2 respectively. The discretized open-loop operator $A \in M_{2N,2N}(\mathbb{R})$ is expressed using a classical upwind scheme :

$$A = \left[\begin{array}{cccc|cccc} -\frac{1}{dx}+M_{11} & 0 & \cdots & 0 & M_{12} & 0 & \cdots & 0 \\ \frac{1}{dx} & -\frac{1}{dx}+M_{11} & \ddots & 0 & 0 & M_{12} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{dx} & -\frac{1}{dx}+M_{11} & 0 & \cdots & \cdots & M_{12} \\ \hline M_{21} & 0 & \cdots & 0 & -\frac{1}{dx}+M_{22} & \frac{1}{dx} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & M_{21} & 0 & 0 & \ddots & -\frac{1}{dx}+M_{22} & \frac{1}{dx} \\ 0 & \cdots & 0 & M_{21}+\frac{H}{dx} & 0 & \cdots & 0 & -\frac{1}{dx}+M_{22} \end{array}\right].$$

Using the function "eig" from python, we plot the spectrum of A :

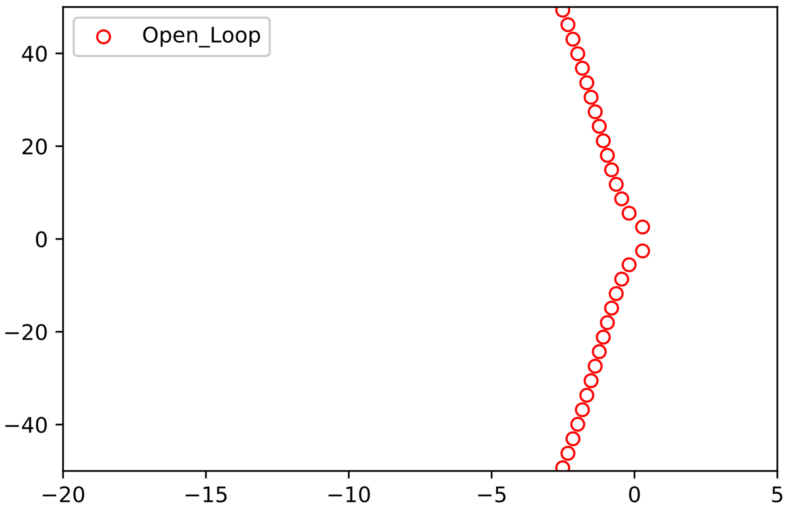


FIGURE 5.1 – The spectrum of the open-loop operator A for $dx = 10^{-3}$

One gets $2N$ simple eigenvalues with two unstable modes. The corresponding eigenvectors are stored in a matrix $V_{ec} \in M_{2N,2N}(\mathbb{C})$ for which eigenvectors are its columns. For complex conjugate eigenvalues, the corresponding couple of complex conjugate eigenvectors (denoted $F, \overline{F}$) is replaced by $\Re(F), \Im(F)$ to work with real matrices. Thus, $V_{ec} \in M_{2N,2N}(\mathbb{R})$ is now a real matrix. For the control operator $B \in M_{2N,1}(\mathbb{R})$, it can be written as :

$$B = \begin{bmatrix} 1/dx \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

To compute the control, matrices A, B are projected on the unstable eigenspace. To do so, V_{ec}^{-1} is computed and we denote the first two columns of V_{ex} by $P := (V_{ec,i,j})_{1 \leq i \leq 2N, 1 \leq j \leq 2}$.

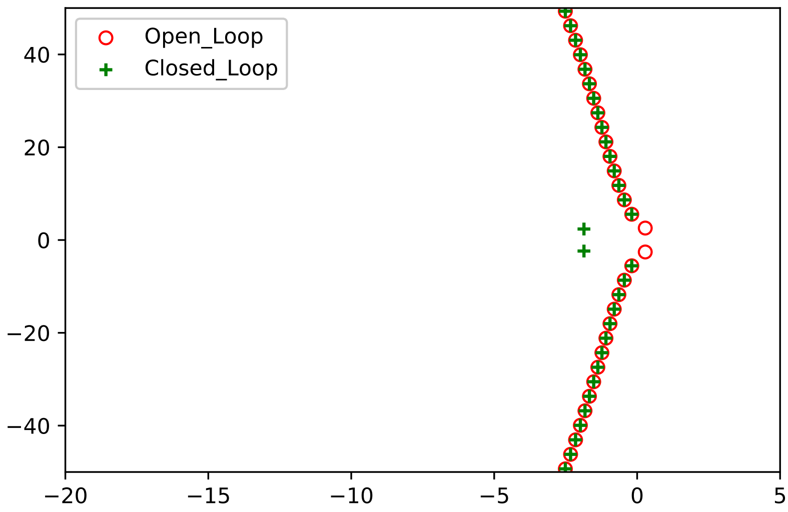
Similarly, $P^\star := (V_{ec}^{-1})_{1\leq i\leq 2N, 1\leq j\leq 2}$ is the first two columns of V_{ex}^{-1}. The projection on the unstable eigenspace writes :

$$A_{proj} = (P^\star)^T AP, \ B_{proj} = (P^\star)^T B.$$

Then, we apply a classic pole placement algorithm to matrices A_{proj}, B_{proj} to move unstable poles to the desired locus. It gives a feedback matrix $K_{proj} \in M_{1,2}(\mathbb{R})$ stabilizing the open-loop. To obtain the non-projected feedback matrix, it suffices to return to the state space :

$$K := K_{proj}(P^\star)^T \in M_{1,2N}(\mathbb{R}).$$

We get the following spectrum for the following closed-loop operator $A + BK$:



.

FIGURE 5.2 – The spectrum of the closed-loop operator $A + BK$ for $dx = 10^{-3}$

One can observe that poles are now all stable which implies that the closed loop system is stable.

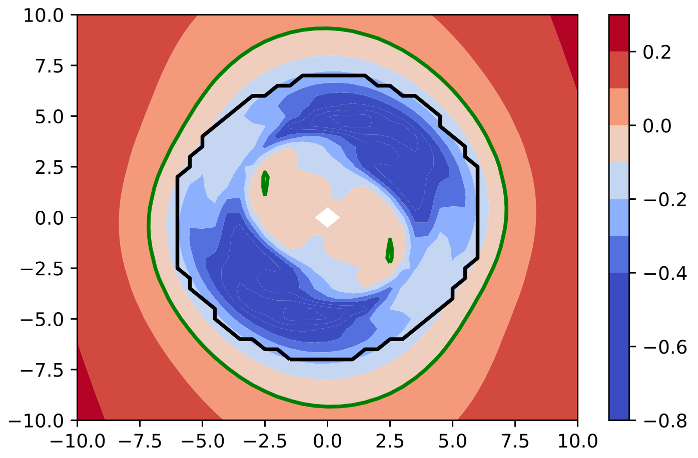
Saturated feedback

Now, the focus is on the case where we saturate the control with a saturation level $\sigma_s := 1$. For computations, we take $dx = 5 \times 10^{-3}$. We then evaluate the relevance of the estimation of the basin of attraction given in Theorem 18. It has been observed that the unstable space is of dimension 2 and is generated by vectors $P_1 := (P_{i,1})_i, P_2 := (P_{i,2})_i$. Hence, to study the basin of attraction, it is sufficient to study initial conditions of the form :

$$R_0 \in xP_1 + yP_2 \in M_{2N,1}(\mathbb{R}) \tag{5.32}$$

where $(x, y) \in \mathbb{R}^2$.

In Figure 5.3, the x-axis corresponds to the variable x in (5.32) and the y-axis corresponds to the variable y in (5.32). To estimate the real basin of attraction, at each initial data given in (5.32) we compute the solution using a time explicit Euler scheme (the time step being $dt = 0.9 \times dx$) and observe the exponential rate of its L^2 norm. This corresponds to the colormap in Figure 5.3. Divergence is associated to the red color whereas convergence corresponds to the blue color. The green curve represents the contour related to a convergence rate equal to zero. Finally, solving matrix inequalities (5.30)-(5.31) from Theorem 18, we give an estimation of the basin of attraction $\mathcal{E}(W^{-1}, 1)$ represented by the closed black curve in Figure 5.3. Note that we have used the package cxvpy for Python to solve the matrix inequalities problem. We remark that the estimation is very closed to the real basin of attraction.



.

FIGURE 5.3 – The basin of attraction and its estimation ($dx = 5 \times 10^{-3}$)

5.6 Conclusion

In this work, a general system of linear balance laws coupled by the domain and the boundary is studied. More precisely, the problem of stabilizability is treated using a spectral method. The main idea is to use a pole placement theorem applied on the unstable finite-dimensional part of the system. As a by-product, the linear control has been saturated to give a more realistic model of controller and a result is given to estimate the basin of attraction. In the final part, numerical simulations are given to illustrate results from the linear and the saturated theory.

Some questions remain open. A priori, it is not guarantied that the control proposed is robust with respect to perturbation. Proving this can be difficult since we do not use Lyapunov methods. Another important question is that the full-state feedback imposes the full observation which may be problematic in real applications. In order to have a realistic model of controller, one should analyze the coupling between the controller and an observer. It is far from being obvious that the stability will be preserved. Finally, the problem of local stability of nonlinear balance laws using pole placement methods was discarded.

165

CHAPITRE 6

LE BACKSTEPPING DISCRÉTISÉ : APPLICATION À UN EXEMPLE GÉNÉRAL DE SYSTÈME LINÉAIRE AVEC DEUX ÉQUATIONS

Abstract.

Dans ce dernier chapitre, nous allons nous concentrer sur un autre type de "full state feedback" : la commande de backstepping. Toujours dans un contexte linéaire, on discrétise, à l'aide d'un schéma upwind, un système de deux équations couplées au sein du domaine et au bord. D'une part, on montre à l'aide d'un exemple qu' appliquer la méthode de backstepping en discrétisant directement les équations du noyau n'est pas forcément pertinent et peut même générer des solutions instables. D'autre part, un schéma bien choisi pour la synthèse du contrôle est exhibé et un résultat de stabilité en temps fini est démontré en fin de chapitre.

La suite de ce chapitre est écrite en anglais.

Sommaire

6.1 Introduction

6.1.1 Literature review

In the chapter, we investigate the boundary stabilization of the same class of linear first-order hyperbolic systems of PDEs, as in previous chapter :

$$\begin{cases} \partial_t R + \Lambda_+(x)\partial_x R &= M_{11}(x)R + M_{12}(x)S \\ \partial_t S - \Lambda_-(x)\partial_x S &= M_{21}(x)R + M_{22}(x)S \\ R(t,0) &= u(t) \\ S(t,1) &= HR(t,1) \end{cases} \tag{6.1}$$

where $R \in \mathbb{R}^{d_1}, S \in \mathbb{R}^{d_2}$, $\Lambda_+, \Lambda_- > 0$ are positive diagonal matrices and $H, M_{11}, M_{12}, M_{21}, M_{22}$ are matrices of appropriate dimensions. To stabilize system (6.1), feedback controls $u(t)$ depending on the boundary values $S(t,0)$ were designed. Lyapunov techniques allows to establish exponential stabilization in Sobolev or C^p spaces when the terms $M_{\bullet\bullet}$ are supposed to be small. Applications to Saint Venant systems are given in [64, 65, 37, 17, 42].

However when the in-domain coupling term $M_{\bullet\bullet}$ is too large, a simple quadratic Lyapunov function does not exist [13, 64]. Moreover, spectral analysis shows that when $M_{\bullet\bullet}$ exceed a certain amplitude, the system is unstable for any control of the form $u(t) = FS(t,0)$ ($F \in M_{d_1 d_2}(\mathbb{R})$) [16, Proposition 5.2]. Note that in [16, Proposition 5.2] this was proven only for $d_1 = d_2 = 1$.

To solve this problem, one can relax the assumption of a feedback control depending only on boundary terms and use a full-state feedback control of the form $u(t) = \int_0^1 \alpha(\xi)R(t,\xi) + \beta(\xi)S(t,\xi)d\xi$ ($\alpha(\xi), \beta(\xi)$ are functions taking their values in $[0,1]$). Here, we focus on backstepping controls which is a particular case of full-state feedback controls. Besides, backstepping techniques were primarily designed for ODEs [72] where the main idea is to find a bijective transformation that maps the system in a simpler to stabilize one. There is a vast literature on the extension of backstepping to parabolic and hyperbolic PDEs [100, 73, 99]. The book [74] is a very pedagogical introduction to the topic.

In this chapter, we focus on the particular case where $d_1 = d_2 = 1$ that is to say that we have only two heterodirectional transport equations. By a change of variable (see [16, p 176]), it is possible to suppress diagonal zeroth order terms to obtain :

$$\begin{cases} \partial_t R + \lambda_+(x)\partial_x R &= \tilde{M}_{12}(x)S \\ \partial_t S - \lambda_-(x)\partial_x S &= \tilde{M}_{21}(x)R \\ R(t,0) &= u(t) \\ S(t,1) &= hR(t,1) \end{cases}$$

where $h \in \mathbb{R}, \lambda_+, \lambda_- > 0, \tilde{M}_{12}, \tilde{M}_{21} : [0,1] \to \mathbb{R}$, $u(t)$ is the control and $x \in [0,1]$. In this chapter, we neglect the space dependence of $\lambda_1, \lambda_2, \tilde{M}_{12}, \tilde{M}_{21}$ and suppose that they are constant. Moreover, in order to ease the reading, the tilda notation of $\tilde{M}_{12}, \tilde{M}_{21}$ is dropped in the rest of the chapter. As a consequence, the system under consideration writes :

$$\begin{cases} \partial_t R + \lambda_+\partial_x R &= M_{12}S \\ \partial_t S - \lambda_-\partial_x S &= M_{21}R \\ R(t,0) &= u(t) \\ S(t,1) &= hR(t,1) \end{cases} \tag{6.2}$$

where $\lambda_+, \lambda_- > 0$ and $M_{12}, M_{21} \in \mathbb{R}$.

More precisely, the focus is on the finite-time stabilization of the system with optimal time in $L^p([0,1])$ ($1 < p \leq \infty$). The problem is to find a full-state feedback control $u(t)$ such that :

$$\forall t \geq T_{\min} := \frac{1}{\lambda_+} + \frac{1}{\lambda_-},\ \|R(t,\cdot)\|_{L^p([0,1])} + \|S(t,\cdot)\|_{L^p([0,1])} = 0.$$

In the continuous setting, this problem is already solved [16, Chapter 7.4]. However to the author's knowledge, no result is known when one discretizes the equations. In [8, Section 3.3], the author designs a backstepping control from the continuous theory and inject it in the discretized closed loop. More precisely, inspired from [68], the author uses an iterative algorithm where the characteristic lines are calculated in order to compute the backstepping kernel (see Section 6.1.2 for the definition of the kernel). However, finding the characteristic lines makes the implementation quite difficult when characteristic velocities are not constant. Here, we do not use the iterative algorithm from [8, Section 3.3], the method rather relies on a finite volume scheme presented in Section 6.3, that is easier to implement.

To simulate the closed loop system, a classic upwind finite volume scheme is given. With an example, we will see that the finite volume scheme used to compute the backstepping control cannot be chosen arbitrarily. It is mandatory to apply the backstepping method directly on the scheme in itself in order to build a control. More precisely, if the schemes are not wisely chosen, then instabilities occur when the in-domain coupling is large. The contribution can be summed up as follows :

— We illustrate on an example that injecting a control synthesized from an arbitrary finite volume scheme does not stabilize the discretized closed loop system.

— We give a numerical framework for the numerical backstepping theory.

— We prove a finite time stabilization result for the discretized system.

Outline : The chapter is organized as follows. In Section 6.1.2, we recall the way to compute the backstepping control in the continuous theory. In Section 6.2, an example is given to show that both schemes for calculating the control and the solution must be wisely chosen. In Section 6.3, we propose another scheme and prove the finite time stabilization of the numerical system using new discretized backstepping techniques. In Section 6.4, numerical illustrations of our results are given. Finally, conclusions and perspectives are proposed in the last part of this chapter.

6.1.2 The continuous backstepping method

In this section, we recall the continuous backstepping procedure without giving any proof. The system (6.2) without boundary condition can be rewritten in the form,

$$\partial_t U + \Lambda \partial_x U = MU$$

where

$$U(t,x) = \begin{pmatrix} R(t,x) \\ S(t,x) \end{pmatrix}, \Lambda = \begin{pmatrix} \lambda_+ & 0 \\ 0 & -\lambda_- \end{pmatrix}, M = \begin{pmatrix} 0 & M_{12} \\ M_{21} & 0 \end{pmatrix}.$$

To find a feedback control, we use the strategy of backstepping. A second order Volterra transform allows to pass from the original system (6.2) to a target system for which finite time extinction is straightforward :

$$\begin{cases} \partial_t U^\star + \Lambda \partial_x U^\star = & 0, \\ U_1^\star(t,0) = & 0, \\ U_2^\star(t,1) = & hU_1^\star(t,1). \end{cases} \tag{6.3}$$

Note that we got rid of the 0th order term and that after a time $t = T_{\min}$, the solution to (6.3) is zero for any initial data.

