Contributions à la stabilisation des systèmes d'evolution

Amaury Hayat

Dauphine | PSL
UNIVERSITÉ PARIS

Habilitation à Diriger des Recherches

Présentée et soutenue publiquement le 8 février 2023

Spécialité : Mathématiques

Présenté par

Amaury HAYAT

Stabilization of 1D evolution systems

Devant le jury composé de :

M.	Jean-Michel Coron	Examinateur
M.	Sylvain Ervedoza	Examinateur
M.	Olivier Glass	Coordinateur
Mme	Paola Goatin	Rapporteure
M.	Miroslav Krstic	Rapporteur
M.	Pierre Rouchon	Examinateur
M.	Emmanuel Trélat	Examinateur
M.	Enrique Zuazua	Rapporteur

Amaury HAYAT :

CERMICS - Ecole des Ponts Paristech

Adresse électronique : amaury.hayat@enpc.fr

Remerciements

Je remercie tout d'abord très chaleureusement Olivier Glass d'avoir accepté d'être le coordinateur de cette HDR et d'avoir fait face à mes (nombreuses) questions tout au long du processus. J'aimerais ensuite vivement remercier Paola Goatin, Miroslav Krstić et Enrique Zuazua d'avoir accepté de rapporter cette HDR. Merci grandement pour votre intérêt et votre soutien. J'ai un profond respect pour vos travaux et vos avis comptent beaucoup à mes yeux. Je remercie aussi Sylvain Ervedoza, Pierre Rouchon et Emmanuel Trélat pour le temps qu'ils prennent pour assister à ma soutenance et faire partie du jury. C'est un grand honneur que vous me faites. J'aimerais enfin remercier tout spécialement Jean-Michel Coron à qui je dois tant. Depuis les débuts de ma thèse jusqu'à aujourd'hui, ses conseils avisés m'ont été très utiles et j'apprécie toujours beaucoup nos discussions, qu'elles soient scientifiques ou non. Il faut le recul de quelques années pour se rendre compte de la chance que j'ai eue d'être formé par quelqu'un aussi bienveillant et amical. J'essaye, à mon échelle, de m'en inspirer lorsque j'encadre de futurs jeunes chercheurs.

Mes remerciements vont aussi à l'ensemble de mes collègues du CERMICS, qui m'ont tous –sans exception– accueilli avec une bienveillance qui fait chaud au coeur. J'aimerais remercier en particulier Gabriel Stoltz, Eric Cancès et Tony Lelièvre, qui ont su m'aiguiller dès mon arrivée sur de nombreux sujets tant concernant la recherche que l'enseignement à l'Ecole des Ponts. Je remercie aussi Virginie Ehrlacher, à qui je dois mes premières collaborations au CERMICS et l'opportunité de co-encadrer un brillant élève. Enfin, j'aimerais remercier aussi tout spécialement Isabelle Simunic, qui prend toujours le temps de discuter avec moi malgré son emploi du temps chargé, et qui me sauve régulièrement la vie administrative, avec l'aide de Stéphanie Bonnel. Au-delà du CERMICS, Je tiens à exprimer ma gratitude à l'Ecole des Ponts Paristech, qui a fait le pari de me donner ma chance en me recrutant à peine sorti de thèse. Je remercie grandement Françoise Prêteux et Sophie Mougard pour cela. Je remercie aussi Jérôme Lesueur pour son soutien à mes projets de recherche et Anthony Briant pour son intérêt pour mes travaux qui me va droit au coeur. Cette Ecole est un bel endroit et je suis reconnaissant d'y travailler.

Plusieurs personnes m'ont aidé à progresser ces dernières années et j'aimerais remercier en particulier Benedetto Piccoli, qui m'a accueilli dans son groupe [1]. Je suis très reconnaissant d'avoir pu travailler avec lui et j'ai beaucoup appris, à la fois en mathématiques et sur le métier de chercheur. Son enthousiasme permanent est un modèle à mes yeux. Merci pour ton soutien et merci de m'avoir permis de rejoindre un projet qui m'a tenu très à coeur pendant trois ans. J'en profite pour remercier à cette occasion Alexandre Bayen ainsi que toute l'équipe de cette aventure incroyable qu'est CIRCLES et particulièrement Jonathan Lee, Sean McQuade, Xiaoqian Gong, Maria Teresa Chiri, Ryan Delorenzo, Tinhinane Mezair, Alexander Keimer, et Sydney Truong. Merci aussi à Zheming An et Nathaniel Merrill pour le travail mené ensemble, et merci enfin à Thibault Liard, grâce à qui j'ai enfin compris ces dessins mystérieux que les gens font pour représenter un problème de Riemann.

Le travail présenté dans ce manuscrit est autant le travail de mes co-auteurs que le mien, et je les remercie vivement pour tous les bons moments que j'ai passés à travailler avec eux. Je pense en particulier à Georges Bastin, et nos séances de travail à Paris, ainsi qu'à Peipei Shang, et sa bonne humeur contagieuse (même quand j'ai un mois de retard sur un projet). J'aimerais remercier tout particulièrement Shengquan Xiang, un mathématicien exceptionnel avec qui j'ai eu la chance de travailler sur des sujets extrêmement divers à Paris,

1. ma non riuscì a convincermi che tutti i grandi matematici francesi della storia erano italiani. La discussione continua.

Lausanne ou Oberwolfach, ainsi que mes fidèles compagnons du backstepping, ces chevaliers de Fredholm qui ne reculent pas face à la longueur des preuves: Christophe Zhang, Ludovick Gagnon, et plus récemment Swann Marx.
Depuis trois ans, j'ai eu l'opportunité de découvrir un monde fascinant à l'interface entre mathématiques et intelligence artificielle. J'étais loin d'imaginer que nos premières discussions avec Guillaume Lample puis notre première collaboration avec Guillaume et François Charton mènerait à tout cela. Un grand merci à eux ainsi qu'à Timothée Lacroix, Xavier Martinet, Aurélien Rodriguez, et toute l'équipe d'Evariste. Travailler avec vous est un vrai plaisir. Quand nous avons commencé, la plupart des mathématiciens étaient sceptiques à l'idée qu'un modèle d'intelligence artificielle puisse un jour faire des mathématiques. Aujourd'hui, j'ai le sentiment que les choses ont commencé à changer. Merci enfin à Harini Desiraju et Alexandre Krajenbrink qui partagent ce sentiment et avec qui, j'espère, nos travaux porteront leurs fruits.

J'aimerais ensuite remercier les élèves que j'ai eu la chance d'encadrer et tout particulièrement Jean Cauvin-Vila, pour sa bonne humeur, sa rigeur et sa gentillesse, et Nathan Lichtlé, dont la thèse est loin d'être finie et pourtant déjà très consistante. C'est un réel plaisir de travailler avec vous. La relève est prometteuse avec Epiphane Loko et Fabian Glöckle dont j'ai déjà pu voir les grandes qualités.

J'aimerais, aussi, remercier mes amis et mes proches qui m'accompagnent depuis de nombreuses années. En commençant par ceux que j'ai rencontrés sur un certain terrain de rugby il y a des années, William et Mathilde, Sylvain et Sophie C., Romain et Sophie M., Quentin et Mathieu, Maxence et Constance, Léa et James, Anrong et Christie, mais aussi Adélie, Romuald, Paul-Henri, Daphné, Rémy, Thomas, Coralie, Alexis, Benjamin, Adrien, Tanguy, Etienne, JB, Louis, Hélène, Aglaé, Jean-Vincent, Cédric, Quentin L., et de nombreux autres que je n'ai pas la place de citer, mais à qui je tiens tout autant et que j'ai toujours plaisir à voir (quelle que soit la fréquence).

Je remercie ma famille et en particulier mes parents Serge et Martine et mes soeurs Lauriane et Amandine. Vous êtes mes premiers soutiens et mes rocs, j'ai de la chance de vous avoir.

Enfin, j'aimerais remercier une personne qui m'est très chère. Merci Gabriela d'avoir changé ma vie, d'avoir confiance en moi et d'être toujours là pour me soutenir au jour le jour. Quelle que soit la journée, le meilleur moment est quand je me lève à tes cotés.

Contents

Chapter 1

Introduction

1.1 Presentation of the thesis

Control theory answers a simple question: "if we can act on a system, what can we make it do?" From a mathematical point of view, we have a system of equations in which we can choose one or several parameters (in a sense to be defined) and we wonder if we can choose these parameters so that the solutions of the system have the behavior we want and, if so, how. These parameters, called *controls*, can be for example a term of an equation, a coefficient, a boundary condition, etc. They can have constraints (or not) on the regularity, the support, the domain and the image, the dimension etc. This theory is often divided into three branches:

— Controllability: this consists in knowing if one can reach any final state from any initial condition, by choosing well the controls. The classical example is the following: for a system of the form

$$\dot{x}(t) = f(x(t), u(t)), \tag{1.1.1}$$

where u is a control parameter that we can choose, T is a given time, x_0 and x_1 are any initial and final state, does there exist for a control u such that if $x(0) = x_0$ then $x(T) = x_1$?

— Optimal control: we are looking for the best way to choose the controls to achieve an objective. "Best" refers to some cost functional over which we are optimizing. For example for a system of the form

$$\dot{x}(t) = f(x(t), u(t)), \tag{1.1.2}$$

and a functional $J(t, x, u)$ that we want to minimize, is there one or more optimal solutions u of this minimization problem and, if so, how to find them?

— Stabilization: we want to know if a given trajectory is asymptotically stable, i.e. that whatever the initial condition of the system, the solution(s) converge(s) to this given trajectory. This convergence can be either only asymptotic, or exponential, or even in finite time. For example: consider the system

$$\dot{x}(t) = f(x(t), u(t)), \tag{1.1.3}$$

is it possible to find u such that, whatever the initial condition x_0, $\lim_{t \to +\infty} x(t) = 0$? This of course depends on f and the constraints on u.

These control problems have many variations in many contexts, in finite or infinite dimensional spaces, for ordinary or partial differential equations, on different types of domains, with or without constraints, etc. The theory, already rich, is still far from being complete, in particular concerning the third category: stabilization. We focus almost exclusively on this branch in this thesis.

The main characteristic of stabilization –compared to controllability and optimal control– is that the control has the form of a feedback. That is, it depends on the state of the system at each time t and not on the initial condition, which is potentially unknown. Formally speaking this means [1]

$$u(t) = \mathcal{F}(t, y(t, \cdot)). \tag{1.1.4}$$

From a practical point of view one can understand the interest: a control that depends on the initial condition will be blind to possible disturbances along the trajectory. These disturbances will almost inevitably occur since the model is not perfect, and will deviate the system from its ideal trajectory. On the other hand, a control that depends on the state of the system at time t can react to these perturbations and thus adapt. This is what automaticians call a closed loop system.

Before diving into the theory, we can see that control theory has a specificity that is relatively rare [2] within mathematics. On the one hand, the use of a control depending itself on the state of the system and the presence of partial differential equations makes the mathematical problem complicated, to the point of giving rise to a very rich and yet still largely incomplete mathematical theory (in particular for non-linear systems as well as rapid or finite time stabilization problems, see Part II). On the other hand, stabilization problems have many direct applications, in industry, engineering, etc. Stabilization is a question that humans have been trying to solve, with or without mathematical tools, since Antiquity [3] to the point that, even nowadays, practice is sometimes ahead of mathematics: the proportional-integral (PI) control used in the regulation of waterways was used in engineering applications years before having the mathematical analysis confirming which control ensures the stability of the system in [21] then [136].

This makes stabilization an abstract and theoretical field, but with extremely practical direct applications. If all mathematical fields have practical applications more or less distant (and sometimes even unsuspected), this direct proximity is not that common. This leads to having very different communities, from mathematicians to engineers, working on similar problems with quite different visions (and sometimes rigors). Although this thesis belongs to the mathematical side, it will try to give a glimpse of this diversity. Parts I–II deal with an abstract and purely mathematical problem, while Part III talks about a very practical problem of road traffic control, yet sometimes just as theoretical, as we will see in Chapter 9.

In order to facilitate the reading, this thesis is divided into four largely independent parts. Each part is summarized below in Section 1.3. In each of them, we first present some background information before presenting some of our results, taken from the articles listed below (Section 1.2). Each part generally focus on one or two articles, for which we will give brief ideas of proofs sufficiently detailed to (hopefully) understand how they work, without giving the whole proof.

1. in some cases the control can also depend on the state of the system at previous times. To simplify, we just use this expression here
2. in my opinion
3. One can note the example of the water clock of Ktésibios (also known as Ctesibius) [90], or much more recently the fire pump of the Perier brothers whose integral regulation system is the ancestor of the PI controls which were theorized mathematically rigorously only towards the beginning of the XX$^{\text{th}}$ century with, among others, the works of Minorski [189].

1.2 Publications and supervisions

1.2.1 List of publications

Publications and pre-publications presented in this thesis

[AH1] Georges Bastin, Jean-Michel Coron, and Amaury Hayat. Input-to-state stability in sup norms for hyperbolic systems with boundary disturbances. *Nonlinear Anal.*, 208:112300, 28, 2021.

[AH2] Amaury Hayat. Global exponential stability and input-to-state stability of semilinear hyperbolic systems for the L^2 norm. *Systems Control Lett.*, 148:Paper No. 104848, 8, 2021.

[AH3] Jean Cauvin-Vila, Virginie Ehrlacher, and Amaury Hayat. Boundary stabilization of one-dimensional cross-diffusion systems in a moving domain: linearized system. *Journal of Differential Equations*, 2022.

[AH4] Ludovick Gagnon, Amaury Hayat, Shengquan Xiang, and Christophe Zhang. Fredholm transformation on laplacian and rapid stabilization for the heat equation. *Journal of Functional Analysis*, 2021.

[AH5] Ludovick Gagnon, Amaury Hayat, Shengquan Xiang, and Christophe Zhang. Fredholm backstepping for critical operators and application to rapid stabilization for the linearized water waves. *arXiv preprint arXiv:2202.08321*, 2022.

[AH6] Jean-Michel Coron, Amaury Hayat, Shengquan Xiang, and Christophe Zhang. Stabilization of the linearized water tank system. *Arch. Ration. Mech. Anal.*, 244(3):1019–1097, 2022.

[AH7] Amaury Hayat, Benedetto Piccoli, and Sydney Truong. Dissipation of traffic jams using a single autonomous vehicle on a ring road. *SIAM Journal on Applied Mathematics*, 2023.

[AH8] Amaury Hayat, Thibault Liard, Francesca Marcellini, and Benedetto Piccoli. A multiscale second order model for the interaction between AV and traffic flows: analysis and existence of solutions. working paper or preprint, January 2021.

[AH9] Francois Charton, Amaury Hayat, and Guillaume Lample. Learning advanced mathematical computations from examples. In *International Conference on Learning Representations*, 2020.

[AH10] Guillaume Lample, Marie-Anne Lachaux, Thibaut Lavril, Xavier Martinet, Amaury Hayat, Gabriel Ebner, Aurélien Rodriguez, and Timothée Lacroix. Hypertree proof search for neural theorem proving. accepted in *Advances in neural information processing systems*, 2022.

Publications and pre-publications posterior to the PhD thesis, but not presented in this thesis.

[AH11] Saleh Albeaik, Alexandre Bayen, Maria Teresa Chiri, Xiaoqian Gong, Amaury Hayat, Nicolas Kardous, Alexander Keimer, Sean T McQuade, Benedetto Piccoli, and Yiling You. Limitations and improvements of the intelligent driver model (IDM). *SIAM Journal on Applied Dynamical Systems*, 21 (3): 1862-1892, 2021.

[AH12] Zheming An, Nathaniel J Merrill, Kwangwon Lee, Rémi Robin, Amaury Hayat, Olivia Zapfe, and Benedetto Piccoli. A Two-Step Model of Human Entrainment: A Quantitative Study of Circadian Period and Phase of Entrainment. *Bulletin of Mathematical Biology*, 83(2):1–29, 2021.

[AH13] Georges Bastin, Jean-Michel Coron, and Amaury Hayat. Feedforward boundary control of $2\times$ 2 nonlinear hyperbolic systems with application to Saint-Venant equations. *European Journal of Control*, 57:41–53, 2021.

[AH14] Georges Bastin, Jean-Michel Coron, and Amaury Hayat. Diffusion and robustness of boundary feedback stabilization of hyperbolic systems. *Mathematics of Control, Signals, and Systems* 2022.

[AH15] François Charton, Amaury Hayat, Sean T McQuade, Nathaniel J Merrill, and Benedetto Piccoli. A deep language model to predict metabolic network equilibria. *arXiv preprint arXiv:2112.03588*, 2021.

[AH16] Amaury Hayat. Boundary stabilization of 1D hyperbolic systems. *Annu. Rev. Control*, 52:222–242, 2021.

[AH17] Amaury Hayat, Xiaoqian Gong, Jonathan Lee, Sydney Truong, Sean McQuade, Nicolas Kardous, Alexander Keimer, Yiling You, Saleh Albeaik, Eugene Vinistky, et al. A holistic approach to the energy-efficient smoothing of traffic via autonomous vehicles. In *Intelligent Control and Smart Energy Management*, pages 285–316. Springer, 2022.

[AH18] Amaury Hayat, Yating Hu, and Peipei Shang. PI control for the cascade channels modeled by general Saint-Venant equations. *preprint*, 2022.

[AH19] Amaury Hayat, Benedetto Piccoli, and Shengquan Xiang. Stability of multi-population traffic flows. *Networks and Heterogeneous media*, 2023.

[AH20] Nicolas Kardous, Amaury Hayat, Sean T McQuade, Xiaoqian Gong, Sydney Truong, Tinhinane Mezair, Paige Arnold, Ryan Delorenzo, Alexandre Bayen, and Benedetto Piccoli. A rigorous multi-population multi-lane hybrid traffic model for dissipation of waves via autonomous vehicles. *The European Physical Journal Special Topics*, pages 1–12, 2022.

PhD Publications

[AH21] Georges Bastin, Jean-Michel Coron, Amaury Hayat, and Peipei Shang. Boundary feedback stabilization of hydraulic jumps. *IFAC J. Syst. Control*, 7:100026, 10, 2019.

[AH22] Georges Bastin, Jean-Michel Coron, Amaury Hayat, and Peipei Shang. Exponential boundary feedback stabilization of a shock steady state for the inviscid Burgers equation. *Math. Models Methods Appl. Sci.*, 29(2):271–316, 2019.

[AH23] Jean-Michel Coron and Amaury Hayat. PI controllers for 1-D nonlinear transport equation. *IEEE Trans. Automat. Control*, 64(11):4570–4582, 2019.

[AH24] Amaury Hayat. Boundary stability of 1-D nonlinear inhomogeneous hyperbolic systems for the C^1 norm. *SIAM J. Control Optim.*, 57(6):3603–3638, 2019.

[AH25] Amaury Hayat. On boundary stability of inhomogeneous 2×2 1-D hyperbolic systems for the C^1 norm. *ESAIM Control Optim. Calc. Var.*, 25:Paper No. 82, 31, 2019.

[AH26] Amaury Hayat. PI controllers for the general Saint-Venant equations. *Journal de l'Ecole Polytechnique*, to appear, 2022

[AH27] Amaury Hayat and Peipei Shang. A quadratic Lyapunov function for Saint-Venant equations with arbitrary friction and space-varying slope. *Automatica J. IFAC*, 100:52–60, 2019.

[AH28] Amaury Hayat and Peipei Shang. Exponential stability of density-velocity systems with boundary conditions and source term for the H^2 norm. *Journal de Mathématiques Pures et Appliquées*, 153:187–212, 2021.

Publications prior PhD thesis

[AH29] Amaury Hayat, Andrew J Hacket-Pain, Hans Pretzsch, Tim T Rademacher, and Andrew D Friend. Modeling tree growth taking into account carbon source and sink limitations. *Frontiers in plant science*, 8:182, 2017.

[AH30] Amaury Hayat, JP Balthasar Mueller, and Federico Capasso. Lateral chirality-sorting optical forces. *Proceedings of the National Academy of Sciences*, 112(43):13190–13194, 2015.

1.2.2 Supervisions

Internships, research assistants and master students

— Sydney Truong, 2019-2020, 8 months, co-advised with Benedetto Piccoli. Sydney Truong received the prize of the best undergraduate research at Rutgers University - Camden for this work.

— Paige Arnold, 2020, 2 months, co-advised with Benedetto Piccoli.

— Nicolas Kardous, 2020, 6 months, co-advised with Alexandre Bayen and Alexander Keimer.

— Epiphane Loko, master thesis defended in September 2021.

— Tinhinane Mezair, co-advised with Benedetto Piccoli, master thesis defended in July 2022.

PhD students

— Jean Cauvin-Vila, started in October 2020 (co-advised with Virginie Ehrlacher).

— Nathan Lichtlé, started in October 2021 (co-advised with Alexandre Bayen).

— Epiphane Loko, started in September 2022 (co-advised with Antoine Chaillet).

1.3 Summary of the thesis

1.3.1 Part 1

This part is in the direct continuity of my PhD work. The goal is to use a Lyapunov approach to determine sufficient conditions of exponential stability. We are looking for conditions under which there exists, for a given system, a basic quadratic Lyapunov function, i.e. a functional similar to an energy that decreases exponentially along the trajectories. The systems we look at are one-dimensional and mostly nonlinear hyperbolic systems where the control is located at the boundaries. These systems model many physical phenomena and are found in many areas such as hydrodynamics, engineering, physics, but also biology or economics [12, 19, 29, 67, 134, 185, 220]. They can be written as

$$\begin{aligned} \partial_t \mathbf{u} + A(\mathbf{u},x)\partial_x \mathbf{u} + B(\mathbf{u},x) = 0, \quad x \in (0,L), \\ \begin{pmatrix} \mathbf{u}_+(t,0) \\ \mathbf{u}_-(t,L) \end{pmatrix} = G \begin{pmatrix} \mathbf{u}_+(t,L) \\ \mathbf{u}_-(t,0) \end{pmatrix}. \end{aligned} \tag{1.3.1}$$

For such systems, these Lyapunov functions take the form

$$V(\mathbf{U}) = \sum_{n=0}^{l} \|F(\cdot)E(\mathbf{U},\cdot)D_n\mathbf{U}(\cdot)\|_{L^p(0,L)}, \quad \forall\, U \in W^{l,p}(0,L), \tag{1.3.2}$$

where $W^{l,q}$ is the Sobolev norm considered, F is a diagonal matrix of weight functions $(f_i)_{i\in\{1,\dots,n\}} \in C^1([0,L];(\mathbb{R}_+^*)^n)$, $E(\mathbf{U},x)$ is a matrix diagonalizing $A(\mathbf{U},x)$, and D_n is a differential operator iteratively defined by $D_0\mathbf{U} = \mathbf{U}$ and

$$D_{n+1}\mathbf{U} = \partial_{\mathbf{U}}(D_{n-1}\mathbf{U})(-A(\mathbf{U},x)\partial_x\mathbf{U} - B(\mathbf{U},x)). \tag{1.3.3}$$

These functions are thus written as norms with weights and the question we ask is:

"Under what condition do weights $(f_i)_{i\in\{1,\dots n\}}$ exist in $C^1([0,L];(\mathbb{R}_+^*)^n)$ such that V is a Lyapunov function for the system under consideration, i.e. decreases exponentially along the trajectories?"

For example, a basic quadratic Lyapunov function (or energy-like Lyapunov function) for the L^2 norm is written

$$\int_0^L \mathbf{u}^T(t,x)Q(x)\mathbf{u}(t,x)dx, \tag{1.3.4}$$

where $Q = \text{diag}(f_1, ...f_n)$ is a diagonal matrix of weights. When such a Lyapunov function exists for a given norm the system is exponentially stable for this norm. These Lyapunov functions are the key ingredient of many works in a general framework [18, 19, 57, 132, 133, 207, 245] or in the framework of particular physical systems (see for example [37, 59, 94, 95, 127, 128, 139]). The Lyapunov approach is a very general method and in particular much more general than the 1D setting. It can work in a multidimensional setting and can give rise to decay rate that are not exponential (see for instance [182, 183]). However, when the system is multi-dimensional there is currently no known systematic way of deriving good Lyapunov functions. Some results exists in particular cases by leveraging a natural energy of the system (see for instance [263]).

Before continuing, one might wonder why the matrix Q of (1.3.4) must be diagonal and not simply positive definite. This is in fact a result shown in [18]: if V is a Lyapunov function for the norm L^2 of the form (1.3.4) then Q is necessarily diagonal. This comes from the fact that the eigenvalues of the system (1.3.1) are distinct. When some eigenvalues are repeated it is possible to find block diagonal Lyapunov functions for the linearized system but they do not necessarily guarantee the stability of the associated nonlinear system, unlike diagonal Lyapunov functions.
The advantage of these simple Lyapunov functions is that they often lead to stability conditions with relatively simple controls, and in addition this stability is often robust. To measure this robustness, we show in Chapter 2 that the sufficient stability conditions that we find also allow us to obtain a more general property than the exponential stability: the *Input-to-State Stability (ISS)*.
Introduced in 1989 in [225] for finite dimensional systems, ISS consists in looking at the resilience of the exponential stability when unknown and unmeasured perturbations occur. This notion was then extended to delay systems (see [44] for a survey of known results) and then generalized to PDEs (see [151, Chapter 1] for more details). Unknown perturbations can then occur in the dynamics or the boundary conditions. Most of the time it is no longer possible to have an exponential stability [4] but it is sometimes possible to show that the deviation from exponential stability is continuous with respect to these unknown perturbations. More precisely we try to show an estimate of the form [5]

$$\|\mathbf{u}(t,\cdot)\|_X \leq C_1 e^{-\gamma t}\|\mathbf{u}_0\|_X + C_2\left(\|\mathbf{d}_1\|_{X_t\times X} + \|\mathbf{d}_2\|_{X_t}\right), \tag{1.3.5}$$

where $\|\cdot\|_X$ is the norm considered (formally) in space, $\|\cdot\|_{X_t}$ the norm considered in time and $\|\cdot\|_{X_t\times X}$ is the norm considered in time and space, $\gamma > 0$ is the exponential decay rate, $\mathbf{d}_1$ are perturbations on the dynamics (which depend on time and space) and $\mathbf{d}_2$ are perturbations at the boundaries (and depend only on time). We can notice that when there are no perturbations, i.e. $\mathbf{d}_1 \equiv \mathbf{d}_2 \equiv 0$, then we recover the definition of exponential stability. This is why ISS is often referred to as a generalization of exponential stability, which is of interest in practical applications where there are always disturbances, either from unknown external elements or from deviations between reality and the chosen mathematical model. ISS for PDE systems is still much less studied than the exponential stability and most of the results are recent. One can quote for example [151, Part I-Part II], where the authors give ISS conditions for a semilinear parabolic or a linear hyperbolic PDE for the norm L^p for any $p \in \mathbb{N}\setminus\{0\}\cup\{+\infty\}$. In particular, it was the most advanced results for ISS of inhomogeneous hyperbolic systems in L^∞ norm before [22]. In [81], the authors study Lyapunov functions for ISS (see Section 2.2) and apply them to ISS of semilinear reaction-diffusion equations for L^p and H^1 norm. In [187], the authors study a linear parabolic system for L^2 norm and in [208] the authors study a non-autonomous linear hyperbolic system and perturbations in the dynamics for L^2 norm. In [82], the authors show an ISS property for semilinear wave equations in the sup norm, as well as a partial ISS property for the L^2 norm. In [232], the authors study linear homogeneous hyperbolic systems in the H^1 norm and

4. exceptions exist, proportional integral controls -which we will not discuss in this thesis- allow to keep the exponential stability even with constant unknown perturbations. The sliding mode generalizes this to time-dependent perturbations (see for example [175])

5. In fact this estimate is even a little restrictive, it is what we call *exponential ISS*. Weaker estimates exist, see for example [151].

show an ISS estimate using a dynamical controller obtained as a solution of an ODE. Other relevant results can be found in [3] where the authors relate ISS for a nonlinear system in the H^p-norm to the behavior of a storage functional and in [190] where the authors reduce ISS problem in general to ISS with respect to constant perturbations alone for monotone nonlinear systems. The case of an inhomogeneous linear system in the L^2 norm has been treated in [106] while the nonlinear case has been treated in [245] with the H^2 norm.

We investigate the following problems

— In Chapter 2, we are interested in ISS of 1D hyperbolic systems in general, and we show that exponential stability conditions for the C^q norms ($q \geq 1$) obtained in [58, 132] also imply ISS. Interestingly, the Lyapunov functions we obtain do not satisfy the differential estimate usually expected for Lyapunov functions, but still allow us to obtain the ISS.

— In Chapter [135], we focus on the ISS of a semilinear Lipschitz system and we show that in this framework, we can not only stabilize the system for the L^2 norm but also find global ISS and global exponential stability conditions.

1.3.2 Part 2

The second part focuses on a method called backstepping and its generalization. Originally, backstepping is a method to stabilize finite dimensional systems, introduced in [42, 157, 234]. It takes advantage of the triangular (or cascade) structure of a system. This method has been adapted to infinite dimensional systems in [60] and then modified to be applied to partial differential equations (see for example [16, 30, 162] and the Section 1.3.2).
The idea of backstepping is simple: find a transformation between the system of interest and a system that is simple to stabilize (for example thanks to a basic quadratic Lyapunov function). If this transformation is invertible, it is enough to find a stabilizing control for the simple system and then to apply the inverse transformation in order to have a control for the original system. The problem is to prove the existence of an invertible transformation between the original system and the simple system. For example if we want to stabilize the following system

$$\begin{array}{l} \partial_t z_1 + \partial_x z_1 + z_2 = 0, \\ \partial_t z_2 - \partial_x z_2 + z_1 = 0. \end{array} \quad \text{on } [0, L], \tag{1.3.6}$$

with boundary conditions

$$\begin{array}{l} z_1(t, 0) = z_2(t, 0), \\ z_2(t, L) = u(t), \end{array} \tag{1.3.7}$$

where $u(t)$ is the feedback control we aim to find, we may want to find T in $\mathcal{L}(L^2, L^2)$ such that $y := Tz$ is solution to the target system

$$\begin{array}{l} \partial_t y_1 + \partial_x y_1 = 0, \\ \partial_t y_2 - \partial_x y_2 = 0. \end{array} \quad \text{on } [0, L], \tag{1.3.8}$$

with boundary conditions

$$\begin{array}{l} y_1(t, 0) = y_2(t, 0), \\ y_2(t, L) = v(t). \end{array} \tag{1.3.9}$$

It can be shown that this target system is easy to stabilize and is exponentially stable [6] if one chooses $v(t) := e^{-\lambda} z_1(t, L)$. Therefore, if T is an isomorphism, z converges exponentially to 0 whatever the initial

6. and with an arbitrary decay rate provided that λ can be chosen arbitrarily

condition and the original system is thus exponentially stable.

If the principle of this method is simple, its practical implementation is much more complicated: a general transformation between two systems belongs a priori to a very large space, potentially difficult to explore, and ensuring invertibility is not always easy. It should also be noted that the control is also a parameter that can be chosen and, in particular, it can be chosen such that the transformation exists. So, the problem can be reformulated as follows: find a control and an invertible transformation such that the image of the original system by the transformation is a stable system. However, this equivalent formulation is not easier to implement. Moreover, we have to be careful that the transformation is done in the space of interest (L^2 for example if we are looking for stability in L^2 norm) and that the regularity of the obtained control is compatible. Finally, since the final feedback control u is obtained by applying a potentially complicated inverse transformation to the feedback control v of the target system, the final feedback control is also potentially complicated. Unlike the controls obtained by the Lyapunov approach presented in Part I, these controls often depend on the state of the system on the whole domain and not only on one point. Nevertheless, this method allows to obtain very impressive stabilization results and to overcome the limits of pointwise local controls. We are interested in two types of backstepping: *Volterra backstepping* and a more recent approach of *generalized backstepping*.

Volterra Backstepping In order to simplify the problem, the transformation is often sought in the form of a Volterra transformation of the second kind, that is

$$(\mathcal{T}U)(x) := U(x) - \int_0^x K(x,y)U(y)dy, \tag{1.3.10}$$

with $K \in L^2([0,L]\times[0,L];\mathbb{R}^{n\times n})$. These transformations have the nice advantage of being always invertible (from L^2 into L^2). Moreover, they often allow to transform a complicated term in the dynamics into a boundary term (see [142, 162] for instance). This type of backstepping, which we will call *backstepping Volterra* in the following, was introduced in the early 2000s. First from a discretized approach [16, 30] then systematized using a Volterra transformation for scalar parabolic linear systems [224] (see also the now-famous course [162] for a in depth explanation). The hyperbolic linear systems with propagation velocities of the same sign were then studied in [161], then with different signs in [91, 237]. Hyperbolic nonlinear systems in general were then studied in [73, 143]. Many special cases have been studied such as the Korteweg de Vries equations, the Saint-Venant Exner equations, or engineering systems such as the motion of a crane [80, 92, 250]. Volterra backstepping has been a huge success and has resulted in more than a thousand articles [7] since the pioneering work of Krstic et al.

In Chapter 4 we look at a parabolic cross-diffusion system where the control is located at the boundaries and the size of the domain increases with time (in a way that depends *a priori* on the control at the boundaries). This particularity induces several difficulties, for instance it imposes that the kernel K of $\mathcal{T}$ given in (1.3.10) depends on time. Showing that $\mathcal{T}(t,\cdot)$ is invertible and continuous from L^2 into L^2 for each fixed time t is then not sufficient to guarantee an exponential stability. Indeed the exponential stability estimate that we can hope to obtain becomes formally [8]

$$\|\mathbf{u}(t,\cdot)\|_{L^2(\Omega(t)} \leq \|\mathcal{T}^{-1}(t,\cdot)\|_{\mathcal{L}(L^2(\Omega(t))}\|\mathcal{T}(0,\cdot)\|_{\mathcal{L}(L^2(\Omega(0))}e^{-\lambda t}\|\mathbf{u}^0\|_{L^2(\Omega(0))}, \tag{1.3.11}$$

where $\Omega(t)$ is the domain considered at time t. Thus, the quantity $\|\mathcal{T}^{-1}(t,\cdot)\|_{\mathcal{L}(L^2(\Omega(t))}$ depends on time and we must guarantee a certain control on it so that it cannot destroy the exponential stability. The asymptotic stabilization of a system where the domain grows with time has been studied in a simpler case in [144, 145], thanks to the considered dynamics and without guaranteeing any bounds on the norm of the transformation, which is one of the main difficulties in our case. We look at the linearized system and we obtain a rapid stabilization result, i.e. exponential stabilization with an arbitrarily large decay rate, and a finite time stabilization result using Volterra backstepping.

7. google scholar count
8. We use the notation $L^2(\Omega(t))$ by a slight abuse of language.

Generalized backstepping For a large number of systems, however, this Volterra backstepping is limiting and does not allow to conclude. This is often the case for systems with an internal control, i.e. when the control is located in the dynamics. To overcome this obstruction, several people have been interested in looking for a more general transformation, not limited to Volterra transformations. The goal is to obtain more powerful results. Of course, this also means that we are again confronted with the difficulties inherent to backstepping, in particular the question of invertibility that Volterra transforms allowed to avoid. A first approach has been proposed by Coron and Lu in [68, 69] to obtain the fast stabilization of the Korteweg-de-Vries equation and then of the Kuramoto-Sivashinsky equation in 2014 and has been adapted in many cases: the Schroedinger equation, a degenerate parabolic equation, etc. A slightly different approach using also a generalized backstepping has been introduced to deal with the case of autonomous or non-autonomous balance laws in [65, 66] but in this case the control is located at the boundaries, which significantly changes the approach.

In Chapter 5 we show how to adapt this backstepping to the heat equation on the torus with internal scalar controls, as well as to the viscous Burgers equation, and we obtain sharp bounds on the considered spaces. We show in particular that the same feedback operator can be used to stabilize the system rapidly in a continuum of H^s spaces (see Remark 5.3.1). We empasize that the controls here are scalar, which means that they cannot depend on the space variable x but only on the time variable t and this complicates the problem.
This new backstepping method relies on several ingredients: the existence of an orthonormal eigenvector basis for the considered differential operator; a condition on the transformation and on the control operator allowing to have a relatively explicit form of the transformation (depending on the control operator) along this basis; the equivalence for a linear operator between being invertible and mapping an orthonormal basis into a Riesz basis; and a transformation which can be separated between a simple invertible part and a part which reduces to a quadratic perturbation in the considered norm. This is called the "quadratically close" behavior in the sense that the image family of the transformation is quadratically close to an orthonormal family (see Definition 5.3.1).
This last point is both central and, unfortunately, limiting. Indeed, it only allows us to study differential operators whose eigenvalues grow to infinity like n^α with $\alpha > 3/2$. This makes inaccessible some systems like the water wave system [4, 5, 167, 168] which corresponds exactly to the critical case $\alpha = 3/2$. The question of the rapid stabilization of the linearized water-waves system using a backstepping method was an open question for several years [9] and finally solved in [110], that we discuss below.

In Chapter 6 we present a way to change this method to avoid requiring the quadratically close behavior. This new method of "compactness-duality" involves showing the compactness of some operators and using, among other things, a duality between ω-independence in H^r and density in H^{-r} to obtain a Riesz basis. An iteration then allows to isolate the singular part of the control operator and to obtain enough regularity so that the transformation exists and, later, is invertible in the desired space(s). This new method allows not only to solve the open question of the linearized water waves system but also to make the method work beyond the critical threshold $\alpha = 3/2$ for any $\alpha > 1$. Again the same feedback allows to obtain stability in a continuum of norms H^s, and more precisely for any $s \in (1-\alpha, \alpha-1)$. These bounds are sharp in the sense that the closed-loop system is not even well posed for $s = (1-\alpha)$.

Finally, there remains the case of the hyperbolic systems for which $\alpha = 1$. In this case, the sharp bound on $(1-\alpha, \alpha-1)$ means that the previous method is doomed in the sense that the feedback obtained will never work. However, in particular cases it is possible to do otherwise. Several works exist in the scalar case with a linear transport equation [256–258], where a very explicit transformation can be found in the form of a convolution, and which the author shows to be a solution to the problem. This is done, among other things, thanks to the Dirichlet convergence theorem. Motivated by this example, we show in Chapter 7 how it is possible to apply a generalized backstepping method for a more complicated hyperbolic system: the Saint-Venant system which models a water-tank. The lack of quadratically close behavior or compactness is compensated by a different target system which allows to show directly the basic properties of Riesz basis

9. The first mention of which I am aware is in this 2017 College de France lecture [56]

without going through the perturbation of an orthonormal family; together with the generalization of an equiconvergence result of Komornic [159, 160] which allows to find a solution to the condition on the control operator. Whether it is possible to adapt this method to hyperbolic linear systems in general, and in particular the crucial step of the target system, is an open question.

It should be noted that there are still many open questions in this area, either for hyperbolic systems in general, non-linear systems, finite time stabilization etc. We list some of these questions at the end of Part II in Chapter 7.

1.3.3 Part 3

In this part, the focus is no longer on a specific mathematical structure but rather on an application: traffic control. The goal is to use autonomous vehicles to influence the overall traffic flow and to remove, as much as possible, traffic jams. This may seem far from the previous parts and the subject of this thesis, but in fact traffic jams, and more precisely stop-and-go waves, have a deep mathematical cause: they are the manifestations of a steady state which becomes unstable above a certain density of vehicles [76].

This problem can be seen at the microscopic scale, by modeling an N cars system by $2N$ ODEs representing the speeds and accelerations of the cars; or at the macroscopic scale by modeling the traffic with hyperbolic partial differential equations. In this last case, traffic jams correspond to shock phenomena (see [107, 213, 219]). It is therefore necessary to consider the non-regular solutions of these equations. The solutions to be considered are typically BV functions, i.e. with bounded variations (see [78, Section 1.7] for a precise definition). Studying non-regular solutions of nonlinear hyperbolic systems is something natural: these systems are known to spontaneously create discontinuities in certain cases, even when the initial condition is C^∞. Therefore no regular solution is possible at long times. BV solutions have quickly become the natural discontinuous solutions for these systems. The idea is to have a notion of function which is in some way a generalized version of a piecewise C^1 function interspersed with jumps. This leads to look at this class of functions where discontinuities are rare (located on manifolds of co-dimension 1). Note that BV solutions of hyperbolic systems are in general not unique. One therefore add an entropy condition [78, 170], which we will talk about later, to find this uniqueness. Under this condition (and often a smallness assumption on the total variation of the function) it is possible to show the existence and uniqueness of solutions in different hyperbolic frameworks, in the scalar case [163]; in the case of a $n \times n$ homogeneous system [34] (preceded by the work of [118] for the existence) for sufficiently small initial data in one space dimension; for a system on a bounded domain [35]; the local existence and uniqueness in the case of a $n \times n$ system with a potentially large initial condition [172], or in the case of some $n \times n$ systems in several space dimensions [8]. Systems modeling road traffic are typical examples of hyperbolic systems whose solutions must be non-regular, and give rise to many interesting questions from a mathematical point of view (see for instance [24, 49, 50, 120, 169]). For more details on the non-regular solutions of hyperbolic systems one can refer to [33, 36, 78]. One can also note results on control and stabilization of systems where the solutions are BV, for example [117] for the controllability of the isentropic Euler equation, or [62] in the case of the stabilization of a 2×2 system with a control located at the boundaries, [198] in the case where there is both a control at the boundaries and a control in the dynamics, and [23, 199] when the control is only at the boundaries for a scalar system. For an overview of some control and stabilization problems for solutions of nonlinear hyperbolic systems one can look at [26, 116].

In Chapter 8 we look at the microscopic scale and are interested in a circular road where N vehicles are modeled by a system of $2N$ ODEs. These dynamics use the model called *Bando-Follow-the-Leader* [17, 113] which is written:

$$\begin{aligned} \dot{x}_i &= v_i \\ \dot{v}_i &= a(V(x_{i+1} - x_i) - v_i) + b\frac{v_{i+1} - v_i}{(x_{i+1} - x_i)^2}, \end{aligned} \tag{1.3.12}$$

where $i \in \{1, ..., N-1\}$. The control is an autonomous vehicle whose dynamics are given by

$$\begin{aligned} \dot{x}_N &= v_N \\ \dot{v}_N &= u(t, x_N, v_N, x_{N+1}, v_{N+1}) \end{aligned} \tag{1.3.13}$$

with the convention $N+1 = 1$ and where u is a feedback control that depends only on the state of the autonomous vehicle and, potentially, on the state of the vehicle right in front of it. We show that adding a single autonomous vehicle makes the entire traffic locally exponentially stable, regardless of the number of vehicles, despite the fact that the autonomous vehicle observes only itself and the vehicle in front of it. More surprisingly, the decay rate also admits a lower bound independent of the number of vehicles.

In Chapter 9, we are interested in the interaction between road traffic and a vehicle that drives differently from the rest of the traffic, for example an autonomous vehicle in a human traffic. The traffic is modeled by the Generalized-Aw-Rascle-Zhang equations, which form a nonlinear hyperbolic system [103]. These equations encompass several of the most common traffic models and have the particularity of having a linearly degenerate propagation speed (see Section 9.2 or [78] for example for more details). It is expressed as follows:

$$\begin{aligned} &\partial_t \rho + \partial_x \left(\rho\, V(\rho, w)\right) = 0, \\ &\partial_t (\rho w) + \partial_x \left(\rho w\, V(\rho, w)\right) = 0, \end{aligned} \tag{1.3.14}$$

where ρ is the car density at a point in space, w is a driving behavior parameter that corresponds to the driver's speed on an empty road, and V is the speed of the cars in the traffic flow. The interaction with the particular vehicle results in a condition on the flow, similar to those introduced in [9, 87, 88, 177].

$$\rho\left(t, y(t)\right)\left(V(\rho(t, y(t)), w(t, y(t))) - \dot{y}(t)\right) \leq \alpha F(\dot{y}), \tag{1.3.15}$$

where $\alpha \in (0, 1)$ and F denotes

$$F(\dot{y}) = \max_{x \in [0, \rho_{\max}],\ w \in [w_{\min}, w_{\max}]} \left(\rho(V(x, w) - \dot{y})\right). \tag{1.3.16}$$

Moreover, the dynamics of the autonomous vehicle are given by

$$\dot{y}(t) = \min(V(\rho(t, y(t)^+), w(t, y(t)^+)), V_b). \tag{1.3.17}$$

Note that the system (1.3.14) is a 2×2 nonlinear hyperbolic system which is already well posed in the space of entropic BV solutions [10] [36]. One may then ask why add a flow condition like (1.3.15). The reason is that, as surprising as it may seem, entropic solutions are not the physical solutions for this system. To understand this we need to look at what the notion of entropic solution means. An entropic solution is a solution where the characteristics can disappear in a shock but not appear from a shock. In other words, an entropic solution is a solution where a microscopic point cannot have a macroscopic influence on the system. This is the problem for road traffic: if a particular vehicle drives differently from the general flow and decides to brake suddenly on the highway, it will have a macroscopic influence on the highway [11]. The physical solutions can therefore produce non-classical shocks at the position of the particular vehicle. Hence, one must abandon the framework of entropic solutions at the location of the particular vehicle. But, to guarantee the uniqueness of the solutions it is necessary to introduce a new physical condition which replaces the entropy condition. The condition (1.3.15), originally proposed by Delle-Monache and Goatin in [87, 88] in the framework of the (scalar) Lighthill-Whitham-Richards system, meets this need. It represents the fact that the flow past the particular vehicle is limited by steric hindrance, with the coefficient α representing this hindrance. In Chapter 9 we study the existence of weak solutions to the system (1.3.14)–(1.3.15) and their regularity. We show that for any initial condition $(\rho^0, \omega^0) \in BV(\mathbb{R})$ the system (1.3.14)–(1.3.15) admits a solution in the

10. an entropic solution here refers to a solution which verifies an entropy inequality such as Lax's or Liu's depending on the case. Other notions of entropic solutions exist, we can refer for example to [78].

11. Intuitively the problem comes from the fact that a single vehicle can create a bottleneck because the width of the road has a size of the same order of magnitude as the vehicles

space of BV solutions (see Theorem 9.2.1 for more details). We use for that a wave-front tracking approach, similar to [176, 177] which consists in constructing a solution by approximation, by looking at a sequence of Riemann problems. We first approach the initial condition by a sequence of piecewise constant functions, then we see each discontinuity as a Riemann problem for which we have an explicit solution and we build the approximated solutions. Finally we show that they converge to a function which is a solution of the problem [12].

1.3.4 Part 4

In this part we move away from classical mathematics to focus on the interactions between mathematics and artificial intelligence (AI). From a mathematician point of view, it is common to see these interactions by focusing on what mathematics can bring to AI. Here, we take the opposite view by asking how AI methods can help mathematics. This manifests in two questions:

1. Is a trained neural network able to predict the solution of an advanced mathematical problem?

2. Is a trained neural network able to solve a mathematical problem and to provide a proof ?

Concerning the first question, we see in Chapter 10 and the Section 10.2 that the answer is yes for several mathematical problems. The first works in this domain are very recent and, if several works have been considering teaching arithmetics to neural networks [13] [150, 233, 255] few have studied solving a symbolic maths problem. We can nevertheless mention works such as [7, 11, 255] which have in common the attempt to learn mathematical representations. Trying to teach an AI to predict the solution of a mathematical problem may seem a bit daring at first. The motivation comes from the fantastic performances of AI models in translation, as for example [165] in which the authors obtain an AI able to translate words from one language to another without ever having seen a rosetta stone. The idea is to see a mathematical problem as a translation problem, where statements are translated into solutions [14].

In [164] the authors train a neural network to guess explicit solutions to differential equations. In [46] we study more complicated problems from control theory and in [47] problems from computational biology. These works are obtained by using Transformers, an architecture introduced in [236] which is characterized by its attention mechanism and its efficiency for translation problems. The mathematical problems considered are difficult to solve, in the sense that they cannot be solved simply by interpolation and therefore require a kind of "understanding" of the problem. We try in particular to predict the answer to the following questions:

— Is a finite dimensional nonlinear system exponentially stable (or exponentially unstable) ? If so, what is the speed of convergence (resp. divergence) ?

— For a given finite dimensional system, is the associated linearized system controllable ? If so, what would be a stabilizing feedback ?

One can notice that these questions are sometimes qualitative ("yes" or "no") and sometimes quantitative with numerical answers (stabilizing feedback for a given system, convergence speed).

Note that the neural network has no mathematical knowledge prior to the training. The network sees everything that is given to it as a sequence of characters: to the network $f(x)$ or $1+1$ are simply token sequences like "f", "(", "x", ")", and "1", " + ", "1" and it does not know any relation between them (not even $1+1=2$). As for its answer, it is also given in the form of a sequence of tokens: "0" or "1" for questions whose answer is qualitative and a more developed sequence of tokens when the answer is numerical and/or symbolic.

As we will see, after training the neural network can predict the answer to these problems with a very high accuracy. We can notice several interesting things: the first one is that a neural network trained on systems with between 2 and 5 variables still has a good accuracy when it tries to predict the result on a system with

12. this is the main difficulty here

13. and have shown that neural networks have difficulty with arithmetics.

14. obviously for many problems this translation is not bijective at all and therefore very asymmetric, unlike translations between two languages.

6 variables even though it has never seen a system with 6 variables and 6 equations before (in particular it has never seen the variable x_6). The second is that, for several of these questions, a mathematician would start by looking at the linearized system, yet knowing the linearized system does not seem to help the neural network [46].

The second question, "Is a trained neural network capable of solving a mathematical problem and providing a proof?", is by far the most difficult. In the Section 10.3 we see that the answer is yes, at least partially. We see how a neural network is trained to prove statements, first in a purely supervised way (we give examples of statements and associated proofs to the neural network during the training phase and then we evaluate it on statements that it has never seen), then using an algorithm called *HyperTree Proof Search* that we present in Section 10.3.1. The goal of this algorithm is to explore the set of possible proofs in an intelligent way by combining at the same time an estimation of the best theorem to use (policy model), an evaluation of the difficulty to prove a statement (critic model), and a procedure of expansion and back propagation of the obtained scores in the graph. The two estimations, policy model and critic model, are themselves partly or totally composed of neural networks that are trained in parallel by the data from the proof exploration. The principle of this architecture is reminiscent of AlphaZero [222], a model trained to play chess widely popularized in 2017, as well as some reinforcement learning architectures [193].
We also present a proof environment for equalities and inequalities that we have created in Python both to be able to test and improve our system but also to be able to more easily generate synthetic data, i.e. to generate theorems and proofs in an automated way. Indeed, the models underlying our procedure generally need a lot of data to be trained, but the existing data of theorems and proofs are human tabulated and therefore rare.
After training, the model is able to prove high school level exercises and sometimes more: it proves for example two problems from the International Mathematical Olympiad. In the Section 10.3.2 we give examples of proofs from the neural network. For example the proof that 7 never divides $2^n + 1$ whatever $n \in \mathbb{N}$ is. These results have largely increased the state-of-the-art in automated proofs using AI and it is, at the moment, the most performing model in the automated proof of theorems by Machine Learning techniques.

Part I

Stability of quasilinear inhomogeneous hyperbolic systems.

Chapter 2

Stabilization and ISS of 1-D hyperbolic systems

2.1 Introduction

In this chapter, we look at stability properties of generic 1-D hyperbolic systems defined on a bounded domain. Such a hyperbolic system can be written as follows:

$$\partial_t \mathbf{Y} + F(\mathbf{Y})\partial_x \mathbf{Y} + S(\mathbf{Y}, x) = 0, \tag{2.1.1}$$
$$\mathcal{B}(\mathbf{Y}(t,\cdot), \mathbf{Y}(t,L), \mathbf{Y}(t,0), t) = 0, \tag{2.1.2}$$

where $\mathbf{Y}$ is the state of the system, $F(\mathbf{Y})$ is a diagonalisable matrix with distinct and real eigenvalues, $S(\mathbf{Y}, x)$ is the source term, and $\mathcal{B}$ represents some abstract boundary conditions which will be precised later on. Let us consider a steady state $\mathbf{Y}^*$. As $F(\mathbf{Y}^*)$ is diagonalizable one can define a matrix P such that

$$P(x)F(\mathbf{Y}^*(x))P^{-1}(x) = \Lambda(x), \tag{2.1.3}$$

where $\Lambda(x)$ is a diagonal matrix with coefficients $(\Lambda_i(x))_{i\in\{1,\dots n\}}$. Let us remark that P and Λ both depend on x if only if $\mathbf{Y}^*$ does. Assuming that $\mathbf{Y}^*$ is a regular steady-state, namely $\mathbf{Y}^* \in C^1([0,L])$, and setting $\mathbf{u}(t,x) = P(x)(\mathbf{Y}(t,x) - \mathbf{Y}^*(x))$ the system (2.1.1) becomes

$$\partial_t \mathbf{u} + A(\mathbf{u}, x)\partial_x \mathbf{u} + B(\mathbf{u}, x) = 0, \tag{2.1.4}$$

where

$$\begin{aligned} A(\mathbf{u}, x) &= P(x)F(\mathbf{Y})P^{-1}(x) = P(x)F(P^{-1}(x)\mathbf{u} + \mathbf{Y}^*(x))P^{-1}(x), \\ B(\mathbf{u}, x) &= P(F(\mathbf{Y})(\mathbf{Y}^*_x + (P^{-1})'\mathbf{u}) + S(\mathbf{Y}, x)), \end{aligned} \tag{2.1.5}$$

and in particular $A(\mathbf{0}, x) = \Lambda(x)$, and $B(\mathbf{0}, x) = 0$, since $\mathbf{Y}^*$ is a steady-state. Obviously, the exponential stability of $\mathbf{Y}^*$ for the system (2.1.1) is equivalent to the exponential stability of $\mathbf{u}^* = 0$ for this new system (2.1.4). Moreover, if $B(\mathbf{u}, x) \equiv 0$ (or equivalently $S(\mathbf{Y}, x) \equiv 0$, since in this case $\mathbf{Y}^*$ is a constant and so is P^{-1}), the system is said to be *homogeneous*.

Without loss of generality we can assume that there exists $m \in \{0, ..., n\}$ such that

$$\Lambda_i > 0 \text{ for any } i \in \{1, ..., m\}, \quad \text{and} \quad \Lambda_i < 0 \text{ for any } i \in \{m+1, ..., n\}. \tag{2.1.6}$$

In order to be well-posed, this system requires to impose some boundary conditions at $x = 0$ and/or $x = L$. A way to see this is that the *incoming information* entering the system at each time t should be prescribed in order to have a well-posed system (see [173, 174] for more details). Let us look at our system to understand what this incoming information is. By assumption $F(\mathbf{Y}^*)$ has non-zero and distinct eigenvalues for any x,

and therefore so do $\Lambda(x)$. From (2.1.5), as long as $\mathbf{Y}^*$ and F are continuous, A is continuous with $\mathbf{u}$. As $A(\mathbf{0},x)=\Lambda(x)$, if $\mathbf{u}$ is small enough, $A(\mathbf{u},x)$ has also distinct and non-zero eigenvalues $\lambda_i(\mathbf{u},x)$ which have the same sign as $\Lambda_i(x)$. This means that at $x=0$ we have to impose the quantities that have a positive propagation speed, and at $x=L$ the quantities that have a negative propagation speed (i.e. $\lambda_i(\mathbf{u},x)>0$). This is translated as

$$\begin{pmatrix} u_1(0) \\ ... \\ u_m(0) \\ u_{m+1}(L) \\ ... \\ u_n(L) \end{pmatrix} = \mathcal{U}(t), \tag{2.1.7}$$

where $\mathcal{U}(t)$ is either the control we impose or some boundary conditions given by the physics of the system. Of course the u_i are not exactly the quantities propagating with speed $\lambda_i(\mathbf{u},x)$ given that $A(\mathbf{u},x)$ is not diagonal. But, assuming that we are close enough to the steady-state $\mathbf{u}^*=0$, the perturbations are small, and the u_i are close to the eigenvectors of $A(\mathbf{u},x)$. In this case, the boundary condition (2.1.7) still impose the incoming information and allow the system to be well-posed (see [20, Chapter 6]). In the following we will denote $\mathbf{u}_+=(u_1,...,u_m)^T$ the vector of components associated to positive propagation speeds, and $\mathbf{u}_-=(u_{m+1},...,u_n)^T$ the vector of components associated to negative propagation speeds. As a consequence (2.1.7) can be written in the compact notation

$$\begin{pmatrix} \mathbf{u}_+(t,0) \\ \mathbf{u}_-(t,L) \end{pmatrix} = \mathcal{U}(t), \tag{2.1.8}$$

A usual example of boundary conditions is given by

$$\begin{pmatrix} \mathbf{u}_+(t,0) \\ \mathbf{u}_-(t,L) \end{pmatrix} = G \begin{pmatrix} \mathbf{u}_+(t,L) \\ \mathbf{u}_-(t,0) \end{pmatrix}, \tag{2.1.9}$$

which expresses that the incoming information is a function of the *output information* (that is, the information in $x=L$ for quantities with positive propagation speeds and in $x=0$ for quantities with negative propagation speeds). Such boundary conditions are the most commonly used when looking at the exponential stabilization of such systems (see for instance [19, 94, 130, 132, 173, 174]). There are the simplest example of a feedback loop, they also require only little measurements of the system (no need to measure inside the system, but only at the boundaries), and are simple to implement in practice.

Let us now recall the definition of exponential stability

Definition 2.1.1. Let X be a Banach space endowed with the norm $\|\cdot\|_X$, that we refer to as the X norm in the following. The steady-state $\mathbf{u}^*=0$ of the system (2.1.4), (2.1.8) is exponentially stable for the X norm if there exist $\gamma>0$, $\eta>0$, and $C>0$ such that for every $\mathbf{u}^0\in X$ satisfying the compatibility conditions [1] and $\|\mathbf{u}^0\|_X\leq\eta$, the Cauchy problem (2.1.4), (2.1.8), ($\mathbf{u}(0,x)=\mathbf{u}^0$) has a unique solution in $C^0([0,+\infty),X)$ and

$$\|\mathbf{u}(t,\cdot)\|_X \leq Ce^{-\gamma t}\|\mathbf{u}^0\|_X,\ \forall t\in[0,+\infty). \tag{2.1.10}$$

Moreover, if $\eta=+\infty$ the system is said *globally* exponentially stable.

We can make several remarks:

— One can note that specifying the norm $\|\cdot\|_X$ considered is needed. Indeed, the exponential stability for different norms are not equivalent for infinite-dimensional systems [70].

— When nothing can be said about η, this definition is only a *local exponential stability*. The reason to work with such a notion is that quasilinear hyperbolic systems are known for spontaneously generating shock waves even if the initial condition is very regular, which makes the global exponential stabilization by boundary controls (and even the global well-posedness) impossible in general. However, we will see that when the system is semilinear we can obtain a global well-posedness provided

1. this very formal given that X is not precised. For the H^p or C^p norm, this correspond to the p order compatibility condition given in [19, (4.136) (see also (4.137)-(4.142))]

a Lipschitz condition on the source term. This is the object of Chapter 3 (see in particular Theorem 3.2.1 and 3.2.2).

When the system is linear, the exponential stability of such a system can be found using spectral mapping theorems (see for instance [180, 194, 214]). Such theorems allow to link the eigenvalues of the differential operator $-A\partial_x - BI_d$ (defined on a domain taking into account the boundary conditions), to the stability of the overall system. This reduces the question of stability to an eigenvalue problem. In particular this allows to use many spectral tools. Unfortunately, for quasilinear systems this approach cannot work: in general the exponential stability of the linearized system does not give any information on the exponential stability –even local– of the quasilinear system (see [70] for a counter-example). Thus other tools are needed.

The study of the exponential stability of nonlinear hyperbolic systems of the form (2.1.4)–(2.1.9) goes back to the pioneering work of [126] in 1984 where Li and Greenberg looked at the exponential stability of 2×2 homogenous systems in the C^1 norm. Their method relies on a careful analysis of the solution along the characteristics. This method was later generalized by Qin, Zhao, De Halleux et al. (among others, [84, 211, 260]) and allowed to study any quasilinear hyperbolic system of the form (2.1.4)–(2.1.9) in the C^1 norm when the system is homogeneous (and even C^p norm for $p \geq 1$). An alternative proof using a so-called basic quadratic Lyapunov function, or energy-like Lyapunov function, was shown in [58]. The case of inhomogeneous systems was treated in [132], and was the first work of my PhD.

