Université de Versailles - Saint-Quentin

THÈSE

présentée en vue de l'obtention du grade de

Docteur de l'Université de Versailles - Saint-Quentin
Mention Mathématiques et Applications

par

Sylvain Ervedoza

Problèmes de contrôle et de stabilisation

Thèse soutenue le 25 novembre 2008

Rapporteurs :	Nicolas Burq
	Marius Tucsnak
Examinateurs :	Jean-Michel Coron
	Benoit Perthame
	Luc Robbiano
	Enrique Zuazua
Directeur de thèse :	Jean-Pierre Puel.

Table des matières

Introduction **i**

I Examples **3**

1 Perfectly Matched Layers in 1-d : Energy decay for continuous and semi-discrete waves **5**

1.1 Introduction 5

1.2 Analysis of the space operator L 10

1.2.1 Inverse of the operator L 11

1.2.2 Analysis of the spectrum : Eigenvalues of L 11

1.2.3 Analysis of the spectrum : Eigenvectors 12

1.3 On the decay of the energy 13

1.3.1 On the decay rate 13

1.3.2 Comments 15

1.3.3 Optimality of the decay rate 15

1.4 On the semi-discrete PML equations 16

1.4.1 Construction of non propagating waves 17

1.4.2 Spectral analysis for constant σ 19

1.4.3 Connections with the theory of stabilization 27

1.5 A semi-discrete viscous PML 27

1.6 Discussion and remarks 33

2 A mixed finite element discretization of a 1d wave equation on nonuniform meshes **37**

2.1 Introduction 37

2.2 Spectral Theory 40

2.2.1 Computations of the eigenvalues for a general mesh 40
2.2.2 Spectral properties on M-regular meshes 44
2.2.3 Proof of Theorem 2.1.2 46
2.2.4 The regularity assumption 48
2.3 Application to the null controllability of the wave equation 50
2.3.1 The continuous setting 50
2.3.2 The semi-discrete setting 51
2.4 Application to the damped wave equation 58
2.4.1 The continuous setting 58
2.4.2 The semi-discrete setting 59
2.5 Further comments 60

II Observability and stabilization properties for time-discrete approximation schemes 65

3 On the observability of time-discrete conservative linear systems 67
3.1 Introduction 67
3.2 The implicit mid-point scheme 71
3.3 General time-discrete schemes 77
3.3.1 General time-discrete schemes for first order systems 77
3.3.2 The Newmark method for second order in time systems 80
3.4 Applications 85
3.4.1 Application of Theorem 3.2.1 85
3.4.2 Application of Theorem 3.3.1 88
3.4.3 Application of Theorem 3.3.2 89
3.5 Fully discrete schemes 91
3.5.1 Main statement 91
3.5.2 Applications 94
3.6 On the admissibility condition 98
3.6.1 The time-continuous setting 98
3.6.2 The time-discrete setting 101
3.7 Further comments and open problems 102

4 Uniform exponential decay for viscous damped systems **107**

4.1 Introduction . . . 107

4.2 Proof of Theorem 4.1.1 . . . 109

4.3 Variants of Theorem 4.1.1 . . . 112

4.3.1 General viscosity operators . . . 112

4.3.2 Wave type systems . . . 113

4.4 Applications . . . 117

4.4.1 The viscous Schrödinger equation . . . 117

4.4.2 The viscous damped wave equation . . . 118

4.5 Further comments . . . 119

5 Uniformly exponentially stable approximations for a class of damped systems **123**

5.1 Introduction . . . 123

5.2 Stabilization of time-discrete systems . . . 128

5.2.1 Observability of time-discrete conservative systems . . . 128

5.2.2 Proof of Theorem 5.1.1 . . . 129

5.2.3 Some variants . . . 136

5.3 Stabilization of time-discrete systems depending on a parameter . . . 137

5.3.1 The general case . . . 138

5.3.2 Stabilization of fully discrete approximation schemes without viscosity . . . 139

5.3.3 Stabilization of fully discrete approximation schemes with viscosity . . . 141

5.4 Applications . . . 144

5.4.1 The time-discrete damped wave equation . . . 144

5.4.2 A fully discrete damped wave equation: The mixed finite element method . . . 145

5.4.3 A fully discrete damped wave equation: A viscous finite difference approximation 149

5.5 Further comments . . . 155

III Admissibility and Observability for finite element discretizations of conservative systems **159**

6 Schrödinger equations **161**

6.1 Introduction . . . 161

6.2 Spectral methods . . . 166
6.2.1 Characterizations of admissibility . . . 167
6.2.2 Characterizations of observability . . . 169
6.3 Proof of Theorem 6.1.3 . . . 171
6.3.1 Admissibility . . . 171
6.3.2 Observability . . . 174
6.4 Examples of applications . . . 176
6.4.1 The 1-d case . . . 176
6.4.2 More general cases . . . 178
6.5 Fully discrete approximation schemes . . . 179
6.6 Controllability properties . . . 180
6.6.1 The continuous setting . . . 180
6.6.2 The space semi-discrete setting . . . 181
6.7 Stabilization properties . . . 187
6.7.1 The continuous setting . . . 187
6.7.2 The space semi-discrete setting . . . 188
6.8 Further comments . . . 189

7 Wave equations **193**

7.1 Introduction . . . 193
7.2 Spectral methods . . . 198
7.2.1 Characterizations of admissibility . . . 199
7.2.2 Characterizations of observability . . . 202
7.3 Proof of Theorem 7.1.3 . . . 208
7.3.1 Admissibility . . . 209
7.3.2 Observability . . . 212
7.4 Examples . . . 213
7.4.1 The 1d wave equation . . . 214
7.4.2 More general cases . . . 215
7.5 Fully discrete approximation schemes . . . 216
7.6 Controllability properties . . . 218

7.6.1 The continuous setting . . . 218
7.6.2 The semi-discrete setting . . . 218
7.7 Stabilization properties . . . 223
7.7.1 The continuous setting . . . 223
7.7.2 The space semi-discrete setting . . . 223
7.8 Other models . . . 224
7.8.1 A wave equation observed through $y(t) = Bu(t)$ . . . 224
7.8.2 Applications to Schrödinger type equations . . . 226
7.9 Further comments . . . 228

IV Miscellaneous **233**

8 Control and stabilization property for a singular heat equation **235**
8.1 Introduction . . . 235
8.2 Null controllability in the case $\mu \leq \mu^*(N)$ . . . 239
8.2.1 Carleman estimate . . . 239
8.2.2 From the Carleman estimate to the Observability inequality . . . 244
8.2.3 Proofs of technical Lemmas . . . 245
8.3 Non uniform stabilization in the case $\mu > \mu^*(N)$ . . . 250
8.3.1 Spectral estimates . . . 251
8.3.2 Proof of Theorem 8.1.2 . . . 252
8.4 Comments . . . 253

Remerciements

Je tiens avant tout à remercier Jean-Pierre Puel, qui a immédiatement éveillé ma curiosité sur des problèmes de contrôle dès le DEA. J'ai particulièrement apprécié sa disponibilité, ses qualités d'écoute, son attention toujours vive pour mes questions, mais aussi ses grandes compétences pédagogiques qui lui ont permis à maintes reprises de m'expliquer des idées très riches. J'ai aussi été très impressionné par sa connaissance de Shanghaï !

J'aimerais également témoigner de ma reconnaissance à Enrique Zuazua, qui est également un des instigateurs de ce travail. Il a eu la gentillesse de m'accueillir à deux reprises à Madrid, et m'a suivi avec beaucoup d'enthousiasme tout au long de cette thèse, la ponctuant de collaborations très fructueuses. Je le remercie aussi d'avoir accepté de faire partie de mon jury.

Je remercie très chaleureusement Nicolas Burq et Marius Tucsnak, dont j'ai utilisé les résultats mathématiques de nombreuses fois, pour m'avoir fait l'honneur d'accepter de rapporter cette thèse. J'aimerais remercier également Marius Tucsnak pour ses conseils avisés et pour m'avoir déjà invité à plusieurs reprises à Nancy.

C'est avec grand plaisir que je remercie Jean-Michel Coron, Benoît Perthame et Luc Robbiano, pour avoir consenti à être membres de mon jury. Chacun a joué pour moi un rôle très important dans ma formation. Merci particulièrement à Jean-Michel Coron pour son intérêt soutenu pour mes travaux de recherche.

Je remercie également les gens qui s'y sont intéressés pour leurs encouragements répétés, ainsi que ceux qui ont pris le temps de m'expliquer leurs résultats. Un grand merci à Takéo Takahashi, Karine Beauchard, Sergio Guerrero, Olivier Glass, Chuang Zheng, Jérôme Le Rousseau, Luc Miller, Pierre Rouchon, Julie Valein, Sorin Micu, Carlos Castro, Marianne Chapouly et Mazyar Mirrahimi.

Je souhaite aussi remercier l'ensemble des membres du département de mathématiques de Versailles. La bonne humeur générale qui règne au sein du département a certainement favorisé le bon déroulement de ma thèse. Merci en particulier aux thésards, Jean-Maxime, Pascal, Claudio, Vianney, Éric, Clémence, Jérémy et les autres, ainsi qu'aux occupants des bureaux voisins, Alexis, Aude, Aurélie, Nicolas, Stéphane, Ariane, Mariane et Mokka pour leur bienveillance. Je pense aussi à ceux qui m'ont apporté leur aide de nombreuses fois et les remercie de leur disponibilité. Merci notamment à Otared, Yvan et Thierry pour les mathématiques et à Jean, Thierry et Frédéric pour les enseignements.

Je remercie aussi mes amis pour leur soutien moral et parfois même mathématique ! Mes camarades de prépa qui me supportent depuis plus de huit ans maintenant, ont su m'aider à résister à la pression d'abord des concours, puis de la thèse : Élodie, Thomas, Oscar, Étienne, Calixte, Yi, et Céline, à qui j'adresse un remerciement tout spécifique. Merci à mes copains normaliens également, pour la plupart en thèse, et qui comprennent mieux que personne les doutes et les difficultés qu'ont pu suscité ce travail de longue haleine : Guillaume, Sylvain, Loïc, Matthieu, Simon, Benjamin, Pierre et Christophe.

Je remercie également ma famille et plus particulièrement mes parents pour m'avoir toujours soutenu et aidé tout au long de mon parcours.

Enfin, j'adresse un grand merci à tous ceux qui m'ont accompagné pendant ces quelques années et que je n'ai pas pu remercier nominalement dans ces quelques lignes.

Introduction

Dans cette thèse, nous nous intéresserons à divers problèmes liés à la contrôlabilité et à la stabilisation de systèmes d'évolution continus et discrets. Dans un premier temps, nous allons décrire le contexte dans lequel se place le présent travail, et pour cela, introduire un formalisme abstrait qui contient tous les problèmes étudiés, et qui sera spécifié par la suite.

De nombreux modèles physiques se mettent sous la forme suivante :

$$\begin{cases} \dot{z} = \mathcal{A}z, \quad t \geq 0, \\ z(0) = z_0, \end{cases} \tag{0.0.1}$$

où $\dot{(\,)}$ désigne la dérivée par rapport au temps, et où $\mathcal{A}$ est un opérateur, en général différentiel. Pour fixer les idées, on suppose que la donnée initiale z_0 appartient à un espace de Hilbert $\mathfrak{X}$, et que l'opérateur $\mathcal{A}$ est un opérateur éventuellement non borné sur $\mathfrak{X}$.

Dans la suite, nous supposerons également que, si z_0 est dans $\mathfrak{X}$, alors la solution $t \mapsto z(t)$ de (0.0.1) existe, est unique, et appartient à l'espace $C([0,T];\mathfrak{X})$ pour tout temps $T > 0$. Pour être plus précis, nous supposons que le problème de Cauchy associé à (0.0.1) est un problème bien posé au sens de Hadamard.

Le système (0.0.1) modélise effectivement de nombreux phénomènes physiques : Citons entre autres les modèles diffusifs (chaleur), les modèles issus de la mécanique quantique (équation de Schrödinger), et de l'étude des systèmes oscillants (ondes). Pour plus d'exemples, nous faisons référence à l'ouvrage [11].

Observabilité. Le premier problème que nous étudions est celui de l'observabilité. On se donne un opérateur B défini sur $\mathcal{D}(\mathcal{A})$, à valeurs dans un espace de Hilbert $\mathcal{Y}$, et nous supposons que nous pouvons observer, pendant un certain temps T, la quantité

$$y(t) = Bz(t), \quad t \in (0,T). \tag{0.0.2}$$

Comme la donnée initiale z_0 est dans $\mathfrak{X}$, la solution $t \mapsto z(t)$ de (0.0.1) appartient à $C([0,T];\mathfrak{X})$, et on ne peut *a priori* pas donner un sens précis à (0.0.2) pour $B \in \mathfrak{L}(\mathcal{D}(\mathcal{A}),\mathcal{Y})$. C'est pourquoi nous demandons à ce que le système (0.0.1)-(0.0.2) soit admissible :

Définition 1 (Admissibilité)**.** Le système (0.0.1)-(0.0.2) est dit admissible si pour tout $T > 0$, il existe une constante K_T telle que toute solution de (0.0.1) avec donnée initiale $z_0 \in \mathcal{D}(\mathcal{A})$ satisfait

$$\int_0^T \|Bz(t)\|_{\mathcal{Y}}^2 \ \mathrm{dt} \leq K_T \|z_0\|_{\mathfrak{X}}^2 . \tag{0.0.3}$$

Dans ce cas, lorsque l'opérateur $\mathcal{A}$ est de domaine dense, ce qui sera toujours vérifié par la suite, par densité de $\mathcal{D}(\mathcal{A})$ dans $\mathfrak{X}$, l'opérateur d'observation peut être étendu en un opérateur continu de $\mathfrak{X}$ à valeurs dans $L^2(0,T;\mathcal{Y})$. En particulier, remarquons que si l'opérateur B appartient à $\mathfrak{L}(\mathfrak{X},\mathcal{Y})$, alors la propriété (0.0.3) est automatiquement satisfaite.

La question est alors de savoir si la connaissance de y nous permet de déterminer, ou non, la fonction z. Si tel est le cas, nous dirons que le système (0.0.1)-(0.0.2) est observable au sens suivant :

Définition 2 (Observabilité)**.** Le système (0.0.1)-(0.0.2) est dit observable au temps $T > 0$ s'il existe une constante $k_T > 0$ telle que toute solution de (0.0.1) satisfait

$$k_T \left\|z(T)\right\|_{\mathfrak{X}}^2 \leq \int_0^T \left\|Bz(t)\right\|_{\mathcal{Y}}^2 \, \mathrm{dt}. \tag{0.0.4}$$

Dans la suite, nous dirons que le système (0.0.1)-(0.0.2) est observable s'il l'est en un certain temps $T > 0$.

Remarquons que ce problème est très pertinent en pratique. En effet, il n'est pas rare que nous ne puissions avoir accès qu'à des données partielles sur certains systèmes complexes. C'est par exemple le cas en météorologie, où les seules informations à notre disposition concernent une petite couche au voisinage de la surface terrestre. Nous faisons référence par exemple à [31] en ce qui concerne ce problème d'assimilation de données.

Il est intéressant de constater que les propriétés d'observabilité sont reliées à deux autres questions tout aussi pertinentes en pratique, celles de la contrôlabilité et de la stabilisation.