More precisely, the Volterra transform is expressed as follows :

$$U^\star(t,x) = U(t,x) - \int_x^1 P(x,\xi)U(t,\xi)d\xi \tag{6.4}$$

where P takes its values in $M_{2,2}(\mathbb{R})$ and is defined on the triangle $\{(x,\xi) \mid 0 \le x \le \xi \le 1\}$.

With such transformation, we express the time derivative of $U^\star$:

$$\begin{aligned}
\partial_t U^\star(t,x) &= \partial_t U(t,x) - \int_x^1 P(x,\xi)\partial_t U(t,\xi)d\xi \\
&= \partial_t U(t,x) - \int_x^1 P(x,\xi)(-\Lambda\partial_\xi U + MU)(t,\xi)d\xi \\
&= \partial_t U(t,x) + P(x,1)\Lambda U(t,1) - P(x,x)\Lambda U(t,x) \\
&\quad - \int_x^1 (\partial_\xi P\Lambda + PM)(x,\xi)U(t,\xi)d\xi.
\end{aligned}$$

For the space derivative :

$$\partial_x U^\star(t,x) = \partial_x U(t,x) + P(x,x)U(t,x) - \int_x^1 \partial_x P(x,\xi)U(t,\xi)d\xi.$$

Gathering previous results, one gets the PDE solved by $U^\star$:

$$\begin{aligned}
\partial_t U^\star(t,x) + \Lambda\partial_x U^\star(t,x) &= \int_x^1 (-\Lambda\partial_x P - \partial_\xi P\Lambda - PM)(x,\xi)U(t,\xi)d\xi \\
&\quad + (M + \Lambda P(x,x) - P(x,x)\Lambda)U(t,x) \\
&\quad + P(x,1)\Lambda U(t,1).
\end{aligned}$$

To get that $U^\star$ solves (6.3), the function P is then chosen as the unique solution to the following system :

$$\left\{\begin{array}{lcl}
\partial_\xi P\Lambda + \Lambda\partial_x P + PM &=& 0 \\
M + \Lambda P(x,x) - P(x,x)\Lambda &=& 0 \\
P(x,1)\Lambda(1,h)^T &=& 0
\end{array}\right. \tag{6.5}$$

which is a well-posed transport system. See [16, chapter 7] or [34] for details. This can be rewritten in the usual PDE form as follows :

$$\left\{\begin{array}{ll}
\lambda_+\partial_\xi P_{11} + \lambda_+\partial_x P_{11} = & -M_{21}P_{12} \\
\lambda_-\partial_\xi P_{12} - \lambda_+\partial_x P_{12} = & +M_{12}P_{11} \\
\lambda_+\partial_\xi P_{21} - \lambda_-\partial_x P_{21} = & -M_{21}P_{22} \\
\lambda_-\partial_\xi P_{22} + \lambda_-\partial_x P_{22} = & +M_{12}P_{21}
\end{array}\right. \tag{6.6}$$

with boundary conditions :

$$\left\{\begin{array}{lcl}
P_{12}(x,x) &=& -\frac{M_{12}}{\lambda_+ + \lambda_-} \\
P_{21}(x,x) &=& \frac{M_{21}}{\lambda_+ + \lambda_-} \\
P_{11}(x,1) &=& \frac{h\lambda_-}{\lambda_+}P_{12}(x,1) \\
P_{22}(x,1) &=& \frac{\lambda_+}{h\lambda_-}P_{21}(x,1).
\end{array}\right. \tag{6.7}$$

Additionally, let us compute the trace of $U^\star$:

$$\begin{cases} U_1^\star(t,0) &= u(t) - \int_0^1 p_{11}(0,\xi)U_1(t,\xi) + p_{12}(0,\xi)U_2(t,\xi)d\xi \\ U_2^\star(t,1) &= U_2(t,1) = hU_1(t,1). \end{cases}$$

Hence to get the same boundary condition at $x = 0$ as in (6.3), the control $u(t)$ is chosen such that :

$$u(t) := \int_0^1 p_{11}(0,\xi)U_1(t,\xi) + p_{12}(0,\xi)U_2(t,\xi)d\xi. \tag{6.8}$$

Equations (6.6)-(6.8) defines the backstepping control allowing finite-time stabilization for the system under study.

6.2 The numerical backstepping approach : the naive way

In this section, a scheme is proposed to compute the backstepping kernel. It is shown by numerical illustrations that the associated control may not stabilize the closed loop system. This is why, the method is called the "naive" method in contrast with the one presented in Section 6.3 for which stability will be proved.

6.2.1 The scheme for the closed-loop system

We discretize the state (R, S) introducing two discretization parameters $dt > 0, dx = 1/N (N \in \mathbb{N})$. The numerical approximation of R, S is piecewise constant on cells $[ndt, (n+1)dt] \times [(j-1)dx, jdx]$ $(n \in \mathbb{N}^\star, 1 \le j \le N)$. To designate its values on such cells, we introduce the sequences $(R_j^n, S_j^n)_{n\in\mathbb{N}^\star, 1\le j\le N-1}$ and the numerical closed loop system is :

$$\forall j : 1 \le j \le N, \quad \begin{cases} \dfrac{R_j^{n+1} - R_j^n}{dt} + \lambda_+ \dfrac{R_j^n - R_{j-1}^n}{dx} &= M_{12}S_j^n \\ \dfrac{S_j^{n+1} - S_j^n}{dt} - \lambda_- \dfrac{S_{j+1}^n - S_j^n}{dx} &= M_{21}R_j^n. \end{cases}$$

For boundary conditions, we impose the ghost cell condition :

$$\begin{cases} R_0^n = & u^n \\ S_{N+1}^n = & hR_N^{n-1}. \end{cases}$$

where u^n is given by :

$$u^n := \sum_{j=1}^{N} (P_{11,N,j}R_j^{n-1} + P_{12,N,j}S_j^{n-1})dx. \tag{6.9}$$

where $P_{11}, P_{12}, P_{21}, P_{22} \in M_{N,N}(\mathbb{R})$ are the discretized version of P that will be defined just after.

6.2.2 Resolution of (6.5)

Before going into the simulations, we need to compute P solving (6.5) using a scheme that will be chosen later.

Equations in (6.5) can be seen as as a system of coupled transport equations on a triangular domain drawn in Figure 6.1 :

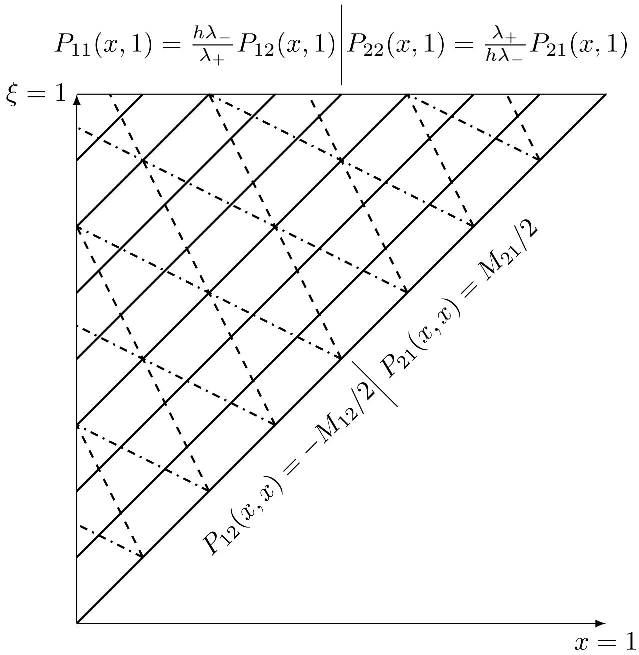


FIGURE 6.1 – The domain where P is defined

Here, the dot-dashed lines corresponds to characteristics for P_{12} and the dashed ones to those of P_{21} while the plain ones correspond to P_{11} and P_{22}.

One can see the variable x as a "time" variable and ξ as a space variable. Boundary conditions are imposed on the diagonal and the upper edge of the triangle. The corresponding "initial data" condition corresponds to the upper right corner of the triangle in Figure 6.1. In the next section, we give a naive method to solve these transport equations on such triangular domain.

6.2.3 A naive scheme to solve (6.5)

To solve system (6.5), a finite difference method is used where x is seen as the time variable while ξ corresponds to the space variable. To do so, we introduce the step $dx = 1/N$ ($N \in \mathbb{N}^*$) and the step $d\xi = 1/N$. The (x,ξ) mesh is then defined by :

$$\forall i \in \mathbb{N}, \quad \begin{array}{c} 1 \leq i \leq N, \\ i \leq j \leq N \end{array} \quad \left\{ \begin{array}{lcl} x_i & := & (i-1/2)dx \\ \xi_j & := & (j-1/2)d\xi. \end{array} \right.$$

As a consequence, the grid is Cartesian (and square) and drawn in Figure 6.2 :

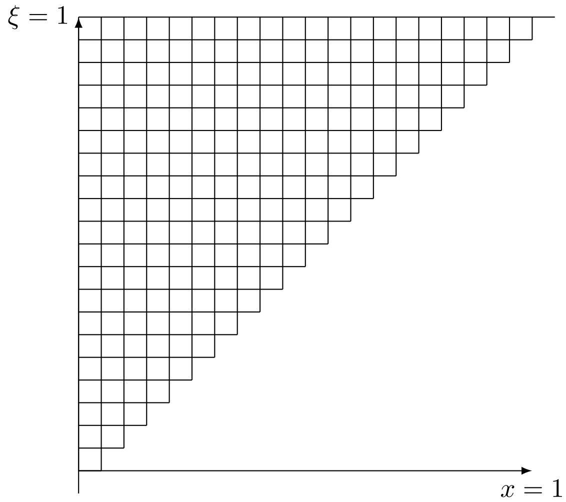


FIGURE 6.2 – The grid for the computation of P

The numerical approximation $P_{i,j}$ of P is piecewise constant on cells of the form $(x_i - dx/2, x_i + dx/2) \times (\xi_j - d\xi/2, \xi_j + d\xi/2)$. It is computed as follows :

— For all $1 \leq j \leq N$, $P_{12,N,j} = -\frac{M_{12}}{\lambda_+ + \lambda_-}$, $P_{21,N,j} = \frac{M_{21}}{\lambda_+ + \lambda_-}$ and $P_{11,N,N} = -\frac{h\lambda_-}{\lambda_+}\frac{M_{12}}{\lambda_+ + \lambda_-}$, $P_{22,N,N} = \frac{\lambda_+}{h\lambda_-}\frac{M_{21}}{\lambda_+ + \lambda_-}$.

— Suppose that $P_{i,\bullet}$ is given for some $i \leq N$. Recalling (6.6), we calculate to calculate P_{i-1} using an upwind scheme :

$$\begin{array}{lll} \forall j \leq i-1 & P_{12,i-1,j} = & -\frac{M_{12}}{\lambda_+ + \lambda_-} \\ \forall j \geq i & P_{12,i-1,j} = & P_{12,i,j} + \frac{\lambda_- dx}{\lambda_+ d\xi}(P_{12,i,j-1} - P_{12,i,j}) + dx\frac{M_{12}}{\lambda_+}P_{11,i,j}. \end{array}$$

$$\begin{array}{lll} \forall j \leq i-1 & P_{21,i-1,j} = & \frac{M_{21}}{\lambda_+ + \lambda_-} \\ \forall j \geq i & P_{21,i-1,j} = & P_{21,i,j} + \frac{\lambda_+ dx}{\lambda_- d\xi}(P_{21,i,j-1} - P_{21,i,j}) - dx\frac{M_{21}}{\lambda_-}P_{22,i,j}. \end{array}$$

$$\begin{array}{lll} & P_{11,i-1,N} = & \frac{h\lambda_-}{\lambda_+}P_{12,i+1,N} \\ \forall i-1 \leq j < N & P_{11,i-1,j} = & P_{11,i,j} + dx/d\xi(P_{11,i,j+1} - P_{11,i,j}) + dx\frac{M_{21}}{\lambda_+}P_{12,i,j}. \end{array}$$

$$\begin{array}{lll} & P_{22,i-1,N} = & \frac{\lambda_+}{h\lambda_-}P_{21,i+1,N} \\ \forall i-1 \leq j < N & P_{22,i-1,j} = & P_{22,i,j} + dx/d\xi(P_{22,i,j+1} - P_{22,i,j}) - dx\frac{M_{12}}{\lambda_-}P_{21,i,j}. \end{array}$$

— When all lines of P are computed, it is important to impose :

$$\forall i, \forall j < i, \star \in \{11, 12, 21, 22\},\ P_{\star,i,j} = 0$$

to have an upper triangular structure.

Remark 33. Note that the scheme for P_{12} exhibits a CFL number equal to $\frac{\lambda_- dx}{\lambda_+ d\xi} = \frac{\lambda_-}{\lambda_+}$ whereas for the scheme for P_{21}, the CFL number is $\frac{\lambda_+}{\lambda_-}$. As a consequence, this naive scheme for P is stable only if $\lambda_+ = \lambda_-$.

6.2.4 Numerical experiments

For numerical experiments, we consider three tests illustrating why such a scheme is not appropriate to get a finite time stabilization. In the three cases, we consider the same initial data :

$$\begin{cases} R_j^0 = & -4\sin(50\frac{j}{N}) \\ S_j^0 = & 2 \times \mathbb{1}_{\frac{j}{N}<0.5} - \cos(50\frac{j}{N}). \end{cases}$$

1. For $M_{12} = 2, M_{21} = -2$, $dt = 0.002$, $\lambda_+ = \lambda_- = 1$ ($T_{\min} = 2$), $dx = d\xi = 0.0022$ ($N = 450$), the energy dynamics is shown in Figure 6.3 :

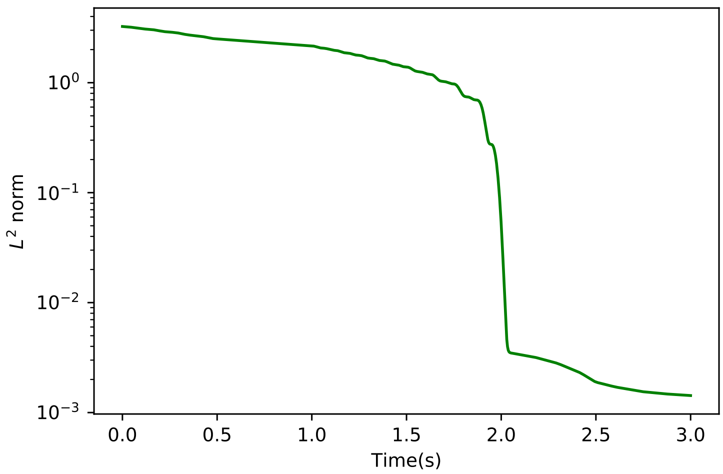
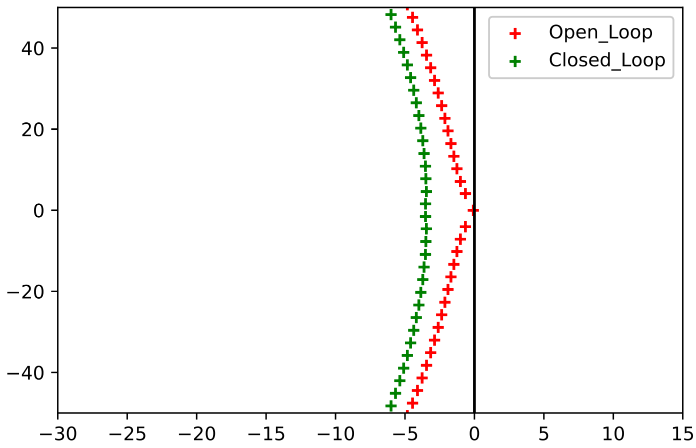


FIGURE 6.3 – The L^2 norm of the solution for case 1

The spectra of the closed-loop and open-loop operators are displayed in Figure 6.4 :

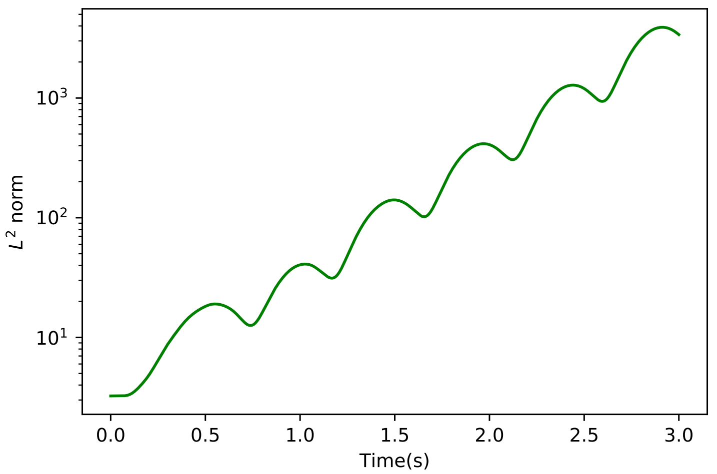


.

FIGURE 6.4 – Spectrums of discretized operators for case 1

Here the action of the control is quite satisfying since in Figure 6.3, one can clearly see that just after the extinction time $T_{\min}$, there is an abrupt decrease of the solution's energy. From a spectral point of view, finite time stabilization is less clear since we have modes of real part around -4 which gives an exponential stability with rate equal to -4 only.