Other norms, often simpler to handle than sup-norms, have also been considered. In [18, 19] the authors deal with the case of a generic system for the H^p norm, for $p \geq 2$, again by using basic quadratic Lyapunov functions. Using a time-delay approach the authors of [70] found a result for a generic homogeneous system in the $W^{2,p}$ norm. Numerous particular cases or specific examples have also been considered, for instance in fluid mechanics [20, 37, 59, 94, 95, 127, 128], road traffic [24, 100, 235, 253, 254], manufacturing [48, 74, 93, 221]. In many of these works, the key tool is to show the existence of a well-chosen basic quadratic Lyapunov function whose definition is given as follows:

Definition 2.1.2. For a system of the form (2.1.4)–(2.1.8) and a Sobolev space $W^{l,p}(0,L)$ where $(l,p) \in \mathbb{N} \times \mathbb{N}\backslash\{0\} \cup \{+\infty\}$, we call *basic quadratic Lyapunov function* for the $W^{l,p}$ norm a function $V \in C^0(W^{l,p}(0,L),\mathbb{R})$ such that

1. $$V(\mathbf{U}) = \sum_{n=0}^{l} \|F(\cdot)E(\mathbf{U},\cdot)D_n\mathbf{U}\|_{L^p(0,L)}, \quad \forall\, U \in W^{l,p}(0,L), \tag{2.1.11}$$
   where $E(\mathbf{U},x)$ is a matrix diagonalizing $A(\mathbf{U},x)$, $F = \mathrm{diag}(f_1,...f_n)$ with f_i positive C^1 functions on $[0,L]$, D_n is the operator defined iteratively by $D_0\mathbf{U} = \mathbf{U}$, and
   $$D_{n+1}\mathbf{U} = \partial_{\mathbf{U}}(D_{n-1}\mathbf{U})(-A(\mathbf{U},x)\partial_x\mathbf{U} - B(\mathbf{U},x)). \tag{2.1.12}$$
2. There exists $\delta > 0$ and $\gamma > 0$ with which, for any $T > 0$ and along any regular solution $\mathbf{u}$ on $[0,T]$ of (2.1.4)–(2.1.8) satisfying $\|\mathbf{u}(t,\cdot)\|_{W^{l,p}} \leq \delta$, one has in a distributional sense [2]
   $$\frac{dV(\mathbf{u}(t,\cdot))}{dt} \leq -\gamma V(\mathbf{u}(t,\cdot)), \quad \text{for all } t \in [0,T]. \tag{2.1.13}$$

These Lyapunov functions can be also referred to as energy-like Lyapunov functions because, when looking at the basic quadratic Lyapunov functions for the L^2 or H^p norms, they look similar to physical energies (see [85, 134, 146]). In particular several physical quantities, such as the mechanical energy, or the physical entropies, have the form of a basic quadratic Lyapunov function (see for instance [61, 195]).

In practical applications, the system can sometimes be subject to unknown disturbances, which can be the results of external events, errors of the model, etc. The system (2.1.4), (2.1.9) then becomes

$$\partial_t\mathbf{u} + A(\mathbf{u},x)\partial_x\mathbf{u} + B(\mathbf{u},x) + \mathbf{d}_1(t,x) = 0, \tag{2.1.14}$$

2. see [121, Definition 3.2.10] for a definition of distributional inequalities

$$\begin{pmatrix} \mathbf{u}_+(t,0) \\ \mathbf{u}_-(t,L) \end{pmatrix} = G \begin{pmatrix} \mathbf{u}_+(t,L) \\ \mathbf{u}_-(t,0) \end{pmatrix} + \mathbf{d}_2(t), \tag{2.1.15}$$

where $\mathbf{d}_1$ are the internal disturbances and $\mathbf{d}_2$ the boundary disturbances. In this case, the previous steady-state $\mathbf{u}^* = 0$ is not a solution of the system anymore and there is usually no hope of stabilizing it. However, it could be interesting to see what is the error we make when trying. In other words, how robust the exponential stability is with respect to $\mathbf{d}_1$ and $\mathbf{d}_2$. This leads to introducing a slightly more general notion than the exponential stability: the *Input-to-State Stability* (or ISS). The notion of ISS was first introduced by Sontag in 1989 [225] for finite dimensional systems. It was later extended to time delay systems, and then generalized to PDEs (see [151, Chapter 1] for more details). The goal is to show an estimate of the form

$$\|\mathbf{u}(t,\cdot)\|_X \leq C_1 e^{-\gamma t}\|\mathbf{u}_0\|_X + C_2\left(\|\mathbf{d}_1\|_{X_t\times X} + \|\mathbf{d}_2\|_{X_t}\right), \tag{2.1.16}$$

where $\|\cdot\|_X$ is the (formal) norm considered in space, $\|\cdot\|_{X_t}$ the norm considered in time and $\|\cdot\|_{X_t\times X}$ is the norm considered in time and space (again formally), and $\gamma > 0$. Note that when $\mathbf{d}_1 \equiv \mathbf{d}_2 \equiv 0$, that is to say when there is no disturbance, we recover the exponential stability. In this sense, this ISS is a generalization of the exponential stability. The estimate (2.1.16) would in fact guarantee the so-called *exponential ISS*, which is a particular case of a less restrictive ISS notion requiring

$$\|\mathbf{u}(t,\cdot)\|_X \leq \sigma(\|\mathbf{u}_0\|_X, t) + \|\alpha(\|\mathbf{d}_1(s,\cdot)\|_X + |\mathbf{d}_2(s)|)\|_{X_t}, \tag{2.1.17}$$

where α belongs to $\mathcal{K}$, the space of strictly increasing functions $\mathbb{R}_+ \to \mathbb{R}_+$ such that $\alpha(0) = 0$; σ is a function such that for any $t \in \mathbb{R}_+$, $x \to \sigma(x,t)$ belongs to $\mathcal{K}$ and for any $x \in \mathbb{R}_+$, $t \to \sigma(x,t)$ is non-increasing and satisfies $\lim_{t\to+\infty}\sigma(x,t) = 0$. A more detailed review about the genesis of ISS notions for PDEs and some variations about these notions can be found in [151, Chapter 1]. In what follows, we will only consider the exponential ISS and we will even consider a stricter notion that requires the influence of the past disturbances to decrease exponentially with time. This is the so-called *exponential ISS with fading memory* defined as follows for the C^p and H^p norms

Definition 2.1.3. A system of the form (2.1.14), (2.1.15) is exponentially ISS with fading memory for the C^q norm (resp. H^p norm) if there exist positive constants $C_1 > 0$, $C_2 > 0$, $\gamma > 0$, and $\eta > 0$ such that, for any $T > 0$, for any $\mathbf{u_0} \in C^q([0,L];\mathbb{R}^n)$ (resp. $\mathbf{u}_0 \in H^p((0,L);\mathbb{R}^n)$) satisfying the q-th order (resp. $p-1$ order) compatibility conditions [3], with

$$\begin{aligned} \|\mathbf{u}_0\|_{C^q} \leq& \eta \quad (\text{resp. } \|\mathbf{u}_0\|_{H^p} \leq \eta) \\ \|\mathbf{d}_2\|_{C^q([0,T])} + \|\mathbf{d}_1\|_{C^q([0,T]\times[0,L])} \leq& \eta, \quad (\text{resp. } \|\mathbf{d}_2\|_{H^p} + \|\mathbf{d}_1\|_{H^p([0,T])\times H^p([0,L]))} \leq \eta), \end{aligned} \tag{2.1.18}$$

there exists a unique solution $\mathbf{u} \in C^0([0,T], C^q([0,L]))$, (resp. $\mathbf{u} \in C^0([0,T]; H^p(0,L))$) such that

$$\begin{aligned} \|\mathbf{u}(t,\cdot)\|_{C^q} \leq C_1 e^{-\gamma t}\|\mathbf{u}_0\|_{C^q} &+ C_2\left(\sum_{k=0}^{q}\sup_{\tau\in[0,t]}\left(e^{-\gamma(t-\tau)}|\mathbf{d}_2^{(k)}(\tau)|\right)\right) \\ &+ C_3\left(\sup_{(\tau,x)\in[0,t]\times[0,L]}\left(e^{-\gamma(t-\tau)}|\partial_t^q\mathbf{d}_1(\tau,x)|\right)\right. \\ &\left.+\sum_{k_1+k_2\leq q-1}\sup_{(\tau,x)\in[0,t]\times[0,L]}\left(e^{-\gamma(t-\tau)}|\partial_t^{k_1}\partial_x^{k_2}\mathbf{d}_1(\tau,x)|\right)\right), \\ \left(\text{resp. } \|\mathbf{u}(t,\cdot)\|_{H^p} \leq C_1 e^{-\gamma t}\|\mathbf{u}_0\|_{H^p} &+ C_2\left(\sum_{k=0}^{p}\|e^{-\gamma(t-\tau)}\mathbf{d}_2^{(k)}(\tau)\|_{L^2(0,t)}\right) \\ &+ C_3\left(\|e^{-\gamma(t-\tau)}\partial_t^p\mathbf{d}_1(\tau,x)\|_{L^2((0,t)\times(0,L))}\right. \\ &\left.+\sum_{k_1+k_2\leq p-1}\|e^{-\gamma(t-\tau)}\partial_t^{k_1}\partial_x^{k_2}\mathbf{d}_1(\tau,x)\|_{L^2((0,t)\times(0,L))}\right)\Big). \end{aligned} \tag{2.1.19}$$

3. see [19, 4.5.2] for a definition of such compatiblity conditions

Moreover, if $\eta = +\infty$ the system is said *globally* exponentially ISS with fading memory.

These estimates may seem complicated but each term is simple: the term proportional to C_1 describe the exponential stability of the system and the terms proportional to C_2 and C_3 describe the ISS behavior and the influence of the disturbances. In each of these terms, the term $e^{-\gamma(t-\tau)}$ ensures that the past disturbances decay exponentially with time. We will refer to C_2 and C_3 as the *ISS gains*.

While exponential stability of nonlinear 1-D hyperbolic systems has been widely studied in the last 40 years [4], ISS of these systems has only received some attention recently and relatively few results exist. One can look at [22, 151] for an overview of ISS for these systems.
In Section 2.2 we present a study of the ISS of a general nonlinear hyperbolic systems in the C^p norms. We show that the sufficient conditions of exponential stability found in [58, 84, 126, 211, 260] and [132] are also sufficient conditions of ISS. But, surprisingly, we are unable to find an ISS Lyapunov function in the sense of Definition 2.2.1 below, because the basic quadratic Lyapunov function we consider cannot satisfy the differential inequality (2.2.2). We also show that these results are an improvement to the most up to date ISS conditions for these systems (see [151]). Later, in Chapter 3 we present a study of a semilinear system where the source term is potentially nonlocal but Lipschitz (in L^2) and the boundary term is Lipschitz. We show that it is possible to obtain an exponential stability result for the L^2 norm, provided some conditions on the source term and the boundary term (while this is impossible in general when the source term is not Lipschitz). Moreover this exponential stability result is global. Finally, we show that this can be extended to a global ISS result.

2.2 ISS of general hyperbolic system for the C^p norm

This section is based on the results of [22], a collaboration with Georges Bastin and Jean-Michel Coron. To study the ISS of nonlinear systems, a practical approach is to use an ISS-Lyapunov function, in a way similar to what is done for exponential stability. These ISS-Lyapunov functions are defined as follows.

Definition 2.2.1. Let $p \geq 0$. The function V is called ISS-Lyapunov function for the C^p norm (resp. H^p norm) of the system (2.1.14)–(2.1.15) if there exists positive constants C, c, η and γ such that for any $T > 0$, any $\|\mathbf{d}_1\|_{C^p \times C^p} + \|\mathbf{d}_2\|_{C^p} \leq \eta$, and any solution $\mathbf{u} \in C^1([0,T], C^p([0,L]))$ of (2.1.14)–(2.1.15) satisfying $\|\mathbf{u}(t,\cdot)\|_{C^p} \leq \eta$ we have

(i) Equivalence with the norm

$$\begin{aligned} &c(\|\mathbf{u}(t,\cdot)\|_{C^p} + \sum_{k_1+k_2\leq p-1} \|\partial_t^{k_1}\partial_x^{k_2}\mathbf{d}_2(t,\cdot)\|_{C^0}) \\ &\leq V(\mathbf{u}(t,\cdot)) \leq C(\|\mathbf{u}(t,\cdot)\|_{C^p} + \sum_{k_1+k_2\leq p-1} \|\partial_t^{k_1}\partial_x^{k_2}\mathbf{d}_2(t,\cdot)\|_{C^0}), \end{aligned} \tag{2.2.1}$$

(ii) Lyapunov estimate

$$\begin{aligned} \frac{dV(\mathbf{u}(t,\cdot))}{dt} \leq -\gamma V(\mathbf{u}(t,\cdot)) + C\Big(&\sum_{k=0}^{p} |\mathbf{d}_2^{(k)}(t)| + \sum_{k_1+k_2\leq p-1} \|\partial_t^{k_1}\partial_x^{k_2}\mathbf{d}_1(t,\cdot)\|_{C^0([0,L])} \\ &+ \|\partial_t^p \mathbf{d}_1(t,\cdot)\|_{C^0([0,L])}\Big), \quad \forall\, t \in [0,T]. \end{aligned} \tag{2.2.2}$$

Remark 2.2.1. *A similar definition holds for the H^p norm by replacing C^p by H^p, and C^0 by L^2 in the equivalence with the norm.*

Of course, if there exists an ISS-Lyapunov function for the C^p norm (resp. H^p norm), the system is exponentially ISS [5] for the C^p norm (resp. H^p norm). A natural question arises:

4. but with still many questions unanswered
5. and in fact even exponentially ISS with fading memory

"Are the basic quadratic Lyapunov functions, given by (2.1.2), also ISS-Lyapunov functions?"

In which case all the exponential stability results obtained for these functions would automatically extend to the ISS. Until recently this question was largely open. If this question arises, it is because very few results exists for the ISS of hyperbolic systems. In fact the most advanced result until two years ago was the result of [151] recalled below (see Theorem 2.2.4) which used a *small gain analysis to obtain* the ISS in sup norms for linear 2×2 systems. Our hope was that the basic quadratic Lyapunov functions could be an additional tool for ISS, potentially more powerful than the current existing methods. And they happened to be. An answer to this question was given in [106] for the stabilization in L^2 norm: this work shows that for linear systems, the basic quadratic Lyapunov functions for the L^2 norm are also basic ISS-Lyapunov functions (for the L^2 norm). This is also true for nonlinear systems considered in the H^2 norm (see [245] and also [22]) and, in fact, in the H^p norm for any $p \geq 2$. In sup norm, a more interesting phenomena appear: we showed in [22] that we are unable to obtain the decrease (2.2.2) with basic C^1 Lyapunov functions because basic C^1 Lyapunov functions are not ISS-Lyapunov functions in the sense of Definition 2.2.1. Nevertheless, it is still possible to show the exponential ISS for the C^1 norm with these functions, and the sufficient conditions we obtain on the system and the control are the same as the conditions obtained in [132] for the exponential stability (and [58, 58, 84, 84, 126, 211, 260] in the homogeneous case). This comes from the fact that a basic quadratic Lyapunov function V still satisfy an estimate of the form

$$\begin{aligned} V(\mathbf{u}(t,\cdot),t) \leq e^{-\gamma(t-s)}V(s) + C_2\left(\sum_{k=0}^{p}\sup_{\tau\in[s,t]}\left(e^{-\gamma(t-\tau)}|\mathbf{d}_2^{(k)}(\tau)|\right)\right) \\ + C_3\left(\sup_{(\tau,x)\in[s,t]\times[0,L]}\left(e^{-\gamma(t-\tau)}|\partial_t^p\mathbf{d}_1(\tau,x)|\right)\right. \\ \left. + \sum_{k_1+k_2\leq p-1}\sup_{(\tau,x)\in[s,t]\times[0,L]}\left(e^{-\gamma(t-\tau)}|\partial_t^{k_1}\partial_x^{k_2}\mathbf{d}_1(\tau,x)|\right)\right) \quad \forall\, s\in[0,T], \end{aligned} \tag{2.2.3}$$

despite not satisfying (2.2.2).

Let us get into more details. For $k \in \mathbb{N}\setminus\{0\}\cup\{+\infty\}$, we define ρ_k as follows

$$\rho_k(M) = \inf\{\|\Delta M\Delta^{-1}\|_k \mid \Delta \in \mathcal{D}_n^+\}. \tag{2.2.4}$$

where $\mathcal{D}_n^+$ is the space of diagonal matrix with positive entries and,

$$\|M\|_k = \sup_{\|\xi\|_k=1}(\|M\xi\|_k), \quad \forall M \in M_n(\mathbb{R}). \tag{2.2.5}$$

where $\|\xi\|_k$ is the usual k-norm for a vector of $\mathbb{R}^n$ and $M_n(\mathbb{R})$ the space of square matrices of size n on $\mathbb{R}$. The first result we show in [22] is the following

Theorem 2.2.1. *Let a homogeneous quasilinear hyperbolic system be of the form (2.1.14), (2.1.15), with A and G of class C^p, with $p \in \mathbb{N}\setminus\{0\}$. If*

$$\rho_\infty(G'(0)) < 1, \tag{2.2.6}$$

then the system is exponentially ISS with fading memory for the C^p norm.

One can remark that this holds irrespective of the system dynamics, and the condition (2.2.6) only depends on the boundary conditions (or the boundary control). This condition is the same as the one found in [58, 84, 211, 260] for the exponential stability in the C^1 norm. When the system is inhomogeneous we show the following:

Theorem 2.2.2. *Let a quasilinear hyperbolic system be of the form (2.1.14)–(2.1.15) with A, B and G of class C^p, with $p \in \mathbb{N}\setminus\{0\}$. Let us denote $M(x) = \partial_{\mathbf{u}}B(0,x)$. Let us assume that the system of differential inequalities*

$$\Lambda_i(x)f_i'(x) \leq -2\left(-M_{ii}(x)f_i(x) + \sum_{k=1,k\neq i}^{n}|M_{ik}(x)|\frac{f_i^{3/2}(x)}{\sqrt{f_k(x)}}\right), \quad i\in\{1,...,n\}, \tag{2.2.7}$$

has a solution $(f_1, ..., f_n) : [0, L] \to \mathbb{R}^n$ *such that* f_i *are positive functions on* $[0, L]$ *for any* $i \in \{1, ..., n\}$ *and that there exists a diagonal matrix* Δ *with positive coefficients such that*

$$\|\Delta G'(0)\Delta^{-1}\|_\infty < \frac{\inf_i \left(\frac{f_i(l_i)}{\Delta_i^2}\right)}{\sup_i \left(\frac{f_i(L-l_i)}{\Delta_i^2}\right)}, \quad i \in \{1, ..., n\}, \tag{2.2.8}$$

where $l_i = L$ *if* $\Lambda_i > 0$ *and* $l_i = 0$ *otherwise. Then the system* (2.1.14)–(2.1.15) *is exponentially ISS with fading memory for the* C^p *norm.*

This time, the conditions depend intrinsically on the system dynamics: there is a first condition (2.2.7) where the control is not involved. Moreover, when this condition is satisfied, the solutions of (2.2.7) are used in the boundary condition (2.2.8). Note that condition (2.2.7) is not always satisfied and relates the source term and the length of the domain. Indeed, as the right-hand side is nonlinear with the f_i, a solution could either explode or reach 0 in finite length. Thus, this condition amounts to giving a bound on the length of the domain or on the amplitude of the source term. This phenomena is frequent for inhomogeneous hyperbolic systems and also appears in other norms (see for instance [18, 19, 58, 132, 133]). Finally, Conditions (2.2.7)–(2.2.8) are also the same conditions as the ones found in [132] for the exponential stability. If the system is semilinear, these theorems can be further extended:

Proposition 2.2.3 (Case of semilinear systems)**.** *If the system* (2.1.14) *is semilinear (i.e.* $A(\mathbf{u}, x) = A(x)$*), then Theorems 2.2.1 and 2.2.2 also hold true for* $p = 0$*.*

Let us look now at the comparison between these conditions and the existing conditions found in [151] using a small-gain analysis. To be in the same framework as [151], we consider a linear 2×2 system of the form

$$\partial_t \begin{pmatrix} u_1(t,x) \\ u_2(t,x) \end{pmatrix} + \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \partial_x \begin{pmatrix} u_1(t,x) \\ u_2(t,x) \end{pmatrix} + \begin{pmatrix} 0 & a(x) \\ b(x) & 0 \end{pmatrix} \begin{pmatrix} u_1(t,x) \\ u_2(t,x) \end{pmatrix} = 0 \tag{2.2.9}$$

$$\begin{pmatrix} \mathbf{u}_1(t,0) \\ \mathbf{u}_2(t,1) \end{pmatrix} = \begin{pmatrix} 0 & k_1 \\ k_2 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_1(t,1) \\ \mathbf{u}_2(t,0) \end{pmatrix} + \mathbf{d}(t), \tag{2.2.10}$$

where $a(x)$ and $b(x)$ are continuous functions in $C^0([0, L])$, $\Lambda_1 > 0$ and $\Lambda_2 < 0$ are constant speed propagations, k_1 and k_2 are constant parameters, and $\mathbf{d} \in L^\infty(\mathbb{R}_+)$ is the boundary disturbance. We assume without loss of generality [6] that $L = 1$. What is showed in [151] is the following

Theorem 2.2.4 ([151])**.** *Consider a system of the form* (2.2.9)–(2.2.10)*, if there exists* $K > 0$ *such that*

$$\begin{aligned} &(|k_1| + |k_2|)\exp(-K) < 1, \\ &\left(\sqrt{\frac{\exp(2K) - \exp K}{|\Lambda_2|K}}B + \sqrt{|k_2|}\right)\left(\sqrt{\frac{1 - \exp(-K)}{\Lambda_1 K}}A + \sqrt{|k_1|}\right) < 1, \\ &\text{where } A := \max_{0 \le z \le 1} |a(z)\exp(2Kz)| \text{ and } B := \max_{0 \le z \le 1} |b(z)\exp(-2Kz)|, \end{aligned} \tag{2.2.11}$$

then the system (2.2.9)–(2.2.10) *is ISS for the* C^0 *norm.*

In this framework Theorem 2.2.2 can be simplified as follows

6. Indeed, one can always rescale the system by setting $y = x/L$. This is not in contradiction with the fact that (2.2.7) depends strongly on the length of the system, because the propagation speed would change accordingly, and the condition (2.2.7) would remain the same overall.

Corollary 1. *For the system* (2.2.9)–(2.2.10) *with a and b constant, the conditions* (2.2.7)–(2.2.8) *of Theorem 2.2.2 are respectively equivalent to*

$$\begin{aligned} &\textit{(interior condition)} \quad \left(\frac{\pi}{2} - \sqrt{\left|\frac{ab}{\Lambda_1\Lambda_2}\right|}\right) \geq 0, \\ &\textit{(boundary conditions)} \quad |k_1| < \sqrt{\left|\frac{a\Lambda_2}{b\Lambda_1}\right|} \tan\left(\frac{\pi}{2} - \sqrt{\left|\frac{ab}{\Lambda_1\Lambda_2}\right|}\right), \\ &\qquad\qquad |k_2| < \left|\frac{b\Lambda_1}{a\Lambda_2}\right| \left(\tan\left(\operatorname{atan}\left(\sqrt{\left|\frac{b\Lambda_1}{a\Lambda_2}\right|}|k_1|\right) + \sqrt{\left|\frac{ab}{\Lambda_1\Lambda_2}\right|}\right)\right)^{-1}. \end{aligned} \tag{2.2.12}$$

The first thing we can show is the following

Proposition 2.2.5. *Consider the system* (2.2.9)–(2.2.10) *with a and b constant. Suppose there exists $K > 0$ such that* (2.2.11) *of Theorem 2.2.4 holds. Then the conditions* (2.2.12) *of Corollary 1, and consequently the two conditions* (2.2.7)–(2.2.8) *of Theorem 2.2.2, are satisfied.*

What is interesting is that this is a strict implication and the converse does not hold in general. In fact, the converse only holds when $a \equiv b \equiv 0$, and in this case the conditions of Theorem 2.2.4 and of Corollary 1 are both equivalent to $|k_1 k_2| < 1$.

Another interesting way to consider these results is to look at the lower bound on the *maximal length of ISS* they give. The maximal length of ISS describes how large the length of the domain can be for the ISS to hold with a given dynamics, irrespectively of the boundary conditions (2.1.15). Formally it is defined as

Definition 2.2.2. Let a system be of the form (2.1.14). We call *maximal length of ISS* the largest length $L_{\max} \geq 0$ such that for any $L \in [0, L_{\max})$, there exists G such that the system (2.1.14), (2.1.15) defined on $[0, L]$ is ISS.

Looking at Theorem 2.2.2, one can see right away that $L_{\max}$ is strictly positive (possibly infinite). In practice Theorem 2.2.2, allows to have a numerical lower bound on $L_{\max}$ as follows: for any $C > 0$ let $L(C) \in (0, +\infty]$ be the largest length such that the maximal solution of

$$\begin{cases} \Lambda_i(x) f_i'(x) = -2\left(-M_{ii}(x) f_i(x) + \displaystyle\sum_{k=1, k\neq i}^{n} |M_{ik}(x)| \frac{f_i^{3/2}(x)}{\sqrt{f_k(x)}}\right), \ x \geq 0, \\ \qquad f_i(0) = C \text{ if } 1 \leq i \leq m, \\ \qquad f_i(0) = 0 \text{ if } m+1 \leq i \leq n, \end{cases} \tag{2.2.13}$$

is defined and remains non-negative on $[0, L(C))$. Then $L(C)$ is a nondecreasing function of $C > 0$ and, for every $C > 0$, and $L(C) \leq L_{\max} \in (0, +\infty]$. Therefore a lower bound of $L_{\max}$ can be estimated in practice by choosing $C > 0$ large enough and by solving numerically system (2.2.13) in order to estimate $L(C)$.

For a 2×2 system, (2.2.13) can be simplified into a linear system (see [133] for more details) and gives a lower bound on $L_{\max}$ that can be shown to be at least as good as the existing result of Theorem 2.2.4. This is given by the following proposition, observing that, both in Theorem 2.2.4 and Corollary 1, the largest length L permitted for a given system is achieved for $k_1 = k_2 = 0$.

Proposition 2.2.6. *Let $k_1 = k_2 = 0$ and assume that the condition* (2.2.11) *of Theorem 2.2.4 holds. Then the conditions* (2.2.7)–(2.2.8) *of Theorem 2.2.2 are satisfied.*

We give some ideas of the proof of Theorem 2.2.1–2.2.2 and Proposition 2.2.5–2.2.6 in the following section.

2.2.1 Ideas of the proof

2.2.1.1 ISS of a 1-D hyperbolic system

We start by the proof of Theorem 2.2.1–2.2.2 for which we only give the main ideas. To simplify, we only consider the C^1 norm, considering the C^q norm with $q > 1$ can be done similarly by a system augmentation (see [132] or [22]). We have the following theorem (see [244]).

Theorem 2.2.7 (Well-posedness). *For any $T > 0$ there exist $C_1(T) > 0$ and $\delta(T) > 0$ such that, for every $\mathbf{d}_1 \in C^1([0,T], C^0([0,L])$, $\mathbf{d}_2 \in C^1([0,T])$, $\mathbf{u_0} \in C^1([0,L];\mathbb{R}^n)$ satisfying the first order compatibility conditions[7] and such that $\|\mathbf{u}_0\|_{C^1} + \|\mathbf{d}\|_{C^1} + \|\mathbf{d}\|_{C^1} \leq \delta(T)$, the system (2.1.14), (2.1.15), with A and B of class C^1, has a unique solution on $[0,T]\times[0,L]$ with initial condition $\mathbf{u_0}$. Moreover one has:*

$$\begin{aligned}\|\mathbf{u}(t,\cdot)\|_{C^1} \leq C_1(T)\Bigg(&\|\mathbf{u}(0,\cdot)\|_{C^1} + \sup_{\tau\in[0,t]}(|\mathbf{d}_2(\tau)|)\\ &+ \sup_{\tau\in[0,t]}(|\mathbf{d}_2'(\tau)|) + \sup_{(\tau,x)\in[0,t]\times[0,L]}(|\mathbf{d}_1(\tau,x)|) + \sup_{(\tau,x)\in[0,t]\times[0,L]}(|\partial_t\mathbf{d}_1(\tau,x)|)\Bigg),\ \forall t\in[0,T].\end{aligned} \tag{2.2.14}$$

The method consists in finding a Lyapunov function for this system. For the exponential stability, the usual basic quadratic Lyapunov function has the form

$$V(\mathbf{U},t) = \|\sum_{i=1}^n \sqrt{f_i}U_i(\cdot)\|_{C^0} + \|\sum_{i=1}^n \sqrt{f_i}(E(\mathbf{U},\cdot)\Theta(\mathbf{U}),t)_i\|_{C^0},\ \ \forall\, \mathbf{U} = (U_1,...,U_n)\in C^1([0,L];\mathbb{R}^n), \tag{2.2.15}$$

where f_i are positive and C^1 functions on $[0,L]$, $E(\mathbf{U},x)$ is a matrix diagonalizing $A(\mathbf{U},x)$, and $\Theta(\mathbf{U},t)$ is given by

$$\Theta(\mathbf{U},t) = A(\mathbf{U},x)\partial_x\mathbf{U} + B(\mathbf{U},x) + \mathbf{d}_1(t,x) \tag{2.2.16}$$

such that for a solution u to the system (2.1.14)–(2.1.15), $\Theta(\mathbf{u},t) = \partial_t\mathbf{u}$ and

$$V(\mathbf{u},t) = \|\sum_{i=1}^n \sqrt{f_i}u_i(t,\cdot)\|_{C^0} + \|\sum_{i=1}^n \sqrt{f_i}(E(\mathbf{u},\cdot)\partial_t\mathbf{u})_i\|_{C^0}. \tag{2.2.17}$$

Unfortunately, because of the perturbations, $V(\mathbf{u},t)$ is not equivalent anymore to the C^1 norm of the solution when $\mathbf{d}_1 \neq 0$ (recall that we are talking of the C^1 norm in x). To remedy this, we add a term to the basic Lyapunov function and define

$$V(\mathbf{u},t) = \|\sum_{i=1}^n \sqrt{f_i}U_i(\cdot)\|_{C^0} + \|\sum_{i=1}^n \sqrt{f_i}(E(\mathbf{U},\cdot)\Theta(\mathbf{U}),t)_i\|_{C^0} + \|\mathbf{d}_1(t,\cdot)\|_{C^0},\ \ \forall\, \mathbf{U} = (U_1,...,U_n)\in C^1([0,L],\mathbb{R}^n), \tag{2.2.18}$$

such that for any solution $\mathbf{u}$ to the system (2.1.14)–(2.1.15) and provided $\|\mathbf{u}\|_{C^1}$ and $\|\mathbf{d}_1\|_{C^1}$ are sufficiently small

$$c(\|\mathbf{u}(t,\cdot)\|_{C^1} + \|\mathbf{d}_1(t,\cdot)\|_{C^0}) \leq V(\mathbf{u}(t,\cdot),t) \leq C(\|\mathbf{u}(t,\cdot)\|_{C^1} + \|\mathbf{d}_1(t,\cdot)\|_{C^0}). \tag{2.2.19}$$

This is still not an equivalence with the norm of the perturbations only but with the sum of the norms of the perturbations and the internal disturbances. However, this is enough to get the result and to conclude. It suffices to show is that there exists an ISS estimate on $V(\mathbf{u},t)$. Indeed, assume that

$$\begin{aligned}V(\mathbf{u}(t,\cdot),t) \leq\ & e^{-\gamma(t-s)}V(s) + C_2\left(\sup_{\tau\in[s,t]}\left(e^{-\gamma(t-\tau)}|\mathbf{d}_2(\tau)|\right) + \sup_{\tau\in[s,t]}\left(e^{-\gamma(t-\tau)}|\mathbf{d}_2'(\tau)|\right)\right)\\ &+ C_3\left(\sup_{(\tau,x)\in[s,t]\times[0,L]}\left(e^{-\gamma(t-\tau)}|\partial_t\mathbf{d}_1(\tau,x)|\right) + \sup_{(\tau,x)\in[s,t]\times[0,L]}\left(e^{-\gamma(t-\tau)}|\mathbf{d}_1(\tau,x)|\right)\right),\ \ \forall\, s\in[0,T]\end{aligned} \tag{2.2.20}$$

7. see [19, (4.137)–(4.132)] or [134, (2.8)] for a definition of compatibility conditions for a hyperbolic system.

then

$$\|\mathbf{u}(t,\cdot)\|_{C^1} \leq \frac{C}{c}e^{-\gamma(t-s)}\|\mathbf{u}(s,\cdot)\|_{C^1} + \frac{C_2}{c}\left(\sup_{\tau\in[s,t]}\left(e^{-\gamma(t-\tau)}|\mathbf{d}_2(\tau)|\right) + \sup_{\tau\in[s,t]}\left(e^{-\gamma(t-\tau)}|\mathbf{d}_2'(\tau)|\right)\right)$$
$$+ \frac{C_3+C}{c}\left(\sup_{(\tau,x)\in[s,t]\times[0,L]}\left(e^{-\gamma(t-\tau)}|\partial_t\mathbf{d}_1(\tau,x)|\right) + \sup_{(\tau,x)\in[s,t]\times[0,L]}\left(e^{-\gamma(t-\tau)}|\mathbf{d}_1(\tau,x)|\right)\right), \quad \forall\, s\in[0,T]. \tag{2.2.21}$$

In fact, it is enough to show this for any C^2 solution of the system $\mathbf{u}$ such that $\sup_{t\in[0,T]}\|\mathbf{u}(t,\cdot)\|_{C^1}$ is small enough, and then use Theorem 2.2.7 and a density argument.

As a function $V(\mathbf{u},t)$ is not very convenient to differentiate, we start by approximating it as in [58, 132] by

$$\begin{aligned} W_p &= W_{1,p} + W_{2,p} \\ W_{1,p} &= \left(\int_0^L \sum_{i=0}^n f_i^p e^{-2p\mu s_i x} u_i^{2p}(t,x)dx\right)^{1/2p}, \\ W_{2,p} &= \left(\int_0^L \sum_{i=0}^n f_i^p e^{-2p\mu s_i x} (E(\mathbf{u},x)\partial_t\mathbf{u}(t,x))_i^{2p}dx\right)^{1/2p}, \end{aligned} \tag{2.2.22}$$

where $p\in\mathbb{N}\setminus\{0\}$. Obviously $W_p\to V(\mathbf{u})$ for any solution of the system. Differentiating $W_{1,p}$ with respect to C^2 solutions, integrating by parts and using a Taylor expansion of $\lambda(\mathbf{u})$ and $E(\mathbf{u},x)$ at the first order, one eventually gets for p large enough and $\|\mathbf{u}(t,\cdot)\|_{C^1}$ sufficiently small

$$\frac{dW_{1,p}}{dt} \leq -I_2 - I_3 - \frac{\mu\alpha_0}{2}W_{1,p} + CW_{1,p}\|\mathbf{u}\|_{C^1}, \tag{2.2.23}$$

where

$$I_2 = \frac{W_{1,p}^{1-2p}}{2p}\left[\sum_{i=1}^n \lambda_i(\mathbf{u},x)f_i^p u_i^{2p}e^{-2p\mu s_i x}\right]_0^L \tag{2.2.24}$$

$$\begin{aligned} I_3 =& W_{1,p}^{1-2p}\int_0^L \sum_{i=1}^n f_i^p(x)u_i^{2p-1}\left(\sum_{k=1}^n M_{ik}u_k\right)e^{-2\mu s_i x}dx \\ &- \frac{W_{1,p}^{1-2p}}{2}\int_0^L\sum_{i=1}^n \lambda_i(\mathbf{u},x)f_i^{p-1}(x)f_i'(x)u_i^{2p}e^{-2\mu s_i x}dx. \end{aligned} \tag{2.2.25}$$

I_3 is relatively easy to deal with and setting $D_p = \left(\int_0^L f_i^p(x)d_{1,i}^{2p}(t,x)dx\right)^{2p}$, we can show that, under the assumptions of Theorem 2.2.2,

$$\frac{dW_{1,p}}{dt} \leq -I_2 - \frac{3\mu\alpha_0}{8}W_{1,p} + CW_{1,p}\|\mathbf{u}\|_{C^1} + \left(\frac{8}{\mu\alpha_0}\right)^{2p-1}\frac{W_{1,p}^{1-2p}}{2p}D_p^{2p}. \tag{2.2.26}$$

We will not detail the proof of this estimate here but one can refer to [22, Section 5]. Let us now look at I_2. Let Δ a matrix of $\mathcal{D}_n^+$ such that (2.2.8) holds and set $\xi=(\xi_1,...,\xi_n)$ defined by

$$\xi_i = \begin{cases} \Delta_i u_i(t,L) \text{ for } i\in[1,m], \\ \Delta_i u_i(t,0) \text{ for } i\in[m+1,n], \end{cases} \tag{2.2.27}$$

and denote $F(\xi)=(F_i)_{i\in\{1,...,n\}} = G\begin{pmatrix}\mathbf{u}_+(L)\\ \mathbf{u}_-(0)\end{pmatrix}$. Using the boundary conditions (2.1.15) and the fact that

$\lambda(\mathbf{u}, x)$ is C^1 with $\mathbf{u}$, we have

$$\begin{aligned} I_2 \geq \frac{W_{1,p}^{1-2p}}{2p} \Bigg[& \sum_{i=1}^{m} (\Lambda_i(L) + O(\xi)) \frac{f_i^p(L)}{\Delta_i^{2p}} \xi_i^{2p} e^{-2p\mu L} \\ & + \sum_{i=m+1}^{n} (|\Lambda_i(0)| + O(\xi)) \frac{f_i^p(0)}{\Delta_i^{2p}} \xi_i^{2p} \\ & - \sum_{i=1}^{m} \left(\Lambda_i(0) + O\left(\sum_{i=1}^{n} (|F_i(\xi)| + |d_i|) \right) \right) f_i^p(0) \left(F_i(\xi) + d_i \right)^{2p} \\ & - \sum_{i=m+1}^{n} \left(|\Lambda_i(L)| + O(\sum_{i=1}^{n} (|F_i(\xi)| + |d_i|)) \right) f_i^p(L) e^{2p\mu L} \left(F_i(\xi) + d_i \right)^{2p}, \Bigg] \end{aligned} \tag{2.2.28}$$

where the O represents a continuous function independent of p such that $O(x)/|x|$ is bounded when $|x|$ tends to 0. As we have a bound on $\mathbf{u}$ (hence on $F_i(\xi)$) and a bound on d_i, a natural idea would be to develop $(F_i(\xi) + d_i)^{2p}$ and bound each of the terms that appear. However, this would pose a problem when making p tends to $+\infty$ as we would end up with an infinite number of small terms and their sum might not be small anymore. Note that this problem does not occur if one want to transpose the same type of result for the H^q norm, where p would be fixed to $p = 1$ and therefore the number of terms in the sum would remain finite. In order to avoid this problem, we use that for any $(a, d) \in \mathbb{R}^2$, and any $\alpha > 0$,

$$(a + d)^{2p} \leq (1 + \alpha)^{2p} a^{2p} + \left(1 + \frac{1}{\alpha} \right)^{2p} d^{2p}. \tag{2.2.29}$$

This can be shown by checking the cases $|a|\alpha > |d|$ and $|a|\alpha \leq |d|$. For simplicity we denote $d_{\max}(t) = \sup_i |d_i(t)|$ and we recall the notation l_i defined in Theorem 8.2.2 by $l_i := L$ if $1 \leq i \leq m$ and $l_i := 0$ if $m + 1 \leq i \leq n$. Using (2.2.29) in (2.2.28), we get

$$\begin{aligned} I_2 \geq \frac{W_{1,p}^{1-2p}}{2p} \Bigg[& \sum_{i=1}^{m} (\Lambda_i(l_i) + O(\xi)) \frac{f_i^p(l_i)}{\Delta_i^{2p}} \xi_i^{2p} e^{-2p\mu L} \\ & + \sum_{i=m+1}^{n} (|\Lambda_i(l_i)| + O(\xi)) \frac{f_i^p(l_i)}{\Delta_i^{2p}} \xi_i^{2p} \\ & - \sum_{i=1}^{n} \left(|\Lambda_i(L - l_i)| + O\left(|F(\xi)| + d_{\max} \right) \right) f_i^p(L - l_i) e^{2p\mu(L - l_i)} \left(1 + \alpha \right)^{2p} F_i^{2p}(\xi) \\ & - \sum_{i=1}^{n} \left(|\Lambda_i(L - l_i)| + O\left(|F(\xi)| + d_{\max} \right) \right) f_i^p(L - l_i) e^{2p\mu(L - l_i)} \left(1 + \frac{1}{\alpha} \right)^{2p} d_{\max}^{2p} \Bigg]. \end{aligned} \tag{2.2.30}$$

Then, proceeding similarly as in [132] we can show that the sum of the first four terms are positive under the assumption of Theorem 2.2.2. Using (2.2.26), showing a similar estimate on $W_{2,p}$, and choosing $\|\mathbf{u}(t, \cdot)\|_{C^1}$ and $d_{\max}$ smaller than $\mu\alpha_0/4$ we conclude that

$$\frac{dW_p}{dt} \leq -\frac{\mu\alpha_0}{8} W_p + \frac{W_p^{1-2p}}{2p} \left(\left(\frac{8}{\mu\alpha_0} \right)^{2p-1} D_p^{2p} + C D_{2,p}^{2p} \left(1 + \frac{1}{\alpha} \right)^{2p} \left(d_{\max}^{2p}(t) + (d'_{\max}(t))^{2p} \right) \right), \tag{2.2.31}$$

Where $D_{2,p} = \left(\sum_{i=1}^{n} |\Lambda_i(L - l_i)| f_i^{2p}(L - l_i) e^{2p\mu(L - l_i)} \right)^{1/2p}$. We see here that when there are no disturbances, i.e. $\mathbf{d}_1 \equiv \mathbf{d}_2 \equiv 0$, we can let $p \to +\infty$ and since $W_p \to V$ for any $t \in [0, T]$ we are able to get the differential inequality

$$\frac{dV(\mathbf{u}, t)}{dt} \leq -\frac{\mu\alpha_0}{8} V(\mathbf{u}, t). \tag{2.2.32}$$

However, in our case we are unable to obtain the corresponding ISS inequality

$$\frac{dV(\mathbf{u},t)}{dt} \leq -\frac{\mu\alpha_0}{8}V(\mathbf{u},t) + C\left(d_{\max}(t) + d'_{\max}(t)\right). \tag{2.2.33}$$

Nevertheless, multiplying (2.2.31) by $2pW_p^{2p-1}$ on both sides, we can use Gronwall Lemma and use to concavity of $x \to x^{1/2p}$ to get, for any $t, s \in [0,T]$,

$$\begin{aligned} W_p(t) \leq e^{-\frac{\mu\alpha_0}{8}(t-s)}W_p(s) + C^{1/2p}\left(\int_s^t e^{-2p\frac{\mu\alpha_0}{8}(t-v)}D_{2p}^{2p}\left(1+\frac{1}{\alpha}\right)^{2p}(d_{\max}^{2p}(v) + (d'_{\max}(v))^{2p})dv\right)^{1/2p} \\ + \left(\int_s^t e^{-2p\frac{\mu\alpha_0}{8}(t-v)}D_p^{2p}\left(\frac{8}{\mu\alpha_0}\right)^{2p-1}dv\right)^{1/2p}. \end{aligned} \tag{2.2.34}$$

Letting $p \to +\infty$ and using the fact that for a continuous function a, $(\int_s^t \sum_{i=1}^n |a_i|^{2p}(v)dv)^{1/2p} \xrightarrow[p\to+\infty]{} \max_{i,x\in[s,t]}|a_i|$, we obtain the desired estimate (2.2.20).

2.2.1.2 Comparison with existing results

We give here the main ideas of the proof of Proposition 2.2.5–2.2.6. The first thing to note is the following proposition, shown in [133, Theorem 3.2] [8]

Proposition 2.2.8. *Let a system be of the form (2.2.9), (2.2.10), with a and b two continuous functions on $[0,1]$ and denote $M := \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$ and $G(\mathbf{u}) = \begin{pmatrix} 0 & k_1 \\ k_2 & 0 \end{pmatrix}\mathbf{u}$. Then the two following are equivalent:*

- *Condition (2.2.7)–(2.2.8) are satisfied.*
- *There exists a solution η on $[0,1]$ to*

$$\begin{cases} \eta' = \left|\frac{a}{\Lambda_1}\right| + \left|\frac{b}{|\Lambda_2|}\right|\eta^2, \\ \eta(0) = |k_1| \end{cases} \tag{2.2.35}$$

such that

$$\eta(1) < |k_2|^{-1}. \tag{2.2.36}$$

When a and b are constant, η can be computed explicitly. Indeed, denoting $c_1 = |a|/\Lambda_1$ and $c_2 = |b|/|\Lambda_2|$, we have

$$\eta(x) = \sqrt{\frac{c_1}{c_2}}\tan(\operatorname{atan}(\sqrt{\frac{c_2}{c_1}}|k_1|) + \sqrt{c_1c_2}x), \text{ on } [0,x_1), \tag{2.2.37}$$

where x_1 is given by

$$x_1 = \frac{\left(\pi/2 - \operatorname{atan}(\sqrt{\frac{c_2}{c_1}}|k_1|)\right)}{\sqrt{c_1c_2}}, \tag{2.2.38}$$

and

$$\lim_{x\to x_1}\eta(x) = +\infty. \tag{2.2.39}$$

Hence, conditions (2.2.7), (2.2.8) of Theorem 2.2.2 are equivalent to

$$\begin{aligned} x_1 = \frac{\left(\pi/2 - \operatorname{atan}(\sqrt{\frac{c_2}{c_1}}|k_1|)\right)}{\sqrt{c_1c_2}} > 1. \\ |k_2| < \eta(1)^{-1} = \left(\sqrt{\frac{c_1}{c_2}}\tan(\operatorname{atan}(\sqrt{\frac{c_2}{c_1}}|k_1|) + \sqrt{c_1c_2})\right)^{-1}. \end{aligned} \tag{2.2.40}$$

8. The conditions stated in [133, Theorem 3.2] are in fact different than (2.2.7)–(2.2.8), but are shown to be equivalent in the same paper (see [133, Section 4])

Assume that there exists $K > 0$ and that (2.2.11) of Theorem 2.2.4 holds. Our goal is to show that (2.2.40) holds as well.

— **Showing that $x_1 > 1$.** From (2.2.11) we obtain

$$\sqrt{\frac{|k_1|}{c_1}} < \left(\frac{1}{\sqrt{c_1 c_2}} - 1\right), \tag{2.2.41}$$

and thus

$$x_1 > \frac{\pi/2 - \operatorname{atan}\left(\left(\frac{1}{(c_1c_2)^{1/4}} - (c_1c_2)^{1/4}\right)^2\right)}{\sqrt{c_1c_2}}. \tag{2.2.42}$$

Note that $\sqrt{c_1c_2} < 1$ from (2.2.41). We can divide the analysis in three cases: $\sqrt{c_1c_2} \in (1/2, 1)$, $\sqrt{c_1c_2} \in (1/3, 1/2)$, and $\sqrt{c_1c_2} < 1/3$. The two first do not bring any problem, and for the third one we can conclude by using that for every $x > 0$,

$$\begin{aligned} \pi/2 - \operatorname{atan}(x) &= \operatorname{atan}(1/x), \\ \operatorname{atan}(x) &\geq x - x^3/3. \end{aligned} \tag{2.2.43}$$

and introducing the function $y \to 1/(1-y)^2 - y/(3(1-y)^6)$ that is strictly increasing then decreasing on $[0, 1/3]$.

— **Showing that $|k_2| < \eta(1)^{-1}$.** This is the hardest part. To prove this, we introduce x_2 such that $\eta(x_2) = |k_2|^{-1}$. Such an x_2 exists given that $\eta(0) = |k_1| < |k_2|^{-1}$ from (2.2.11) and given that $\eta \to +\infty$ when $x \to x_1$. In fact

$$x_2 = \frac{\operatorname{atan}(\sqrt{\frac{c_1}{c_2}}|k_2|^{-1}) - \operatorname{atan}(\sqrt{\frac{c_2}{c_1}}|k_1|)}{\sqrt{c_1c_2}}. \tag{2.2.44}$$

As η is strictly increasing, it suffices to show that $x_2 > 1$ to prove the result. From (2.2.11) we show that

$$x_2 > \frac{\operatorname{atan}(\sqrt{\frac{c_2}{c_1}}|k_2|^{-1}) - \operatorname{atan}\left(\left(\frac{1}{(c_2c_1)^{1/4} + \left(\frac{c_1}{c_2}\right)^{1/4}\sqrt{|k_2|}} - (c_1c_2)^{1/4}\right)^2\right)}{\sqrt{c_1c_2}}. \tag{2.2.45}$$

which depends a priori of three parameters $|k_2|$, c_1 and c_2 but can be simplified by setting $X := \sqrt{c_1/c_2}|k_2|$ and $Y := \sqrt{c_1c_2}$, $Z = \sqrt{X} + \sqrt{Y}$, which gives

$$\begin{aligned} x_2 &> \frac{\operatorname{atan}\left(\frac{1}{(Z-\sqrt{Y})^2}\right) - \operatorname{atan}\left(\left(\frac{1}{Z} - \sqrt{Y}\right)^2\right)}{Y} \\ &= \frac{\frac{\pi}{2} - \left[\operatorname{atan}\left((Z-\sqrt{Y})^2\right) + \operatorname{atan}\left(\left(\frac{1}{Z} - \sqrt{Y}\right)^2\right)\right]}{Y}, \end{aligned} \tag{2.2.46}$$

and $Z \in \left(\sqrt{Y}, 1/\sqrt{Y}\right)$. Then we can use that $\operatorname{atan}(a - x) \leq \operatorname{atan}(a) - x/(1 + a^2)$ for any $x \in [0, a]$ and (2.2.43) to get

$$x_2 > \frac{2(Z + Z^3)}{1 + Z^4}\frac{1}{\sqrt{Y}} - 1. \tag{2.2.47}$$

By showing that $x \to x + x^3/(1 + x^4)$ is increasing on $[0, 1]$ and decreasing on $[1, +\infty)$, and recalling that $Z \in \left(\sqrt{Y}, 1/\sqrt{Y}\right)$ we obtain that

$$x_2 > \frac{2(1 + \sqrt{Y}^2)}{1 + \sqrt{Y}^4} - 1 \geq 1, \tag{2.2.48}$$

since $\sqrt{Y} \leq 1$.

This ends the first statement of Proposition 2.2.5. Showing that the converse does not hold when $a \neq 0$ or $b \neq 0$ is a consequence from taking $k_2 = \eta^{-1}(1) - \varepsilon$ with ε sufficiently small and showing by contradiction that (2.2.11) does not hold. This comes from the fact that x_2 can be made as close to 1 as desired by selecting ε sufficiently small, while, if (2.2.11) holds, there exists $c > 0$ independent of ε such that $x_2 - 1 > c$. We do not detail here the proof of Proposition 2.2.6, which relies on similar estimates as the proof of Proposition 2.2.5.

Chapter 3

Global L^2 exponential stability and ISS of a semilinar system

3.1 Introduction

In this chapter, we present the work of [135], where we study the exponential stability and ISS of a Lipschitz semilinear hyperbolic system in the L^2 norm.

Let us first have a look at a general hyperbolic system without disturbances, namely

$$\begin{aligned} &\partial_t \mathbf{u} + A(\mathbf{u},x)\partial_x \mathbf{u} + B(\mathbf{u},x) = 0, \\ &\begin{pmatrix} \mathbf{u}_+(t,0) \\ \mathbf{u}_-(t,L) \end{pmatrix} = G \begin{pmatrix} \mathbf{u}_+(t,L) \\ \mathbf{u}_-(t,0) \end{pmatrix}. \end{aligned} \tag{3.1.1}$$

Before [135], the only general result that existed for a general 1-D hyperbolic system for a H^p norm is the following, showed in [19, Chapter 6] (see [18] for the 2×2 case in a linear framework).

Theorem 3.1.1 ([19]). *Consider a system of the form (3.1.1) where A, B and G are of class C^p. If there exists $Q \in C^1([0,L], \mathcal{D}_n^+(\mathbb{R}))$, where $\mathcal{D}_n^+(\mathbb{R})$ is the space of definitive positive diagonal matrices, such that*

— *the matrix*

$$-(Q\Lambda)'(x) + Q(x)\partial_{\mathbf{u}} B(\mathbf{0},x) + \partial_{\mathbf{u}} B(\mathbf{0},x)^T Q(x)^T \tag{3.1.2}$$

is positive definite for any $x \in [0,L]$,

— *the matrix*

$$\begin{pmatrix} \Lambda_+(L)Q_+(L) & 0 \\ 0 & -\Lambda_-(0)Q_-(0) \end{pmatrix} - G'(0)^T \begin{pmatrix} \Lambda_+(0)Q_+(0) & 0 \\ 0 & -\Lambda_-(L)Q_-(L) \end{pmatrix} G'(0) \tag{3.1.3}$$

is semi-definite positive.

then the system (3.1.1) is exponentially stable for the H^p norm for any $p \in \mathbb{N} \setminus \{0,1\}$.

When the system is semilinear, i.e. $A(\mathbf{u},x) = A(\mathbf{0},x) = \Lambda(x)$, this result also holds for the H^1 norm, but cannot hold in general for the L^2 norm. When, in addition, the system is linear then this result also holds for the L^2 norm. In [132] was shown a (local) exponential stability result for the C^0 norm when the system is semilinear, while [99] showed a (semi-global) exponential stability result for the same norm provided that the transport term is constant and the source term is separable [1] and Lipschitz with some condition on its Lispchitz bound. Thus, usually, getting an exponential stability result in the L^2 norm is out of reach for semilinear systems [2].

1. meaning that for any $i \in \{1,...,N\}$, $B_i(\mathbf{u},x)$ only depends on u_i.
2. It is worth noting that [99] also showed a well-posedness result in the L^2 norm in their framework, primarily to study the effect of saturating boundary conditions. Besides, some result exists in particular cases, see for instance [264] for the semilinear wave equations in a multidimensional framework where the stabilization is obtained in H^1 norm for the solution y, hence for the L^2 norm for the total state $\mathbf{u} = (y, \partial_x y)$.

What we show in this chapter is that this stability can be recovered, when the source term is Lipschitz and with some condition on the size of the source. In addition, the stabilization is global and holds even if B is nonlocal, i.e. depends on $\mathbf{u}$ on the entire domain $[0, L]$ and not only on $\mathbf{u}(x)$. Nonlocal source terms are not only a mathematical curiosity but are also found in many important phenomena as population dynamics, material sciences, flocking, traffic flow [25, 28, 230]. Finally, we consider the system with disturbances given by (3.2.8)–(3.2.9) (which is similar to (2.1.14)–(2.1.15) in a semilinear nonlocal framework) and we show that these results can be extended to the exponential ISS in the L^2 norm.

In what follows, we look at the system (3.1.1) where $A(\mathbf{u}, x) = \Lambda(x)$ (i.e. the system is semilinear) and B is not anymore a function from $\mathbb{R}^n \times [0, L]$ but can be a non-local source term from $L^2((0, L); \mathbb{R}^n) \times [0, L]$ to $\mathbb{R}^n$, i.e. $B(\mathbf{u}, x)$ stands for $B(\mathbf{u}(t, \cdot), x)$. In the following we will assume that B is Lipschitz in the following sense[3]: for any $\mathbf{u}$, $\mathbf{v} \in L^2((0, L); \mathbb{R}^n)$,

$$\|B(\mathbf{u}, \cdot) - B(\mathbf{v}, \cdot)\|_{L^2} \leq C_B \|\mathbf{u} - \mathbf{v}\|_{L^2}, \tag{3.1.4}$$

where C_B is a positive constant independent of $\mathbf{u}$ and $\mathbf{v}$. Of course this assumption is satisfied if B is local, i.e. takes argument in $\mathbb{R}^n \times [0, L]$, and is Lipschitz with respect to its first argument, with a Lipschitz constant $C_B^{(1)}(x)$ that might depend on x but as a L^2 function. We also assume that the boundary operator G appearing in (3.1.1) is Lipschitz. The fact that G is Lipschitz implies that there exists a matrix K such that[4]

$$\left| G_i \begin{pmatrix} \mathbf{u}_+(t, L) \\ \mathbf{u}_-(t, 0) \end{pmatrix} \right| \leq \sum_{j=1}^{m} K_{ij} |u_j(t, L)| + \sum_{j=m+1}^{n} K_{ij} |u_j(t, 0)|. \tag{3.1.5}$$

The matrix $K = C_G I$, where I is the identity matrix and C_G the Lipschitz constant of G would work. However, there might be other matrices K satisfying (3.1.5) and some could lead to potentially better conditions in our results (see Theorem 3.2.1). The first thing we can note is that this system is globally well-posed:

Theorem 3.1.2. *For any $T > 0$ and any $\mathbf{u}_0 \in L^2(0, L)$ the Cauchy problem (3.1.1) with $A(\mathbf{u}, x) = \Lambda$, B satisfying (3.1.4), G satisfying (3.1.5), and initial condition $\mathbf{u}(0, \cdot) = \mathbf{u}_0$ has a unique solution $\mathbf{u} \in C^0([0, T], L^2(0, L))$. Moreover,*

$$\|\mathbf{u}(t, \cdot)\|_{L^2} \leq C(T) \|\mathbf{u}_0\|_{L^2}, \ \forall\ t \in [0, T], \tag{3.1.6}$$

where $C(T)$ is a constant depending only on T.

This theorem is showed by extending the result of [99, Theorem A.1] to the case where B is nonlocal, Λ depends on x and the eigenvalues of Λ might have different sign. It allows to show the existence of a ζ-dissipative nonlinear semigroup as defined in [191]. From the existence of this nonlinear semigroup we deduce the existence of a unique integral solution (see [191] for a definition) to the Cauchy problem for any initial condition $\mathbf{u}^0 \in L^2(0, L)$. Besides, when the initial condition belongs to $H^1(0, L)$ the integral solution satisfies the boundary conditions and the estimate (3.1.6). Then using a density argument we show that this unique integral solution is also the unique weak L^2 solution in the following sense

Definition 3.1.1. Let $\mathbf{u}_0 \in L^2(0, L)$. We say that $\mathbf{u} \in C^0([0, +\infty); L^2(0, L))$ is an L^2 solution of the Cauchy problem (2.1.4), (2.1.9), $\mathbf{u}(0, \cdot) = \mathbf{u}_0$, if for every $T > 0$ there exists a sequence of functions $\mathbf{u}_{0,n} \in H^1(0, L)$ satisfying (2.1.9) and such that

$$\begin{aligned} &\mathbf{u}_{0,n} \to \mathbf{u}_0 \text{ in } L^2(0, L), \\ &\mathbf{u}_n \to \mathbf{u} \text{ in } C^0([0, T], L^2(0, L)), \end{aligned} \tag{3.1.7}$$

3. Note that B can also be seen as a function from $L^2((0, L); \mathbb{R}^n)$ to $L^2((0, L); \mathbb{R}^n)$, hence (3.1.4) corresponds exactly to the definition of a Lipschitz function from $L^2((0, L); \mathbb{R}^n)$ to itself.