Contrôlabilité. Nous nous intéressons désormais au problème suivant : pour une donnée initiale $z_0 \in \mathfrak{X}$, trouver un contrôle $v \in L^2(0,T;\mathcal{Y})$ tel que la solution de

$$\begin{cases} \dot{z} = \mathcal{A}z + Cv, \quad t \in (0,T), \\ z(0) = z_0, \end{cases} \tag{0.0.5}$$

soit nulle au temps $T > 0$:

$$z(T) = 0. \tag{0.0.6}$$

Ici, l'opérateur C, qui décrit les possibilités d'actions sur le système (0.0.5), est un opérateur continu de $\mathcal{Y}$ dans $\mathcal{D}(\mathcal{A})^*$.

Remarquons que, en utilisant la linéarité du système (0.0.5), le problème ci-dessus, dit de contrôlabilité à zéro, est équivalent au problème de contrôlabilité sur les trajectoires. Là encore, il s'agit donc d'une question physiquement pertinente puisqu'il s'agit de décrire l'action que l'on peut exercer sur un système donné.

Il est désormais classique que la contrôlabilité à zéro est équivalente à l'observabilité du système adjoint. C'est le contenu de la méthode HUM (*Hilbert Uniqueness Method*) introduite dans [27].

Considérons le problème adjoint (rétrograde)

$$\begin{cases} \dot{w} = -\mathcal{A}^* w, \quad t \in (0,T). \\ w(T) = w_T \in \mathfrak{X}, \end{cases} \tag{0.0.7}$$

et les propriétés d'admissibilité et d'observabilité suivantes : il existe des constantes $k_T > 0$ et $K_T > 0$ telles que toute solution de (0.0.7) satisfait

$$k_T \left\|w(0)\right\|_{\mathfrak{X}}^2 \leq \int_0^T \left\|C^* w(t)\right\|_{\mathcal{Y}}^2 \, \mathrm{dt} \leq K_T \left\|w_T\right\|_{\mathfrak{X}}^2. \tag{0.0.8}$$

Supposons que les propriétés d'admissibilité et d'observabilité (0.0.8) sont vérifiées. Supposons également que le système (0.0.7) satisfait la propriété d'unicité rétrograde suivante, qui sera vérifiée dans tous les exemples que nous traiterons ci-après : toute solution w de (0.0.7) satisfaisant $w(0) = 0$ est identiquement nulle.

Introduisons alors la fonctionnelle $\mathcal{J}$ définie pour $w_T \in \mathfrak{X}$ par

$$\mathcal{J}(w_T) = \frac{1}{2}\int_0^T \|C^* w(t)\|_{\mathcal{Y}}^2 \ \mathrm{dt} + < w(0), z_0 >_{\mathfrak{X}}, \tag{0.0.9}$$

où w est la solution de (0.0.7) associée à w_T. Cette fonctionnelle est strictement convexe, et, au vu de la propriété (0.0.8), est également coercive dans la norme

$$\|w_T\|_{obs}^2 = \int_0^T \|C^* w(t)\|_{\mathcal{Y}}^2 \ \mathrm{dt}.$$

La fonctionnelle $\mathcal{J}$ admet donc un unique minimum w_T^* dans le complété $\bar{\mathfrak{X}}$ de $\mathfrak{X}$ pour la norme $\|\cdot\|_{obs}$. Remarquons qu'alors il existe une unique application Θ continue de $\bar{\mathfrak{X}}$ sur $L^2(0,T;\mathcal{Y})$ qui coïncide avec $w_T \mapsto C^* w(t)$ pour $w_T \in \mathfrak{X}$.

Le contrôle v de (0.0.5) de norme $L^2(0,T;\mathcal{Y})$ minimale est alors donné par

$$v(t) = \Theta w_T^*. \tag{0.0.10}$$

Remarquons que, lorsque le système (0.0.7) est conservatif, l'hypothèse (0.0.8) implique $\bar{\mathfrak{X}} = \mathfrak{X}$. Il s'ensuit que, dans ce cas, $\Theta w_T^* = C^* w^*(t)$, où w^* est la solution de (0.0.7) associée à w_T^*. De même, la même simplification peut être faite lorsque $\mathfrak{X}$ est de dimension finie puisqu'alors toutes les normes sont équivalentes.

Stabilisation. Pour cette question, nous nous limitons aux cas où l'opérateur $\mathcal{A}$ est antisymétrique, et où l'opérateur B appartient à $\mathfrak{L}(\mathfrak{X},\mathcal{Y})$.

Considérons alors le système amorti

$$\begin{cases} \dot{w} = \mathcal{A}w - B^*Bw, \quad t \geq 0, \\ w(0) = w_0 \in \mathfrak{X}. \end{cases} \tag{0.0.11}$$

Un tel système modélise de nombreux systèmes physiques comportant un terme de stabilisation de type *feedback*, par exemple les ondes amorties.

Il s'agit en effet d'un système amorti puisque l'énergie des solutions w de (0.0.11), définie par

$$E(t) = \frac{1}{2}\|w(t)\|_{\mathfrak{X}}^2, \tag{0.0.12}$$

satisfait la loi de dissipation

$$\frac{dE}{dt}(t) = -\|Bw(t)\|_{\mathcal{Y}}^2. \tag{0.0.13}$$

Nous nous interrogeons alors sur la possibilité de décroissance exponentielle des solutions. Pour être plus précis, nous voulons savoir s'il existe des constantes strictement positives M et $\nu > 0$ telles que toutes les solutions de (0.0.11) satisfont

$$E(t) \leq M\,E(0)\exp(-\nu t), \quad t \geq 0. \tag{0.0.14}$$

Il est désormais bien connu (cf. [27, 22]) que la décroissance exponentielle de l'énergie pour les solutions de (0.0.11) au sens de (0.0.14) est également équivalente à l'observabilité du système (0.0.1)-(0.0.2).

Problématique. Il existe de nombreuses situations concrètes où l'on a besoin de considérer non pas un système mais une famille de systèmes, pour lesquels on aimerait avoir des propriétés d'observabilité uniformes, afin d'en déduire divers types de résultats de contrôlabilité et de stabilisation.

C'est par exemple le cas lorsque l'on s'intéresse à des systèmes discrétisés en espace et/ou en temps.

Pour fixer les idées, considérons un système continu (0.0.1)-(0.0.2) admissible et observable au sens de (0.0.3) et (0.0.4), et supposons de plus que l'opérateur $\mathcal{A}$ est antisymétrique.

Observabilité discrète. Introduisons les opérateurs $\mathcal{A}_h$ et B_h correspondants aux discrétisations des opérateurs $\mathcal{A}$ et B sur un maillage de taille $h > 0$. Le système (0.0.1)-(0.0.2) est alors approché par

$$\begin{cases} \dot{z}_h = \mathcal{A}_h z_h, \quad t \geq 0, \\ z_h(0) = z_{0h} \in \mathfrak{X}_h, \end{cases} \qquad y_h(t) = B_h z_h(t), \quad t \in (0,T). \tag{0.0.15}$$

Ici, l'espace $\mathfrak{X}_h$ correspond à une approximation de dimension finie de $\mathfrak{X}$. L'opérateur B_h est à priori à valeurs dans un certain espace $\mathcal{Y}_h$ qui correspond également à une approximation de $\mathcal{Y}$ dans un sens raisonnable. Pour l'instant, nous restons volontairement imprécis, mais des affirmations précises seront données plus tard dans le corps du manuscrit. Comme nous avons supposé $\mathcal{A}$ antisymétrique, nous nous intéressons uniquement à des discrétisations qui préservent cette propriété, et supposons donc que pour tout $h > 0$, l'opérateur $\mathcal{A}_h$ est antisymétrique sur $\mathfrak{X}_h$.

Il est alors naturel de s'interroger sur les propriétés d'admissibilité et d'observabilité des systèmes (0.0.15). Il peut arriver que, pour tout $h > 0$, il existe des solutions z_h des systèmes (0.0.15) telles que $B_h z_h(t) = 0$ pour tout t (cf. contre-exemple d'Otared Kavian, explicité dans [44, p.72]). Dans ce cas, cette réponse négative à la continuation unique nie les propriétés d'observabilité pour les systèmes discrets (0.0.15).

Cependant, ce n'est pas le seul problème qui peut intervenir pour l'observabilité des systèmes (0.0.15). En effet, par exemple en dimension un, il est en général facile de montrer que les seules solutions z_h de (0.0.15) qui satisfont $B_h z_h(t) = 0$ pour tout t dans un intervalle de temps sont les solutions nulles. Notamment, on déduit alors dans ce cas que, pour tout $h > 0$, le système (0.0.15) est admissible et observable, et ce en tout temps : pour tout $h > 0$ et pour tout $T > 0$, il existe des constantes positives $k_{T,h} > 0$ et $K_{T,h} > 0$ telles que toute solution z_h de (0.0.15) satisfait

$$k_{T,h} \left\| z_h(T) \right\|_{\mathfrak{X}_h}^2 \leq \int_0^T \left\| B_h z_h(t) \right\|_{\mathcal{Y}_h}^2 \, \mathrm{dt} \leq K_{T,h} \left\| z_{0h} \right\|_{\mathfrak{X}_h}^2 . \tag{0.0.16}$$

Nous allons voir ci-dessous que, lorsque la propriété d'observabilité (0.0.16) n'est pas satisfaite uniformément en $h > 0$, c'est-à-dire lorsque $\lim_{h \to 0} k_{T,h} = 0$, les procédures de calcul des contrôles sur les systèmes discrets (0.0.15) peuvent donner des résultats biaisés, voire faux, pour le système continu.

Contrôles discrets. Sous la condition (0.0.16), pour tout $h > 0$, le système

$$\begin{cases} \dot{z}_h = \mathcal{A}_h z_h + B_h^* v_h, \quad t \in (0,T). \\ z_h(0) = z_{0h} \in \mathfrak{X}_h, \end{cases} \tag{0.0.17}$$

est contrôlable, c'est-à-dire qu'il existe une fonction v_h dans $L^2(0,T;\mathcal{Y}_h)$ telle que la solution de (0.0.17) satisfait

$$z_h(T) = 0.$$

En fait, suivant la méthode HUM décrite ci-dessus, on peut même calculer le contrôle à zéro v_h qui minimise la norme $L^2(0,T;\mathcal{Y}_h)$.

Il est alors naturel de penser que, si les données initiales z_{0h} convergent vers z_0, alors les contrôles v_h devraient converger vers le contrôle v de (0.0.5) (avec $C = B^*$). Cela est en fait faux en pratique dans de multiples situations, comme le montrent les simulations numériques concernant l'équation des ondes unidimensionnelle disponibles dans l'article [44].

Comme souligné dans [44], la norme des contrôles v_h peut exploser quand $h \to 0$. En effet, les systèmes discrétisés (0.0.15) ont une dynamique différente de celle du système continu (0.0.1), notamment aux hautes fréquences, cf. [38].

Dans ce cadre, de nombreux travaux récents (cf. [44] et sa bibliographie) ont été consacrés à mettre au point des techniques permettant de calculer sur les systèmes discrétisés (0.0.15) des pseudo-contrôles v_h pour (0.0.17) qui convergent, lorsque les données initiales z_{0h} convergent vers z_0, vers un contrôle admissible pour le système continu (0.0.5) (toujours avec $C = B^*$).

Ces méthodes consistent essentiellement en des mécanismes de filtrage qui permettent d'éliminer les hautes fréquences parasites introduites lors de la discrétisation. Afin de prouver la convergence des contrôles v_h des systèmes discrétisés (0.0.17) vers un contrôle v du système continu, la méthode la plus courante consiste à trouver des classes de données pour lesquelles on peut prouver les inégalités (0.0.16) avec des constantes $k_{T,h}$ et $K_{T,h}$ indépendantes de $h > 0$.

En d'autres termes, il s'agit de déterminer, pour tout $h > 0$, un sous-espace $\tilde{\mathfrak{X}}_h \subset \mathfrak{X}_h$ de données, globalement invariant par l'équation (0.0.15), tel qu'il existe un temps $T > 0$ et des constantes positives $k_T > 0$ et $K_T > 0$, indépendants de $h > 0$, tels que toute solution z_h de (0.0.15) ayant pour donnée initiale $z_{0h} \in \tilde{\mathfrak{X}}_h$ satisfait

$$k_T \left\| z_h(T) \right\|_{\mathfrak{X}_h}^2 \leq \int_0^T \left\| B_h z_h(t) \right\|_{\mathcal{Y}_h}^2 \, \mathrm{dt} \leq K_T \left\| z_{0h} \right\|_{\mathfrak{X}_h}^2 . \tag{0.0.18}$$

Les méthodes utilisées jusqu'à présent pour démontrer les inégalités (0.0.18) reposent sur des techniques de multiplicateurs (inspirées de [25] et directement effectuées sur les systèmes discrétisés (0.0.15)), ou sur des propriétés de séparation spectrale basées sur [24].

Les Chapitres 2, 6, et 7 présentent des études détaillées de ces questions sur divers exemples.

Au Chapitre 2 (correspondant à [12]), nous considérons l'équation des ondes unidimensionnelle discrétisée sur des maillages non uniformes en utilisant la méthode des éléments finis mixtes, dont nous présentons une étude détaillée des propriétés d'observabilité. Cette question avait déjà été traitée dans les travaux [8, 9] dans le cas des maillages uniformes, ce qui permettait d'utiliser des méthodes de multiplicateurs, ou, en dimension un, des méthodes spectrales. Ici, dû au manque d'uniformité du maillage, le spectre est moins explicite, mais nous arrivons tout de même à prouver des propriétés de séparation du spectre, puis d'équirépartition des vecteurs propres, qui permettent de démontrer les propriétés (0.0.18) dans tout l'espace $\mathfrak{X}_h$, uniformément en $h > 0$. À notre connaissance, c'est l'étude spectrale la plus précise menée jusqu'à présent pour des systèmes discrétisés sur des maillages non uniformes. Signalons que les questions d'observabilité pour l'équation des ondes unidimensionnelle discrétisée avec la méthode des éléments finis sur des maillages non uniformes ont été traitées dans [34], mais la question de l'optimalité de la réponse apportée dans [34] est encore largement ouverte.

Aux Chapitres 6 et 7 (correspondants à [13, 14]), nous étudions des systèmes abstraits généraux représentant les équations de Schrödinger et des ondes discrétisées selon la méthode des éléments finis. La méthode que nous utilisons est une méthode spectrale basée sur les travaux [6, 29, 33], que nous adaptons pour des systèmes discrétisés. Cela fournit une approche robuste pour étudier les propriétés d'observabilité des discrétisations en espaces des systèmes admissibles et observables. Notamment, nos

résultats s'appliquent en n'importe quelle dimension et sans condition sur la structure des maillages, ce qui généralise grandement les résultats connus jusqu'à présent (cf. [44] et sa bibliographie). Cependant, comme dans [34], nous ne savons pas si nos résultats sont optimaux. Cette question est largement ouverte.

Stabilisation discrète. Lorsque les opérateurs B et B_h sont bornés sur $\mathfrak{X}$ et $\mathfrak{X}_h$ respectivement, on peut s'interroger sur les propriétés de décroissance de l'énergie des systèmes discrétisés

$$\begin{cases} \dot{w}_h = \mathcal{A}_h w_h - B_h^* B_h w_h, \quad t \geq 0, \\ w_h(0) = w_{0h} \in \mathfrak{X}_h, \end{cases} \tag{0.0.19}$$

ainsi que de leur uniformité.