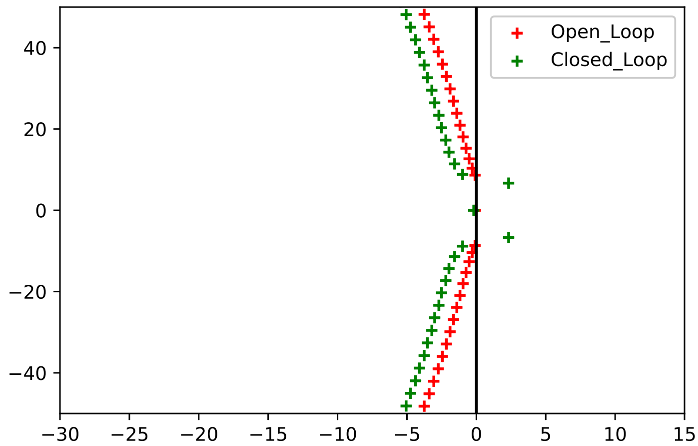
2. In fact when the coupling term M is too large, it is possible to exhibit the lack of effectiveness of the control synthesized in the previous section. As an example, if we take $M_{12} = 8, M_{21} = -8$ keeping the same discretization parameters $(dt, dx...)$, we obtain Figure 6.5 :



.

FIGURE 6.5 – The L^2 norm of the solution for case 2

In that case, the spectrum of the closed-loop and open-loop operators are shown in Figure 6.6 :



.

FIGURE 6.6 – Spectra of discretized operators for case 2

One clearly sees that the closed-loop system is unstable even if the "continuous" (*ie* without discretization) version is finite-time stable.

There is a more efficient way to compute the backstepping control and this is the object of the next section.

6.3 The numerical backstepping approach

In all this section, an initial data (R^0, S^0) is taken in $(L^\infty([0,1]))^2$. In order to remedy the problem of effectiveness of the control highlighted in case 2, we apply directly the backstepping method on the discretized open-loop system. In addition, we consider two different grids for R and S to avoid problems of CFL pinpointed in Remark 33.

6.3.1 The scheme

Let N, α be integers. In this section, we consider two different space grids (see Figure 6.7), a coarse grid with N cells, and a fine grid obtained by dividing each coarse cell into α finer cells. Using the notation $dx_+ := \frac{1}{\alpha N}, dx_- := \frac{1}{N}$ and introducing the time step $dt > 0$, one defines the space-time grid by :

$$\begin{cases} \forall 1 \leq i_c \leq N, & x^c_{i_c} := (i_c - 1/2)dx_- \\ \forall 1 \leq i_f \leq \alpha N, & x^f_{i_f} := (i_f - 1/2)dx_+ \\ \forall n \geq 0, & t^n := ndt. \end{cases}$$

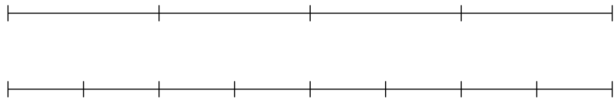


FIGURE 6.7 – The space grids for $\alpha = 2$

Moreover, the time step $dt > 0$ is given such that the following CFL conditions are satisfied :

$$\left\{ \begin{array}{lcl} \nu_+ := & \frac{\lambda_+ dt}{dx_+} \leq 1, \\ \nu_- := & \frac{\lambda_- dt}{dx_-} \leq 1. \end{array} \right. \tag{6.10}$$

The following definition will be useful to pass from the coarse grid to the the finer one and vice-versa.

Definition 19. *For all* $1 \leq i_c \leq N$,

$$N_f(i_c) := \left\{ 1 \leq i_f \leq \alpha N \mid x^f_{i_f} \in [x^c_{i_c} - dx_-/2, x^c_{i_c} + dx_-/2] \right\}.$$

Moreover, $1 \leq N_c(i_f) \leq N$ *is the unique index such that* $[x^f_{i_f} - dx_+/2, x^f_{i_f} + dx_+/2] \subset [x_{N_c(i_f)} - dx_-/2, x_{N_c(i_f)} + dx_-/2]$.

The numerical approximation $(R^n)_n \in (\mathbb{R}^{\alpha N})^{\mathbb{N}}$ is piecewise constant on cells of the form $]x^f_{i_f} - dx_+/2, x^f_{i_f} + dx_+/2[$ whereas $(S^n)_n \in (\mathbb{R}^N)^{\mathbb{N}}$ is piecewise constant on cells $]x^c_{i_c} - dx_-/2, x^c_{i_c} + dx_-/2[$. It is computed as follows :

— The initial data is given by $R^0_i = \dfrac{1}{dx_+} \displaystyle\int_{x^f_{i_f} - dx_+/2}^{x^f_{i_f} + dx_+/2} R^0(x)dx$ and $S^0_i = \dfrac{1}{dx_-} \displaystyle\int_{x^c_{i_c} - dx_-/2}^{x^c_{i_c} + dx_-/2} S^0(x)dx$.

— If we assume that (R^n, S^n) is given at time t^n, we compute (R^{n+1}, S^{n+1}) with :

$$\left\{ \begin{array}{lcl} R^{n+1} & = & R^n + dt(-\lambda_+ \partial^+_x R^n + M_{12} \Pi_{f \leftarrow c} S^n + B u^n) \\ S^{n+1} & = & S^n + dt(\lambda_- \partial^-_x S^n + M_{21} \Pi_{c \leftarrow f} R^n + B_2 R^n) \end{array} \right. \tag{6.11}$$

where operators are defined below.

(i) The positive transport operator $\partial^+_x \in M_{\alpha N, \alpha N}(\mathbb{R})$ is :

$$\partial^+_x := \begin{pmatrix} 1/dx_+ & 0 & \cdots & \cdots & 0 \\ -1/dx_+ & 1/dx_+ & \cdots & \cdots & \vdots \\ 0 & -1/dx_+ & 1/dx_+ & \cdots & \vdots \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & -1/dx_+ & 1/dx_+ \end{pmatrix},$$

whereas the negative one $\partial^-_x \in M_{N,N}(\mathbb{R})$ is :

$$\partial^-_x := \begin{pmatrix} -1/dx_- & 1/dx_- & \cdots & \cdots & 0 \\ 0 & -1/dx_- & 1/dx_- & \cdots & \vdots \\ 0 & 0 & -1/dx_- & 1/dx_- & \vdots \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & -1/dx_- \end{pmatrix}.$$

(ii) The projection from the coarse grid towards the fine one $\Pi_{f\leftarrow c} \in M_{\alpha N,N}(\mathbb{R})$ is introduced here. To define its action, we take a coarse cell indexed by $1 \le i_c \le N$ and a coarse vector $S \in \mathbb{R}^N$:

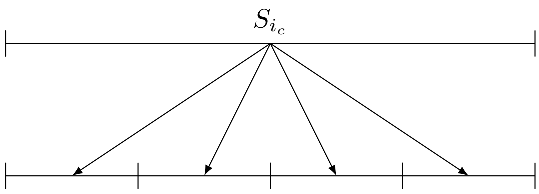


FIGURE 6.8 – The definition of $\Pi_{f\leftarrow c}$

The fine vector $\Pi_{f\leftarrow c}S \in \mathbb{R}^{\alpha N}$ is constructed by copying the value of S in the coarse cell into the associated fine cells *ie* :

$$\forall j_f \in N_f(i_c), (\Pi_{f\leftarrow c}S)_{j_f} = S_{i_c}.$$

For the projection from the fine grid towards the coarse one, the operator $\Pi_{c\leftarrow f} \in M_{N,\alpha N}(\mathbb{R})$ is constructed here. To define its action, we take a fine vector R and a coarse cell indexed by $1 \le i_c \le N$

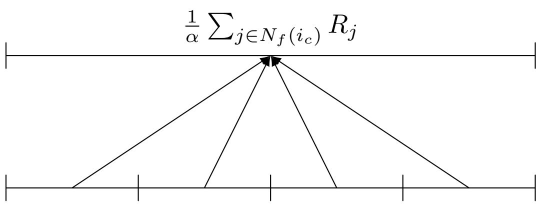


FIGURE 6.9 – The definition of $\Pi_{c\leftarrow f}$

The coarse value $(\Pi_{c\leftarrow f}R)_{i_c}$ is computed using the arithmetic mean of the values of R in the fine cells corresponding to the neighborhood of the coarse cell i_c. Obviously, we have the following dual property :

$$\langle R, \Pi_{f\leftarrow c}S\rangle_f dx_- = \langle \Pi_{c\leftarrow f}R, S\rangle_c dx_+ \tag{6.12}$$

where $\langle\cdot,\cdot\rangle_f, \langle\cdot,\cdot\rangle_c$ are the respective canonical scalar products in $\mathbb{R}^{\alpha N}, \mathbb{R}^N$.
At a matrix level, this is equivalent to :

$$\Pi_{f\leftarrow c} = \alpha \Pi^T_{c\leftarrow f}. \tag{6.13}$$

(iii) The discretized boundary control operator $B \in \mathbb{R}^{\alpha N}$ is given below :

$$B := \begin{pmatrix} \lambda_+/dx_+ \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

(iv) The boundary transfer operator $B_2 \in M_{N,\alpha N}(\mathbb{R})$ is given by :

$$B_{2,N,\alpha N} := \frac{h\lambda_-}{dx_-}$$

and all other coefficients of B_2 are set to zero.

6.3.2 The numerical backstepping method

We apply the backstepping method to the discretized system (6.11). We look for a backstepping transformation $\mathcal{T}$ as a Volterra transform of the second kind, of the following form :

$$\begin{pmatrix} R^\star \\ S^\star \end{pmatrix} = \mathcal{T}\begin{pmatrix} R \\ S \end{pmatrix} \iff \left\{ \begin{array}{lcl} R^\star & = & R - P_{11}Rdx_+ - P_{12}Sdx_- \\ S^\star & = & S \end{array} \right.$$

The structure of P_{11}, P_{12} is upper triangular in a sense that will be defined later.

The system verified by $R^{\star,n} := R^n - P_{11}R^n dx_+ - P_{12}S^n dx_-$ is calculated below :

$$\begin{array}{rcl} \dfrac{R^{\star n+1} - R^{\star n}}{dt} & = & \dfrac{R^{n+1} - R^n}{dt} - P_{11}(-\lambda_+\partial_x^+ R^n + M_{12}\Pi_{f\leftarrow c}S^n + Bu^n)dx_+ \\ & & -P_{12}(\lambda_-\partial_x^- S^n + M_{21}\Pi_{c\leftarrow f}R^n + B_2R^n)dx_- \end{array}$$

Moreover,

$$\lambda_+\partial_x^+ R^{\star n} = \lambda_+\partial_x^+ R^n - \lambda_+\partial_x^+(P_{11}R^n dx_+ + P_{12}S^n dx_-).$$

Thus, the equation for $R^\star$ reads :

$$\begin{array}{rcl} \dfrac{R^{\star n+1} - R^{\star n}}{dt} + \lambda_+\partial_x^+ R^{\star n} & = & M_{12}\Pi_{f\leftarrow c}S^n + Bu^n - \lambda_+\partial_x^+(P_{11}R^n dx_+ + P_{12}S^n dx_-) \\ & & -P_{11}(-\lambda_+\partial_x^+ R^n + M_{12}\Pi_{f\leftarrow c}S^n + Bu^n)dx_+ \\ & & -P_{12}(\lambda_-\partial_x^- S^n + M_{21}\Pi_{c\leftarrow f}R^n + B_2R^n)dx_- \\ & = & -\Gamma_{11}R^n - \Gamma_{12}S^n + (B - P_{11}Bdx_+)u^n, \end{array} \tag{6.14}$$

where the terms Γ_{11}, Γ_{12} are given below :

$$\begin{array}{lcl} \Gamma_{11} & := & \lambda_+\partial_x^+ P_{11}dx_+ - \lambda_+P_{11}\partial_x^+ dx_+ + P_{12}(M_{21}\Pi_{c\leftarrow f} + B_2)dx_-, \\ \Gamma_{12} & := & \lambda_+\partial_x^+ P_{12}dx_- + \lambda_-P_{12}\partial_x^- dx_- + M_{12}P_{11}\Pi_{f\leftarrow c}dx_+ - M_{12}\Pi_{f\leftarrow c}. \end{array} \tag{6.15}$$

In what follows, we give a scheme that construct a P such that most of the terms of Γ are equal to zero. To do so, we see the column index of P (the "x" variable) as a time variable whereas row's index is seen as the space variable (the "ξ" variable). The scheme is given below :

— We begin by setting the last lines of P_{11}, P_{12} :

$$\begin{array}{ll} \forall 1 \le j_f \le \alpha N, & P_{11,\alpha N,j_f} = 0 \\ \forall 1 \le j_c \le N, & P_{12,\alpha N,j_c} = 0 \end{array} \tag{6.16}$$

— Suppose that P_{12}, P_{11} is given at rows $i_f, \cdots, \alpha N$, then we calculate P_{12,i_f-1,j_c} by imposing $\Gamma_{12,i_f,j_c} = 0$ for $j_c \ge N_c(i_f - 1)$. From (6.15), we deduce :

$$\lambda_+ \frac{P_{12,i_f,j_c} - P_{12,i_f-1,j_c}}{dx_+} dx_- - \lambda_- \frac{P_{12,i_f,j_c} - P_{12,i_f,j_c-1}}{dx_-} dx_- \\ + M_{12}[P_{11}\Pi_{f\leftarrow c}]_{i_f,j_c} dx_+ - M_{12}[\Pi_{f\leftarrow c}]_{i_f,j_c} = 0. \quad (6.17)$$

It is easy to compute $[P_{11}\Pi_{f\leftarrow c}]_{i_f,j_c}$ using local indexes. Using (6.13), one gets :

$$[P_{11}\Pi_{f\leftarrow c}]_{i_f,j_c} = [\Pi^T_{f\leftarrow c} P^T_{11}]_{j_c,i_f} = \alpha[\Pi_{c\leftarrow f} P^T_{11}]_{j_c,i_f} = \sum_{i\in N_f(j_c)} P_{11,i_f,i}.$$

Then, we compute P_{11,i_f-1,j_f} imposing $\Gamma_{11,i_f,j_f} = 0$ for $j_f \geq i_f - 1$:

$$\lambda_+ \frac{P_{11,i_f,j_f} - P_{11,i_f-1,j_f}}{dx_+} dx_+ - \lambda_+ \frac{P_{11,i_f,j_f} - P_{11,i_f,j_f+1}}{dx_+} dx_+ \\ + M_{21}[P_{12}\Pi_{c\leftarrow f}]_{i_f,j_f} dx_- + [P_{12}B_2]_{i_f,j_f} dx_- = 0. \quad (6.18)$$

It is easy to calculate $[P_{12}\Pi_{c\leftarrow f}]_{i_f,j_f}$ using local indexes. Using (6.13), one obtains :

$$[P_{12}\Pi_{c\leftarrow f}]_{i_f,j_f} = [\Pi^T_{c\leftarrow f} P^T_{12}]_{j_f,i_f} = \frac{1}{\alpha}[\Pi_{f\leftarrow c} P^T_{12}]_{j_f,i_f} = \frac{1}{\alpha} P_{12,i_f,N_c(j_f)}.$$

For $[P_{12}B_2]_{i_f,j_f}$, we also have :

$$[P_{12}B_2]_{i_f,j_f} = \frac{h\lambda_-}{dx_-} P_{12,i_f,N}\delta_{j_f=\alpha N}.$$

Algorithm 1 Calculate P_{12}

$P_{12,i_f,j_c} \leftarrow 0$ for all $1 \leq i_f \leq \alpha N, 1 \leq j_c \leq N$.
for $2 \leq i_f \leq \alpha N$ (step = -1) **do**
  for $N_c(i_f - 1) \leq j_c \leq N$ **do**
    if $j_c = N_c(i_f)$ **then**
      $P_{12,i_f-1,j_c} \leftarrow -\frac{M_{12}}{\alpha\lambda_+}$
    end if
    $P_{12,i_f-1,j_c} \leftarrow P_{12,i_f-1,j_c} + (1 - \frac{\lambda_-}{\lambda_+}\frac{dx_+}{dx_-})P_{12,i_f,j_c} + \frac{\lambda_-}{\lambda_+}\frac{dx_+}{dx_-}P_{12,i_f,j_c-1}$ (transport)
    $P_{12,i_f-1,j_c} \leftarrow P_{12,i_f-1,j_c} + \frac{M_{12}}{\alpha\lambda_+}\sum_{j_f\in N_f(j_c)} P_{11,i_f,j_f} dx_+$ (exchange)
  end for
end for

Algorithm 2 Calculate P_{11}

$P_{11,i_f,j_f} \leftarrow 0$ for all $1 \leq i_f \leq \alpha N, 1 \leq j_f \leq \alpha N$.
for $2 \leq i_f \leq \alpha N$ (step = -1) **do**
  $P_{11,i_f-1,\alpha N} \leftarrow \frac{h\lambda_-}{\lambda_+} P_{12,i_f,N}$
  for $i_f - 1 \leq j_f \leq \alpha N$ **do**
    if $j_f \neq \alpha N$ **then**
      $P_{11,i_f-1,j_f} \leftarrow P_{11,i_f-1,j_f} + P_{11,i_f,j_f+1}$ (transport)
    end if
    $P_{11,i_f-1,j_f} \leftarrow P_{11,i_f-1,j_f} + M_{21} P_{12,i_f,N_c(j_f)} \frac{dx_-}{\alpha\lambda_+}$ (exchange)
  end for
end for

Remark 34. The algorithms immediately give that :

$$\begin{cases} \forall j_f < i_f, & P_{11,i_f,j_f} = 0 \\ \forall j_c < N_c(i_f), & P_{12,i_f,j_c} = 0. \end{cases} \tag{6.19}$$

6.3.3 Convergence properties of the kernel

In what follows and until the end of the chapter, C designates a constant independent on the discretization. Here we present a boundedness result for P which will be useful later :

Proposition 5. *If $\frac{\lambda^-}{\lambda^+}\frac{dx_+}{dx_-} \leq 1$, there exists $C_\infty > 0$ independent on N such that :*

$$\|P_{11}\|_{L^\infty([0,1]^2)} + \|P_{12}\|_{L^\infty([0,1]^2)} \leq C_\infty. \tag{6.20}$$

Remark 35. The condition :

$$\frac{\lambda^-}{\lambda^+}\frac{dx_+}{dx_-} \leq 1$$

can be interpreted as a CFL condition. It imposes an asymmetry of grids in the sense that if $\frac{\lambda_-}{\lambda_+} > 1$, the grid for R (where the control is applied) needs to be finer than the one for S.