4. Recall that we denoted $\mathbf{u}^+ = (u_i)_{i \in \{1, \dots, m\}}$ and $\mathbf{u}^- = (u_i)_{i \in \{m+1, \dots, n\}}$ as in Section 2.1.

where $\mathbf{u}_n \in C^0([0,T], H^1(0,L))$ is a weak solution of (2.1.4), (2.1.9) with initial condition $\mathbf{u}_{0,n}$, i.e. $\mathbf{u}_n$ satisfies (2.1.9) and for any $\phi \in C^1([0,T]; C_c^1((0,L);\mathbb{R}^n))$ we have

$$\begin{aligned}&\int_0^L \int_0^T \partial_t \phi^\top \mathbf{u}_n + \partial_x \phi^\top \Lambda(x)\mathbf{u}_n + \phi^\top(\Lambda_x \mathbf{u}_n - B(\mathbf{u}_n, x))dt\, dx \\ &= \int_0^L \left[\phi(\cdot,x)^\top \mathbf{u}_n(\cdot,x)\right]_0^T dx.\end{aligned} \tag{3.1.8}$$

Remark 3.1.1. *As noted in [99], this definition is slightly different from the usual definition given in [19, Definition A.3] when looking at linear systems. The usual definition given in [19, Definition A.3] consists in finding a solution to the weak formulation with test functions satisfying particular boundary conditions which corresponds to the adjoint of the boundary conditions of the system The reason for this difference comes from the nonlinear boundary conditions which may prevent the existence of the adjoint boundary conditions. Of course, in the linear case, a solution in the sense of [19, Definition A.3] is also a solution in the sense of Definition 3.1.1.*

3.2 Main results

Our main result is the following

Theorem 3.2.1. *Let a semilinear system be of the form (3.1.1), where $A(\mathbf{u},\cdot) = \Lambda \in C^1([0,L])$ and B is Lipschitz with respect to $\mathbf{u}$. If there exist $K \in M_n(\mathbb{R})$ satisfying (3.1.5), $J \in C^1([0,L]; M_n(\mathbb{R}))$ where $J(x)$ is a diagonal matrix with positive coefficients, and $M \in C^0([0,L]; M_n(\mathbb{R}))$, such that the following conditions are satisfied*

1. *(Interior condition)*

$$-(J^2\Lambda)' + J^2 M + M^\top J^2 \tag{3.2.1}$$

is positive definite and there exists $D \in C^1([0,L]; M_n(\mathbb{R}))$ where $D(x)$ is a diagonal matrix with positive coefficients, such that

$$C_g < \frac{\lambda_m}{2\max_{i,x}(D_i)\max_{i,x}(D_i J_i^2)}, \tag{3.2.2}$$

where C_g is the Lipschitz constant of $g := B - M$ and λ_m denotes the smallest eigenvalue of

$$-D(J^2\Lambda)'D + DJ^2 M D + DM^\top J^2 D, \tag{3.2.3}$$

2. *(Boundary condition) the matrix*

$$\begin{aligned}&\begin{pmatrix} J_+^2(L)\Lambda_+(L) & 0 \\ 0 & J_-^2(0)|\Lambda_-(0)| \end{pmatrix} \\ &-K^\top \begin{pmatrix} J_+^2(0)\Lambda_+(0) & 0 \\ 0 & J_-^2(L)|\Lambda_-(L)| \end{pmatrix} K\end{aligned} \tag{3.2.4}$$

is positive semidefinite,

then the system is globally exponentially stable for the L^2 norm.

Moreover the gain (or cost) of the estimate is $\|J^{-1}\|_{L^\infty}\|J\|_{L^\infty}$ and an admissible decay rate is $\mu := \lambda_m(2\max_{i,x}(D_i J_i^2))^{-1} - C_g \max_{i,x}(D_i)$, namely the following estimate holds

$$\|\mathbf{u}(t,\cdot)\|_{L^2} \leq \|J^{-1}\|_{L^\infty}\|J\|_{L^\infty} e^{-\mu t}\|\mathbf{u}^0\|_{L^2}. \tag{3.2.5}$$

We can observe that the conditions do not directly depend on the Lipschitz constant of B but rather on the Lipschitz constant of $g := B - M$ where M is a linear matrix that can be chosen, provided is satisfies the conditions (3.2.2)–(3.2.4). This makes the formulation of Theorem 3.2.1 a little more complicated than the same statement with $M = 0$ but it has a two advantages:

— Finding $M \neq 0$ such that (3.2.1)–(3.2.3) hold allows to find a potentially less restrictive Lipschitz bound on the size of the source term B.

— If B is a local linear operator we recover the existing result of Bastin and Coron [19, Chapter 6] by taking $M = B$ and $g := 0$.

Besides, note that λ_m can be easily numerically solved for practical applications. Finally, when B is local we can define a space dependent Lipschitz constant $C_B(x)$ belonging to $L^2(0, L)$, and thus we can also define a space dependent Lipschitz constant $C_g(x)$ of $g := B - M$ and the condition (3.2.2) can be slightly improved as follows:

$$C_g(x) < \frac{\lambda_m(x)}{\max_i(J_i^2)(x)} \quad \text{or} \quad C_g(x) < \mu_m(x)\frac{\max_i(J_i)(x)}{\inf_i(J_i)(x)}, \tag{3.2.6}$$

where $\lambda_m(x)$ and $\mu_m(x)$ are the smallest eigenvalues at a given x of the matrix given by (3.2.1) and (3.2.3) respectively. The proof is based on the existence a basic quadratic Lyapunov function of the form

$$\int_0^L (J(x)\mathbf{u}(t,x))^T (J(x)\mathbf{u}(t,x))dx, \tag{3.2.7}$$

where $J \in C^1([0, L]; \mathcal{D}_n^+)$.

Theorem 3.2.1 can be extended to ISS. Consider now the system with internal and boundary disturbances, as (2.1.14)–(2.1.15) in Chapter 2, but semilinear with B potentially nonlocal

$$\partial_t \mathbf{u} + \Lambda(x)\partial_x \mathbf{u} + B(\mathbf{u}, x) + \mathbf{d}_1(t, x) = 0, \tag{3.2.8}$$

$$\begin{pmatrix} \mathbf{u}_+(t,0) \\ \mathbf{u}_-(t,L) \end{pmatrix} = G \begin{pmatrix} \mathbf{u}_+(t,L) \\ \mathbf{u}_-(t,0) \end{pmatrix} + \mathbf{d}_2(t). \tag{3.2.9}$$

We have the following result.

Theorem 3.2.2. *Let a system be of the form (3.2.8)–(3.2.9) where $\Lambda \in C^1([0, L])$, $\mathbf{d}_1 \in L^2((0, T) \times (0, L))$, $\mathbf{d}_2 \in H^1([0, T])$ and B is Lipschitz with respect to $\mathbf{u}$. If the condition (3.2.2) is satisfied and the matrix defined by (3.2.4) is positive definite, then the system is globally strongly ISS with fading memory for the L^2 norm (see Definition[5] 2.1.3).*

Remark 3.2.1. *As expected the conditions of this theorem are very similar to the conditions of Theorem 3.2.1. The only difference is that the matrix given in (3.2.4) has to be definite positive to handle the boundary disturbances and not only semi-definite positive. Finally, in both Theorems 3.2.1 and 3.2.2 the gains (also sometimes called costs) C (resp. C_1, C_2) defined in (2.1.10) (resp. (2.1.19)) can be computed explicitely as a function of K, B and Λ.*

3.3 Illustrations

We present here some numerical simulations to illustrate Theorem 3.2.1. We consider a system inspired from [19, Section 5.6],

$$\begin{aligned} \partial_t u_1 + \partial_x u_1 &= cL^{-1} \sin\left(\int_0^L u_2(t,x)dx\right) \\ \partial_t u_2 - \partial_x u_2 &= cL^{-1} \sin\left(\int_0^L u_1(t,x)dx\right) \end{aligned} \tag{3.3.1}$$

5. which can be directly extended to this framework

with the boundary conditions

$$\begin{aligned} u_1(t,0) - u_2(t,0) &= 0 \\ u_1(t,L) - u_2(t,L) &= k u_1(t,L) \end{aligned} \tag{3.3.2}$$

Here, one boundary condition can be controlled through a parameter k to be chosen, while the other one is imposed. This system is genuinely non-local. One can check that $\mathbf{u}^* = 0$ is a steady-state and that, for any $c \in \mathbb{R}$, the open-loop system (i.e. $k = 0$) is unstable. Indeed there is a continuum of non-zero travelling wave solutions to (3.3.1)–(3.3.2) of the form

$$\begin{cases} u_1(t,x) = a \sin\left(\dfrac{2\pi}{L}(t-x)\right) \\ u_2(t,x) = a \sin\left(\dfrac{2\pi}{L}(t+x)\right) \end{cases} \tag{3.3.3}$$

with $a > 0$. We can apply Theorem 3.2.1 and deduce that, as long as $|c|L < 1/2$, there exists a globally stabilizing feedback given as follows:

$$k \in \left[1 - \sqrt{\frac{\varepsilon}{\varepsilon + 2L}}, 1 + \sqrt{\frac{\varepsilon}{\varepsilon + 2L}}\right] \quad \text{with} \quad \varepsilon = \frac{3}{4}(|c|^{-1} - 2L). \tag{3.3.4}$$

This is obtained by setting $M = 0$, $D = Id$, $J = (\sqrt{L + \varepsilon - x}, \sqrt{L + \varepsilon + x})$. Then $-(J^2\Lambda)' = Id$ and therefore is positive definite with smallest eigenvalue 1. We have $\max_{i,x}(J_i^2) = \varepsilon + 2L$ and observing that the condition (3.2.2) becomes

$$|c| < \frac{1}{\varepsilon + 2L}, \tag{3.3.5}$$

which holds thanks to the choice of ε. Finally conditions (3.2.4) becomes

$$(1-k)^2 \leq \frac{\varepsilon}{\varepsilon + 2L}, \tag{3.3.6}$$

which holds from the definition of k given by (3.3.4). On Figure 3.1 we represent the L^2 norm with respect to time of a solution to system (3.3.1)–(3.3.2) for different values of k when $c = 1/4$ and $L = 1$. In blue we represent the open loop situation $k = 0$, in green the closed-loop situation with $k = 3/4$, and in red with $k = 1/2$, both satisfying (3.3.4).

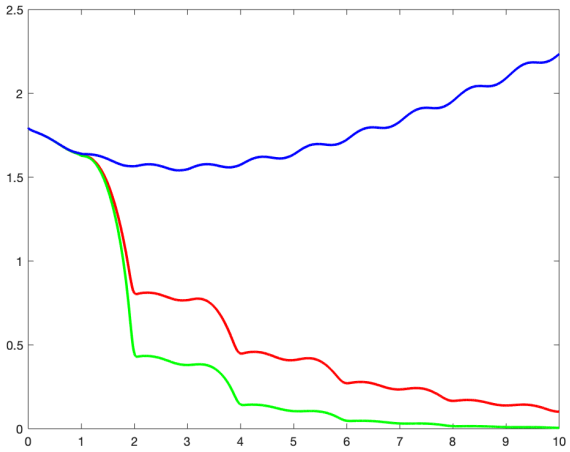


Figure 3.1 – Stability of the system (3.3.1)–(3.3.2) in open-loop (blue) and closed-loop with $k = 3/4$ (green) and $k = 1/2$ (red). horizontal axis represents time, and vertical axis represents the L^2 norm of the solution with initial condition $u_{1,0}(x) = \sqrt{2\pi x}$ and $u_{2,0}(x) = e^{-2\pi x}$.

Part II

Backstepping problems

Chapter 4

Stabilization of a cross-diffusion system by a backstepping method

In some cases, a direct use of basic quadratic Lyapunov functions is bound to fail. For example, if we consider the following system:

$$\begin{aligned} &\partial_t u_1 + \partial_x u_1 + u_2 = 0, \\ &\partial_t u_1 - \partial_x u_1 + u_1 = 0. \end{aligned} \quad \text{on } [0, L], \tag{4.0.1}$$

with boundary conditions

$$\begin{aligned} u_1(t, 0) &= u_2(t, 0), \\ u_2(t, L) &= k u_1(t, L), \end{aligned} \tag{4.0.2}$$

where $L = 2\pi$. In this case we can apply Theorem 1 of [18] which tells us that there exists a basic quadratic Lyapunov function for the norm L^2 if and only if the equation

$$\begin{aligned} \eta' &= 1 + \eta^2, \\ \eta(0) &= 0 \end{aligned} \tag{4.0.3}$$

has a solution on $[0, L]$. Clearly this equation has a simple maximal solution which is the tangent function and which exists only on $[0, \pi/2)$. With $L = 2\pi$ this eliminates the possibility of having a basic quadratic Lyapunov function, whatever the control at the boundaries. In fact for this example Bastin and Coron have shown with a spectral approach that it is even useless to look for a local static control of the form $u_2(t, L) = k u_1(t, L)$ for this linear system when $L > \pi$ [19, Section 5.6]. This motivates the search for more efficient controls, even if they are potentially more complicated.

To do this, backstepping is a very efficient method. Its principle is simple: transform the original system potentially hard to stabilize into a target system whose stabilization or stability is easy to show. The problem is to prove the existence of an invertible transformation between the original system and this simpler target system. Once this transformation is found, we just have to find a stabilizing control for the target system, and then apply the inverse transformation to have a control for the original system. Obviously, since we have applied a potentially complicated inverse transformation, the control of the original system is also potentially complicated.

4.1 Rapid and finite-time stabilization of a cross-diffusion problem

In this section, we present the results of [43], written with Jean Cauvin-Vila and Virginie Ehrlacher.

4.1.1 Backstepping: the Volterra approach

As we have seen, backstepping consists in finding two things:
— A target system, easy to stabilize
— An isomorphism mapping the original system into this target system.
The search for an isomorphism without any other condition is something complicated, because the space of isomorphisms on the working space is very large. This is why the first approaches to infinite dimensional backstepping were restricted to searching for the isomorphism in a very particular form: a Volterra transformation of the second kind, i.e. a T transformation of the form

$$\begin{aligned} T &: L^2 \to L^2, \\ (TU)(x) &= U(x) - \int_0^x K(x,y)U(y)dy, \quad \text{a.e. } [0,L], \end{aligned} \tag{4.1.1}$$

where $K \in L^2((0,L)^2)$ is a kernel to be determined. The choice of this type of transformation is not a coincidence: we find the same triangular structure as in the finite dimensional backstepping. Indeed, in view of the form (4.1.1), at a given point $x \in [0,L]$ the value $TU(x)$ depends only on the values of $u(t,\cdot)$ before the point x. Moreover, in this form, the K-dependent term is a Hilbert-Schmidt operator and the transformation T is therefore a compact perturbation of the identity. The transformation T is in fact naturally an isomorphism, which is very convenient. The infinite dimensional backstepping method has thus extensively (and almost exclusively) used this type of transformation and has known a great expansion in the last fifteen years (see Section 1.3). The main difficulty is to identify a good target system [1], and to show the existence of a solution K to a PDE system on a triangular domain [2].

4.1.2 A 1D cross-diffusion system

Cross-diffusion systems are widespread in real-life and appear in many fields: population dynamics, material physics, evolution of biological tissues, chemistry etc. They represent for example the coupled diffusion between several chemical species. A classical cross-diffusion system on a domain Ω and a time interval $[0,T]$ can be written as

$$\partial_t u - \operatorname{div}(A(u)\nabla u) = 0 \quad \text{on } [0,T]\times\Omega, \tag{4.1.2}$$

where $u \in \mathbb{R}^{n+1}$ is the vector representing the concentrations of each species which must therefore be positive and verify the condition $\sum_{i=0}^n u_i = 1$. $A(u)$ is a matrix $(n+1)\times(n+1)$ called the cross-diffusion matrix (or sometimes diffusion matrix). Since the concentration of the species u_0 is completely imposed by the condition $u_0 = 1 - \sum_{i=1}^n u_i$ we can re-express the system as a system of the form (4.1.2) where $u = (u_i)_{i\in\{1,\dots,n\}} \in \mathbb{R}^n$ (so without the 0 component) and $A(u)$ is now an $n\times n$ matrix. In this framework, the solutions verify the positivity condition $u_i > 0$, as well as $\sum_{i=0}^n u_i < 1$. These systems have been studied a lot, because of their interest in modeling, and it appears that the physical examples often have an entropic structure [38, 147–149]. In this section we are interested in a system of the form

$$\begin{cases} \partial_t u - \partial_x(A(u)\partial_x u) = 0, \quad \text{for } x \in (0,e(t)),\ t\in[0,T] \\ (A(u)\partial_x u)(t,0) = 0, \\ (A(u)\partial_x u)(t,e(t)) + e'(t)u(t,e(t)) = (\varphi_i(t))_{i\in\{1,..,n\}}, \\ u(0,x) = u^0(x), \quad x\in(0,e_0), \\ e'(t) = \sum_{i=0}^n \varphi_i(t), \end{cases} \tag{4.1.3}$$

where A admits an entropic structure, that is to say if we define

$$\mathcal{D} = \left\{ (u_1,\cdots,u_n) \in (\mathbb{R}_+^*)^n, \quad \sum_{i=1}^n u_i < 1 \right\} \subset (0,1)^n,$$

1. note that it is not always possible to transform the system into a homogeneous system [19]
2. precisely due to the triangular structure of the transformation

then

— there exists $h \in C^2(\bar{\mathcal{D}})$, a bounded from below strictly convex function such that its derivative Dh is invertible from $\bar{\mathcal{D}}$ to $\mathbb{R}^n$;

— there exists $\alpha > 0$, and $1 \geq m_i > 0$ such that for all $z = (z_1, \cdots, z_n)^T \in \mathbb{R}^n$ and $u = (u_1, \cdots, u_n)^T \in \mathcal{D}$,

$$z^T D^2 h(u) A(u) z \geq \alpha \sum_{i=1}^{n} u_i^{2m_i - 2} z_i^2.$$

and we denote $M(u) := A(u)(D^2 h(u))^{-1}$.

In the following we will assume additionally that $M(u)$ is symmetrical. We notice that this system has an additional complexity compared to (4.1.2): the domain depends on time and the dynamics of $e(t)$ is coupled to the state of the system u. This system models for example the Physical Vapor Deposition (PVD) process, used in material physics and in industry. This process allows to create thin layers of materials on a conductive substrate and has a large number of applications, from photovoltaic cells to the finishing of car shells or locks. The chemical elements to be deposited are introduced in gaseous form and, as they are deposited, the size of the layer increases. At the same time the temperature causes a diffusion in the layer between the different species. The vector u represents the concentrations of the different components, $A(u)$ their interactions, while $varphi_i$ represent the flows injected during the deposition process. This system was introduced in [15] where the authors showed the global existence of weak solutions to the nonlinear system by exploiting the entropic structure. They also showed that, when the fluxes are constant, the concentrations (normalized by the size of the domain) converge in long time to a uniform stationary state. Nevertheless this convergence is slow and in $1/\sqrt{t}$. This can be understood intuitively: when the fluxes are constant the domain increases with time and, the more time passes, the less the initial state influences the concentrations since one keeps adding material which has the "good" concentrations. Thus, at long time the initial state, whatever it is, is diluted in the new material added to the system. And at infinity the composition of the system is the same as that of the flows. This convergence is therefore mainly a consequence of the growth of the domain, and if one select a given piece of finite size in the layer, there is no guarantee that the concentrations on this piece converge well. The goal of [43] is to find a feedback control for $(\varphi_i)_{i\in[1,n]}$ which allows to overcome this problem by obtaining an exponential stabilization. Moreover, we would also like to stabilize the size of the domain $e(t)$ at the same time. We are therefore interested in the local stabilization of trajectories corresponding to a uniform stationary state $\bar{u}$ and a domain size $\bar{e}(t)$. Note that if $\bar{u}$ is a constant then consequently $\bar{e}(t) = a + bt$, where a and b are constants. In [43] we treat the case of the linear system, which already presents several difficulties.

4.1.3 Main results

Let us start by introducing the linearized system corresponding to (4.1.3) around a target trajectory $(\bar{u}, \bar{e}(t))$. By abuse of language we will call this trajectory $(\bar{u}, \bar{e}(t))$ a steady state even if e depends on time[3]. We denote $V = \bar{e}'(t) = \sum_{i=0}^{n} \bar{\varphi}_i$. The linearized system is then written

$$\begin{cases} \partial_t v - A(\bar{u})\partial_{xx}^2 v = 0, & \text{on } (0, \bar{e}(t)), \\ A(\bar{u})\partial_x v(t, \bar{e}(t)) + V v(t, \bar{e}(t)) = \psi(t), & \\ A(\bar{u})\partial_x v(t, 0) = 0, & \\ v(0, x) = v^0(x), & \text{for } x \in (0, \bar{e}_0). \end{cases} \tag{4.1.4}$$

3. We use this name because it is the trajectory corresponding to a steady state for u. There is no steady state for (u, e) because of the physical condition $u > 0$, $\varphi_i \geq 0$.

where $v = u - \bar{u}$ et $\psi_i = \varphi_i - \bar{\varphi}_i - \sum_{k=0}^{n} \bar{\varphi}_k$ for $i \in \{1, .., n\}$. Besides, denoting $\hat{e}(t) = e(t) - \bar{e}(t)$, the equation on $e(t)$ becomes

$$\hat{e}'(t) = \sum_{i=0}^{n} \varphi_i - \bar{\varphi}_i =: \theta(t). \tag{4.1.5}$$

We will denote in the following

$$\|v\|_{L^2(0,e(t))} = \left(\int_0^{e(t)} |v|^2 dx \right)^{1/2} \tag{4.1.6}$$

We will use several times the slight abuse of notation $L^2(0, \bar{e}(t))$ or $C^0([0,T]; L^2((0, \bar{e}(t))))$ to refer to a solution defined on $\cup_{t\geq 0}\{t\} \times (0, \bar{e}(t))$ (see [43] for more rigorous definitions). Since the size of the domain depends on time and diverges at infinity, a question arises *a priori* concerning the stability: do we seek to obtain

$$\frac{1}{e(t)} \int_0^{e(t)} |v(t,x)|^2 dx \to 0 \tag{4.1.7}$$

or

$$\int_0^{e(t)} |v(t,x)|^2 dx \to 0. \tag{4.1.8}$$

Clearly, it is easier to obtain (4.1.7) than (4.1.8) and the results of [15] illustrate this difference: in [15] the authors manage to obtain (4.1.7) but do not manage to obtain (4.1.8). In reality this question will not arise in our case because we will be looking for exponential stability, or even finite time stability, for which the stabilities of $e(t)^{-1}\|v\|_{L^2(0,e(t))}$ and $\|v\|_{L^2(0,e(t))}$ are equivalent as the growth of $e(t)$ is slower than any exponential growth.
The available controls for these systems are the $n+1$ flows ϕ_i, $i \in \{0, ..., n\}$ or equivalently ψ and θ. Concerning θ we will look for it in the form of a feedback control depending on time t and $\hat{e}(t)$. Concerning ψ, we will look for the feedback control in the following form:

$$\psi(t) = F_l(t)v(t, \bar{e}(t)) + F_{nl}(t)v(t, \cdot), \tag{4.1.9}$$

where $F_l(t) \in \mathbb{R}^{n\times n}$ represents the local part of the feedback, similar to that studied in Chapters 2–3, and $F_{nl}(t) : L^2((0, \bar{e}(t)) \to \mathbb{R}^n$ is a potentially non-local continuous operator (but more regular, since it is applicable on functions that are only L^2).
The exponential stabilization that we want to obtain is defined as follows:

Definition 4.1.1. Let $\lambda > 0$. A target state $(\bar{u}, \bar{e})$ of (4.1.3) is said to be exponentially stabilizable in L^2 with decay rate λ if there exist constants $C_{\bar{u},\lambda}, C_{\bar{e},\lambda} > 0$ independent of time and operators F_l and F_{nl} such that for any τ_1, $T > 0$,

a) $F_l \in L^\infty_{loc}(\mathbb{R}^*_+; \mathbb{R}^{n\times n})$ and $F_{nl} \in L^2((\tau_1, T); \mathcal{L}(L^2((0, \bar{e}(t))), \mathbb{R}^n))$, and the continuity constant of $F_{nl}(t)$ is uniformly bounded on $t \in [\tau_1, T]$, and for any $v^{\tau_1} \in L^2(0, \bar{e}(\tau_1))$, the linearized system (4.1.4) with initial condition v^{τ_1} and feedback control (4.1.9) has a unique weak solution $v \in C^0((\tau_1, T), L^2(0, \bar{e}(t)))$ and

$$\|v(t)\|_{L^2(0,\bar{e}(t))} \leq C_{\bar{u},\lambda} e^{-\lambda(t-\tau_1)/2} \|v^{\tau_1}\|_{L^2(0,\bar{e}(\tau_1))}, \text{ for all } t \in [\tau_1, T]. \tag{4.1.10}$$

b) There exists a function $\Theta \in L^1_{\text{loc}}\left(\mathbb{R}^*_+; \mathcal{C}^0(\mathbb{R})\right)$ such that, for any $\hat{e}_{\tau_1} \in \mathbb{R}$, $\hat{e}$ is well-defined by (4.1.5) with $\theta(t) = \Theta(t, \hat{e}(t))$ and satisfies:

$$|\hat{e}(t)| \leq C_{\bar{e},\lambda} e^{-\lambda(t-\tau_1)/2} |\hat{e}_{\tau_1}|, \text{ for all } t \in [\tau_1, T]. \tag{4.1.11}$$

The main results we obtain are the following.

Theorem 4.1.1 (Rapid stabilization)**.** *Let $\lambda > 0$ and $(\bar{u}, \bar{e})$ a target state. If λ is large enough, then, $(\bar{u}, \bar{e})$ is exponentially stabilizable in L^2 with decay rate $\lambda/2$. Moreover, A is diagonalizable and the feedback controls F_l, F_{nl} and Θ can be constructed as follows, for any $\tau_1 > 0$, $T > 0$:*

— *for any* $t \geq \tau_1$ *and* $\hat{e} \in \mathbb{R}$, $\Theta(t, \hat{e}) = -\lambda \hat{e}$;

— *for any* $t \geq \tau_1$, $1 \leq i \leq n$,

$$F_{l,\lambda}(t) = Q(\bar{u})^{-1} diag\left(\sigma_i k_\lambda^{\sigma_i}(\bar{e}(t), \bar{e}(t))\right)_{i\in\{1,\dots,n\}} Q(\bar{u}),$$

$$F^i_{nl,\lambda}(t) : v \mapsto Q(\bar{u})^{-1} \left(\int_0^{\bar{e}(t)} \left[\sigma_i \partial_x k_\lambda^{\sigma_i}(\bar{e}(t), y) + V k_\lambda^{\sigma_i}(\bar{e}(t), y)\right] Q(\bar{u}) v(y) dy \right)_{i\in\{1,\dots,n\}},$$

where $(\sigma_i)_{i\in\{1,\dots,n\}}$ *are the eigenvalues of* $A(\bar{u})$, $Q(\bar{u})^{-1}$ *is a matrix diagonalising* $A(\bar{u})$ *such that* $Q(\bar{u})A(\bar{u})Q(\bar{u})^{-1}$ *is diagonal and* $k_\lambda^{\sigma_i}$ *is the (unique) solution to the kernel equations*

$$\begin{cases} \partial^2_{xx} k_\lambda^{\sigma_i}(x,y) - \partial^2_{yy} k_\lambda^{\sigma_i}(x,y) = \dfrac{\lambda}{\sigma_i} k_\lambda^{\sigma_i}(x,y), \quad (x,y) \in \left\{(x,y) \in (\mathbb{R}_+)^2, \quad 0 < y \leq x\right\}, \\ \partial_y k_\lambda^{\sigma_i}(x,0) = 0 \quad x \in (0, +\infty), \\ k_\lambda^{\sigma_i}(x,x) = -\dfrac{\lambda}{2\sigma_i} x \quad x \in (0, +\infty), \end{cases} \tag{4.1.12}$$

This result shows that with a backstepping approach, it is possible to obtain a rapid stabilization, i.e. exponential with a decay rate as large as desired. We can in fact go further and show that it is possible to perfectly stabilize the system in finite time. The definition of stabilization in finite time is similar to the Definition 4.1.1 by replacing the exponential stability estimate by a stability condition in the broad sense and a convergence of v and $\hat{e}$ to 0 in a finite time T (see [43]).

Theorem 4.1.2 (Finite-time stabilization)**.** *Let* $(\bar{u}, \bar{e})$ *be a target state. Then, the system is stabilizable in* L^2 *for any finite time* $T > 0$.

This second result follows from the Theorem 4.1.1 and uses a technique already used for example in [71] and [75, 251], which consists in increasing the decay rate on shorter and shorter time steps and whose sum converges to a finite value, while the value of the decay rate converges to $+\infty$. For this, we must have a good estimate of the dependence of the pre-factor $C_{\bar{u},\lambda}$ of (4.1.10) on λ and this dependence should be sub-exponential [4]. In our case we can show that $C_{\bar{u},\lambda} = C' e^{\bar{e}_{\tau_1}\sqrt{\lambda}}$, where C' is a constant independent of λ.

4.1.4 Ideas of proof

We would like to find a Volterra transformation $\mathcal{T}$ of the form

$$\mathcal{T}(t)Z = Z(x) - \int_0^x K_\lambda(t,x,y) Z(x) dx, \quad x \in (0, \bar{e}(t)) \tag{4.1.13}$$

which goes from L^2 to itself and which transforms the system (4.1.4) in

$$\begin{cases} \partial_t v - A(\bar{u})\partial^2_{xx} v + \lambda v = 0, & \text{on } (0, \bar{e}(t)), \\ A(\bar{u})\partial_x v(t, \bar{e}(t)) + V v(t, \bar{e}(t)) = 0, \\ A(\bar{u})\partial_x v(t,0) = 0, \\ v(0,x) = v^0(x), \quad \text{for } x \in (0, \bar{e}_0), \end{cases} \tag{4.1.14}$$

i.e. essentially the same system with homogeneous boundary conditions and an additional damping in λ which will give the exponential decay. The main feature of our system is that it is non-autonomous and the domain is time dependent. This means that the backstepping transformation $\mathcal{T}$ depends *a priori* on time, and that its norm, as well as the norm of its inverse transformation too. We can classically show that this inverse transformation has the form

$$(\mathcal{T}^{-1}(t)W)(x) = W(x) + \int_0^x L_\lambda(t,x,y) W(y) dy, \tag{4.1.15}$$

4. or *a minima* less strong than the exponential decay of the estimate

for $x \in (0, \bar{e}(t))$. Thus since the exponential stabilization estimate that we obtain with the backstepping method depends on the norm of $\mathcal{T}$ and its inverse, it is necessary to check that they do not grow too fast and that they do not compensate the exponential decrease.
The second difficulty comes from the determination of $\mathcal{T}$. If we try to find a solution for $\mathcal{T}$ we will end up with equations on K, called kernel equations. But here, the kernel equations are also time dependent and involve derivatives in t, x, and y which makes them quite complicated. However, one can find an explicit time dependence and show that the solutions of these equations can be written as the restriction to a moving domain of a function independent of t. We then fall back on the kernel equations (4.1.12) which are more classical (see [71, 162] for instance).
We can then give an estimate of the norm of the transformation and the norm of the inverse transformation, by showing the following bounds on K_λ and L_λ

$$\|K_\lambda(t,\cdot,\cdot)\|^2_{H^1(D_t)} \leq C \exp\left(\tilde{c}\bar{e}(t)\sqrt{\frac{\lambda}{\inf\limits_{i\in\{1,\dots,n\}}|\sigma_i|}}\right), \tag{4.1.16}$$

$$\|L_\lambda(t,\cdot,\cdot)\|^2_{H^1(D_t)} \leq C\left(\frac{\lambda}{\inf\limits_{i\in\{1,\dots,n\}}|\sigma_i|}\right)^2 e^{\tilde{c}\bar{e}(t)}, \tag{4.1.17}$$

where $\tilde{c}$ and C are constants independent of λ and t, and D_t is the moving triangular domain, i.e., $D_t = \{(x,y) \in \mathbb{R}^2 : 0 < y \leq x < \bar{e}(t)\}$. These bounds then allow us to conclude both on the exponential stability, and finite time stability with an iteration procedure similar to [71, 205, 251].

Chapter 5

A more general backstepping: application to the heat equation

Backstepping using a Volterra transform is a very powerful tool. In several cases, however, it is not enough to stabilize the system. Indeed, consider for instance the system

$$\begin{cases} \partial_t u - \Delta u = v_1(t)\phi_1 + v_2(t)\phi_2, & (t,x) \in (0,+\infty) \times \mathbb{T}, \\ u|_{t=0} = u_0(x), & x \in \mathbb{T}, \end{cases} \tag{5.0.1}$$

where $\mathbb{T}$ is the one-dimensional torus, and v_1, v_2 are the (scalar) controls. In this case the backstepping approach with a Volterra transform will not be truly helpful. This intuitively comes from the fact that the Volterra transform is moving the difficulty to the boundaries and then the difficulty is tackled using the boundary controls. When the control is internal, however, the interest of the Volterra transform is sometimes limited. For this reason, we would like to investigate a more general backstepping that does not limit itself to Volterra transforms. This has inherent difficulties: we cannot count anymore on the invertibility and the cascade structure of the transformation. But, on the positive side, we may get stronger results.

In Section 5.2.1 we give a general overview of the method that we detail in Section 5.3 on the particular example of the heat equation with internal scalar controls given in (5.0.1). Then in Chapter 6 we show the limitations of this method and we present a *compactness-duality* method to overcome them. Finally, in Chapter 7 we detail an example of hyperbolic systems on which we manage to apply the same approach despite the fact that none of the methods presented in this chapter and Chapter 6 can be applied. This example is the water-tank system, modelled by the Saint–Venant equations.

5.1 Finite-dimensional systems

Before diving into the generalized backstepping, we start by taking a step back to look at a finite-dimensional system in order to get an intuition of the approach. A linear finite-dimensional system has the form

$$\dot{X} = AX + Bu, \tag{5.1.1}$$

where $X \in \mathbb{R}^n$ is the state, $A \in \mathbb{R}^{n\times n}$ is the operator, $B \in \mathbb{R}^{n\times 1}$ and $u \in \mathbb{R}^1$ is the scalar control. For such systems, it is a known fact that the system is stabilisable if it is controlable [55, Corollary 10.12]. And, thanks to the famous Kalman rank criterion, this system is controlable if and only if the pair (A, B) is controlable, i.e.

$$\mathrm{Span}\{A^i B | i \in \{0, ..., n-1\}\} = \mathbb{R}^n. \tag{5.1.2}$$

When this is satisfied, an explicit feedback $u(t) = KX$ can be found using for instance the formula derived from the Gramian (see [55, Chapter 10])

$$K = -B^T C_T^{-1}, \tag{5.1.3}$$

where

$$C_T = \int_0^T e^{-tA} B B^T e^{-tA^T} dt. \tag{5.1.4}$$

Then $(A + BK)$ has eigenvalues with only strictly negative real part, hence the system with feedback $u(t) = KX$ is exponentially stable (see [184, Theorem 3.1], [155]). In fact, one can even go further and show that for any polynomial $P \in \mathbb{R}_n[X]$ there exists K such that P is the characteristic polynomial of $(A + BK)$ [248]. Therefore for any $\lambda > 0$, one can find K such that $(A + BK)$ has only eigenvalues with real part lower than $-\lambda$ and hence the system is exponentially stable with decay rate (at least) λ. This is called pole-shifting.

In fact the pole shifting can be seen differently: we can see it as trying to transform the system (5.1.1) with $u = KX$ into the target system

$$\dot{Y} = (A - \lambda)Y, \tag{5.1.5}$$

with an invertible transformation T. This amounts to set $Y = TX$ and to show that it is possible to find T and K such that Y is solution to (5.1.5). Using (5.1.1), this means that

$$\dot{Y} = T(A + BK)X = (A - \lambda)Y = (A - \lambda)TX. \tag{5.1.6}$$

Setting $\tilde{A} = A - \lambda Id$, this amounts to showing that

$$T(A + BK) = \tilde{A}T. \tag{5.1.7}$$

This equation is not well-posed in the sense that if there exists a solution (T, K), then there exists an infinite number of solutions, for instance (aT, K) with $a \in \mathbb{R}$. Thus, there is some room for maneuvers on T. On the other hand, this equation is quadratic in (T, K) because of the term TBK that appears when developing the left-hand side. This makes it more complicated to solve. As a consequence, it is very tempting to impose an additional condition [1]

$$TB = B. \tag{5.1.8}$$

Then the system to solve becomes

$$\begin{aligned} TA + BK &= \tilde{A}T, \\ TB &= B, \end{aligned} \tag{5.1.9}$$

and it can be shown that this system has a unique solution (T, K) (see for instance [64, Section 2.2]). Overall, this illustrates that, in finite dimension, the generalized backstepping approach that we want to use is simply another way to see the pole-placement theorem, which is directly linked to the controllability of the system.

5.2 Infinite-dimensional systems

In infinite dimension, the linear systems of PDEs we are looking at can be formally reformulated as:

$$\partial_t f(t) = \mathcal{A} f(t) + B u(t), \tag{5.2.1}$$

where $\mathcal{A}$ is a differential operator defined on some domain $D(\mathcal{A})$, which contains the information about the boundary conditions, B is an operator associated to the control and u is the scalar control. For instance the following linear transport PDE

$$\begin{aligned} \partial_t f - \partial_x f &= \phi(x) u_1(t), \\ f(t, 0) &= k f(t, 1), \end{aligned} \tag{5.2.2}$$

1. In fact there is a deeper reason to consider this specific additional condition $TB = B$, other than making (5.1.7) linear. This appears when considering (A, B) under canonical form and using that $(\tilde{A}, B)$ is controllable. More details can be found in [64, Section 2.2].

with $\phi \in L^2(0,1)$ can be reformulated as (5.2.1) with $\mathcal{A} : f \to \partial_x f$, defined on $D(\mathcal{A}) = \{f \in H^1(0,1) \ : f(0) = kf(1)\}$ and $B : m \to m\phi$ defined on $\mathbb{R}$ with value in $L^2(0,1)$.

The target system can be reformulated similarly in

$$\partial_t g = \tilde{A} g, \tag{5.2.3}$$

where $\tilde{A}$ is defined on some $D(\tilde{\mathcal{A}})$. Formally, the problem we would like to solve is finding an isomorphism T and a feedback operator K such that, for a solution f to (5.2.1) with $u(t) = Kf(t)$, Tf is a solution to (5.2.3). This problem formulates very similarly to its the finite dimensional counterpart presented above and, once again, the formal operator problem that we would like to solve is finding (T, K) such that

$$T(\mathcal{A} + BK) = \tilde{\mathcal{A}} T, \tag{5.2.4}$$

in a sense to be defined. This problem is ill-posed again in the sense that there is no uniqueness [2]. Thus the finite dimensional approach suggests to add a condition $TB = B$ (in a sense to be defined) to simplify the problem and solve

$$T\mathcal{A} + BK = \tilde{\mathcal{A}} T \tag{5.2.5}$$

instead of solving directly $T(\mathcal{A} + BK) = \tilde{\mathcal{A}} T$. As a final remark before diving into the method, we can note that overall we are fundamentally trying to use the controllability of the system to show the existence of a transformation which, as a consequence, shows that the system can be rapidly stabilized, with an explicit feedback.

5.2.1 Overview of the method

The general spirit of the method is the following:

- First assume the feedback operator K is fixed and the condition $TB = B$ holds, and find a candidate transform T mapping the original system to the target one. This amounts to finding a solution T to the operator equality (5.2.5) described above.
- Then show a condition on K such that T is an isomorphism.
- Show that there exists K such that $TB = B$ holds and express it.

The first step is usually straightforward, while the second step is crucial. For the second step there is a useful characterization when working on a Hilbert space

Lemma 5.2.1. *Let T be a mapping from a Hilbert space X to itself. T is an isomorphism if and only if there exists an orthonormal basis $(f_n)_{n\in\mathbb{Z}}$ of X such that its image by T, denoted $(Tf_n)_{n\in\mathbb{Z}}$ is a Riesz basis of X.*

This reduces the problem to studying the property of $(Tf_n)_{n\in\mathbb{Z}}$ for some orthonormal basis $(f_n)_{n\in\mathbb{Z}}$. It remains the problem of showing that $(Tf_n)_{n\in\mathbb{Z}}$ is a Riesz basis. In the original approach introduced in [68, 69] (see also [63, 109, 111]) this was shown by using the following Lemma

Lemma 5.2.2. *Let $\mathcal{I} \subseteq \mathbb{Z}$ and $(\xi_n)_{n\in\mathcal{I}}$ be quadratically close to an orthonormal basis $(e_n)_{n\in\mathcal{I}}$ of X. Suppose that $(\xi_n)_{n\in\mathcal{I}}$ is either dense in X or ω-independent in X, then $(\xi_n)_{n\in\mathcal{I}}$ is a Riesz basis of X.*

The proper definitions of Riesz basis, dense, ω-independent, and quadratically close are given below in Definition 5.3.1. The key point here is then to show that the family $(Tf_n)_{n\in\mathbb{Z}}$ is quadratically close to some orthonormal basis, likely derived from $(f_n)_{n\in\mathbb{Z}}$. This is what we discuss in more details in Section 5.3 for the heat equation. As we will see it later on in Section 5.3 (see for instance (5.3.55)) the quadratically close property relies a lot on the fact that the eigenvalues of the operator $\mathcal{A}$ are increasing quickly enough (so that the gap between two eigenvalues increase quickly enough as well). In fact it amounts to showing that

$$\sum_{n\in\mathbb{N}^*} \sum_{p\neq n} \left(\frac{1}{\lambda_p + \lambda - \lambda_n} \right)^2 < +\infty, \tag{5.2.6}$$

2. if (T, K) is solution, (aT, K) is again a solution for $a \in \mathbb{R}^*$

which holds as soon as

$$\lambda_n \sim n^\alpha, \text{ with } \alpha > 3/2. \tag{5.2.7}$$

However, this can be very limiting as it does not work for operators with a slower eigenvalue growth. Indeed for $\alpha \leq 3/2$, (5.2.6) would fail. This excludes for instance hyperbolic systems where the growth of the eigenvalues typically scales with n (hence $\alpha = 1$), but also physical systems like the water-wave equations which correspond to the critical case $\alpha = 3/2$. This last problem is an open question mentionned in 2017 in [56]. In these cases, proving that $(Tf_n)_{n\mathbb{Z}}$ is quadratically close to $(f_n)_{n\in\mathbb{Z}}$ is most likely vain and there is a need for a new approach.

In Section 5.3 we present how this generalized approach can be adapted to the rapid stabilization of a heat equation on a torus in a sharp functional framework. In Chapter 6 we give a new approach: the compactness-duality method which relies on Fredholm's alternative and allows to overcome the quadratically close argument and to deal with the cases $\alpha \in (1, 3/2]$ for skew-adjoint operators. In particular, this allows to answer the question of the water-waves equation. Finally in Chapter 7 we focus on hyperbolic systems where $\alpha = 1$ and we see how to deal with a particular 2×2 system studied in [53, 54, 72, 98, 200, 206] consisting of Saint-Venant equations and modelling a water-tank.

5.3 Stabilization of a heat equation on a torus with two scalar controls

This Section is taken from [109], a collaboration with Ludovick Gagnon, Shengquan Xiang and Christophe Zhang. We consider the following heat equation on the torus

$$\begin{aligned} \partial_t y(t) - \Delta y(t) &= \Phi u(t), \quad t \in (0, +\infty), \\ y(0) &= y_0 \in H^s(\mathbb{T}), \end{aligned} \tag{5.3.1}$$

Where $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ is the one-dimensional torus, $s \in \mathbb{R}_+$, $u(t) \in \mathbb{R}^d$ are real-valued scalar controls belonging to $L^2((0, +\infty); \mathbb{R}^d)$, and Φ is a linear application on $\mathbb{R}^d$ which can be represented by the vector $(\phi_1, ..., \phi_d)$.

The first question that arises is: what is the smallest number d of controls such that the system is controllable? We show the following in [109, Section 3]:

Theorem 5.3.1. *If $d = 1$, then for any $T > 0$ the system (5.3.1) is not controllable.*

Note here that the controls are scalar, which means that they do not depend on the space variable. When the control also depends on x, it is then of infinite dimension and the room for maneuver is much larger (see [108, 171], or for instance [102, 105]).
Theorem 5.3.1 means that $d \geq 2$ is necessary. In fact, $d \geq 2$ is also sufficient and, if $d = 2$ a necessary condition on (ϕ_1, ϕ_2) is that

$$\begin{pmatrix} \langle \phi_1, e^{inx} \rangle & \langle \phi_1, e^{-inx} \rangle \\ \langle \phi_2, e^{inx} \rangle & \langle \phi_2, e^{-inx} \rangle \end{pmatrix} \text{ is invertible, for any } n \in \mathbb{N}^*. \tag{5.3.2}$$

A simple example for which this is satisfied is, for instance,

$$\begin{aligned} &\phi_1 \in \overline{\text{Span}_{\mathbb{R}}\{\sin(nx)\}_{n\in\mathbb{N}}} \text{ with } \langle \phi_1, \sin(nx) \rangle \neq 0, \quad \forall n \in \mathbb{N}^* \\ &\phi_2 \in \overline{\text{Span}_{\mathbb{R}}\{\cos(nx)\}_{n\in\mathbb{N}}} \text{ with } \langle \phi_2, \cos(nx) \rangle \neq 0, \quad \forall n \in \mathbb{N}. \end{aligned} \tag{5.3.3}$$

In the following we will consider such (ϕ_1, ϕ_2), to simplify. We introduce the notation

$$H^{s-}(\mathbb{T}) = \cup_{\varepsilon>0} H^{s-\varepsilon}(\mathbb{T}) \tag{5.3.4}$$

and

$$H^{s+}(\mathbb{T})' = \cup_{\varepsilon>0} (H^{s+\varepsilon}(\mathbb{T}))', \tag{5.3.5}$$

where $(H^s(\mathbb{T}))'$ is the dual of $H^s(\mathbb{T})$. We have the following [109]:

Theorem 5.3.2 (Controllability). *Assume that $d = 2$ and that there exist $-\infty < \alpha \leq \beta < 1/2$ and $c, C > 0$ such that*

$$cn^{\alpha} \leq |\langle \phi_1, \sin(nx)\rangle|, |\langle \phi_2, \cos(nx)\rangle| \leq Cn^{\beta}, \forall n \in \mathbb{N}^*, \tag{5.3.6}$$

and $\langle \phi_2, 1\rangle \neq 0$, then the system (5.3.1) is exact null controllable in $L^2(\mathbb{T})$. If, in addition,

$$cn^{-s} \leq |\langle \phi_1, \sin(nx)\rangle|, |\langle \phi_2, \cos(nx)\rangle| \leq Cn^{-s}, \ \forall n \in \mathbb{N}^*, \tag{5.3.7}$$

and $\langle \phi_2, 1\rangle \neq 0$, then the system is exact null controllable in $H^{s+1/2-}(\mathbb{T})$.

For $d > 2$ we can simply set all the other scalar controls to 0 and the system is again controllable under these conditions. Thus, in the following we will only consider $d = 2$, without loss of generality. The system reads

$$\begin{cases} \partial_t y(t) - \Delta y(t) = u_1(t)\phi_1 + u_2(t)\phi_2, \ \ x \in \mathbb{T}, \ \ t \in (0, +\infty), \\ u(0) = u_0, \ \ x \in \mathbb{T}. \end{cases} \tag{5.3.8}$$

Our main results are the following:

Theorem 5.3.3 (Rapid stabilization). *Let $s \in \mathbb{R}_+$ and $\phi_1, \phi_2 \in H^{s-1/2-}$ such that (5.3.7) holds. For any $\lambda > 0$, there exist K_1 and K_2 bounded feedback functionals on $H^{s+1/2+}$ such that for any $y_0 \in H^{s+r}$ with $r \in (-1/2, 1/2)$, the equation (5.3.8) with $u_1 = K_1(y)$ and $u_2 = K_2(y)$ has a unique solution satisfying*

$$y \in C^0([0, +\infty); H^{s+r}(\mathbb{T})) \cap L^2_{loc}((0, +\infty); H^{s+r+1}(\mathbb{T})) \cap H^1_{loc}((0, +\infty); H^{s+r-1}(\mathbb{T})). \tag{5.3.9}$$

Moreover, we have the following exponential stability estimate,

$$\|y(t, \cdot)\|_{H^{s+r}} \leq Ce^{-\lambda t}\|y_0\|_{H^{s+r}}, \quad \forall t \in [0, +\infty), \tag{5.3.10}$$

where $C = C_r(\lambda, s)$ is a constant independent of y_0.

We can make several interesting remarks

Remark 5.3.1 (Uniformity of the feedback with r). *The feedback laws K_1 and K_2 do not depend on $r \in (-1/2, 1/2)$, which means that for a given s, the same feedback law can stabilize the system in any of the H^{s+r} spaces with $r \in (-1/2, 1/2)$.*

Remark 5.3.2 (Case $a_0 = 0$). *The condition $a_0 \neq 0$ is necessary to have the controllability and stabilizability of the system. Otherwise, one can see that $\int_{\mathbb{T}} y(t, x)dx$ is conserved. However, if $a_0 = 0$ a stabilization is still possible in some sense, but, instead of converging to 0, the system will converge to the constant steady-state $y^* \equiv \int_{\mathbb{T}} y_0(x)dx$.*

We can extend this results and show that the feedback obtained in Theorem 5.3.3 can also stabilize locally nonlinear systems, such as the viscous Burgers' equation.

Theorem 5.3.4 (Rapid stabilization). *Let $\phi_1, \phi_2 \in H^{-1/2-}$, such that (5.3.7) holds with $s = 0$. For any $\lambda > 0$, there exists K_1 and K_2 bounded feedback functionals on $H^{1/2+}$ such that, for any $y_0 \in L^2$, the equation*

$$\begin{cases} \partial_t y - \Delta y + \partial_x(y^2/2) = K_1(y)\phi_1 + K_2(y)\phi_2, \\ y(0) = y_0, \end{cases} \tag{5.3.11}$$

has a unique solution

$$y \in C^0([0, +\infty); L^2(\mathbb{T})) \cap L^2_{loc}((0, +\infty); H^1(\mathbb{T})) \cap H^1_{loc}((0, +\infty); H^{-1}(\mathbb{T})). \tag{5.3.12}$$

Moreover, there exists $\delta > 0$ such that for any $\|y_0\|_{L^2} < \delta$, we have the following exponential stability estimate

$$\|y(t, \cdot)\|_{L^2} \leq Ce^{-\lambda t}\|y_0\|_{L^2}, \quad \forall t \in [0, +\infty), \tag{5.3.13}$$

where $C = C(\lambda)$ is a constant independent of y_0.

In the following, we give an idea of the proofs of Theorem 5.3.3 and 5.3.4. We start by introducing some definitions.

5.3.1 Functional setting and definitions

Functional setting Let us remark that $(e^{inx})_{n\in\mathbb{Z}}$ are eigenfunctions of the Laplacian on $\mathbb{T}$ associated to the eigenvalues $\lambda_n = \lambda_{-n} = -n^2$. This degeneracy of the eigenvalues is the main reason why the system needs at least two scalar controls to be controllable. This means that we can construct the following orthonormal basis of real-valued eigenfunctions of the Laplacian

$$\begin{aligned} f_n^1 := \frac{1}{\sqrt{\pi}}\sin(nx), \;\; f_n^2 := \frac{1}{\sqrt{\pi}}\cos(nx), \text{ associated to } \lambda_n := -n^2, \; \forall n \in \mathbb{N}^*, \\ f_0^2 := \frac{1}{\sqrt{2\pi}}, \text{ associated to } \lambda_0 := 0. \end{aligned} \tag{5.3.14}$$

Also, on $\mathbb{T}$ there are no boundary thus $H^s(\mathbb{T})$ coincides with the closure of $\mathrm{Span}\{n^{-s}e^{inx}\}_{n\in\mathbb{Z}} = \mathrm{Span}\{(n^{-s}\sin(nx), n^{-s}\cos(nx))\}_{n\in\mathbb{N}}$. This motivates the definition of the following spaces

$$L^2(\mathbb{T}) = L_1^2 \oplus L_2^2, \;\; L_1^2 := \overline{\mathrm{Span}_{\mathbb{R}}\{\sin nx\}}_{n\in\mathbb{N}^*}, \;\; L_2^2 := \overline{\mathrm{Span}_{\mathbb{R}}\{\cos nx\}}_{n\in\mathbb{N}}. \tag{5.3.15}$$

Thus, L_1^2 describes the odd functions, L_2^2 describes the even functions and both subspaces of L^2 are endowed with the L^2 norm. Similarly, we can define, for any $s \in \mathbb{R}_+$ and $i \in \{1, 2\}$,

$$H_i^s = \{a \in H^s(\mathbb{T}) \mid a = \sum_{n\in\mathbb{N}} a_n f_n^i\}, \tag{5.3.16}$$

where we recall that, with such a basis $(f_n^1, f_n^2)_{n\in\mathbb{N}}$,

$$H^s(\mathbb{T}) = \left\{ a = \sum_{n\in\mathbb{N}^*} a_n^1 f_n^1 + \sum_{n\in\mathbb{N}} a_n^2 f_n^2 \;\middle|\; a_n^i \in \mathbb{R} \text{ and } \sum_{n\in\mathbb{N}^*} n^{2s}\left((a_n^1)^2 + (a_n^2)^2\right) < +\infty \right\}, \;\; \forall\, s \in \mathbb{R}_+, \tag{5.3.17}$$

and

$$\|a\|_{H^m} = \left((a_0^2)^2 + \sum_{n\in\mathbb{N}} n^{2s}((a_n^1)^2 + (a_n^2)^2) \right)^{1/2}. \tag{5.3.18}$$

We can define the inner product $\langle\cdot,\cdot\rangle_{H_i^{s-m}, H_i^{s+m}}$ as follows

$$\langle f, g\rangle_{H_1^{s-m}, H_1^{s+m}} = \sum_{n\in\mathbb{N}} (n^{s-m}\langle f, f_n^1\rangle)(n^{s+m}\langle g, f_n^1\rangle). \tag{5.3.19}$$

where $\langle\cdot,\cdot\rangle$ refers to the usual scalar product in L^2. We also introduce the Schwartz space on the torus $\mathcal{S}(\mathbb{T})$ and decompose it similarly in its odd and even part

$$\begin{aligned} \mathcal{S}_1 &:= \{a \in \mathcal{S}(\mathbb{T}) : \langle a, f_n^2\rangle = 0, \;\; \forall\, n \in \mathbb{N}\}, \\ \mathcal{S}_1 &:= \{a \in \mathcal{S}(\mathbb{T}) : \langle a, f_n^1\rangle = 0, \;\; \forall\, n \in \mathbb{N}^*\}. \end{aligned} \tag{5.3.20}$$

and we define

$$\mathcal{S}_1' := \left\{a \in \mathcal{S}' \mid \langle a, f_n^2\rangle = 0, \; \forall n \in \mathbb{N}\right\}, \tag{5.3.21}$$

$$\mathcal{S}_2' := \left\{a \in \mathcal{S}' \mid \langle a, f_n^1\rangle = 0, \; \forall n \in \mathbb{N}^*\right\}, \tag{5.3.22}$$

In the following we will refer to S_1' and S_2' as the dual of S_1 and S_2 respectively (even if, strictly speaking, it is not the dual of $\mathcal{S}_k$, but the quotient space $\mathcal{S}'/\mathcal{S}_{3-k}$). Finally, we say that $\mathcal{L}$ is a bounded linear map from $H^{s+}(\mathbb{T})$ to $\mathbb{R}$ if

$$\text{For any } \varepsilon > 0, \; \mathcal{L} \text{ is a linear map from } H^{s+\varepsilon}(\mathbb{T}) \text{ to } \mathbb{R}. \tag{5.3.23}$$

A similar definition holds for H_i^{s+}, with $i = 1$ or 2.

Families of functions in a Hilbert space As mentioned in introduction of this section, to prove that our candidate backstepping transform T is indeed an isomorphism, we will use Lemma 5.2.1 and show that there exists an orthonormal basis $(f_n)_{n\in\mathbb{N}}$ such that $(Tf_n)_{n\in\mathbb{N}}$ is a Riesz basis. To do so, we will use Lemma 5.2.2. We recall here some definitions about families of functions in Hilbert spaces.

Definition 5.3.1. Let X be a Hilbert space. A family of vectors $\{\xi_n\}_{n\in\mathcal{I}}$, where $\mathcal{I}=\mathbb{Z}$, $\mathbb{N}$, or $\mathbb{N}^*$ is said to be

(1) **Minimal** in X, if for every $k\in\mathcal{I}$, $\xi_k\notin\overline{\text{Span}\{\xi_i; i\in\mathcal{I}-\{k\}\}}$.
(2) **Dense** in X, if $\overline{\text{Span}\{\xi_i; i\in\mathcal{I}\}}=X$.
(3) **ω-independent** in X, if

$$\sum_{k\in\mathcal{I}} c_k\xi_k=0 \text{ in } X \text{ with } \{c_n\}_{n\in\mathcal{I}}\in\ell^2(\mathcal{I})\Longrightarrow c_n=0,\ \forall n\in\mathcal{I}. \tag{5.3.24}$$

(4) **Quadratically close** to a family of vector $\{e_n\}_{n\in\mathcal{I}}$, if

$$\sum_{k\in\mathcal{I}}\|\xi_k-e_k\|_X^2<+\infty. \tag{5.3.25}$$

(5) **Riesz basis** of X, if it is the image by an isomorphism (on X) of some orthonormal basis.
(5)' **Riesz basis** of X (equivalent definition), if it is dense in X and if there exist $C_1,C_2>0$ such that for any $(a_n)_{n\in\mathcal{I}}\in\ell^2(\mathcal{I})$ we have

$$C_1\sum_{k\in\mathcal{I}}|a_k|^2\leq\|\sum_{k\in\mathcal{I}}a_k\xi_k\|_X^2\leq C_2\sum_{k\in\mathcal{I}}|a_k|^2. \tag{5.3.26}$$

5.3.2 Outline of the strategy

We give here the general idea of the proof. Given that $\phi_1\in H_1^{s-1/2-}$ and $\phi_2\in H_2^{s-1/2-}$, the previous definitions motivate us to divide the state of the system u in two components

$$y(t,\cdot)=y_1(t,\cdot)+y_2(t,\cdot), \tag{5.3.27}$$

where $y_1\in L_2^1$ and $y_2\in L_2^2$. Then one can check that

$$\partial_t y_i(t)-\Delta y_i(t)=u_i(t)\phi_i,\ \ \forall t\in(0,+\infty),\ i\in\{1,2\}, y_i(0)=y_i^0, \tag{5.3.28}$$

where $y^0=y_1^0+y_2^0$ is the decomposition of the initial condition in odd and even functions. The logic is that the odd part of the solution is then controlled by the odd part of the control $u_1\phi_1$ and the even part of the solution is controlled by the even part of the control $u_2\phi_2$. This gives us two problems defined on H_i^s. What we are showing in [109] is that each of these two systems can be stabilized rapidly using a feedback control, and this gives the main result Theorem 5.3.3.

For this, we would like to map each of these systems to the target systems

$$\begin{cases}\partial_t z_i(t)-\Delta z_i(t)-\lambda z_i=0,\ \ \forall t\in(0,+\infty),\ i\in\{1,2\},\\ z_i(0)=z_i^0,\end{cases} \tag{5.3.29}$$

where $\lambda>0$ is an arbitrary constant. Indeed, this would bring the result as one can easily check that any H_i^m solution of (5.3.29) satisfies

$$\|z_i\|_{H^m}\leq e^{-\lambda t}\|z_i^0\|_{H_i^m}. \tag{5.3.30}$$

Therefore, our goal will be to show that for any $\lambda>0$ and any $i\in\{1,2\}$, there exists an isomorphism $T_i(\lambda):H_i^s\to H_i^s$ as well as a feedback $u_i(t):=K_i(\lambda)y_i(t,x)$ such that for any initial condition $y_i^0\in H_i^s$, there exists unique solution y_i of (5.3.28) and $T_i(\lambda)y_i$ is a solution to (5.3.29). In fact, we would like more that: for s and λ given, we would like the well-posedness of y_i and the exponential stability result to hold in

H_1^{s+r} for any $r \in (-1/2, 1/2)$, while $K_i(\lambda)$ and $T_i(\lambda)$ still only depend on s and λ. This requires in particular K_i to be a functional on $H^{s+1/2+}$ and $T_i(\lambda)$ to be an isomorphism on H^{s+r} for any $r \in (-1/2, 1/2)$.