L'énergie des solutions w_h de (0.0.19) est donnée par

$$E_h(t) = \frac{1}{2} \|w_h(t)\|^2_{\mathfrak{X}_h}. \tag{0.0.20}$$

Comme dans le cas ci-dessus, lorsque les inégalités (0.0.16) sont satisfaites, pour tout $h > 0$, il existe des constantes positives M_h et ν_h telles que les solutions de (0.0.19) satisfont

$$E_h(t) \leq M_h\, E_h(0) \exp(-\nu_h t), \quad t \geq 0. \tag{0.0.21}$$

Mais la décroissance n'est pas, en général, uniforme. On peut notamment avoir des cas où ν_h tend vers 0 quand $h \to 0$.

Il est alors naturel de se demander si l'on peut modifier le système (0.0.19) de façon à obtenir des systèmes discrétisés exponentiellement stables uniformément en h.

À nouveau, nous nous référons à [44] et à sa bibliographie pour divers travaux concernant cette question. L'idée générale consiste à introduire un terme de viscosité numérique dans (0.0.19) de façon à amortir efficacement les hautes fréquences parasites introduites lors de la discrétisation.

Les Chapitres 1, 4 et 5 proposent une étude de ces questions.

Au Chapitre 1 (correspondant à [17]), nous étudions les propriétés spectrales fines des discrétisations spatiales des équations du modèle *Perfectly Matched Layers* unidimensionnelles, qui constituent une variante de l'équation des ondes amorties. En particulier, nous mettons en évidence l'existence de valeurs propres parasites qui correspondent à des vecteurs propres hautes fréquences localisés dans la zone où le terme d'amortissement n'est pas actif, ce qui prouve en particulier que la quantité ν_h dans (0.0.21) tend vers 0 quand $h \to 0$. Cette description précise du spectre des systèmes amortis discrétisés est, à notre connaissance, la première à montrer ce phénomène explicitement. Nous montrons alors, en s'inspirant des travaux [37, 34], qu'en introduisant un terme de viscosité numérique correctement choisi dans les équations discrétisées, on peut obtenir des systèmes discrétisés exponentiellement stables, uniformément en $h > 0$.

Au Chapitre 4 (correspondant à [18]), nous exhibons, pour des systèmes continus abstraits (0.0.11), plusieurs formes d'opérateurs de viscosité pour lesquels les phénomènes d'*overdamping* n'apparaissent pas. La méthode que nous utilisons a l'avantage de traiter séparément basses et hautes fréquences, utilisant aux basses fréquences les propriétés d'observabilité des systèmes (0.0.1)-(0.0.2), et aux hautes fréquences les propriétés dissipatives des systèmes visqueux sans amortissement. En particulier, cela fournit des résultats robustes et généraux qui peuvent s'appliquer aussi bien dans le contexte des équations discrétisées en espace, comme aux Chapitres 1, 6 et 7, que pour les équations discrétisées en temps, ou même en temps et en espace, cf. Chapitre 5, généralisant ainsi les résultats de [37, 34] sur les propriétés de stabilisation des systèmes semi-discrétisés en espace.

Ainsi, au Chapitre 5 (correspondant à [19]), nous donnons une méthode systématique qui permet de mettre au point, pour des systèmes qui ne sont observables qu'aux basses fréquences, des variantes visqueuses de (0.0.1) pour lesquelles on peut guarantir des propriétés de stabilisation uniformes en $h > 0$.

Des problèmes similaires se posent lorsque l'on considère des systèmes discrétisés en temps, où des solutions parasites hautes fréquences perturbent les propriétés d'observabilité et d'admissibilité des systèmes discrétisés, et, notamment, le bon fonctionnement de la méthode HUM pour calculer numériquement des contrôles approchés pour les systèmes continus.

Au Chapitre 3 (correspondant à [16]), nous prouvons donc, pour un système conservatif (0.0.1)-(0.0.2) admissible et observable, des propriétés d'observabilité uniformes pour les systèmes discrétisés en temps, dans une classe filtrée. Là encore, nous utilisons les résultats spectraux de [6, 29] pour obtenir une méthode robuste, qui s'applique pour de nombreux systèmes et de nombreuses discrétisations en temps. Ainsi, nos résultats s'appliquent également à des familles de systèmes uniformément observables pour lesquels nous pouvons déduire pour les familles de systèmes discrets en temps correspondants des propriétés d'observabilité uniformes en le paramètre de discrétisation en temps. En particulier, si l'on considère une famille de systèmes discrétisés en espace qui sont uniformément observables en le paramètre de discrétisation en espace, alors les systèmes totalement discrétisés correspondants satisfont des propriétés d'observabilité uniformes en les paramètres de discrétisation en espace et en temps. Cet argument permet ainsi de découpler les problèmes liés à la discrétisation en espace de ceux liés à la discrétisation en temps, permettant par exemple de déduire des résultats des Chapitres 2, 6 et 7 des propriétés d'observabilité pour les systèmes totalement discrétisés correspondants, uniformément en les paramètres de discrétisation en espace et en temps. A notre connaissance, ce résultat est le premier qui donne, de façon systématique, des résultats d'observabilité pour des systèmes discrétisés en temps à partir des propriétés d'observabilité des systèmes continus en temps correspondants.

Au Chapitre 5 (correspondant à [19]), nous combinons les résultats du Chapitre 3 avec ceux du Chapitre 4, pour obtenir une approche générale et robuste qui fournit, pour des systèmes continus exponentiellement stables, des discrétisations en temps et en espace uniformément exponentiellement stables. Comme indiqué ci-dessus, la méthode abstraite que nous développons généralise et étend les résultats obtenus au Chapitre 1 ainsi que dans [37, 34] pour des systèmes discrétisés en espace.

Ci-dessous, nous présentons, pour la commodité du lecteur, le plan que nous avons adopté.

Dans la Partie I, nous étudions deux systèmes modélisant des équations des ondes unidimensionnelles, tout d'abord le système PML (pour *Perfectly Matched Layers*) discrétisé sur des grilles uniformes, puis un système classique d'ondes unidimensionnel, discrétisé selon une méthode d'éléments finis mixtes, mais sur des maillages non uniformes. Dans ces deux cas, en utilisant conjointement des méthodes de multiplicateurs et des méthodes spectrales, nous prouvons des résultats qui sont, en un sens que nous préciserons, optimaux.

Dans la Partie II, nous considérons des systèmes conservatifs abstraits, que nous supposons admissibles et observables, et prouvons des propriétés d'admissibilité et d'observabilité pour leurs discrétisations en temps. Notre méthode est basée sur des techniques spectrales. En particulier, nous utilisons de manière décisive la caractérisation spectrale de l'observabilité de systèmes conservatifs donnée dans [6, 29]. Nous expliquons aussi comment ces résultats s'interprètent dans le cadre des systèmes amortis.

Dans la Partie III, nous étudions les propriétés d'observabilité de systèmes abstraits discrétisés selon la méthode des éléments finis. Notre méthode, à nouveau basée sur des critères spectraux, nous

permet d'obtenir des résultats très généraux, qui, à notre connaissance, sont les premiers à pouvoir s'appliquer instantanément en n'importe quelle dimension et pour n'importe quel maillage régulier.

Dans la Partie IV, nous présentons un travail relié à cette thématique correspondant à [15], mais dans le cadre assez différent d'une équation de la chaleur avec un potentiel singulier $-\mu/|x|^2$. Cependant, notre approche est là encore basée sur des considérations d'uniformité des propriétés d'observabilité pour des potentiels réguliers de la forme $-\mu/(|x|^2+|\epsilon|^2)$. Nos méthodes reposent alors sur une inégalité de Carleman pour prouver un résultat positif lorsque $\mu \leq \mu^*(N)$, où $\mu^*(N)$ est la constante de Hardy en dimension N, et sur des méthodes spectrales afin de prouver un résultat négatif lorsque $\mu > \mu^*(N)$.

Dans la suite de cette introduction, nous présentons plus précisément le contenu de chaque partie de cette thèse.

Partie I. Étude précise d'équations d'ondes discrétisées en espace

Dans cette partie, nous présentons, pour deux modèles d'équations des ondes, des études exhaustives et optimales des propriétés d'observabilité et de dissipation de systèmes discrétisés. En effet, les deux exemples étudiés sont suffisamment explicites en dimension un d'espace pour mettre en évidence avec précision les phénomènes parasites qui apparaissent aux hautes fréquences.

Chapitre 1. La méthode *Perfectly Matched Layers* (PML).

Lorsque l'on résout numériquement un problème d'équation des ondes en domaine extérieur en temps grand, il est nécessaire de limiter le domaine de calcul à cause des capacités finies de calcul numérique. Il est alors nécessaire d'introduire des conditions limites sur la frontière nouvellement formée, qui peuvent éventuellement perturber la solution à l'intérieur du domaine de calcul, à cause de phénomènes de réflexion.

La méthode PML, introduite par Bérenger dans [2] en 1994, consiste à entourer le domaine de calcul d'une couche dans laquelle les équations sont modifiées afin de dissiper l'énergie qui y entre, de telle sorte que l'énergie réfléchie est petite, voire nulle. Depuis, cette méthode a démontré son efficacité dans de nombreux problèmes concrets [39].

Nous nous proposons donc d'étudier précisément le modèle PML en dimension un d'espace et d'expliquer son efficacité.

Considérons le système du premier ordre suivant, équivalent à l'équation des ondes sur $(0,\infty)$:

$$\begin{cases} \partial_t P + \partial_x V = 0 \quad \text{dans } (0,\infty)\times(0,\infty), \\ \partial_t V + \partial_x P = 0 \quad \text{dans } (0,\infty)\times(0,\infty), \\ P(0,t) = 0, \quad P(x,0) = P_0(x), \quad V(x,0) = V_0(x), \end{cases} \tag{0.0.22}$$

où P_0 et V_0 sont des fonctions de $L^2(\mathbb{R})$ à support dans $(0,1)$.

ll est alors bien connu que l'énergie des solutions se propage à vitesse 1. En particulier, la solution $t \mapsto (P,V)(t)$ de (0.0.22) est nulle dans $(0,1)$ pour tout temps $t > 2$.

Considérons alors le système déduit de (0.0.22) par la méthode PML dans le cas où la zone de

calcul (i.e. la zone qui nous intéresse) est $(0,1)$:

$$\begin{cases} \partial_t P + \partial_x V + \chi_{(1,2)}\sigma P = 0 \quad \text{dans } (0,2)\times(0,T), \\ \partial_t V + \partial_x P + \chi_{(1,2)}\sigma V = 0 \quad \text{dans } (0,2)\times(0,T), \\ P(0,t) = P(2,t) = 0, \quad P(x,0) = P_0(x), \quad V(x,0) = V_0(x), \end{cases} \tag{0.0.23}$$

où P_0 et V_0 sont dans $L^2(0,2)$ et à support dans $(0,1)$, $\chi_{(1,2)}$ est la fonction caractéristique de l'intervalle $(1,2)$, et $\sigma = \sigma(x)$ est une fonction positive dans $L^\infty(1,2)$.

Le système (0.0.23) correspond en fait à un système dissipatif, puisque l'énergie

$$E(t) = \frac{1}{2}\left\|P(t)\right\|^2_{L^2(0,2)} + \frac{1}{2}\left\|V(t)\right\|^2_{L^2(0,2)}$$

satisfait

$$\frac{dE}{dt}(t) = -\int_1^2 \sigma\Big(|P(t)|^2 + |V(t)|^2\Big)\, dx.$$

Au vu de la propriété de propagation de l'énergie pour (0.0.22), il est naturel de s'attendre à ce que l'énergie des solutions de (0.0.23) décroisse, et nous pouvons voir le taux de décroissance de cette énergie comme une mesure de l'efficacité de la méthode PML.

Dans un premier temps, nous prouvons donc que l'énergie des solutions de (0.0.23) est exponentiellement décroissante. Nous présentons deux méthodes pour prouver ce résultat. L'une est basée sur une décomposition spectrale explicite de l'opérateur spatial dans (0.0.23), et l'autre sur la méthode des caractéristiques (ce qui est proche de la preuve de la formule de D'Alembert). Par ces méthodes assez explicites, nous prouvons le théorème suivant :

Théorème 3. *Les solutions de* (0.0.23) *avec donnée initiale dans* $L^2(0,2)^2$ *(pas forcément à support dans* $(0,1)$*) satisfont*

$$E(t) \le E(0)\exp\Big((4-t)\int_1^2 \sigma\Big), \quad t \ge 0.$$

On en déduit alors que la norme $L^1(1,2)$ de σ mesure l'efficacité de la méthode PML pour le système (0.0.23), confirmant ainsi les résultats [3, 5, 4].

Dans un deuxième temps, nous étudions les discrétisations en espace de (0.0.23) de type différences finies. Pour $h = 1/N > 0$, nous considérons

$$\begin{cases} \partial_t P_j + \dfrac{V_{j+1/2} - V_{j-1/2}}{h} + \sigma_j P_j = 0, \quad j \in \{1,\dots,2N-1\}, \\ \partial_t V_{j+1/2} + \dfrac{P_{j+1} - P_j}{h} + \sigma_{j+1/2} V_{j+1/2} = 0, \quad j \in \{0,\dots,2N-1\}, \\ P_0 = P_{2N} = 0. \end{cases} \tag{0.0.24}$$

Ici, P_j et σ_j sont des approximations de P et $\chi_{(1,2)}\sigma$ respectivement aux points $x_j = jh$, et $V_{j+1/2}$ et $\sigma_{j+1/2}$ de V et σ aux points $x_{j+1/2} = (j+1/2)h$.

Pour ce système, nous prouvons que l'énergie des solutions (P,V) de (0.0.24), donnée par

$$E_h(t) = \frac{h}{2}\sum_{j=0}^{2N-1}\big(|P_j(t)|^2 + |V_{j+1/2}(t)|^2\big), \tag{0.0.25}$$

n'est pas exponentiellement décroissante uniformément en $h > 0$:

Théorème 4. *Il n'existe pas de constantes strictement positives M et ν telles que, pour tout $h > 0$, les solutions de* (0.0.24) *satisfont*

$$E_h(t) \leq M\, E_h(0) \exp(-\nu t), \quad t \geq 0. \tag{0.0.26}$$

Pour cela, nous construisons des solutions de (0.0.24) localisées en dehors de la zone où l'amortissement est actif, et dont l'énergie ne peut donc pas décroître exponentiellement vite. Cette construction est basée sur celles des ondes gaussiennes [32].

Dans le cas où σ est constant sur $(1,2)$, nous fournissons également une description spectrale détaillée de l'opérateur spatial intervenant dans (0.0.24). Ainsi, nous prouvons que les fonctions propres basses fréquences sont équiréparties dans les zones $(0,1)$ et $(1,2)$, tandis que les fonctions propres hautes fréquences sont concentrées, soit dans $(0,1)$, soit dans $(1,2)$. En particulier, l'existence de fonctions propres hautes fréquences concentrées dans $(0,1)$ nie également la décroissance exponentielle de l'énergie uniformément en $h > 0$, puisque les solutions associées ne rentrent pas dans la zone où l'amortissement est effectif.