Proof. The scheme verified by P_{12} gives :

$$\begin{aligned} \forall j_c > N_c(i_f), |P_{12,i_f-1,j_c}| \leq \quad & (1 - \tfrac{\lambda_-}{\lambda_+}\tfrac{dx_+}{dx_-})|P_{12,i_f,j_c}| + \tfrac{\lambda_-}{\lambda_+}\tfrac{dx_+}{dx_-}|P_{12,i_f,j_c-1}| \\ & + \frac{|M_{12}|}{\alpha\lambda_+} \sum_{j\in N_f(j_c)} |P_{11,i_f,j}| dx_+. \end{aligned} \tag{6.21}$$

For $j_c \leq N_c(i_f) - 1$, we have :

$$|P_{12,i_f-1,j_c}| = \quad 0, \tag{6.22}$$

whereas for $j_c = N_c(i_f)$, it holds :

$$|P_{12,i_f-1,j_c}| \leq \quad \left(1 - \frac{\lambda_-}{\lambda_+}\frac{dx_+}{dx_-}\right)|P_{12,i_f,j_c}| + \left|\frac{M_{12}}{\lambda_+\alpha}\right| + \frac{|M_{12}|}{\lambda_+\alpha}\sum_{j\in N_f(j_c)} |P_{11,i_f,j}| dx_+. \tag{6.23}$$

Combining (6.21)-(6.23), it holds :

$$\|P_{12,i_f-1,\cdot}\|_{L^\infty([0,1])} = \quad \max\left\{\|P_{12,i_f,\cdot}\|_{L^\infty([0,1])}, \left|\frac{M_{12}}{\lambda_-}\right|\right\} + C\|P_{11,i_f,\cdot}\|_{L^\infty([0,1])}dx_+.$$

Similarly, it holds :

$$\|P_{11,i_f-1,\cdot}\|_{L^\infty([0,1])} \leq \max\left\{\|P_{11,i_f,\cdot}\|_{L^\infty([0,1])}, \frac{h\lambda_-}{\lambda_+}\|P_{12,i_f,\cdot}\|_{L^\infty([0,1])}\right\} + C\|P_{12,i_f,\cdot}\|_{L^\infty([0,1])}dx_-.$$

Denoting
$A_{i_f} := \max\{\frac{h\lambda_-}{\lambda_+}\|P_{12,i_f,\cdot}\|_{L^\infty([0,1])}, \|P_{11,i_f,\cdot}\|_{L^\infty([0,1])}, \left|\frac{hM_{12}}{\lambda_+}\right|\}$, it holds :

$$A_{i_f-1} \leq (1 + Cdx_-)A_{i_f}$$

and hence :

$$\forall 1 \leq i_f \leq \alpha N,\ A_{i_f} \leq e^{i_f C dx_-} \leq e^{C\alpha}.$$

□

Finally, in order to exhibit the exact target system, we need to see which term of Γ is zero :

Lemma 28. *The matrix Γ_{12} is zero except for the first row. Moreover, the term Γ_{11,i_f,j_f} is non zero only for $j_f \in [\![i_f - \alpha, i_f - 2]\!]$ and $i_f = 1$. Hence, Γ_{11} can be decomposed as $\Gamma_{11} =: \tilde{\Gamma}_{11} + \bar{\Gamma}_{11}$ where $\tilde{\Gamma}_{11}$ is the subdiagonal part of Γ_{11} and $\bar{\Gamma}_{11} = \Gamma_{11} - \tilde{\Gamma}_{11}$ is non zero only on its first row. Furthermore,*

$$\sup_{i_f, j_f} |\tilde{\Gamma}_{11,i_f,j_f}| \leq Cdx_-.$$

with $C > 0$ is independent on the discretization.

In order to see which entry is zero, we represent Γ_{11}, Γ_{12} for $\alpha = 3$ putting a cross on each non zero entry in Figures 6.10-6.11 :

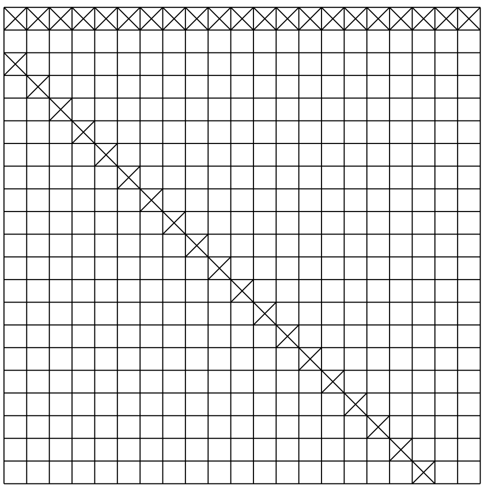


FIGURE 6.10 – The non zero coefficients for Γ_{11} ($\alpha = 3$)

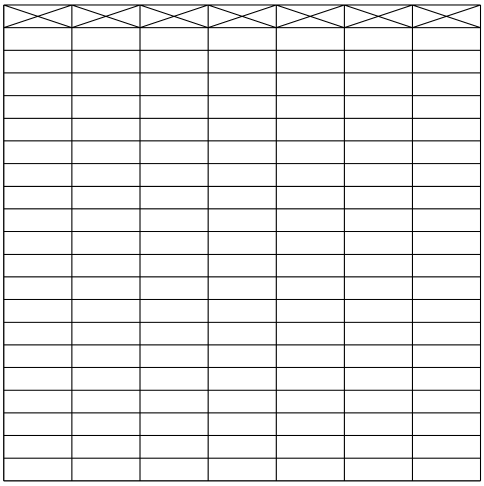


FIGURE 6.11 – The non zero coefficients for Γ_{12} ($\alpha = 3$)

Proof. Let $i_f \geq 2$. For $j_c \geq N_c(i_f - 1)$ and by Algorithm 1, we have $\Gamma_{12,i_f,j_c} = 0$. For $j_c \leq N_c(i_f) - 2$, we have :

1. $j_c < N_c(i_f - 1)$ and $P_{12,i_f-1,j_c} = 0$.
2. $j_c < N_c(i_f)$ and $P_{12,i_f,j_c} = 0$.
3. For all $j_f \in N_f(j_c)$, we have $j_f < i_f$ (because $j_c \leq N_c(i_f) - 2$ by assumption) and hence $P_{11,i_f,j_f} = 0$.

By the definition of Γ_{12}, $\Gamma_{12,i_f,j_c} = 0$.

Now for $j_c = N_c(i_f) - 1$, then points 2 and 3 of the previous case hold. If $j_c = N_c(i_f - 1)$, then by Algorithm 1, $\Gamma_{12,i_f,j_c} = 0$. Otherwise, $j_c < N_c(i_f - 1)$ implying that $P_{12,i_f-1,j_c} = 0$. Thus, by the definition of Γ_{12,i_f,j_c}, one gets $\Gamma_{12,i_f,j_c} = 0$.

For Γ_{11,i_f,j_f}, the only terms that could be non zero are terms such that $i_f - \alpha + 1 \leq j_f < i_f - 1$ and by the definition of Γ_{11} :

$$|\Gamma_{11,i_f,j_f}| \leq C \|P_{12}\|_{L^\infty([0,1]^2)} dx_- \leq C dx_-$$

where we have used Proposition 5.

□

Remark 36. If $\alpha \leq 2$ then :

$$\sup_{i_f > 1, j_f} |\tilde{\Gamma}_{11,i_f,j_f}| = 0.$$

Indeed, the condition $i_f - \alpha + 1 \leq j_f < i_f - 1 \iff i_f - \alpha + 1 \leq j_f \leq i_f - 2$ is empty.

From the computations of the kernel P, we obtained that all the entries of Γ_{11}, Γ_{12} are zero except for the first row and some diagonal terms in Γ_{11}. By (6.14), in order to get rid of the contribution of these first rows, we impose a control u^n of the form :

$$u^n = \frac{1}{\lambda_+ - [P_{11}B]_{11} dx_+^2} \Big(\sum_{j_f=1}^{\alpha N} \Gamma_{11,1,j_f} R^n_{j_f} dx_+ + \sum_{j_c=1}^{N} \Gamma_{12,1,j_c} S^n_{j_c} dx_+ \Big) \tag{6.24}$$

and thus, the final target system is :

$$\left\{\begin{array}{lcl}\dfrac{R^{\star n+1}-R^{\star n}}{dt}+\lambda_+\partial_x^+R^{\star n} & = & \tilde{\Gamma}_{11}R^n\\ \dfrac{S^{\star n+1}-S^{\star n}}{dt}-\lambda_-\partial_x^-S^{\star n} & = & B_2R^n+M_{21}\Pi_{c\leftarrow f}R^n\end{array}\right. \tag{6.25}$$

where $\tilde{\Gamma}_{11}$ is the subdiagonal part of Γ_{11} (see Figure 6.10).

Next proposition asserts that the discrete Voltera transform of the second kind is invertible and continuous uniformly with respect to the parameter of discretization. This will be useful to express the target system uniquely in terms of star variables. Before presenting the result, it is needed to introduce the L^p $(1 < p < \infty)$ norm for vectors $R \in \mathbb{R}^{\alpha N}, S \in \mathbb{R}^N$:

$$\|R\|_{L^p([0,1])} = \sqrt[p]{\sum_{j_f=1}^{\alpha N}|R_{j_f}|^p dx_+} \text{ and } \|S\|_{L^p([0,1])} = \sqrt[p]{\sum_{j_c=1}^{N}|S_{j_c}|^p dx_-}.$$

Proposition 6. *Let $1 < p \leq \infty$. There exists $C > 0$ independent on the discretization such that if $dx_- < C$, then the operator $\mathcal{T}$ is invertible as an operator from $L^p([0,1])$ into $L^p([0,1])$. Moreover, there exists a $C_{back,p} > 0$ independent on the discretization such that for all $R \in \mathbb{R}^{\alpha N}, S \in \mathbb{R}^N$:*

$$\frac{1}{C_{back,p}}\|\mathcal{T}(R,S)\|_{L^p([0,1])} \leq \|(R,S)\|_{L^p([0,1])} \leq C_{back,p}\|\mathcal{T}(R,S)\|_{L^p([0,1])}. \tag{6.26}$$

Moreover, there exists $L_{11} \in M_{\alpha N,\alpha N}(\mathbb{R}), L_{12} \in M_{\alpha N,N}(\mathbb{R})$ such that :

$$\forall R^\star \in \mathbb{R}^{\alpha N}, S^\star \in \mathbb{R}^N,\ \mathcal{T}^{-1}\begin{pmatrix}R^\star\\ S^\star\end{pmatrix} = \begin{pmatrix}R^\star + L_{11}R^\star dx_+ + L_{12}S^\star dx_-\\ S^\star\end{pmatrix}.$$

where (L_{11}, L_{12}) is upper triangular in the sense of (6.19).

Proof. The first inequality is easy to prove owing Proposition 5 (the proof is left to the reader).

For the second one, we use a fixed point argument. Let $(R^k)_k$ be the sequence such that $R^0 = 0_{\mathbb{R}^{\alpha N}}$ and :

$$\forall k \geq 0,\ R^{k+1} \quad = \quad R^\star + P_{11}R^k dx_+ + P_{12}S^\star dx_-.$$

We estimate $R^{k+1} - R^k$ in a weighted space. More precisely, let $\gamma > 0$ and for all $1 \leq i_f \leq \alpha N$

$$\begin{aligned}(R^{k+1}_{i_f} - R^k_{i_f})e^{\gamma x^f_{i_f}} &= \sum_{j_f} P_{11,i_f,j_f}(R^k_{j_f} - R^{k-1}_{j_f})e^{\gamma x^f_{i_f}}dx_+\\ &= \sum_{j_f} P_{11,i_f,j_f}e^{-\gamma(x^f_{j_f}-x^f_{i_f})}(R^k_{j_f} - R^{k-1}_{j_f})e^{\gamma x^f_{j_f}}dx_+.\end{aligned}$$

For the case $p = \infty$, we have :

$$\|R^{k+1} - R^k\|_{L^\infty_\gamma([0,1])} \quad \leq \quad \sup_{i_f}\sum_{j_f}|P_{11,i_f,j_f}|e^{-\gamma(x^f_{j_f}-x^f_{i_f})}dx_+\|R^k - R^{k-1}\|_{L^\infty_\gamma([0,1])}. \tag{6.27}$$

where L^∞_γ is the weighted L^∞ norm :

$$\forall R \in \mathbb{R}^{\alpha N},\ \|R\|_{L^\infty_\gamma([0,1])} := \max|R_{i_f}|e^{\gamma x^f_{i_f}}.$$

Because of the uniform estimate from Proposition 5 and the upper triangular structure (6.19), one gets :

$$\begin{array}{rcl}\sup_{i_f}\sum_{j_f}|P_{11,i_f,j_f}|e^{-\gamma(x^f_{j_f}-x^f_{i_f})}dx_+ \leq & C_\infty \sup_{i_f}\sum_{j_f\geq i_f} e^{-\gamma(x^f_{j_f}-x^f_{j_f})}dx_+ \\ = & C_\infty \sup_{i_f}\sum_{j_f\geq i_f} e^{-\gamma(x^f_{j_f}-x^f_{i_f})}dx_+ \\ \leq & C_\infty \frac{dx_+}{1-e^{-\gamma dx_+}} = \frac{C_\infty}{\gamma}\frac{\gamma dx_+}{1-e^{-\gamma dx_+}}.\end{array} \tag{6.28}$$

Using the boundedness around 0 of the function $x \mapsto \frac{x}{1-e^{-x}}$ and taking $dx_+ \leq C/\gamma$:

$$\sup_{i_f}\sum_{j_f}|P_{11,i_f,j_f}|e^{-\gamma(x^f_{j_f}-x^f_{i_f})}dx_+ \leq \frac{2C_\infty}{\gamma}.$$

With $4C_\infty \leq \gamma$ and $dx_+ \leq C/\gamma$ (with $dx_+ < \frac{C}{4C_\infty}$) :

$$\forall n > 0, \|R^{k+1}-R^k\|_{L^\infty_\gamma([0,1])} \leq \frac{1}{2}\|R^k - R^{k-1}\|_{L^\infty_\gamma([0,1])}.$$

For the case $1 < p < \infty$, the conjugate exponent is $q := \frac{p}{p-1}$ and :

$$\|R^{k+1}-R^k\|_{L^p_\gamma([0,1])} \leq \sqrt[p]{\sum_{i_f}\left(\sum_{j_f}|P_{11,i_f,j_f}|^q e^{-\gamma q(x^f_{j_f}-x^f_{i_f})}dx_+\right)^{\frac{p}{q}}dx_+}\|R^k-R^{k-1}\|_{L^p_\gamma([0,1])}$$

where :

$$\forall R \in \mathbb{R}^{\alpha N},\ \|R\|_{L^p_\gamma([0,1])} := \sqrt[p]{\sum_{j_f}|R_{j_f}|^p e^{p\gamma x^f_{i_f}}}.$$

Using Proposition 5, (6.19) and for $dx_+ \leq C/\gamma$, we get by similar computations as in (6.28), that :

$$\sqrt[p]{\sum_{i_f}\left(\sum_{j_f}|P_{11,i_f,j_f}|^q e^{-\gamma q(x^f_{j_f}-x^f_{i_f})}dx_+\right)^{\frac{p}{q}}dx_+} < C_\infty\left(\frac{2}{q}\right)^{1/q}\gamma^{-1/q}$$

and for $\frac{2^{q+1}C^q_\infty}{q} \leq \gamma$ and $dx_+ \leq C/\gamma$:

$$\forall n > 0, \|R^{k+1}-R^k\|_{L^p_\gamma([0,1])} \leq \frac{1}{2}\|R^n - R^{k-1}\|_{L^p_\gamma([0,1])}.$$

As a consequence, for $1 < p \leq \infty$ the series $\sum_{k=0}^\infty R^{k+1}-R^k$ is convergent and the sequence $(R^k)_k$ converges in $L^p([0,1])$ towards a limit $R \in L^p([0,1])$. Moreover, we have for γ large enough :

$$\begin{array}{rcl}\|R\|_{L^p_\gamma([0,1])} & \leq & \sum_{k=0}^\infty \|R^{k+1}-R^k\|_{L^p_\gamma([0,1])} \\ & \leq & \left(\sum_{k=0}^\infty \frac{1}{2^k}\right)\|R^1\|_{L^p_\gamma([0,1])} \\ & \leq & \left(\sum_{k=0}^\infty \frac{1}{2^k}\right)(\|R^\star\|_{L^p_\gamma([0,1])} + \|S^\star\|_{L^p_\gamma([0,1])}).\end{array}$$

Using the fact that $\|\cdot\|_{L^p([0,1])} \leq \|\cdot\|_{L^p_\gamma([0,1])} \leq e^\gamma \|\cdot\|_{L^p([0,1])}$, one finally gets :

$$\|R\|_{L^p([0,1])} \leq \left(\sum_{k=0}^{\infty} \frac{1}{2^k}\right) e^\gamma (\|R^\star\|_{L^p([0,1])} + \|S^\star\|_{L^p([0,1])}).$$

This proves (6.26) taking $\gamma = \frac{2^{q+1}C_\infty^q}{q}$ which does not depend on the discretization.