As mentioned in Section 5.2, what we are looking for, formally, is $T_i(\lambda)$ and $K_i(\lambda)$ such that (5.2.4) holds. Denoting $\mathcal{A} = \Delta$, this is

$$T_i\mathcal{A} + T_i\phi_i K_i = \mathcal{A}' T_i = (\mathcal{A} - \lambda Id)T_i, \tag{5.3.31}$$

$$T_i\phi_i = \phi_i, \tag{5.3.32}$$

in a sense to be defined (in particular (5.3.31) and (5.3.32) might not have a meaning in a strong sense). The existence of $T_i(\lambda)$, $K_i(\lambda)$ such that this holds is given by the following proposition.

Proposition 5.3.5 (Main proposition)**.** *Let the countable set*

$$\mathcal{N} := \{i^2 - j^2 : i, j \in \mathbb{N}\}, \tag{5.3.33}$$

and denote $a_n^1 = \langle \phi_1, \sin(nx)\rangle$ *for* $n \in \mathbb{N}^*$ *and* $a_n^2 = \langle \phi_2, \cos(nx)\rangle$ *for* $n \in \mathbb{N}$*. Let* $s \in \mathbb{R}$ *and* $i \in \{1, 2\}$ *and assume that*

$$cn^{-s} < |a_n^i| < Cn^{-s}, \quad \text{for } i \in \{1, 2\}, \quad \text{for } n \in \mathbb{N}^*, \tag{5.3.34}$$

$$a_0^2 \neq 0. \tag{5.3.35}$$

Then for any $\lambda \notin \mathcal{N}$*, there exists a sequence* $(K_n^i(\lambda))_n$ *satisfying*

$$\begin{aligned} K_0^2(\lambda) &\neq 0, \\ cn^s < |K_n^i(\lambda)| &< Cn^s, \quad \text{for } i \in \{1, 2\}, \quad \text{for } n \in \mathbb{N}^*, \\ \{(\lambda + a_n^i K_n^i(\lambda))n^r\}_n &\in \ell^2, \quad \forall r \in [0, 1/2), \\ K_i(\lambda) \text{ is a bounded functional on } &H_i^{s+1/2+}, \end{aligned}$$

such that the linear operator $T_i(\lambda)$ *defined as follows*

$$T_i(\lambda) : \mathcal{S}_i \to \mathcal{S}_i', \tag{5.3.36}$$

$$f_n^i \mapsto -K_n^i(\lambda) \sum_p \frac{a_p f_p^i}{p^2 + \lambda - n^2}, \tag{5.3.37}$$

$$f_n^{3-i} \mapsto 0, \tag{5.3.38}$$

can be linearly extended to $H^{s-3/2+}$*, and*

$$T_i(\lambda) \text{ is an isomorphism on } H_i^{s+m} \text{ for any } m \in (-3/2, 3/2), \tag{5.3.39}$$

$$T_i(\lambda)\phi_i = \phi_i \text{ in } H_i^{s-1/2-}, \tag{5.3.40}$$

and moreover, for any $r \in (-1/2, 1/2)$*, for any* $\chi \in H_i^{s+r+1}$ *we have*

$$(T_i(\lambda)A + T_i(\lambda)\phi_i K_i(\lambda))\chi = (AT_i(\lambda) - \lambda T_i(\lambda))\chi \text{ in } H_i^{s+r-1}. \tag{5.3.41}$$

We remark several things about this proposition:

- First, λ cannot belong to $\mathbb{R}$ but only to $\mathbb{R} \setminus \{\mathcal{N}\}$, where $\mathcal{N}$ is a countable set. This, however, is not limiting as λ can still be as large as desired.
- Second, we can note that the coefficients a_n^i of ϕ_i have to satisfy an estimate (5.3.34), which corresponds to the condition (5.3.7) we required to ensure the controllability in $H^{s+1/2-}$.
- Third, the operator equalities (5.3.32) only holds in $H^{s-1/2-}$ while (5.3.31) only holds for functions in H^{s+r+1} for any $r \in (-1/2, 1/2)$. In fact $T_i\phi_i$ could not hold in H_i^s, given that $\phi_i \in H_i^{s-1/2-}$, and $T_i\mathcal{A} + \phi_i K_i = \mathcal{A} - \lambda Id$ could not hold when applied on less regular functions either, given that K_i is only defined on $H_i^{s+1/2+}$.

— Fourth, the bounds we get on m and r are optimal with respect to the method.

Thanks to the decomposition $H^s = H_1^s \oplus H_2^s$, Proposition 5.3.5 gives the following corollary

Corollary 2. *Under the assumption of Proposition 5.3.5, the transformation T defined as*

$$Tf = T_1 f_1 + T_2 f_2, \quad \text{for all } f = f_1 + f_2 \in S_1 \oplus S_2, \tag{5.3.42}$$

can be linearly extended on $H^{s-3/2+}$. Moreover,

$$T \text{ is an isomorphism on } H^{s+m} \text{ for any } m \in (-3/2, 3/2),$$

and, for any $r \in (-1/2, 1/2)$, for any $\chi \in H^{s+r+1}$,

$$(TA + TBK)\chi = (AT - \lambda T)\chi \quad \text{in } H^{s+r-1}, \tag{5.3.43}$$

$$TB = B \quad \text{in } H^{s-1/2-}. \tag{5.3.44}$$

Once this corollary is proved, Theorem 5.3.3 follows provided that the system (5.3.8) is well-posed with the feedback control we constructed. We will study the well-posedness later in Section 5.3.6.1.

5.3.3 Ideas of the proof

In this Section we give some ideas about how to prove Proposition 5.3.5. In the following, we deal with the case $i = 1$ and we will drop the index i for clarity. Namely, we denote T_1, K_1, f_n^1, H_1^s by T, K, f_n and H^s. We also denote the coefficients of ϕ by a_n, namely

$$a_n = \langle \phi, \sin(nx) \rangle, \quad \forall\, n \in \mathbb{N}^*. \tag{5.3.45}$$

5.3.3.1 Constructing a candidate transform T

Following the general summary of the method given in Section 5.2.1, we start by constructing a candidate transform T by assuming that the following operator equality holds

$$T\mathcal{A} + \phi K = (\mathcal{A} - \lambda Id)T, \tag{5.3.46}$$

for K to be defined. As K is a linear operator, it is entirely defined by the family $(K_n)_{n\in\mathbb{N}^*} := (Kf_n)_{n\in\mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}^*}$. Projecting (5.3.46) on the eigenfunctions $(f_n)_{n\in\mathbb{N}^*}$, this becomes

$$T(\mathcal{A} f_n) + K_n \phi = \mathcal{A}(Tf_n) - \lambda(Tf_n). \tag{5.3.47}$$

As f_n is an eigenvector of $\mathcal{A} = \Delta$, this gives

$$\lambda_n(Tf_n) + K_n\phi = \mathcal{A}(Tf_n) - \lambda(Tf_n), \tag{5.3.48}$$

which is a differential equation on $h_n := (Tf_n)$. We can now project this equation on f_p for [3] $p \in \mathbb{N}^*$, which gives, using the fact that $\mathcal{A} = \Delta$ is self adjoint.

$$\begin{aligned} \lambda_n \langle h_n, f_p \rangle + K_n \langle \phi, f_p \rangle &= \langle \Delta h_n, f_p \rangle - \lambda \langle h_n, f_p \rangle, \\ &= (\lambda_p - \lambda)\langle h_n, f_p \rangle, \end{aligned} \tag{5.3.49}$$

Hence, as long as $\lambda \in \mathbb{R} \setminus \mathcal{N}$, and using the fact that $\langle \varphi, f_p \rangle = a_p$

$$\langle h_n, f_p \rangle = \frac{-K_n a_p}{\lambda_n - \lambda_p + \lambda}. \tag{5.3.50}$$

We have now characterized $h_n = Tf_n$. We remark that Tf_n is imposed, up to the feedback operator K which remains to be chosen. In order to show that this is an isomorphism we will show that $(Tf_n)_{n\in\mathbb{N}^*}$ is quadratically close to an orthonormal basis derived from f_n, and this will give a condition on K. For this we introduce the families $(q_n)_{n\in\mathbb{N}^*}$ and $(g_n)_{n\in\mathbb{N}^*}$ defined by:

$$q_n := \frac{h_n}{K_n} = \sum_{p\in\mathbb{N}^*} \frac{-a_p}{\lambda_n - \lambda_p + \lambda} f_p, \quad g_n := \sum_{p\in\mathbb{N}^*} \frac{f_p}{\lambda_n - \lambda_p + \lambda}. \tag{5.3.51}$$

3. In general, at this stage we would like to project on a family of eigenfunctions of $\tilde{\mathcal{A}}^*$, the adjoint of the operator $\tilde{\mathcal{A}}$ of the target system. However, here $\tilde{\mathcal{A}} = \Delta - \lambda I_d$ and is self-adjoint so the family of eigenfunctions is again $(f_n)_{n\in\mathbb{N}^*}$

5.3.4 Riesz basis property

Obviously there exists an isomorphism linking g_n and q_n (which is itself $(Tf_n/K_n)_{n\in\mathbb{N}^*}$). The interest of studying q_n is that it does not depend on K, while q_n does not depends on ϕ either. What we show is the following

Lemma 5.3.1. *Let $s \geq 0$. Let $a_n \neq 0$ such that $cn^{-s} < |a_n| < Cn^{-s}$. Let $\lambda \notin \mathcal{N}$. The following properties hold:*

(1) $\{g_n\}_{n\in\mathbb{N}^*}$ is a Riesz basis of L^2_1.

(2) Let $m \in (-3/2, 3/2)$. Then $\{n^{-m}g_n\}_{n\in\mathbb{N}^*}$ is a Riesz basis of H^m_1.

(3) Let $m \in (-3/2, 3/2)$. Then $\{n^{-m}q_n\}_{n\in\mathbb{N}^*}$ is a Riesz basis of H^{s+m}_1.

(4) Let $m \in (-3/2, 3/2)$. If $K_n := Kf_n$ satisfies $|K_n| < Cn^s$, then the transformation $T : H^{s+m}_1 \to H^{s+m}_1$ is bounded. Moreover, if

$$cn^s < |K_n| < Cn^s, \tag{5.3.52}$$

then the transformation $T : H^{s+m}_1 \to H^{s+m}_1$ is an isomorphism.

Let us remark here that all the choices of s and m in the above are sharp.

Before giving some details about the proof of 5.3.1 we can make a few remarks:

— (1) and (2) are dealing with the fact that $(g_n)_{n\in\mathbb{N}^*}$ is a Riesz basis. As $(g_n)_{n\in\mathbb{N}^*}$ does not depend on K or ϕ, this is really only a property of the operator $\mathcal{A} = \Delta$.

— The fact that $(n^{-m}q_n)_{n\in\mathbb{N}^*}$ is a Riesz basis of H^{s+m}_1 is a direct consequence from the fact that $(n^{-m}g_n)_{n\in\mathbb{N}^*}$ is a Riesz basis of H^m_1 and the fact that there exists an isomorphism between g_n and q_n defined by

$$\tau : n^{-m}f_n \to n^{-m}a_nf_n \tag{5.3.53}$$

which is also an isomorphism from H^m to H^{s+m} given that with $cn^{-s} < |a_p| < Cn^{-s}$.

— Finally, as $n^{-(s+m)}f_n$ is an orthonormal basis of H^{s+m}_1, the point (4) can be reduced to looking at where $(T(n^{-(s+m)}f_n))_{n\in\mathbb{N}^*}$ belongs and whether or not $(T(n^{-(s+m)}f_n))_{n\in\mathbb{N}^*}$ is a Riesz basis of H^{s+m}_1. Therefore, this is a consequence of the fact that $(n^{-m}q_n)_{n\in\mathbb{N}^*}$ is a Riesz basis of H^{s+m}_1, the fact that $(Tf_n) = K_nq_n$, and the upper and lower bounds of K_n given by (5.3.52). In particular, assuming that K_n/n^s is uniformly bounded by above and below then $(\frac{K_n}{n^s}(n^{-m}q_n))$ is a Riesz basis of H^{s+m}_1 which means that $(T(n^{-(s+m)}f_n))_{n\in\mathbb{N}^*}$ is a Riesz basis of H^{s+m}_1.

We give below some ideas about how to prove (1). As announced before, we are going to use Lemma 5.2.2 to show that $(g_n)_{n\in\mathbb{N}^*}$ is a Riesz basis of L^2_1. This requires two things: showing that this family is dense or ω-independent, and showing that this family is quadratically close to an orthonormal basis of L^2_1.

— **quadratically close**. We first show that $(g_n)_{n\in\mathbb{N}^*}$ is quadratically close to $(f_n/\lambda)_{n\in\mathbb{N}^*}$, which is (obviously) an orthonormal basis of L^2_1. Given Definition 5.3.1 and the definition of $(g_n)_{n\in\mathbb{N}^*}$, this amounts to showing that

$$\left(\sum_{p\in\mathbb{N}^*}\frac{f_p}{\lambda_n-\lambda_p+\lambda}\right)_{n\in\mathbb{N}^*} \text{ is quadratically close to } \left(\frac{f_n}{\lambda}\right)_{n\in\mathbb{N}^*} \text{ in } L^2_1. \tag{5.3.54}$$

or that

$$\sum_{n\in\mathbb{N}^*}\sum_{p\neq n}\left(\frac{1}{\lambda_p+\lambda-\lambda_n}\right)^2 = \sum_{n\in\mathbb{N}^*}\sum_{p\neq n}\left(\frac{1}{p^2+\lambda-n^2}\right)^2 < +\infty, \tag{5.3.55}$$

Once (5.3.55) is showed for $\lambda \in \mathbb{N}^*$, it can then be extended to $\lambda \in \mathbb{R} \setminus \mathcal{N}$. The fact that we deal with $\lambda_n = n^2$ and $\lambda_p = p^2$ here is important, as this is what allows the sum in (5.3.55) to converge. Indeed, one can check that if we had $\lambda_n = n$ and $\lambda_p = p$ in (5.3.55) instead, the sum would not converge. This is where the main limitation of this method appears. We will come back to this point later in Chapters 6 and 7.

— **dense or ω-independent** Showing that $(g_n)_{n\in\mathbb{N}^*}$ is either dense in L^2_1 or ω-independent can be done by noticing first that

$$\mathcal{A}^{-1}g_n = \Delta^{-1}g_n = (n^2-\lambda)^{-1}g_n - (n^2-\lambda)^{-1}\mathcal{A}^{-1}h, \tag{5.3.56}$$

and assuming that $(g_n)_{n\in\mathbb{N}^*}$ is not ω-independent (otherwise the proof is done). Then, we deduce the existence of

$$\sum_{n\in\mathbb{N}^*} c_n g_n = 0, \text{ in } L^2_1. \tag{5.3.57}$$

The preceding formula is well-defined since, thanks to the quadratically close property (5.3.54),

$$\sum_{n\in\mathbb{N}^*} c_n g_n = \sum_{n\in\mathbb{N}^*} c_n \frac{f_n}{\lambda} + \sum_{n\in\mathbb{N}^*} c_n \left(g_n - \frac{f_n}{\lambda}\right) \tag{5.3.58}$$

converges in L^2_1. Next, by applying $\mathcal{A}^{-1}$ to this equality we conclude

$$\sum_{n\in\mathbb{N}^*} c_n k_n g_n = \sum_{n\in\mathbb{N}^*} c_n k_n \mathcal{A}^{-1}h, \text{ in } L^2_1, \tag{5.3.59}$$

where we have used the fact $\sum_n c_n k_n$ converges. We then iterate and show that

$$\sum_{n\in\mathbb{N}^*} c_n k_n^m g_n = \sum_{i=1}^{m} C_{m+1-i}\mathcal{A}^{-i}h, \tag{5.3.60}$$

with

$$C_l := \sum_{n\in\mathbb{N}^*} c_n k_n^l < +\infty. \tag{5.3.61}$$

From this point, there are only two possibilities: either there exists $m \geq 1$ such that $C_m \neq 0$ and we can show that $\{g_n\}_{n\in\mathbb{N}^*}$ is dense, or $C_m = 0$ for any $m \in \mathbb{N}^*$ and we get contradiction using that the complex function

$$\tilde{G}(z) = \sum_{n\in\mathbb{N}^*} c_n k_n e^{k_n z} \tag{5.3.62}$$

is holomorphic (and hence identically equal to 0 from (5.3.61)). In both cases we deduce that $(g_n)_{n\in\mathbb{N}^*}$ is either dense in L^2_1 or ω-independent.

5.3.5 Smoothing effect

Before going any further, recall that we want a result that ensures the stabilisation on a whole range of space H_1^{s+r} for any $r \in (-1/2, 1/2)$, while T and K should only depend on λ and s. For this we need the following smoothing effects:

Lemma 5.3.2. *Let $s \geq 0$. Let $a_n \neq 0$ such that $cn^{-s} < |a_n| < Cn^{-s}$. Let $\lambda \notin \mathcal{N}$. The following properties hold:*

(1) Let $r \in [0, 1/2)$. Then, q_n has the following smoothing property,

$$\sum_{n\in\mathbb{N}^*} \|q_n - a_n f_n/\lambda\|^2_{H_1^{s+r}} < +\infty. \tag{5.3.63}$$

(2) Let $r \in [0, 1/2)$. Similar smoothing effect also holds in the space H^{-1+s},

$$\sum_{n\in\mathbb{N}^*} \|n(q_n - a_n f_n/\lambda)\|^2_{H_1^{-1+s+r}} < +\infty. \tag{5.3.64}$$

(3) Let $s = 0$ and $r \in [0, 1/2)$. Then,

$$\sum_{n\in\mathbb{N}^*} \left(q_n - \frac{a_n f_n}{\lambda}\right) \in H_1^r. \tag{5.3.65}$$

Remark 5.3.3 (Sharpness of the functional setting)**.** *The choice of the bounds for r in the above is sharp.*

The point (1) and (2) follows from direct, although careful, estimations that will not be detailed here. Concerning point (3), note that $\sum_{n\in\mathbb{N}^*}(q_n - \frac{a_n f_n}{\lambda})$ belongs to H^{-1} a priori, given that $(nq_n)_{n\in\mathbb{N}^*}$ is a Riesz basis of H^{-1} from Lemma 5.3.1. The goal is to show that it actually belongs to H^r for $r \in [0, 1/2)$ even though $\sum_{n\in\mathbb{N}^*} q_n$ does not belong to $H^{-1/2}$ and $\sum_{n\in\mathbb{N}^*} \frac{a_n f_n}{\lambda}$ does not either. The smoothing comes from the cancelation of the singular parts. In other words f_n/λ contains the singular parts of q_n. From the definition of q_n given in (5.3.51), this is equivalent to showing that

$$\left\| \sum_n \sum_{p\neq n} \frac{a_p f_p}{p^2 + \lambda - n^2} \right\|^2_{L^2} < +\infty. \tag{5.3.66}$$

Note that this cannot be deduced from the quadratically close inequality (5.3.55) which would correspond in this case to

$$\sum_n \left\| \sum_{p\neq n} \frac{a_p f_p}{p^2 + \lambda - n^2} \right\|^2_{L^2} < +\infty. \tag{5.3.67}$$

Therefore (5.3.66) needs a more careful estimate that we will not detail here but that can be found in [109, Section 4].

5.3.6 Finding the feedback operator K

With Lemmas 5.3.1-5.3.2, we have a candidate transform T defined by its action on f_n. And we have a condition on K given by (5.3.52) under which our candidate T is an isomorphism. The next step is to find K such that this candidate transform is suitable. In particular, to check that we can obtain $T\phi = \phi$, at least weakly, provided some additional conditions on K.

Let assume to simplify that $s = 0$. It is clear that $T\phi = \phi$ cannot hold in L^2_1 given that ϕ only belongs to $H^{-1/2-}$, however it might be possible to find K such that it holds in $H_1^{-1/2-}$ (or at least H^σ for $\sigma \in (-3/2, -1/2)$ given that T is only defined for $\sigma > 3/2$). Let us start by showing this equality in H^{-1}. We know from Lemma 5.3.1 that T is defined on H^s_1 for any $s \in (-3/2, 3/2)$, hence it is defined on H^{-1} and

$$T\phi = -\sum_{n\in\mathbb{N}^*} a_n K_n q_n = \sum_{n\in\mathbb{N}^*} \frac{a_n K_n}{n}(nq_n). \tag{5.3.68}$$

From Lemma 5.3.1, $(nq_n)_{n\in\mathbb{N}^*}$ is a Riesz basis of H^{-1}, therefore there exists a unique family $(c_n)_{n\in\mathbb{N}^*} \in \ell^2$ such that

$$\phi = \sum_{n\in\mathbb{N}^*} c_n(nq_n). \tag{5.3.69}$$

This, together with (5.3.68), means that setting

$$K_n = -\frac{nc_n}{a_n} \tag{5.3.70}$$

gives

$$T\phi = \phi \ \text{ in } H_1^{-1}. \tag{5.3.71}$$

And, in fact, as ϕ is more regular than H^{-1} and belongs to $H^{-1/2-}$, one can get for any $\sigma \in (-3/2, -1/2)$

$$(n^{\sigma+1}c_n)_{n\in\mathbb{N}^*} \in \ell^2. \tag{5.3.72}$$

So, with the same choice of K_n,

$$T\phi = \phi \ \text{ in } H_1^{\sigma}, \quad \text{for all } \sigma \in (-3/2, 1/2). \tag{5.3.73}$$

We just showed $T\phi = \phi$ as intended and this sets K_n. However, recall that we need K_n to satisfy some uniform upper and lower bounds to ensure that T is an isomorphism (see (5.3.52)). In the current case $s = 0$, this means that we need to check that there exists $C > 0$ and $c > 0$ such that

$$c < |K_n| < C, \ \ \forall n \in \mathbb{N}^*. \tag{5.3.74}$$

To show the upper bound, the first thing we can note is that, for any $\varepsilon > 0$, $n^{1/2-\varepsilon}c_n \in \ell^2$, a_n is uniformly bounded by assumption and $|n^{1/2-\varepsilon}c_n| = n^{1/2-\varepsilon}|a_nK_n|/n$. However, this is not enough to show directly that K_n is uniformly bounded by above. Indeed, let us look for instance at $b_n := \log n$, we can easily observe that

$$\left(n^{1/2-\varepsilon}\frac{b_n}{n}\right)_{n\in\mathbb{N}^*} = \left(\frac{b_n}{n^{1/2+\varepsilon}}\right)_{n\in\mathbb{N}^*} \in \ell^2, \ \forall \varepsilon > 0. \tag{5.3.75}$$

So, to show that K_n is uniformly bounded, we need to use more information. And in particular we use the smoothing effect stated in Lemma 5.3.2. We first define

$$d_n = -a_nK_n - \lambda = nc_n - \lambda. \tag{5.3.76}$$

As $T\phi = \phi$ in H_1^{-1}, we have

$$\sum_n (\lambda + d_n)q_n = \sum_n a_nf_n \ \text{ in } H_1^{-1}, \tag{5.3.77}$$

thus using that $\lambda\sum_n \frac{1}{n}(nq_n) \in H_1^{-1}$,

$$\sum_n d_nq_n = \sum_n (a_nf_n - \lambda q_n) \ \text{ in } H_1^{-1}. \tag{5.3.78}$$

But, in fact, thanks to the smoothing effect of Lemma 5.3.2 (3) with $r = 0$, we have

$$\sum_n d_nq_n = \sum_n (a_nf_n - \lambda q_n) \in L_1^2. \tag{5.3.79}$$

As $(q_n)_{n\in\mathbb{N}^*}$ is a Riesz basis of L_1^2,

$$(d_n)_{n\in\mathbb{N}^*} \in \ell^2. \tag{5.3.80}$$

From (5.3.76) and the fact that a_n is uniformly bounded, this implies that K_n is uniformly bounded.

From Lemma 5.3.1, this upper bound is enough to show that T is a well-defined and bounded linear operator from H_1^m to H_1^m where $m \in (-3/2, 3/2)$. However, without more information on K_n we do not know yet that T is an isomorphism on these spaces. For this, we need a lower bound on K_n as in (5.3.74). Instead of finding this lower bound now, we show first that T is an isomorphism on H_1^{-1} and then we find a lower bound on K_n as a consequence. To show that T is an isomorphism on H_1^{-1}, we show that T is a Fredholm operator (of index 0) from H_1^{-1} to itself, as stated in this Lemma

Lemma 5.3.3. *Let $r \in [0, 1/2)$. The operator*

$$\begin{aligned} \tilde{T} &:= T - Id : L_1^2 \to H_1^r, \\ (\textit{resp. } \tilde{T} &:= T - Id : H_1^{-1} \to H_1^{-1+r}) \end{aligned} \tag{5.3.81}$$

is a continuous operator.

In particular, $\tilde{T} : L^2_1 \to L^2_1$ (resp. $T : H_1^{-1} \to H_1^{-1}$), is a compact operator and T is a Fredholm operator of index 0 on L^2_1 (resp. on H_1^{-1}).

The second part of the Lemma is a direct consequence of the first one since H^r is compact in L^2_1 (resp. H_1^{-1+r} is compact in H_1^{-1}). To show this Lemma, we proceed as follows: first we can check that T and K thus defined are indeed a solution of the operator equalities (5.3.31)–(5.3.32). The $T\phi = \phi$ equality (5.3.32) holds in a weak sense and is given by (5.3.73). The operator equality (5.3.31) also holds in a weak sense:

Proposition 5.3.6. *Let T constructed by Lemma 5.3.1 and $(K_n)_{n\in\mathbb{N}^*}$ be chosen as (5.3.70). For any $\chi \in H_1^{r+1}$ we have*

$$(TA + T\phi K)\chi = (AT - \lambda T)\chi \text{ in } H_1^{r-1}. \tag{5.3.82}$$

In particular we can consider $r = 0$, then the equality holds in H_1^{-1}.

Then we show the following technical lemma, which comes from a direct estimation and the fact that $(d_n)_{n\in\mathbb{N}^*}$ defined in (5.3.76) belongs to ℓ^2.

Lemma 5.3.4. *Let $r \in [0, 1/2)$. There exists a constant $C > 0$ such that*

$$\left\| \sum_n b_n (q_n - \frac{a_n f_n}{\lambda}) \right\|^2_{H_1^r} \leq C \sum_n b_n^2, \quad \forall (b_n) \in \ell^2(\mathbb{N}). \tag{5.3.83}$$

Lemma 5.3.3 can be showed from (5.3.73), Proposition 5.3.6 and Lemma 5.3.4.

Given Lemma 5.3.3, T is a Fredholm operator of index 0. Hence, it suffices to show that $\ker(T^*) = \{0\}$ to conclude that it is an isomorphism in H_1^{-1}. This is done in three steps:

1) Showing that there exists $\rho \in \mathbb{C}$ such that

$$A + \phi K + \lambda Id + \rho Id : H_1^1 \to H_1^{-1}, \quad \text{and} \quad A + \rho Id : H_1^1 \to H_1^{-1}$$
$$\text{are invertible.}$$

2) Showing that for such a complex number ρ, $\ker T^*$ is stable under $(A + \rho Id)^{-1}$. As $\ker T^*$ is finite-dimensional (recall that T is Fredholm), this means that either $\ker(T^*) = \{0\}$ or $(A + \rho Id)^{-1}$ has an eigenvector in $\ker T^*$, *i.e.* there exists $h \in \ker T^*$ and μ ($\mu \neq 0$ from the invertibility of $A + \rho Id$) such that
$$(A + \rho Id)^{-1} h = \mu h. \tag{5.3.84}$$
This implies that $h \in H_1^1$ and $(A + \rho Id)h = \frac{1}{\mu} h$ and therefore that h is an eigenvector of A. Since the eigenspaces of A have dimension 1, this means that $h = f_k$ for some $k \in \mathbb{N}^*$.
3) By adapting the $T\phi_1 = \phi_1$ condition, we show that $h = f_k$ is not in $\ker T^*$, hence $\ker T^* = \{0\}$.

Once this is done, we conclude that T is an isomorphism from H_1^{-1} to itself. This implies that K_n is uniformly bounded by below, i.e. the lower bound of (5.3.74) holds. From this we deduce that T is in fact an isomorphism from H_1^m into itself for any $m \in (-3/2, 3/2)$. And this conclude the proof of Proposition 5.3.5.

5.3.6.1 Well-posedness of the system

Once Proposition 5.3.5 and Corollary 2 are proven, it only remains to show the well-posedness of the system (5.3.8) with the feedback control we defined. This is what we do in this section. More precisely, what we show is the following

Lemma 5.3.5. *Let $k \in \{1,2\}$. Let $y_0 \in L^2_k$, $\phi \in H^{-1}_k$, and $K_k : H^{3/4}_k \to \mathbb{R}$ be bounded. The equation*

$$\begin{cases} \partial_t y - \Delta y = K_k(y)\phi_k, \\ y(0) = y_0, \end{cases} \tag{5.3.85}$$

has a unique solution satisfying the equation in $L^2_{loc}((0,+\infty); H^{-1}_k)$, and

$$y(t) \in C^0([0,+\infty); L^2_k) \cap L^2_{loc}((0,+\infty); H^1_k) \cap H^1_{loc}((0,+\infty); H^{-1}_k). \tag{5.3.86}$$

from which we then deduce

Proposition 5.3.7. *Let $k \in \{1,2\}$, $r \in (-1/2, 1/2)$, $y_0 \in H^r_k$, $\phi \in H^{-1/2-}_k$, and $K_k : H^{1/2+}_k \to \mathbb{R}$ be bounded. The equation (5.3.85) has a unique solution such that the equation is satisfied in $L^2_{loc}((0,+\infty); H^{r-1}_k)$, and*

$$y(t) \in C^0([0,+\infty); H^r_k) \cap L^2_{loc}((0,+\infty); H^{r+1}_k) \cap H^1_{loc}((0,+\infty); H^{r-1}_k). \tag{5.3.87}$$

Using the fact that $H^m = H^m_1 \bigoplus H^s_m$ for any $m \in \mathbb{R}$ we deduce directly the well-posedness needed for Theorem 5.3.3. Note also that we only need to prove the well-posedness for $s = 0$ and then use the isomorphism $D_s : H^s \to L^2$ defined by

$$D_s : n^{-s} f_k \to f_k. \tag{5.3.88}$$

Lemma 5.3.5 differs from the classical well-posedness of the heat equation because the operator K is non-local. Nevertheless, it can be shown with a classical fixed point in the norm

$$\|z\|_{C^0([0,T];L^2_1)} + \|z\|_{L^2((0,T);H^1_1)}, \tag{5.3.89}$$

and using the fact that $K_k \in \mathcal{L}(H^{3/4}_1, \mathbb{R})$ (rather than $\mathcal{L}(H^{3/4}_1, \mathbb{R})$) so $\|Kz\|_{L^2(0,T)}$ can be bounded by $\|z\|_{L^2((0,T);H^{3/4}_1)}$ rather than $\|z\|_{L^2((0,T);H^1_1)}$. Then we conclude using the following Lemma

Lemma 5.3.6. *The following estimate holds*

$$\|z\|_{L^{\frac{8}{3}}(0,T;H^{\frac{3}{4}}_1)} \leq \|z\|^{\frac{1}{4}}_{L^\infty(0,T;L^2_1)} \|z\|^{\frac{3}{4}}_{L^2(0,T;H^1_1)} \leq (\|z\|_{C^0([0,T];L^2_1)} + \|z\|_{L^2((0,T);H^1_1)}). \tag{5.3.90}$$

This enables to obtain a contraction mapping, hence the existence of a solution for a small time horizon. We extend it on a large time domain by showing that no blow-up happens.

Finally these proofs can be extended to the nonlinear viscous Burgers' equation, by noting first that

$$\langle y, \partial_x(y^2) \rangle = 0, \tag{5.3.91}$$

since we work on the torus. And by noting that,

$$\begin{aligned} \|\partial_x(y^2)\|^2_{L^2(0,T;H^{-1})} &\leq \|yy\|^2_{L^2(0,T;L^2)}, \\ &\leq \int_0^T \|y(t,\cdot)\|^2_{L^2} \|y(t,\cdot)\|^2_{L^\infty} dt, \\ &\leq C \int_0^T \|y(t,\cdot)\|^2_{L^2} \|y(t,\cdot)\|^2_{H^{1/2}} dt, \end{aligned} \tag{5.3.92}$$

which, from Gagliardo-Nirenberg interpolation, gives

$$\begin{aligned} &\leq C \int_0^T \|y(t,\cdot)\|^3_{L^2} \|y(t,\cdot)\|_{H^1} dt, \\ &\leq C T^{\frac{1}{2}} \|y\|^3_{C^0([0,T];L^2)} \|y\|_{L^2(0,T;H^1)}. \end{aligned} \tag{5.3.93}$$

The rest of the proof is very similar to the proof of the linear case.

Chapter 6

Compactness-duality method and water-wave equations

This Chapter is taken from [110], a collaboration with Ludovick Gagnon, Shengquan Xiang and Christophe Zhang.

6.1 Introduction

When dealing with the heat equation, our backstepping approach used the fact that the eigenvalues of the evolution operator $\mathcal{A}$ are growing quickly enough. Indeed, we used that $\lambda_n = n^2$ to ensure that

$$\sum_{n\in\mathbb{N}^*}\sum_{p\neq n}\left(\frac{1}{\lambda_p+\lambda-\lambda_n}\right)^2 < +\infty \tag{6.1.1}$$

This was used to show eventually that $(Tf_n)_{n\in\mathbb{N}}$ is quadratically close to an orthonormal basis and finally that it is a Riesz basis, using Lemma 5.2.2. Looking at (6.1.1), this is expected to work again as long as $\lambda_n \sim n^\alpha$ with $\alpha > 3/2$. However, when the eigenvalues λ_n are not growing fast enough with n, this inequality fails. Indeed, one can check for instance that when $\lambda_n = n^{3/2}$,

$$\sum_{n\in\mathbb{N}^*}\sum_{p\neq n}\left(\frac{1}{\lambda_p+\lambda-\lambda_n}\right)^2 = +\infty \tag{6.1.2}$$

This means that $(Tf_n)_{n\in\mathbb{N}}$ is not anymore quadratically close to the orthonormal basis easily derived from $(f_n)_{n\in\mathbb{N}^*}$ and this is an incitation to use another approach.

Coincidentally, the critical case $n = 3/2$ corresponds to a well-known system: the linearized capillarity-gravity water-wave equations which were studied and derived in [4, 5, 167, 168] and take the form

$$\begin{aligned} &\partial_t y = \mathcal{A}y + Bu(t), \quad \text{on } \mathbb{T}, \\ &\mathcal{A} = -i\left((g-\partial_x^2)|D_x|\tanh(l|D_x|)\right)^{1/2}, \end{aligned} \tag{6.1.3}$$

where g is a constant of gravity, l is the height of the water, $u = (u_1, u_2)$ is a two-dimensional control operator and $B = (B_1, B_2)$ are space dependent functions. Here they will be characterised by

$$B_1 = \sum_{n\in\mathbb{N}^*} a_n^1 f_n^1, \quad B_2 = \sum_{n\in\mathbb{N}} a_n^2 f_n^2, \tag{6.1.4}$$

where $(f_n^1))_{n\in\mathbb{N}^*}$ and $(f_n^2)_{n\in\mathbb{N}}$ are given in (5.3.14). As in Chapter 5, a two-dimensional control is the minimum needed for exact controllability and the system is exact controllable in L^2 if there exists c_1, $c_2 > 0$ such that

$$a_0 \neq 0 \ \text{ and } \ c_1 < |a_n^i| < c_2, \quad \text{for } i \in \{1,2\},\ n \in \mathbb{N}^*. \tag{6.1.5}$$

Knowing whether System (6.1.3) can be rapidly stabilized using a backstepping approach was an open question introduced in [56]. We introduce in the following an approach to solve this question.

6.2 Main result

In [110] we show that one can get rid of the quadratically close argument with a new method: the *compactness-duality method*. With this method, one can show the following result

Theorem 6.2.1. *Let $\alpha > 1$. Let $B \in (H^{-3/4})^2$ of the form (6.1.4) satisfying the controllability assumption (6.1.5). Let $h(s)$ a real valued-function satisfying*
- *$|n_1 - n_2| n_1^{\alpha-1} \lesssim |h(n_1) - h(n_2)|$ for any $(n_1, n_2) \in \mathbb{N}^*$.*
- *$s^\alpha \lesssim |h(s)| \lesssim s^\alpha$ for any $s \in [1, +\infty)$.*

Then, for any $\lambda > 0$, there exists a bounded linear operator $K \in \mathcal{L}(H^{3/4}; \mathbb{C}^2)$ and an operator T such that, for any $r \in (1/2 - \alpha, \alpha - 1/2)$, T is an isomorphism from $H^r(\mathbb{T})$ to itself and maps the system

$$\partial_t u = i\, h(|D_x|)u + BK(u),\ x \in \mathbb{T}, \tag{6.2.1}$$

to the system

$$\partial_t v = i\, h(|D_x|)v - \lambda v,\ x \in \mathbb{T}.$$

In particular the system (6.2.1) is exponentially stable in $H^r(\mathbb{T})$ with decay rate λ, for any $r \in (1/2 - \alpha, \alpha - 1/2)$.

The rapid stabilization of the water waves system (6.1.3) is a direct application of this theorem with $\alpha = 3/2$ and $h(s) = -\left((g - s^2)s \tanh(ls)\right)^{1/2}$.

Remark 6.2.1 (Regularity)**.** *The following points are worth noting.*
- *Similarly as for the heat equation, even if the regularity of B is fixed, the same feedback operator K works for a continuum of spaces $H^r(\mathbb{T})$.*
- *The bound on this continuum is sharp in the sense that for $r = \alpha - 1/2$ the system does not even generate a strongly continuous semigroup.*

6.3 Strategy of the proof

Definitions and notations We keep the notations of Chapter 5 (see in particular Section 5.3.1). We look for T of the form $T = T_1 \bigoplus T_2$ corresponding to the odd and even part of the transform. To simplify, we again only look at the odd part, i.e. T_1, B_1, K_1, f_n^1, and H_1^s and we drop the i for clarity. As in Chapter 5, the first step is to look for T satisfying the operator equalities

$$\begin{aligned} T\mathcal{A} + BK &= (\mathcal{A} - \lambda Id)T, \\ TB &= B, \end{aligned} \tag{6.3.1}$$

potentially in a weak sense to be defined. The eigenvalues of $\mathcal{A}$ associated to the orthonormal family of eigenvectors $(f_n)_{n\in\mathbb{N}}$ are denoted λ_n, and we observe that T is entirely defined by $(h_n)_{n\in\mathbb{N}} = (Tf_n)_{n\in\mathbb{N}}$ which necessarily satisfies

$$Tf_n = (-K(f_n)) \sum_{p\in\mathbb{N}^*} \frac{a_p f_p}{\lambda_n - \lambda_p + \lambda}, \tag{6.3.2}$$

just like in Chapter 5 (see (5.3.51)). Here again, we remark that once the feedback operator K is chosen, T is completely determined. Finally we define the operator S as follows

$$Sf_n = \sum_{p\in\mathbb{N}^*} \frac{f_p}{\lambda_n - \lambda_p + \lambda}, \tag{6.3.3}$$

The interest behind defining such an operator is to decouple the intrinsic properties of the transform T with the influence of the control K. The influence of the operator B is also removed in S, and deducing the

invertibility of T from the invertibility of S can later be done simply, thanks to the controllability condition (6.1.5). The strategy of the method is the following:

Step 1 Show that S is a Fredholm operator from $H^r \to H^r$ for any $r \in (1/2-\alpha, \alpha-1/2)$.

Step 2 Show that $(Sf_n)_{n\in\mathbb{N}^*}$ is a Riesz basis for L^2 using a duality argument and the fact that S is Fredholm.

Step 3 Further show that $(n^{-r}Sf_n)_{n\in\mathbb{N}^*}$ is a Riesz basis for H^r for any $r \in (1/2-\alpha, \alpha-1/2)$ by showing it is ω-independent using a duality argument between the density of $(n^{-r}q_n)_{n\in\mathbb{N}^*}$ in H^r and the ω-independence of $(n^r\overline{q_n})_{n\in\mathbb{N}^*}$ in H^{-r}.

Step 4 Provide an explicit candidate of $(K_n)_{n\in\mathbb{N}}$ which satisfies $TB = B$ in $H^{-\alpha/2}$ sense. Show that $(|K_n|)_{n\in\mathbb{N}}$ is bounded from above and that $a_nK_n = -(\lambda + k_n)$ for any $n \in \mathbb{N}^*$, where $(k_n n^\varepsilon)_{n\in\mathbb{N}^*} \in l^\infty$ for any $\varepsilon \in [0, \varepsilon_1(\alpha))$ where $\varepsilon_1(\alpha)$ is a constant that can be computed.

Step 5 Show that T is bounded from H^r in itself for $r \in (1/2-\alpha, \alpha-1/2)$ and the first operator equality of (6.3.1) holds in $\mathcal{L}(H^{\alpha/2}; H^{-\alpha/2})$.

Step 6 Show that T is a Fredholm operator from $H^{-\alpha/2}$ to $H^{-\alpha/2}$.

Step 7 Show that T is an isomorphism from $H^{-\alpha/2}$ to $H^{-\alpha/2}$ using a Fredholm argument and spectral arguments in $H^{-\alpha/2}$.

Step 8 Show that T is an isomorphism from L^2 to L^2 and in fact an isomorphism from H^r to itself for any $r \in (1/2-\alpha, \alpha-1/2)$.

Let us briefly discuss step 6 to 8. At first sight, it seems odd to prove the invertibility in $H^{-\alpha/2}$ (step 6-7) and not in L^2 for instance. The main motivation is to avoid working in the space $D(A+BK) := \{f \in L^2 : (A+BK)f \in L^2\}$ before proving the invertibility of T. Indeed, $D(A+BK)$ does not *a priori* have the nice properties we can expect from Sobolev spaces, such as the density of C^∞ functions. This comes from the lack of regularity of B. This implies we are not able to conclude that $f_n \in D(A+BK)$ for any $n \in \mathbb{N}^*$. For this reason, it is easier to first prove the invertibility in a weaker space, namely $H^{-\alpha/2}$, but with nice Sobolev spaces properties before deducing the invertibility in the required spaces. Once this is done, the invertibility of T in H^r will allow to construct an equivalent norm to H^r, which will allow to prove that $D(A+BK)$ is a Hilbert space. This last observation would have been hard to show without the invertibility of T. As a final comment, let us underline that, even though our setting is close to the linearized bilinear Schrödinger equation, one cannot decouple the real and imaginary part of the solution to deal directly with the space $D(A+BK)$ as was done in [63] for the Schrödinger equation.

6.4 Ideas of the proof

We focus on the first four steps for which we give the key elements. The full proof can be found in [110] where each step corresponds to a section for readability.

Step 1: the spirit of this step is simply to write S as $S = Id/\lambda + S_c$ where

$$S_c : n^{-r}f_n \to n^{-r} \sum_{p\in\mathbb{N}^*\setminus\{n\}} \frac{f_p}{\lambda_n - \lambda_p + \lambda}, \tag{6.4.1}$$

and to show that S_c is compact. This eventually amounts to showing the following estimate

Lemma 6.4.1. *For any $s < \alpha - 1$ we have*

$$\sum_{n\in\mathbb{N}^*\setminus\{p\}} \frac{n^s}{|\lambda_n - \lambda_p|} \lesssim p^{1-\alpha+s}\log(p) + p^{-\alpha},\ \forall p \in \mathbb{N}^*. \tag{6.4.2}$$

Step 2-3: the goal of these steps is to show that S is invertible in $H^r(\mathbb{T})$. Since S is a Fredholm operator in $H^r(\mathbb{T})$, it is enough to show that $\ker(S^*) = \{0\}$, where S^* is the adjoint of S. However, since $\dim(\text{coker}(S)) = \dim(\ker(S^*)) = \dim(\ker(S)) < +\infty$, it is enough to show that $\ker(S) = \{0\}$. This is shown to be equivalent to showing that $(n^{-r}Sf_n)_{n\in\mathbb{N}}$ is ω-independent in H^r. From that point, one has to separate the case $r = 0$ and $r \neq 0$. For $r = 0$ we show the following Lemma.

Lemma 6.4.2. *The families $(Sf_n)_{n\in\mathbb{N}^*}$ and $(\overline{Sf_n})_{n\in\mathbb{N}^*}$ satisfy the following:*
(i) $(Sf_n)_{n\in\mathbb{N}^*}$ is either ω-independent in L^2 or L^2-dense.
(ii) $(\overline{Sf_n})_{n\in\mathbb{N}^*}$ is either ω-independent in L^2 or L^2-dense.
(iii) $(Sf_n)_{n\in\mathbb{N}^*}$ is ω-independent in $L^2 \iff (\overline{Sf_n})_{n\in\mathbb{N}^*}$ is ω-independent in L^2.
(iv) $(Sf_n)_{n\in\mathbb{N}^*}$ is L^2-dense $\iff (\overline{Sf_n})_{n\in\mathbb{N}^*}$ is L^2-dense.
(v) $(Sf_n)_{n\in\mathbb{N}^*}$ is L^2-dense $\iff (\overline{Sf_n})_{n\in\mathbb{N}^*}$ is ω-independent in L^2.
(vi) $(\overline{Sf_n})_{n\in\mathbb{N}^*}$ is L^2-dense $\iff (Sf_n)_{n\in\mathbb{N}^*}$ is ω-independent in L^2.

In particular $(Sf_n)_{n\in\mathbb{N}^*}$ is both L^2 dense and ω-independent.

For $r \neq 0$ the situation is more delicate, and trying to show the result for one given r at the time would be a mistake. In fact, the key step is to look at r and $-r$ at the same time and to use the following duality between ω-independence in H^{-r} and density in H^r.

Lemma 6.4.3. *For $r > 0$, if $(n^{-r}Sf_n)_{n\in\mathbb{N}^*}$ in $H^r(\mathbb{T})$, then $(n^rSf_n)_{n\in\mathbb{N}^*}$ is ω-independent in $H^{-r}(\mathbb{T})$.*

Step 4: The spirit of this step is to solve the $TB = B$ condition of (6.3.1) in some weak sense and to see what regularity can be deduced on the K_n. The first thing to observe is that there exists a unique sequence $(K_n)_{n\in\mathbb{N}^*}$ such that for any $\varepsilon \in (0, 1/2)$ $TB = B$ holds in a weak sense in $H^{-1/2-\varepsilon}$ and [1] $(K_n a_n n^{-1/2-\varepsilon})_{n\in\mathbb{N}^*} \in l^2$. Then we can set $k_n = -(a_nK_n + \lambda)$ and the goal is to show a better regularity on $(k_n)_{n\in\mathbb{N}^*}$ and deduce that $(K_n)_{n\in\mathbb{N}^*}$ is l^∞ and that $k : n^{-r}f_n \to k_n\tau Sn^{-r}f_n$ is a compact operator from H^r to itself. This would mean that $-\lambda$ is the singular part of a_nK_n. Note that if $(k_n)_{n\in\mathbb{N}^*}$ is very regular, l^∞ is the best one can get for $(K_n)_{n\in\mathbb{N}^*}$ since it is the regularity of λ. To do so, the strategy is to replace K_n in the weak $TB = B$ by its expression with k_n, simplify and observe that the resulting equation has to hold in a more regular space. For $\alpha = 3/2$ this is enough to conclude.

For $\alpha \in (1, 3/2)$ the situation is more delicate. In this case one has to give a better asymptotic estimate on K_n and to decompose it as

$$-K_n = \lambda + e_n^{(1)} + e_n^{(2)} + ... + e_n^{(p)} + k_n^{(p)}, \tag{6.4.3}$$

where $p \in \mathbb{N}$. Then one can show by induction that $(e_n^{(p)})_{n\in\mathbb{N}^*} \in l^\infty$, that the regularity of $(k_n^{(p+1)})_{n\in\mathbb{N}}$ is strictly better than the regularity of $(k_n^{(p)})_{n\in\mathbb{N}}$, and that for a given $\alpha \in (1, 3/2)$ there exists a finite $p \in \mathbb{N}$ such that $k^{(p)} : n^{-r}f_n \to k_n^{(p)}\tau Sn^{-r}f_n$ is a compact operator [2].

6.5 Well-posedness of the system

It remains to show that the closed loop-system (6.1.3) with $u(t) = K(y)$ is well-posed in $H^r(\mathbb{T})$. This is less standard than for the heat equation. Let us assume to simplify that $\alpha = 3/2$, the case $\alpha \neq 3/2$ can be done similarly. The first thing to show is that $D(A + BK)$ is a Hilbert space. For this we show that $D(A + BK) \subset H^{1-\varepsilon}$ for any $\varepsilon > 0$, we extend the first operator equality of (6.3.1), and we show that

1. Recall that $a_n = \langle \phi, f_n \rangle$.
2. It can also be shown that $p \to +\infty$ when $\alpha \to 1$.

$D(A+BK) = T^{-1}(H^{3/2})$. From this we deduce that it is a Hilbert space, dense in L^2. Then one can show that T is actually an isomorphism not only on H^r for $r \in (-1,1)$ but also from $T^{-1}(H^{3/2})$ to $H^{3/2}$. We will not detail it here. The rest of the proof for the well-posedness in L^2 is more standard and relies on Lumer-Philips theorem. For the well-posedness and exponential stability in H^r with $r \in (-1,1)$ one needs to deal with two situations:

— (regular situation) when $r \in (-1,-1/2)$ in which case $D_r(A+BK) := \{f \in H^r : (A+BK)f \in H^r\}$ is simply $H^{r+3/2}$ and can contain regular functions like the f_n (recall that $T^{-1}(H^{r+3/2}) = H^{r+3/2}$ in this case since T is an isomorphism from H^s to itself for $s \in (-1,1)$),

— (singular situation) when $r \in [-1/2,1)$ where $D_r(A+BK) = T^{-1}(H^{r+3/2})$ and cannot contain regular functions like the f_n. This includes the case $r = 0$ and the proof of the well-posedness in L^2 can be adapted to this situation.

Chapter 7

Stabilization of a hyperbolic system with $\lambda_n \sim n$: the water tank system

In this section, we study a case where the system is hyperbolic and $\lambda_n \sim n$. In this case, none of the two previous approaches of Chapter 5–6 can apply. This section is taken from [64], a collaboration with Jean-Michel Coron, Shengquan Xiang and Christophe Zhang.

7.1 Formulation of the problem

We consider the homogeneous Saint-Venant equations in a water tank, subjected to an acceleration and given by

$$\begin{cases} \partial_t H + \partial_x(HV) = 0, \\ \partial_t V + V\partial_x V + g\partial_x H = -U(t). \end{cases} \tag{7.1.1}$$

where $U(t)$ is the acceleration applied to the water-tank. Given that we consider a water tank, the boundary conditions satisfy

$$V(t,0) = V(t,L) = 0. \tag{7.1.2}$$

This implies, in particular and together with (7.1.1), that $\int_0^L H(t,x)dx$ is constant with time. This represents the conservation of mass. Our goal is to study and stabilize the linearized equations around the steady-states (H^*, V^*) corresponding to a small acceleration $U(t) = \gamma > 0$. The motivation to consider $\gamma \neq 0$ is that these linearized equations are not controllable, hence not stabilizable, when $\gamma = 0$ (see [53, 54]), but they are when $\gamma > 0$. These steady-states are given by $V^* = 0$ and $H^* = H^\gamma(x)$ with

$$\begin{aligned} H^\gamma(x) =& 1 - \gamma\left(x - \frac{L}{2}\right), \ \forall x \in [0,L], \\ & \int_0^L H^\gamma(x)dx = L. \end{aligned} \tag{7.1.3}$$

Denoting $h = (H - H^*)$, $v = (V - V^*)$ and $u(t) = -(U(t) - \gamma)$, the linearized equations are

$$\partial_t \begin{pmatrix} h \\ v \end{pmatrix} + \begin{pmatrix} 0 & H^\gamma \\ 1 & 0 \end{pmatrix} \partial_x \begin{pmatrix} h \\ v \end{pmatrix} + \begin{pmatrix} 0 & -\gamma \\ 0 & 0 \end{pmatrix} \begin{pmatrix} h \\ v \end{pmatrix} = u(t) \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{7.1.4}$$

with the boundary conditions

$$v(t,0) = v(t,L) = 0. \tag{7.1.5}$$

And the conservation of mass in the tank becomes

$$\frac{d}{dt}\int_0^L h(t,x)dx = 0. \tag{7.1.6}$$

We will assume from now on that $\int_0^L h(0,x)dx = 0$, which means that the mass of the water in the initial state is the same as the mass of the water at the steady-state. The reason behind is clear: if this condition is not satisfied, there is no chance to show any convergence to the steady-state given that the mass of the water remains constant in the system. So in the following we will consider the controllability and stabilizability of the system only within states satisfying this condition.

Before stating our results, let us introduce the operator

$$\mathcal{L} = \begin{pmatrix} 0 & H^\gamma \\ 1 & 0 \end{pmatrix} \partial_x + \begin{pmatrix} 0 & -\gamma \\ 0 & 0 \end{pmatrix} Id, \tag{7.1.7}$$

associated to (7.1.4) and defined on the domain

$$D(\mathcal{L}) := \left\{ \begin{pmatrix} h \\ v \end{pmatrix} \in (H^1([0,L],\mathbb{C})^2 \text{ such that (7.1.5) holds} \right\}. \tag{7.1.8}$$

We can show that there is a family of eigenvectors of $\mathcal{L}$ that form a Riesz basis of $(L^2(0,L;\mathbb{C}))^2$. We denote this family by $(h_n^\gamma, v_n^\gamma)_{n\in\mathbb{Z}}$. We also denote by $\mathcal{D}_\gamma$ the space of finite linear combinations of $(h_n^\gamma, v_n^\gamma)_{n\in\mathbb{Z}}$ and $\mathcal{D}'_\gamma$ its dual. What we show is the following rapid stabilization result

Theorem 7.1.1. *For any $\mu > 0$, there exists $\gamma_0 > 0$ such that, for any $\gamma \in (0,\gamma_0)$, there exists a feedback law u which stabilizes the system (7.1.4)–(7.1.5) with decay rate (at least) μ.*

More precisely, let us define for $\nu > 0$ the feedback operator F_1^γ belonging to $\mathcal{D}'_\gamma$ and given by

$$\begin{aligned} \langle (h_n^\gamma, v_n^\gamma)^T, F_1^\gamma \rangle &= -\frac{\tanh(4\mu L)}{H^\gamma(0)} \frac{(h_n^\gamma)^2(0)}{\int_0^L \frac{L}{L_\gamma\sqrt{1-\gamma(x-\frac{L}{2})}} \exp\left(-\int_0^x \frac{3\gamma}{4(1-\gamma(x-\frac{L}{2}))} ds\right) v_n^\gamma(x)dx}, \ \forall\, n \in \mathbb{Z}^*, \\ \langle (h_0^\gamma, v_0^\gamma)^T, F_1^\gamma \rangle &= -2\frac{\tanh(4\mu L)}{H^\gamma(0)} \frac{(h_0^\gamma)^2(0)}{\nu}, \end{aligned} \tag{7.1.9}$$

and the control feedback u_2 defined by

$$u_2'(t) = \frac{\nu L}{L^\gamma} \langle (h_0^\gamma \ v_0^\gamma)^T, F_1^\gamma \rangle \left(u_2(t) + \left\langle \begin{pmatrix} h \\ v \end{pmatrix}(t,\cdot), F_1^\gamma \right\rangle \right), \tag{7.1.10}$$

with $L_\gamma := \frac{2}{\gamma}\left(1 - \sqrt{1-\gamma\frac{L}{2}}\right)$. There exists $\nu \neq 0$ such that the feedback law u defined by

$$u(t) := \left\langle \begin{pmatrix} h \\ v \end{pmatrix}(t,\cdot), F_1^\gamma \right\rangle + u_2(t), \tag{7.1.11}$$

stabilizes the system (7.1.4)–(7.1.5) exponentially in H^1 norm with decay rate μ.

Despite its complicated formulation, the method is constructive and the feedback law F is explicit [1]. One can remark also that this feedback has two parts: a proportional (given by the feedback operator F_1^γ) and an integral (given by the control u_2, defined as the solution of a first-order ODE). This comes from the fact that we will actually study a dynamic extension of the system. Before applying the backstepping approach, we transform the system and show that, using a change of variables, this system is equivalent to

$$\partial_t \zeta + \Lambda \partial_x \zeta + \delta J \zeta = u_\gamma(t) \exp\left(\int_0^x \delta(y)dy\right) \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{7.1.12}$$

$$\begin{cases} \zeta_1(t,0) = -\zeta_2(t,0), \\ \zeta_2(t,L) = -\zeta_1(t,L), \end{cases} \tag{7.1.13}$$

1. Nevertheless, this result is not completely quantitative if we lack information on the Riesz basis.

with

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, J = \begin{pmatrix} 0 & \frac{1}{3} \\ -\frac{1}{3} & 0 \end{pmatrix},$$
$$\delta(x) = -\frac{3}{4}\gamma\Big(1 + \frac{1}{2}\gamma\left(\frac{L}{2} + x\right) + O(\gamma^2)\Big), \tag{7.1.14}$$

and the conservation of mass becoming

$$\int_0^L \left(\sqrt{1 + \frac{\gamma L}{2}} - \frac{\gamma}{2}\frac{L_\gamma}{L}x\right)^{1/2} (\zeta_1(x) - \zeta_2(x))\frac{L}{L_\gamma}dx = 0. \tag{7.1.15}$$

In fact this system has a drawback: defining

$$\mathcal{I} = \exp\left(\int_0^x \delta(y)dy\right)\begin{pmatrix}1\\1\end{pmatrix}, \tag{7.1.16}$$

and setting $f_0 = (f_{0,1}, f_{0,2})^T$ as a non-zero solution of

$$\begin{aligned} &\Lambda\partial_x f_0 + \delta J f_0 = 0, \\ &f_{0,1}(0) = -f_{0,2}(0), \end{aligned} \tag{7.1.17}$$

we can check that $f_{0,1}(x) = -f_{0,2}(x)$ for any $x \in [0, L]$. This implies that

$$\langle \mathcal{I}, f_0 \rangle = 0. \tag{7.1.18}$$

This means that the control has no effect on the direction given by the vector f_0. This is linked to the conservation of mass and corresponds to the fact that the control cannot add or remove mass. To overcome this issue we introduce the following virtual control operator

$$\mathcal{I}_\nu = \exp\left(\int_0^x \delta(y)dy\right)\begin{pmatrix}1\\1\end{pmatrix} + \nu f_0, \tag{7.1.19}$$

where f_0 is the solution of[2] (7.1.17) with norm 1. And we introduce the corresponding dynamic extension of our system

$$\begin{cases} \partial_t Z + \Lambda\partial_x Z + \delta(x)JZ = u_\gamma^\nu(t)\mathcal{I}_\nu, \\ Z_1(t,0) = -Z_2(t,0), \quad \forall t \geq 0, \\ Z_1(t,L) = -Z_2(t,L), \quad \forall t \geq 0. \end{cases} \tag{7.1.20}$$

We will now try to find u_γ^ν in the form of a feedback law $u_\gamma^\nu = \langle Z(t,\cdot), F\rangle$ (where the sense and the spaces involved in these brackets have to be defined). This will allow us to recover a feedback law for the original system, of the form (7.1.11). Just like for the heat equation and the water-wave equations, we can write this system in a more abstract way:

$$\partial_t Z = -\mathcal{A}Z + \langle Z(t,\cdot), F\rangle\mathcal{I}_\nu, \quad Z(t,\cdot) \in D(\mathcal{A}), \tag{7.1.21}$$

where $\mathcal{A} = \Lambda\partial_x + \delta(x)J$ and $D(\mathcal{A})$ is the domain of $\mathcal{A}$ defined by

$$D(\mathcal{A}) := \left\{(f_1, f_2) \in (H^1)^2, \quad f_1(0) + f_2(0) = 0, \quad f_1(L) + f_2(L) = 0\right\}. \tag{7.1.22}$$

Let $\mu > 0$, the target system will be

$$\partial_t z = -\tilde{\mathcal{A}}z, \quad z(t,\cdot) \in D(\tilde{\mathcal{A}}), \tag{7.1.23}$$

where $\tilde{\mathcal{A}}$ is still $\Lambda\partial_x + \delta(x)J$ but is defined on the domain

$$D(\widetilde{\mathcal{A}}) := \left\{(f_1, f_2) \in (H^1)^2, \quad f_1(0) + e^{-2\mu L}f_2(0) = 0, \ f_1(L) + f_2(L) = 0\right\}. \tag{7.1.24}$$

The boundary conditions induced by this new domain provide an exponential dissipation and one can show that this system is exponentially stable at rate as close to μ as desired:

2. f_0 is in fact the normal eigenfunction of the system associated with the eigenvalue 0, as we will see later on.