Dans un troisième et dernier temps, nous étudions une variante de (0.0.24), inspirée de [37, 36] dans laquelle un terme de viscosité numérique a été ajouté :

$$\begin{cases} \partial_t P_j + \dfrac{V_{j+1/2} - V_{j-1/2}}{h} + \sigma_j P_j - h^2 (\Delta_h P)_j = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j \in \{1, \dots, 2N\}, \\ \partial_t V_{j+1/2} + \dfrac{P_{j+1} - P_j}{h} + \sigma_{j+1/2} V_{j+1/2} - h^2 (\Delta_h V)_{j+1/2} = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j \in \{0, \dots, 2N-1\}, \\ P_0 = P_{2N} = 0, \quad V_{-1/2} = V_{1/2}, \quad V_{2N-1/2} = V_{2N+1/2}. \end{cases} \tag{0.0.27}$$

où Δ_h correspond à l'opérateur Laplacien discrétisé

$$(\Delta_h A)_j = \frac{1}{h^2}(A_{j+1} + A_{j-1} - 2A_j).$$

Dans ce cas, par une méthode des multiplicateurs, nous prouvons que l'énergie des solutions de (0.0.27) décroît exponentiellement, uniformément en $h > 0$:

Théorème 5. *Il existe des constantes strictement positives M et ν telles que, pour tout $h > 0$, les solutions de* (0.0.27) *satisfont* (0.0.26).

De plus, ce résultat est optimal, dans la mesure où l'on ne peut pas espérer des résultats similaires avec un terme visqueux plus petit, à cause de l'existence de vecteurs propres pour (0.0.24) localisés dans $(0,1)$.

Nous étudions également la possibilité de rétablir le taux de décroissance de l'énergie du système continu (0.0.23) en augmentant le terme de viscosité numérique, et donnons un résultat partiel dans cette direction. En effet, nous démontrons, sous certaines hypothèses qui seront précisées au cours du Chapitre 1, qu'il est possible de choisir le terme de viscosité numérique de façon à ce que, pour tout $h > 0$, il existe une constante M_h telle que l'énergie $E_h(t)$ des solutions de (0.0.27), définie par (0.0.25), satisfait

$$E_h(t) \leq M_h\, E_h(0) \exp\Big(- \Big(\int_1^2 \sigma - o_{h\to 0}(1) \Big) t \Big), \quad t \geq 0.$$

Chapitre 2. La méthode des éléments finis mixtes sur des maillages non uniformes

Ce chapitre propose l'étude des propriétés d'observabilité de l'équation des ondes unidimensionnelle, discrétisée par la méthode des éléments finis mixtes, mais sur des maillages non uniformes. A notre connaissance, c'est à ce jour le seul exemple où la théorie a pu être effectuée à ce niveau de détail pour des maillages non uniformes.

Considérons l'équation des ondes unidimensionnelle

$$\begin{cases} \partial_{tt}^2 u - \partial_{xx}^2 u = 0, & (x,t) \in (0,1)\times\mathbb{R}, \\ u(0,t) = u(1,t) = 0, & t \in \mathbb{R}, \\ u(x,0) = u^0(x),\ \partial_t u(x,0) = u^1(x), & x \in (0,1), \end{cases} \tag{0.0.28}$$

avec $(u^0, u^1) \in H_0^1(0,1)\times L^2(0,1)$.

L'énergie des solutions de (0.0.28), donnée par

$$E(t) = \frac{1}{2}\ \|\partial_t u(t)\|_{L^2(0,1)}^2 + \frac{1}{2}\|\partial_x u\|_{L^2(0,1)}^2,$$

est constante.

De plus, il est bien connu (cf. [27, 25]) que pour tout temps $T > 2$, il existe des constantes strictement positives k_T et K_T telles que les solutions de (0.0.28) satisfont

$$k_T E(0) \le \int_0^T |\partial_x u(0,t)|^2\ \mathrm{dt} \le K_T E(0).$$

Discrétisons (0.0.28) sur un maillage non uniforme $\mathcal{S}_n$ donné par $n+2$ points

$$0 = x_{0,n} < x_{1,n} < \cdots < x_{n,n} < x_{n+1,n} = 1, \qquad h_{j+1/2,n} = x_{j+1,n} - x_{j,n}, \quad j \in \{0,\cdots,n\}. \tag{0.0.29}$$

La méthode des éléments finis mixtes donne alors le système

$$\begin{cases} \dfrac{h_{j-1/2,n}}{4}(\ddot{u}_{j-1,n} + \ddot{u}_{j,n}) + \dfrac{h_{j+1/2,n}}{4}(\ddot{u}_{j,n} + \ddot{u}_{j+1,n}) \\ \qquad\qquad = \dfrac{u_{j+1,n} - u_{j,n}}{h_{j+1/2,n}} - \dfrac{u_{j,n} - u_{j-1,n}}{h_{j-1/2,n}}, \quad j = 1,\cdots n,\ t \in \mathbb{R}, \\ u_{0,n}(t) = u_{n+1,n}(t) = 0, \quad t \in \mathbb{R}, \\ u_j(0) = u_{j,n}^0, \qquad \dot{u}_j(0) = u_{j,n}^1, \qquad j = 1,\cdots,n. \end{cases} \tag{0.0.30}$$

L'énergie des solutions u_n de (0.0.30), donnée par

$$E_n(t) = \frac{1}{2}\sum_{j=0}^{n} h_{j+1/2,n}\left(\frac{u_{j+1,n}(t) - u_{j,n}(t)}{h_{j+1/2,n}}\right)^2 + \frac{1}{2}\sum_{j=0}^{n} h_{j+1/2,n}\left(\frac{\dot{u}_{j+1,n}(t) + \dot{u}_{j,n}(t)}{2}\right)^2, \tag{0.0.31}$$

est alors constante.

D'après les travaux [8, 9], pour des maillages uniformes, en tout temps $T > 2$, on peut trouver des constantes strictement positives k_T et K_T telles que les solutions de (0.0.30) (toujours sur des maillages uniformes) satisfont

$$k_T E_n(0) \le \int_0^T \left(\left|\frac{u_{1,n}(t)}{h_{1/2,n}}\right|^2 + |\dot{u}_{1,n}(t)|^2\right)\ \mathrm{dt} \le K_T E_n(0). \tag{0.0.32}$$

Nous démontrons que ce résultat s'étend en fait pour une large classe de maillages non uniformes. Introduisons la notion de régularité d'un maillage :

Définition 6. Soit un maillage $\mathcal{S}_n$ donné par $n+2$ points comme dans (0.0.29). La régularité de $\mathcal{S}_n$ est définie par

$$\mathrm{Reg}(\mathcal{S}_n) = \frac{\max_j\{h_{j+1/2,n}\}}{\min_j\{h_{j+1/2,n}\}}. \tag{0.0.33}$$

Pour $M \geq 1$, on dira qu'un maillage $\mathcal{S}_n$ est de régularité M si $\mathrm{Reg}(\mathcal{S}_n) \leq M$.

Nous démontrons alors le résultat suivant :

Théorème 7. *Soit $M \geq 1$ et $(\mathcal{S}_n)$ une suite de maillages de régularité M.*
Alors, pour tout temps $T > 2$, il existe des constantes strictement positives k_T et K_T telles que les solutions de (0.0.30) *satisfont, uniformément en n, les estimées* (0.0.32).

De façon similaire, nous prouvons le même type de résultat en ce qui concerne une observation distribuée sur un sous-intervalle $\omega \subset (0,1)$.

La preuve du Théorème 7 est basée sur l'étude spectrale de (0.0.30), qui se trouve être particulièrement explicite. Notamment, il est possible de démontrer que les valeurs propres λ_n^k de l'opérateur en (0.0.30) satisfont, pour n'importe quel maillage, la propriété suivante, dite de séparation spectrale, ou de *spectral gap* :

$$\min_{k\in\{1,\cdots,n-1\}}\{\lambda_n^{k+1} - \lambda_n^k\} \geq \pi.$$

En particulier, le Lemme d'Ingham (cf. [24]) sur les séries trigonométriques montre qu'il suffit alors de prouver des propriétés uniformes d'observabilité sur les fonctions propres. En utilisant l'expression explicite des fonctions propres, nous arrivons alors à montrer (0.0.32), à condition que les maillages soient M réguliers.

De plus, nous montrons aussi que la condition de M régularité sur les maillages est, en un certain sens, optimale pour les propriétés d'admissibilité et d'observabilité discrètes.

Enfin, nous exhibons les applications du Théorème 7 pour des problèmes de contrôlabilité et de stabilisation, basées sur la dualité HUM et sur les résultats [22, 27] présentés ci-dessus.

Partie II. Discrétisation en temps de systèmes conservatifs

Dans cette partie, nous considérons un couple d'opérateurs $(\mathcal{A}, B)$, et étudions les discrétisations en temps de (0.0.1)-(0.0.2).

Nous supposons, dans toute cette partie, que le système (0.0.1)-(0.0.2) est admissible et observable.

Dans toute cette partie, nous supposons également que l'opérateur $\mathcal{A}$ est anti-adjoint, et à résolvente compacte. Il s'ensuit que le spectre de $\mathcal{A}$ est constitué uniquement de valeurs propres $i\mu_j$, avec $\mu_j \in \mathbb{R}$. De plus, les vecteurs propres Φ_j correspondants peuvent être choisis de façon à former une base orthonormale de $\mathfrak{X}$.

Notre but est de formuler, de la façon la plus générale possible, des propriétés d'observabilité et d'admissibilité uniformes en le paramètre de discrétisation en temps.

Chapitre 3. Propriétés d'observabilité

Pour fixer les idées, considérons la discrétisation standard de (0.0.1)-(0.0.2) :

$$\begin{cases} \dfrac{z^{k+1}-z^k}{\triangle t} = \mathcal{A}\Big(\dfrac{z^{k+1}+z^k}{2}\Big), \quad \text{dans } \mathfrak{X},\ k\in\mathbb{Z}, \\ z^0 = z_0, \end{cases} \qquad y^k = Bz^k, \quad k\triangle t \geq 0. \tag{0.0.34}$$

Remarquons que l'énergie des solutions de (0.0.34), définie par

$$E^k = \frac{1}{2}\left\|z^k\right\|_{\mathfrak{X}}^2,$$

est constante.

Introduisons alors, pour $s>0$, les classes filtrées

$$\mathcal{C}_s(\mathcal{A}) = \text{vect}\{\Phi_j \text{ tels que les valeurs propres correspondantes } i\mu_j \text{ vérifient } |\mu_j|\leq s\}. \tag{0.0.35}$$

Pour le système (0.0.34), nous prouvons le théorème suivant :

Théorème 8. *Supposons que $B\in\mathfrak{L}(\mathcal{D}(\mathcal{A}),\mathcal{Y})$. Fixons $\delta>0$.*
Alors il existe un temps $T_\delta>0$, et des constantes strictement positives k_δ et K_δ telles que, pour tout $\triangle t>0$, les solutions z de (0.0.34) *avec donnée initiale $z_0\in\mathcal{C}_{\delta/\triangle t}(\mathcal{A})$ satisfont :*

$$k_\delta \left\|z_0\right\|_{\mathfrak{X}}^2 \leq \triangle t \sum_{k\triangle t\in[0,T_\delta]} \left\|B\Big(\frac{z^k+z^{k+1}}{2}\Big)\right\|_{\mathcal{Y}}^2 \leq K_\delta \left\|z_0\right\|_{\mathfrak{X}}^2. \tag{0.0.36}$$

Le résultat d'observabilité uniforme (0.0.36) est optimal au vu de [43]. En effet, il est prouvé dans [43] que, pour l'équation des ondes, on ne peut pas espérer de résultat d'observabilité uniforme en $\triangle t$ dans des classes filtrées $\mathcal{C}_{1/(\triangle t)^{1+\epsilon}}(\mathcal{A})$ avec $\epsilon>0$.

La preuve du Théorème 8 est basée sur une méthode spectrale introduite dans [6, 29]. Dans [6, 29], il est en effet prouvé que l'observabilité de (0.0.1)-(0.0.2) est équivalente à l'existence de deux constantes positives m et M telles que

$$M^2\left\|(\mathcal{A}-i\omega)z\right\|_{\mathfrak{X}}^2 + m^2\left\|Bz\right\|_{\mathcal{Y}}^2 \geq \left\|z\right\|_{\mathfrak{X}}^2, \qquad \forall z\in\mathcal{D}(\mathcal{A}),\ \forall\omega\in\mathbb{R}. \tag{0.0.37}$$

La démonstration de l'inégalité d'observabilité dans (0.0.36) suit essentiellement celle donnée dans [6, 29] pour montrer que l'estimée de la résolvante (0.0.37) implique l'observabilité de (0.0.1)-(0.0.2).

Pour prouver l'inégalité d'admissibilité dans (0.0.36), nous introduisons un nouveau critère spectral équivalent à l'admissibilité de (0.0.1)-(0.0.2). À nouveau, en suivant la preuve du cas continu (0.0.1)-(0.0.2), nous démontrons l'inégalité d'admissibilité dans (0.0.36).

La méthode que nous développons pour prouver le Théorème 8 présente de nombreux intérêts.

Notre méthode s'applique en effet à de nombreux schémas numériques, et pas seulement à des systèmes discrétisés selon (0.0.34). Pour être plus précis, nous prouvons que, pour une large gamme de méthodes de discrétisation en temps incluant entre autres la méthode de Newmark et la méthode

de Gauss d'ordre quatre, des propriétés d'observabilité et d'admissibilité sont vérifiées uniformément en $\triangle t > 0$.

Grâce aux estimées explicites sur les constantes intervenant dans le Théorème 8, nous pouvons également considérer les propriétés d'admissibilité et d'observabilité des discrétisations en temps de familles de systèmes uniformément admissibles et observables. Notamment, si les systèmes (0.0.15) sont admissibles et observables au sens de (0.0.18) uniformément en $h > 0$ dans la classe $\tilde{\mathfrak{X}}_h$, alors, pour tout $\delta > 0$, il existe un temps $T_\delta > 0$, et des constantes strictement positives k_δ et K_δ tels que, pour tout $h, \triangle t > 0$, les solutions z_h de

$$\begin{cases} \dfrac{z_h^{k+1} - z_h^k}{\triangle t} = \mathcal{A}_h\Big(\dfrac{z_h^{k+1} + z_h^k}{2}\Big), \quad \text{dans } \ \mathfrak{X}_h, \ k \in \mathbb{Z}, \\ z_h^0 = z_{0h}, \end{cases}$$

avec $z_{0h} \in \mathcal{C}_{\delta/\triangle t}(\mathcal{A}_h) \cap \tilde{\mathfrak{X}}_h$ satisfont

$$k_\delta \left\| z_{0h} \right\|_{\mathfrak{X}_h}^2 \leq \triangle t \sum_{k\triangle t \in [0, T_\delta]} \left\| B_h\Big(\frac{z_h^k + z_h^{k+1}}{2}\Big) \right\|_{\mathcal{Y}_h}^2 \leq K_\delta \left\| z_{0h} \right\|_{\mathfrak{X}_h}^2 . \tag{0.0.38}$$

Ce résultat permet de déduire instantanément des propriétés d'admissibilité et d'observabilité uniformes pour des systèmes totalement discrétisés à partir de l'étude des systèmes semi-discrétisés en espace (et donc continus en temps) correspondants.

A notre connaissance, il n'existait auparavant que très peu de références bibliographiques sur les propriétés d'observabilité de systèmes conservatifs discrétisés en temps avant notre travail, sinon l'article [30] qui étudie l'équation des ondes totalement discrétisée en dimension un, et l'article [43] qui étudie l'équation des ondes dans un domaine borné $\Omega \subset \mathbb{R}^d$ semi-discrétisée en temps, mais avec une méthodologie qui ne permet pas d'envisager facilement des extensions aux cas complètement discrétisés.