To finish the proof of the proposition, it suffices to write that :

$$\begin{pmatrix} R \\ S \end{pmatrix} = \begin{pmatrix} R^\star \\ S^\star \end{pmatrix} + \begin{pmatrix} P_{11}dx_+ & P_{12}dx_- \\ 0 & 0 \end{pmatrix} \begin{pmatrix} R \\ S \end{pmatrix}$$

and by induction, it holds :

$$\begin{pmatrix} R \\ S \end{pmatrix} = \begin{pmatrix} R^\star \\ S^\star \end{pmatrix} + \sum_{k=1}^{\infty} \begin{pmatrix} P_{11}dx_+ & P_{12}dx_- \\ 0 & 0 \end{pmatrix}^k \begin{pmatrix} R^\star \\ S^\star \end{pmatrix}.$$

Owing the strict upper triangular structure of $P_{\bullet,\bullet}$, we get that :

$$\begin{pmatrix} L_{11}dx_+ & L_{12}dx_- \\ 0 & 0 \end{pmatrix} := \sum_{k=1}^{\infty} \begin{pmatrix} P_{11}dx_+ & P_{12}dx_- \\ 0 & 0 \end{pmatrix}^k$$

with L_{11} strict upper triangular and L_{12} verifying (6.19) (easy to prove by induction).

□

6.3.4 Proof of the finite time stabilization result

Owing Proposition 6, it is possible to write the target system (6.25) as :

$$\begin{cases} \dfrac{R^{\star n+1} - R^{\star n}}{dt} + \lambda_+ \partial_x^+ R^{\star n} &= \tilde{\Gamma}_{11} R^n \\ \dfrac{S^{\star n+1} - S^{\star n}}{dt} - \lambda_- \partial_x^- S^{\star n} &= (B_2 + M_{21}\Pi_{c\leftarrow f})(R^{\star n} + L_{11}R^{\star n}dx_+ + L_{12}S^{\star n}dx_-). \end{cases} \tag{6.29}$$

Then using Lemma 28, next lemma shows that the right hand side of the equation for $R^\star$ in (6.29) can be neglected.

Lemma 29. *For all $T > 0$ and $1 < p < \infty$, there exists $C > 0$ independent on the parameters of discretization such that :*

$$\forall n : ndt \leq T, \|R^{\star n} - \tilde{R}^{\star n}\|^p_{L^p([0,1])} + \|S^{\star n} - \tilde{S}^{\star n}\|^p_{L^p([0,1])} \leq C dx_-^p \sum_{m=0}^{n} (\|R^{\star m}\|^p_{L^p([0,1])} + \|S^{\star m}\|^p_{L^p([0,1])}) dt.$$

where $\tilde{R}^\star, \tilde{S}^\star$ satisfies the same system as (6.29) without the term Γ_{11}, that is

$$\begin{cases} \dfrac{\tilde{R}^{\star n+1} - \tilde{R}^{\star n}}{dt} + \lambda_+ \partial_x^+ \tilde{R}^\star_n = & 0 \\ \dfrac{\tilde{S}^{\star n+1} - \tilde{S}^{\star n}}{dt} - \lambda_- \partial_x^- \tilde{S}^{\star n} = & (B_2 + M_{21}\Pi_{c\leftarrow f})(\tilde{R}^{\star n} + L_{11}\tilde{R}^{\star n}dx_+ + L_{12}\tilde{S}^{\star n}dx_-), \end{cases} \tag{6.30}$$

with initial data :

$$\begin{cases} \tilde{R}^{\star 0} = & R^{\star 0} \\ \tilde{S}^{\star 0} = & S^{\star 0}. \end{cases}$$

Before going into the proof of Lemma 29, we will prove the following preliminary result :

Lemma 30. *For all $T > 0$, there exists $C > 0$ independent on the discretization such that for every $(f^n)_n \in (\mathbb{R}^{\alpha N})^{\mathbb{N}}$ and every $(\tilde{R}^{\star n}, \tilde{S}^{\star n})_n$ satisfying :*

$$\begin{cases} \dfrac{\tilde{R}^{\star n+1} - \tilde{R}^{\star n}}{dt} + \lambda_+ \partial_x^+ \tilde{R}^{\star n} = & f^n \\ \dfrac{\tilde{S}^{\star n+1} - \tilde{S}^{\star n}}{dt} - \lambda_- \partial_x^- \tilde{S}^{\star n} = & (B_2 + M_{21}\Pi_{c\leftarrow f})(\tilde{R}^{\star n} + L_{11}\tilde{R}^{\star n} dx_+ + L_{12}\tilde{S}^{\star n} dx_-), \end{cases} \tag{6.31}$$

then :

$$\forall n : ndt \leq T, \|\tilde{R}^{\star n}\|^p_{L^p([0,1])} + \|\tilde{S}^{\star n}\|^p_{L^p([0,1])} \leq C \left(\|\tilde{R}^{\star 0}\|^p_{L^p([0,1])} + \|\tilde{S}^{\star 0}\|^p_{L^p([0,1])} + \sum_{k=0}^{n} \|f^n\|^p_{L^p([0,1])} dt \right).$$

Proof. As this is classical, we give only a rapid sketch of the proof here.

Let $1 < p < \infty$, then by the convexity of $x \to |x|^p$:

$$\forall j_f \geq 2,\ |\tilde{R}^{\star n+1}_{j_f}|^p \leq (1+dt)^{p-1}\big((1-\nu_+)|\tilde{R}^{\star n}_{j_f}|^p + \nu_+ |\tilde{R}^{\star n}_{j_f - 1}|^p + dt|f^n_{j_f}|^p\big).$$

Hence, summing over j_f and multiplying by dx_+, one gets :

$$\|\tilde{R}^{\star n+1}\|^p_{L^p([0,1])} \leq (1+dt)^{p-1}\big(\|\tilde{R}^{\star n}\|^p_{L^p([0,1])} + dt\|f^n\|^p_{L^p([0,1])} - \nu_+ |\tilde{R}^{\star n}_{\alpha N}|^p dx_+\big).$$

For $S^\star$, the method is similar :

$$\begin{aligned} \forall j_c \leq N-1,\ |\tilde{S}^{\star n+1}_{j_c}|^p \leq\ & (1+Cdt)^{p-1}\Big((1-\nu_-)|\tilde{S}^{\star n}_{j_c}|^p + \nu_- |\tilde{S}^{\star n}_{j_c+1}|^p \\ & + dt|M_{21}|\big(|[\Pi_{c\leftarrow f} L_{11}\tilde{R}^{\star n}]_{j_c} dx_+|^p + |[\Pi_{c\leftarrow f} L_{12}\tilde{S}^{\star n}]_{j_c} dx_-|^p \\ & + |[\Pi_{c\leftarrow f}\tilde{R}^{\star n}]_{j_c}|\big)\Big). \end{aligned}$$

Then, summing over j_c, multiplying by dx_- and owing the fact that $L_{11}dx_+, L_{12}dx_-$ are bounded as operators from L^p to L^p((Proposition 6), one gets :

$$\|\tilde{S}^{\star n+1}\|^p_{L^p([0,1])} \leq (1+Cdt)^p\big(\|\tilde{S}^{\star n}\|^p_{L^p([0,1])} + dt\|\tilde{R}^{\star n}\|^p_{L^p([0,1])} + C|\tilde{R}^{\star n}_{\alpha N}|^p dx_-\big)$$

where the last term corresponds to the boundary condition at $x = 1$.

Thus, it is possible to take an $\eta > 0$ such that $\mathcal{E}^n := \|\tilde{R}^{\star n}\|^p_{L^p([0,1])} + \eta\|\tilde{S}^{\star n}\|^p_{L^p([0,1])}$ verifies :

$$\mathcal{E}^{n+1} \leq (1+Cdt)^p(\mathcal{E}^n + \|f^n\|^p_{L^p([0,1])} dt)$$

which immediately gives the result of the Lemma by induction. □

Now we are able to prove Lemma 29 :

Proof of Lemma 29. By using Lemma 28, we have :

$$\|\tilde{\Gamma}_{11} R^n\|_{L^p([0,1])} \leq C dx_- \|R^n\|_{L^p([0,1])}.$$

By Proposition 6, one can bound $\|R^n\|_{L^p([0,1])}$ by $\|R^{\star n}\|_{L^p([0,1])} + \|S^{\star n}\|_{L^p([0,1])}$ to get :

$$\|\tilde{\Gamma}_{11} R^n\|_{L^p([0,1])} \leq C dx_-(\|R^{\star n}\|_{L^p([0,1])} + \|S^{\star n}\|_{L^p([0,1])})$$

where we may have changed the constant C.

Applying Lemma 30 to $(\tilde{R}^{\star n}-R^{\star n},\tilde{S}^{\star n}-S^{\star n})_n$, we can conclude easily. This ends the proof of Lemma 29.

□

As a consequence, the energy dynamics of $R^\star, S^\star$ can be estimated by the one of $\tilde{R}^\star, \tilde{S}^\star$. Indeed,

$$\begin{aligned}\|R^{\star n}\|^p_{L^p([0,1])}+\|S^{\star n}\|^p_{L^p([0,1])} \leq\ & C\Big(\|R^{\star n}-\tilde{R}^{\star n}\|^p_{L^p([0,1])}+\|S^{\star n}-\tilde{S}^{\star n}\|^p_{L^p([0,1])}\\ &+\|\tilde{R}^{\star n}\|^p_{L^p([0,1])}+\|\tilde{S}^{\star n}\|^p_{L^p([0,1])}\Big)\\ \leq\ & Cdx_-^p\sum_{m=0}^{n}(\|R^{\star m}\|^p_{L^p([0,1])}+\|S^{\star m}\|^p_{L^p([0,1])})dt\\ &+C\Big(\|\tilde{R}^{\star n}\|^p_{L^p([0,1])}+\|\tilde{S}^{\star n}\|^p_{L^p([0,1])}\Big)\end{aligned}$$

and thus by Grönwall inequality [67, p.1], there exists a constant C independent on the parameters of discretization such that :

$$\begin{aligned}\|R^{\star n}\|^p_{L^p([0,1])}+\|S^{\star n}\|^p_{L^p([0,1])} \leq\ & C\Big(\|\tilde{R}^{\star n}\|^p_{L^p([0,1])}+\|\tilde{S}^{\star n}\|^p_{L^p([0,1])}\Big)\\ &+Cdx_-^p\sum_{m=0}^{n}e^{Cdx_-^p(n-m)dt}(\|\tilde{R}^{\star m}\|^p_{L^p([0,1])}+\|\tilde{S}^{\star m}\|^p_{L^p([0,1])})dt.\end{aligned}$$

Hence for all $T>0$ fixed and $1<p<\infty$, there exists a constant $C>0$ such that :

$$\begin{aligned}\forall ndt\leq T,\ \|R^{\star n}\|_{L^p([0,1])}+\|S^{\star n}\|_{L^p([0,1])} \leq\ & C(\|\tilde{R}^{\star n}\|_{L^p([0,1])}+\|\tilde{S}^{\star n}\|_{L^p([0,1])}\\ &+dx_-\|R^{\star 0}\|_{L^p([0,1])}+\|S^{\star 0}\|_{L^p([0,1])})\end{aligned} \tag{6.32}$$

where we used Lemma 30 applied to $(\tilde{R}^{\star n},\tilde{S}^{\star n})_n$.

From now on, we do not make the difference between $R^\star, S^\star$ and $\tilde{R}^\star, \tilde{S}^\star$ keeping the non tilda notation. To prove a finite time stabilization result, the first step is to prove an extinction result for $R^\star$.

Proposition 7. *The following estimate holds :*

$$\forall n\geq 0,\ \|R^{\star n}\|_{L^\infty([0,1])}\leq e^{\frac{1}{\sqrt{dt}}(\frac{1}{\lambda_+}-(1-C\sqrt{dt})ndt)}\|R^{\star 0}\|_{L^\infty([0,1])}.$$

Proof. We use a Lyapunov argument using the Lyapunov functional firstly introduced in [28] in a continuous setting and then adapted to a discrete framework [48] :

$$V_\gamma(R^{\star n}):=\sup_{1\leq j_f\leq \alpha N}|R^{\star n}_{j_f}e^{-\frac{\gamma}{\lambda_+}x^f_{j_f}}|$$

where $\gamma>0$ will be chosen later.

Using the scheme verified by $R^\star$ (identified with $\tilde{R}^\star$) (6.30), it holds for $j_f>1$:

$$\begin{aligned}|R^{\star n+1}_{j_f}e^{-\frac{\gamma}{\lambda_+}x^f_{j_f}}| \leq\ & (1-\nu_+)|R^{\star n}_{j_f}e^{-\frac{\gamma}{\lambda_+}x^f_{j_f}}|+\nu_+|R^{\star n}_{j_f-1}e^{-\frac{\gamma}{\lambda_+}x^f_{j_f}}|\\ \leq\ & (1-\nu_+)V_\gamma(R^{\star n})+\nu_+e^{-\frac{\gamma}{\lambda_+}dx_+}V_\gamma(R^{\star n})\\ =\ & (1-\nu_+(1-e^{-\frac{\gamma}{\lambda_+}dx_+}))V_\gamma(R^{\star n})\end{aligned}$$

and for $j_f = 1$:

$$\begin{aligned} |R^{\star n+1}_{j_f} e^{-\frac{\gamma}{\lambda_+} x^f_{j_f}}| &\leq (1-\nu_+)|R^{\star n}_{j_f} e^{-\frac{\gamma}{\lambda_+} x^f_{j_f}}| \\ &\leq (1-\nu_+) V_\gamma(R^{\star n}). \end{aligned}$$

As a consequence,

$$V_\gamma(R^{\star n+1}) \leq \Big(1 - \nu_+(1 - e^{-\frac{\gamma}{\lambda_+} dx_+})\Big) V_\gamma(R^{\star n}).$$

As $e^{-x} \leq 1 - x + \frac{x^2}{2}$ for $x \geq 0$, it holds :

$$V_\gamma(R^{\star n+1}) \leq \left(1 - \gamma dt + \frac{1}{2\nu_+}(\gamma dt)^2\right) V_\gamma(R^{\star n}).$$

Owing the fact that $1 + x \leq e^x$, one gets :

$$V_\gamma(R^{\star n+1}) \leq e^{-\gamma dt + \frac{1}{2\nu_+}(\gamma dt)^2} V_\gamma(R^{\star n}).$$

Thus, for $\gamma dt \leq 2\nu_+ \varepsilon$ where ε will be fixed later :

$$\forall n \geq 0, \ V_\gamma(R^{\star n}) \leq e^{-(1-\varepsilon)\gamma n dt} V_\gamma(R^{\star 0}).$$

To finish the proof, we use the fact that :

$$e^{-\frac{\gamma}{\lambda_+}} \| \cdot \|_{L^\infty([0,1])} \leq V_\gamma(\cdot) \leq \| \cdot \|_{L^\infty([0,1])}$$

to conclude that :

$$\forall n \geq 0, \ \|R^{\star n}\|_{L^\infty([0,1])} \leq e^{\gamma(\frac{1}{\lambda_+} - (1-\varepsilon) n dt)} \|R^{\star 0}\|_{L^\infty([0,1])}$$

For $\gamma := 1/\sqrt{dt}$ and $\varepsilon = \frac{1}{2\nu_+}\sqrt{dt}$, we have :

$$\forall n \geq 0, \ \|R^{\star n}\|_{L^\infty([0,1])} \leq e^{\frac{1}{\sqrt{dt}}(\frac{1}{\lambda_+} - (1 - \frac{1}{2\nu_+}\sqrt{dt}) n dt)} \|R^{\star 0}\|_{L^\infty([0,1])}.$$

This finishes the proof of the proposition.

□

A similar analysis for $S^\star$ is given, to deduce the extinction result for all the system.