Proposition 7.1.2. *For any $\lambda_0 \in (0, \mu)$, there exists $\gamma_s(\lambda_0) > 0$ such that for any $\gamma \in (0, \gamma_s)$ and $\lambda \in [0, \lambda_0]$, the target system (7.1.23)–(7.1.24) is exponentially stable with decay rate λ (for the H^p norm, for any $p \in \mathbb{N}$). Moreover γ_s can be chosen continuous and decreasing with respect to λ_0.*

This can be shown, for instance, using basic quadratic Lyapunov functions.

7.2 Setting-up the backstepping: some functional properties and Riesz basis

Let us now start the backstepping approach. We aim to map the original system to target system, in other words, we want to find a transform T and a linear feedback operator F such that, formally,

$$T(-\mathcal{A} + \mathcal{I}_\nu F) = -\tilde{\mathcal{A}} T \tag{7.2.1}$$

$$T\mathcal{I}_\nu = \mathcal{I}_\nu, \tag{7.2.2}$$

which is equivalent with the current notations of (5.3.31)–(5.3.32) in Chapter 5 or (6.3.1) in Chapter 6. Looking back at the general method in introduction of Section 5.2.1, the backstepping method starts by considering an orthonormal basis of the considered functional space and then applying the operator equality (7.2.1) on this orthonormal basis to find a candidate transform T. One of the key point here will be to use an orthonormal basis $(f_n)_{n\in\mathbb{Z}}$ of $L^2(0,1;\mathbb{C}^2)$ whose elements are eigenvectors of $\mathcal{A}$ (thus belonging to $D(\mathcal{A})$) rather than a fourier basis $(e^{inx})_{n\in\mathbb{Z}}$ (for the heat equation and the water-wave equations such a question did not arise since $(e^{inx})_{n\in\mathbb{Z}}$ is also a basis of eigenvectors for the operator). Such a basis exists (see [216] for instance, and observe that $\mathcal{A}$ is skew-adjoint) and we list here some of its properties.

Proposition 7.2.1. *Denoting $(f_n)_{n\in\mathbb{Z}}$ an orthonormal basis of eigenvector of $\mathcal{A}$ and $(\mu_n)_{n\in\mathbb{Z}}$ the associated eigenvalues, one has*

(i) The eigenvalues are purely imaginary and

$$\mu_n = \frac{i\pi n}{L} + O(1), \quad \forall n \in \mathbb{Z}, \tag{7.2.3}$$

(ii)

$$\mu_{-n} = \overline{\mu_n} = -\mu_n, \quad \forall n \in \mathbb{Z}. \tag{7.2.4}$$

In particular, $\mu_0 = 0$.

(iii)

$$f_{-n} = (f_{-n,1}, f_{-n,2}) = \overline{f_n} = (-f_{n,2}(\cdot), -f_{n,1}(\cdot)), \quad \forall n \in \mathbb{Z}. \tag{7.2.5}$$

In particular, $f_{n,1}(0), f_{n,1}(L) \in \mathbb{R}$, and

$$f_{0,1}(x) + f_{0,2}(x) = 0, \forall x \in [0, L]. \tag{7.2.6}$$

Still looking at the general method, after applying the operator equality to f_n, we will want to project it again on a basis of the considered Hilbert space (see for instance (5.3.49) for the heat equation). Since f_n are not eigenvectors of $\tilde{\mathcal{A}}$ because they do not belong to $D(\tilde{\mathcal{A}})$, it would be interesting to see whether a basis of eigenvectors of $\tilde{\mathcal{A}}$ exists. The answer is yes, but the situation is slightly more complicated: due to the lack of symmetry in the boundary conditions induced by $D(\tilde{\mathcal{A}})$ (see the definition of $D(\tilde{\mathcal{A}})$ given by (7.1.24)), we can only say that the eigenvectors $\tilde{f}_n$ of $\tilde{\mathcal{A}}$ form a Riesz basis of $L^2((0,1);\mathbb{C}^2)$, and this basis is not necessarily orthonormal. However, there still exists a biorthogonal family $(\tilde{\phi}_n)_{n\in\mathbb{Z}}$ such that

$$f = \sum_{n\in\mathbb{Z}} \langle f, \tilde{\phi}_n \rangle \tilde{f}_n, \quad \text{for any } f \in L^2((0,1),\mathbb{C}^2), \tag{7.2.7}$$

and such that this biorthogonal family $(\tilde{\phi}_n)_{n\in\mathbb{Z}}$ is a family of eigenvectors of the adjoint operator $\tilde{\mathcal{A}}^*$. Also, still using [216], we have again the following property

$$\tilde{\mu}_n = \mu + \frac{i\pi n}{L} + O(1), \tag{7.2.8}$$

where $\tilde{\mu}_n$ denotes the eigenvalue of $\tilde{\mathcal{A}}$ associated to $\tilde{f}_n$. In [64] we show a proposition analogous to Proposition 7.2.1 for this eigensystem.

Proposition 7.2.2. *$(f_n, \phi_n, \mu_n)_{n\in\mathbb{Z}}$ satisfies the following properties*

(i)

$$\widetilde{\mu}_{-n} = \overline{\widetilde{\mu}_n}, \quad \forall n \in \mathbb{Z}. \tag{7.2.9}$$

(ii)

$$\begin{aligned} \widetilde{f}_{-n} &= \overline{\widetilde{f}_n}, \\ \widetilde{\phi}_{-n} &= \overline{\widetilde{\phi}_n}, \quad \forall n \in \mathbb{Z}. \end{aligned} \tag{7.2.10}$$

We conclude this section by defining $D(\mathcal{A}^s)$ as follows

$$D(\mathcal{A}^s) := \left\{ \alpha \in (L^2)^2, \quad \sum_{n\in\mathbb{Z}} (1+|\mu_n|^{2s})|\langle \alpha, f_n\rangle|^2 < \infty \right\} \subset (H^s)^2, \quad s \in [0, +\infty), \tag{7.2.11}$$

endowed with the following norm

$$\|\alpha\|^2_{D(\mathcal{A}^s)} = \sum_{n\in\mathbb{Z}} (1+|\mu_n|^{2s})|\langle \alpha, f_n\rangle|^2, \quad \forall \alpha \in D(\mathcal{A}^s). \tag{7.2.12}$$

Similarly we have

$$D(\widetilde{\mathcal{A}}^s) := \left\{ \alpha \in (L^2)^2, \quad \sum (1+|\widetilde{\mu}_n|^{2s})|\langle \alpha, \widetilde{\phi}_n\rangle|^2 < \infty \right\}, \quad s \in [0, +\infty), \tag{7.2.13}$$

endowed with the norm

$$\|\alpha\|^2_{D(\widetilde{\mathcal{A}}^s)} = \sum (1+|\widetilde{\mu}_n|^{2s})|\langle \alpha, \widetilde{\phi}_n\rangle|^2, \quad \forall \alpha \in D(\widetilde{\mathcal{A}}^s). \tag{7.2.14}$$

7.3 Controllability

Before trying to stabilize the system, it is natural to wonder whether it is controllable. There is another motivation to look at controllability: as we already mentioned, generalized backstepping is really about using the controllability to construct an explicit feedback law. Thus, we expect the controllability estimates to play an essential role at some point. In the example of the heat equation presented in Chapter 5 the condition on the control operator we required for the controllability was $a_0^2 \neq 0$ with $(a_n^i)_{n\in\mathbb{N}}$ defined by (5.3.45) and for any $i \in \{1, 2\}$ and $n \in \mathbb{N}^*$,

$$cn^{-s} < |a_n^i| < Cn^{-s}, \tag{7.3.1}$$

which was used to show the existence of a transform T and a feedback K later on. In Chapter 6 the condition (6.1.5) was also used to derive the feedback operator K later on. In [64], we show the controllability of the original (virtual) system (7.1.20), which is equivalent to showing the controllability of the system excluding in the direction f_0 describing the change of mass. We also show the controllability of the target system. The controllability is shown using a moment method as well as an expansion in γ of the eigenvectors, Kato's method for asymptotic calculation (see [153, Chapter 2]), and the derivation of relatively long asymptotic estimates. We will not detail here the long analysis leading to the controllability of both systems but, rather, we will highlight the conditions and estimates it brings on the control operator. What can be highlighted, among others, are the following estimates and expressions that we obtain:

— From the controllability of the original system, we get

$$m|\mu_n^{-1}| \leq |\langle \mathcal{I}, f_n\rangle| \leq M|\mu_n^{-1}|, \quad \forall n \in \mathbb{Z}^*, \tag{7.3.2}$$

where m and M are positive constants independent of n. (7.3.2) describes the controllability of system (7.1.12)–(7.1.13) without the direction f_0 and implies that

$$m(1+|\mu_n|)^{-1} \leq |\langle \mathcal{I}_\nu, f_n\rangle| \leq M(1+|\mu_n|)^{-1}, \quad \forall n \in \mathbb{Z}. \tag{7.3.3}$$

We have in fact a more precise expression

$$\langle \mathcal{I}_\nu, f_n\rangle = \frac{2\left(f_{n,1}(0) - e^{\int_0^L \delta} f_{n,1}(L)\right)}{\mu_n} + \frac{1}{\mu_n}\langle \delta(x) J\mathcal{I}, f_n\rangle, \quad \forall n \in \mathbb{Z}^*, \tag{7.3.4}$$

and we also obtain

$$m \leq |f_{n,1}(0)| \leq M, \quad \forall n \in \mathbb{Z}. \tag{7.3.5}$$

— From the controllability of the target system, we get

$$m(1+|\tilde{\mu}_n|)^{-1} \leq |\langle \mathcal{I}_\nu, \tilde{\phi}_n\rangle| \leq M(1+|\tilde{\mu}_n|)^{-1}, \quad \forall n \in \mathbb{Z}, \tag{7.3.6}$$

and

$$m \leq \left|\widetilde{\phi}_{n,1}(0)\right| \leq M, \quad \forall n \in \mathbb{Z}. \tag{7.3.7}$$

7.4 Finding a candidate T

We are now ready to start the backstepping method. As in Chapter 5–6, we start by finding a candidate transform T by assuming that the operator equalities (7.2.1)–(7.2.2) hold and applying (7.2.1) on an orthonormal basis of eigenfunctions to see what condition it implies on T. We apply it to our basis $(f_n)_{n\in\mathbb{Z}}$, the family of eigenfunctions of $\mathcal{A}$, and we obtain, assuming (7.2.2) holds,

$$-T\mathcal{A}f_n + \mathcal{I}_\nu\langle f_n, F\rangle = -\tilde{\mathcal{A}}Tf_n. \tag{7.4.1}$$

As f_n is an eigenvector of $\mathcal{A}$ this becomes

$$-\mu_n(Tf_n) + \mathcal{I}_\nu K_n = -\tilde{\mathcal{A}}(Tf_n), \tag{7.4.2}$$

where $K_n = \langle f_n, F\rangle$ are the coefficients of the feedback control on the basis $(f_n)_{n\in\mathbb{Z}}$. We have now an equation on $(Tf_n)_{n\in\mathbb{Z}}$. However, this is still a differential equation because of the operator $\tilde{\mathcal{A}}$ in (7.4.2) and we would like to simplify this again. Therefore, we project this equation again on an orthonormal basis and, this time, we choose the basis $(\tilde{\phi}_n)_{n\in\mathbb{Z}}$ of eigenvectors of $\tilde{\mathcal{A}}^*$, the adjoint of $\tilde{\mathcal{A}}$. We are motivated by the fact that we would like to get rid of the differential operator $\tilde{\mathcal{A}}$ in (7.4.2) but this operator is applied on the left so the projection will naturally make its adjoint $\tilde{\mathcal{A}}^*$ appear (note that this question did not appear in Chapter 5 and 7 since the operators where self or anti adjoint). We have

$$\begin{aligned} -\mu_n\langle (Tf_n), \tilde{\phi}_p\rangle + \langle \mathcal{I}_\nu, \tilde{\phi}_p\rangle K_n &= -\langle \tilde{\mathcal{A}}(Tf_n), \tilde{\phi}_p\rangle, \\ &= -\langle (Tf_n), \tilde{\mathcal{A}}^*\tilde{\phi}_p\rangle, \\ &= -\langle (Tf_n), \overline{\tilde{\mu}_p}\tilde{\phi}_p\rangle, \\ &= -\tilde{\mu}_p\langle (Tf_n), \tilde{\phi}_p\rangle, \end{aligned} \tag{7.4.3}$$

Hence we obtain the following characterisation of $(Tf_n)_{n\in\mathbb{Z}}$

$$\langle (Tf_n), \tilde{\phi}_p\rangle = \frac{\langle \mathcal{I}_\nu, \tilde{\phi}_p\rangle K_n}{\mu_n - \tilde{\mu}_p}, \quad \forall\, n \in \mathbb{Z},\ \forall\, p \in \mathbb{Z}. \tag{7.4.4}$$

Observe that, as in the previous chapters, when the K_n are fixed, (7.4.4) defines entirely T as $(\tilde{\phi}_p)_{p\in\mathbb{Z}}$ is a biorthogonal family associated to a basis of L^2 and f_n is an orthonormal basis of L^2. One can already observe that this implies the following condition on K

$$K_n \neq 0, \tag{7.4.5}$$

the condition $\langle \mathcal{I}_\nu, \tilde{\phi}_p\rangle \neq 0$ resulting from the controllability estimate (7.3.6) of the target system.

The next step is to show that the candidate transform T is an isomorphism provided some condition on F. Looking at Theorem 7.1.1 we would like T to be an isomorphism for solutions with H^1 regularity, and to map the original system to the target system. This means that we would like T to be an isomorphism from $D(\mathcal{A})$ to $D(\tilde{\mathcal{A}})$. Let $\alpha \in D(\mathcal{A})$, one has

$$\alpha = \sum_{n\in\mathbb{Z}} \langle \alpha, f_n\rangle f_n, \tag{7.4.6}$$

and

$$T\alpha = \sum_{n\in\mathbb{Z}} \langle \alpha, f_n\rangle (Tf_n). \tag{7.4.7}$$

Since $\alpha \in D(\mathcal{A})$, $(1+|n|)\langle \alpha, f_n\rangle \in \ell^2$. So, looking at the expression (7.4.7), we want $(1+|n|)^{-1}(Tf_n)$ to be a Riesz basis of $D(\tilde{\mathcal{A}})$ in order to conclude that T is an isomorphism from $D(\mathcal{A})$ to $D(\tilde{\mathcal{A}})$. However, because $|\lambda_n| \sim |n|$ we cannot hope to use a quadratically close property between $(Tf_n)_{n\in\mathbb{Z}}$ and an orthonormal basis derived from f_n or $\tilde{f}_n$ together with Lemma 5.2.2. Instead, we will show directly that $(Tf_n)_{n\in\mathbb{Z}}$ is a Riesz basis by checking the definition given by Definition 5.3.1 (5)' and that we recall here

Definition 7.4.1. A family $(\xi_n)_{n\in\mathcal{Z}}$ is a Riesz basis of a Hilbert space X if and only if it is dense in X and there exists $C_1, C_2 > 0$ such that for any $(a_n)_{n\in\mathbb{Z}} \in \ell^2$,

$$C_1 \sum_{k\in\mathbb{Z}} |a_k|^2 \leq \| \sum_{k\in\mathbb{Z}} a_k \xi_k \|_X^2 \leq C_2 \sum_{k\in\mathbb{Z}} |a_k|^2. \tag{7.4.8}$$

To show that $(Tf_n)_{n\in\mathbb{Z}}$ satisfies the assumptions of this definition, we introduce an auxiliary family k_n defined by

$$\langle k_n, \tilde{\phi}_p\rangle := \frac{1}{\tilde{\mu}_p - \mu_n}. \tag{7.4.9}$$

One can observe that k_n does not depend on the feedback control F and the control operator $\mathcal{I}_\nu$. This is the analogous of q_n and Sf_n in Chapter 5–6. The key lemma is the following

Lemma 7.4.1. *The family $(k_n)_{n\in\mathbb{Z}}$ is a Riesz basis of $L^2((0,1);\mathbb{C}^2)$.*

To prove this, note that, because f_n are eigenfunctions of $\mathcal{A}$

$$\mu_n \langle f_n, \widetilde{\phi}_p\rangle = \left\langle \mathcal{A} f_n, \widetilde{\phi}_p \right\rangle. \tag{7.4.10}$$

Here we can use an integration by parts (which amounts to taking the adjoint of $\mathcal{A}$), but as $\tilde{\phi}_p$ does not belong to $D(\mathcal{A}^*)$, there will be non-zero boundary terms:

$$\left\langle \mathcal{A} f_n, \widetilde{\phi}_p \right\rangle = \left\langle f_n, \mathcal{A}^* \widetilde{\phi}_p \right\rangle + f_{n,2}(0)\overline{\widetilde{\phi}_{p,2}(0)} - f_{n,1}(0)\overline{\widetilde{\phi}_{p,1}(0)}. \tag{7.4.11}$$

Using now that $\mathcal{A}$ and $\tilde{\mathcal{A}}$ are the same operators considered on different domain (hence $\mathcal{A}^*$ and $\tilde{\mathcal{A}}^*$ are as well), and that $\tilde{\phi}_p$ are eigenfunctions of $\tilde{\mathcal{A}}^*$

$$\begin{aligned} \left\langle \mathcal{A} f_n, \widetilde{\phi}_p \right\rangle &= \left\langle f_n, \widetilde{\mathcal{A}}^* \widetilde{\phi}_p \right\rangle + f_{n,2}(0)\overline{\widetilde{\phi}_{p,2}(0)} - f_{n,1}(0)\overline{\widetilde{\phi}_{p,1}(0)} \\ &= \widetilde{\mu_p} \langle f_n, \widetilde{\phi}_p\rangle - f_{n,1}(0)\overline{\widetilde{\phi}_{p,1}(0)} \left(1 - e^{-2\mu L}\right). \end{aligned} \tag{7.4.12}$$

From (7.4.10)–(7.4.12), the fact that f_n are eigenvectors of $\mathcal{A}$, and the fact that $(\widetilde{\phi}_n)_{n\in\mathbb{Z}}$ is the biorthonormal family associated to $(\widetilde{f}_n)_{n\in\mathbb{Z}}$, we deduce that

$$f_n = \sum_{p\in\mathbb{Z}} \langle f_n, \widetilde{\phi}_p\rangle \widetilde{f}_p = \sum_{p\in\mathbb{Z}} \frac{f_{n,1}(0)\overline{\widetilde{\phi}_{p,1}(0)}\left(1 - e^{-2\mu L}\right)}{\widetilde{\mu_p} - \mu_n} \widetilde{f}_p \quad \forall n \in \mathbb{Z}. \tag{7.4.13}$$

Introducing the operator τ defined by

$$\tau \widetilde{f}_p := \overline{\widetilde{\phi}_{p,1}(0)}(1 - e^{-2\mu L})\widetilde{f}_p, \quad \forall p \in \mathbb{Z}, \tag{7.4.14}$$

we have finally

$$f_n = f_{n,1}(0)\tau k_n, \quad \forall n \in \mathbb{Z}. \tag{7.4.15}$$

Because $\mu > 0$, and thanks to the controllability estimate (7.3.7) on the target system, τ is an isomorphism. Thanks to the controllability estimate (7.3.5), $f_{n,1}(0)$ is uniformly bounded by above and below, which means that $(\tau^{-1} f_n)_{n\in\mathbb{Z}}$ is a Riesz basis of $L^2((0,1);\mathbb{C}^2)$ and so is $\left(\tau^{-1} f_n / f_{n,1}(0)\right)_{n\in\mathbb{Z}} = (k_n)_{n\in\mathbb{Z}}$. Note that this relies a lot on the fact that $\mathcal{A}$ and $\tilde{\mathcal{A}}$ are the same operators but defined on different domains with one boundary condition that differs. Indeed, this is what allows f_n to be expressed as the image of k_n by a simple isomorphism. This is the reason why we chose such a target system (7.1.23)–(7.1.24) with a boundary damping instead of the usual target system, where one adds an internal damping to the original system.

From the Riesz basis property of $(k_n)_{n\in\mathbb{Z}}$ we can show the following proposition

Proposition 7.4.1. *The family*

$$\left(\frac{1}{K_n}(Tf_n)\right)_{n\in\mathbb{Z}} \tag{7.4.16}$$

is a Riesz basis of $D(\widetilde{\mathcal{A}})$.

Showing this can be done by checking directly Definition 7.4.1, and noting that, by density, the property (7.4.8) only needs to be checked for any $(a_n)_{n\in\mathbb{Z}}$ with finite support. Both assumptions of Definition 7.4.1 (density and estimate (7.4.8)) follow from the fact that they hold for $(k_n)_{n\in\mathbb{Z}}$ as it is a Riesz basis of $L^2((0,1);\mathbb{C}^2)$, together with the controllability estimate (7.3.6) of the target system which imposes that there exists $m, M > 0$ such that

$$m \leq (1 + |\widetilde{\mu}_p|^2) \left|\left\langle \mathcal{I}_\nu, \widetilde{\phi}_p \right\rangle\right|^2 \leq M. \tag{7.4.17}$$

From Proposition 7.4.1 we deduce directly that, if there exists $c, C > 0$ such that

$$c(1 + |n|) \leq |K_n| \leq C(1 + |n|), \quad \forall n \in \mathbb{Z}, \tag{7.4.18}$$

then $(1 + |n|)^{-1}(Tf_n)$ is a Riesz basis of $D(\tilde{\mathcal{A}})$, which is what we want.

Hence, provided that condition (7.4.18) on the feedback control F holds, the transform T is an isomorphism from $D(\mathcal{A})$ to $D(\widetilde{\mathcal{A}})$. Besides, we can check that if F is real-valued, then T maps real-valued functions to real-valued functions. The condition (7.4.18) can be interpreted as the growth needed on the feedback law so that (Tf_n) is not too regular (in which case T would not map the whole $D(\tilde{\mathcal{A}})$ from $D(\mathcal{A})$) but still regular enough (in which case the image of $D(\mathcal{A})$ by T would not be in $D(\tilde{\mathcal{A}})$).

7.5 Applying the transform T

We now have a good candidate T for the backstepping transform, and a condition on the feedback F so that it is an isomorphism. The next step is to show that T indeed satisfies the operator equalities (7.2.1)–(7.2.2), possibility in a weak sense, and provided some addtionnal conditions on F (which have to be compatible with the condition (7.4.18) we already have).

One can see right away that $T\mathcal{I}_\nu = \mathcal{I}_\nu$ cannot hold in a strong sense, simply because $\mathcal{I}_\nu$ is a (constant) function that do not belong to $D(\mathcal{A})$. Thus what we will want is a weak $T\mathcal{I}_\nu = \mathcal{I}_\nu$ equality defined as follows

$$\langle T\mathcal{I}_\nu^N, \tilde{\phi}_m \rangle \xrightarrow[N\to+\infty]{} \langle \mathcal{I}_\nu, \tilde{\phi}_m \rangle, \quad \forall m \in \mathbb{Z}, \tag{7.5.1}$$

where $\mathcal{I}_\nu^N$ is the orthogonal projection of $\mathcal{I}_\nu$ on the $2N+1$ dimensional space $\mathrm{Span}\{f_{-N}, ..., f_N\}$ given by

$$\mathcal{I}_\nu^N = \sum_{n=-N}^{N} \langle \mathcal{I}_\nu, f_n \rangle f_n. \tag{7.5.2}$$

Note that $\mathcal{I}_\nu^N$ belongs to $D(\mathcal{A})$, given that it is a finite sum of elements of $D(\mathcal{A})$, while $\mathcal{I}_\nu^N \to \mathcal{I}_\nu$ in $L^2((0,1);\mathbb{C}^2)$. As $(\tilde{\phi}_m)_{m\in\mathbb{Z}}$ also form a basis of $L^2((0,1);\mathbb{C}^2)$, the condition (7.5.1) is indeed a weak form of the formal $T\mathcal{I}_\nu = \mathcal{I}_\nu$.

Concerning the operator equality (7.2.1), it is clear that it cannot be applied *a priori* to any function of $L^2((0,1);\mathbb{C}^2)$ or even $D(\mathcal{A})$, since T is only properly defined on $D(\mathcal{A})$. Indeed for a function $\alpha \in D(\mathcal{A})$, $\mathcal{A}\alpha$ belongs to L^2 hence $T\mathcal{A}\alpha$ is not properly defined. To find the right space in which considering the operator equality, it is actually easier to use its original form before simplifying with $T\mathcal{I}_\nu = \mathcal{I}_\nu$ and look directly at

$$T(-\mathcal{A} + \mathcal{I}_\nu F) = -\tilde{\mathcal{A}}T \tag{7.5.3}$$

This suggest to look at the operator equality applied to functions α belonging to a domain D_F defined as

$$D_F = \{\alpha \in D(\mathcal{A}), \quad -\mathcal{A}\alpha + \langle \alpha, F\rangle \mathcal{I}_\nu \in D(\mathcal{A})\}\,. \tag{7.5.4}$$

This way, both sides of (7.5.3) applied to α are well defined since $T\alpha \in D(\tilde{\mathcal{A}})$ and $(-\mathcal{A} + \mathcal{I}_\nu F)\alpha \in D(\mathcal{A})$. This space D_F might seem peculiar, but we will see, when looking at the well-posedness [3], that it is intrinsically the right space to consider (see Lemma 7.6.1).

Let $\alpha \in D_F$ and $\alpha^{(N)}$ its truncation on $\mathrm{Span}\{f_{-N}, ...f_N\}$ defined as in (7.5.2). We can observe that

$$\begin{aligned}\langle T(-\mathcal{A}\alpha^{(N)} + \langle \alpha, F\rangle \mathcal{I}_\nu^{(N)}), \widetilde{\phi}_m\rangle &= \langle -\mathcal{A}\alpha^{(N)} + \langle \alpha, F\rangle \mathcal{I}_\nu^{(N)}, T^*\widetilde{\phi}_m\rangle_{D(\mathcal{A})\times D(\mathcal{A})'} \\ &\xrightarrow[N\to\infty]{} \langle -\mathcal{A}\alpha + \langle \alpha, F\rangle \mathcal{I}_\nu, T^*\widetilde{\phi}_m\rangle_{D(\mathcal{A})\times D(\mathcal{A})'} \\ = \langle T(-\mathcal{A}\alpha + \langle \alpha, F\rangle \mathcal{I}_\nu), \widetilde{\phi}_m\rangle,\end{aligned} \tag{7.5.5}$$

and $-\tilde{\mathcal{A}}T\alpha^N \to -\tilde{\mathcal{A}}T\alpha$ in $L^2((0,1),\mathbb{C}^2)$. So showing (7.5.3) on D_F reduces to showing that

$$\langle T(-\mathcal{A}\alpha^{(N)} + \langle \alpha, F\rangle \mathcal{I}_\nu^{(N)}) + \tilde{\mathcal{A}}T\alpha^{(N)}, \widetilde{\phi}_m\rangle \to 0, \quad \forall m \in \mathbb{Z},\ \forall \alpha \in D_F. \tag{7.5.6}$$

As $T\mathcal{A}\alpha^{(N)} = \sum_{n=-N}^{N} \mu_n \langle \alpha, f_n\rangle (Tf_n)$, observing from the definition of (Tf_n) (see (7.4.4)) that

$$\mu_n(Tf_n) = \tilde{\mathcal{A}}(Tf_n) + K_n\mathcal{I}_\nu, \tag{7.5.7}$$

and recalling that $K_n = \langle f_n, F\rangle$, showing (7.5.6) reduces to showing that

$$\langle \langle \alpha, F\rangle T\mathcal{I}_\nu^{(N)} - \langle \alpha^{(N)}, F\rangle \mathcal{I}_\nu, \widetilde{\phi}_m\rangle \to 0, \quad \forall m \in \mathbb{Z},\ \forall \alpha \in D_F. \tag{7.5.8}$$

Assuming that (7.5.1) holds, this amounts to showing that

$$\langle \alpha^{(N)}, F\rangle \to \langle \alpha, F\rangle, \quad \forall m \in \mathbb{Z},\ \forall \alpha \in D_F. \tag{7.5.9}$$

We give some insight about how to prove (7.5.1) and (7.5.9) in the two following paragraphs.

Proof of the weak $T\mathcal{I}_\nu = \mathcal{I}_\nu$ Let us look at $\langle T\mathcal{I}_\nu^N, \tilde{\phi}_m\rangle$. The first step is to express Tf_n on the basis $(\tilde{f}_p)_{n\in\mathbb{Z}}$ using the expression (7.4.4), then use the fact that $\langle \tilde{f}_p, \tilde{\phi}_m\rangle = 0$ for $p \neq m$ to make k_n appear. Finally we use the isomorphism linking k_n to f_n to get

$$\langle T\mathcal{I}_\nu^N, \tilde{\phi}_m\rangle = -\frac{\langle \mathcal{I}_\nu, \widetilde{\phi}_m\rangle}{\overline{\widetilde{\phi}_{m,1}(0)}(1 - e^{-2\mu L})} \sum_{n=-N}^{N} \langle \mathcal{I}_\nu, f_n\rangle \frac{K_n}{f_{n,1}(0)} \left\langle f_n, \widetilde{\phi}_m \right\rangle. \tag{7.5.10}$$

Taking a step back and looking at the general method, we expect that the "$TB = B$" condition (or here $T\mathcal{I}_\nu = \mathcal{I}_\nu$) would add enough condition on $K_n = \langle f_n, F\rangle$ to fully set F. This is reflected in (7.5.10) which

3. see in particular Proposition 7.6.1 below.

depends on K_n and where we have to choose these coefficients in order to ensure $\langle T\mathcal{I}_\nu^{(N)}, \tilde{\phi}_m\rangle \to \langle \mathcal{I}_\nu, \tilde{\phi}_m\rangle$, or, in other words,

$$\frac{-1}{\overline{\widetilde{\phi}_{m,1}(0)}(1-e^{-2\mu L})}\sum_{n=-N}^{N}\langle \mathcal{I}_\nu, f_n\rangle\frac{K_n}{f_{n,1}(0)}\left\langle f_n, \widetilde{\phi}_m\right\rangle \to 1, \tag{7.5.11}$$

when $N \to +\infty$. This requires a kind of equiconvergence result, similar to the Dirichlet sum for instance. In [64] we show the following

Proposition 7.5.1. *Let us denote*

$$\sigma_\mu(f,x) = \sum_{|Im(\mu_p)|<\mu}\langle f, f_p\rangle f_p(x) \tag{7.5.12}$$

and

$$p_\mu(f,x) = \sum_{|Im(\mu_p^{(0)})|<\mu}\langle f, E_p\rangle E_p(x) \tag{7.5.13}$$

where $\mu_p^{(0)}$ are the eigenvalues of $\mathcal{A}$ when $\gamma = 0$ and

$$(E_p)_{p\in\mathbb{Z}} = \left(\left(e^{\frac{i\pi px}{L}}, -e^{\frac{-i\pi px}{L}}\right)\right)_{p\in\mathbb{Z}}. \tag{7.5.14}$$

One has for any compact $K_c \subset [0,L)$

$$\lim_{\mu\to+\infty}\sup_{x\in K_c}|\sigma_\mu(f,x) - p_\mu(f,x)| = 0. \tag{7.5.15}$$

This is done by generalizing a powerful result by Komornic [159, 160]. Looking at (7.5.10), we want to use the estimate (7.5.15) at $x = 0$. This incitate to choose $K_n = C(f_{n,1}(0))^2/\langle \mathcal{I}_\nu, f_n\rangle$ where C is some constant. In fact choosing

$$\langle f_n, F\rangle := -2\tanh(\mu L)\frac{(f_{n,1}(0))^2}{\langle \mathcal{I}_\nu, f_n\rangle}, \quad \forall n \in \mathbb{Z}, \tag{7.5.16}$$

gives the result. Note that thanks to the controllability estimates (7.3.3) and (7.3.5), the feedback F satisfies the condition (7.4.18) so that T is an isomorphism.

Proof of $\langle \alpha^N, F\rangle \to \langle \alpha, F\rangle$ to get the operator equality (7.5.3) We can note that this limit is not trivial, because F lacks of continuity on $D(\mathcal{A})$. More precisely, denoting $\mathcal{E}$ the space of finite linear combinaisons of $(f_n)_{n\in\mathbb{Z}}$ and $\mathcal{E}'$ its dual, we show the following

Lemma 7.5.1. *$F \in \mathcal{E}'$ defined by (7.5.16) defines a linear form on $D(\mathcal{A}^2)$ which is continuous for $\|\cdot\|_{D(\mathcal{A}^2)}$ but not for $\|\cdot\|_{D(\mathcal{A})}$.*

Therefore, to obtain the limit $\langle \alpha^N, F\rangle \to \langle \alpha, F\rangle$, we need to investigate more precisely the regularity of F. This can now be done, since F is completely defined by (7.5.16). Let us define h by

$$\langle f_n, h\rangle = -\frac{\tanh(\mu L)}{\tau_n^{\mathcal{I}}}f_{n,1}(0)\mu_n, \tag{7.5.17}$$

where

$$\tau_n^{\mathcal{I}} = \left(e^{\int_0^L \delta}\frac{f_{n,1}(L)}{f_{n,1}(0)} - 1\right), \quad \forall n \in \mathbb{Z}. \tag{7.5.18}$$

This choice of $\tau_n^{\mathcal{I}}$ comes from the expression of $\langle \mathcal{I}_\nu, f_n\rangle$ given by (7.3.4). What we show is the following

Proposition 7.5.2. *Let X^s be defined as*

$$\{f \in (L^2_{(0)})^2, \quad (\tau^{\mathcal{I}})^{-1}(\Lambda\partial_x f + \delta(x)Jf) \in (H^{s-1})^2\}, \ \forall s \in [0,+\infty). \tag{7.5.19}$$

The linear form h defined by (7.5.17) defines the following linear form on $X^2 \cap D(\mathcal{A})$, continuous for $\|\cdot\|_{X^2}$:

$$\langle \alpha, h \rangle = -\tanh(\mu L) \frac{\left(\mathcal{A}(\tau^{\mathcal{I}})^{-1}\alpha\right)_1(0) - \left(\mathcal{A}(\tau^{\mathcal{I}})^{-1}\alpha\right)_2(0)}{2}. \tag{7.5.20}$$

Moreover, $\widetilde{F} := F - h$ is continuous for $\|\cdot\|_{D(\mathcal{A})}$, so that F is actually defined on $X^2 \cap D(\mathcal{A})$, and is continuous for $\|\cdot\|_{X^2}$, but not for $\|\cdot\|_{D(\mathcal{A})}$.

This means that h is the singular part of F, i.e. the part that limits its regularity [4]. Hence, since $F - h$ is continuous on $D(\mathcal{A})$ the problem is reduced to showing $\langle \alpha^N, h\rangle \to \langle \alpha, h\rangle$.

For this, a first step is to show that $\langle \alpha, h\rangle$ has a meaning for $\alpha \in D_F$, namely that $D_F \subset X^2 \cap D(\mathcal{A})$. Then, we use the fact that $f_{n,1}(0) = f_{-n,1}(0)$ to express $\langle \alpha^N, h\rangle$ as

$$\langle \alpha^N, h\rangle = -\frac{\tanh(\mu L)}{2} \sum_{n=-N}^{N} \left(\frac{\langle \alpha, f_n\rangle}{\tau_n^{\mathcal{I}}}\mu_n + \frac{\langle \alpha, f_{-n}\rangle}{\tau_{-n}^{\mathcal{I}}}\mu_{-n} \right) f_{n,1}(0). \tag{7.5.21}$$

Defining $\tau^{\mathcal{I}} : f_n \to \tau_n^{\mathcal{I}} f_n$, which can be shown to be an isomorphism from $D(\mathcal{A}^2)$ to itself, and σ by

$$\langle \sigma(\beta), f_n\rangle = \langle \beta, f_{-n}\rangle, \quad \forall\, \beta \in L^2((0,1);\mathbb{C}^2), \tag{7.5.22}$$

we have

$$\frac{\langle \alpha, f_n\rangle}{\tau_n^{\mathcal{I}}}\mu_n + \frac{\langle \alpha, f_{-n}\rangle}{\tau_{-n}^{\mathcal{I}}}\mu_{-n} = \left\langle \mathcal{A}((\tau^{\mathcal{I}})^{-1}\alpha - \sigma((\tau^{\mathcal{I}})^{-1}\alpha)), f_n \right\rangle. \tag{7.5.23}$$

which means that

$$\langle \alpha^N, h\rangle = -\frac{\tanh(\mu L)}{2} \sum_{n=-N}^{N} \left\langle \mathcal{A}((\tau^{\mathcal{I}})^{-1}\alpha - \sigma((\tau^{\mathcal{I}})^{-1}\alpha)), f_n \right\rangle f_{n,1}(0). \tag{7.5.24}$$

The last step is to show that for $\beta \in (H^2)^2 \cap D(\mathcal{A})$, we have $\beta - \sigma(\beta) \in D(\mathcal{A}^2)$. Once this is done, we show that for $\alpha \in D_F \subset X^2 \cap D(\mathcal{A})$, we have $(\tau^I)^{-1}\alpha \in (H^2)^2 \cap D(\mathcal{A})$, which means that $(\tau^{\mathcal{I}})^{-1}\alpha - \sigma((\tau^{\mathcal{I}})^{-1}\alpha) \in D(\mathcal{A}^2)$, hence $\mathcal{A}((\tau^{\mathcal{I}})^{-1}\alpha - \sigma((\tau^{\mathcal{I}})^{-1}\alpha)) \in D(\mathcal{A})$. And, as consequence the sum of the right-hand side of (7.5.24) converges absolutely when $N \to +\infty$ and $\langle \alpha^N, h\rangle \to \langle \alpha, h\rangle$.

7.6 Well-posedness of the closed-loop system

Now that we have found an isomorphism T and a feedback control F that maps the original system to the (exponentially stable) target system. The only thing that remains to be done is to show the well-posedness of the closed loop system in some sense. Let us note first that the target system (7.1.23)–(7.1.24) is well-posed in $D(\tilde{\mathcal{A}})$ (see for instance [19, Appendix A]) and there exists a basic quadratic Lyapunov function to the system (7.1.23)–(7.1.24), which has the form

$$V(Z) = \|\Theta(x)Z\|^2_{L^2(0,L)} + \|\Theta(x)(\tilde{\mathcal{A}}Z)\|^2_{L^2(0,L)}, \quad \forall\, Z \in H^p(0,L), \tag{7.6.1}$$

where Θ is a positive C^1 function on $[0,L]$. By the definition of a basic Lyapunov function, V is equivalent to the square of the H^1 norm, thus its square root is a norm itself on $D(\mathcal{A})$. This implies that $\tilde{\mathcal{A}}$ generates a contraction semigroup on $D(\tilde{\mathcal{A}})$ for this norm (the domain of $\tilde{\mathcal{A}}$ as infinitesimal generator is then $D(\tilde{\mathcal{A}}^2)$). Let us denote by $(\tilde{S}(t))_{t\geq 0}$ the semigroup thus generated. Intuitively, to define a solution to our original system and show the well-posedness, we would like to apply T to the initial condition, then find the solution

4. This situation is similar to what we have done with the heat equation in Lemma 5.3.2 where we isolate the singular part of q_n.

for the target system, and define the solution of the original system at time t as being T^{-1} applied to the solution of the target system at time t. This leads to define the following semigroup

$$\begin{aligned} S:\ \mathbb{R}_+ &\to \mathcal{L}(D(\mathcal{A})) \\ t &\mapsto T^{-1}\widetilde{S}(t)T. \end{aligned} \tag{7.6.2}$$

What we can show is that the closed-loop operator $-\mathcal{A}+\mathcal{I}_\nu F$ indeed generates this semigroup $(S(t))_{t\geq 0}$, and thus that the original closed-loop system is well-posed. More precisely we can show the following

Proposition 7.6.1. *The mapping $(S(t))_{t\geq 0}$ defines an exponentially stable C^0-semigroup on $D(\mathcal{A})$ with decay rate $\mu/2$, and its infinitesimal generator is the unbounded operator $-\mathcal{A}+\mathcal{I}_\nu F$ defined on D_F given by*

$$D_F = \left\{ \alpha \in D(\mathcal{A}) | \alpha = \sum_{n\in\mathbb{Z}} \alpha_p h_p,\ (\widetilde{\mu}_p \alpha_p)_{p\in\mathbb{Z}} \in \ell^2 \right\}. \tag{7.6.3}$$

Moreover, this semigroup is real-valued on real-valued functions.

Showing that $(S(t))_{t\geq 0}$ is a C^0 semigroup follows directly from its definition (7.6.2), the continuity of T and T^{-1}, and the fact that $(\tilde{S}(t))_{t\geq 0}$ is a C^0 semigroup. By definition, its infinitesimal generator is the operator $\mathcal{L}$ defined by

$$\mathcal{L}x = \lim_{t\to 0^+} \frac{1}{t}(S(t)-Id)x, \tag{7.6.4}$$

on the domain $D_\mathcal{L}$ such that this limit has a sense. The exponential stability is a direct consequence of the fact that $(\tilde{S}(t))_{t\geq 0}$ is not only a contraction semigroup but also an exponentially stable semigroup on $D(\tilde{\mathcal{A}})$ with decay rate (at least) $\mu/2$, which results again from the existence of a basic Lyapunov function for the H^1 norm and the continuity of T. Indeed, one has for $\alpha \in D(\mathcal{A})$ and $t \geq 0$,

$$\|\tilde{S}(t)\alpha\|_{D(\mathcal{A})} \leq Ce^{-\mu t/2}\|\alpha\|_{D(\mathcal{A})}, \tag{7.6.5}$$

where C is independent of α. Hence, from (7.6.2),

$$\|S(t)\alpha\|_{D(\mathcal{A})} \leq Ce^{-\mu t/2}\|T^{-1}\|_{\mathcal{L}(D(\tilde{\mathcal{A}}),D(\mathcal{A}))}\|T\|_{\mathcal{L}(D(\mathcal{A}),D(\tilde{\mathcal{A}}))}\|\alpha\|_{D(\mathcal{A})}, \tag{7.6.6}$$

and S is exponentially stable. It remains to show that $(\mathcal{L}, D_\mathcal{L}) = (-\mathcal{A}+\mathcal{I}_\nu F, D_F)$. For this we need to show the following Lemma

Lemma 7.6.1. *The domain D_F satisfies the following equality:*

$$D_F = T^{-1}D(\widetilde{\mathcal{A}}^2). \tag{7.6.7}$$

This lemma explains why D_F appears: although the set D_F might seem peculiar at first, it is really the analogous of $D(\tilde{\mathcal{A}}^2)$ for the original closed-loop system. It also explains why it was natural to show the operator equality (7.5.3) for any $\alpha \in D_F$ in the previous section. In order to show Lemma 7.6.1, we will need an additional Lemma.

Lemma 7.6.2. *The operator $-\mathcal{A}+\mathcal{I}_\nu F$ admits a Riesz basis of eigenvectors in $D(\mathcal{A})$, given by*

$$h_p := T^{-1}\frac{\widetilde{f}_p}{\widetilde{\mu}_p}, \quad \forall p \in \mathbb{Z}, \tag{7.6.8}$$

with corresponding eigenvalues $(-\widetilde{\mu}_p)_{p\in\mathbb{Z}}$.

The fact that $(h_p)_{n\in\mathbb{Z}}$ is a Riesz basis of $D(\mathcal{A})$ is a direct consequence from the fact that T^{-1} is an isomorphism from $D(\tilde{\mathcal{A}})$ to $D(\mathcal{A})$ and the fact that $(\tilde{f}_n/\tilde{\mu}_n)_{n\in\mathbb{Z}}$ is a Riesz basis of $D(\tilde{\mathcal{A}})$. This last claim comes from the fact that $(\tilde{f}_n)_{n\in\mathbb{Z}}$ is a basis of $L^2((0,1);\mathbb{C}^2)$ of elements of $D(\tilde{\mathcal{A}})$ and $(|\tilde{\mu}_n|)^{-1} \sim (1+|n|)^{-1}$. The difficulty is to show that $h_p \in D_F$. The functions h_p are likely regular enough, but requiring that $\mathcal{A}h_p + \langle h_p, F\rangle\mathcal{I}_\nu$ belongs to $D(\mathcal{A})$ implies that $\mathcal{A}h_p + \langle h_p, F\rangle\mathcal{I}_\nu$ satisfies the correct boundary conditions imposed by $D(\mathcal{A})$.

This is not obvious as $\mathcal{I}_\nu$ does not satisfy them. This can nevertheless be shown by expressing $\tilde{f}_p$ as a function of h_p, and then decompose h_p on f_n to get

$$\langle h_p, f_n\rangle = a_{n,p} = -\frac{\overline{\widetilde{\phi}_{p,1}(0)}(1-e^{-2\mu L})}{\widetilde{\mu}_p\langle \mathcal{I}_\nu, \widetilde{\phi}_p\rangle}\frac{f_{n,1}(0)}{\langle f_n, F\rangle}\langle \widetilde{f}_p, f_n\rangle. \tag{7.6.9}$$

Using this and the equiconvergence given in Proposition 7.5.1, we can compute $\langle h_p, F\rangle$ and then the coefficients $\langle -\mathcal{A}h_p + \mathcal{I}_\nu\langle h_p, F\rangle, f_n\rangle$ as functions of $\langle \tilde{f}_p, f_n\rangle$. Using then the same trick as for (7.4.10)–(7.4.12) we are eventually able to obtain that

$$\langle -\mathcal{A}h_p + \mathcal{I}_\nu\langle h_p, F\rangle, f_n\rangle = -\frac{\widetilde{f}_{p,1}(0)\overline{\widetilde{\phi}_{p,1}(0)}(1-e^{4\mu L})}{2\langle \mathcal{I}_\nu, \widetilde{\phi}_p\rangle}\frac{\langle \mathcal{I}_\nu, f_n\rangle}{\widetilde{\mu}_p - \mu_n}, \tag{7.6.10}$$

which means that

$$-\mathcal{A}h_p + \mathcal{I}_\nu\langle h_p, F\rangle = \sum_{n\in\mathbb{Z}}\left(-\frac{\widetilde{f}_{p,1}(0)\overline{\widetilde{\phi}_{p,1}(0)}(1-e^{4\mu L})}{2\langle \mathcal{I}_\nu, \widetilde{\phi}_p\rangle}\right)\langle \mathcal{I}_\nu, f_n\rangle\frac{f_n}{\widetilde{\mu}_p - \mu_n}. \tag{7.6.11}$$

Recall that $|\mu_n| \sim (1+|n|)$ and, from the controllability estimate (7.3.5), we have $(\langle \mathcal{I}_\nu, f_n\rangle)_{n\in\mathbb{Z}} \in \ell^2$. Therefore, as $\left(\frac{f_n}{\widetilde{\mu}_p-\mu_n}\right)_{n\in\mathbb{Z}}$ is a Riesz basis of $D(\mathcal{A})$, we have $-\mathcal{A}h_p + \mathcal{I}_\nu\langle h_p, F\rangle \in D(\mathcal{A})$ and $h_p \in D_F$. Lemma 7.6.1 then follows from the fact that

$$D_F = \left\{\alpha \in D(\mathcal{A}) | \alpha = \sum_{n\in\mathbb{Z}}\alpha_p h_p,\ (\widetilde{\mu}_p\alpha_p)_{p\in\mathbb{Z}} \in \ell^2\right\}, \tag{7.6.12}$$

and from the operator equality (7.5.3) applied to $\alpha \in D_F$.

From this, we can finish the proof of Proposition 7.6.1. Let $\alpha \in D_F$, then $T\alpha \in D(\tilde{\mathcal{A}}^2)$ and, because $-\tilde{\mathcal{A}}$ is the infinitesimal generator of $(\tilde{S}(t))_{t\geq 0}$ with domain $D(\tilde{\mathcal{A}}^2)$ we have, by definition,

$$\frac{\widetilde{S}(t)T\alpha - T\alpha}{t} \xrightarrow[t\to 0^+]{D(\tilde{\mathcal{A}})} -\tilde{\mathcal{A}}T\alpha. \tag{7.6.13}$$

Since $\alpha \in D_F$ we can use the operator equality (7.5.3) and $\tilde{S}(t) = TS(t)T^{-1}$ to get

$$\frac{TS(t)\alpha - T\alpha}{t} \xrightarrow[t\to 0^+]{D(\tilde{\mathcal{A}})} T(-\mathcal{A} + \mathcal{I}_\nu F)\alpha \tag{7.6.14}$$

Since T^{-1} is a continuous linear application from $D(\tilde{\mathcal{A}})$ to $D(\mathcal{A})$, we have

$$\frac{S(t)\alpha - \alpha}{t} \xrightarrow[t\to 0^+]{D(\tilde{\mathcal{A}})} (-\mathcal{A} + \mathcal{I}_\nu F)\alpha, \tag{7.6.15}$$

and hence $(-\mathcal{A} + \mathcal{I}_\nu F)$ is the infinitesimal generator of $(S(t))_{t\geq 0}$. Finally, as $(\tilde{S}(t))_{t\geq 0}$ is real valued on real-valued function, and as T^{-1} and T map real valued function to real valued function, we deduce that $(S(t))_{t\geq 0}$ also maps real valued function to real valued function. This concludes the well-posedness in $D(\mathcal{A})$ of the original closed-loop system.

7.7 Open questions and perspectives

The generalized backstepping (or F-equivalence) approach is fairly new. Therefore there are many interesting open-questions yet to be answered, and we give below just a few of them.

— The compactness-duality approach presented in Chapter 6 works for anti-adjoint operators. Is it possible to extend it to other operators, like self-adjoint operators, or more general operators? If not is there a new method that could work?

— Is it possible to obtain a finite-time stabilization using the approaches of Chapters 5–7? This is known to be true for a scalar hyperbolic system [256], but the question remains largely open in general. The key step for this would be to obtain an explicit dependency on the lower bound of $|K_n|$, the coefficients of the feedback operator. So far, the way this lower bound is obtained in Chapter 5–6 is very indirect: we show that T is an isomorphism in a weak space and we deduce that there exists a bound on K_n. Being able to have an explicit estimate with respect to λ would be very interesting.

— Is it possible to extend the results presented in 7 to quasilinear systems, at least locally? Even for the transport equation this is an open question. This problem is harder as one may think given the lack of robustness of spectral properties for hyperbolic systems (in particular the fact that in general the exponential stability of the linear system does not give any information on the local exponential stability of the associated nonlinear system [70]).

— Can the approach of Section 7 be generalized to all linear (controllable) hyperbolic systems?

— Is there a general abstract framework that could encompass all the cases $\alpha > 1$ or even $\alpha \geq 1$?

— Is there a way to extend this approach in two space dimensions, namely $x \in \mathbb{R}^2$? At least for the heat equation where the eigenvalues of the Laplacian are easily described. And, if so, it is possible to extend it to $x \in \mathbb{R}^n$ for any $n \in \mathbb{N}^*$?

Part III

Control of traffic flow

Chapter 8

Control of traffic flows: microscopic approach

8.1 Introduction

This chapter starts with a less mathematical and more practical question: what happens when many cars are placed on a road with the same speed and spacing? The empirical answer is known to everyone who has ever driven a car during rush hour: a traffic jam forms. A *stop-and-go wave* phenomenon is formed, where cars alternate between speeding up and slowing down, often keeping their speed low. This kind of traffic jam is what is sometimes called a "ghost" traffic jam: it has no apparent cause, no accidents on the road, no lane reductions, etc.

This traffic jam actually has a very mathematical underlying cause: as we will see later, above a certain density of cars, the uniform steady state (i.e., the state of smooth traffic where all cars have the same speed) is not a stable state. This chapter focuses on making this uniform steady state stable, using autonomous vehicles that play the role of controls.

Being able to stabilize the traffic flow has attracted interest for decades. Many strategies have been considered as using ramp metering or junctions as a boundary control of the traffic system [19, 101, 129, 158, 196, 228]. Another classical approach is to allow a variable speed limit and to use it as a controller (see for instance [119, 140]). With the democratization of autonomous vehicles (briefly AVs) and means of communication, the idea of using AVs as a means of control to regulate traffic has gained momentum in recent years [76, 231, 241, 242, 262] or [89, 223] for a more detailed review, and [27, 51, 79, 112, 178, 201, 210] for examples of traffic control where the traffic is modelled by a hyperbolic PDE). Of course, designing theoretically a control requires first to choose an accurate model [1] for the system and, first of all, a scale.

Mathematically, there are three possible points of view to study this problem:

— the microscopic approach, where each car is modeled separately by one or more differential equations.

— the mean field approach, which consists in passing the microscopic model to the higher scale. To do this, we try to make the number of vehicles tend towards infinity and to define a limit solution to the system in the form of a measure. This allows to capture the main characteristics of the microscopic model while describing the traffic in an aggregated way as the solution of a partial differential equation.

1. There are also model-free initiatives to derive a controller using AI tools. In particular deep-reinforcement learning trained on real trajectories with well defined reward functions and policies (see for instance [179, 238, 239]). Here, we let them aside and only look at the mathematical approaches.

— the macroscopic approach, where road traffic is represented by macroscopic quantities (e.g., density and speeds) that are solutions of partial differential equations, while autonomous vehicles are represented by solutions of ordinary differential equations.

In this section we will focus on the first and third approaches. More details about the second approach can be found for instance in [86, 122, 124, 141, 197, 209].

8.2 Microscopic approach

A first way to model traffic is to represent each vehicle separately. Physically, human drivers (and the algorithms implemented in AVs) act through the command of the cars on the acceleration of the vehicle. It is therefore natural to model the vehicles as follows

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= f(t, x, , x_l, v_l, v), \end{aligned} \tag{8.2.1}$$

where x is the position of the given vehicle, v its velocity, and f corresponds to the action on the acceleration which can depend on x, on v but also on the time t and external parameters, like the position and the velocity of the vehicle in front of it, denoted respectively by x_l and v_l. The function f can therefore take into account both the driver's action and physical phenomena (such as friction). This function is called a *car-following model*. Given the industrial stakes and the low mathematical entry cost of the microscopic approach, many car-following models have been developed over the years, with the objective of fitting the reality as well as possible, either in general or for a particular phenomena. In the following, we will focus on one model: the *Bando-Follow the Leader* (Bando-FTL) model also sometimes called *Optimal Velocity - Follow the Leader* [17, 113]. This model has the advantage of being able to represent the stop-and-go waves and the traffic jam phenomena we are interested in in this chapter. Another model, the *Intelligent Driver Model* (IDM), most used in engineering, can also represent these stop-and-go waves, but will not be studied here. It was studied in [6] where we show that, surprisingly, this model as such is ill-posed. On the other hand, the well-posedness of Bando-FTL has been studied in [123]. This section taken from [138], a collaboration with Benedetto Piccoli and Sydney Truong.

8.2.1 The Bando-FTL model

As its name indicates, this model is in fact the sum of two parts, the *Bando* model (or *Optimal velocity*) and the *Follow-the Leader* model. The Bando model, introduced in [17], is given by

$$\dot{v} = k(V(x_l - x) - v), \tag{8.2.2}$$

where k is a positive coefficient, V is a so-called optimal speed function, such that $V(0) = 0$, $\lim_{h \to +\infty} V(h) \in \mathbb{R}$, and x_l is the position of the leading vehicle. V represents the driver's preferred speed as a function of the distance to the vehicle ahead. This speed is usually taken strictly increasing and equal to a hyperbolic tangent ratio (see [17] or (8.2.38)). We will not use this assumption in the following and we will only assume that V is strictly increasing and C^2. The Follow-The-Leader model, introduced in [113], is given by

$$\dot{v} = k \frac{v_l - v}{(x_l - x)^2} \tag{8.2.3}$$

where k is still a positive coefficient, x_l is the position of the leading vehicle and v_l its speed. This model reflects the driver's preference to have the same speed as the vehicle in front of him, and the closer this vehicle is, the stronger this preference is. These two models have drawbacks when taken separately since they account for different phenomena that both exist : the Follow-the-Leader model prevents collisions but all distances between vehicles can give an equilibrium and it has no locally unstable equilibrium, so it is not able to represent real traffic alone. The Bando model, on the other hand, has difficulty preventing collisions,

which makes it rather unsuitable for use alone . This is why a commonly used version consists in considering a linear combination of these two models, it is the Bando-FTL model. Each car is modeled as follows:

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= a(V(x_l - x) - v) + b\frac{v_l - v}{(x_l - x)^2}, \end{aligned} \tag{8.2.4}$$

where a et b are two parameters that represent respectively the weight of the Bando part and the Follow-the Leader part.

8.2.2 Steady-states and stabilisation

We now consider a circular road of length $L > 0$ with $N \in \mathbb{N}^*$ vehicles represented by their positions and speeds $(x_i, v_i)_{i\in\{1,...,N\}}$. The system is then described by

$$\begin{aligned} \dot{x}_i &= v_i \\ \dot{v}_i &= a(V(x_{i+1} - x_i) - v_i) + b\frac{v_{i+1} - v_i}{(x_{i+1} - x_i)^2}, \quad i \in \{1, ..., N\}, \end{aligned} \tag{8.2.5}$$

with the convention $x_{N+1} = x_1$ et $v_{N+1} = v_1$ on a circular road with N cars. In reality, we are not interested in stabilizing the positions and speeds of the vehicles but rather their spacings and speeds. We can therefore reformulate the system (8.2.5) as follows by denoting $h_i = x_{i+1} - x_i$,

$$\begin{aligned} \dot{h}_i &= v_{i+1} - v_i \\ \dot{v}_i &= a(V(h_i) - v_i) + b\frac{v_{i+1} - v_i}{(h_i)^2}, \quad i \in \{1, ..., N\}. \end{aligned} \tag{8.2.6}$$

Given that we do not want any collision the state-space considered is $(h_i, v_i)_{i\in\{1,...,N\}} \in (0, +\infty)^N \times \mathbb{R}_+^N$ which corresponds in the original coordinates (x, v) to the following:

$$\begin{aligned} &(x_i(t))_{i\in\{1,...,N\}} \in \mathcal{O}_N := \{(z_i)_{i\in\{1,...,N\}} \in \mathbb{R}/L\mathbb{Z} \mid 0 \le z_1 < z_2 < ... < z_n \le L\}, \\ &(v_i(t))_{i\in\{1,...,N\}} \in \mathbb{R}_+^N. \end{aligned} \tag{8.2.7}$$

Clearly, there is only one possible steady state of this system in h and v, taking into account that the road is circular hence $x_{N+1} = x_1$ and $\sum_{i=1}^N h_i = L$. This steady state is given by

$$\bar{v} = V(L/N), \quad h_i = L/N. \tag{8.2.8}$$

Depending on the parameters a, b, and the function V, this system can have stable or unstable dynamics. More precisely we have the following proposition:

Proposition 8.2.1 ([76]). *Let d be the steady-state spacing, i.e. $d = L/N$. If*

$$\frac{b}{2} + (\frac{a}{d^2}) < V'(d) \tag{8.2.9}$$

then the system (8.2.6) is unstable.

This explains (qualitatively) the phenomenon we talked about in the introduction: the steady state becomes unstable for some densities. This phenomenon has in fact been clearly measured experimentally [227] and reproduced in the case of a circular road [76].