Chapitre 4. Limites visqueuses de systèmes exponentiellement stables

Ici, nous délaissons temporairement les problèmes introduits par les méthodes de discrétisation, afin de nous concentrer sur l'étude des différents termes de viscosité que nous pouvons introduire dans (0.0.11) de façon à préserver les propriétés dissipatives des systèmes ainsi obtenus.

Les méthodes que nous développons dans ce chapitre sont en fait des versions simplifiées de celles utilisées au Chapitre 5 pour des systèmes discrétisés en temps. Leur principal intérêt est qu'elles permettent de prouver des résultats de stabilisation y compris pour des systèmes (0.0.11) dont le système conservatif associé (0.0.1)-(0.0.2) est seulement observable aux basses fréquences.

Ici, ainsi qu'au Chapitre 5, nous supposons que B appartient à $\mathfrak{L}(\mathfrak{X}, \mathcal{Y})$. Rappelons que l'opérateur $\mathcal{A}$ est supposé anti-adjoint.

Rappelons aussi que, dans ce cas, le système (0.0.1)-(0.0.2) est observable si et seulement si le système (0.0.11) est exponentiellement stable.

Le but de ces deux chapitres est donc de fournir des méthodes de discrétisation en temps de (0.0.11) de façon à conserver la décroissance exponentielle de l'énergie des systèmes discrétisés uniformément en le pas de temps.

Pour cela, il est nécessaire de réinterpréter les résultats du Chapitre 3 en termes de stabilisation. Formellement, le Théorème 8 indique que les basses fréquences (jusqu'à l'ordre $1/\triangle t$) sont efficacement amorties par l'opérateur B^*B. Nous allons donc introduire dans les équations un terme visqueux qui aura pour but de dissiper efficacement les hautes fréquences, de la même manière qu'au Chapitre 1.

Il faut alors éviter des phénomènes d'*overdamping*, qui peuvent apparaître pour ces équations dissipatives (cf. [10]), et qui pourraient empêcher des propriétés de stabilisation uniformes. Nous nous intéressons donc, dans un premier temps sur des modèles continus, aux divers types de viscosité $\mathcal{V}$ qui n'introduisent pas d'effet d'*overdamping*.

Nous introduisons donc, pour $\varepsilon > 0$, les systèmes

$$\dot{z} = \mathcal{A}z + \varepsilon \mathcal{V}_\varepsilon z - B^*Bz, \quad t \geq 0, \qquad z(0) = z_0 \in \mathfrak{X}, \tag{0.0.39}$$

où $\mathcal{V}_\varepsilon$ est un terme de viscosité qui peut dépendre de ε, et que nous préciserons plus tard.

L'énergie des solutions z de (0.0.39), définie par

$$E(t) = \frac{1}{2} \|z(t)\|^2_{\mathfrak{X}},$$

satisfait la loi de décroissance

$$\frac{dE}{dt}(t) = - \|Bz\|^2_{\mathcal{Y}} + \varepsilon < \mathcal{V}_\varepsilon z, z >_{\mathfrak{X}}, \quad t \geq 0.$$

Rappelons que, sous nos hypothèses, quand $\varepsilon = 0$, le système (0.0.39), qui correspond alors au système sans terme visqueux (0.0.11), est exponentiellement stable.

Pour énoncer notre résultat, nous introduisons la projection orthogonale $\pi_{1/\sqrt{\varepsilon}}$ dans $\mathfrak{X}$ sur $\mathcal{C}_{1/\sqrt{\varepsilon}}(\mathcal{A})$.

Nous prouvons alors que, pour une large classe de termes de viscosité, le système (0.0.39) est exponentiellement stable uniformément en ε :

Théorème 9. *Supposons que les opérateurs de viscosité $\mathcal{V}_\varepsilon$ satisfont*

1. *$\mathcal{V}_\varepsilon$ est un opérateur auto-adjoint défini négatif.*
2. *La projection $\pi_{1/\sqrt{\varepsilon}}$ et l'opérateur $\mathcal{V}_\varepsilon$ commutent.*
3. *Il existe des constantes strictement positives c et C telles que pour tout $\varepsilon > 0$,*

$$\sqrt{\varepsilon} \left\| \left(\sqrt{-\mathcal{V}_\varepsilon} \right) z \right\|_{\mathfrak{X}} \leq C \|z\|_{\mathfrak{X}}, \ \forall z \in \mathcal{C}_{1/\sqrt{\varepsilon}}(\mathcal{A}), \ \text{et } \sqrt{\varepsilon} \left\| \left(\sqrt{-\mathcal{V}_\varepsilon} \right) z \right\|_{\mathfrak{X}} \geq c \|z\|_{\mathfrak{X}}, \ \forall z \in \mathcal{C}_{1/\sqrt{\varepsilon}}(\mathcal{A})^\perp.$$

Alors l'énergie des solutions de (0.0.39) *est exponentiellement décroissante au sens de* (0.0.14)*, uniformément en le paramètre de viscosité $\varepsilon \geq 0$.*

Des exemples d'opérateurs visqueux satisfaisant les hypothèses du Théorème 9 sont

$$\varepsilon \mathcal{V}_\varepsilon = \varepsilon \mathcal{A}^2, \quad \varepsilon \mathcal{V}_\varepsilon = \frac{\varepsilon \mathcal{A}^2}{I - \varepsilon \mathcal{A}^2}, \quad , \varepsilon \mathcal{V}_\varepsilon = \sqrt{\varepsilon} |\mathcal{A}|, \quad \ldots.$$

La preuve du Théorème 9 est basée sur celle de [22], qui lie les propriétés de décroissance exponentielle de l'énergie des solutions de (0.0.11) à l'observabilité du système (0.0.1)-(0.0.2). Dans notre cas cependant, à cause du caractère éventuellement non borné de l'opérateur de viscosité $\mathcal{V}_\varepsilon$, nous ne pouvons pas nous ramener à l'inégalité d'observabilité (0.0.4) pour (0.0.1)-(0.0.2).

Adaptant [22], nous étudions plutôt le système visqueux suivant

$$\dot{u} = \mathcal{A}u + \varepsilon \mathcal{V}_\varepsilon u, \quad t \in \mathbb{R}, \qquad u(0) = u_0 \in \mathfrak{X}, \tag{0.0.40}$$

et démontrons alors qu'il existe un temps $T > 0$ et une constante strictement positive k_T indépendants de ε tels que les solutions u de (0.0.40) satisfont

$$k_T \left\| u_0 \right\|_{\mathfrak{X}}^2 \leq \int_0^T \left\| Bu(t) \right\|_{\mathcal{Y}}^2 \, \mathrm{dt} + \varepsilon \int_0^T \left\| \Big(\sqrt{-\mathcal{V}_\varepsilon} \Big) u(t) \right\|_{\mathfrak{X}}^2 \, \mathrm{dt}. \tag{0.0.41}$$

Cette inégalité d'observabilité est en effet équivalente à la propriété de stabilisation uniforme pour les systèmes (0.0.39).

Pour démontrer (0.0.41), nous utilisons un argument de découplage des solutions du système (0.0.40) en basses et hautes fréquences.

Aux basses fréquences, en utilisant la méthode de [22], comme $\mathcal{V}_\varepsilon$ se comporte comme un opérateur borné, nous démontrons (0.0.41) à partir des propriétés d'observabilité du système (0.0.1)-(0.0.2).

Pour les hautes fréquences, nous utilisons la dissipation induite par le terme de viscosité dans (0.0.40) pour obtenir l'inégalité (0.0.41).

Chapitre 5. Approximations uniformément exponentiellement stables de systèmes dissipatifs

Il s'agit ici d'essayer d'appliquer les résultats du Chapitre 3 pour mettre au point des schémas numériques semi-discrets en temps pour lesquels nous pouvons garantir la décroissance exponentielle de l'énergie, uniformément en le paramètre de discrétisation en temps $\triangle t$.

La méthode que nous avons mise au point au Chapitre 4 sert de base à cette partie. En effet, au Chapitre 4, nous n'utilisons les propriétés d'observabilité du système (0.0.1)-(0.0.2) qu'aux basses fréquences, les hautes fréquences étant traitées via l'introduction d'un terme de viscosité.

Pour la discrétisation en temps introduite en (0.0.34), nous avons précisément démontré que les propriétés d'observabilité de (0.0.34) sont satisfaites aux basses fréquences $\mathcal{C}_{1/\triangle t}(\mathcal{A})$.

En conséquence, nous allons introduire dans les schémas numériques que nous allons considérer un terme de viscosité numérique qui amortit efficacement les fréquences qui sont de l'ordre de $1/\triangle t$ et plus, sans changer la dynamique du système aux basses fréquences. Ainsi, nous allons être amenés à considérer des discrétisations formelles de

$$\dot{z} = \mathcal{A}z - B^*Bz + (\triangle t)^2 \mathcal{A}^2 z, \tag{0.0.42}$$

menant par exemple au schéma numérique

$$\left\{ \begin{array}{rcl} \dfrac{\tilde{z}^{k+1} - z^k}{\triangle t} & = & \mathcal{A}\Big(\dfrac{z^k + \tilde{z}^{k+1}}{2} \Big), \quad k \in \mathbb{N}, \\ \dfrac{z^{k+1} - \tilde{z}^{k+1}}{\triangle t} & = & -B^*Bz^{k+1} + (\triangle t)^2 \mathcal{A}^2 z^{k+1}, \quad k \in \mathbb{N}, \\ z^0 & = & z_0. \end{array} \right. \tag{0.0.43}$$

Remarquons que pour obtenir le système discrétisé (0.0.43), nous avons décomposé l'opérateur $\mathcal{A} - B^*B + (\triangle t)^2 \mathcal{A}^2$ en une partie « conservative » $\mathcal{A}$ et une partie « dissipative » $-B^*B + (\triangle t)^2 \mathcal{A}^2$

que nous avons discrétisées différemment. En effet, le schéma du point milieu est approprié pour la discrétisation de systèmes conservatifs puisqu'il préserve la propriété de conservation de l'énergie. Cependant, ce schéma numérique n'est pas adapté à la discrétisation de systèmes dissipatifs, car il ne préserve pas les propriétés de dissipation des hautes fréquences. C'est pourquoi nous préférons utiliser une méthode d'Euler implicite pour discrétiser l'opérateur de dissipation $-B^*B + (\triangle t)^2\mathcal{A}^2$.

Nous pouvons alors démontrer, en raffinant l'argument utilisé au Chapitre 4, le théorème suivant :

Théorème 10. *Il existe des constantes strictement positives $\mu > 0$ et $\nu > 0$ telles que pour tout $\triangle t > 0$, les solutions z de* (0.0.43) *satisfont*

$$\left\|z^k\right\|_{\mathfrak{X}}^2 \leq \mu \left\|z^0\right\|_{\mathfrak{X}}^2 \exp(-\nu k \triangle t), \quad k \in \mathbb{N}.$$

De même qu'au Chapitre 4, nous obtenons des résultats similaires pour des termes de viscosité plus généraux, ainsi que pour certaines autres formes de discrétisations de (0.0.42).

De même qu'au Chapitre 3, nos résultats s'appliquent également pour des familles d'opérateurs $(\mathcal{A}_h, B_h)$ uniformément observables (en $h > 0$) au sens de (0.0.18) dans une classe $\tilde{\mathfrak{X}}_h = \mathcal{C}_{\eta/h^\sigma}(\mathcal{A}_h)$ pour η et σ des constantes strictement positives indépendantes de $h > 0$, et telles que $\sup_h \|B_h\|_{\mathfrak{L}(\mathfrak{X}_h, \mathcal{Y}_h)} < \infty$. Sous ces hypothèses en effet, en posant $\varepsilon = \min\{(\triangle t)^2, h^{2\sigma}\}$, il existe des constantes strictement positives $\mu > 0$ et $\nu > 0$ indépendantes de $h > 0$ telles que, pour tout $h, \triangle t > 0$ les solutions z_h de

$$\left\{\begin{array}{rcl} \dfrac{z_h^{k+1} - z_h^k}{\triangle t} & = & \mathcal{A}_h\Big(\dfrac{z_h^k + z_h^{k+1}}{2}\Big), \quad k \in \mathbb{N}, \\ \dfrac{z_h^{k+1} - \tilde{z}_h^{k+1}}{\triangle t} & = & -B_h^* B_h z^{k+1} + \varepsilon \mathcal{A}_h^2 z_h^{k+1}, \quad k \in \mathbb{N}, \\ z_h^0 & = & z_{0h} \in \mathfrak{X}_h. \end{array}\right. \tag{0.0.44}$$

satisfont

$$\left\|z_h^k\right\|_{\mathfrak{X}_h}^2 \leq \mu \left\|z_h^0\right\|_{\mathfrak{X}_h}^2 \exp(-\nu k \triangle t), \quad k \in \mathbb{N}.$$

Dans ce cas, le terme de viscosité numérique est ajusté de façon à amortir efficacement les fréquences de l'ordre de $1/\sqrt{\varepsilon}$ et plus, à partir desquelles les propriétés d'observabilité du système conservatif correspondants ne sont plus assurées.

Nous présentons également quelques applications précises de nos résultats, notamment pour des équations des ondes amorties.

Partie III. Discrétisation en espace de systèmes conservatifs

Dans cette partie, nous considérons deux modèles abstraits conservatifs discrétisés selon la méthode des éléments finis, correspondants à des équations de type Schrödinger et ondes, que nous écrivons de façon générique sous les formes

$$\left\{\begin{array}{l} i\dot{z} = A_0 z, \quad t \geq 0, \\ z(0) = z_0, \end{array}\right. \qquad y(t) = Bz(t),\ t \geq 0, \tag{0.0.45}$$

et, respectivement,

$$\left\{\begin{array}{l} \ddot{u} + A_0 u = 0, \quad t \geq 0, \\ u(0) = u_0, \quad \dot{u}(0) = u_1. \end{array}\right. \qquad y(t) = B\dot{u}(t), \quad t \geq 0, \tag{0.0.46}$$

Dans les deux cas, A_0 est un opérateur auto-adjoint défini positif sur un espace de Hilbert H, et l'opérateur B est supposé appartenir à $\mathfrak{L}(\mathcal{D}(A_0^\kappa), \mathcal{Y})$, avec $\kappa < 1/2$, $\mathcal{Y}$ étant un espace de Hilbert.

Pour décrire la méthode des éléments finis, pour tout $h > 0$, nous nous donnons un espace vectoriel V_h de dimension finie n_h et une application linéaire injective $\pi_h : V_h \to H$. Pour tout $h > 0$, l'application π_h induit alors un produit scalaire naturel $< \cdot, \cdot >_h = < \pi_h \cdot, \pi_h \cdot >_H$ sur V_h^2.

Nous supposons que, pour tout $h > 0$, $\pi_h(V_h) \subset \mathcal{D}(A_0^{1/2})$. Nous définissons alors l'opérateur $A_{0h} : V_h \to V_h$ correspondant à la discrétisation de l'opérateur A_0 par

$$< A_{0h}\phi_h, \psi_h >_h = < A_0^{1/2}\pi_h\phi_h, A_0^{1/2}\pi_h\psi_h >_H, \quad \forall (\phi_h, \psi_h) \in V_h^2, \tag{0.0.47}$$

ce qui est équivalent à poser $A_{0h} = \pi_h^* A_0 \pi_h$.