Theorem 19. *For $T > T_{\min}$, there exists a constant C independent on the discretization such that for all n with $\frac{T_{\min} + C dt^{1/8}}{1 - C\sqrt{dt}} \leq n dt \leq$ and for all $(R^{\star 0}, S^{\star 0}) \in L^\infty([0,1])^2$:*

$$\|R^{\star n}\|_{L^\infty([0,1])} + \|S^{\star n}\|_{L^\infty([0,1])} \leq C e^{-\frac{1}{2dt^{1/8}}} (\|R^{\star 0}\|_{L^\infty([0,1])} + \|S^{\star 0}\|_{L^\infty([0,1])}).$$

Proof. By Proposition 7, it holds :

$$\forall n dt \leq T, \|R^{\star n}\|_{L^\infty([0,1])} \leq e^{\frac{1}{\sqrt{dt}}(\frac{1}{\lambda_+} - (1 - C\sqrt{dt}) n dt)} \|R^{\star 0}\|_{L^\infty([0,1])}. \tag{6.33}$$

In order to have estimates on $S^\star$, we again proceed by a Lyapunov argument. This time the Lyapunov function is the norm L^∞_γ defined by :

$$\|S^{\star n}\|_{L^\infty_\gamma([0,1])} := \sup_{i_c} |S^{\star n}_{i_c}| e^{\frac{\gamma}{\lambda_-}(x^c_{i_c} - 1)}$$

where $\gamma > 0$. Note that in the proof of Proposition 6, we have proven that for $C \leq \gamma \leq \frac{C}{\sqrt{dx_-}}$:

$$\|L_{11}dx_+\|_{\mathcal{L}(L^\infty_\gamma([0,1]))} + \|L_{12}dx_-\|_{\mathcal{L}(L^\infty_\gamma([0,1]))} \leq C$$

uniformly in dt, γ. Then, by similar computations as in the proof of Proposition 7 :

$$\begin{aligned} \|S^{\star n+1}\|_{L^\infty_\gamma([0,1])} \leq \quad & (1-\nu_-(1-e^{-\frac{\gamma}{\lambda_-}dx_-}))\|S^{\star n}\|_{L^\infty_\gamma([0,1])} + C|R^{\star n}_{\alpha N}| \\ & +(\|L_{11}dx_+\|_{\mathcal{L}(L^\infty_\gamma([0,1]))}\|R^{\star n}\|_{L^\infty_\gamma([0,1])} + \|L_{12}dx_-\|_{\mathcal{L}(L^\infty_\gamma([0,1]))}\|S^{\star n}\|_{L^\infty_\gamma([0,1])})dt. \end{aligned}$$

For $ndt \leq T$,

$$\|S^{\star n+1}\|_{L^\infty_\gamma([0,1])} \leq \quad (1-(\gamma-C)dt + C(\gamma dt)^2)\|S^{\star n}\|_{L^\infty_\gamma([0,1])} + C|R^{\star n}_{\alpha N}|.$$

For γ such that $C(\frac{1}{\gamma}+\gamma dt) \leq \varepsilon$ (ε will be fixed later), it holds :

$$\|S^{\star n+1}\|_{L^\infty_\gamma([0,1])} \leq \quad e^{-\gamma dt(1-\varepsilon)}\|S^{\star n}\|_{L^\infty_\gamma([0,1])} + Ce^{\frac{1}{\sqrt{dt}}(\frac{1}{\lambda_+}-(1-C\sqrt{dt})ndt)}\|R^{\star 0}\|_{L^\infty([0,1])}.$$

where we have used (6.33). Let us introduce $n_1 := E\left(\frac{1/\lambda_+ + dt^{1/4}}{dt(1-C\sqrt{dt})}\right)+1$ and for $n_1 \leq n \leq E(T/dt)$, we easily get (by induction) :

$$\begin{aligned} \|S^{\star n+1}\|_{L^\infty_\gamma([0,1])} \leq \quad & C\|R^{\star 0}\|_{L^\infty([0,1])}\sum_{k=n_1}^n e^{-\gamma(n-k)dt(1-\varepsilon)}e^{\frac{1}{\sqrt{dt}}(\frac{1}{\lambda_+}-(1-C\sqrt{dt})kdt)} \\ & +e^{-\gamma(n-n_1)dt(1-\varepsilon)}\|S^{\star n_1}\|_{L^\infty_\gamma([0,1])} \\ \leq \quad & C\|R^{\star 0}\|_{L^\infty([0,1])}e^{-\frac{1}{dt^{1/4}}}\sum_{k=n_1}^n e^{-\gamma(n-k)dt(1-\varepsilon)} + e^{-\gamma(n-n_1)dt(1-\varepsilon)}\|S^{\star n_1}\|_{L^\infty_\gamma([0,1])} \\ \leq \quad & \frac{C}{dt}\|R^{\star 0}\|_{L^\infty([0,1])}e^{-\frac{1}{dt^{1/4}}} + e^{-\gamma(n-n_1)dt(1-\varepsilon)}\|S^{\star n_1}\|_{L^\infty_\gamma([0,1])}. \end{aligned}$$

To finish the proof, we use the fact that :

$$e^{-\frac{\gamma}{\lambda_-}}\|\cdot\|_{L^\infty([0,1])} \leq \|\cdot\|_{L^\infty_\gamma([0,1])} \leq \|\cdot\|_{L^\infty([0,1])}$$

to conclude that :

$$\|S^{\star n+1}\|_{L^\infty([0,1])} \leq \quad \frac{C}{dt}\|R^{\star 0}\|_{L^\infty([0,1])}e^{\frac{\gamma}{\lambda_-}-\frac{1}{dt^{1/4}}} + e^{\frac{\gamma}{\lambda_-}-\gamma(n-n_1)dt(1-\varepsilon)}\|S^{\star n_1}\|_{L^\infty([0,1])}.$$

For $\gamma = \frac{\lambda_-}{2dt^{1/4}}, \varepsilon = Cdt^{1/4}$, one gets :

$$\|S^{\star n+1}\|_{L^\infty([0,1])} \leq \quad \frac{C}{dt}\|R^{\star 0}\|_{L^\infty([0,1])}e^{-\frac{1}{2dt^{1/4}}} + e^{\frac{1}{2dt^{1/4}}-\frac{\lambda_-}{2dt^{1/4}}(n-n_1)dt(1-\tilde{C}\sqrt{dt})}\|S^{\star n_1}\|_{L^\infty([0,1])}.$$

For $ndt \geq \frac{T_{\min}+dt^{1/8}/\lambda_-}{1-\tilde{C}\sqrt{dt}} \iff (n-n_1)dt \geq \frac{1}{1-\tilde{C}\sqrt{dt}} \times (1+dt^{1/8}) \times \frac{1}{\lambda_-}$:

$$\begin{aligned} \|S^{\star n+1}\|_{L^\infty([0,1])} \leq \quad & \frac{C}{dt}\|R^{\star 0}\|_{L^\infty([0,1])}e^{-\frac{1}{2dt^{1/4}}} + e^{-\frac{1}{2dt^{1/8}}}\|S^{\star n_1}\|_{L^\infty([0,1])} \\ \leq \quad & \frac{C}{dt}\|R^{\star 0}\|_{L^\infty([0,1])}e^{-\frac{1}{2dt^{1/4}}} + Ce^{-\frac{1}{2dt^{1/8}}}\|S^{\star 0}\|_{L^\infty([0,1])}. \end{aligned}$$

Hence,

$$\|S^{\star n+1}\|_{L^\infty([0,1])} \leq Ce^{-\frac{1}{2dt^{1/8}}}(\|R^{\star 0}\|_{L^\infty([0,1])} + \|S^{\star 0}\|_{L^\infty([0,1])}).$$

□

Because of the perturbation term $\tilde{\Gamma}_{11}$ (see also (6.32)), we lose the convergence in $e^{-\frac{1}{2dt^{1/8}}}$. The final result is given in the next corollary.

Corollary 3. *For $T > T_{\min}$ and $1 < p < \infty$, there exists a constant C independent on the discretization such that for all n with $\frac{T_{\min}+Cdt^{1/8}}{1-C\sqrt{dt}} \leq ndt \leq$ and for all $(R^0, S^0) \in L^\infty([0,1])^2$:*

$$\|R^n\|_{L^p([0,1])} + \|S^n\|_{L^p([0,1])} \leq Cdx_-(\|R^0\|_{L^\infty([0,1])} + \|S^0\|_{L^\infty([0,1])}).$$

Proof. Immediate from (6.32), Proposition 6 and Theorem 19. □

6.4 A numerical example

6.4.1 Comparison with the naive method

Here we give a numerical example with $h = 1$, $M_{12} = 2$, $M_{21} = 3$, $\lambda_+ = \lambda_- = 1$, $dt = 1/1000$. For the naive way, we have only one discretization *ie* $\alpha = 1$ for which we take a $CFL = 0.8$. The naive way gives a converged result for the kernel represented in Figure 6.12.

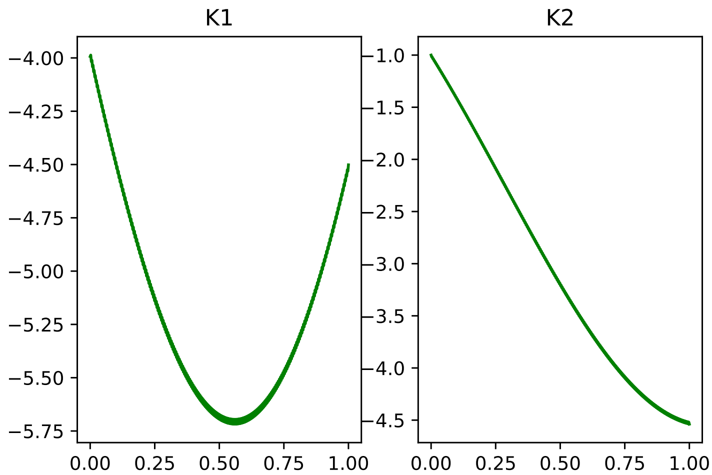
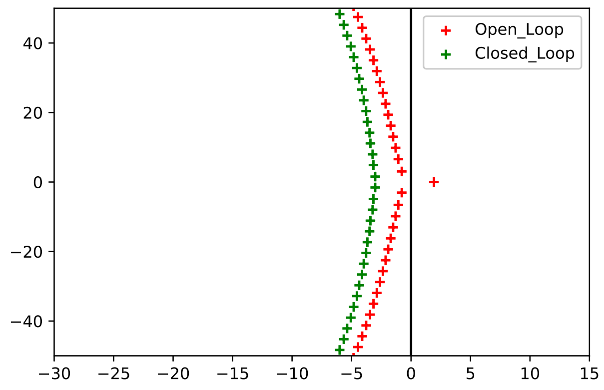


FIGURE 6.12 – The kernels of the closed-loop operator for the naive method

The kernels K_1, K_2 represents the functions such that :

$$\forall n, u^n = K_1 R dx_+ + K_2 S dx_-.$$

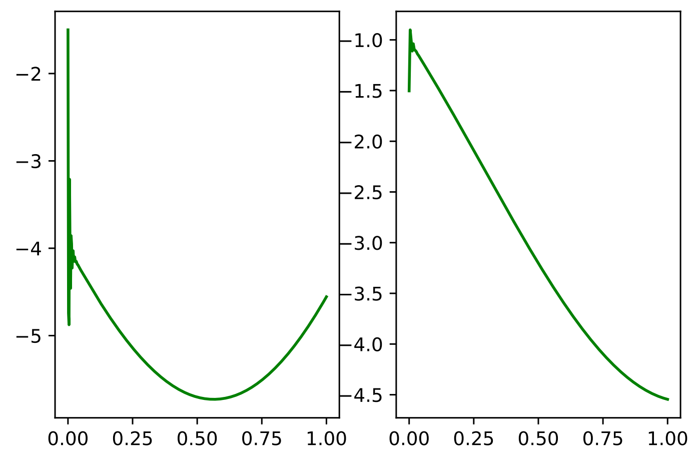
The spectrum is given in Figure 6.13.



.

FIGURE 6.13 – The spectrum of the closed loop operator for the naive method

For our scheme with $\alpha = 2, dt = 1/1000, dx_+ = 1/800, dx_- = 1/400$, the kernels are represented in Figure 6.14 :



.

FIGURE 6.14 – The kernels of the closed-loop operator for less naive method

However, the spectrum of the closed loop system is much closer to its "continuous" equivalent whose spectrum is rejected at infinity. This is displayed in Figure 6.15.

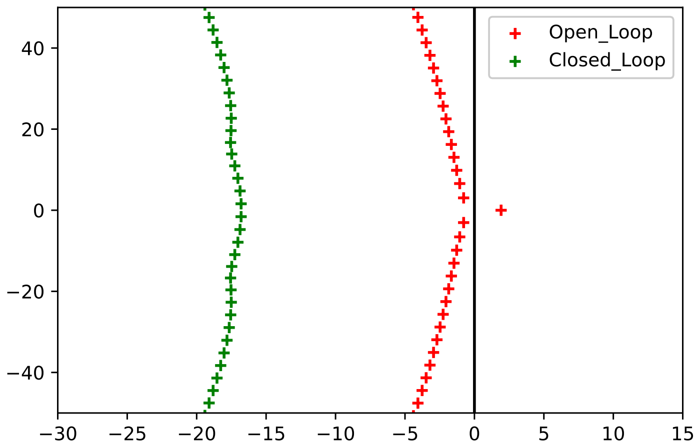
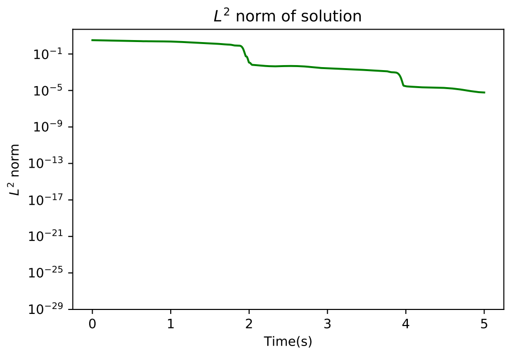


FIGURE 6.15 – The spectrum of the closed loop operator for the less naive method

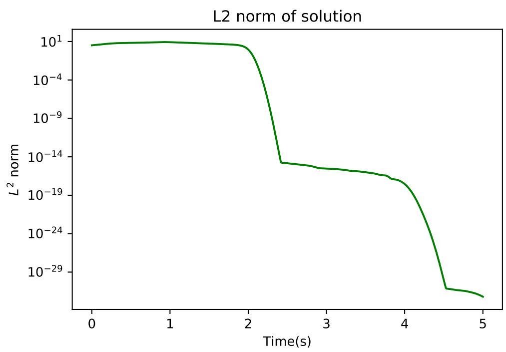
Indeed, we have a much stronger rate of convergence for our method than the the naive one. To see this, take an initial data :

$$\left\{ \begin{array}{lcl} R^0(x) & = & -4\sin(50x) \\ S^0(x) & = & 2 \times \mathbb{1}_{x\leq 0.5} - \cos(50x). \end{array} \right. \tag{6.34}$$

and look at the evolution of the L^2 norm of the solution as time goes by. The comparison between both methods is shown in Figure 6.16.



(a) The naive method ($CFL = 0.8$)



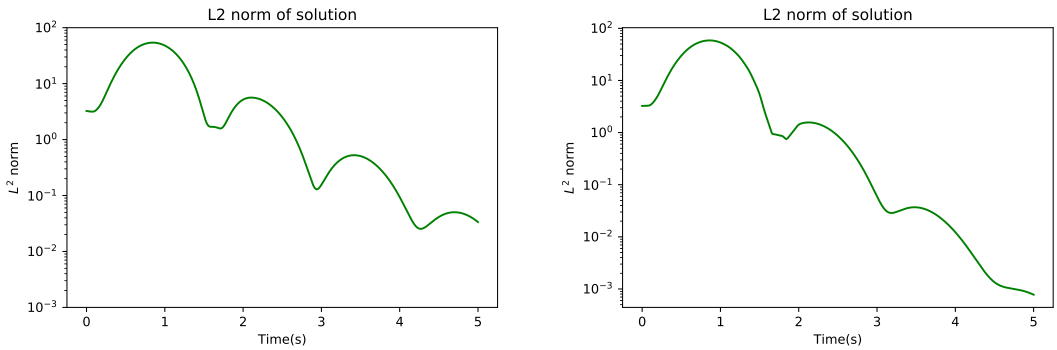
(b) Our method ($\nu_+ = 0.8, \nu_- = 0.4$)

FIGURE 6.16 – The L^2 norm of the solution (log10 scale)

Our method gives much more dissipation than the naive one. This is why it represents better the behaviour of the "continuous" solution which extinguishes at a finite time ($T_f = 2$).