We now consider a control which consists of an autonomous vehicle. We represent this autonomous vehicle by (x_{N+1}, v_{N+1}) (and we therefore abandon the convention $x_{N+1} = x_1$ and $v_{N+1} = v_1$ for $x_{N+2} = x_1$ and $v_{N+2} = v_1$). Its dynamics is given by

$$\begin{aligned} \dot{x}_{N+1} &= v_{N+1} \\ \dot{v}_{N+1} &= u(t), \end{aligned} \tag{8.2.10}$$

where $u(t)$ is our control. We can note that the appearing of this new vehicle gives rise to new accessible steady states, which form a continuum. Indeed, provided that $u(t)$ can be chosen, for any $(\bar{h}, \bar{v})$ satisfying

$$\bar{h} < \frac{L}{N}, \quad \bar{v} = V(\bar{h}), \tag{8.2.11}$$

the state

$$\begin{aligned} h_i &= \bar{h}, \quad \forall i \in \{1, ..., N\}, \\ v_i &= \bar{v}, \quad \forall i \in \{1, ..., N+1\} \end{aligned} \tag{8.2.12}$$

describes a steady-state[2]. In the following we will aim to define a feedback control u which stabilizes such a steady state of the system (8.2.6)–(8.2.10) that we will denote by $(\bar{v}, \bar{h})$. Ideally, we would like to have a control as simple as possible, depending only on local measurements around the AV. We will therefore first look for a feedback control of the form

$$u(t) = -k(v_{N+1} - \bar{v}), \tag{8.2.13}$$

where k is a control parameter. We denote $\mathbf{h} = (h_1, ..., h_N)$, $\mathbf{v} = (v_1, ..., v_{N+1})$ and we define the vectors $\bar{\mathbf{h}} = (\bar{h}, ..., \bar{h}) \in \mathbb{R}^n$ and $\bar{\mathbf{v}} = (\bar{v}, ..., \bar{v}) \in \mathbb{R}^{n+1}$. The main result of this section is the following

Theorem 8.2.2. *Let $(\bar{v}, \bar{h})$ be a steady-state as described by (8.2.11) and such that Proposition 8.2.1 applies (i.e. (8.2.9) holds and the corresponding open-loop system is unstable). If $k > 0$ then the system (8.2.6), (8.2.10) with control feedback (8.2.13) is locally exponentially stable around $(\bar{v}, \bar{h})$.*
Moreover, there exists a uniform decay rate $\gamma_{uniform}$ independent of N, $(\bar{v}, \bar{h})$ and L that can be achieved, and for any $\gamma \in (0, \gamma_{uniform})$, there exists a characteristic time $\tau > 0$, independent of N, and $\varepsilon > 0$ such that for any initial conditions $(\mathbf{v}_0, \mathbf{h}_0)$ satisfying

$$\|(\mathbf{v}_0 - \bar{\mathbf{v}}, \mathbf{h}_0 - \bar{\mathbf{h}})\| \leq \varepsilon, \tag{8.2.14}$$

we have

$$\|\mathbf{h}(t) - \bar{\mathbf{h}}, \mathbf{v}(t) - \bar{\mathbf{v}}\| \leq e^{N\tau} e^{-\gamma t} \|\mathbf{h}_0 - \bar{\mathbf{h}}, \mathbf{v}_0 - \bar{\mathbf{v}}\|. \tag{8.2.15}$$

Finally, for a given steady-state $(\bar{v}, d)$, the supremum value of the achievable decay rate is

$$\gamma_{\max} = \min\left(k, \frac{1}{2}\left(\frac{a}{d^2} + b - Re\left(\sqrt{\left(\frac{a}{d^2} + b\right)^2 - 4bV'(d)}\right)\right)\right). \tag{8.2.16}$$

We can make several remarks on this theorem:

Remark 8.2.1 (Uniformity in N)**.** *This result holds for any number of cars N while there is always a single AV. Besides, note that one can achieve an exponential decay rate that is uniform both with respect to N and L and the steady-state considered. And the control gain k can also be made independent of N, L and the steady-state considered.*

This uniformity in N may seem a bit strange, nevertheless it becomes more logical when we look at the lower bound that we are able to obtain on the basin of attraction that strongly decreases with N (see Proposition 8.2.3 below).

Remark 8.2.2 (Relaxation time)**.** *The total relaxation time τ_R, which can be seen as a characteristic time needed to stabilize the system, is defined as*

$$\tau_R := \ln(C_{inf}), \tag{8.2.17}$$

where C_{inf} is the infimum of the values of C for a given γ such that (8.2.15) holds. As it could be expected this total relaxation time is not necessarily uniform in N. However, from (8.2.15), there exists τ independent of N such that $\tau_R \leq N\tau$, and therefore τ_R is, at most, linear with N. This at-most-linear dependency seems intuitively to be the best we could hope, given the finite speed of propagation of the information in the system.

2. Of course $u(t)$ has to be compatible with this. Note that h_{N+1} is imposed from the fact that the road is circular, i.e. $\sum_{i=0}^{N+1} h_i = L$

The reason why γ_{uniform} may be different from $\gamma_{\max}$ is that $\gamma_{\max}$ depends on $\bar{h}$ which implicitly depends on N since $\bar{h} \leq L/N$ and could tend to 0 when $N \to +\infty$. Also, one can note that L could have any dependency in N so far. The two following remarks address the case where $\bar{h}$ can stay away from 0, either because L scales with N or because we assume (or require) a minimal safety distance on the road, hence there exists $d_1 > 0$ such that $\bar{h} \geq d_1$ for any N considered.

Remark 8.2.3 (Case of a road length that depends on N)**.** *If L does not depend on N, then, as $\bar{h} \leq L/N$, the value $\gamma_{\max}$ might depend indirectly of N (while $\gamma_{uniform}$ is still uniform in N). However, if we assume that there exists α such that $L \geq \alpha N$, which is a physically reasonable assumption given that a road with a fixed length cannot hold an infinite number of cars, then $(\bar{v}, \bar{h})$ can be chosen independently of N and $\gamma_{\max}$ is also uniform in N (but still depends on the steady-state chosen).*

Remark 8.2.4 (Case of a safety distance $d_{\min}$)**.** *Whether L depends or N or not, if we assume that the desired steady-state satisfies $\bar{h} \geq d_{\min} > 0$ (which simply means that the steady-state headway has to be larger than some safety distance), then there exists again an achievable decay rate independent of N, L, $(\bar{v}, \bar{h})$ without restricting to the steady-states that are unstable in open-loop. This can be showed in the same way as the existence of $\gamma_{uniform}$ (see [138, (3.17)]).*

Proposition 8.2.3. *There exists a lower bound η on the basin of attraction which satisfies:*

$$\eta < \eta_0 e^{-\alpha N}, \tag{8.2.18}$$

where η_0 and α are constants independent of N (but may depend on $\bar{h}$).

Before giving some ideas of proofs, we can note that the condition $k > 0$ is necessary:

Proposition 8.2.4. *Let $(\bar{v}, \bar{h})$ be an admissible steady-state as in (8.2.11). If $k \leq 0$, then the system (8.2.6), (8.2.10) with control feedback (8.2.13) is not asymptotically stable around $(\bar{v}, \bar{h})$.*

This proposition is immediately deduced from the form of the control (8.2.13).

8.2.2.1 Ideas for the proof of Theorem 8.2.2

Exponential stability and decay rate. First of all let us notice that we are trying to stabilize 2N+1 variables which are $(h_1, ..., h_N)$ and $(v_1, ..., v_{N+1})$. Indeed we do not need to try to stabilize the variable $h_{N+1} := x_1 - x_{N+1}$ which will be anyway imposed by the fact that the road has a fixed length L, hence

$$\sum_{i=1}^{N+1} h_i = L. \tag{8.2.19}$$

We set the following change of variables

$$\begin{aligned} y_{2p+1} &= h_{p+1} - d, \ 0 \leq p \leq N-1 \\ y_{2p} &= v_{p+1} - v_p, \ 1 \leq p \leq N \\ y_{2N+1} &= \bar{v} - v_{N+1}. \end{aligned} \tag{8.2.20}$$

and the system (8.2.6), (8.2.10), (8.2.13), becomes

$$\dot{\mathbf{y}} = f(\mathbf{y}, k), \tag{8.2.21}$$

where

$$\begin{aligned} f_{2p+1}(y,k) &= y_{2p+2}, \ 0 \leq p \leq N-1 \\ f_{2p}(y,k) &= a\left[\frac{y_{2p+2}}{(d+y_{2p+1})^2} - \frac{y_{2p}}{(d+y_{2p-1})^2}\right] + b[V(d+y_{2p+1}) - V(d+y_{2p-1}) - (y_{2p})], \ 1 \leq p \leq N-1, \\ f_{2N}(y,k) &= k y_{2N+1} - a\left[\frac{y_{2N}}{(d+y_{2N-1})^2}\right] - b[V(d+y_{2N-1}) + y_{2N} + y_{2N+1} - \bar{v}] \\ f_{2N+1}(y,k) &= -k y_{2N+1}. \end{aligned} \tag{8.2.22}$$

The exponential stability for $k > 0$ is then observed relatively easily. The key point being that the $N+1$-th vehicle breaks the natural (unstable) feedback loop of the uncontrolled system. The system is thus purely cascaded: for any $p \in \{0, ..., N-2\}$, the dynamics of $(y_{2p+1},\ y_{2p+2})$ depends only on itself and the following coordinates and more precisely only on $(y_{2p+1},\ y_{2p+2},\ y_{2p+3},\ y_{2p+4})$. Similarly the dynamics of $(y_{2N-1},\ y_{2N})$ depend only on $(y_{2N-1},\ y_{2N},\ y_{2N+1})$ and the dynamics of y_{2N+1} depends only on itself. This can be seen very well if we look at the Jacobian matrix $\partial_y f(\mathbf{0}, k)$ which is given by

$$\partial_y f(\mathbf{0}, k) = \begin{pmatrix} A & B & 0 & ... & 0 & \\ 0 & A & B & 0 & ... & \\ ... & ... & ... & ... & ... & \\ 0 & ... & 0 & A & B & 0 \\ 0 & ... & 0 & 0 & A & \begin{pmatrix} 0 \\ k-b \end{pmatrix} \\ 0 & ... & 0 & 0 & 0 & -k \end{pmatrix}, \tag{8.2.23}$$

where A and B are the 2×2 blocks

$$\begin{aligned} A &= \begin{pmatrix} 0 & 1 \\ -bV'(d) & -(\frac{a}{d^2} + b) \end{pmatrix}, \\ B &= \begin{pmatrix} 0 & 0 \\ bV'(d) & \frac{a}{d^2} \end{pmatrix}. \end{aligned} \tag{8.2.24}$$

The exponential stability and the associated exponential decay rate $\gamma_{\max}$ are direct consequences of the Jordan-Chevalley decomposition, of the fact that A is trigonalizable with strictly negative eigenvalues, and of the fact that $k > 0$. We are then able to find a uniform decay rate γ_{uniform} by showing that, if the target steady state corresponds to an unstable state in open loop, then d admits a strictly positive lower bound independent of N and L. The continuity of the decay rate obtained previously allows then to conclude.

Estimation of the relaxation time. Estimating the relaxation time requires a little more work. Let us look at the linearized system first. A simple way is to trigonalize the system, possible since A is trigonalizable, to get back to

$$\dot{\xi} = \begin{pmatrix} \Lambda_1 & B_2 & 0 & ... & 0 & \\ 0 & \Lambda_1 & B_2 & 0 & ... & \\ ... & ... & ... & ... & ... & \\ 0 & ... & 0 & \Lambda_1 & B_2 & 0 \\ 0 & ... & 0 & 0 & \Lambda_1 & \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\ 0 & ... & 0 & 0 & 0 & -k \end{pmatrix} \xi, \tag{8.2.25}$$

where Λ_1 is upper triangular with negative eigenvalues $\lambda_1 < \lambda_2$. We then use the following ISS estimate: for a 2×2 system with two external inputs of the form

$$\frac{d}{dt} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \Lambda_1 \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} + B_2 \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{8.2.26}$$

we have, for any $\varepsilon > 0$,

$$|\mathbf{q}(t)| \leq |\mathbf{q}(0)| M(\varepsilon) e^{-(\mathrm{Re}(|\lambda_2|-\varepsilon))t} + 2M(\varepsilon)\|B_2\|_\infty \int_0^t e^{-(|\mathrm{Re}(\lambda_2)|-\varepsilon)(t-s)} |\mathbf{w}(s)|\, ds,\ \forall\, t \geq 0, \tag{8.2.27}$$

Setting $\gamma_{\max} = \min(k, |\mathrm{Re}(\lambda_2)|)$ (which is in fact exactly the $\gamma_{\max}$ of Theorem 8.2.2), this gives

$$|\mathbf{q}(s)| e^{(\gamma_{\max}-\varepsilon)t} \leq M(\varepsilon)|\mathbf{q}(0)| + 2M(\varepsilon)\|B_2\|_\infty \int_0^t e^{(\gamma_{\max}-\varepsilon)s} |\mathbf{w}(s)|\, ds. \tag{8.2.28}$$

This estimate allows us to trace the cascade system in ξ, estimating first (ξ_1, ξ_2), then (ξ_2, ξ_3), etc. We finally obtain an estimate of the form

$$\left(|\xi_{2N+1}(t)| + \sum_{j=0}^{N-1} |(\xi_{2j+1}(t), \xi_{2j+2}(t))^T|\right) e^{(\gamma_{\max}-\varepsilon)t} \leq \left(|\xi_{2N+1}(0)| + \sum_{i=0}^{N} |(\xi_{2i+1}(0), \xi_{2i+2}(0))^T|\right) M(\varepsilon) \times \left(\sum_{j=0}^{N} \frac{(M(\varepsilon)\max(2\|B_2\|_\infty, |c_1|, |c_2|)t)^j}{j!}\right), \tag{8.2.29}$$

For all $\gamma \in (0, \gamma_{\max})$, we can then conclude by setting $\varepsilon = (\gamma_{\max} - \gamma)/2$ and by showing that

$$\sup_{t\in[0,+\infty)} \left| e^{-\varepsilon t} \sum_{j=0}^{N} \frac{(2M(\varepsilon)\max(\|B_2\|_\infty, |c_1|, |c_2|)t)^j}{j!} \right| \leq 1 + C_0 \sum_{j=1}^{N} \frac{1}{\sqrt{2\pi j}} \left(\frac{C_1}{\varepsilon}\right)^j \leq C_0 e^{\tau N}, \tag{8.2.30}$$

where C_0 is a numerical constant independent of N and

$$\tau = \ln(4M(\varepsilon)\max(\|B_2\|_\infty, |c_1|, |c_2|)/(\gamma_{\max} - \gamma)). \tag{8.2.31}$$

A similar procedure can be done with the nonlinear system by treating the nonlinear terms as perturbations, since we are looking at a local exponential stability. More details can be found in [138].

Estimation of the basin of attraction. The goal of this paragraph is to obtain a lower bound on the basin of attraction of the form $\eta = \eta_0^N$ with η_0 an explicit constant (less than 1) and independent of N. This lower bound decreases exponentially fast with N. The idea of the proof is first to show that for all $T > 0$, there exists such a bound of the form $\eta = \eta_0^N$ on the initial conditions such that the trajectories of the nonlinear system exist and are regular on $[0, T]$. Then, we obtain a bound (depending on T) on the basin of attraction and we end up choosing a T_1 such that this bound also works on $[0, +\infty)$. For this, we select T_1 such that the state at T_1 has a norm smaller than the bound required on the initial condition, thanks to the exponential decay. These arguments are classical and we will not detail them here.

8.2.2.2 Numerical simulations and multilane experiments

Semi-global stabilization and multilane setting Theorem 8.2.2 is a local stabilization result. In practice we would like to be able to have more than this: when the traffic is already congested, we often have quite high speed variations compared to the steady state and the assumption of small perturbations does not hold anymore. Moreover, we would also like to be able to stabilize real roads with several lanes. We can model these roads in the following way: the system consists of J lanes each obeying a system of the form (8.2.6) with N_j vehicles where j is the lane number, and a vehicle i passes from lane j to lane $j_1 \in \{j-1, j+1\}$ when [3]:

$$\tilde{a}_i^{j_1} > a_i^j + \Delta \tag{8.2.32}$$

$$\tilde{a}_i^{j_1} > -\Delta, \quad \tilde{a}_{\text{fol}}^{j_1}(i) > -\Delta \tag{8.2.33}$$

where $a_i^j = \dot{v}_i^j$ is the current acceleration in lane j of the i-th vehicle given by (8.2.6), while $\tilde{a}_i^{j_1}$ is the expected acceleration it would have in the new lane. That is, $\tilde{a}_i^{j_1}$ is the acceleration that the i-th vehicle would have if it were in lane j_1 instead with its current speed and location. Finally $\tilde{a}_{\text{fol}}^{j_1}(i)$ is the expected acceleration of the follower of the i-th vehicle in lane j_1, that is, the acceleration that would have the vehicle right behind the i-th vehicle in the new lane if the lane switching occurs.

3. This model is inspired from [154], see also [124, 125, 152]

The problem is then much more complicated because the system becomes hybrid: at fixed times a vehicle changes lane with threshold effects. And not only do the numbers of vehicles N_j change discretely but the acceleration of a vehicle behind a lane-changing vehicle is discontinuous at that time. The control can also be more complex since we can choose times when the autonomous vehicle changes lane. Even if we are not (yet) able to provide a mathematical answer to this problem, we can still try to modify our control heuristically and observe the effects in numerical simulations.

Modification of the control The most problematic point with disturbances that are not small is that there is nothing in this case that prevents the autonomous vehicle following the control law (8.2.13) from colliding with the vehicle in front of it. To avoid this we make two modifications to the control

— Quasi-stationary target state. Instead of directly stabilizing the target state $(\bar{v}, \bar{h})$, we start by stabilizing a slower steady state $(d, \bar{v}_d)$ with $d < \bar{h}$ (hence $\bar{v}_d < \bar{v}$) and we slowly increase the target velocity $\bar{v}_d$ until reaching $\bar{v}$. The control law becomes

$$\begin{aligned} u(t) &= k(\bar{v}_d(t) - v_{N+1}) + Z, \\ \dot{Z} &= (\bar{v}_d(t) - v_{N+1}), \end{aligned} \tag{8.2.34}$$

where

$$\begin{cases} v_d(t) = v_{\min} + (\bar{v} - v_{\min})t/\bar{t}, & \text{for } t \in [0, \bar{t}] \\ v_d(t) = \bar{v}, & \text{for } t \geq \bar{t}. \end{cases} \tag{8.2.35}$$

— Safety mechanism. When the autonomous vehicle is too close to the vehicle in front of it, the control law is changed to

$$u(t) = -k(v_{N+1} - \min(v_l, \bar{v}_d(t))). \tag{8.2.36}$$

Other variations [4] of safety mechanisms have been used for example in [152], but we will not detail them here. In order to take into account the different lanes, we also add a lateral control, which consists in making the autonomous vehicle change lane according to the state of the system. This lateral control works as follows: let Δt_1 and Δt_2 be time parameters to be chosen; $x_{i,j}$ and $v_{i,j}$ the position and speed of the i-th vehicle in the j lane; $j_0 \in \{1, ...J\}$ the lane of the autonomous vehicle at time t^-; i_0 its number of vehicles and t_0 the last time it changed lane (0 if it never changed lane), the autonomous vehicle changes lane if and only if

— $t > \Delta t_2 + t_0$

— $t > \Delta t_1$ and there exists $j \in \{1, ..., J\} \setminus \{j_0\}$ such that

$$\begin{aligned} &\int_{t-\Delta t_1}^{t} \frac{1}{N_j} \sum_{i=1}^{N_j} v_{i,j}^2(s) - \frac{1}{N_j^2} \left(\sum_{i=1}^{N_j} v_{i,j}(s) \right)^2 ds \\ &\qquad > c_1 + \int_{t-\Delta t_1}^{t} \frac{1}{N_{j_0}} \sum_{i=1}^{N_{j_0}} v_{i,j_0}^2(s) - \frac{1}{N_{j_0}^2} \left(\sum_{i=1}^{N_{j_0}} v_{i,j_0}(s) \right)^2 ds; \end{aligned} \tag{8.2.37}$$

If $j > j_0$ then the target lane will be $j_1 = j + 1$ and if $j < j_0$ it will be $j_1 = j - 1$.

— the condition (8.2.33) is satisfied with $i = i_0$, $j = j_0$ and $j_1 \in \{j - 1, j + 1\}$.

In less mathematical terms, this means that the autonomous vehicle changes lane if and only if:

— The autonomous vehicle has not changed lanes in the last Δt_2 seconds.

4. sometimes more effective

— The average speed variance in another lane is higher than the average speed variance in this lane by some threshold $c_1/\Delta t_1$ (the average is performed on the vehicles and the last Δt_1 seconds).

— The safety conditions are satisfied.

Numerical simulations We consider a circular road with 1 to 3 lanes, and the inner most lane has length $260m$, to correspond to the real experiment conducted in [226]. When there are three lanes, outermost lane has length $L_1 = 298m$. For the parameters of the model, we will consider the following values: $a = 20$, $b = 0.5$ which were obtained in [202] from the calibration to real data. We consider the usual V function for the Bando model [17]

$$V(h) = V_{\max} \frac{\tanh(\frac{h-l_v}{d_0} - 2) + \tanh(2)}{1 + \tanh(2)}, \tag{8.2.38}$$

where h is the headway between two vehicles, l_v is the length of a vehicle and $d_0 = 2.5$ is a characteristic length. For each lane j, the steady state we are trying to reach is $\bar{h}^j = (L^j/N^j)$, $\bar{v} = V(L^j/N^j)$, where L^j is the length of lane j and N^j is the total number of vehicule in the lane j.

On Figure 8.1 we show an example of numerical simulation. We represent the speed variance in each lane and the number of vehicles as a function of time. The simulation starts with 25 vehicles in each lane and lasts 1500s. At the beginning no vehicle is controlled, and at $t = 750s$ one of the vehicles of the middle lane becomes an autonomous vehicle and the control (8.2.34) with the lateral control described above is activated. We see that with these simple modifications, the control seems to be able to stabilize the system: shortly after $750s$ the variance of the speed of the three lanes decreases sharply, and the number of vehicles stabilizes.

To confirm this experiment we performed 50 simulations with randomly perturbed initial conditions. The averaged results are presented in Table 8.1. Additional numerical simulations can be found in [138, Section 5], where we can see for instance that the modified controller perfectly stabilize the system even for relatively large initial perturbations. These results are encouraging regarding the ability of this control to ensure a semi-global stabilization and to stabilize the hybrid multi-lane system.

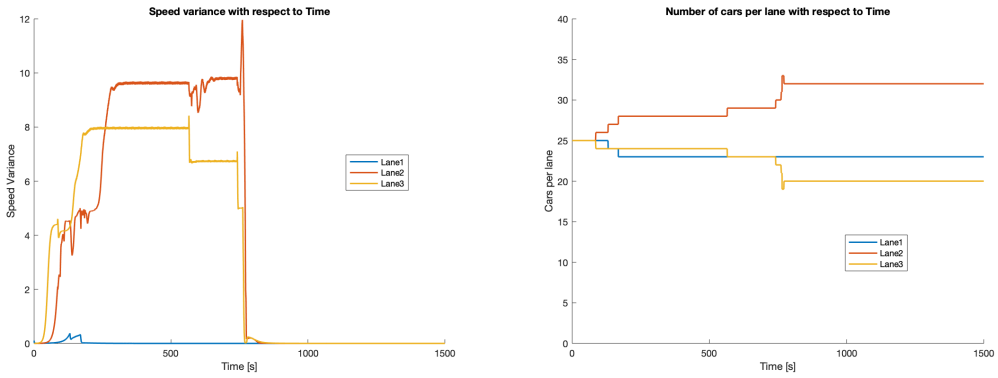


Figure 8.1 – **Left:** Speed variance with respect to time in a three lanes ring-road. **Right:** Number of vehicles per lane with respect to time. Control starts at $t = 750s$.

Time	$700s$ (before control)	$1500s$ (after control)
Speed variance lane 1 $(m^2.s^{-2})$	2.88	0.55
Speed variance lane 2 $(m^2.s^{-2})$	3.27	0.03
Speed variance lane 3 $(m^2.s^{-2})$	4.59	0.14
Total speed variance $(m^2.s^{-2})$	**3.58**	**0.24**
Number of lane-change per minute	**1.64**	**0.34**

Table 8.1 – Speed variance, average speed, energy consumption per distance travelled and number of lane-changes per minute before activation of the control ($t = 700s$) and after ($t = 1500s$). Control is activated at $t = 750s$ and all quantities are averaged on 50 simulations. At $t = 0$, each lane has 25 vehicles and the AV is in the middle lane.

8.3 Open-questions

Many related problems remain open:

— Is is possible to show theoretically the effectiveness of the controller (8.2.34) in the multilanes framework ? The difficulty is that the system is hybrid and even the notion of existence of solutions become much more complicated a priori. However, in this case there is a cooling time Δt before two lane changes by the same vehicle. This means that there cannot be an accumulation point of switching times and the solution can be seen as a piecewise continuous solution of a regular system. This is likely to simplify greatly the analysis and gives a hope that some results could be shown.

— Is it possible to extend the theoretical result to a global stability, or at least a semi-global stability (with a reasonable basin of attraction) ? In particular, can we show that the control described in Section 8.2.2.2 achieves such a semi-global stability and, if so, with which bound on the basin of attraction ?

— Related to this last question, does there exist a semi-global Lyapunov function for the (nonlinear) Bando-Follow the leader system ? Since this system is in cascade, one could start by studying the interaction between two cars only.

Chapter 9

Control of traffic flow: macroscopic approach

A second way to model traffic is to consider the traffic as a whole, represented at each point by a density ρ and a velocity v. The traffic dynamics then takes the form of one or more partial differential equations on ρ and v (or equivalent variables) and the resulting system is a hyperbolic nonlinear system of conservation laws or balance laws.
Like many nonlinear hyperbolic systems, the traffic system has an interesting feature: it can naturally give rise to discontinuous solutions, even when starting from a regular initial condition [78]. This phenomenon can be observed very well on the Burgers equation [170], where there is no continuous solution at any time for a certain initial condition, as well as on the LWR equation that we will see just after. These discontinuities, called shocks, then propagate in the system. An important particular case is when the shock is associated with a propagation velocity of the system that changes sign, going for example from positive to negative or the opposite. This example will be found in Section 9.1.

Shocks make the analysis of hyperbolic systems complicated, that is why, when we talk about stabilization of hyperbolic systems [1], it is common to consider only strong solutions without shock. To do this, we study a problem where we start from a regular initial condition close to a regular steady state and where the control ensures that the solution always remains regular and close to this steady state. This avoids difficulties due to shocks in the analysis. For many hyperbolic systems this approach can be considered physically relevant. However, for a traffic system, ignoring shocks means ignoring the traffic jams that form, i.e. ignoring the essence of the problem. Therefore, it can rarely be ignored. The existence and characterization of non-regular solutions of hyperbolic systems is something that has been studied a lot, in particular in the case of systems of conservation equations. One can refer for that to the discussion in Section 1.3.3.

In our case we try to stabilize road traffic with some autonomous vehicles that constitute the control. The question then arises as to how to model these autonomous vehicles and their interaction with the rest of the traffic. We consider that the autonomous vehicles are discrete and have a dynamic given by an ODE that depends on the traffic state where they are located. Their action on the rest of the traffic will be given by a steric hindrance factor: if an autonomous vehicle is slower than the rest of the traffic, the number of cars that can pass at this location is lower than if it was going at the same speed as the rest of the traffic. The traffic flow is thus locally reduced. This leads to an interesting ODE/PDE system that couples two scales: the macroscopic scale that models the traffic, and the microscopic scale that models the AVs.
Independently of its practical interest, this system has a large mathematical interest: surprisingly at first sight, the entropic BV solutions, which usually represent the physical solutions of hyperbolic systems, are not the physically relevant solutions of this system and there could be non-classical shocks appearing. This can be understood heuristically: with entropic solutions, a point element cannot have a macroscopic influence on the system, while in reality a vehicle that decides to brake suddenly on a road would have an influence

1. especially by the boundaries

on a macroscopic scale. From an analysis point of view this means that most tools for studying BV entropic solutions are no longer accessible, and this renders the analysis harder as we will see in Section 9.2. From a control point of view, however, this is precisely the reason why our control approach might work: we want a control that act at a microscopic scale to have a macroscopic impact on the system. And this would not be possible if the relevant solutions were BV entropic. We will study the existence of solutions to this system in Section 9.2.

9.1 The Lighthill-Whitham-Richard equation: a first model

First introduced in [181, 215] in 1955 and 1956, this equation describes the traffic by a density ρ and a velocity $v = V(\rho)$ taken as a function of ρ. The function ρ satisfies a conservation of mass, which gives the LWR equation

$$\partial_t \rho + \partial_x(\rho V(\rho)) = 0. \tag{9.1.1}$$

In this model, V is further assumed to be a C^2 decreasing function of ρ. This seems reasonable and corresponds to the intuitive impression that the more cars on the road, the lower the velocity of the overall traffic. V is often assumed to satisfy $2V'(\rho) + \rho V''(\rho) \leq 0$ which implies that $\rho \to \rho V(\rho)$ is concave. Besides it is also often assumed the existence of $\rho_{\max} > 0$ such that $V(\rho_{\max}) = 0$ which means that the road is so packed that the cars have to stop.

As mentioned previously, the solutions of (9.1.1) could present a discontinuity in finite-time even if the initial condition is smooth. On the other hand, just like for Burger equation, it is a well-known fact that one cannot ensure the well-posedness in $L^\infty(\mathbb{R})$ or $L^1(\mathbb{R})$ either, because the uniqueness would fail [170]. To recover the well-posedness one usually introduce the notion of entropic solutions [170, Section 3] (see also [78]) by requiring additionnally an *entropy condition*, at a discontinuity located in $x_s(t)$

$$f'(x_s^-(t)) \geq \dot{x}_s \geq f'(x_s^+(t)), \tag{9.1.2}$$

where f is the flow of the equation, x_s^+ (resp. x_s^-) refers to the right (resp. left) limit in x_s, and $s = \dot{x}_s(t)$ is the speed of the discontinuity. This simply means that the propagation speed of the wave at the left of the discontinuity is faster than the speed of the discontinuity, which is itself faster than the propagation speed of the wave at the right of the discontinuity. In other words, discontinuities can only merge together. This ensures that information cannot spontaneously appear from a single point. In the case of System (9.1.1), this entropy condition is simply

$$V(\rho(t,x_s^-)) + \rho(t,x_s^-)V'(\rho(t,x_s^-)) \geq \dot{x}_s \geq V(\rho(t,x_s^+)) + \rho(t,x_s^+)V'(\rho(t,x_s^+)) \tag{9.1.3}$$

Besides, we can deduce the dynamics of a shock from the equation (9.1.1), this leads to the Rankine-Hugoniot conditions which general form (for a scalar equation) is

$$\dot{x}_s = \frac{f(\rho(t,x_s^+(t)) - f(\rho(t,x_s^-(t)))}{\rho(t,x_s^+(t)) - \rho(t,x_s^-(t))}, \tag{9.1.4}$$

where f is again the flow of the equation. Here, it becomes

$$\dot{x}_s = \frac{\rho(t,x_s^+)V(\rho(t,x_s^+)) - \rho(t,x_s^-)V(\rho(t,x_s^-))}{\rho(t,x_s^+) - \rho(t,x_s^-)}. \tag{9.1.5}$$

Such a discontinuity curve x_s is called a *shock*, and an *entropic shock* or *classical shock* when it also satisfies the entropy condition (9.1.2). Analogous conditions exists when considering not only a single equation of conservation but a system of conservation laws (see for instance [33, 78]).

As (9.1.1) is a scalar system there is only one quantity ρ propagating and the propagation speed is given by $\lambda(\rho) = \rho V'(\rho) + V(\rho)$. By the concavity assumption $\rho \to \rho V'(\rho) + V(\rho)$ is decreasing. Since $V(\rho_{\max}) < 0$ and $V' < 0$ we can deduce the existence of a critical density $\rho_c \in (0, \rho_{\max})$ such that

$$\lambda(\rho_c) = \rho V'(\rho_c) + V(\rho_c) = 0, \tag{9.1.6}$$

and for any $\rho \in [0, \rho_c)$ the propagation speed is positive and the system is said to be in *free-flow*, while for any $\rho \in (\rho_c, \rho_{\max}]$ the propagation speed is negative and the system is said to be in *congested regime*.

Finally, let us remark that there can only be two types of steady-states to (9.1.1): constant steady states, corresponding to a steady flow, or shock steady-states. Indeed, $\partial_x(\rho V(\rho)) \equiv 0$ implies that there exists a constant C such that $\rho V(\rho) = C$. Since $V(\rho) + \rho V'(\rho)$ is positive for $\rho = 0$, and $\rho \to \rho V(\rho)$ is concave by assumption, this equation $\rho V(\rho) = C$ has only 0, 1 or 2 solutions for a given C. This implies that ρ is either constant or switches discontinuously between two values. In this case, ρ is constant between two discontinuity points and the discontinuity curves satisfy $\dot{x}_s \equiv 0$. Among the shock steady-states, those which satisfy the entropy conditions corresponds to the transition between a free-flow traffic and a congested traffic and, therefore, only have one shock.

9.2 Existence of solutions to the Generalized Aw-Rascle-Zhang equations

This section present the results of [137], a work done with Thibault Liard, Francesca Marcellini and Benedetto Piccoli.

In this Section we consider a second order model of traffic flow, namely the *Generalized Aw-Rascle-Zhang equations (GARZ)* [103]

$$\begin{aligned} &\partial_t \rho + \partial_x \left(\rho\, V(\rho, w)\right) = 0, \\ &\partial_t (\rho w) + \partial_x \left(\rho w\, V(\rho, w)\right) = 0. \end{aligned} \tag{9.2.1}$$

As their name indicates, these equations are a generalization of the ARZ model [2], introduced in [12, 259], which gave a large impulse to second order models. Compared to the LWR equations, the velocity $v = V(\rho, w)$ is now a function of ρ and another parameter w defining a driving behaviour which correspond to the drivers' velocity on an empty road. We assume the following

- The function $(\rho, w) \mapsto V(\rho, w)$ is C^2 $([0, \rho_{max}] \times [w_{\min}, w_{\max}])$.
- The vehicles never drive backwards on the road, namely $V(\rho, w) \geq 0$ for any $(\rho, w) \in [0, \rho_{max}] \times [w_{\min}, w_{\max}]$,
- $V(0, w) = w$ for any $w \in [w_{\min}, w_{\max}]$, i.e. w is each driver's speed on an empty road.
- $\frac{\partial^2(\rho V(\rho,w))}{\partial \rho^2} < 0$ for any $(\rho, w) \in [0, \rho_{max}] \times [w_{\min}, w_{\max}]$ and $\frac{\partial V}{\partial \rho}(\rho, w) < 0$.
- $\frac{\partial V}{\partial w}(\rho, w) > 0$, for any $(\rho, w) \in [0, \rho_{max}] \times [w_{\min}, w_{\max}]$ which simply means that if a driver goes faster than another on an empty road this driver will also go faster than the other on a non-empty road.
- $V(\rho_{\max}, w) = 0$ for any $w \times [w_{\min}, w_{\max}]$, at maximal density $\rho_{\max}$, the speed of each driver is zero. The density $\rho_{\max}$ therefore corresponds to the density at which the traffic is completely packed.

This system (9.2.1) is hyperbolic and has two propagation speeds. Away from the vacuum (i.e. away from $\rho = 0$), these propagation speeds are given by

$$\begin{aligned} \lambda_1(\rho, w) &= V(\rho, w) + \rho \frac{\partial V}{\partial \rho}(\rho, w) \\ \lambda_2(\rho, w) &= V(\rho, w) \end{aligned} \tag{9.2.2}$$

and are associated to the eigenvectors $r_1 = (1, 0)$ and $r_2 = (-\partial_w V(\rho, w), \partial_\rho V(\rho, w))$. One can check that $\nabla\lambda_2(\rho, w).r_2(\rho, w) = 0$ which means that the second propagation speed is said to be *linearly degenerate*, while the first one satisfies $\nabla\lambda_1(\rho, w).r_1(\rho, w) \neq 0$ and hence is said to be *genuinely nonlinear*.

We now couple this traffic system with an AV described by its location $y(t)$, and the following dynamics

$$\dot{y}(t) = \min(V(\rho(t, y(t)^+), w(t, y(t)^+)), V_b) \tag{9.2.3}$$

2. see also the Collapsed-GARZ equations [104]

where V_b is the control velocity that we assume constant in this section, while the min comes from the fact that the AV cannot go faster than the traffic velocity at its location, otherwise it would crash in a leading vehicle. In turns the AV has an effect on the traffic given by the following flow condition:

$$\rho\left(t, y(t)\right)\left(V(\rho(t, y(t)), w(t, y(t))) - \dot{y}(t)\right) \leq \alpha F(\dot{y}), \tag{9.2.4}$$

where $\alpha \in (0, 1)$ and F denotes

$$F(\dot{y}) = \max_{x\in[0,\rho_{\max}], w\in[w_{\min}, w_{\max}]} \left(\rho(V(x, w) - \dot{y})\right), \tag{9.2.5}$$

The condition (9.2.4) only applies when $y(t) \neq V(\rho(t, y(t)), w(t, y(t)))$ which means that the AV is slower than the bulk traffic. It represents the fact that the AV is a local obstacle on the road when it is going slower than the rest of the traffic and the associated steric hindrance is representend by the constant $\alpha < 1$. The smaller α, the higher the hindrance. The maximum flow that can pass locally at point $y(t)$ is then what would be the maximum possible flow on a road of width α rather than 1 and this is what the right-hand side of (9.2.4) represents.

Overall, the total system is (9.2.1), (9.2.3)–(9.2.4). Let us pause a moment to note a potentially surprising fact: it was shown that the equations (9.2.1) alone are well-posed in the framework of BV entropic solutions (see [36] for instance), and (9.2.3) is well-posed on its own in the space of absolutely continuous solutions as long as ρ, w are fixed and belong to L^1_{loc}. So it would look like (9.2.1), (9.2.3) is well-posed which would suggest that (9.2.4) is either redundant or make the system ill-posed. The answer to this paradox is that (9.2.1), (9.2.3) is indeed well-posed for solutions (ρ, w) that are BV entropic, but those solutions are not the relevant physical solutions (in particular the AV would not have any impact on the traffic). This means that in this system the relevant solutions are not necessarily entropic, especially at the location of the AV. This explains why we also have the condition (9.2.4): we expect it to replace in some sense the entropy condition at the location of the AV.

In the following, we are going to show that for any initial condition $(\rho_0, w_0, y^0) \in BV(\mathbb{R}; [0, \rho_{\max}] \times [w_{\min}, w_{\max}]) \times \mathbb{R}$ there exists a solution $(\rho, w, y) \in L^\infty([0, +\infty); BV(\mathbb{R}; [0, \rho_{\max}] \times [w_{\min}, w_{\max}])) \times W^{1,1}_{loc}([0, +\infty); \mathbb{R})$ to the system (9.2.1), (9.2.3)–(9.2.4) that is entropic on $(-\infty, y(t))$ and $(y(t), +\infty)$. The rigorous definition and statement are given below. Before this we introduce a last notation: since $\partial_\rho V(\rho, w) < 0$ on $[0, \rho_{\max}] \times [w_{\min}, w_{\max}]$, there exists a function R such that for any $\rho \in [0, \rho_{\max}]$ and $w \in [w_{\min}, w_{\max}]$,

$$\rho = R(V(\rho, w), w), \tag{9.2.6}$$

which means that there exists a unique $\rho \in [0, \rho_{\max}]$ associated to a speed v and a parameter w with $v \in [0, w]$ and $w \in [w_{\min}, w_{\max}]$.

The definition of a weak solution of (ρ, w, y) to the system (9.2.1), (9.2.3)–(9.2.4), entropic on $(-\infty, y(t))$ and $(y(t), +\infty)$ is given by

Definition 9.2.1. The couple

$$((\rho, w), y) \in C^0\left([0, +\infty); L^1_{\text{loc}}(\mathbb{R}; [0, \rho_{\max}] \times [w_{\min}, w_{\max}]\right) \times W^{1,1}_{\text{loc}}\left([0, +\infty); \mathbb{R}\right)$$

is a weak solution to the system (9.2.1), (9.2.3)–(9.2.4) if

1. the function (ρ, w) is a weak solution of (9.2.1), i.e for all $\varphi \in C^1_c(\mathbb{R}^2, \mathbb{R})$;

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}} \rho\left[\partial_t\varphi + V(\rho, w)\partial_x\varphi\right]\begin{pmatrix}1\\ w\end{pmatrix} dx\, dt + \int_{\mathbb{R}} \rho_0 \begin{pmatrix}1\\ w_0\end{pmatrix}\varphi(0, x) = 0 \tag{9.2.7}$$

2. The function ρ is an entropy admissible solution of (9.2.1), i.e for every $k \in [0, V(0, w_{\max})]$, for all

$\varphi \in C_c^1(\mathbb{R}^2, \mathbb{R}_+)$, it holds

$$\begin{aligned}
&\int_{\mathbb{R}_+}\int_{\mathbb{R}} \mathcal{E}_k(v(t,x),w(t,x))\partial_t\varphi + Q_k(v(t,x),w(t,x))\partial_x\varphi dx\, dt \\
&+ \int_{\mathbb{R}} \mathcal{E}_k(v_0,w_0)\varphi(0,x)\, dx \\
&+ \int_{\mathbb{R}_+} R(v(t,y(t)),w(t,y(t)))(v(t,y(t))-\dot{y})\left[\frac{k-\dot{y}}{\alpha F(\dot{y})} - \frac{1}{R(k,w(t,y(t)))}\right]^+ \varphi(t,y(t))dt \geq 0,
\end{aligned} \tag{9.2.8}$$

where we denote $(v,w) = (V(\rho,w),w)$, R is defined by (9.2.6) (extended by $R(k,w)=0$ if $k \geq w$), and $(\mathcal{E}_k, \mathcal{Q}_k)$ is the entropy pair defined by

$$\mathcal{E}_k(v,w) = \begin{cases} 0 \text{ if } v \leq k \\ 1 - \frac{R(v,w)}{R(k,w)}, \quad \text{if } v > k, \end{cases} \tag{9.2.9}$$

$$\mathcal{Q}_k(v,w) = \begin{cases} 0 \text{ if } v \leq k \\ k - \frac{R(v,w)v}{R(k,w)}, \quad \text{if } v > k. \end{cases} \tag{9.2.10}$$

3. For every $t \in \mathbb{R}^+$,
$$y(t) = y_0 + \int_0^t \min\left(V_b, V(\rho(t,y(t)^+), w(t,y(t)^+))\right) ds\,. \tag{9.2.11}$$

4. the constraint in (9.2.4) is satisfied, namely for a.e. $t \in \mathbb{R}^+$
$$\lim_{x \to y(t)\pm} \rho\,(t,x)\,(V(\rho(t,x),w(t,x)) - \dot{y}(t)) - \alpha F(\dot{y}) \leq 0\,; \tag{9.2.12}$$

Remark 9.2.1 (Entropy pairs and non-classical shock)**.** *The entropy pairs $(\mathcal{E}_k, \mathcal{Q}_k)$ are the same as in [9]. Here the term of the third line of (9.2.8) differs from the usual entropy condition to compensate for the potential non-classical shocks that would occur at $y(t)$. Note that all other non-classical shocks are prohibited with this condition and therefore the solution is entropic in a classical sense on $(-\infty, y(t))$ and $(y(t), +\infty)$. Besides, one can show that the last integral term of (9.2.8) also ensures that any solution maximize the flux when non-classical shock occurs, i.e. condition (9.2.12) becomes an equality. This is similar to the result of [9, Section 3].*

With this definition we can now state our main result

Theorem 9.2.1. *Let $(\rho_0, w_0, y_0) \in BV(\mathbb{R}; [0,\rho_{\max}] \times [w_{\min}, w_{\max}]) \times \mathbb{R}$, and assume that $V_b < w_{\min}$. Then the Cauchy problem (9.2.1), (9.2.3)–(9.2.4), $(\rho(0,\cdot), w(0,\cdot), y(0)) = (\rho_0, w_0, y_0)$ admits a solution $(\rho, w, y) \in C^0([0,+\infty); BV(\mathbb{R}; [0,\rho_{\max}] \times [w_{\min}, w_{\max}])) \times W^{1,1}_{loc}([0,+\infty); \mathbb{R})$ in the sense of Definition 9.2.1.*

To show this result, we are going to construct a sequence of approximate solutions (ρ^n, w^n, y^n) using a wave front-tracking algorithm and show that this sequence of solution converges to a solution (ρ, w, y). We describe briefly the waves induced by GARZ equations and the principle of the wave front tracking algorithm in Section 9.2.1. Then in Section 9.2.2 we give an idea of the proof of Theorem 9.2.1.

9.2.1 Riemann problem and wave front tracking algorithm

Our goal is to create a solution to (9.2.1), (9.2.3)–(9.2.4) for an initial condition $(\rho_0, w_0, y_0) \in BV(\mathbb{R}; [0,\rho_{\max}] \times [w_{\min}, w_{\max}]) \times \mathbb{R}$. As (ρ_0, w_0) is a BV function, it can be approximated by a sequence of piecewise constant functions (ρ_0^n, w_0^n) which converges to (ρ_0, w_0) in L^1_{loc}. The strategy is the following: we would like to be able to

— create a solution (ρ^n, w^n, y^n) associated to the piecewise constant initial condition (ρ_0^n, w_0^n, y_0),
— then show that this solution converges in $L^1_{loc}(\mathbb{R}_+ \times \mathbb{R})$ to some (ρ, w, y) (up to a subsequence) when $n \to +\infty$, which belongs to $C^0([0,+\infty); BV(\mathbb{R}; [0,\rho_{\max}] \times [w_{\min}, w_{\max}])) \times W^{1,1}_{loc}([0,+\infty); \mathbb{R})$.

— and finally show that this limit satisfies the system (9.2.1), (9.2.3)–(9.2.4) in a weak sense (see Definition 9.2.1).

Constructing a solution (ρ^n, w^n, y^n) associated to a piecewise constant initial condition (ρ_0^n, w_0^n, y_0) requires us to define and solve a so-called *Riemann problem* for the system (9.2.1), (9.2.3)–(9.2.4). This is the object of the following subsection.

9.2.1.1 A constrained Riemann problem

Let us start by introducing what is usually called a *Riemann problem* with initial data for the equations (9.2.1). The Riemann problem consists in solving the equation (9.2.1) for an initial data of the following form

$$\rho(0,x) = \begin{cases} \rho_l & \text{if } x < 0, \\ \rho_r & \text{if } x > 0, \end{cases} \quad \text{and} \quad w(0,x) = \begin{cases} w_l & \text{if } x < 0, \\ w_r & \text{if } x > 0, \end{cases} \tag{9.2.13}$$

for any states $U_l := (\rho_l, w_l) \in [0, \rho_{\max}] \times [w_{\min}, w_{\max}]$ and $U_r := (\rho_r, w_r) \in [0, \rho_{\max}] \times [w_{\min}, w_{\max}]$. Observe that the constant parts of the initial condition obviously satisfy the equations (9.2.1), so what matters is how the discontinuity will evolve and propagate over time. From the Rankine-Hugoniot conditions (whose scalar case is recalled in (9.1.4)), if there is a discontinuity located at $x(t)$, then

$$\begin{aligned} &\dot{x}(t)\,[\rho]_{x(t)^-}^{x(t)^+} - [\rho V(\rho, w)]_{x(t)^-}^{x(t)^+} = 0, \\ &\dot{x}(t)\,[\rho w]_{x(t)^-}^{x(t)^+} - [\rho V(\rho, w) w]_{x(t)^-}^{x(t)^+} = 0, \end{aligned} \tag{9.2.14}$$

where we denoted $[f]_{x^-}^{x^+} = f(t, x^+) - f(t, x^-)$. In the following we will denote $\rho(t, x(t)) = \rho(x(t))$ and $w(t, x(t)) = w(x(t))$ for clarity. Using the first equation in the second one, and the fact that $[fg]_{x^-}^{x^+} = [f]_{x^-}^{x^+} g(a^+) + f(a^-)[g]_{x^-}^{x^+}$, we deduce that (see [137, Appendix B])

$$\text{either } w(x(t)^+) = w(x(t)^-) \text{ or } V(\rho(x(t)^+), w(x(t)^+)) = V(\rho(x(t)^-), w(x(t)^-)) \text{ or } \rho(x(t)^+) = \rho(x(t)^-) = 0, \tag{9.2.15}$$

and in the second case $\dot{x}(s) = V(\rho(x(t)), w(x(t)))$. This means that there are only a few possible waves:

— If $w(x(t)^-) = w(x(t)^-)$, then the discontinuity travels at speed

$$\dot{x}(t) = \frac{\rho(x(t)^+) V(\rho(x(t)^+), w(x(t)^+) - \rho(x(t)^-) V(\rho(x(t)^-), w(x(t)^-))}{\rho(x(t)^+) - \rho(x(t)^-)}. \tag{9.2.16}$$

Note that this is exactly the shocks of the scalar LWR equation (9.1.1) where the velocity $V(\rho)$ is replaced by $V(\rho, w^+) = V(\rho, w^-)$. Therefore these shocks are called a *1-wave*.

— If $V(\rho(x(t)^-), w(x(t)^-) = V(\rho(x(t)^+), w(x(t)^+)$, then the discontinuity travels at speed

$$\dot{x}(t) = V(\rho(x(t)^-), w(x(t)^-). \tag{9.2.17}$$

This is the type of discontinuities that do not exist in the scalar LWR system and really rely on the fact that this is a second order system. Therefore these shocks are called *2-waves*.

— If $\rho(x(t)^+) = \rho(x(t)^-) = 0$, the shock connects two state that have a zero density, therefore this is called a *V-wave* (which stands for vacuum wave).

These waves will be the constitutive elements of the solution to the Riemann problem. This solution is given as follows:

— If $w_l = w_r$ and $\rho_+ > \rho_-$ then the discontinuity corresponds to a classical shock and at anytime U_l is connected to U_r by a so called *1-wave* travelling at speed

$$\sigma(\rho_l, \rho_r) = \frac{\rho^l V(\rho^l, w^l) - \rho^r V(\rho^r, w^r)}{\rho^l - \rho^r}. \tag{9.2.18}$$

The solution is then simply

$$\begin{aligned}(\rho, w)(t, x) &= (\rho_l, w_l) \quad \text{if } x < \sigma(\rho_l, \rho_r)t,\\ (\rho, w)(t, x) &= (\rho_r, w_r) \quad \text{if } x > \sigma(\rho_l, \rho_r)t.\end{aligned} \tag{9.2.19}$$

— If $w_l = w_r =: w$ and $\rho_+ > \rho_-$ then the discontinuity corresponds to a non-classical shock. This situation typically appears when the Riemann problem approximates a function that moves continuously from ρ_- to ρ_+. In this case we would like to avoid having too large non-classical shocks so instead of generating a single wave, we approximate this continuous behavior by generating a fan of k waves. We define intermediate states $U_i = (\rho_i, w)$ such that U_l is connected to U_1 by a *1-wave*, U_2 is a connected to U_3 by a *1-wave* and so on up to U_{k-1} that is connected to U_r by a *1-wave*. This leads to a fan of *1-wave* emerging from the origin. The number of waves k and the intermediate states are chosen such that the amplitude between two states are lower than a bound to be specified. The solution is then

$$\begin{aligned}(\rho, w)(t, x) &= (\rho_l, w_l) \quad \text{if } x < \sigma(\rho_l, \rho_1)t,\\ (\rho, w)(t, x) &= (\rho_l, w_l) \quad \text{if } \sigma(\rho_i, \rho_{i+1})t < x < \sigma(\rho_{i+1}, \rho_{i+2})t,\\ (\rho, w)(t, x) &= (\rho_l, w_l) \quad \text{if } \sigma(\rho_{k-2}, \rho_{k-1})t < x < \sigma(\rho_{k-1}, \rho_r)t,\\ (\rho, w)(t, x) &= (\rho_r, w_r) \quad \text{if } x > \sigma(\rho_{k-1}, \rho_r)t.\end{aligned} \tag{9.2.20}$$

We call each of these k waves a rarefaction shock (often called rarefaction front) and the ensemble of k waves is called a rarefaction fan.

— If $V(\rho_l, w_l) = V(\rho_r, w_r)$ then at anytime U_l is connected to U_r by a so called *2-wave* travelling at speed $V(\rho_l, w_l)$. The solution is simply

$$\begin{aligned}(\rho, w)(t, x) &= (\rho_l, w_l) \quad \text{if } x < V(\rho_l, w_l)t,\\ (\rho, w)(t, x) &= (\rho_r, w_r) \quad \text{if } x > V(\rho_l, w_l)t.\end{aligned} \tag{9.2.21}$$

— if $\rho_l = \rho_r = 0$ then U_l is connected to U_r by a so called *V-wave* (physically this is an empty wave since the density is equal to 0). The speed of the V-wave satisfies $s = w_r$. The solution is simply

$$\begin{aligned}(\rho, w)(t, x) &= (0, w_l) \quad \text{if } x < st,\\ (\rho, w)(t, x) &= (0, w_r) \quad \text{if } x > st.\end{aligned} \tag{9.2.22}$$

— In other cases, U_l cannot be connected to U_r simply by a single wave. So at any time $t > 0$ there is a least an intermediate state U_m between U_l and U_r. There are two subcases:
  — If $w_l > w_r$ or $\rho_r \geq R(V(\rho_l, w_l), w_r)$, where R is defined in (9.2.6), then

$$U_m = (R(V(\rho_r, w_r), w_l), w_l), \tag{9.2.23}$$

  and $R(V(\rho_r, w_r) \in [0, \rho_{\max}]$. Thus, U_l is connected to U_m by a 1-wave and U_m is connected to U_r by a 2-wave.
  — If $w_l < w_r$ and $\rho_r < R(V(\rho_l, w_l), w_r)$ it is not possible to have only a single intermediate state between U_l and U_r and there are in fact two intermediate states $U_{m,1} = (0, w_l)$ and $U_{m,2} = (0, V(\rho_r, w_r))$ between U_l and U_r. U_l is connected to $U_{m_1} := (0, w_l)$ by a 1-wave, U_{m_1} is connected to U_{m_2} by a V-wave and U_{m_2} is connected to U_r by a 2-wave.

One can look at [33, 177] for more details. We denote $\mathcal{RS}(U_l, U_r)$ this solution to the Riemann problem associated with initial state (U_l, U_r). We also denote $\mathcal{RS}_\rho$ and $\mathcal{RS}_w$ its components, such that $\mathcal{RS} = (\mathcal{RS}_\rho, \mathcal{RS}_w)$

Now, we would like to solve the Riemann problem associated to the full problem (9.2.1), (9.2.3)–(9.2.4). Outside of the location of the AV, this problem will be the same as the Riemann problem for GARZ equations (9.2.1) that we just described. Thus we focus on the case where the AV is located at the discontinuity and we consider an initial condition of the from (9.2.13) with $y_0 = 0$. The goal is now to find a solution associated to this initial condition that satisfies again (9.2.1) but also (9.2.3) and (9.2.4). This is called a *constrained*

Riemann problem. We denote $\mathcal{RS}_c(U_l, U_r)$ the solution of this constrained Riemann problem. We also denote by f the function

$$f : (\rho, w) \to \rho V(\rho, w). \tag{9.2.24}$$

Three cases are possible:

— If $V_b \geq V(\mathcal{RS}(U_l, U_r)(t, y(t)))$. In this case the traffic is too slow compared to the desired speed V_b of the AV. Hence, the AV has to adapt and drive at the same speed as the other cars in the flow. The solution of the constrained Riemann problem at the AV's location is given by

$$RS_c(U_l, U_r)(t, x) = RS(U_l, U_r)(t, x)y(t) = V(RS(U_l, U_r)(V_b))t \tag{9.2.25}$$

— If $f(\mathcal{RS}(U_l, U_r)(t, y(t))) \leq \alpha F(w_l) + V_b \mathcal{RS}_\rho((U_l, U_r)(t, y(t)))$ and $V_b < V(\mathcal{RS}(U_l, U_r)(t, y(t)))$. In this case the AV is going slower than the traffic but the traffic is not too congested and therefore the AV can be passed with no difficulty and its presence has not effect on the traffic. The solution of the constrained Riemann problem at the AV's location is given by

$$RS_c(U_l, U_r)(t, x) = RS(U_l, U_r)(t, x)y(t) = V_b t \tag{9.2.26}$$

— If $f(\mathcal{RS}(U_l, U_r)(t, y(t))) > \alpha F(w_l) + V_b \mathcal{RS}_\rho((U_l, U_r)(t, y(t)))$. In this case the AV is limiting the flow that can go through locally at its location $y(t)$. This results in a non-classical shock at the location of the AV given by

$$\begin{aligned} \mathcal{RS}_c(U_l, U_r)(t, x) &= \begin{cases} \mathcal{RS}\left(U_l, (\hat{\rho}(w_l), w_l)\right)(t, x) & \text{if} \quad x < y(t) \\ \mathcal{RS}\left((\check{\rho}(w_l), w_l), U_r\right)(t, x) & \text{if} \quad x > y(t), \end{cases} \\ y(t) &= V_b t. \end{aligned} \tag{9.2.27}$$

where $\check{\rho}(w)$ and $\hat{\rho}(w)$ are the two shock densities defined as the solutions of

$$\alpha F(\dot{y}) + V_b \rho = \rho V(\rho, w), \tag{9.2.28}$$

and such that $\hat{\rho}(w) > \check{\rho}(w)$. The fact that (9.2.28) has exactly two solutions comes from the concavity of the function $\rho \to \rho V(\rho, w)$ which vanishes in $\rho = 0$ and $\rho = \rho_{\max}$.

These densities $\check{\rho}(w)$ and $\hat{\rho}(w)$ also correspond to the nonclassical shock densities of the LWR model when w is a constant equal to $w_l = w_r$ [87, 88, 176, 177]. In this case, the left state U_l is connected by a classical shock to a state with density $\hat{\rho}$ which is connected by a nonclassical shock to a state with density $\check{\rho}$ at the location of the AV, which is itself connected to the right state U_r by a classical shock.

This constrained Riemann problem and its solution $\mathcal{RS}_c(U_l, U_r)$ is the stepping stone for the wave-front tracking algorithm that we now describe briefly.