Nous sommes alors amenés à étudier les propriétés d'observabilité des systèmes discrétisés suivants, discrétisations respectives de (0.0.45) et de (0.0.46) :

$$\begin{cases} i\dot{z}_h = A_{0h} z_h, \quad t \geq 0, \\ z_h(0) = z_{0h} \in V_h, \end{cases} \qquad y_h(t) = B\pi_h z_h(t), \ t \geq 0, \tag{0.0.48}$$

et

$$\begin{cases} \ddot{u}_h + A_{0h} u_h = 0, \quad t \geq 0, \\ u_h(0) = u_{0h}, \quad \dot{u}_h(0) = u_{1h}. \end{cases} \qquad y_h(t) = B\pi_h \dot{u}_h(t), \quad t \geq 0, \tag{0.0.49}$$

La convergence des schémas numériques (0.0.48) et (0.0.49) se déduit alors des propriétés de π_h (cf. [35]) : Notamment, on suppose qu'il existe des constantes positives C_0 et $\theta > 0$ telles que

$$\begin{cases} \left\| A_0^{1/2}(\pi_h\pi_h^* - I)\phi \right\|_H \leq C_0 \left\| A_0^{1/2}\phi \right\|_H, \quad \forall \phi \in \mathcal{D}(A_0^{1/2}), \\ \left\| A_0^{1/2}(\pi_h\pi_h^* - I)\phi \right\|_H \leq C_0 h^\theta \left\| A_0\phi \right\|_H, \quad \forall \phi \in \mathcal{D}(A_0). \end{cases} \tag{0.0.50}$$

En pratique, $\theta = 1$ quand A_0 est l'opérateur de Laplace avec conditions aux limites de Dirichlet pour des éléments finis P1 sur des triangulations régulières.

Comme A_{0h} défini par (0.0.47) est un opérateur auto-adjoint défini positif, son spectre est formé d'une suite de valeurs propres

$$0 < \lambda_1^h \leq \lambda_2^h \leq \cdots \leq \lambda_{n_h}^h, \tag{0.0.51}$$

et de vecteurs propres $(\Psi_j^h)_{1 \leq j \leq n_h}$ que nous pouvons prendre normalisés dans V_h. On introduit alors, pour $s > 0$, l'espace filtré

$$\mathcal{F}_h(s) = \text{vect}\Big\{ \Psi_j^h \text{ tels que les valeurs propres correspondantes satisfont } |\lambda_j^h| \leq s \Big\}.$$

Chapitre 6. Équations de type Schrödinger

Pour les équations de type Schrödinger (0.0.45), nous obtenons les résultats suivants concernant les propriétés d'admissibilité et d'observabilité de (0.0.48) :

Théorème 11. *Posons*

$$\sigma = \theta \min\Big\{ 2(1 - 2\kappa), \frac{2}{5} \Big\}. \tag{0.0.52}$$

Admissibilité : *Supposons que le système* (0.0.45) *est admissible.*
Alors, quels que soient $\eta > 0$ et $T > 0$, il existe une constante positive $K_{T,\eta} > 0$ telle que pour tout $h > 0$, toute solution de (0.0.48) *avec donnée initiale*

$$z_{0h} \in \mathcal{F}_h(\eta/h^\sigma) \tag{0.0.53}$$

satisfait

$$\int_0^T \|B\pi_h z_h(t)\|_{\mathcal{Y}}^2 \ \mathrm{dt} \leq K_{T,\eta} \|z_{0h}\|_h^2 . \tag{0.0.54}$$

Observabilité : *Supposons que le système* (0.0.45) *est admissible et observable.*
Alors il existe une constante $\epsilon > 0$, un temps T^* et une constante strictement positive $k_* > 0$ tels que pour tout $h > 0$, toute solution de (0.0.48) *avec donnée initiale*

$$z_{0h} \in \mathcal{F}_h(\epsilon/h^\sigma) \tag{0.0.55}$$

satisfait

$$k_* \|z_{0h}\|_h^2 \leq \int_0^{T^*} \|B\pi_h z_h(t)\|_{\mathcal{Y}}^2 \ \mathrm{dt}. \tag{0.0.56}$$

La preuve du Théorème 11 est basée sur des caractérisations spectrales. La propriété d'admissibilité (0.0.54) est déduite du critère spectral introduit au Chapitre 3, que nous reformulons sous la forme d'une estimée de résolvante puis d'une inégalité d'interpolation. La propriété d'observabilité (0.0.56), quant à elle, est déduite d'une relecture des inégalités de résolvantes (0.0.37) introduites dans [6, 29] en termes d'inégalités d'interpolation.

L'intérêt majeur de ce résultat est qu'il ne fait intervenir ni la structure du maillage ni la dimension, et donc fournit une méthode robuste pour traiter les questions d'admissibilité et d'observabilité des systèmes discrétisés. Il est toutefois à noter que ce résultat n'est probablement pas optimal, mais cette question reste, pour l'instant, largement ouverte.

Nous détaillons aussi quelques exemples d'applications du Théorème 11, que nous combinons avec les résultats démontrés précédemment aux Chapitres 3 et 5.

Notamment, nous déduisons du Théorème 11 et des résultats du Chapitre 3 des propriétés d'admissibilité et d'observabilité uniformes en les paramètres de discrétisation en espace et en temps pour des discrétisations en temps déduites de (0.0.48).

Nous montrons aussi comment ce théorème s'applique en théorie du contrôle, en proposant deux procédés permettant de calculer numériquement des approximations des contrôles HUM du système continu. Ces procédés sont tout deux basés sur des mécanismes de filtrage, l'un impliquant de connaître une méthode efficace de filtrage au niveau discret, l'autre via une méthode de régularisation de Tychonoff basée sur les travaux [21, 44].

Enfin, nous combinons les résultats du Chapitre 5 avec le Théorème 11 pour fournir, lorsque B est dans $\mathfrak{L}(H, \mathcal{Y})$, des systèmes discrétisés déduits de

$$i\dot{z} = A_0 z - iB^* B z, \quad t \geq 0,$$

dont l'énergie est exponentiellement décroissante, uniformément en les paramètres de discrétisation.

Chapitre 7. Équations de type ondes

Pour les équations de type ondes (0.0.46), nous obtenons les résultats suivants concernant les propriétés d'admissibilité et d'observabilité de (0.0.49) :

Théorème 12. *Posons*

$$\varsigma = \theta \min\Big\{2(1-2\kappa), \frac{2}{3}\Big\}. \tag{0.0.57}$$

Admissibilité : *Supposons que le système* (0.0.46) *est admissible.*
Alors, quels que soient $\eta > 0$ et $T > 0$, il existe une constante positive $K_{T,\eta} > 0$ telle que pour tout $h > 0$, toute solution de (0.0.49) *avec donnée initiale*

$$(u_{0h}, u_{1h}) \in \mathcal{F}_h(\eta/h^\varsigma)^2 \tag{0.0.58}$$

satisfait

$$\int_0^T \|B\pi_h \dot{u}_h(t)\|_{\mathcal{Y}}^2 \ \mathrm{dt} \leq K_{T,\eta}\Big(\Big\|A_{0h}^{1/2} u_{0h}\Big\|_h^2 + \|u_{1h}\|_h^2 \Big). \tag{0.0.59}$$

Observabilité : *Supposons que le système* (0.0.46) *est admissible et observable.*
Alors il existe une constante $\epsilon > 0$, un temps T^* et une constante strictement positive $k_* > 0$ tels que pour tout $h > 0$, toute solution de (0.0.49) *avec donnée initiale*

$$(u_{0h}, u_{1h}) \in \mathcal{F}_h(\epsilon/h^\varsigma)^2 \tag{0.0.60}$$

satisfait

$$k_*\Big(\Big\|A_{0h}^{1/2} u_{0h}\Big\|_h^2 + \|u_{1h}\|_h^2 \Big) \leq \int_0^{T^*} \|B\pi_h \dot{u}_h(t)\|_{\mathcal{Y}}^2 \ \mathrm{dt}. \tag{0.0.61}$$

Là encore, notre preuve est basée sur des critères spectraux, que nous écrivons sous la forme d'inégalités d'interpolation. Cette fois-ci cependant, la méthode spectrale que nous utilisons pour démontrer la propriété d'observabilité (0.0.61) est basée sur une version précisée des résultats [28, 33, 40]. À nouveau, ce résultat présente l'intérêt de s'appliquer dans un grand nombre de situations concrètes, mais son optimalité n'est pas garantie.

Nous donnons également quelques applications du Théorème 12, comme précédemment. En utilisant les résultats du Chapitre 3, nous déduisons des propriétés d'observabilité pour des discrétisations en espace et en temps de (0.0.46). De même qu'au Chapitre 6, nous donnons aussi des applications du Théorème 12 pour ce qui concerne des problèmes de contrôle et de stabilisation.

Enfin, nous déduisons du Théorème 12 une amélioration du Théorème 11 dans le cas où le système (0.0.46) est admissible et observable. Pour cela, nous utilisons, au niveau discret, une variante des résultats de [29] qui prouvent, notamment, que si le système (0.0.46) est observable, alors le système (0.0.45) est observable.

Partie IV : Chapitre 8. Étude d'une équation de la chaleur avec potentiel singulier

Dans cette partie, nous considérons un problème assez différent de ceux considérés jusqu'à présent, puisque nous allons étudier une équation continue de type parabolique. Cela dit, les thématiques centrales de contrôle, de stabilisation et d'observabilité restent les mêmes.

Fixons un domaine régulier $\Omega \subset \mathbb{R}^N$ avec $N \geq 3$ tel que $0 \in \Omega$, et un sous ouvert non-vide $\omega \subset \Omega$. Nous nous proposons d'étudier les propriétés de contrôle et de stabilisation de l'équation

$$\begin{cases} \partial_t u - \Delta_x u - \dfrac{\mu}{|x|^2} u = f, & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) = u_0(x), & x \in \Omega, \end{cases} \tag{0.0.62}$$

où $u_0 \in L^2(\Omega)$. La fonction $f \in L^2(0,T;H^{-1}(\Omega))$ est le contrôle, que nous supposons à support dans $\bar{\omega}$ (au sens des distributions).

Avant d'aller plus loin, il est nécessaire de préciser que la définition même d'une solution de (0.0.62) n'est pas claire, le caractère bien posé du problème étant lié à la valeur du paramètre μ. Quand $\mu \leq \mu^*(N) = (N-2)^2/4$, en utilisant l'inégalité de Hardy

$$\forall u \in H_0^1(\Omega), \qquad \mu^*(N) \int_\Omega \frac{u^2}{|x|^2} \ \mathrm{dx} \leq \int_\Omega |\nabla u|^2 \ \mathrm{dx}, \tag{0.0.63}$$

on peut démontrer que le problème de Cauchy pour (0.0.62) est bien posé (cf. [1, 42]). Au contraire, pour $\mu > \mu^*(N)$, l'équation (0.0.62) n'admet pas de solution si les données u_0 et f sont positives, même localement en temps [1, 7].

Dans un premier temps, nous étudions le cas $\mu \leq \mu^*(N)$. Dans ce cas, nous prouvons que le système (0.0.62) peut être contrôlé à zéro avec un contrôle $f \in L^2(0,T;L^2(\omega))$.

Théorème 13. *Soit μ un nombre réel tel que $\mu \leq \mu^*(N)$.*
Pour tout sous-ouvert $\omega \subset \Omega$ non-vide, pour tout $T > 0$ et $u_0 \in L^2(\Omega)$, il existe un contrôle $f \in L^2((0,T)\times\omega)$ tel que la solution u de (0.0.62) satisfait $u(T) = 0$. De plus, il existe une constante C_T telle que

$$\|f\|_{L^2((0,T)\times\omega)} \leq C_T \|u_0\|_{L^2(\Omega)}. \tag{0.0.64}$$

Le même résultat a déjà été prouvé dans [41] dans le cas où l'ouvert ω encercle la singularité, condition géométrique non triviale dont nous montrons ici qu'elle n'est pas nécessaire. Remarquons aussi que ce résultat est connu pour l'équation de la chaleur sans potentiel (i.e. $\mu = 0$ dans (0.0.62)) [20, 26], ou lorsque le potentiel est dans $L^{2N/3}(\Omega)$, cf. [23]. Ici, le potentiel $1/|x|^2$ que nous considérons n'est pas dans $L^{N/2}(\Omega)$, et ces résultats ne s'appliquent donc pas.

Pour démontrer le Théorème 13, nous prouvons des propriétés d'observabilité sur le système adjoint à l'aide d'inégalités de Carleman. Les inégalités de Carleman que nous démontrons sont inspirées des travaux précédents [41] et [20].

Pour être plus précis, nous montrons qu'il est possible de choisir une fonction poids σ qui coïncide au voisinage de la singularité avec celle introduite dans [41], tandis que nous la choisissons comme dans [20] loin de la singularité. Ce choix nous permet de contourner la condition géométrique nécessaire dans [41] : dans [41], la preuve est basée sur une décomposition des solutions en harmoniques sphériques, qui permet de se ramener ainsi à l'étude d'équations radiales unidimensionnelles.

Dans un second temps, nous considérons le cas $\mu > \mu^*(N)$. Rappelons que dans ce cas, le problème de Cauchy est mal posé, puisqu'il y a explosion complète instantanée des solutions de (0.0.62) pour $u_0 > 0$ et $f = 0$, cf. [1]. Cependant, cela ne répond pas à la question suivante : étant donné $u_0 \in L^2(\Omega)$, peut-on trouver une fonction $f \in L^2((0,T);H^{-1}(\Omega))$ à support dans $\bar{\omega}$ telle qu'il existe une solution $u \in L^2(0,T;H_0^1(\Omega))$?

Nous allons répondre à cette question par la négative. Pour cela, nous considérons, pour $\varepsilon > 0$, les systèmes approchés

$$\begin{cases} \partial_t u - \Delta_x u - \dfrac{\mu}{|x|^2+\varepsilon^2} u = f, & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) = u_0(x), & x \in \Omega. \end{cases} \tag{0.0.65}$$

Pour $\varepsilon > 0$, le problème de Cauchy dans (0.0.65) est bien posé. Nous nous proposons alors d'étudier les fonctionnelles

$$J^\varepsilon_{u_0}(f) = \frac{1}{2} \iint\limits_{\Omega\times(0,T)} |u(x,t)|^2 \, \mathrm{dx} \, \mathrm{dt} + \frac{1}{2} \int\limits_0^T \|f(t)\|^2_{H^{-1}(\Omega)} \, \mathrm{dt}, \tag{0.0.66}$$

définies pour $f \in L^2((0,T);H^{-1}(\Omega))$ à support dans $\bar{\omega}$, et où u est la solution correspondante de (0.0.65).