6.4.2 The effect of the perturbation term

In order to observe the effect of the perturbation term $\tilde{\Gamma}_{11}$, we need to take a larger zeroth order term M. The parameters are now $h = 1$, $M_{12} = 8$, $M_{21} = -8$, $\lambda_+ = \lambda_- = 1$, $dt = \frac{1}{1000}$, $dx_+ = 1/800, dx_- = 1/200$ ($\alpha = 4$) and the initial data is given by (6.34). The Figure 6.17 gives energy dynamics :



(a) $dt = 1/1000, dx_+ = 1/800, dx_- = 1/200$ (b) $dt = 1/4000, dx_+ = 1/3200, dx_- = 1/800$

FIGURE 6.17 – When the perturbation term is non zero

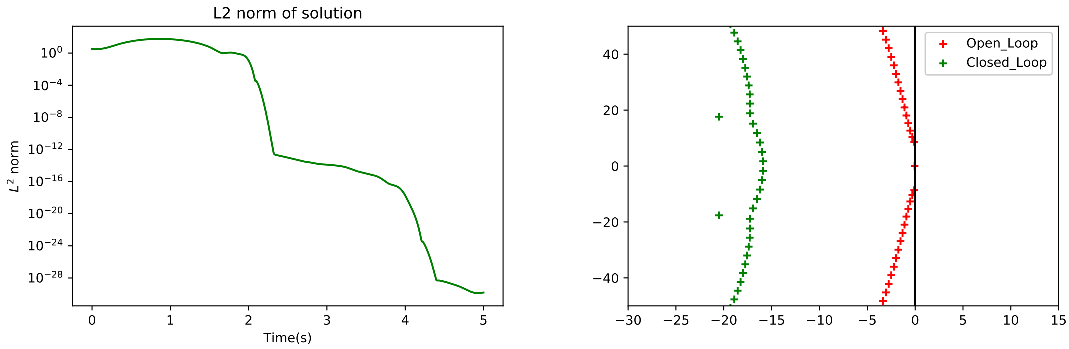
One can observe that we are far from a finite time stabilization picture since after a time $t > T_{\min} = 2$, the solution has still a lot of energy. Moreover, when the discretization is finer (Figure 6.17-b), the picture shows a stronger decay of the energy but still, one cannot see a clear extinction at $t = T_{\min}$.

In order to explain this, we need to come back to the results of Lemma 29. In fact, when the zeroth order term M is large, the constant C from Lemma 29 may be very large. Besides, we can estimate it heuristically :

$$C \simeq e^{M_{21} T_{\min}} \simeq 10^7.$$

Hence, when M is large, we need to discretize a lot (in this example dx_- should be at least 10^{-7}) for $(R^\star, S^\star)$ to be close to $(\tilde{R}^\star, \tilde{S}^\star)$.

As a consequence, the effect of the perturbation term $\tilde{\Gamma}_{11}$ can be important if dx_- is not small enough. To confirm this, now take a different discretization for which the perturbation term $\tilde{\Gamma}_{11}$ is zero. After Remark 36, this is the case when the fine grid is two times finer than the coarse one ($\alpha = 2$). Take for example, $dt = 1/1000, dx_+ = 1/800, dx_- = 1/400$ to obtain Figure 6.18 :



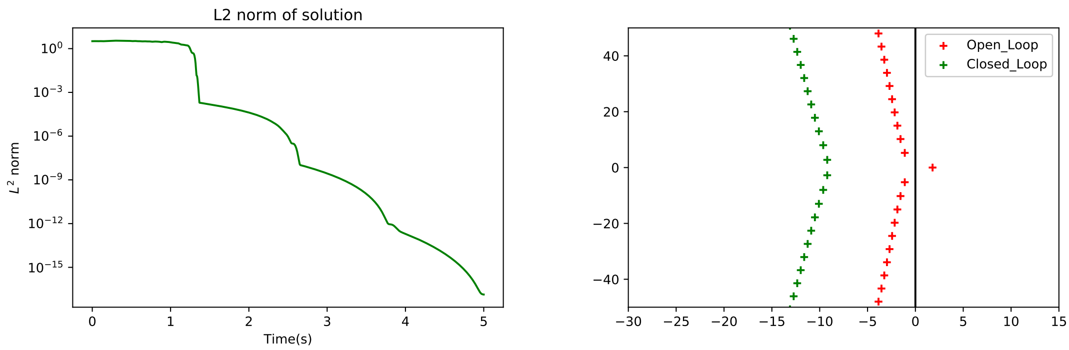
(a) The energy (b) The spectrum

FIGURE 6.18 – When the perturbation term is zero

Here, the finite time extinction is clear and after $t \simeq T_{\min}$, the solution is of the order of at least $e^{-\frac{1}{2dt^{1/8}}}$ (see Theorem 19).

6.4.3 The case of different velocities

For completeness, we illustrate that our method works for the case were $\lambda_1 \neq \lambda_2$,$\alpha > 2$ and when zeroth order term M is moderate. Taking $\lambda_+ = 1$, $\lambda_- = 3$, $M_{12} = 2, M_{21} = 3$, $dt = 1/1000$, $dx_+ = dt$ and $dx_- = 4dx_+$ ($\alpha = 4$), one gets Figure 6.19 :



(a) The energy (b) The spectrum

FIGURE 6.19 – When velocities are different

6.5 Conclusion

In this chapter, it is established that designing a discretized backstepping control must be done with the appropriate scheme. This is why we discretized the system first and then designed a backstepping control adapted to the numerical open-loop system. Doing so, an approximate finite-time stabilization result was shown for the numerical system.

The natural question which comes into play is how can we construct a similar control when the numbers d_1, d_2 of equations are larger. We can quote papers [41] where $d_1 = 1$ and $d_2 \in$

$\mathbb{N}^\star$ and [68] where $d_1, d_2 \in \mathbb{N}^\star$ for the continuous theory. In these papers, the authors design a backstepping control to have exponential stabilization. Another interesting question is the influence of the choice of the scheme. Here we took the upwind scheme but it well known that it is very diffusive and gives poor results in simulation. It would be interesting to see if we can extend the numerical backstepping method to higher order schemes like the slope limiter [79] one for example. This is not an easy task since slope limiter schemes are nonlinear even if the PDE system is linear. Finally, it would be interesting to generalize the method to a larger class of grids where the coarse grid is not a subgrid of the fine one.

CHAPITRE 7

CONCLUSION GÉNÉRALE ET PERSPECTIVES

Tout le travail fait dans cette thèse a consisté à établir rigoureusement la stabilisation exponentielle ou en temps fini de systèmes hyperboliques scalaires. On peut décomposer ce manuscrit en deux parties bien distinctes :

— Les chapitres 2, 3, 4 se focalisent plutôt sur des méthodes de Lyapunov. Dans les chapitres 2 et 3, on utilise des méthodes de Lyapunov "continues" pour exhiber un bassin d'attraction dans L^∞ lorsque la commande est saturée. Alors que dans le chapitre 4, la fonctionnelle de Lyapunov est discrète et on prouve la stabilité exponentielle d'un schéma à limiteur de pente modélisant un système de lois de conservation scalaires avec feedback linéaire.

— Les chapitres 6 et 5, quant à eux, vont utiliser des techniques linéaires. Dans le chapitre 5, on utilise des techniques de placement de poles pour stabiliser des systèmes linéaires avec couplage dans le domaine. Un bassin d'attraction est également identifié dans L^2 lorsque l'on sature la commande. Pour le chapitre 6, on discrétise deux équations de transport couplées et un résultat de stabilité en temps fini est démontré en utilisant la méthode de backstepping discrétisé, contribution essentielle de cette thèse.

Présentons quelques extensions possibles des travaux exposés dans ce manuscrit.

Stabilisation de systèmes généraux. Les équations étudiées dans ce manuscrit ne font intervenir qu'un terme de transport, certes non linéaire pour certains chapitres, mais scalaire. Il serait intéressant de développer des méthodes de stabilisation pour des systèmes hyperboliques physiques où un couplage intervient aussi dans le terme de transport. La plupart des travaux sur le sujet ne prouvent qu'une stabilisation locale puisque les espaces fonctionnels considérés sont souvent mal adaptés aux équations qui présentent des solutions très irrégulières. De plus, les techniques utilisées proviennent le plus souvent du linéaire et donc ne permettent pas de prendre en compte les phénomènes de chocs. On citera quand même les travaux [21], [32] et [5] qui n'utilisent pas le carcan du linéaire pour prouver des résultats de stabilité BV. On peut soulever deux difficultés principales pour établir un résultat de stabilisation au bord :

— La première et la plus évidente provient du caractère bien posé du problème. Il est très difficile, à ce jour, de trouver des résultats d'existence et surtout d'unicité pour des systèmes hyperboliques généraux. On donne quelques articles [97, 3, 2, 4, 27, 44] et la remarque [32, remarque 1.3] pour illustrer ce genre de problématique. Le plus souvent, on aura des solutions où on a imposé une condition de petitesse de la condition initiale pour garantir l'existence en temps long. Une idée pour avancer serait de considérer des systèmes un peu plus simple où on a pas besoin de ce genre d'hypothèse mais où on conserve quand même le couplage au sein du transport : les systèmes de classe Temple [66, p. 235].

— L'autre difficulté est le fait de ne pas avoir un sens de propagation bien défini. A chaque fois, on s'est placé dans un contexte favorable où le sens de propagation était fixé et ainsi, l'effet d'une condition au bord était évident. Pour des problèmes plus physiques, on aura des conditions aux bords beaucoup moins intuitives et pour définir un contrôle, il s'agira de construire des solveurs de Riemann au bord. On citera quelques articles [19, 6, 7] pour donner des pistes de réflexion sur ce genre de problématique.

Une idée intéressante serait de considérer la version cinétique d'un système hyperbolique [91].

$$\partial_t f_\varepsilon(t,x,v) + v\partial_x f_\varepsilon(t,x,v) = \frac{1}{\varepsilon} Q(f_\varepsilon(t,x,v)) \tag{7.1}$$

où $v \in \mathbb{R}$ est la variable des vitesses, Q un terme de collision non linéaire et $\varepsilon > 0$ est supposé petit. Cette représentation correspond à une vision mésoscopique du système physique étudié. Par exemple, pour Saint-Venant, la hauteur et le débit se déduisent aisément de f_ε : $H(t,x) =$

$\int_v f_\varepsilon dv, HV(t,x) = \int_v f_\varepsilon dv$. Cette représentation a deux avantages, le premier étant qu'il est très facile de définir une condition au bord dans (7.1). En effet, il suffit de définir f en $x = 0, v > 0$ et en $x = 1, v < 1$. Le deuxième est que le sens de propagation est bien défini par la vitesse v et on peut aisément identifier l'information qui se propage vers la gauche de celle qui se propage vers la droite. Par contre, on a un gros inconvénient qui provient du terme de couplage raide $\frac{1}{\varepsilon}Q(f_\varepsilon(t,x,v))$ qui reste difficile à traiter via une étude de Lyapunov ou une méthode de backstepping.

BIBLIOGRAPHIE

[1] F. Alabau-Boussouira, Y. Privat, and E. Trélat, *Nonlinear damped partial differential equations and their uniform discretizations*, J. Funct. Anal., 273 (2017), pp. 352–403.

[2] D. Amadori, *Initial-boundary value problems for nonlinear systems of conservation laws*, NoDEA Nonlinear Differential Equations Appl., 4 (1997), pp. 1–42.

[3] D. Amadori and R. Colombo, *Continuous dependence for* 2×2 *conservation laws with boundary*, J. Differential Equations, 138 (1997), pp. 229 – 266.

[4] D. Amadori and R.-M. Colombo, *Viscosity solutions and standard Riemann semigroup for conservation laws with boundary*, Rend. Sem. Mat. Univ. Padova, 99 (1998), pp. 219–245.

[5] F. Ancona and A. Marson, *Asymptotic stabilization of systems of conservation laws by controls acting at a single boundary point*, in Control methods in PDE-dynamical systems, vol. 426 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2007, pp. 1–43.

[6] B. Andreianov, K.-H. Karlsen, and N.-H. Risebro, *A theory of* L^1*-dissipative solvers for scalar conservation laws with discontinuous flux*, Arch. Ration. Mech. Anal., 201 (2011), pp. 27–86.

[7] B. P. Andreianov, G. M. Coclite, and C. Donadello, *Well-posedness for vanishing viscosity solutions of scalar conservation laws on a network*, Discrete Contin. Dyn. Syst., 37 (2017), pp. 5913–5942.

[8] J. Auriol, *Robust design of backstepping controllers for systems of linear hyperbolic PDEs*, PhD thesis, Mines ParisTech, Paris, 2018.

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

[10] J. M. Ball and M. Slemrod, *Feedback stabilization of distributed semilinear control systems*, Appl. Math. Optim., 5 (1979), pp. 169–179.

[11] M. K. Banda and M. Herty, *Numerical discretization of stabilization problems with boundary controls for systems of hyperbolic conservation laws*, Math. Control Relat. Fields, 3 (2013), pp. 121–142.

[12] C. Bardos, A. Y. le Roux, and J.-C. Nédélec, *First order quasilinear equations with boundary conditions*, Comm. Partial Differential Equations, 4 (1979), pp. 1017–1034.

[13] G. Bastin and J.-M. Coron, *On boundary feedback stabilization of non-uniform linear 2×2 hyperbolic systems over a bounded interval*, Systems Control Lett., 60 (2011), pp. 900–906.

[14] ——, *On boundary feedback stabilization of non-uniform linear 2×2 hyperbolic systems over a bounded interval*, Systems Control Lett., 60 (2011), pp. 900–906.

[15] ——, *Dissipative boundary conditions for one-dimensional quasilinear hyperbolic systems : Lyapunov stability for the C^1 norm*, SIAM J. Control Optim, 53 (2015), pp. 1464–1483.

[16] ——, *Stability and boundary stabilization of 1-D hyperbolic systems*, vol. 88 of Prog. in Nonlinear Differential Equations and their Applications, Birkhäuser/Springer, 2016.

[17] G. Bastin, J.-M. Coron, and B. d'Andréa Novel, *On Lyapunov stability of linearised Saint-Venant equations for a sloping channel*, Netw. Heterog. Media, 4 (2009), pp. 177–187.

[18] P. Bernard and M. Krstic, *Adaptive output-feedback stabilization of non-local hyperbolic PDEs*, Automatica J. IFAC, 50 (2014), pp. 2692–2699.

[19] S. Blandin, X. Litrico, M. L. Delle Monache, B. Piccoli, and A. Bayen, *Regularity and Lyapunov stabilization of weak entropy solutions to scalar conservation laws*, IEEE Trans. Automat. Control, 62 (2017), pp. 1620–1635.

[20] A. Bressan, *Hyperbolic systems of conservation laws*, vol. 20 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, UK, 2000. The one-dimensional Cauchy problem.

[21] A. Bressan and G.-M. Coclite, *On the boundary control of systems of conservation laws*, SIAM J. Control Optim., 41 (2002), pp. 607–622.

[22] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, New York, 2011.

[23] S. Buckley and M. Leverett, *Mechanism of fluid displacement in sands*, Trans. AIME, 146 (1942).

[24] P. J. Campo and M. Morari, *Robust control of processes subject to saturation nonlinearities*, Computers Chemical Engineering, 14 (1990), pp. 343 – 358.

[25] F. Castillo, E. Witrant, C. Prieur, and L. Dugard, *Boundary observers for linear and quasi-linear hyperbolic systems with application to flow control*, Automatica J. IFAC, 49 (2013), pp. 3180–3188.

[26] C. Chainais-Hillairet, *Second-order finite-volume schemes for a non-linear hyperbolic equation : error estimate*, Math. Methods Appl. Sci., 23 (2000), pp. 467–490.

[27] R.-M. Colombo and G. Guerra, *On general balance laws with boundary*, J. Differential Equations, 248 (2010), pp. 1017–1043.

[28] J.-M. Coron, *On the null asymptotic stabilization of the two-dimensional incompressible Euler equations in a simply connected domain*, SIAM J. Control Optim., 37 (1999), pp. 1874–1896.

[29] ——, *Control and nonlinearity*, vol. 136 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2007.

[30] J.-M. Coron, G. Bastin, and B. d'Andréa Novel, *Dissipative boundary conditions for one-dimensional nonlinear hyperbolic systems*, SIAM J. Control Optim., 47 (2008), pp. 1460–1498.

[31] J.-M. Coron, B. d'Andréa Novel, and G. Bastin, *A strict Lyapunov function for boundary control of hyperbolic systems of conservation laws*, IEEE Trans. on Automat. Control, 52 (2004), pp. 2–11.

[32] J.-M. Coron, S. Ervedoza, S. S. Ghoshal, O. Glass, and V. Perrollaz, *Dissipative boundary conditions for 2×2 hyperbolic systems of conservation laws for entropy solutions in BV*, J. Differential Equations, 262 (2017), pp. 1–30.

[33] J.-M. Coron and H.-M. Nguyen, *Dissipative boundary conditions for nonlinear 1-D hyperbolic systems : sharp conditions through an approach via time-delay systems*, SIAM Journal on Mathematical Analysis, 47 (2015), pp. 2220–2240.

[34] J.-M. Coron, R. Vazquez, M. Krstic, and G. Bastin, *Local exponential H^2 stabilization of a 2×2 quasilinear hyperbolic system using backstepping*, SIAM J. Control Optim., 51 (2013), pp. 2005–2035.

[35] J.-M. Coron and Z. Wang, *Output feedback stabilization for a scalar conservation law with a nonlocal velocity*, SIAM J. Math. Anal., 45 (2013), pp. 2646–2665.

[36] C. Dafermos, *Hyperbolic conservation laws in continuum physics*, vol. 325 of Wiss., Springer-Verlag, Berlin, fourth ed., 2016.

[37] J. de Halleux, C. Prieur, J.-M. Coron, B. Novel, and G. Bastin, *Boundary feedback control in networks of open channels*, Automatica J. IFAC, 39 (2003), pp. 1365–1376.

[38] F. Di Meglio, *Dynamic and control of slugging in oil production*, PhD thesis, Mines ParisTech, Paris, 2011.