9.2.1.2 Wave-front tracking algorithm

The wave-front tracking consists in constructing a solution to (9.2.1), (9.2.3)–(9.2.4) from the solution to the constrained Riemann Problem. The algorithm presented here is based on an algorithm introduced in [32] (see also [32, 33, 177], one can also look at the earlier algorithm proposed in [77] where both the initial condition and the flux while we only discretize the initial condition). We only present its principle, the rigorous formulation can be found in the previous references. We start by approximating (ρ_0, w_0) by a sequence of piecewise constant functions (ρ_0^n, w_0^n) which converges to (ρ_0, w_0) in BV. To do so we choose an approximation mesh that is roughly given [3] by $(2^{-n}\mathbb{N} \cap [0, 1])\rho_{\max}$ for ρ and $(2^{-n}\mathbb{N} \cap [0, 1])(w_{\max} - w_{\min}) + w_{\min}$ for w. As (ρ_0^n, w_0^n) has a finite number of discontinuities, we can denote its discontinuity points by

3. In fact one needs to change slightly the mesh and add potentially several points to take into account the non-classical shocks, more details can be found for instance in [177, Section 2.2]

$(x^0_{n,i})_{i\in\{1,...,N_n\}}$, where $N_n \in \mathbb{N}$ with $x^0_{n,1} < x^0_{n,2} < .... < x^0_{n,N_n}$ and where we also included $y(0)$ in the discontinuity points, even if the solution is continuous at $y(0)$. This means that there exists $i \in \{1, ..., N_n\}$ such that $y(0) = x^0_{i,n}$. Since (ρ^n_0, w^n_0) is piecewise constant, constructing the solution only amounts to seeing how the discontinuities propagate, at least for sufficiently small times. At each discontinuity we can propagate a wave following the solution of the Riemann problem described in the previous section. It could happen that the discontinuity corresponds to a non-classical shock, and this situation can occur even if (ρ, w) has no non-classical shocks. This is because (ρ^n_0, w^n_0) is piecewise constant but the original function (ρ_0, w_0) is not necessarily. This implies that the continuous changes of (ρ_0, w_0), like rarefaction waves, are translated in discontinuities in (ρ^n_0, w^n_0) and these discontinuities are not necessarily entropic. In this case we propagate a fan of k waves as described in the Riemann solver and we choose k such that the maximal amplitude of a non-classical shock in the fan is the smallest possible given the mesh (in our case of the order of $\rho_{\max}2^{-n}$). This rarefaction fan is the way continuous changes are represented in the approximate solution.
As long as the discontinuities propagate without interacting, the solution is simply the sum of N_n Riemann problems [4] and we can denote $(x_{n,i}(t))_{i\in\{1,...,N_n\}}$ the location of the discontinuities describing the solution of each Riemann problem and $y^n(t)$ associated. As N_n is finite and the propagation speed of each discontinuity is also finite, there exists a time $t_{n,1} \in \mathbb{R}^*_+ \cup \{+\infty\}$ such that for any $t \in [0, t_{1,n})$

$$x_{n,1}(t) < x_{n,2}(t) < ... < x_{n,N_n}(t), \tag{9.2.29}$$

and, if $t_{n,1} < +\infty$, there exists $i \in \{1, ..., N_n\}$ such that

$$x_{n,i}(t_{n,1}) = x_{n,i+1}(t_{n,1}) \tag{9.2.30}$$

At this point we have $\rho(x_{n,i}(t_{n,1})^-) = \rho^l_i$ and $\rho(x_{n,i}(t_{n,1})^+) = \rho^r_{i+1}$ so we have a new Riemann problem starting at $x_{n,i}(t_{n,1})$ with left state ρ^l_i and right state ρ^r_{i+1}. This can be done for any $i \in \{1, ..., N_n\}$ such that (9.2.30) holds. We have again a piecewise constant function (ρ^n, w^n) so we can perform the same algorithm with initial condition $(\rho^n(t_{n,1}, \cdot), w^n(t_{n,1}, \cdot))$ instead of (ρ^n_0, w^n_0). We denote again the (new) discontinuity points by $(x^0_{n,i})_{i\in\{1,...,N_n\}}$ and extend the solution (ρ^n, w^n, y^n) up to $t_{n,2}$ where two discontinuities interact again, and so on.

9.2.2 Ideas of the proof

In this section we give some ideas of the proof of Theorem 9.2.1. We start by showing the convergence of (ρ^n, w^n, y^n) up to a subsequence when $n \to +\infty$ to a function (ρ, w, y) which belongs to $L^\infty([0, +\infty); BV(\mathbb{R}; [0, \rho_{\max}] \times [w_{\min}, w_{\max}])) \times W^{1,1}_{loc}([0, +\infty); \mathbb{R})$, when $n \to +\infty$. Then we talk about how to show that this limit is a solution to (9.2.1), (9.2.3)–(9.2.4) in the sense of Definition 9.2.1, which is the main difficulty.

9.2.2.1 Convergence of (ρ^n, w^n, y^n) to (ρ, w, y)

We first show the following Lemma

Lemma 9.2.1. *Let $(\rho_0, w_0, y_0) \in BV(\mathbb{R}; [0, \rho_{\max}] \times [w_{\min}, w_{\max}]) \times \mathbb{R}$ and (ρ^n, w^n, y^n) be an approximate solution of (9.2.1) constructed by the wave-front tracking method described in Section 9.2.1, with initial conditions (ρ_0, w_0, y_0). Then, there exists $C > 0$ such that, for any $t \in \mathbb{R}_+$,*

$$TV(w^n(t, \cdot)) + TV(V(\rho^n(t, \cdot), w^n(t, \cdot)) \leq C, \tag{9.2.31}$$

where TV refers to the total variation.

4. More precisely $N_n - 1$ Riemann problems and 1 constrained Riemann problem: even if $y(0)$ is not located on a discontinuity we included it in the $x^0_{n,i}$ and we consider it as the solution of a constrained Riemann problem with $\rho^l = \rho^r$ and $w^l = w^r$. The function $y^n(t)$ is simply affine in this case.

This is proved by introducing and studying the following function $\Gamma(t)$

$$\Gamma(t) = TV(w^n(t,\cdot)) + TV(V(\rho^n(t,\cdot); w^n(t,\cdot))) + \gamma(t) + C_1 TV(w^n(\cdot, y^n(\cdot)), [t, +\infty)), \tag{9.2.32}$$

where γ is given by

$$\gamma(t) = \begin{cases} -2|\hat{v}(w^n(t, y(t)-)) - \check{v}(w^n(t, y(t)))|, & \text{if } \begin{cases} w^n(t, y_n(t)-) = w^n(t, y_n(t)), \\ \rho^n(t, y_n(t)-) = \hat{\rho}(w^n(t, y(t)-)), \\ \rho^n(t, y_n(t)) = \check{\rho}(w^n(t, y(t)-)), \end{cases} \\ 0 & \text{otherwise.} \end{cases} \tag{9.2.33}$$

and

$$C_1 = 2\left(\sup_{w\in[w_{\min}, w_{\max}]} \frac{d}{dw} V(\check{\rho}(w), w) + \sup_{w\in[w_{\min}, w_{\max}]} \frac{d}{dw} V(\hat{\rho}(w), w) \right) > 0, \tag{9.2.34}$$

where we recall that $\check{\rho}(w)$ and $\hat{\rho(w)}$ are the non-classical shock densities given by (9.2.28). The existence of such a finite and positive constant comes from the fact that $w \to V(\hat{\rho}(w), w)$ and $w \to V(\check{\rho}(w), w)$ are C^1 functions of w. This is a consequence of our assumptions given at the beginning of this section, between (9.2.1) and (9.2.2). We have the following property of Γ (see [137, Appendix A] for a proof)

$$\Gamma(t) \le \Gamma(0), \quad \forall\, t \in \mathbb{R}_+. \tag{9.2.35}$$

In addition, as w^n does not change in a nonclassical shock, then we can also show the following:

$$t \to TV(w^n(t,\cdot)) \text{ is a constant function on } \mathbb{R}_+. \tag{9.2.36}$$

This allows to prove Lemma 9.2.1. The convergence up to a subsequence then follows from Lemma 9.2.1, the finite propagation speeds of the waves, Helly's theorem, and Arzela - Ascoli's theorem and we have:

Lemma 9.2.2. *Let (ρ^n, w^n, y^n) be an approximate solution of (9.2.1) constructed by the wave-front tracking method described in Section 9.2.1. Then, up to a subsequence, we have the following convergences*

$$(\rho^n, w^n) \to (\rho, w), \quad \text{in } L^1_{loc}(\mathbb{R}_+ \times \mathbb{R}; [0, \rho_{\max}] \times [w_{\min}, w_{\max}]), \tag{9.2.37}$$

$$y^n(\cdot) \to y(\cdot), \quad \text{in } L^\infty_{loc}(\mathbb{R}_+; \mathbb{R}), \tag{9.2.38}$$

$$\dot{y}^n(\cdot) \to \dot{y}(\cdot), \quad \text{in } L^1_{loc}(\mathbb{R}_+; \mathbb{R}), \tag{9.2.39}$$

for some $(\rho, w) \in C^0(\mathbb{R}_+; L^1(\mathbb{R}; [0, \rho_{\max}] \times [w_{\min}, w_{\max}]))$ and $y \in W^{1,1}_{loc}(\mathbb{R}_+; \mathbb{R})$ with Lipschitz constant V_b. Moreover, there exists $C > 0$ such that $TV(\rho(t,\cdot)) < C$ and $TV(w(t,\cdot)) < C$ for all $t \ge 0$.

9.2.2.2 The limit (ρ, w, y) is a solution of the system

Thanks to Lemma 9.2.2, we have a candidate (ρ, w, y) for the solution. To show that (ρ, w, y) is indeed a solution, we need to check that it satisfies Definition 9.2.1. Note that (ρ^n, w^n) satisfies Definition 9.2.1 with (2) being replaced by

$$\begin{aligned} &\int_{\mathbb{R}_+} \int_{\mathbb{R}} \mathcal{E}_k(v(t,x), w(t,x)) \partial_t \varphi + Q_k(v(t,x), w(t,x)) \partial_x \varphi dx\, dt \\ &+ \int_{\mathbb{R}} \mathcal{E}_k(v_0, w_0) \varphi(0,x)\, dx \\ &+ \int_{\mathbb{R}_+} R(v(t,y(t)), w(t,y(t)))(v(t,y(t)) - \dot{y}) \left[\frac{k - \dot{y}}{\alpha F(\dot{y}))} - \frac{1}{R(k, w(t,y(t)))} \right]^+ \varphi(t,y(t)) dt \ge C\, TV((\rho^n(t,\cdot), w^n(t,\cdot)))2^{-n}, \end{aligned} \tag{9.2.40}$$

where the term on the right-hand side corresponds to the fact that the approximated solution can have rarefaction shocks, that are themselves non-classical shocks but with a small amplitude.
As $(\rho^n, w^n) \to (\rho, w)$ in $L^1_{loc}(\mathbb{R}_+ \times \mathbb{R}, [0, \rho_{\max}] \times [w_{\min}, w_{\max}])$ we can use the dominated convergence theorem and pass to the limit in (9.2.7) to get that (ρ, w) satisfies (9.2.7) with initial condition (ρ_0, w_0). It remains to show (9.2.8), (9.2.11) and (9.2.12).

Showing that (ρ, w, y) satisfies (9.2.8) and (9.2.12) uses the fact that (ρ^n, w^n, y^n) is a solution of the system in the sense of Definition 9.2.1, together with choosing an appropriate test function and successive dominated convergence theorem in the same fashion as in [9, 177]. More details can be found in [137].

Showing that (ρ, w, y) satisfies (9.2.11) is the main difficulty. If there were no AV and the solution were entropic, this would not be such a problem. But the issue is that the existence of nonclassical shock makes that the usual tools do not work here. Indeed, at first one could try to show that $V(\rho(t, y(t)+), w(t, y(t)+))$ is close to $V(\rho^n(t, y(t)+), w^n(t, y(t)+))$ for n large enough. But this is, in fact, hopeless. What saves us is the fact that what we really have to show is not this, but rather only that $\min(V_b, V(\rho(t, y(t)+), w(t, y(t)+)))$ is close to $\min(V_b, V(\rho^n(t, y(t)+), w^n(t, y(t)+)))$ for n large enough. In other words, it does not matter if $V(\rho(t, y(t)+), w(t, y(t)+))$ is not close to $V(\rho^n(t, y(t)+), w^n(t, y(t)+))$ for n large enough, as long as this only happens when both $V(\rho(t, y(t)+), w(t, y(t)+)))$ and $V(\rho^n(t, y(t)+), w^n(t, y(t)+))$ are either larger or close to V_b.

To show this, first note that there exists a negligible set $\mathcal{N}_0$ such that, for any $t \in \mathbb{R}_+ \setminus \mathcal{N}_0$,
- $\lim_{n \to +\infty} (\rho^n(t, x), w^n(t, x)) = (\rho(t, x), w(t, x))$ for almost every $x \in \mathbb{R}$.
- $s \to y(s)$ is a differentiable function at time $s = t$,
- $\lim_{n \to +\infty} y^n(t) = y(t)$,
- $\dot{y}^n(t) = \min(V_b, V(\rho^n(t, y^n(t)), w^n(t, y^n(t))))$ for any $n \in \mathbb{N}$.

Therefore, it suffices to show that

$$\lim_{n \to +\infty} \min(V_b, V(\rho^n(t, y^n(t)), w^n(t, y^n(t)))) = \min(V_b, V(\rho(t, y(t)^+), w(t, y(t)^+))), \tag{9.2.41}$$

and (9.2.11) will follow. To simplify the notations, let us define $\rho_\pm := \lim_{x \to y(t)^\pm} \rho(t, x)$ and $w_\pm := \lim_{x \to y(t)^-} w(t, x)$. Also, for $w \in [w_{\min}, w_{\max}]$, we define $\rho^*(w)$ as the (unique) density such that

$$V_b = V(\rho^*(w), w). \tag{9.2.42}$$

A similar analysis for the LWR equations (9.1.1) was conducted in [177] where the authors divided the steps in three cases:

1. $(\rho_+, \rho_-) \in (\rho^*, \rho_{\max}]$, in this case no nonclassical shock can occur since $\rho > \rho^*$.
2. $(\rho_+, \rho_-) \in [0, \rho^*]$, in this case $V(\rho, w) \geq V_b$ when $\rho \leq \rho^*$ and thus the minimum of the right-hand side of (9.2.41) is dominated by V_b.
3. $\rho_+ \leq \rho^* < \rho_-$ or $\rho_- \leq \rho^* < \rho_+$, which are the remaining cases.

However, in our case it is impossible to proceed like this directly. Indeed, in the LWR model there is no w, thus ρ^* can be defined similarly as (9.2.42) but in this case ρ^* is a constant as V does not depend on an w. In our case ρ^* depends on w which can be discontinuous as well and thus $\rho^*(w_+)$ may not necessarily have the same value as $\rho^*(w_-)$. Besides, it could be that even if (ρ^n, w^n) is close to (ρ, w) for some x, it could be instantaneously be brought away by a 2-wave, which would not exist when looking at the LWR analogous.

Looking at (9.2.1) and following an argument very close to the one used to derive the Rankine-Hugoniot conditions, we can show that there exists a negligible space $\mathcal{N}$ such that $\mathcal{N}_0 \subset \mathcal{N}$ and for any $t \in \mathbb{R}_+ \setminus \mathcal{N}$,

$$w_+ = w_- \quad \text{or} \quad V(\rho_+, w_+) = V(\rho_-, w_-) \quad \text{or} \quad \rho_+ = \rho_- = 0. \tag{9.2.43}$$

In the vacuum case $\rho_+ = \rho_- = 0$ it is relatively easy to show that (9.2.41) holds, hence it can be discarded in the following without loss of generality. From there we can show the following Lemma:

Lemma 9.2.3. *Let $t \in \mathbb{R}_+ \setminus \mathcal{N}$ and $\varepsilon > 0$. Let $(\rho_+, \rho_-) \in ([0, \rho_{max}])^2$ and $(w_+, w_-) \in ([w_{\min}, w_{\max}])^2$. There exists $\delta > 0$ such that, for $n \in \mathbb{N}$ large enough, if $x \in (\min(y^n, y) - \delta, \min(y^n, y))$ two cases can occur:*

$$\begin{aligned} &V(\rho^n(t,x), w^n(t,x)) \in \mathcal{B}_\varepsilon(V(\rho_-, w_-)), \\ \text{or } &V(\rho^n(t,x), w^n(t,x)) \in [V_b - 2\varepsilon, +\infty) \text{ and } V(\rho_-, w_-) \in [V_b - \varepsilon, +\infty). \end{aligned} \tag{9.2.44}$$

And, for $n \in \mathbb{N}$ large enough, if $x \in (\max(y^n, y), \max(y^n, y) + \delta)$,

$$\begin{aligned} &V(\rho^n(t,x), w^n(t,x)) \in \mathcal{B}_\varepsilon(V(\rho_+, w_+)), \\ \text{or } &V(\rho^n(t,x), w^n(t,x)) \in [V_b - 2\varepsilon, +\infty) \text{ and } V(\rho_+, w_+) \in [V_b - \varepsilon, +\infty). \end{aligned} \tag{9.2.45}$$

where $\mathcal{B}_r(a)$ stands for the ball centered in a of radius ε.

This Lemma shows that in a short spatial area before $\min(y, y^n)$ and after $\max(y, y^n)$, either the velocity $V(\rho^n, w^n)$ is very close to $V(\rho, w)$ or both are above (or close) to V_b. This is precisely what we want. Note that this is not obvious, as the L^1_{loc} convergence of (ρ^n, w^n) to (ρ, w) only gives information almost everywhere in x, and this is all our problem: we want a convergence at a precise location $y(t)$. This lemma is illustrated in Figure 9.1, which is taken from [137].

Among others, the proof of Lemma 9.2.3 relies on the fact that, when w is fixed, one cannot go instantly from a density to another very different density using a rarefaction shock. In other words, as soon as there is a rarefaction shock, there is an incompressible minimal distance to move from a density to another. This means that the only arbitrarily large variations of density that can happen correspond to a shock and therefore has to satisfy some entropy condition (unless in the special case where it is a non-classical shock due to the interaction with the AV). This is expressed more rigorously by the following Lemma.

Lemma 9.2.4. *Let $t \in \mathbb{R}_+ \setminus \{0\}$ and $(x_1, x_2) \in \mathbb{R}^2$ with $x_1 < x_2$. Suppose that w^n is constant between (x_1, x_2) and there exists $c > 0$ (independent of n) such that $\rho^n(x_1) \geq \rho^n(x_2-) + c$. Suppose in addition that there is no non-classical shock occurring in (x_1, x_2). Then there exists $\beta' > 0$ such that*

$$|x_2 - x_1| > \beta' t |\rho^n(x_1) - \rho^n(x_2-)| \tag{9.2.46}$$

where β' is a constant independent of t, x_1 and x_2.

Given the definition of the mesh and the wave-front tracking algorithm described above in Section 9.2.1.2, we can also estimate the maximal change of velocity that can happen in a rarefaction shock:

Lemma 9.2.5. *Suppose that a rarefaction shocks occur in $x_1 \in \mathbb{R}$, then there exists a constant C_0 independent of x_1, n, ρ^n, w^n and depending only on V such that*

$$V(\rho^n(x_1^+), w^n(x_1^+)) - V(\rho^n(x_1^-), w^n(x_1^-)) \leq \frac{C_0 \rho_{\max}}{2^n} \tag{9.2.47}$$

Schematically, the proof of Lemma 9.2.3 can be decomposed as follows:

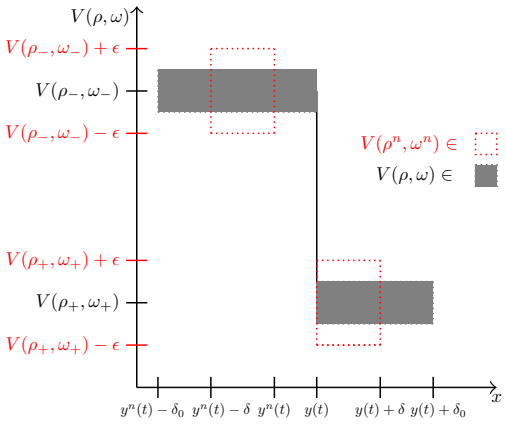
— Given that (ρ, w) has bounded variation we can show that for every $\epsilon > 0$ there exists $\delta_0 > 0$ such that for any $x \in (\min(y^n, y) - \delta_0, \min(y^n, y))$,

$$(\rho(t,x), w(t,x)) \in \mathcal{B}_{\varepsilon/2M}(\rho_-, w_-), \tag{9.2.48}$$

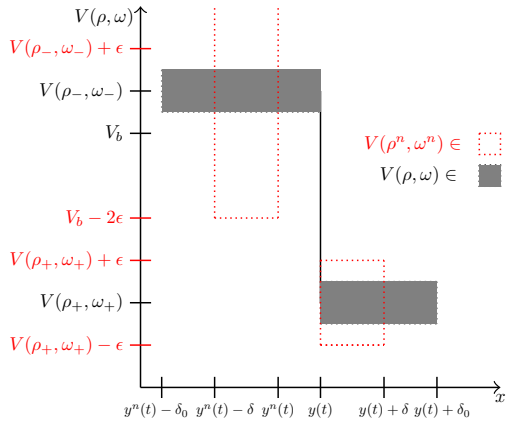
where M is the Lipschitz constant of V.

— We start by the case $x < \min(y^n, y)$. If $V(\rho_-, w_-) < V_b - \varepsilon$ we prove Lemma 9.2.3 by contradiction, showing that, if it does not hold, then by a diagonal argument there exist three sequences (x_n, z_n^1, z_n^2) such that

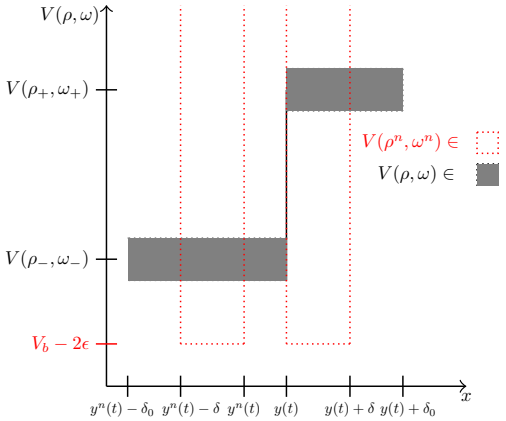
$$\begin{aligned} &\lim_{n \to +\infty} x_n = \lim_{n \to +\infty} z_n^1 = \lim_{n \to +\infty} z_n^2 = y(t), \\ &z_n^1 < x_n < z_n^2 < \min(y^n, y), \ \forall\, n \in \mathbb{N}, \end{aligned} \tag{9.2.49}$$



Case 1: $V(\rho_-,w_-) < V_b - \epsilon$ and $V(\rho_+,w_+) < V_b - \epsilon$



Case 2: $V(\rho_-,w_-) \geq V_b - \epsilon$ and $V(\rho_+,w_+) < V_b - \epsilon$



Case 3: $V(\rho_-,w_-) \geq V_b - \epsilon$ and $V(\rho_+,w_+) \geq V_b - \epsilon$

Figure 9.1 – Illustration of Lemma 9.2.3; let $t \in \mathbb{R}_+ \setminus \mathcal{N}$, $\varepsilon > 0$, $(\rho_+,\rho_-) \in ([0,\rho_{max}])^2$ and $(w_+,w_-) \in ([w_-,w_+])^2$ with $y^n(t) < y(t)$. The approximate speed $V(\rho^n(t,\cdot),w^n(t,\cdot))$ over $[y^n(t)-\delta,y^n(t)]\cup[y(t),y(t)+\delta]$ belongs to the area surrounded by the dotted lines (...) and $\rho(t,\cdot)$ over $[y(t)-\delta_0,y(t)+\delta_0]$ belongs to the shaded zone.

and for all $n \in \mathbb{N}$,

$$\begin{aligned} &V(\rho^n(x_n),w^n(x_n)) \in \mathbb{R} \setminus \mathcal{B}_\varepsilon(V(\rho_-,w_-)),\\ &(\rho^n,w^n)(z_n^1+) \in \mathcal{B}_{3\varepsilon/4M}(\rho_-,w_-) \text{ and } (\rho^n,w^n)(z_n^2-) \in \mathcal{B}_{3\varepsilon/4M}(\rho_-,w_-)\\ &V(\rho^n,w^n)(z_n^1+) \in \mathcal{B}_{3\varepsilon/4}(V(\rho_-,w_-)) \text{ and } V(\rho^n,w^n)(z_n^2-) \in \mathcal{B}_{3\varepsilon/4}(V(\rho_-,w_-)). \end{aligned} \tag{9.2.50}$$

This means that the velocity needs to change quickly from z_n^1 to x_n and from x_n to z_n^2 while all these three points can be arbitrarily close.

— We study the change between z_n^1 and x_n. We show that the connection between the two points cannot be the result of a V-wave, a 2-wave, or a non-classical shock. Once these cases are excluded we conclude that there has to be a 1-wave changing rapidly the density between z_n^1 and x_n. However, given (9.2.49) this has to be a 1-wave corresponding to a rarefaction wave. And, from Lemma 9.2.4, this implies a minimal distance between z_n^1 and x_n, which gives a contradiction with the fact that they can be arbitrarily close.

— If $V(\rho_-,w_-) < V_b - \varepsilon$ we proceed similarly by contradiction. In this case non-classical shocks cannot occur either since $x < \min(y^n, y(t))$.

— The same holds with the case $x > \max(y^n, y)$ by symmetry.

After proving Lemma 9.2.3, we study what happens between $\min(y^n, y)$ and $\max(y^n, y)$. For this, we separate the cases $w_+ = w_-$ and $V(\rho_+, w_+) = V(\rho_-, w_-)$.

First case: $V(\rho_+, w_+) = V(\rho_-, w_-)$. In this case, we have the following Lemma.

Lemma 9.2.6. *Let $t \in \mathbb{R}_+ \setminus \mathcal{N}$ and $\varepsilon > 0$. Assume that $V(\rho_-, w_-) = V(\rho_+, w_+)$, then for $n \in \mathbb{N}$ large enough and $x \in (\min(y^n(t), y(t)), \max(y^n(t), y(t)))$,*

— *if $V(\rho_+, w_+) < V_b - \varepsilon/2$, then*

$$V(\rho^n(t,x), w^n(t,x)) \in \mathcal{B}_\varepsilon(V(\rho_+, w_+)), \tag{9.2.51}$$

— *if $V(\rho_+, w_+) \geq V_b - \varepsilon/2$, then*

$$V(\rho^n(t,x), w^n(t,x)) \in [V_b - 2\varepsilon, +\infty). \tag{9.2.52}$$

The proofs of Lemma 9.2.7 uses mostly the same tools as the proof of Lemma 9.2.3 and can be found in [137]. This allows to prove the desired convergence (9.2.41). Indeed, let $\varepsilon > 0$, for $n_0 > 0$ large enough we have

— if $V(\rho^+, w^+) < V_b - \varepsilon/2$, for any $x \in (\min(y^n, y) - \delta, \max(y^n, y) + \delta) \setminus \{y(t), y^n(t)\}$ we have

$$V(\rho^n(x), w^n(x)) \in \mathcal{B}_\varepsilon(V(\rho_+, w_+)). \tag{9.2.53}$$

— If $V(\rho^+, w^+) \geq V_b - \varepsilon/2$ then for any $x \in (\min(y^n, y) - \delta, \max(y^n, y) + \delta) \setminus \{y(t), y^n(t)\}$ we have

$$V(\rho^n(x), w^n(x)) \geq V_b - \varepsilon. \tag{9.2.54}$$

Hence, in both cases

$$|\min(V(\rho^n(y^n(t)+), w^n(y^n(t)+)), V_b) - \min(V_b, V(\rho^+, w^+))| \leq \varepsilon, \quad \forall\, n \geq n_0. \tag{9.2.55}$$

As ε was chosen arbitrarily (with n_0 which potentially tend to $+\infty$ when $\varepsilon \to 0$), this implies

$$\lim_{n \to +\infty} \min(V_b, V(\rho^n(t, y^n(t)), w^n(t, y^n(t)))) = \min(V_b, V(\rho(t, y(t)^+), w(t, y(t)^+))), \tag{9.2.56}$$

which is the desired convergence.

Second case: $w_+ = w_-$. Recall that ρ^* is defined by (9.2.42). As $V_b < V(0, w)$ for any $w \in [w_{\min}, w_{\max}]$ by assumption, and as V is a decreasing function with $V(\rho_{\max}, w) = 0$ for any $w \in [w_{\min}, w_{\max}]$, (9.2.42) defines $\rho^*(w)$ uniquely. It also implies that

$$\begin{aligned} &V_b < V(\rho), \quad \forall\, \rho \in [0, \rho^*(w)), \\ &V_b < V(\rho), \quad \forall\, \rho \in (\rho^*(w), \rho_{\max}]. \end{aligned} \tag{9.2.57}$$

We can now state the following lemma

Lemma 9.2.7. *Let $t \in \mathbb{R}_+ \setminus \mathcal{N}$ and $\varepsilon > 0$. Assume that $w_- = w_+$, and let w denote this value. Then for $n \in \mathbb{N}$ large enough : assume $x \in (\min(y^n, y), \max(y^n, y))$, we have the following cases*

— *if $(\rho_-, \rho_+) \in [0, \rho^*(w)]^2$, then*

$$V(\rho^n, w^n) \in [V_b - \varepsilon, +\infty), \tag{9.2.58}$$

— *if $(\rho_-, \rho_+) \in (\rho^*(w), \rho_{max}]^2$, then*

$$V(\rho^n, w^n) \in [V(\max(\rho_-, \rho_+), w) - \varepsilon, V(\min(\rho_-, \rho_+), w) + \varepsilon], \tag{9.2.59}$$

— *if* $\rho_- \geq \rho^*(w) > \rho_+$ *or* $\rho_+ \geq \rho^*(w) > \rho_-$, *then*

$$V(\rho^n, w^n) \in [V(\max(\rho_-, \rho_+), w) - \varepsilon, V(\min(\rho_-, \rho_+), w) + \varepsilon] \cup [V_b - \varepsilon, +\infty). \tag{9.2.60}$$

With this and using Lemma 9.2.3, we can show the desired convergence (9.2.41) in each of the three cases of Lemma 9.2.7. Showing the convergence in the first case is quick and will not be detailed here. The difficult cases are the second and third one. Here we give an idea of the proof for the second case $(\rho_-, \rho_+) \in (\rho^*(w), \rho_{max}]^2$. The tools used for the third case are similar.

Sketch of proof when $(\rho_-, \rho_+) \in (\rho^*(w), \rho_{max}]^2$:

— First, we can show that the desired convergence (9.2.41) holds if $y^n \geq y(t)$ for an infinite set of index, so we can restrict ourselves to the case $y^n < y(t)$ except for a finite set of index.

— Then we can show by contradiction that $V(\rho_-, w) \geq V(\rho_+, w)$.

— Finally, we study the time-space area where the solution belongs to $B_\varepsilon(V(\rho_+, w_+))$. We show that there exists $t_n > t$ such that

$$V(\rho^n(s, y^n(s)), w^n(s, y^n(s))) \in B_\varepsilon(V(\rho_+, w_+)), \quad \text{for any } s \geq t_n, . \tag{9.2.61}$$

We also show that $t_n \to t$ when n goes to $+\infty$.

This is rigorously expressed as follows: define the triangle $\mathcal{T}_0$ by

$$\mathcal{T}_0 := \left\{(s, x) \in [t, t_f) \times \left(w_{\max}(s - t) + y^n(t) - \delta, \partial_\rho f(\rho_{\max}, w_{\max})(s - t) + y(t) + \delta\right)\right\}, \tag{9.2.62}$$

where t_f is the closing point of the triangle defined by

$$t_f = \frac{y(t) - y^n(t) + 2\delta}{w_{\max} - \partial_\rho f(\rho_{\max}, w_{\max})}. \tag{9.2.63}$$

Let us also define $t_{y^n} > t$ the time at which $y^n(s)$ gets out of the triangle, i.e. the time t_{y^n} such that

$$\begin{aligned} &(s, y^n(s)) \in \mathcal{T}_0, \ \forall s \in [t, t_{y^n}), \\ &(t_{y^n}, y^n(t_{y^n})) \notin \mathcal{T}_0. \end{aligned} \tag{9.2.64}$$

Obviously $t_{y^n} \leq t_f$ since the triangle closes at t_f. With these notations, we have the following lemma.

Lemma 9.2.8. *Let* $t \in \mathbb{R}_+ \setminus \mathcal{N}$ *and* $\varepsilon > 0$. *Assume that* $(\rho_-, \rho_+) \in (\rho^*(w), \rho_{\max}]$ *with* $\rho_- \neq \rho_+$. *Assume also that* ε *is small enough such that* $\min(\rho_-, \rho_+) - \varepsilon > \rho^*(w)$, *and that* $y^n < y(t)$ *for any* $n \geq n_1$. *Let* $\delta > 0$ *be given by Lemma 9.2.3. Then for any* $n \geq n_1$, *there exists* $t_{\xi^n} > t$ *and a piecewise linear function* ξ^n *such that*

$$(s, \xi^n(s)) \in \mathcal{T}_0, \ \forall s \in [t, t_{\xi^n}), \tag{9.2.65}$$

and for any $(s, x) \in \{[t, t_{\xi^n}) \times \mathbb{R} \mid x > \xi^n(s)\} \cap \mathcal{T}_0$

$$V(\rho^n(s, x), w^n(s, x)) \in \mathcal{B}_\varepsilon(V(\rho_+, w_+)). \tag{9.2.66}$$

Besides, if we denote t_{y^n} *the time at which* $y^n(\cdot)$ *exits the triangle, there exists* $c > 0$ *independent of* n *such that* $\min(t_{y^n}, t_{\xi^n}) - t \geq c$ *and there exists* $t_n > 0$ *such that* $\xi^n(t_n) = y^n(t_n)$ *and* $\lim_{n \to +\infty} t_n = t$.

As announced, this Lemma shows the existence of a time-space area $\{[t, t_{\xi^n}) \times \mathbb{R} \mid x > \xi^n(s)\} \cap \mathcal{T}_0$ in which $V(\rho^n(s, x), w^n(s, x))$ belongs to $\mathcal{B}_\varepsilon(V(\rho_+, w_+))$. What is shown in this Lemma is that y^n always enters this area before exiting the triangle which implies that after the time t_n, $V(\rho^n(s, y^n(s)), w^n(s, y^n(s)))$ belongs to $\mathcal{B}_\varepsilon(V(\rho_+, w_+))$. Moreover $t_n \to t$ when $n \to +\infty$.

— Using Lemma 9.2.8 we deduce that there exists c independent of n such that for any $s \in (t, t+c)$, there exists n large enough such that

$$V(\rho^n(s, y^n(s)), w^n(s, y^n(s))) \in B_\varepsilon(V(\rho_+, w_+)). \tag{9.2.67}$$

Thus, as (ρ^n, w^n, y^n) is a solution of the system and in particular of (9.2.3), we have for any $s \in (t, t+c)$

$$y^n(s) - y^n(t) = \int_t^s \min(V_b, V(\rho^n(\tau, y^n(\tau)), w^n(\tau, y^n(\tau))))d\tau. \tag{9.2.68}$$

Thus we deduce using this and (9.2.67) that

$$\left| \frac{y^n(s) - y^n(t)}{s-t} - \min(V_b, V(\rho_+, w_+)) \right| \leq \varepsilon. \tag{9.2.69}$$

and using the convergence of y^n to y, and the continuity of V, we obtain for any $s \in (t, t+c)$

$$\left| \frac{y(s) - y(t)}{s-t} - \min(V_b, V(\rho_+, w_+)) \right| \leq \varepsilon. \tag{9.2.70}$$

— We conclude by recalling that y is differentiable in t from the definition of $\mathcal{N}$ and that (9.2.69) holds true for any $\varepsilon > 0$ as long as s is close enough to t. Hence

$$\dot{y}(t) = \min(V_b, V(\rho_+, w_+)) = \min(V_b, V(\rho(t, y(t)+), w(t, y(t)+))), \tag{9.2.71}$$

which exactly (9.2.11) and ends the proof.

9.2.3 Open-questions

This analysis raises several open questions. In particular:

— Does the system (9.2.1), (9.2.3)–(9.2.4) present stop-and-go waves when there is no AV?

— Similar to the previous question, is it possible to find self-sustained travelling waves (so-called jamitons) for a circular road modelled by GARZ equations, just like it is for other models (see [107, 213, 219])?

— If so, is it possible to smooth these stop-and go waves using a feedback control?

These three questions are a current work in progress with Shengquan Xiang and Benedetto Piccoli. Going further, other interesting questions could be raised:

— If a feedback can be derived from this system, can it be translated in the microscopic framework and, if so, would it guarantee the stability of the traffic described by microscopic models such as the Bando-FTL model studied in Section 8.2 ?

— The system (9.2.1), (9.2.3)–(9.2.4) is intrinsically multilane as the AV can be passed by other vehicles from the bulk traffic. This multilane property lies in the coefficient $\alpha \in (0, 1)$ in (9.2.4), which represents the proportion of space left on the road when the AV is blocking a lane. What happens in the single lane limit, namely if $\alpha \to 0$?

Part IV

(Deep) Learning mathematics

"Le cri du révolutionnaire Rabaut-Saint-Etienne est bien connu : « Notre histoire n'est pas notre code ». Toutefois, l'histoire des pages qui suivent est bien celle de notre code. "

– Un économiste anonyme

Chapter 10

Learning mathematics with AI

10.1 Introduction

Having intelligent computers able to solve complicated problems on their own has been a sci-fi fantasy for almost as long as computers have existed. The progress of AI in the last 20 years has made this a reality for a number of tasks and has revolutionized some areas such as vision [240, 243] or translation and natural language processing [165, 236]. However, if it is usually conceivable that an AI could translate words, play chess or process data as well or better than humans, it is often hard to believe that they could perform abstract mathematics on their own [1]. This is what we investigate in this chapter. We look at two aspects of this question:

— Can an AI predict a solution to an abstract mathematical problem? This is the object of Section 10.2.

— Can an AI prove a theorem and give a proof? This is the object of Section 10.3.

As this chapter is not purely dealing with mathematics but rather with potential applications of AI to mathematics, we will keep it relatively short and introductory.

10.2 Predicting solutions to abstract maths problems

10.2.1 Problems considered

This section is taken from [46], a collaboration with François Charton and Guillaume Lample. The motivation behind this is the amazing ability of deep language models developed in the last ten years (and in particular since 2014) to solve translation problems between languages [14, 165, 229] and even to learn [2] grammatical structures without any prior knowledge [212]. The idea is to look at mathematics as a translation problem: a statement in which one needs to understand the structure and the meaning, translated in a solution. The first work of this kind by Charton and Lample in [164] showed that neural networks could predict explicit solutions to ODE when they have one, with an accuracy comparable to computer algebras like Mathematica. In other words, the pattern recognition allowed by these networks is so good that it can somehow map the link between equations and solutions in most cases. It is also worth noting that, contrary to Mathematica, the neural network has no *a priori* mathematical knowledge and no built-in rules. In [46] we investigate more complex problems where solving through pattern recognition looks harder or less intuitive. We look at the following problems

1. at least from a mathematician's perspective
2. and somehow understand

(P1) Given the nonlinear system

$$\dot{x} = f(x), \tag{10.2.1}$$

where $x \in \mathbb{R}^n$, f is C^1 around a given $x^* \in \mathbb{R}^n$ such that $f(x^*) = 0$, is x^* an exponentially stable equilibrium? If so, what is the decay rate?

(P2) Given the nonlinear control system

$$\dot{x} = f(x, u), \tag{10.2.2}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, f is C^1 around a given $(x^*, u^*) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $f(x^*, u^*) = 0$, is the linearized system locally controllable around (x^*, u^*) ?

(P3) Assume the same framework as (P2). If the linearized system is locally controllable around (x^*, u^*), what would be a stabilizing feedback $u(t) = u^* + K(x(t) - x^*)$, where $K \in \mathbb{R}^{m \times n}$? In other words, can we give an example of K such that

$$\dot{x} = f(x, Kx) \tag{10.2.3}$$

is locally exponentially stable around x^* ?

Intuitively, it may be conceivable that a pattern recognition could allow good results when recognizing explicit solutions to ODE or integrating functions with an explicit primitive. Being able to solve the three control problems above through pattern recognition seems harder to believe [3]. Indeed, these problems cannot be solved by simple interpolation, and seem to require a deeper understanding of the maths behind. This raises a more philosophical question: is it possible to learn maths from example? And, if so, is it possible in these cases that a neural network can learn the theorems behind and grasp the mathematical structures?

10.2.2 Representations, encoding and automatic generation of training data

10.2.2.1 Representation and data generation

In order to train the neural network, one needs some data, and preferably a lot of them. Using humain tabulated data would be definitely too limiting and for this reason we needed to find a way to automatically generate a dataset of statement and solutions for these different problems. To do so, we represent mathematical expressions as trees (see Fig. 10.1) where

— each leaf is a variable or a number (integer or float)

— each internal node is an operator that can be unary (typically for one-variable functions like exp, ln, cos, but also for differentiation operators etc.) or binary (typically for addition, multiplication, division, etc.)

In this work we only consider the usual trigonometric, exponential and algebraic functions as unary operators but this could be extended to other special functions. This representation is also used in [164] and one can find a related representation for instance in [11]. This representation allows to easily sample random expressions by sampling randomly a tree-shape, and then filling the nodes and leaves uniformly at random. Using this, a formal differentiation and Python libraries, we are able to automatically generate differential systems with or without control and to compute the solutions of the different problems (P1)–(P3). The details of the procedure are given in [46] and the Python environment can be found at in the associated repository MathsFromExamples. This environment is very modular so that it can be relatively easily adapted to other mathematical problems [4]. This allows us to generate a dataset of over 100 millions examples of systems and solutions to the different problems (P1)–(P3). This dataset is also available in the repository MathsFromExamples. 100 millions of examples might seem a lot at first, but this is only a very tiny subset

3. at least from my point of view.
4. An interested reader can definitely clone the code and try.

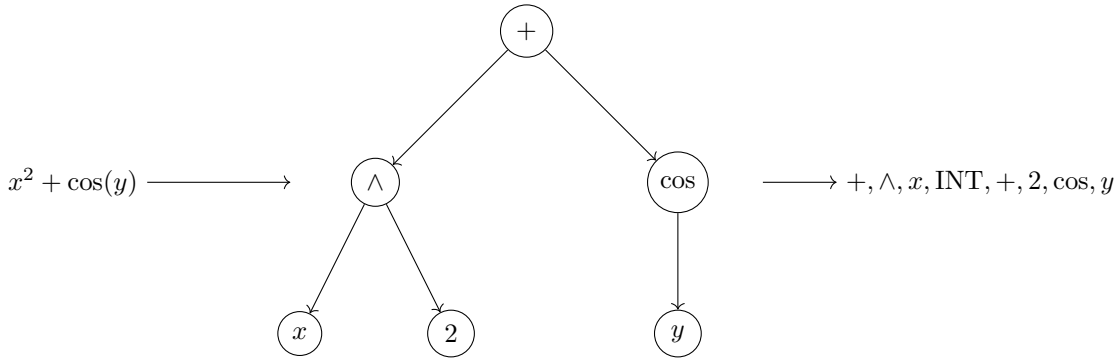


Figure 10.1 – **Representation of mathematical expressions** The mathematical expression $x^2 + \cos(y)$ (left) represented as a tree (center) and then encoded as a string of tokens in prefix order (right).

of all the possible systems. Formally, we are generating at random dynamics f that belongs to an infinite-dimensional space [5]. Given that the number of functions we use are finite and the resolution of the numbers we take is finite as well, the space of possible functions we can generate with our procedure becomes finite, but still with a very large cardinal (much larger than the number of atoms in the universe). Hence, the 100 million examples are a very tiny subpart. We can note that, during the random generation of these 100 million examples, we never encountered a single duplicate.

10.2.2.2 Encoding and architecture

We used a classical Transformer architecture [236] with dimensionality between 68 and 512 and 1 to 6 hidden layers. The details can be found in [46]. The model takes a sequence of tokens in input, and outputs another sequence of tokens. Therefore, the statements are encoded as follows from the mathematical representation:

— **Mathematical expressions** are seen as trees as described earlier, and each tree is then translated into a string of tokens using a prefix enumeration (also called normal Polish notation). This consists in enumerating the tree from the root node, and writing each node before its children, listed from left to right. In this list, each operator or variable is represented by a token. For instance:

$$xy + z \quad \rightarrow \quad \text{`+','}\times\text{', `}x\text{', `}y\text{','}z\text{',} \tag{10.2.4}$$

— **A matrix or a vector** is decomposed in rows and columns which are placed end-to-end in a single string and separated by special tokens.

— **An integer or a float** is represented by a special token, a sign and one or several digits

$$-4 \quad \rightarrow \quad \text{`INT','-','4'}$$
$$1.67 \quad \rightarrow \quad \text{`FLOAT','+','1','.','6','7'.}$$

For the solutions, if the answer is qualitative (yes or no), the encoding is simply 0 or 1. When the answer is quantitative the encoding is the same as for the statements. This encoding is summarized on Fig. 10.1.

10.2.3 Results

During generation, the data are separated into a training and an evaluation dataset with (of course) no overlap. The neural network learns on the training dataset and, after training, is evaluated on the evaluation dataset to see its ability to generalize what it has learned on problems it has never seen before, i.e. whether it has understood some maths or just learned by heart. We discuss here the results obtained for the different problems.

5. which also means that the universal approximation theorem does not apply.

Decay rate of the linearized system For the first task, the model was asked to predict if the system is exponentially stable. For this task the evaluation dataset is selected using a rejection sampling in order to have a balanced distribution of 50% exponentially stable systems and 50% of exponentially unstable systems (other systems for instance with a polynomial decay or growth where the highest real part of the eigenvalues of the linearized system is zero are discarded). The neural network shows an impressive accuracy of around 95% for systems with 3 to 5 equations. The results are summarized in Table 10.1.

Table 10.1 – **Accuracy of predictions of stability (chance level: 50%)**

	Two equations	Three equations	Four equations	Five equations	Overall
Accuracy	**98.2**	**97.3**	**95.9**	**94.1**	**96.4**

Then, the neural network is asked to give an estimate of the decay rate. We consider this estimate to be correct if it falls within a 10% tolerance of the actual value. The results are summarized in Table 10.2. As one can see, the results deteriorate quickly with the number of equations, but remain pretty good when considering that the neural network has a priori no knowledge of maths before training.

Table 10.2 – **Prediction of local convergence speed (within 10%).**

	Two equations	Three equations	Four equations	Five equations	Six equations	Overall
8 layers, dim 1024	**96.3**	**90.4**	**86.2**	**82.7**	**77.3**	**86.6**

Controllability In this second problem, the model is given nonlinear control systems with n equations and m controls, n and m being chosen at random, and is asked to say whether the linearized system is controllable or not. This would be typically done in mathematics by linearizing the system and using Kalman criterion. The model is again evaluated on a dataset with 50% of system with a controllable linearized system and 50% of systems with a non-controllable linearized system. Once again the model manages to find the answer with a striking accuracy above 97%.

Next, the model is asked to provide a stabilizing feedback matrix when the linearized system is controllable. Namely, for a system of the form (10.2.1), the model is asked, if the system is controllable, to provide a matrix K such that the control $u(t) = u^* + K(x(t) - x^*)$ makes the system exponentially stable. The results are summarized in the second line of Table 10.3. In the first line of Table 10.3 we look at another criteria: we check whether the matrix that the model outputs is within a 10% range (in l^1 norm) of the classical stabilizing matrix that would be constructed from the controllability Gramian matrix (see [184] or [55, Theorem 10.16]). We do so since, in the training dataset, the feedback stabilizing matrices that are given as examples are derived from the controllability Gramian matrices. Interestingly when the number of equations increases the neural network hardly ever predict a feedback matrix close to the reference feedback matrix of the training dataset, but still outputs a valid solution to the stabilization problem with a reasonably high accuracy.

Table 10.3 – **Prediction of feedback matrices - Approximation vs. correct mathematical feedback.**

	Three equations	Four equations	Five equations	Six equations	Overall
Prediction within 10%	**50.0**	**9.3**	**2.1**	**0.4**	**15.8**
Correct feedback matrix	87.5	77.4	58.0	41.5	66.5

10.2.4 Discussion and open questions

Neural networks with a Transformer architecture seem to learn to predict solutions to some advanced maths problems [6] with a high accuracy, even though these problems looked unlikely learnable by examples. Yet the model achieves over 95% accuracy on qualitative tasks and between 50% and 85% accuracy on quantitative tasks.
Several things can be noted: not only the model is able to generalize from the training dataset to the whole set of possible functions, but in addition the model is able to generalize also pretty well to a biased distribution of examples. Indeed, evaluating on systems that have larger expression, or have a biased distribution of operators (for instance no trigonometric functions or on the contrary only trigonometric functions) still gives very high results. More surprising: when trained on examples with 2 to 5 equations the model has a good accuracy when evaluated on systems of 6 equations, even though it has never seen a system of 6 equations before (and it has never seen the variable x_6 before either). This suggests that the model indeed learns the maths behind, at least in some sense.
A similar approach was used for other problems, such as some graph problems arising from computational biology [47] or linear algebra [45].
Finally, an interesting open question would be to extend these results to problems that are hard to solve with the current mathematical theories but where the candidate solution can be checked easily. One of such problems is finding Lyapunov functions for a given system, for which no general method exists. This problem may be NP-hard in general [2] which might be out of reach for the neural network [252]. But there could be a large subset of dynamics that encompass many systems usually found in mathematical applications and where this obstruction does not occur anymore.

10.3 Teaching AI to prove theorems

Having computers doing maths and proving theorems has been a longstanding dream, both in sci-fi and academics. In 1976 a turn takes place with the computed-assisted proof of the four colors theorem [10]. Since then, several theorems were proved using a computer-based proof in some sense. In the most recent examples, one can cite Keller's conjecture in dimension 7 where the computer assisted proof takes a monstruous size (200 Gb) [31]. In many of these examples, the computer is used for its ability to explore large computations or cases disjunctions that would be unfeasible by humans. However, this remained localized to specific applications and still far from having an AI able to grasps mathematical rules in general and prove involved theorems on its own.
In parallel, formal verification of proof and proof assistants have known a large interest since the 1960s (see for instances [40, 114]). Many proving environments were build, such as the iconic *Coq* or *Isabelle*, but also more recent environments like *Lean*. In these environments the proof is automatically checked and certified correct by the computer. This has a large interest in a scientifical area like mathematics where "almost correct" is incorrect and where the modern proofs tend to be increasingly large, complex, and harder to check [39, 186]. Its adoption by mathematicians has been limited by the fact that the users' proofs have to obey a rigorous syntax, far from written mathematics as we know. This syntax remains relatively tedious, despite efficient proof assistants and many improvements in the last decades. However, a large amount of the classical mathematical theories and even significant aspects of very recent theories have been formalized in the last few years by a small (but drastically increasing) community. In *Lean* one can look for instance at [41, 52, 96, 97, 131, 246]). In [41], in particular the authors formalize the perfectoid spaces introduced by Peter Scholze in [217] and an on-going work involving many researchers aims to formalize (and certify) the proof one of Scholze's theorems as part of the liquid tensor experiment [218] [7].
From an artificial intelligence perspective, being able to teach mathematical reasoning to a computer that would automatically prove theorems has been a classical interest since the 1950s [83, 115]. However, the interest for applying modern techniques of deep-learning to this goal has been very recent with many progresses in the last two years. One can cite for instance the efforts of [249], a first attempt of inequality benchmark for theorem proving, [13] an attempt to guide proof in a classical automated prover thanks to

6. by "advanced", we mean problems that would be typically taught in a master in mathematics or applied mathematics
7. this work has been announced to be finished on July 14th, 2022 here.

a deep-learning approach, GPT-f [204] and following works [203] using a huge language model designed for translation and trained on formal mathematical proofs. These are just a few of the numerous works that have been done in the last two years. In this section, we present the work of [166], which currently holds the state-of-the-art performances in deep-learning proving and in generation of synthetic theorems. This is a collaboration with Guillaume Lample, Marie-Anne Lachaux, Thibault Lavril, Xavier Martinet, Gabriel Ebner, Aurélien Rodriguez, and Timothée Lacroix.

10.3.1 Environment and proof exploration

To obtain a neural network whose proofs are automatically checked, we work within a formal proving environment. We considered three of such environments: *Metamath* [188], *Lean* [192] and *Equations*, an environment we built to deal with equations and inequations, and that can also generate synthetic theorems. In each of them the prover works backward: we start with a statement to prove and a set of assumptions, and we would like to apply theorems until there is nothing left to prove. An example is given in Fig. 10.2. Formally, applying a theorem to a statement can either create new statements to prove or no statement at all, which indicates that the proof of the original statement is finished. All three environments have their strength and weaknesses: *Metamath* has only a single type of tactic [8]: substitution. All theorems are an iteration of substitutions, which makes it very simple to use but the proof sizes are quickly very large. *Lean* on the other hand is more complex and much richer. It is based on type theory and allows, for instance, automatic inferences by using metavariables. Its tactic are powerful but a prover in *Lean* is harder to implement. Finally, *Equations* is very modular, embedded in Python which makes it flexible, and the theorems that can be used in the proofs are user-specified and easy to design. It has two types of tactics: *assertion*, which assert that a statement is correct if some assumptions are satisfied, and *transformation* which replace the expression by an equivalent expression, potentially provided some assumptions. On the other hand, *Equations* is restricted to equalities and inequalities with only a "for all" quantifier which limits it to relatively simple mathematics.

In classical proof theory, one can represent the set of possible proofs of this statement as a huge (infinite) directed hypertree $\mathcal{T}$ where:

— the root node is the statement to prove,
— the edges starting from this node are all the possible tactics that can be applied to this node (i.e. theorems, substitutions and where to apply them),
— the children nodes obtained from these edges are the statements that remain to be proved after applying this tactic. This can be either 0, one or several statements.

A branch stops if and only if it encounters an empty node, which means that the statement from which the edge originate is proved. If one can find a subtree starting from the root node and ending on empty nodes only in all the branches, then the initial statement is correct and this subtree is a proof. Therefore, formally, proving a theorem is all about exploring correctly this huge hypertree.

Proof exploration: similarities and differences Exploring a graph in clever way is a problem that can be found in many areas and many approaches exist for this, for instance [1, 156, 247]. A similar problem occurs, for example, when playing chess: a chess game can be seen as one branch of a huge tree starting from the initial position and representing all possible games. This motivates us to use an approach inspired from AlphaZero's Monte-Carlo Tree Search (MCTS) [222]. AlphaZero is the IA model that became famous for playing chess much better than any human after only several days of training. However, there are several differences in our framework that make the problem much harder than playing chess:

— When playing a move in chess, there is still only one chessboard and one position. While, when applying a theorem to a goal statement, this might result in several statement that have to be proven.

8. a tactic here refers to a proof operation

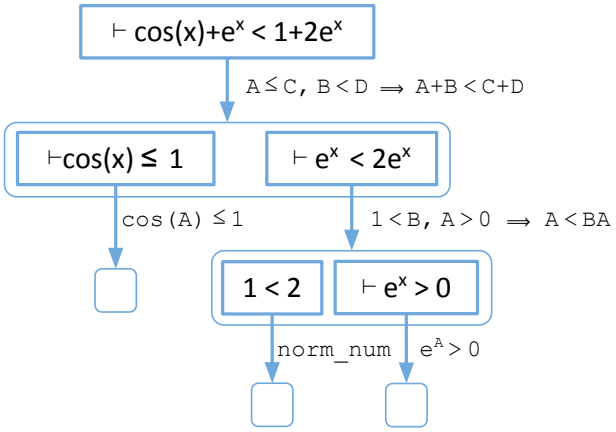


Figure 10.2 – **An example of proof-tree** for $\cos(\mathbf{x}) + \mathbf{e^x} < \mathbf{1} + \mathbf{2e^x}$ in *Equations*. The arrows correspond to the application of a theorem. *norm_num* refers to a numerical evaluation.

— While in chess the number of possible moves is of the order of 100, the number of possible applications of theorems on a mathematical statement is usually many order of magnitudes above (when finite). This makes completely impossible to explore even all the possible next few steps when doing maths, in contrast with chess. Also, an unsuccessful tactic gives much less information than a bad move in chess.

— In chess, one does not need to play perfectly to win, it is just about playing better than the opponent. Hence, a bad or sub-optimal move does not necessarily lead to losing the game. In maths, however, the proof must be valid and it happens more often to be in a position that cannot succeed, because we have an assumption to prove that happens to be wrong, leading to a branch that can never be proved.

HyperTree Proof Search algorithm In [166] we introduce an algorithm called *HyperTree Proof Search (HTPS)*. This algorithm assumes that we have at our disposal two estimators: a *tactic estimator*, that estimates the best theorems and substitutions to apply to a given statement, and a *critic* that estimates the probability that the model eventually manages to find a proof of the statement. We present here an introduction to this algorithm. More details, for instance concerning the training of the tactic estimator and critic models, the architecture, and the evaluation of the overall algorithm can be found in [166]. Let us introduce two quantities: $N(t, g)$ the number of times a tactic t (i.e. a theorem and where to apply it) has been applied to a goal g and a total value $W(t, g)$ which we define later on but represents in some sense the value attributed to a given tactic t for a given goal g during the entire search procedure. The algorithm works in three steps:

— **Selection**: Using the policy, we select a tactic given the current statement and we apply it. The policy depends on the tactic estimator, the visit count function $N(\cdot, g)$ and the total value function $W(\cdot, g)$, where g denotes the original statement to prove. After applying the theorem there remains to prove either no statement (the proof is finished), one or several statements, that we denote by $(g_0^1, ..., g_{n_1}^1)$. Then we can again select a tactic to apply to each of them, using the policy, and repeat this a given number of times m. This leads to a number of statements (hopefully easier than the original statement) to be proven $(g_0^m, ..., g_{n_m}^m)$.

— **Expansion**: We use again the policy to suggest several tactics that would help proving the goals $(g_0^m, ..., g_{n_m}^m)$.

— **Back-propagation**: We give a value to each goal g_i^m: if a tactic suggested proves it, we give to this goal a value $v_{g_i^m} = 1$; if there is no valid tactic suggested or the goal is obviously wrong we give it the value 0; and in any other cases we give it a value estimated by the critic model. Then, we back-propagate the value to the nodes of the previous steps $g_0^{m-j}, ..., g_{n_{m-j}}^{m-j}$ for $j \in \{1, ..., n\}$ iteratively as follows: for a node g_i^j linked by the chosen tactic to the nodes $(g_{i_1}^{j+1}, ..., g_{i_2}^{j+1})$ with respective values $(v(g_{i_1}^{j+1}), ..., v(g_{i_2}^{j+1}))$ we set

$$v(g_i^j) = \prod_{k=i_1}^{i_2} v(g_k^{j+1}). \tag{10.3.1}$$

Finally, for each node g_i^j where a tactic t has been applied, we update $N(t, g_i^j)$ the number of times this node appeared and t was applied and we add the value $v(g)$ to the total value $W(t, g_i^j)$. Then we estimate value of the tactic for this node as the ratio $Q(t, g_i^j) = W(g_i^j, t)/N(g_i^j, t)$.

The larger $N(g_i^j, t)$, the more weight is attributed in the policy to the estimated value of the tactic rather than the value given from the tactic estimator model, and conversely. This is summarized in Fig. 10.3 that is taken from [166] (and a more detailed description can be found in [166]).

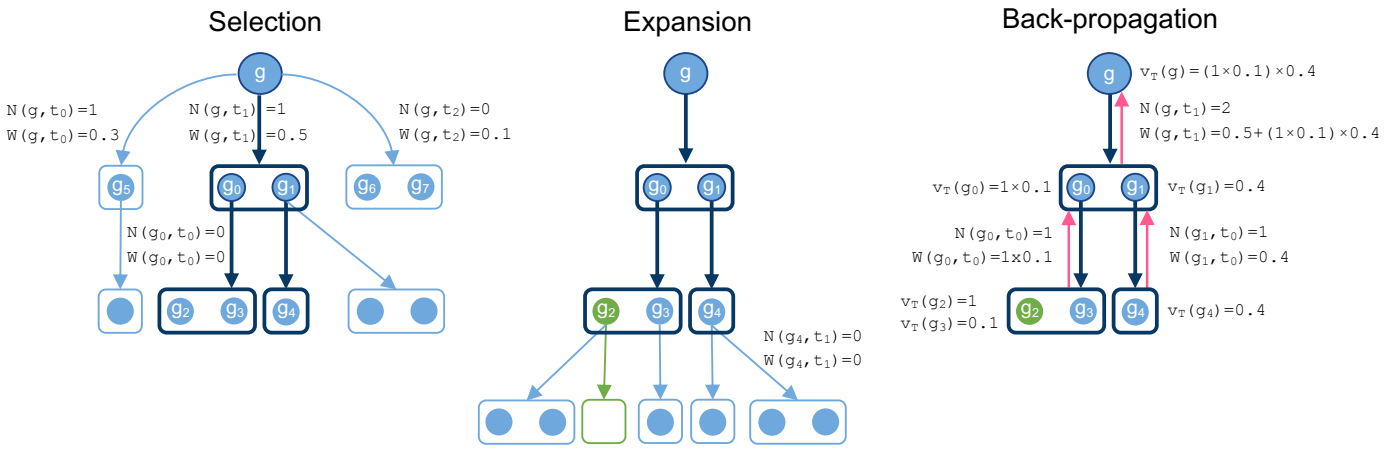


Figure 10.3 – **HyperTree Proof Search**. We aim at finding a proof of g. Being able to prove either $\{g_5\}$, $\{g_0, g_1\}$, or $\{g_6, g_7\}$ would lead to a proof of g using respectively tactic t_0, t_1, or t_2. Guided by the search policy, we first select a hypertree whose leaves are unexpanded nodes. The selected nodes are then expanded, adding new tactics and nodes to the hypergraph. Finally, during back-propagation we evaluate the node values of the hypertree, starting from the leaves back to the root and update the quantities N and W.