Nous démontrons alors le résultat suivant :

Théorème 14. *Soit $\mu > \mu^*(N)$. Supposons que $0 \notin \bar{\omega}$.*
Alors il n'existe pas de constante C telle que pour tout $\varepsilon > 0$ et pour tout $u_0 \in L^2(\Omega)$,

$$\inf_{\substack{f \in L^2((0,T);H^{-1}(\Omega)) \\ f \text{ à support dans } \bar{\omega}}} J^\varepsilon_{u_0}(f) \leq C \|u_0\|^2_{L^2(\Omega)}. \tag{0.0.67}$$

La preuve de ce théorème est basée sur une étude spectrale des opérateurs

$$L^\varepsilon = -\Delta_x - \frac{\mu}{|x|^2+\varepsilon^2}$$

sur Ω avec conditions aux limites de Dirichlet. En particulier, nous étudions la première valeur propre λ^ε_0, dont nous montrons qu'elle tend vers $-\infty$. Nous étudions alors le vecteur propre correspondant ϕ^ε_0, dont nous montrons qu'il est de plus en plus localisé au voisinage de 0 quand $\varepsilon \to 0$. Nous en déduisons alors que

$$\inf_{\substack{f \in L^2((0,T);H^{-1}(\Omega)) \\ f \text{ à support dans } \bar{\omega}}} J^\varepsilon_{\phi^\varepsilon_0}(f) \underset{\varepsilon\to 0}{\longrightarrow} +\infty,$$

ce qui suffit à conclure la preuve du Théorème 14.

Notes : Chaque chapitre présenté ci-après correspond à un article effectué dans le cadre de ma thèse. En conséquence, chaque chapitre introduit ses propres notations et peut être lu indépendamment des autres. Il peut arriver que certaines notations aient des significations différentes dans différents chapitres.

Dans l'introduction, nous avons cherché à donner une vision globale de l'ensemble de la thèse. Il s'ensuit que certaines notations utilisées dans les différents chapitres qui suivent ont été modifiées.

Bibliographie

[1] P. Baras and J. A. Goldstein. The heat equation with a singular potential. *Trans. Amer. Math. Soc.*, 284(1) :121–139, 1984.

[2] J.-P. Bérenger. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.*, 114(2) :185–200, 1994.

[3] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. An exact bounded PML for the Helmholtz equation. *C. R. Math. Acad. Sci. Paris*, 339(11) :803–808, 2004.

[4] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. Numerical simulation of time-harmonic scattering problems with an optimal PML. *Sci. Ser. A Math. Sci. (N.S.)*, 13 :58–71, 2006.

[5] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. An optimal perfectly matched layer with unbounded absorbing function for time-harmonic acoustic scattering problems. *J. Comput. Phys.*, 223(2) :469–488, 2007.

[6] N. Burq and M. Zworski. Geometric control in the presence of a black box. *J. Amer. Math. Soc.*, 17(2) :443–471 (electronic), 2004.

[7] X. Cabré and Y. Martel. Existence versus explosion instantanée pour des équations de la chaleur linéaires avec potentiel singulier. *C. R. Acad. Sci. Paris Sér. I Math.*, 329(11) :973–978, 1999.

[8] C. Castro and S. Micu. Boundary controllability of a linear semi-discrete 1-d wave equation derived from a mixed finite element method. *Numer. Math.*, 102(3) :413–462, 2006.

[9] C. Castro, S. Micu, and A. Münch. Numerical approximation of the boundary control for the wave equation with mixed finite elements in a square. *IMA J. Numer. Anal.*, 28(1) :186–214, 2008.

[10] S. Cox and E. Zuazua. The rate at which energy decays in a damped string. *Comm. Partial Differential Equations*, 19(1-2) :213–243, 1994.

[11] R. Dautray and J.-L. Lions. *Analyse mathématique et calcul numérique pour les sciences et les techniques. Tome 1.* Collection du Commissariat à l'Énergie Atomique : Série Scientifique. Masson, Paris, 1984. With the collaboration of M. Artola, M. Authier, P. Bénilan, M. Cessenat, J.-M. Combes, A. Gervat, H. Lanchon, B. Mercier, C. Wild and C. Zuily.

[12] S. Ervedoza. Observability of the mixed finite element method for the 1d wave equation on non-uniform meshes. *To appear in ESAIM : COCV*, 2008. *Cf Chapitre 2.*

[13] S. Ervedoza. Admissibility and observability for Schrödinger systems : Applications to finite element approximation schemes. *To be published*, 2008. *Cf Chapitre 6.*

[14] S. Ervedoza. Admissibility and observability for Wave systems : Applications to finite element approximation schemes. *To be published*, 2008. *Cf Chapitre 7.*

[15] S. Ervedoza. Control and stabilization properties for a singular heat equation with an inverse-square potential. *To appear in Comm. in PDE*, 2008. *Cf Chapitre 8.*

[16] S. Ervedoza, C. Zheng, and E. Zuazua. On the observability of time-discrete conservative linear systems. *J. Funct. Anal.*, 254(12) :3037–3078, June 2008. *Cf Chapitre 3.*

[17] S. Ervedoza and E. Zuazua. Perfectly matched layers in 1-d : Energy decay for continuous and semi-discrete waves. *Numer. Math.*, 109(4) :597–634, 2008. *Cf Chapitre 1.*

[18] S. Ervedoza and E. Zuazua. Uniform exponential decay for viscous damped systems. *To appear in Proc. of Siena "Phase Space Analysis of PDEs 2007", Special issue in honor of Ferrucio Colombini*, 2008. *Cf Chapitre 4.*

[19] S. Ervedoza and E. Zuazua. Uniformly exponentially stable approximations for a class of damped systems. *To appear in J. Math. Pures Appl.*, 2008. *Cf Chapitre 5.*

[20] A. V. Fursikov and O. Y. Imanuvilov. *Controllability of evolution equations*, volume 34 of *Lecture Notes Series*. Seoul National University Research Institute of Mathematics Global Analysis Research Center, Seoul, 1996.

[21] R. Glowinski, C. H. Li, and J.-L. Lions. A numerical approach to the exact boundary controllability of the wave equation. I. Dirichlet controls : description of the numerical methods. *Japan J. Appl. Math.*, 7(1) :1–76, 1990.

[22] A. Haraux. Une remarque sur la stabilisation de certains systèmes du deuxième ordre en temps. *Portugal. Math.*, 46(3) :245–258, 1989.

[23] O. Y. Imanuvilov and M. Yamamoto. Carleman inequalities for parabolic equations in Sobolev spaces of negative order and exact controllability for semilinear parabolic equations. *Publ. Res. Inst. Math. Sci.*, 39(2) :227–274, 2003.

[24] A. E. Ingham. Some trigonometrical inequalities with applications to the theory of series. *Math. Z.*, 41(1) :367–379, 1936.

[25] V. Komornik. *Exact controllability and stabilization.* RAM : Research in Applied Mathematics. Masson, Paris, 1994. The multiplier method.

[26] G. Lebeau and L. Robbiano. Contrôle exact de l'équation de la chaleur. *Comm. Partial Differential Equations*, 20(1-2) :335–356, 1995.

[27] J.-L. Lions. *Contrôlabilité exacte, Stabilisation et Perturbations de Systèmes Distribués. Tome 1. Contrôlabilité exacte*, volume RMA 8. Masson, 1988.

[28] K. Liu. Locally distributed control and damping for the conservative systems. *SIAM J. Control Optim.*, 35(5) :1574–1590, 1997.

[29] L. Miller. Controllability cost of conservative systems : resolvent condition and transmutation. *J. Funct. Anal.*, 218(2) :425–444, 2005.

[30] A. Münch. A uniformly controllable and implicit scheme for the 1-D wave equation. *M2AN Math. Model. Numer. Anal.*, 39(2) :377–418, 2005.

[31] J.-P. Puel. Une approche non classique d'un problème d'assimilation de données. *C. R. Math. Acad. Sci. Paris*, 335(2) :161–166, 2002.

[32] J. V. Ralston. Solutions of the wave equation with localized energy. *Comm. Pure Appl. Math.*, 22 :807–823, 1969.

[33] K. Ramdani, T. Takahashi, G. Tenenbaum, and M. Tucsnak. A spectral approach for the exact observability of infinite-dimensional systems with skew-adjoint generator. *J. Funct. Anal.*, 226(1) :193–229, 2005.

[34] K. Ramdani, T. Takahashi, and M. Tucsnak. Uniformly exponentially stable approximations for a class of second order evolution equations—application to LQR problems. *ESAIM Control Optim. Calc. Var.*, 13(3) :503–527, 2007.

[35] P.-A. Raviart and J.-M. Thomas. *Introduction à l'analyse numérique des équations aux dérivées partielles*. Collection Mathématiques Appliquées pour la Maitrise. Masson, Paris, 1983.

[36] L. R. Tcheugoué Tebou and E. Zuazua. Uniform boundary stabilization of the finite difference space discretization of the $1-d$ wave equation. *Adv. Comput. Math.*, 26(1-3) :337–365, 2007.

[37] L.R. Tcheugoué Tébou and E. Zuazua. Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity. *Numer. Math.*, 95(3) :563–598, 2003.

[38] L. N. Trefethen. Group velocity in finite difference schemes. *SIAM Rev.*, 24(2) :113–136, 1982.

[39] S. V. Tsynkov. Numerical solution of problems on unbounded domains. A review. *Appl. Numer. Math.*, 27(4) :465–532, 1998. Absorbing boundary conditions.

[40] M. Tucsnak and G. Weiss. Observation and control for operator semigroups, 2008.

[41] J. Vancostenoble and E. Zuazua. Null controllability for the heat equation with singular inverse-square potentials. *J. Funct. Anal.*, 254(7) :1864–1902, 2008.

[42] J. L. Vazquez and E. Zuazua. The Hardy inequality and the asymptotic behaviour of the heat equation with an inverse-square potential. *J. Funct. Anal.*, 173(1) :103–153, 2000.

[43] X. Zhang, C. Zheng, and E. Zuazua. Exact controllability of the time discrete wave equation. *Discrete and Continuous Dynamical Systems*, 2007.

[44] E. Zuazua. Propagation, observation, and control of waves approximated by finite difference methods. *SIAM Rev.*, 47(2) :197–243 (electronic), 2005.

Part I

Examples

Chapter 1

Perfectly Matched Layers in 1-d : Energy decay for continuous and semi-discrete waves

Joint work with Enrique Zuazua.

Abstract: In this paper we investigate the efficiency of the method of Perfectly Matched Layers (PML) for the 1-d wave equation. The PML method furnishes a way to compute solutions of the wave equation for exterior problems in a finite computational domain by adding a damping term on the matched layer. In view of the properties of solutions in the whole free space, one expects the energy of solutions obtained by the PML method to tend to zero as $t \to \infty$, and the rate of decay can be understood as a measure of the efficiency of the method. We prove, indeed, that the exponential decay holds and characterize the exponential decay rate in terms of the parameters and damping potentials entering in the implementation of the PML method. We also consider a space semi-discrete numerical approximation scheme and we prove that, due to the high frequency spurious numerical solutions, the decay rate fails to be uniform as the mesh size parameter h tends to zero. We show however that adding a numerical viscosity term allows us to recover the property of exponential decay of the energy uniformly on h. Although our analysis is restricted to finite differences in 1-d, most of the methods and results apply to finite elements on regular meshes and to multi-dimensional problems.

1.1 Introduction

When numerically solving wave propagation problems in unbounded domains, because of the finite computational possibilities, one has to truncate the computational domain. This makes it necessary to choose boundary conditions at the newly formed exterior boundary. These boundary conditions are relevant, for example in problems arising in acoustics and electrodynamics, since they may have a significant impact on the whole solution due to reflections.
In order to avoid those spurious reflections a natural method, introduced by Engquist and Majda in [21], is based on the use of the so-called transparent boundary conditions. The transparent boundary conditions are often of non-local nature, depend on the geometry of the domain, etc. However, in spite of the simple implementation of lowest order absorbing boundary conditions, good accuracy is

only achieved for higher order ones [6]. For the state of the art, we refer to the survey article [35].
An alternate approach, proposed by Bérenger in [10] in 1994, is the so-called method of the Perfectly Matched Layers (PML). The idea consists in surrounding the computational domain by a layer and extending the equation to it adding damping terms designed to dissipate the energy entering in it, such that no spurious reflection waves are created. This method, first introduced to deal with Maxwell's equations, has been successfully adapted to many other wave models, see the survey article [31].

This article aims to develop a complete rigorous analysis in 1-d for the PML model associated to the scalar wave equation. Our work is inspired by the existing literature on the control and stabilization of waves.
More precisely, the object of this paper is twofold. First, we analyze the continuous 1-d wave equation to accurately describe the efficiency of the PML method in terms of the various parameters entering in it and second, we consider semi-discrete numerical approximation schemes. The study of this system has first been developed by a plane wave analysis (see for instance [13]), where explicit formulas were given for the amplitudes of the reflected and transmitted waves around the interface. Latter, Fourier and energy techniques were also used in [1, 17, 26, 36] for analyzing the PML method for the Helmholtz equation. Very few papers [8, 9, 4] deal with the stability of the time-dependent PML system.
To be more precise, we consider the wave equation in an unbounded domain of the form $(0,\infty)$ with homogeneous Neumann boundary conditions at $x = 0$ and initial data in $L^2(0,\infty)$ with compact support:

$$\begin{cases} \partial_{tt}^2 u - \partial_{xx}^2 u = 0, \ x > 0, \ t > 0, \\ \partial_x u(0,t) = 0. \end{cases} \tag{1.1.1}$$

In the hyperbolic form, considering the physical variables $P = -\partial_x u$ and $V = \partial_t u$, the system under consideration can be written as follows

$$\begin{cases} \partial_t P + \partial_x V = 0 \quad \text{in } (0,\infty)\times(0,T), \\ \partial_t V + \partial_x P = 0 \quad \text{in } (0,\infty)\times(0,T), \\ P(0,t) = 0, \\ P(x,0) = P_0(x), \quad V(x,0) = V_0(x). \end{cases} \tag{1.1.2}$$

Its solution can be computed explicitly by the method of characteristics (which gives D'Alembert's formula). Since we assume the initial data (P_0, V_0) to be compactly supported, for instance in $(0,a)$ for some $a > 0$, it follows that the solutions (P,V) will vanish in $(0,a)$ for $t \geq 2a$, which is the time needed for waves to go from $x = a$ to $x = 0$ and back to $x = a$ after reflection. The fact that P and V reach the zero state in time $t = 2a$ in $(0,a)$ can be seen on u, that stabilizes to the constant $\int_0^a V_0(x)\,dx$ for $t \geq 2a$ on the interval $(0,a)$.