[39] F. Di Meglio, D. Bresch-Pietri, and U. J. F. Aarsnes, *An adaptive observer for hyperbolic systems with application to underbalanced drilling*, IFAC Proceedings Volumes, 47 (2014), pp. 11391 – 11397. 19th IFAC World Congress.

[40] F. Di Meglio, G. Kaasa, N. Petit, and V. Alstad, *Slugging in multiphase flow as a mixed initial-boundary value problem for a quasilinear hyperbolic system*, in Proceedings of the 2011 American Control Conference, 2011, pp. 3589–3596.

[41] F. Di Meglio, R. Vazquez, and M. Krstic, *Stabilization of a system of $n+1$ coupled first-order hyperbolic linear pdes with a single boundary input*, IEEE Transactions on Automatic Control, 58 (2013), pp. 3097–3111.

[42] A. Diagne, G. Bastin, and J.-M. Coron, *Lyapunov exponential stability of 1-D linear hyperbolic systems of balance laws*, Automatica J. IFAC, 48 (2012), pp. 109–114.

[43] R.-J. DiPerna, *Global existence of solutions to nonlinear hyperbolic systems of conservation laws*, J. Differential Equations, 20 (1976), pp. 187–212.

[44] C. Donadello and A. Marson, *Stability of front tracking solutions to the initial and boundary value problem for systems of conservation laws*, NoDEA Nonlinear Differential Equations and Appl., 14 (2007), pp. 569–592.

[45] J. Doyle, B. Francis, and A. Tannenbaum, *Feedback Control Theory*, 2009.

[46] S. Dudret, K. Beauchard, F. Ammouri, and P. Rouchon, *Stability and asymptotic observers of binary distillation processes described by nonlinear convection/diffusion models*, in 2012 American Control Conference (ACC), 2012, pp. 3352–3358.

[47] M. Dus, *Bv exponential stability for systems of scalar conservation laws using saturated feedback.* preprint (to appear in SIAM SICON), 2020.

[48] ——, *Exponential stability of a general slope limiter scheme for scalar conservation laws subject to a dissipative boundary condition.* Preprint, Jan. 2021.

[49] M. Dus, F. Ferrante, and C. Prieur, *On L^∞ stabilization of diagonal semilinear hyperbolic systems by saturated boundary control*, ESAIM Control Optim. Calc. Var., 26 (2020), pp. Paper No. 23, 34.

[50] S. Ervedoza, *Observability properties of a semi-discrete 1D wave equation derived from a mixed finite element method on nonuniform meshes*, ESAIM Control Optim. Calc. Var., 16 (2010), pp. 298–326.

[51] S. Ervedoza and E. Zuazua, *Uniformly exponentially stable approximations for a class of damped systems*, J. Math. Pures Appl. (9), 91 (2009), pp. 20–48.

[52] N. Espitia, A. Tanwani, and S. Tarbouriech, *Stabilization of boundary controlled hyperbolic PDEs via Lyapunov-based event triggered sampling and quantization*, in Proc. 56th IEEE Conf. Decision and Control, 2017, pp. 1266–1271.

[53] R. Eymard, T. Gallouët, and R. Herbin, *Finite volume methods*, in Solution of Equation in $\mathbb{R}^n$ (Part 3), Techniques of Scientific Computing (Part 3), vol. 7 of Handbook of Numerical Analysis, Elsevier, 2000, pp. 713 – 1018.

[54] F. Ferrante, A. Cristofaro, and C. Prieur, *Boundary observer design for cascaded ODE–hyperbolic PDE systems : a matrix inequalities approach*, Automatica J. IFAC, 119 (2020), pp. 109027, 9.

[55] M. Garavello and B. Piccoli, *Traffic flow on networks*, vol. 1 of AIMS Series on Applied Mathematics, American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2006. Conservation laws models.

[56] L. Gearhart, *Spectral theory for contraction semigroups on Hilbert space*, Trans. Amer. Math. Soc., 236 (1978), pp. 385–394.

[57] S. Gerster and M. Herty, *Discretized feedback control for systems of linearized hyperbolic balance laws*, Math. Control Relat. Fields, 9 (2019), pp. 517–539.

[58] M. Ghousein, E. Witrant, V. Bhanot, and P. Petagna, *Adaptive boundary observer design for linear hyperbolic systems ; application to estimation in heat exchangers*, Automatica J. IFAC, 114 (2020), pp. 108824, 13.

[59] J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*, Comm. Pure Appl. Math., 18 (1965), pp. 697–715.

[60] S. K. Godunov, *A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics*, Mat. Sb. (N.S.), 47 (89) (1959), pp. 271–306.

[61] S. Goldstein, *On diffusion by discontinuous movements, and on the telegraph equation*, Quart. J. Mech. Appl. Math., 4 (1951), pp. 129–156.

[62] J. K. Hale and S. M. Verduyn Lunel, *Introduction to functional-differential equations*, vol. 99 of Applied Mathematical Sciences, Springer-Verlag, New York, 1993.

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

[64] A. Hayat, *Boundary stability of 1-D nonlinear inhomogeneous hyperbolic systems for the C^1 norm*, SIAM J. Control Optim., 57 (2019), pp. 3603–3638.

[65] ——, *On boundary stability of inhomogeneous 2×2 1-D hyperbolic systems for the C^1 norm*, ESAIM Control Optim. Calc. Var., 25 (2019), p. 31.

[66] H. Holden and N. H. Risebro, *Front tracking for hyperbolic conservation laws*, vol. 152 of Applied Mathematical Sciences, Springer, Heidelberg, second ed., 2015.

[67] J. M. Holte, *Discrete gronwall lemma and applications*, tech. rep., MAA north central section meeting at und, 2009.

[68] L. Hu, F. Di Meglio, R. Vazquez, and M. Krstic, *Control of homodirectional and general heterodirectional linear coupled hyperbolic pdes*, IEEE Transactions on Automatic Control, 61 (2016), pp. 3301–3314.

[69] M. Kac, *A stochastic model related to the telegrapher's equation*, Rocky Mountain J. Math., 4 (1974), pp. 497–509. Reprinting of an article published in 1956.

[70] T. Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[71] D. Kröner, S. Noelle, and M. Rokyta, *Convergence of higher order upwind finite volume schemes on unstructured grids for scalar conservation laws in several space dimensions*, Numer. Math., 71 (1995), pp. 527–560.

[72] M. Krstic, P. V. Kokotovic, and I. Kanellakopoulos, *Nonlinear and Adaptive Control Design*, John Wiley and Sons, Inc., USA, 1st ed., 1995.

[73] M. Krstic and A. Smyshlyaev, *Backstepping boundary control for first-order hyperbolic pdes and application to systems with actuator and sensor delays*, Systems & Control Letters, 57 (2008), pp. 750 – 758.

[74] M. Krstic and A. Smyshlyaev, *Boundary control of PDEs*, vol. 16 of Advances in Design and Control, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008. A course on backstepping designs.

[75] S. N. Kružkov, *First order quasilinear equations with several independent variables*, Mat. Sb. (N.S.), 81 (123) (1970), pp. 228–255.

[76] I. Lasiecka and T.-I. Seidman, *Strong stability of elastic control systems with dissipative saturating feedback*, vol. 48, 2003, pp. 243–252. Optimization and control of distributed systems.

[77] P. D. Lax, *Hyperbolic systems of conservation laws. II*, Comm. Pure Appl. Math., 10 (1957), pp. 537–566.

[78] ——, *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1973. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 11.

[79] R. J. LeVeque, *Finite volume methods for hyperbolic problems*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2002.

[80] T. Li, R. Bopeng, and J. Yi, *Semi-global C^1 solution and exact boundary controllability for reducible quasilinear hyperbolic systems*, ESAIM : M2AN, 34 (2000), pp. 399–408.

[81] T. T. Li, *Global Classical Solutions For Quasilinear Hyperbolic Systems*, vol. 32 of RAM : Research in Applied Mathematics, Masson, Paris ; John Wiley & Sons, Ltd., Chichester, 1994.

[82] M. Lichtner, *Spectral mapping theorem for linear hyperbolic systems*, Proc. Amer. Math. Soc., 136 (2008), pp. 2091–2101.

[83] M. J. Lighthill and G. B. Whitham, *On kinematic waves. II. A theory of traffic flow on long crowded roads*, Proc. Roy. Soc. London Ser. A, 229 (1955), pp. 317–345.

[84] A. Marica and E. Zuazua, *Localized solutions for the finite difference semi-discretization of the wave equation*, C. R. Math. Acad. Sci. Paris, 348 (2010), pp. 647–652.

[85] S. Marx, E. Cerpa, C. Prieur, and V. Andrieu, *Global stabilization of a Korteweg–de Vries equation with saturating distributed control*, SIAM J. Control Optim., 55 (2017), pp. 1452–1480.

[86] I. Miyadera, *Nonlinear semigroups*, vol. 109 of Translations of Mathematical Monographs, American Mathematical Society, Providence, RI, 1992. Translated from the 1977 Japanese original by Choong Yun Cho.

[87] A. F. Neves, H. S. Ribeiro, and O. Lopes, *On the spectrum of evolution operators generated by hyperbolic systems*, J. Funct. Anal., 67 (1986), pp. 320–344.

[88] S. Osher and E. Tadmor, *On the convergence of difference approximations to scalar conservation laws*, Math. Comp., 50 (1988), pp. 19–51.

[89] E. Y. Panov, *Existence of strong traces for generalized solutions of multidimensional scalar conservation laws*, J. Hyperbolic Differ. Equ., 2 (2005), pp. 885–908.

[90] V. Perrollaz, *Asymptotic stabilization of entropy solutions to scalar conservation laws through a stationary feedback law*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 30 (2013), pp. 879–915.

[91] B. Perthame, *Kinetic formulation of conservation laws*, vol. 21 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, 2002.

[92] C. Prieur and F. Ferrante, *Boundary control design for linear conservation laws in the presence of energy-bounded measurement noise*, Proceedings of the 57th IEEE Conference on Decision and Control, (2018), pp. 6550–6555.

[93] C. Prieur, S. Tarbouriech, and J. M. Gomes da Silva Jr., *Wave equation with cone-bounded control laws*, IEEE Trans. Automat. Control, 61 (2016), pp. 3452–3463.

[94] J. Prüss, *On the spectrum of C_0-semigroups*, Trans. Amer. Math. Soc., 284 (1984), pp. 847–857.

[95] T. H. Qin, *Global smooth solutions of dissipative boundary value problems for first order quasilinear hyperbolic systems*, Chinese Ann. Math. Ser. B, 6 (1985), pp. 289 – 298.

[96] M. Renardy, *On the type of certain C_0-semigroups*, Comm. Partial Differential Equations, 18 (1993), pp. 1299–1307.

[97] M. Sablé-Tougeron, *Méthode de Glimm et problème mixte*, Ann. Inst. H. Poincaré Anal. non linéaire, 10 (1993), pp. 423–443.

[98] M. Slemrod, *Feedback stabilization of a linear control system in Hilbert space with an a priori bounded control*, Math. Control Signals Systems, 2 (1989), pp. 265–285.

[99] A. Smyshlyaev, B. Guo, and M. Krstic, *Arbitrary decay rate for euler-bernoulli beam by backstepping boundary feedback*, IEEE Transactions on Automatic Control, 54 (2009), pp. 1134–1140.

[100] A. Smyshlyaev and M. Krstic, *Adaptive control of parabolic PDEs*, Princeton University Press, Princeton, NJ, 2010.

[101] P. K. Sweby, *High resolution schemes using flux limiters for hyperbolic conservation laws*, SIAM J. Numer. Anal., 21 (1984), pp. 995–1011.

[102] S. Tarbouriech, G. Garcia, J. M. Gomes da Silva Jr., and I. Queinnec, *Stability and stabilization of linear systems with saturating actuators*, Springer, London, 2011. With a foreword by Ian Postlethwaite.

[103] L. R. Tébou and E. Zuazua, *Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity*, Numer. Math., 95 (2003), pp. 563–598.

[104] L. T. Tebou and E. Zuazua, *Uniform boundary stabilization of the finite difference space discretization of the $1-d$ wave equation*, Adv. Comput. Math., 26 (2007), pp. 337–365.

[105] C. Tisdell, *Existence of solutions to first-order periodic boundary value problems*, J. Mathematical Analysis and Applications, 323 (2006), pp. 1325 – 1332.

[106] E. Trélat, *Stabilization of Semilinear PDEs, and Uniform Decay under Discretization*, London Mathematical Society Lecture Note Series, Cambridge University Press, 2017, p. 31–76.

[107] M. Tucsnak and G. Weiss, *Observation and control for operator semigroups*, Birkhäuser Advanced Texts : Basler Lehrbücher., Birkhäuser Verlag, Basel, 2009.

[108] B. Van Leer, *Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method [J. Comput. Phys.* **32** *(1979), no. 1, 101–136]*, vol. 135, 1997, pp. 227–248. With an introduction by Ch. Hirsch, Commemoration of the 30th anniversary {of J. Comput. Phys.}.

[109] A. Vasseur, *Strong traces for solutions of multidimensional scalar conservation laws*, Arch. Ration. Mech. Anal., 160 (2001), pp. 181–193.

[110] R. Vazquez, M. Krstic, and J. Coron, *Backstepping boundary stabilization and state estimation of a 2 x 2 linear hyperbolic system*, in 2011 50th IEEE Conference on Decision and Control and European Control Conference, 2011, pp. 4937–4942.

[111] Xu, C.-Z. and Sallet, G., *Exponential stability and transfer functions of processes governed by symmetric hyperbolic systems*, ESAIM Control Optim. Calc. Var., 7 (2002), pp. 421–442.

[112] E. Zuazua, *Propagation, observation, and control of waves approximated by finite difference methods*, SIAM Review, 47 (2005), pp. 197–243.

Résumé

Dans cette thèse, nous étudions le problème de stabilisation au bord de systèmes généraux d'équations aux dérivées partielles hyperboliques. Plus précisément, l'étude se focalise sur des systèmes où le transport est uniquement scalaire et où le sens de propagation de l'information est fixé. En outre, le contrôle choisi sera la plupart du temps sous la forme d'une loi de retour d'état (ou feedback) linéaire que l'on perturbera éventuellement par l'effet d'une saturation. Le travail est séparé en deux parties bien distinctes ; l'une se concentre sur des méthodes de Lyapunov, tandis que l'autre va plutôt utiliser des techniques propres au linéaire.

Pour la première partie, deux travaux principaux sont présentés. Dans un premier temps, nous ne considérons que des équations de transport linéaires à vitesses positives et cherchons à stabiliser exponentiellement le système dans L^∞ grâce à un feedback linéaire saturé. La méthode consiste à utiliser des techniques classiques de Lyapunov afin d'exhiber un bassin d'attraction et d'en donner une estimation fine. On généralise ensuite ce travail dans un cadre BV pour les systèmes de lois de conservation scalaires couplées au bord. Secondement, un système de lois de conservation scalaires à vitesses positives est discrétisé en utilisant un schéma à limiteur de pente. En s'inspirant des méthodes issues du cadre continu, une fonctionnelle de Lyapunov discrète est étudiée pour prouver la stabilisation exponentielle BV par feedback linéaire de la solution discrète.

Pour la seconde partie, deux études sont également exposées mais cette fois-ci, dans un cadre totalement linéaire. D'une part, il s'agit d'établir la possibilité de construire un feedback issu d'un placement de pôles pour stabiliser exponentiellement des edps hyperboliques linéaires avec couplage au bord et dans le domaine. D'autre part, nous développons une théorie du backstepping discrétisé pour stabiliser en temps fini un schéma numérique modélisant un système 2×2 avec couplage au bord et au sein du domaine.

Mots-Clés

Stabilisation, équations de transport, Backstepping, saturation, schémas numériques

Abstract

In this thesis, we study the problem of boundary stabilization of general hyperbolic systems of partial differential equations. More precisely, the analysis focuses on systems where the transport term is scalar and for which the information propagates in a fixed direction. In addition, the chosen control is most of the time a state feedback law for which a saturation is possibly applied. The work is divided into two distinct parts, one focusing on Lyapunov techniques while the other one uses the linearity of the problem.

In the first part of the thesis, two main works are presented. In the first one, only linear transport equations with positive velocities are considered. The main goal is to design a saturated linear feedback in order to stabilize exponentially the open-loop system in L^∞. The method consists of using classical Lyapunov techniques to exhibit a basin of attraction for which a fine estimate is given. We also extend this work to nonlinear scalar conservation laws in a BV framework.
In the other work, thanks to a slope limiter scheme, a system of scalar conservation laws is discretized. Inspired by "continuous" Lyapunov methods, a discrete Lyapunov functional is studied to prove the exponential BV stabilization of the discrete solution using a linear feedback.

In the second part of the thesis, two works are exposed as well, this time in a full linear framework. On the one hand, we study systems of linear transport equations of arbitrary dimension, coupled on the domain and at the boundary. Designing a controller from a pole placement algorithm, the exponential stabilization is proved in L^2. On the other hand, we develop a numerical Backstepping theory in order to stabilize in finite time a numerical scheme modeling a 2×2 linear system with in domain and boundary couplings.

Keywords

Stabilization, transport equations, Backstepping, saturation, numerical schemes