Here, both the tactic estimator and the critic model are assumed to be fixed. However, they are both neural network models themselves, that we can train as HTPS is applied. At first, the tactic estimator and critic model are obtained from a supervised training on a dataset of existing proofs. Then, we improve them by using the successful proof searches as training target for the tactic estimator model. For the critic, we select hypertrees of proofs and we train the critic model to guess the value $v(g) = 0$ or $v(g) = 1$ respectively for the statement that were shown to be invalid or on the contrary successfully proved, and the value $v(g)$ computed with (10.3.1) for internal nodes of the tree. This procedure is called *online training*.

Data generation and the *Equations* environment One of the main difficulties about teaching a deep-learning model to prove a theorem is the lack of existing data. A deep language model, which is the basis of our approach, generally requires a few hundred thousands or even a few millions examples. Here, even the largest theorem database contains less than 100 000 theorems [9]. This is already a lot from a mathematician

9. although this number is quickly increasing

and a human point of view, but quite small from a machine learning standpoint. Being able to enrich this dataset is an important problematic. One possible way is to be able to generate synthetic theorems, i.e. computer-generated theorems, that could serve as additional datapoints. They could either be generated with or without their proofs and both cases would be useful, either to train the proof search or the underlying deep language model.

We created an environment called *Equations* as a simpler analogue to existing proving environments, handling only equations and inequalities. The (human-made) theorems that can be used in the proof are user-specified, as well as the number of variables, the functions allowed (e.g. exp, ln, cos), the length of the expressions, the probability distribution, the type of variables and constants (integers, natural number or real numbers), the sign of integers involved etc. *Equations* comes with a random theorem generator using (human-made) theorems to build a proof, with mainly two modes: random walk and proof graph generation ([166] for more details). Its modularity allows it to be compatible with other existing proof environments and it can be used to generate synthetic theorems in other languages, from instance in *Lean*. This works as follows:

— We built a parser that reads *Mathlib*, the dataset of *Lean*'s theorems and extract all the (human-made) theorems compatible with equations and inequalities. Then, these theorems are are converted in *Equations*.
— Theorems are randomly generated with their proof in *Equations*.
— We built a translator from *Equations* to *Lean* both for the statements and the proofs, and the automatically generated theorems are translated in *Lean*

The simplicity of *Equations* makes goals and tactics easy to understand, helping with interpretability and debugging of the HTPS and the online training.

Let us also note that generating meaningful theorems is a complicated task, even only because defining what it means for a theorem to be meaningful is already a hard task. More generally, defining a "good theorem" is complicated: for sure, it is easy to identify behaviors we would not like to see in a theorem, but it is complicated to define explicit criteria that could be automatically checked to identify such a good theorem just from the statement and/or the proof. Another difficulty comes from the generation itself. We do not necessarily need synthetic theorems to be as good as the human-made theorems. However, learning on these synthetic theorems needs to be useful for proving the human-made theorem. This means, for instance, that they need to include a certain diversity of human-made theorems in their proof. A naive random generation where the (human-made) theorems used in the proof are simply sampled and applied when possible might not be enough because it would result in always calling in the proof the same (human-made) theorems with no assumptions. This because the likelihood of being able to spontaneously generate at random the assumptions needed for a more complicated theorem is negligible.

10.3.2 Results

In order to evaluate the resulting neural network (referred to as *Evariste*), we use three different benchmarks.

For *Metamath* we look at all the theorems that were tabulated in the database, we isolate the set of theorems that are not used by any other theorems and then we randomly select some of them to be an evaluation dataset that the neural network will never see. After training, *Evariste* using manages to prove up to 82% of the theorems.

For *Lean* we use a benchmark dataset called miniF2F, introduced in [261], that typically consists of high-school olympiad problems. On this benchmark *Evariste* shows an accuracy above 40%, meaning that more than 40% of these exercises were solved by the model, sometimes after several tries. One can look at [166] for more results, and in particular for a description of how the results evolve with respect to the training set and the training time (which can be consequent: the best version of *Evariste* was training during 1620

GPU hours). Among others exercises, *Evariste* manages to solve two exercises derived from the International Mathematical Olympiads, an example of proof is given in Fig. 10.4 (taken from [166]).

On *Equations* we use as a benchmark a set of 144 classical identities, mostly coming from trigonometry, exponential properties or algebra. One thing to note is that the model only has access to basic rules: for instance, for the cosine function it only knows that $\cos(0) = 1$, $\cos(\pi/2) = 0$ and $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ for any $(a, b) \in \mathbb{R}^2$. Hence showing that $\cos^2(x) + \sin^2(x) = 1$ requires already a proof with around 10 steps and something as simple as $(x - y) - (x + y) + 2y$ would require around 20 steps. This number might seem high but recall that for a formal environment, everything is a step and no step can be skipped [10], for instance $x + y = y + x$ is one step. After training, the model is able to prove more than 78% of the identities, some of them with above 100 proof steps. In Figure 10.3.2 we give some examples of the identities proved as well as the number of steps required. One interesting thing to note is that, on *Equations*, the model only learns to prove on synthetically generated theorems but is eventually able to prove human-made theorems [11], suggesting that the synthetically generated theorems were rich enough.

```
theorem imo_1964_p1_2 (n : ℕ) : ¬7∣2^n + 1 =
begin
  rw nat.dvd_iff_mod_eq_zero,
  rewrite [nat.add_mod, nat.mod_eq_of_lt],
  obviously,
  apply nat.strong_induction_on n,
  induction n,
  {
    intros n IH,
    cases n,
    norm_num,
    cases n,
    norm_num,
    rw [nat.succ_eq_add_one, pow_succ],
    rw [nat.succ_eq_add_one, pow_succ],
    induction n,
    norm_num,
    rw [nat.succ_eq_add_one, pow_succ],
    norm_num [nat.mul_mod, ←mul_assoc],
    contrapose! IH,
    refine ⟨n_n, nat.lt_succ_iff.mpr _, IH⟩,
    exact nat.le_succ_of_le (nat.le_succ _),
  },
  exact n_ih,
end
```

Figure 10.4 – **A proof of a problem (taken from IMO 1964) found by the model.** The model shows that for any value of $n \in \mathbb{N}$, $2^n + 1$ is not divisible by 7, by showing that $2^n \bmod 7 + 1 \neq 0$, and $2^n \bmod 7 + 1 < 7$. The second part of the proof uses strong induction and the fact that $2^n \equiv 2^{n+3} \mod 7$. This proof was automatically cleaned after generation.

10. Despite the appearance of tactics acting as shortcut (such as the tactics *ring* and *linarith* in Lean or *ring_simplify*, *auto*, etc. in Coq), this makes writing a proof in a formal environment pretty tedious, and this is, in my opinion, one of the main reasons why formal proofs are so little used by mathematicians today.

11. This is often called out-of-domain generalization

Identity	Proof size		Proof depth	
	First	Minimal	First	Minimal
$\cos(x+y)\cos(x-y) = \cos(x)^2 - \sin(y)^2$	117	64	117	64
$\sin(x+y)\sin(y-x) = \cos(x)^2 - \cos(y)^2$	118	64	118	63
$\lvert\sinh(x/2)\rvert = \sqrt{(\cosh(x)-1)/2}$	86	53	61	36
$\sin(x+y)\sin(x-y) = \sin(x)^2 - \sin(y)^2$	183	66	183	65
$\cosh(x)^2 = 1 + \cosh(2x)/2$	87	40	71	32
$\cosh(2x) = 2\cosh(x)^2 - 1$	78	42	62	33
$\cosh(2x) = \cosh(x)^2 + \sinh(x)^2$	97	72	80	64
$\tanh(x) - \tanh(y) = \sinh(x-y)/(\cosh(x)\cosh(y))$	154	135	85	81
$\tanh(x) + \tanh(y) = \sinh(x+y)/(\cosh(x)\cosh(y))$	162	144	95	91
$\sqrt{1+\sinh(x)^2} = \cosh(x)$	82	70	76	62
$\sin(x)^3 = (3\sin(x) - \sin(3x))/4$	72	58	63	49
$\sin(3x) = 3\sin(x) - 4\sin(x)^3$	80	56	71	47
$\cosh(3x) = 4\cosh(x)^3 - 3\cosh(x)$	204	105	176	79
$\cosh(x)^3 = (3\cosh(x) + \cosh(3x))/4$	162	106	137	79
$\sin(4x) = \cos(x)(4\sin(x) - 8\sin(x)^3)$	73	73	60	60
$\cos(\pi/3) = \sin(\pi/6)$	26	17	26	17

Table 10.4 – **Examples of identities solved.** Some of the 144 identities proved by the model. We indicate the size and the depth of the first proof found by the model and the minimal proof found by the model during online training.

Bibliography

[1] Bruce Abramson and Richard E Korf. A model of two-player evaluation functions. In *AAAI*, pages 90–94, 1987.

[2] Amir Ali Ahmadi and Pablo A Parrilo. Stability of polynomial differential equations: Complexity and converse lyapunov questions. *arXiv preprint arXiv:1308.6833*, 2013.

[3] Mohamadreza Ahmadi, Giorgio Valmorbida, and Antonis Papachristodoulou. Dissipation inequalities for the analysis of a class of PDEs. *Automatica J. IFAC*, 66:163–171, 2016.

[4] T. Alazard, P. Baldi, and D. Han-Kwan. Control of water waves. *J. Eur. Math. Soc. (JEMS)*, 20(3):657–745, 2018.

[5] T. Alazard, N. Burq, and C. Zuily. On the water-wave equations with surface tension. *Duke Math. J.*, 158(3):413–499, 2011.

[6] Saleh Albeaik, Alexandre Bayen, Maria Teresa Chiri, Xiaoqian Gong, Amaury Hayat, Nicolas Kardous, Alexander Keimer, Sean T McQuade, Benedetto Piccoli, and Yiling You. Limitations and improvements of the intelligent driver model (idm). *arXiv preprint arXiv:2104.02583*, 2021.

[7] Miltiadis Allamanis, Pankajan Chanthirasegaran, Pushmeet Kohli, and Charles Sutton. Learning continuous semantic representations of symbolic expressions. In *International Conference on Machine Learning*, pages 80–88. PMLR, 2017.

[8] Luigi Ambrosio, François Bouchut, and Camillo De Lellis. Well-posedness for a class of hyperbolic systems of conservation laws in several space dimensions. *Comm. Partial Differential Equations*, 29(9-10):1635–1651, 2004.

[9] Boris Andreianov, Carlotta Donadello, and Massimiliano Daniele Rosini. A second-order model for vehicular traffics with local point constraints on the flow. *Mathematical Models and Methods in Applied Sciences*, 26(04):751–802, 2016.

[10] Kenneth Appel and Wolfgang Haken. Every planar map is four colorable. *Bulletin of the American mathematical Society*, 82(5):711–712, 1976.

[11] Forough Arabshahi, Sameer Singh, and Animashree Anandkumar. Combining symbolic expressions and black-box function evaluations in neural programs. In *International Conference on Learning Representations*, 2018.

[12] Abdellahi Bechir Aw and Michel Rascle. Resurrection of "second order" models of traffic flow. *SIAM J. Appl. Math.*, 60(3):916–938, 2000.

[13] Eser Aygün, Zafarali Ahmed, Ankit Anand, Vlad Firoiu, Xavier Glorot, Laurent Orseau, Doina Precup, and Shibl Mourad. Learning to prove from synthetic theorems. *arXiv preprint arXiv:2006.11259*, 2020.

[14] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[15] Athmane Bakhta and Virginie Ehrlacher. Cross-diffusion systems with non-zero flux and moving boundary conditions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 52(4):1385–1415, July 2018.

[16] Andras Balogh and Miroslav Krstic. Infinite dimensional backstepping-style feedback transformations for a heat equation with an arbitrary level of instability. *European journal of control*, 8(2):165–175, 2002.

[17] Masako Bando, Katsuya Hasebe, Akihiro Nakayama, Akihiro Shibata, and Yuki Sugiyama. Dynamical model of traffic congestion and numerical simulation. *Physical review E*, 51(2):1035, 1995.

[18] Georges Bastin and Jean-Michel Coron. On boundary feedback stabilization of non-uniform linear 2×2 hyperbolic systems over a bounded interval. *Systems & Control Letters*, 60(11):900–906, 2011.

[19] Georges Bastin and Jean-Michel Coron. *Stability and Boundary Stabilisation of 1-D Hyperbolic Systems*. Number 88 in Progress in Nonlinear Differential Equations and Their Applications. Springer International, 2016.

[20] Georges Bastin and Jean-Michel Coron. A quadratic Lyapunov function for hyperbolic density-velocity systems with nonuniform steady states. *Systems & Control Letters*, 104:66–71, 2017.

[21] Georges Bastin and Jean-Michel Coron. Exponential stability of PI control for Saint-Venant equations with a friction term. *Methods Appl. Anal.*, 26(2):101–112, 2019.

[22] Georges Bastin, Jean-Michel Coron, and Amaury Hayat. Input-to-state stability in sup norms for hyperbolic systems with boundary disturbances. *Nonlinear Anal.*, 208:112300, 28, 2021.

[23] Georges Bastin, Jean-Michel Coron, Amaury Hayat, and Peipei Shang. Exponential boundary feedback stabilization of a shock steady state for the inviscid Burgers equation. *Math. Models Methods Appl. Sci.*, 29(2):271–316, 2019.

[24] Alexandre Bayen, Maria Laura Delle Monache, Mauro Garavello, Paola Goatin, and Benedetto Piccoli. *Control Problems for Conservation Laws with Traffic Applications Modeling, Analysis, and Numerical Methods*. Springer, 2022.

[25] Alexandre Bayen, Alexander Keimer, Lukas Pflug, and Tanya Veeravalli. Modeling multi-lane traffic with moving obstacles by nonlocal balance laws. *Preprint*, 09 2020.

[26] Alexandre Bayen, Maria Laura Delle Monache, Mauro Garavello, Paola Goatin, and Benedetto Piccoli. Boundary control of conservation laws exhibiting shocks. In *Control Problems for Conservation Laws with Traffic Applications*, pages 5–37. Springer, 2022.

[27] Nikolaos Bekiaris-Liberis and Argiris I Delis. Pde-based feedback control of freeway traffic flow via time-gap manipulation of acc-equipped vehicles. *IEEE Transactions on Control Systems Technology*, 29(1):461–469, 2020.

[28] Henri Berestycki, Grégoire Nadin, Benoit Perthame, and Lenya Ryzhik. The non-local Fisher–KPP equation: travelling waves and steady states. *Nonlinearity*, 22(12):2813, 2009.

[29] Giulia Bertaglia and Lorenzo Pareschi. Hyperbolic models for the spread of epidemics on networks: kinetic description and numerical methods. *ESAIM: Mathematical Modelling and Numerical Analysis*, 55(2):381–407, 2021.

[30] Dejan M. Bošković, Andras Balogh, and Miroslav Krstić. Backstepping in infinite dimension for a class of parabolic distributed parameter systems. *Math. Control Signals Systems*, 16(1):44–75, 2003.

[31] Joshua Brakensiek, Marijn Heule, John Mackey, and David Narváez. The resolution of keller's conjecture. *Journal of Automated Reasoning*, pages 1–24, 2022.

[32] Alberto Bressan. Global solutions of systems of conservation laws by wave-front tracking. *Journal of mathematical analysis and applications*, 170(2):414–432, 1992.

[33] Alberto Bressan. *Hyperbolic systems of conservation laws: the one-dimensional Cauchy problem*, volume 20. Oxford University Press on Demand, 2000.

[34] Alberto Bressan. Hyperbolic systems of conservation laws in one space dimension. In *International Congress of Mathematicians, Beijing 2002*, volume 1, pages 159–178, 2002.

[35] Alberto Bressan and Giuseppe Maria Coclite. On the boundary control of systems of conservation laws. *SIAM J. Control Optim.*, 41(2):607–622, 2002.

[36] Alberto Bressan, Graziano Crasta, and Benedetto Piccoli. *Well-posedness of the Cauchy problem for $n \times n$ systems of conservation laws*, volume 694. American Mathematical Soc., 2000.

[37] Federico Bribiesca Argomedo, Christophe Prieur, Emmanuel Witrant, and Sylvain Brémond. A strict control Lyapunov function for a diffusion equation with time-varying distributed coefficients. *IEEE Trans. Automat. Control*, 58(2):290–303, 2013.

[38] Martin Burger, Marco Di Francesco, Jan-Frederik Pietschmann, and Bärbel Schlake. Nonlinear Cross-Diffusion with Size Exclusion. *SIAM Journal on Mathematical Analysis*, 42(6):2842–2871, January 2010.

[39] Kevin Buzzard. What makes a mathematician tick?(invited talk). In *10th International Conference on Interactive Theorem Proving (ITP 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[40] Kevin Buzzard. Proving theorems with computers. *Notices of the American Mathematical Society*, 67(11):1, 2020.

[41] Kevin Buzzard, Johan Commelin, and Patrick Massot. Formalising perfectoid spaces. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 299–312, 2020.

[42] Christopher I. Byrnes and Alberto Isidori. New results and examples in nonlinear feedback stabilization. *Systems Control Lett.*, 12(5):437–442, 1989.

[43] Jean Cauvin-Vila, Virginie Ehrlacher, and Amaury Hayat. Boundary stabilization of one-dimensional cross-diffusion systems in a moving domain: linearized system. *Journal of Differential Equations*, 350:251–307, 2023.

[44] Antoine Chaillet, Iasson Karafyllis, Pierdomenico Pepe, and Yuan Wang. The iss framework for time-delay systems: a survey. *arXiv preprint arXiv:2206.06167*, 2022.

[45] François Charton. Linear algebra with transformers. *arXiv preprint arXiv:2112.01898*, 2021.

[46] Francois Charton, Amaury Hayat, and Guillaume Lample. Learning advanced mathematical computations from examples. In *International Conference on Learning Representations*, 2020.

[47] François Charton, Amaury Hayat, Sean T McQuade, Nathaniel J Merrill, and Benedetto Piccoli. A deep language model to predict metabolic network equilibria. *arXiv preprint arXiv:2112.03588*, 2021.

[48] Jixun Chu, Peipei Shang, and Zhiqiang Wang. Controllability and stabilization of a conservation law modeling a highly re-entrant manufacturing system. *Nonlinear Analysis*, 189:111577, 2019.

[49] Giuseppe Maria Coclite, Nicola De Nitti, Alexander Keimer, and Lukas Pflug. Singular limits with vanishing viscosity for nonlocal conservation laws. *Nonlinear Analysis*, 211:112370, 2021.

[50] GM Coclite, N De Nitti, A Keimer, and L Pflug. On existence and uniqueness of weak solutions to nonlocal conservation laws with BV kernels. *Z. Angew. Math. Phys., to appear. Also available on ResearchGate*, 2021.

[51] Rinaldo M Colombo, Paola Goatin, and Massimiliano D Rosini. On the modelling and management of traffic. *ESAIM: Mathematical Modelling and Numerical Analysis*, 45(5):853–872, 2011.

[52] Johan Commelin and Robert Y Lewis. Formalizing the ring of witt vectors. In *Proceedings of the 10th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 264–277, 2021.

[53] Jean-Michel Coron. Local controllability of a 1-D tank containing a fluid modeled by the shallow water equations. *ESAIM Control Optim. Calc. Var.*, 8:513–554, 2002. A tribute to J. L. Lions.

[54] Jean-Michel Coron. Some open problems on water tank control systems. In *Elliptic and Parabolic Problems*, pages 179–188. Springer, 2005.

[55] Jean-Michel Coron. *Control and nonlinearity*, volume 136 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2007.

[56] Jean-Michel Coron. *Stabilisation en temps fini*, 2017.

[57] Jean-Michel Coron and Georges Bastin. Dissipative boundary conditions for one-dimensional quasi-linear hyperbolic systems: Lyapunov stability for the C^1-norm. *SIAM J. Control Optim.*, 53(3):1464–1483, 2015.

[58] Jean-Michel Coron and Georges Bastin. Dissipative boundary conditions for one-dimensional quasi-linear hyperbolic systems: Lyapunov stability for the C^1-norm. *SIAM Journal on Control and Optimization*, 53(3):1464–1483, 2015.

[59] Jean-Michel Coron, Georges Bastin, and Brigitte d'Andréa Novel. Dissipative boundary conditions for one-dimensional nonlinear hyperbolic systems. *SIAM Journal on Control and Optimization*, 47(3):1460–1498, 2008.

[60] Jean-Michel Coron and Brigitte d'Andréa Novel. Stabilization of a rotating body beam without damping. *IEEE Trans. Automat. Control*, 43(5):608–618, 1998.

[61] Jean-Michel Coron, Brigitte d'Andréa Novel, and Georges Bastin. A Lyapunov approach to control irrigation canals modeled by Saint-Venant equations. In *CD-Rom Proceedings, Paper F1008-5, ECC99, Karlsruhe, Germany*, pages 3178–3183, 1999.

[62] Jean-Michel Coron, Sylvain Ervedoza, Shyam Sundar Ghoshal, Olivier Glass, and Vincent Perrollaz. Dissipative boundary conditions for 2×2 hyperbolic systems of conservation laws for entropy solutions in BV. *J. Differential Equations*, 262(1):1–30, 2017.

[63] Jean-Michel Coron, Ludovick Gagnon, and Morgan Morancey. Rapid stabilization of a linearized bilinear 1-D Schrödinger equation. *J. Math. Pures Appl. (9)*, 115:24–73, 2018.

[64] Jean-Michel Coron, Amaury Hayat, Shengquan Xiang, and Christophe Zhang. Stabilization of the linearized water tank system. *Arch. Ration. Mech. Anal.*, 244(3):1019–1097, 2022.

[65] Jean-Michel Coron, Long Hu, and Guillaume Olive. Finite-time boundary stabilization of general linear hyperbolic balance laws via fredholm backstepping transformation. *Automatica*, 84:95–100, 2017.

[66] Jean-Michel Coron, Long Hu, Guillaume Olive, and Peipei Shang. Boundary stabilization in finite time of one-dimensional linear hyperbolic balance laws with coefficients depending on time and space. *Journal of Differential Equations*, 271:1109–1170, 2021.

[67] Jean-Michel Coron, Matthias Kawski, and Zhiqiang Wang. Analysis of a conservation law modeling a highly re-entrant manufacturing system. *Discrete Contin. Dyn. Syst. Ser. B*, 14(4):1337–1359, 2010.

[68] Jean-Michel Coron and Qi Lü. Local rapid stabilization for a korteweg–de vries equation with a neumann boundary control on the right. *Journal de Mathématiques Pures et Appliquées*, 102(6):1080–1120, 2014.

[69] Jean-Michel Coron and Qi Lü. Fredholm transform and local rapid stabilization for a Kuramoto-Sivashinsky equation. *J. Differential Equations*, 259(8):3683–3729, 2015.

[70] Jean-Michel Coron and Hoai-Minh Nguyen. Dissipative boundary conditions for nonlinear 1-D hyperbolic systems: sharp conditions through an approach via time-delay systems. *SIAM J. Math. Anal.*, 47(3):2220–2240, 2015.

[71] Jean-Michel Coron and Hoai-Minh Nguyen. Null controllability and finite time stabilization for the heat equations with variable coefficients in space in one dimension via backstepping approach. *Arch. Ration. Mech. Anal.*, 225(3):993–1023, 2017.

[72] Jean-Michel Coron, Hoai-Minh Nguyen, and Armand Koenig. Lack of local controllability for a water-tank system when the time is not large enough. working paper or preprint, February 2022.

[73] Jean-Michel Coron, Rafael Vazquez, Miroslav Krstic, and Georges Bastin. Local exponential H^2 stabilization of a 2×2 quasilinear hyperbolic system using backstepping. *SIAM J. Control Optim.*, 51(3):2005–2035, 2013.

[74] Jean-Michel Coron and Zhiqiang Wang. Output feedback stabilization for a scalar conservation law with a nonlocal velocity. *SIAM J. Math. Anal.*, 45(5):2646–2665, 2013.

[75] Jean-Michel Coron and Shengquan Xiang. Small-time global stabilization of the viscous Burgers equation with three scalar controls. *J. Math. Pures Appl. (9)*, 151:212–256, 2021.

[76] Shumo Cui, Benjamin Seibold, Raphael Stern, and Daniel B Work. Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1336–1341. IEEE, 2017.

[77] Constantine M Dafermos. Polygonal approximations of solutions of the initial value problem for a conservation law. *Journal of mathematical analysis and applications*, 38(1):33–41, 1972.

[78] Constantine M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 2010.

[79] Chiara Daini, Paola Goatin, Maria Laura Delle Monache, and Antonella Ferrara. Centralized traffic control via small fleets of connected and automated vehicles. In *ECC 2022-European Control Conference*, 2022.

[80] Brigitte d'Andréa Novel, Iván Moyano, and Lionel Rosier. Finite-time stabilization of an overhead crane with a flexible cable. *Math. Control Signals Systems*, 31(2):Art. 6, 19, 2019.

[81] Sergey Dashkovskiy and Andrii Mironchenko. Input-to-state stability of infinite-dimensional control systems. *Math. Control Signals Systems*, 25(1):1–35, 2013.

[82] Sergey Dashkovskiy and Vitalii Slyn'ko. Robust stability of a perturbed nonlinear wave equation. *IFAC-PapersOnLine*, 53(2):3168–3173, 2020.

[83] Martin Davis and Hilary Putnam. A computing procedure for quantification theory. *Journal of the ACM (JACM)*, 7(3):201–215, 1960.

[84] Jonathan de Halleux, Christophe Prieur, Jean-Michel Coron, Brigitte d'Andréa Novel, and Georges Bastin. Boundary feedback control in networks of open channels. *Automatica. A Journal of IFAC, the International Federation of Automatic Control*, 39(8):1365–1376, 2003.

[85] Belhassen Dehman, Gilles Lebeau, and Enrique Zuazua. Stabilization and control for the subcritical semilinear wave equation. *Ann. Sci. École Norm. Sup. (4)*, 36(4):525–551, 2003.

[86] Marcello Delitala and Andrea Tosin. Mathematical modeling of vehicular traffic: a discrete kinetic theory approach. *Mathematical Models and Methods in Applied Sciences*, 17(06):901–932, 2007.

[87] Maria Laura Delle Monache and Paola Goatin. A front tracking method for a strongly coupled PDE-ODE system with moving density constraints in traffic flow. *Discrete Contin. Dyn. Syst. Ser. S*, 7(3):435–447, 2014.

[88] Maria Laura Delle Monache and Paola Goatin. Scalar conservation laws with moving constraints arising in traffic flow modeling: an existence result. *Journal of Differential equations*, 257(11):4015–4029, 2014.

[89] Maria Laura Delle Monache, Thibault Liard, Anaïs Rat, Raphael Stern, Rahul Bhadani, Benjamin Seibold, Jonathan Sprinkle, Daniel B Work, and Benedetto Piccoli. Feedback control algorithms for the dissipation of traffic waves with autonomous vehicles. In *Computational Intelligence and Optimization Methods for Control Engineering*, pages 275–299. Springer, 2019.

[90] Vincent Deluz. De la clepsydre animée à l'horloge mécanique à automates, entre antiquité et moyen âge. In *Autour des machines de Vitruve. L'ingénierie romaine: textes, archéologie et restitution.*, pages 243–pages. Presses universitaires de Caen, 2015.

[91] Florent Di Meglio, Rafael Vazquez, and Miroslav Krstic. Stabilization of a system of $n+1$ coupled first-order hyperbolic linear PDEs with a single boundary input. *IEEE Trans. Automat. Control*, 58(12):3097–3111, 2013.

[92] Ababacar Diagne, Mamadou Diagne, Shuxia Tang, and Miroslav Krstic. Backstepping stabilization of the linearized *saint-venant-exner* model. *Automatica J. IFAC*, 76:345–354, 2017.

[93] Mamadou Diagne, Peipei Shang, and Zhiqiang Wang. Feedback stabilization for the mass balance equations of an extrusion process. *IEEE Trans. Automat. Control*, 61(3):760–765, 2016.

[94] Markus Dick, Martin Gugat, and Günter Leugering. Classical solutions and feedback stabilization for the gas flow in a sequence of pipes. *Netw. Heterog. Media*, 5(4):691–709, 2010.

[95] Markus Dick, Martin Gugat, and Günter Leugering. A strict H^1-Lyapunov function and feedback stabilization for the isothermal Euler equations with friction. *Numer. Algebra Control Optim.*, 1(2):225–244, 2011.

[96] Floris van Doorn, Gabriel Ebner, and Robert Y Lewis. Maintaining a library of formal mathematics. In *International Conference on Intelligent Computer Mathematics*, pages 251–267. Springer, 2020.

[97] Floris van Doorn, Jakob von Raumer, and Ulrik Buchholtz. Homotopy type theory in lean. In *International Conference on Interactive Theorem Proving*, pages 479–495. Springer, 2017.

[98] François Dubois, Nicolas Petit, and Pierre Rouchon. Motion planning and nonlinear simulations for a tank containing a fluid. In *1999 European Control Conference (ECC)*, pages 3232–3237. IEEE, 1999.

[99] Mathias Dus, Francesco Ferrante, and Christophe Prieur. On L^∞ stabilization of diagonal semilinear hyperbolic systems by saturated boundary control. *ESAIM Control Optim. Calc. Var.*, 26:Paper No. 23, 34, 2020.

[100] Nicolas Espitia, Jean Auriol, Huan Yu, and Miroslav Krstic. Traffic flow control on cascaded roads by event-triggered output feedback. *Internat. J. Robust Nonlinear Control*, 32(10):5919–5949, 2022.

[101] Nicolas Espitia, Jean Auriol, Huan Yu, and Miroslav Krstic. Traffic flow control on cascaded roads by event-triggered output feedback. *International Journal of Robust and Nonlinear Control*, 2022.

[102] Caroline Fabre, Jean-Pierre Puel, and Enrike Zuazua. Approximate controllability of the semilinear heat equation. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 125(1):31–61, 1995.

[103] Shimao Fan, Michael Herty, and Benjamin Seibold. Comparative model accuracy of a data-fitted generalized aw-rascle-zhang model. *Networks & Heterogeneous Media*, 9(2), 2014.

[104] Shimao Fan, Ye Sun, Benedetto Piccoli, Benjamin Seibold, and Daniel B Work. A collapsed generalized aw-rascle-zhang model and its model accuracy. *arXiv preprint arXiv:1702.03624*, 2017.

[105] Enrique Fernández-Cara and Enrique Zuazua. The cost of approximate controllability for heat equations: the linear case. *Advances in Differential equations*, 5(4-6):465–514, 2000.

[106] Francesco Ferrante and Christophe Prieur. Boundary control design for conservation laws in the presence of measurement disturbances. *Mathematics of Control, Signals, and Systems*, 33(1):49–77, 2021.

[107] Morris R Flynn, Aslan R Kasimov, J-C Nave, Rodolfo R Rosales, and Benjamin Seibold. Self-sustained nonlinear waves in traffic flow. *Physical Review E*, 79(5):056113, 2009.

[108] AV Fursikov and O Yu Imanuvilov. Controllability of evolution equations, lect. *Notes Ser*, 34, 1996.

[109] Ludovick Gagnon, Amaury Hayat, Shengquan Xiang, and Christophe Zhang. Fredholm transformation on laplacian and rapid stabilization for the heat equation. *arXiv preprint arXiv:2110.04028*, 2021.

[110] Ludovick Gagnon, Amaury Hayat, Shengquan Xiang, and Christophe Zhang. Fredholm backstepping for critical operators and application to rapid stabilization for the linearized water waves. *arXiv preprint arXiv:2202.08321*, 2022.

[111] Ludovick Gagnon, Pierre Lissy, and Swann Marx. A fredholm transformation for the rapid stabilization of a degenerate parabolic equation. *SIAM Journal on Control and Optimization*, 59(5):3828–3859, 2021.

[112] Mauro Garavello, Paola Goatin, Thibault Liard, and Benedetto Piccoli. A multiscale model for traffic regulation via autonomous vehicles. *Journal of Differential Equations*, 269(7):6088–6124, 2020.

[113] Denos C Gazis, Robert Herman, and Richard W Rothery. Nonlinear follow-the-leader models of traffic flow. *Operations research*, 9(4):545–567, 1961.

[114] P. C. Gilmore. A proof method for quantification theory: Its justification and realization. *IBM J. Res. Dev.*, 4(1):28–35, jan 1960.

[115] Paul C Gilmore. A proof method for quantification theory: Its justification and realization. *IBM Journal of research and development*, 4(1):28–35, 1960.

[116] Olivier Glass. Some questions of control in fluid mechanics. In *Control of Partial Differential Equations*, pages 131–206. Springer, 2012.

[117] Olivier Glass. On the controllability of the non-isentropic 1-D Euler equation. *J. Differential Equations*, 257(3):638–719, 2014.

[118] James Glimm. Solutions in the large for nonlinear hyperbolic systems of equations. *Comm. Pure Appl. Math.*, 18:697–715, 1965.

[119] Paola Goatin, Simone Göttlich, and Oliver Kolb. Speed limit and ramp meter control for traffic flow networks. *Engineering Optimization*, 48(7):1121–1144, 2016.

[120] Paola Goatin and Nicolas Laurent-Brouty. The zero relaxation limit for the aw–rascle–zhang traffic flow model. *Zeitschrift für angewandte Mathematik und Physik*, 70(1):1–24, 2019.

[121] François Golse. *Distributions, analyse de Fourier, équations aux dérivées partielles*. École polytechnique, 2008.

[122] François Golse. On the dynamics of large particle systems in the mean field limit. In *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, pages 1–144. Springer, 2016.

[123] Xiaoqian Gong and Alexander Keimer. On the well-posedness of the "bando-follow the leader" car following model and a time-delayed version". *Preprint*, 2022. Researchgate DOI: 10.13140/RG.2.2.22507.62246.

[124] Xiaoqian Gong, Benedetto Piccoli, and Giuseppe Visconti. Mean-field limit of a hybrid system for multi-lane multi-class traffic. *arXiv preprint arXiv:2007.14655*, 2020.

[125] Xiaoqian Gong, Benedetto Piccoli, and Giuseppe Visconti. Mean-field of optimal control problems for hybrid model of multilane traffic. *IEEE Control Systems Letters*, 2020.

[126] J. M. Greenberg and Tatsien Li. The effect of boundary damping for the quasilinear wave equation. *J. Differential Equations*, 52(1):66–75, 1984.

[127] Martin Gugat, Markus Dick, and Günter Leugering. Gas flow in fan-shaped networks: classical solutions and feedback stabilization. *SIAM J. Control Optim.*, 49(5):2101–2117, 2011.

[128] Martin Gugat and Michaël Herty. Existence of classical solutions and feedback stabilization for the flow in gas networks. *ESAIM Control Optim. Calc. Var.*, 17(1):28–51, 2011.

[129] Martin Gugat, Michael Herty, Axel Klar, and Günter Leugering. Optimal control for traffic flow networks. *Journal of optimization theory and applications*, 126(3):589–616, 2005.

[130] Martin Gugat, Günter Leugering, and Ke Wang. Neumann boundary feedback stabilization for a nonlinear wave equation: A strict H^2-Lyapunov function. *Math. Control Relat. Fields*, 7(3):419–448, 2017.

[131] Alena Gusakov, Bhavik Mehta, and Kyle A Miller. Formalizing hall's marriage theorem in lean. *arXiv preprint arXiv:2101.00127*, 2021.

[132] Amaury Hayat. Boundary stability of 1-D nonlinear inhomogeneous hyperbolic systems for the C^1 norm. *SIAM J. Control Optim.*, 57(6):3603–3638, 2019.

[133] Amaury Hayat. On boundary stability of inhomogeneous 2×2 1-D hyperbolic systems for the C^1 norm. *ESAIM Control Optim. Calc. Var.*, 25:Paper No. 82, 31, 2019.

[134] Amaury Hayat. Boundary stabilization of 1D hyperbolic systems. *Annu. Rev. Control*, 52:222–242, 2021.

[135] Amaury Hayat. Global exponential stability and input-to-state stability of semilinear hyperbolic systems for the L^2 norm. *Systems Control Lett.*, 148:Paper No. 104848, 8, 2021.

[136] Amaury Hayat. PI controllers for the general Saint-Venant equations. *Journal de l'École polytechnique—Mathématiques*, 9:1431–1472, 2022.

[137] Amaury Hayat, Thibault Liard, Francesca Marcellini, and Benedetto Piccoli. A multiscale second order model for the interaction between AV and traffic flows: analysis and existence of solutions. working paper or preprint, January 2021.

[138] Amaury Hayat, Benedetto Piccoli, and Sydney Truong. Dissipation of traffic jams using a single autonomous vehicle on a ring road. *SIAM Journal on Applied Mathematics*, 2023.

[139] Amaury Hayat and Peipei Shang. A quadratic Lyapunov function for Saint-Venant equations with arbitrary friction and space-varying slope. *Automatica J. IFAC*, 100:52–60, 2019.

[140] Andreas Hegyi, Bart De Schutter, and Johannes Hellendoorn. Optimal coordination of variable speed limits to suppress shock waves. *IEEE Transactions on intelligent transportation systems*, 6(1):102–112, 2005.

[141] Michael Herty and Lorenzo Pareschi. Fokker-planck asymptotics for traffic flow models. *Kinetic & Related Models*, 3(1):165, 2010.

[142] Long Hu, Florent Di Meglio, Rafael Vazquez, and Miroslav Krstic. Control of homodirectional and general heterodirectional linear coupled hyperbolic PDEs. *IEEE Trans. Automat. Control*, 61(11):3301–3314, 2016.

[143] Long Hu, Rafael Vazquez, Florent Di Meglio, and Miroslav Krstic. Boundary exponential stabilization of 1-dimensional inhomogeneous quasi-linear hyperbolic systems. *SIAM Journal on Control and Optimization*, 57(2):963–998, 2019.

[144] Mojtaba Izadi, Javad Abdollahi, and Stevan S. Dubljevic. PDE backstepping control of one-dimensional heat equation with time-varying domain. *Automatica*, 54:41–48, April 2015.

[145] Mojtaba Izadi and Stevan Dubljevic. Backstepping output-feedback control of moving boundary parabolic pdes. *European Journal of Control*, 21:27–35, 2015.

[146] Stéphane Jaffard, Marius Tucsnak, and Enrique Zuazua. Singular internal stabilization of the wave equation. *journal of differential equations*, 145(1):184–215, 1998.

[147] Ansgar Jüngel. The boundedness-by-entropy principle for cross-diffusion systems. *Nonlinearity*, 28(6):1963–2001, June 2015.

[148] Ansgar Jüngel and Ines Viktoria Stelzer. Entropy structure of a cross-diffusion tumor-growth model. *Mathematical Models and Methods in Applied Sciences*, 22(07):1250009, July 2012.

[149] Ansgar Jüngel and Ines Viktoria Stelzer. Existence Analysis of Maxwell–Stefan Systems for Multicomponent Mixtures. *SIAM Journal on Mathematical Analysis*, 45(4):2421–2440, January 2013.

[150] Łukasz Kaiser and Ilya Sutskever. Neural gpus learn algorithms. *arXiv preprint arXiv:1511.08228*, 2015.

[151] Iasson Karafyllis and Miroslav Krstic. *Input-to-State Stability for PDEs*. Springer, 2018.

[152] Nicolas Kardous, Amaury Hayat, Sean T McQuade, Xiaoqian Gong, Sydney Truong, Tinhinane Mezair, Paige Arnold, Ryan Delorenzo, Alexandre Bayen, and Benedetto Piccoli. A rigorous multi-population multi-lane hybrid traffic model for dissipation of waves via autonomous vehicles. *The European Physical Journal Special Topics*, pages 1–12, 2022.

[153] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[154] Arne Kesting, Martin Treiber, and Dirk Helbing. General lane-changing model MOBIL for car-following models. *Transportation Research Record*, 1999(1):86–94, 2007.

[155] David Kleinman. An easy way to stabilize a linear constant system. *IEEE Transactions on Automatic Control*, 15(6):692–692, 1970.

[156] Donald E Knuth and Ronald W Moore. An analysis of alpha-beta pruning. *Artificial intelligence*, 6(4):293–326, 1975.

[157] Daniel E Koditschek. Adaptive techniques for mechanical systems. *Proc. 5th. Yale University Conference*, page pp. 259–265, 1987.

[158] Oliver Kolb, Simone Göttlich, and Paola Goatin. Capacity drop and traffic control for a second order traffic model. *Networks and Heterogeneous Media*, 12(4):663–681, 2017.

[159] Vilmos Komornik. An equiconvergence theorem for the Schrödinger operator. *Acta Math. Hungar.*, 44(1-2):101–114, 1984.

[160] Vilmos Komornik. On the equiconvergence of eigenfunction expansions associated with ordinary linear differential operators. *Acta Math. Hungar.*, 47(1-2):261–280, 1986.

[161] Miroslav Krstic and Andrey Smyshlyaev. Backstepping boundary control for first-order hyperbolic PDEs and application to systems with actuator and sensor delays. *Systems Control Lett.*, 57(9):750–758, 2008.

[162] Miroslav Krstic and Andrey Smyshlyaev. *Boundary Control of PDEs: A Course on Backstepping Designs*, volume 16 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[163] Stanislav N Kružkov. First order quasilinear equations in several independent variables. *Mathematics of the USSR-Sbornik*, 10(2):217, 1970.

[164] Guillaume Lample and François Charton. Deep learning for symbolic mathematics. In *International Conference on Learning Representations*, 2019.

[165] Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. Word translation without parallel data. In *International Conference on Learning Representations*, 2018.

[166] Guillaume Lample, Marie-Anne Lachaux, Thibaut Lavril, Xavier Martinet, Amaury Hayat, Gabriel Ebner, Aurélien Rodriguez, and Timothée Lacroix. Hypertree proof search for neural theorem proving. *to appear in Advances in neural information processing systems*, 2022.

[167] D. Lannes. Well-posedness of the water-waves equations. *J. Am. Math. Soc.*, 18(3):605–654, 2005.

[168] D. Lannes. *The water waves problem. Mathematical analysis and asymptotics*, volume 188. Providence, RI: American Mathematical Society (AMS), 2013.

[169] Nicolas Laurent-Brouty, Guillaume Costeseque, and Paola Goatin. A macroscopic traffic flow model accounting for bounded acceleration. *SIAM Journal on Applied Mathematics*, 81(1):173–189, 2021.

[170] Peter D. Lax. *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.

[171] Gilles Lebeau and Luc Robbiano. Contrôle exact de léquation de la chaleur. *Communications in Partial Differential Equations*, 20(1-2):335–356, 1995.

[172] Marta Lewicka. Well-posedness for hyperbolic systems of conservation laws with large BV data. *Arch. Ration. Mech. Anal.*, 173(3):415–445, 2004.

[173] Tatsien Li. *Controllability and observability for quasilinear hyperbolic systems*, volume 3 of *AIMS Series on Applied Mathematics*. American Institute of Mathematical Sciences (AIMS), Springfield, MO; Higher Education Press, Beijing, 2010.

[174] Tatsien Li and Wen Ci Yu. *Boundary value problems for quasilinear hyperbolic systems*. Duke University Mathematics Series, V. Duke University, Mathematics Department, Durham, NC, 1985.

[175] Thibault Liard, Ismaïla Balogoun, Swann Marx, and Franck Plestan. Boundary sliding mode control of a system of linear hyperbolic equations: a lyapunov approach. *Automatica*, 135:109964, 2022.

[176] Thibault Liard and Benedetto Piccoli. Well-posedness for scalar conservation laws with moving flux constraints. *SIAM Journal on Applied Mathematics*, 79(2):641–667, 2019.

[177] Thibault Liard and Benedetto Piccoli. On entropic solutions to conservation laws coupled with moving bottlenecks. *Communications in mathematical sciences*, 2021.

[178] Thibault Liard, Raphael Stern, and Maria Laura Delle Monache. A PDE-ODE model for traffic control with autonomous vehicles. 2020. working paper or preprint.

[179] Nathan Lichtlé, Eugene Vinitsky, George Gunter, Akash Velu, and Alexandre M Bayen. Fuel consumption reduction of multi-lane road networks using decentralized mixed-autonomy control. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 2068–2073. IEEE, 2021.

[180] Mark Lichtner. Spectral mapping theorem for linear hyperbolic systems. *Proc. Amer. Math. Soc.*, 136(6):2091–2101, 2008.

[181] Michael James Lighthill and Gerald Beresford Whitham. On kinematic waves ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 229(1178):317–345, 1955.

[182] Wei-Jiu Liu and Enrique Zuazua. Decay rates for dissipative wave equations. *Ricerche di Matematica*, 48(240):61–75, 1999.

[183] Gabriela López Ruiz. Boundary layer formation in the quasigeostrophic model near nonperiodic rough coasts. *arXiv preprint arXiv:2108.02705*, 2021.

[184] Dahlard L Lukes. Stabilizability and optimal control. *Funkcial. Ekvac*, 11(39-50):17, 1968.

[185] Sophie Marbach, Hiroaki Yoshida, and Lydéric Bocquet. Osmotic and diffusio-osmotic flow generation at high solute concentration. I. mechanical approaches. *The Journal of Chemical Physics*, 146(19):194701, 2017.

[186] Patrick Massot. Why formalize mathematics? Technical report, University Paris-Saclay, 2021.

[187] Frédéric Mazenc and Christophe Prieur. Strict lyapunov functions for semilinear parabolic partial differential equations. *Mathematical Control and Related Fields*, 1(2):231–250, 2011.

[188] Norman Megill and David A Wheeler. *Metamath: a computer language for mathematical proofs*. Lulu. com, 2019.

[189] Nicolas Minorsky. Directional stability of automatically steered bodies. *Naval Engineers Journal*, 32(2), 1922.

[190] Andrii Mironchenko, Iasson Karafyllis, and Miroslav Krstic. Monotonicity methods for input-to-state stability of nonlinear parabolic PDEs with boundary disturbances. *SIAM J. Control Optim.*, 57(1):510–532, 2019.

[191] Isao Miyadera. *Nonlinear semigroups*, volume 109 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1992. Translated from the 1977 Japanese original by Choong Yun Cho.

[192] Leonardo de Moura, Soonho Kong, Jeremy Avigad, Floris van Doorn, and Jakob von Raumer. The lean theorem prover (system description). In *International Conference on Automated Deduction*, pages 378–388. Springer, 2015.

[193] Arun Nair, Praveen Srinivasan, Sam Blackwell, Cagdas Alcicek, Rory Fearon, Alessandro De Maria, Vedavyas Panneershelvam, Mustafa Suleyman, Charles Beattie, Stig Petersen, et al. Massively parallel methods for deep reinforcement learning. *arXiv preprint arXiv:1507.04296*, 2015.

[194] Aloisio Freiria Neves, Hermano de Souza Ribeiro, and Orlando Lopes. On the spectrum of evolution operators generated by hyperbolic systems. *J. Funct. Anal.*, 67(3):320–344, 1986.

[195] JM Nieto-Villar, R Quintana, and J Rieumont. Entropy production rate as a lyapunov function in chemical systems: Proof. *Physica Scripta*, 68(3):163, 2003.

[196] Markos Papageorgiou, Habib Hadj-Salem, and F Middelham. Alinea local ramp metering: Summary of field results. *Transportation research record*, 1603(1):90–98, 1997.

[197] SL Paveri-Fontana. On boltzmann-like treatments for traffic flow: a critical review of the basic model and an alternative proposal for dilute traffic analysis. *Transportation research*, 9(4):225–235, 1975.

[198] Vincent Perrollaz. Asymptotic stabilization of entropy solutions to scalar conservation laws through a stationary feedback law. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 30(5):879–915, 2013.

[199] Vincent Perrollaz. Asymptotic stabilization of stationnary shock waves using a boundary feedback law. *arXiv preprint arXiv:1801.06335*, 2018.

[200] Nicolas Petit and Pierre Rouchon. Dynamics and solutions to some control problems for water-tank systems. *IEEE Trans. Automat. Control*, 47(4):594–609, 2002.

[201] Giulia Piacentini, Paola Goatin, and Antonella Ferrara. Traffic control via platoons of intelligent vehicles for saving fuel consumption in freeway systems. *IEEE Control Systems Letters*, 5(2):593–598, 2020.

[202] Hannah Pohlmann and Benjamin Seibold. Simple control options for an vehicle used to dissipate traffic waves. Technical report, 2015.

[203] Stanislas Polu, Jesse Michael Han, Kunhao Zheng, Mantas Baksys, Igor Babuschkin, and Ilya Sutskever. Formal mathematics statement curriculum learning. *arXiv preprint arXiv:2202.01344*, 2022.

[204] Stanislas Polu and Ilya Sutskever. Generative language modeling for automated theorem proving. *arXiv preprint arXiv:2009.03393*, 2020.

[205] Andrey Polyakov, Jean-Michel Coron, and Lionel Rosier. On boundary finite-time feedback control for heat equation. *IFAC-PapersOnLine*, 50(1):671–676, 2017.

[206] Christophe Prieur and Jonathan de Halleux. Stabilization of a 1-d tank containing a fluid modeled by the shallow water equations. *Systems & Control Letters*, 52(3):167–178, 2004.

[207] Christophe Prieur and Frédéric Mazenc. ISS-Lyapunov functions for time-varying hyperbolic systems of balance laws. *Mathematics of Control, Signals, and Systems*, 24(1-2):111–134, 2012.

[208] Christophe Prieur and Frédéric Mazenc. ISS-Lyapunov functions for time-varying hyperbolic systems of balance laws. *Math. Control Signals Systems*, 24(1-2):111–134, 2012.

[209] Ilya Prigogine and Robert Herman. *Kinetic theory of vehicular traffic.* American Elsevier Publishing Co., New York, 1971.

[210] Jie Qi, Shurong Mo, and Miroslav Krstic. Delay-compensated distributed pde control of traffic with connected/automated vehicles. *IEEE Transactions on Automatic Control*, 2022.

[211] Tie Hu Qin. Global smooth solutions of dissipative boundary value problems for first order quasilinear hyperbolic systems. *Chinese Ann. Math. Ser. B*, 6(3):289–298, 1985. A Chinese summary appears in Chinese Ann. Math. Ser. A **6** (1985), no. 4, 514.

[212] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. Technical report, Open-AI, 2019.

[213] Rabie Ramadan, Rodolfo Ruben Rosales, and Benjamin Seibold. Structural properties of the stability of jamitons. In *Mathematical Descriptions of Traffic Flow: Micro, Macro and Kinetic Models*, pages 35–62. Springer, 2021.

[214] Michael Renardy. On the type of certain C_0-semigroups. *Comm. Partial Differential Equations*, 18(7-8):1299–1307, 1993.

[215] Paul I Richards. Shock waves on the highway. *Operations research*, 4(1):42–51, 1956.

[216] David L. Russell. Canonical forms and spectral determination for a class of hyperbolic distributed parameter control systems. *J. Math. Anal. Appl.*, 62(1):186–225, 1978.

[217] Peter Scholze. Perfectoid spaces. *Publications mathématiques de l'IHÉS*, 116(1):245–313, 2012.

[218] Peter Scholze. Liquid tensor experiment. *Experimental Mathematics*, pages 1–6, 2021.

[219] Benjamin Seibold, Morris R Flynn, Aslan R Kasimov, and Rodolfo R Rosales. Constructing set-valued fundamental diagrams from jamiton solutions in second order traffic models. *Networks & Heterogeneous Media*, 8(3):745, 2013.

[220] Peipei Shang. Cauchy problem for multiscale conservation laws: Application to structured cell populations. *Journal of Mathematical Analysis and applications*, 401(2):896–920, 2013.

[221] Peipei Shang and Zhiqiang Wang. Analysis and control of a scalar conservation law modeling a highly re-entrant manufacturing system. *Journal of Differential Equations*, 250(2):949–982, 2011.

[222] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[223] Silvia Siri, Cecilia Pasquale, Simona Sacone, and Antonella Ferrara. Freeway traffic control: A survey. *Automatica*, 130:109655, 2021.

[224] Andrey Smyshlyaev and Miroslav Krstic. Closed-form boundary state feedbacks for a class of 1-D partial integro-differential equations. *IEEE Trans. Automat. Control*, 49(12):2185–2202, 2004.

[225] Eduardo D. Sontag. Smooth stabilization implies coprime factorization. *IEEE Trans. Automat. Control*, 34(4):435–443, 1989.

[226] Raphael E Stern, Shumo Cui, Maria Laura Delle Monache, Rahul Bhadani, Matt Bunting, Miles Churchill, Nathaniel Hamilton, Hannah Pohlmann, Fangyu Wu, Benedetto Piccoli, et al. Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. *Transportation Research Part C: Emerging Technologies*, 89:205–221, 2018.

[227] Yuki Sugiyama, Minoru Fukui, Macoto Kikuchi, Katsuya Hasebe, Akihiro Nakayama, Katsuhiro Nishinari, Shin-ichi Tadaki, and Satoshi Yukawa. Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam. *New journal of physics*, 10(3):033001, 2008.

[228] Xiaotian Sun, Laura Muñoz, and Roberto Horowitz. Highway traffic state estimation using improved mixture kalman filters for effective ramp metering control. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, volume 6, pages 6333–6338. IEEE, 2003.

[229] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[230] Eitan Tadmor and Changhui Tan. Critical thresholds in flocking hydrodynamics with non-local alignment. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20130401, 22, 2014.

[231] Alireza Talebpour and Hani S Mahmassani. Influence of connected and autonomous vehicles on traffic flow stability and throughput. *Transportation Research Part C: Emerging Technologies*, 71:143–163, 2016.

[232] Aneel Tanwani, Christophe Prieur, and Sophie Tarbouriech. Disturbance-to-state stabilization and quantized control for linear hyperbolic systems. *arXiv preprint arXiv:1703.00302*, 2017.

[233] Andrew Trask, Felix Hill, Scott E Reed, Jack Rae, Chris Dyer, and Phil Blunsom. Neural arithmetic logic units. *Advances in neural information processing systems*, 31, 2018.

[234] John Tsinias. Sufficient Lyapunov-like conditions for stabilization. *Math. Control Signals Systems*, 2(4):343–357, 1989.

[235] Liudmila Tumash, Carlos Canudas-de Wit, and Maria Laura Delle Monache. Boundary control design for traffic with nonlinear dynamics. *IEEE Trans. Automat. Control*, 67(3):1301–1313, 2022.

[236] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[237] Rafael Vazquez, Jean-Michel Coron, Miroslav Krstic, and Georges Bastin. Local exponential H^2 stabilization of a 2×2 quasilinear hyperbolic system using backstepping. *50th IEEE Conference on Decision and Control and European Control Conference, Orlando*, pages 1329–1334, 2011.

[238] Eugene Vinitsky, Nathan Lichtle, Kanaad Parvate, and Alexandre Bayen. Optimizing mixed autonomy traffic flow with decentralized autonomous vehicles and multi-agent rl. *arXiv preprint arXiv:2011.00120*, 2020.

[239] Eugene Vinitsky, Kanaad Parvate, Aboudy Kreidieh, Cathy Wu, and Alexandre Bayen. Lagrangian control through deep-rl: Applications to bottleneck decongestion. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 759–765. IEEE, 2018.

[240] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018, 2018.

[241] Jiawei Wang, Yang Zheng, Qing Xu, Jianqiang Wang, and Keqiang Li. Controllability analysis and optimal control of mixed traffic flow with human-driven and autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[242] Meng Wang, Winnie Daamen, Serge P Hoogendoorn, and Bart van Arem. Cooperative car-following control: Distributed algorithm and impact on moving jam features. *IEEE Transactions on Intelligent Transportation Systems*, 17(5):1459–1471, 2015.

[243] Zhengwei Wang, Qi She, and Tomas E Ward. Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR)*, 54(2):1–38, 2021.

[244] Zhiqiang Wang. Exact controllability for nonautonomous first order quasilinear hyperbolic systems. *Chinese Ann. Math. Ser. B*, 27(6):643–656, 2006.

[245] Gediyon Y Weldegiyorgis and Mapundi K Banda. Input-to-state stability of non-uniform linear hyperbolic systems of balance laws via boundary feedback control. *Applied Mathematics & Optimization*, 84(3):2701–2726, 2021.

[246] Eric Wieser and Utensil Song. Formalizing geometric algebra in lean. *Advances in Applied Clifford Algebras*, 32(3):1–26, 2022.

[247] Mark HM Winands, Yngvi Björnsson, and Jahn-Takeshi Saito. Monte-carlo tree search solver. In *International Conference on Computers and Games*, pages 25–36. Springer, 2008.

[248] W Wonham. On pole assignment in multi-input controllable linear systems. *IEEE transactions on automatic control*, 12(6):660–665, 1967.

[249] Yuhuai Wu, Albert Qiaochu Jiang, Jimmy Ba, and Roger Grosse. Int: An inequality benchmark for evaluating generalization in theorem proving. *arXiv preprint arXiv:2007.02924*, 2020.

[250] S. Xiang. Null controllability of a linearized Korteweg–de Vries equation by backstepping approach. *SIAM J. Control Optim.*, 57(2):1493–1515, 2019.

[251] Shengquan Xiang. Small-time local stabilization for a Korteweg–de Vries equation. *Systems Control Lett.*, 111:64–69, 2018.

[252] Gal Yehuda, Moshe Gabel, and Assaf Schuster. It's not what machines can learn, it's what we cannot teach. In *International conference on machine learning*, pages 10831–10841. PMLR, 2020.

[253] Huan Yu, Jean Auriol, and Miroslav Krstic. Simultaneous downstream and upstream output-feedback stabilization of cascaded freeway traffic. *Automatica J. IFAC*, 136:Paper No. 110044, 10, 2022.

[254] Huan Yu, Mamadou Diagne, Liguo Zhang, and Miroslav Krstic. Bilateral boundary control of moving shockwave in LWR model of congested traffic. *IEEE Trans. Automat. Control*, 66(3):1429–1436, 2021.

[255] Wojciech Zaremba, Karol Kurach, and Rob Fergus. Learning to discover efficient mathematical identities. *Advances in Neural Information Processing Systems*, 27, 2014.

[256] Christophe Zhang. Finite-time internal stabilization of a linear 1-D transport equation. *Systems Control Lett.*, 133:104529, 8, 2019.

[257] Christophe Zhang. *internal control and stabilization of some 1-D hyperbolic systems.* PhD thesis, Sorbonne Université, 2019.

[258] Christophe Zhang. Internal rapid stabilization of a 1-D linear transport equation with a scalar feedback. *Math. Control Relat. Fields*, 12(1):169–200, 2022.

[259] H Michael Zhang. A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research Part B: Methodological*, 36(3):275–290, 2002.

[260] Y.-C. Zhao. *Classical Solutions for Quasilinear Hyperbolic Systems.* PhD thesis, Fudan University, 1986.

[261] Kunhao Zheng, Jesse Michael Han, and Stanislas Polu. Minif2f: a cross-system benchmark for formal olympiad-level mathematics. *arXiv preprint arXiv:2109.00110*, 2021.

[262] Yang Zheng, Jiawei Wang, and Keqiang Li. Smoothing traffic flow via control of autonomous vehicles. *IEEE Internet of Things Journal*, 7(5):3882–3896, 2020.

[263] Enrike Zuazua. Exponential decay for the semilinear wave equation with locally distributed damping. *Communications in Partial Differential Equations*, 15(2):205–235, 1990.

[264] Enrike Zuazua. Uniform stabilization of the wave equation by nonlinear boundary feedback. *SIAM Journal on Control and Optimization*, 28(2):466–477, 1990.