The goal of the PML method, when applied to this 1-d model, is to reproduce this very property of P, V but by solving a problem in a bounded domain. For convenience, we translate the domain $(0,\infty)$ where waves propagate to $(-1,\infty)$ and focus on the restriction of solutions on the compact domain $(-1,0)$. This can be done, by scaling, without loss of generality. Then, solutions (P,V) with initial data compactly supported in $(-1,0)$ vanish on $(-1,0)$ for $t \geq 2$ and we expect that the approximate solutions, obtained by the PML method in a bounded domain, will reproduce this property. A way of measuring how small is the restriction of the approximate solutions to $(-1,0)$ is analyzing the time decay properties of its energy as $t \to \infty$.
The PML method is designed to give an accurate approximation of the solutions of (1.1.2) in $(-1,0)$, by solving the following system on the domain $(-1,1)$, in which the space-layer $(0,1)$ has been added:

$$\begin{cases} \partial_t P + \partial_x V + \chi_{(0,1)}\sigma P = 0 \quad \text{in } (-1,1)\times(0,T), \\ \partial_t V + \partial_x P + \chi_{(0,1)}\sigma V = 0 \quad \text{in } (-1,1)\times(0,T), \\ P(-1,t) = P(1,t) = 0, \\ P(x,0) = P_0(x), \quad V(x,0) = V_0(x), \end{cases} \tag{1.1.3}$$

where $(P_0, V_0) \in L^2(-1,0)^2$ have been extended by 0 in $(0,1)$.
Here σ is a positive function defined on $(0,1)$, which is assumed to be in $L^1(0,1)$. Note that within the added layer $(0,1)$ the equations in (1.1.3) have been modified by adding the terms involving the dissipative potential σ. Throughout the paper the function σ is extended on $(-1,1)$ by zero in $(-1,0)$. Actually, one can recover most of the results presented here in the case where the added space-layer is $(0,r)$ by a scaling argument, which maps $(-1,r)$ to $(-1,1)$ and by considering functions σ in (1.1.3) vanishing in $(-1, 2/(1+r))$.
We analyze (1.1.3) for all initial data though, as we have said, the relevant ones in the context of the PML method are those with compact support in $(-1,0)$. Recall that the true solution (P,V) of (1.1.2) vanishes in $(-1,0)$ for $t > 2$ when the initial data have support in $(-1,0)$. So we expect the energy of the PML solutions localized in $(-1,0)$ to be small when $t > 2$. Then the exponential decay rate of the restriction of solutions of (1.1.3) to $(-1,0)$ is a way of measuring the efficiency of the PML method and the chosen damping potential σ. Actually, as we shall see, it coincides with the decay rate of the total energy of solutions. Thus, most of the paper will be devoted to analyze the latter.
System (1.1.3) is well-posed, and the total energy of solutions

$$E(t) = E(P(t), V(t)) = \frac{1}{2}\int_{-1}^{1} \left(|P(t,x)|^2 + |V(t,x)|^2\right)\, dx \tag{1.1.4}$$

is dissipated according to the following law

$$\frac{dE}{dt}(t) = -\int_0^1 \sigma(x)\left(|P(t,x)|^2 + |V(t,x)|^2\right)\, dx. \tag{1.1.5}$$

This last equation shows the well-posedness of the 1-d PML equations in the space

$$(P,V) \in C([0,\infty); L^2(-1,1)^2).$$

As far as we know, the problem of the exponential decay of the energy for the PML method has not been addressed in detail so far. In [8, 9] it was stated that a first order energy of solutions for Maxwell's PML model with a constant σ decays, but no decay rate was given.
In our analysis we will follow the techniques of [19], which, actually, in the present setting, can be applied more simply. Note that system (1.1.3) and its dissipative properties are similar to those of the classical damped wave equation:

$$\begin{cases} \partial_{tt}^2 w - \partial_{xx}^2 w + 2a(x)\partial_t w = 0 \quad \text{in } (-1,1)\times(0,T), \\ w(-1,t) = w(1,t) = 0. \end{cases} \tag{1.1.6}$$

In this case, the energy dissipation law reads :

$$\frac{d}{dt}\Big(\frac{1}{2}\int_{-1}^{1} \left(|\partial_t w|^2 + |\partial_x w|^2\right)\, dx\Big) = -2\int_{-1}^{1} a(x)|\partial_t w|^2\, dx. \tag{1.1.7}$$

For system (1.1.6), it is well-known that the energy decays exponentially as $t \to \infty$ provided $a \geq 0$ is strictly positive on some subinterval. Moreover, in [19] the exponential decay rate was characterized

as the spectral abscissa, for $a \in BV(-1,1)$.
Actually, in the special case where σ is constant, the PML equations (1.1.3) in $(0,1)$ read as follows:

$$\partial_{tt}^2 u - \partial_{xx}^2 u + 2\sigma\partial_t u + \sigma^2 u = 0 \quad \text{in } (0,1)\times(0,T), \tag{1.1.8}$$

which is a dispersive variant of system (1.1.6), since (1.1.8) contains the extra term $\sigma^2 u$. As we shall see, the presence of this added dispersive term simplifies the spectral analysis of the system.

We define the exponential decay rate of solutions of (1.1.3) as a function of σ, by

$$\omega(\sigma) = \sup\{\omega :\ \exists C,\ \forall (P_0,V_0)\in (L^2(-1,1))^2,\ \forall t,\ E(t)\le CE(P_0,V_0)\exp(-\omega t)\}. \tag{1.1.9}$$

For each $\omega \le \omega(\sigma)$, we define $C(\omega)$ as the best constant such that

$$\forall (P_0,V_0)\in (L^2(-1,1))^2,\ \forall t,\ E(t)\le C(\omega)E(P_0,V_0)\exp(-\omega t). \tag{1.1.10}$$

Note that this actually measures the decay rate of the energy of solutions of (1.1.3) in the whole domain, not only in $(-1,0)$. However, we will prove that the decay rates of the energy of the restriction of solutions of (1.1.3) to $(-1,0)$ and in the whole domain coincide.
Let us also define the space operator L by

$$\begin{array}{l} L(P,V) = (\partial_x V + \chi_{(0,1)}\sigma P, \partial_x P + \chi_{(0,1)}\sigma V), \\ \mathcal{D}(L) = H_0^1(-1,1)\times H^1(-1,1). \end{array} \tag{1.1.11}$$

This unbounded operator on $L^2(-1,1)$ is the generator of a semi-group of contractions solving the equations (1.1.3). We prove that the decay rate $\omega(\sigma)$ satisfies $\omega(\sigma) = 2S(\sigma)$, where $S(\sigma)$ is the spectral abscissa, defined in terms of $\Lambda(L)$, the spectrum of the operator L, as follows:

$$S(\sigma) = \sup\{Re(\lambda)\ |\lambda\in\Lambda(L)\}. \tag{1.1.12}$$

This is done by means of a complete description of the spectrum of L, that also shows that $\omega(\sigma)$ coincides with

$$I = \int_0^1 \sigma(x)\ dx, \tag{1.1.13}$$

which is a measure of the total damping entering in the system.
This result confirms the ones in [11, 13, 14] about the efficiency of taking a singular damping $\sigma \notin L^1$ for the PML method for the Helmholtz equation. Our characterization (1.1.13) of the decay rate as the integral of σ confirms that, when taking σ singular, the decay rate may be made arbitrarily large.

In the second part of this article, we investigate the decay of the energy for the following semi-discrete finite-difference approximation scheme for PML:

$$\begin{cases} \partial_t P_j + \dfrac{V_{j+1/2}-V_{j-1/2}}{h} + \sigma_j P_j = 0, \quad j\in\{-N+1,\dots,N-1\}, \\ \partial_t V_{j+1/2} + \dfrac{P_{j+1}-P_j}{h} + \sigma_{j+1/2}V_{j+1/2} = 0, \quad j\in\{-N,\dots,N-1\}, \\ P_{-N} = P_N = 0. \end{cases} \tag{1.1.14}$$

The notations we employ are the classical ones for finite differences: $h = 1/N$, for some $N\in\mathbb{N}$, is the mesh size, $x_j = jh$, $j = -N,\cdots,N$ constitute the mesh points and P_j and $V_{j+1/2}$ are approximations of P on x_j and of V on $(x_j+x_{j+1})/2$. We approximate the function σ by a piecewise constant function

taking the value $\sigma_{j+1/2}$ on each (x_j, x_{j+1}) and denote by σ_j the mean value of $\sigma_{j-1/2}$ and $\sigma_{j+1/2}$.
The energy $E_h(t)$ of the semi-discrete system (1.1.14) is given by

$$E_h(t) = \frac{h}{2} \sum_{j=-N}^{N-1} \big(|P_j(t)|^2 + |V_{j+1/2}(t)|^2 \big), \tag{1.1.15}$$

and can be interpreted as a discretization of the continuous energy E in (1.1.4). It decays exponentially as $t \to \infty$. But, as we shall see, the decay rate is not uniform on h. This is due to the spurious high frequency numerical oscillations whose group velocity is close to zero. The effect of these spurious oscillations has already been noticed in a number of articles in connection with the qualitative properties of numerical waves since [34] and further developed in the survey article [39]. We give a precise analysis of the spectrum in terms of h and σ, when σ is a constant on $(0, 1)$, that will further clarify this lack of uniform (on h) exponential decay.

Inspired by [33], in order to remedy this lack of uniform decay, we consider the following viscous scheme, which is again convergent of order 2:

$$\begin{cases} \partial_t P_j + \dfrac{V_{j+1/2} - V_{j-1/2}}{h} + \sigma_j P_j - h^2 (\Delta_h P)_j = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j \in \{N+1, \dots, N-1\}, \\ \partial_t V_{j+1/2} + \dfrac{P_{j+1} - P_j}{h} + \sigma_{j+1/2} V_{j+1/2} - h^2 (\Delta_h V)_{j+1/2} = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j \in \{-N, \dots, N-1\}, \\ P_{-N} = P_N = 0, \quad V_{-N-1/2} = V_{-N+1/2}, \quad V_{N-1/2} = V_{N+1/2}. \end{cases} \tag{1.1.16}$$

Here and in the sequel Δ_h denotes the classical discretization of the Laplace operator:

$$(\Delta_h A)_j = \frac{1}{h^2}(A_{j+1} + A_{j-1} - 2A_j).$$

The energy of this modified system is further dissipated by the added numerical viscosity terms:

$$\frac{dE_h}{dt}(t) = -h \sum_{j=-N+1}^{N} \sigma_j |P_j|^2 - h \sum_{j=-N}^{N-1} \sigma_{j+1/2} |V_{j+1/2}|^2 \\ - h^3 \sum_{j=-N}^{N-1} \Big(\Big(\frac{P_{j+1} - P_j}{h} \Big)^2 + \Big(\frac{V_{j+1/2} - V_{j-1/2}}{h} \Big)^2 \Big). \tag{1.1.17}$$

In particular, the viscosity terms provide an efficient dissipation on the high frequency waves and, accordingly, as we shall see in Theorem 1.5.1, the decay rate is uniform on h.
Furthermore, we prove in Theorem 1.5.3 that the decay rate of the energy of the semi-discrete approximation schemes (1.1.16) coincides with the continuous one, that is I, under an appropriate choice of the viscosity parameter. In other words, we can recover the dynamical properties of the continuous PML at the semi-discrete level.
This numerical technique of adding numerical viscosity provides a way to keep the PML method accurate at the semi-discrete level. Inspired on previous work on the control of waves ([39]), we may expect that other remedies will also allow preserving the uniform (on h) decay properties of the energy, for instance a mixed-finite element method as in [5, 16] or a multi-grid scheme as in [22, 24].
Actually, most of the results presented here at the semi-discrete level have a very wide range of validity, and can be extended to different approximation schemes, for instance using finite elements, and even in higher dimension. In particular, the construction in subsection 1.4.1 works and proves that in

general the discrete energy cannot be uniformly exponentially decaying, if a numerical viscosity is not added everywhere in the domain, including the part where the PML is not effective.

Here is a brief overview on the PML method and its possible applications. The mathematical analysis of the continuous model was done in [26, 17, 36], where it was proved that the solution of the continuous PML for the Helmholtz equation with an infinite layer corresponds exactly to the unbounded solution in the computational domain. Moreover, it was also stated that, if the layer is bounded but large enough, solutions provide a good approximation in the computational domain. Moreover, it was proved in [14, 11, 13] that when the layer is bounded, the PML method for the Helmholtz equation recovers the exact solution in the computational domain if we choose a radial damping potential $\sigma \notin L^1$. Unfortunately, it was proved in [1] that the PML method is only weakly well-posed for Maxwell's equations in the sense that the functions involved in the splitting induced by the PML method do not stay in the same functional space as the initial data, thus requiring smoother initial data. This also implies that instabilities may arise under small perturbations. A number of articles has been devoted to gain a better comprehension of these problems on well-posedness and instabilities in the continuous case ([8, 7, 28, 37]). New absorbing layers were also proposed in the continuous case for Maxwell's equations and advective acoustics, in particular, in [2, 3, 31, 9, 4] for which well-posedness and stability have been successfully proved. Note however that this phenomenon does not appear in 1-d, as follows from (1.1.5). On the semi-discrete level, very few results are available. We refer however to [32] for a study of the accuracy of the discretized Helmholtz-PML equations and to [15] for an analysis of the convergence of the finite element PML approximations towards the continuous PML system in the case of the time-harmonic electromagnetic scattering problem.

The structure of the present paper is the following. In section 1.2, we carefully analyze the spectral properties of the space operator L, by using a shooting method. This will allow us to give an explicit formula for its spectrum in Theorem 1.2.1. In section 1.3, we prove that the quantities I, S, and $\omega(\sigma)$ above coincide. We will also prove that the inequality (1.1.10) holds for $\omega = \omega(\sigma)$ and give some estimates on the best constant $C(\omega(\sigma))$ in this inequality. We also give an explicit representation formula for the solutions of the continuous PML equations and deduce the optimality of our estimates. In section 1.4, we address the same issues for the space semi-discrete system. We show that the high frequency spurious numerical solutions are responsible for a lack of uniform exponential decay of the energy and, in the special case where σ is constant, we give an asymptotic description of the spectrum of the discretized operator. Finally, in section 1.5, we consider the viscous scheme (1.1.16) and prove the exponential decay of the energy, uniformly in h.

1.2 Analysis of the space operator L

The aim of this section is to give a complete description of the spectral properties of L defined as in (1.1.11).

Theorem 1.2.1. *Let $\sigma \in L^1(0,1)$ be a non-trivial and non-negative function. Then:*

1. *The operator L has a compact inverse.*
2. *The spectrum of the operator L coincides with the set of the eigenvalues*

$$\lambda_k = \frac{1}{2}\Big(\int_0^1 \sigma(x)\ dx + ik\pi \Big),\ k \in \mathbb{Z}. \tag{1.2.1}$$

3. The eigenvectors (P_k, V_k) form a Riesz basis of $L^2(-1,1)^2$.

Let us first remark that the first statement implies that the spectrum is discrete. The interest of the second statement is that it provides an explicit description of the eigenvalues. The last claim allows characterizing the decay rate in terms of the spectral abscissa. The following subsections will be devoted to the proof of each of these three statements.

1.2.1 Inverse of the operator L

Consider the system

$$(P,V) \in \mathcal{D}(L) \quad ; \quad L(P,V) = (f,g).$$

where f and g are two given functions in $L^2(-1,1)$.

To solve this problem, we consider $Q = P + V$ and $R = V - P$ that satisfy

$$\partial_x Q + \sigma(x)Q(x) = f(x) + g(x), \quad \partial_x R - \sigma(x)R(x) = f(x) - g(x). \tag{1.2.2}$$

Introducing the boundary conditions $P = 0$ at $x = \pm 1$, this yields

$$Q = R, \quad x = \pm 1. \tag{1.2.3}$$

Then straightforward computations show that equations (1.2.2)-(1.2.3) have a unique solution if and only if $I \neq 0$, which is true since σ is a non-trivial non-negative function.
By (1.2.2) and (1.2.3) we deduce that L^{-1} defines a bounded operator

$$L^{-1} : L^2(-1,1)^2 \to H_0^1(-1,1) \times H^1(-1,1),$$

which turns out to be compact as an operator from $L^2(-1,1)^2$ into itself.

1.2.2 Analysis of the spectrum : Eigenvalues of L

The system characterizing the spectrum is as follows:

$$\begin{cases} \partial_x V + \sigma P = \lambda P, \ \partial_x P + \sigma V = \lambda V, \ x \in (-1,1), \\ P(-1) = P(1) = 0. \end{cases}$$

Using the functions Q and R as in the previous section gives

$$Q(x) = Q(-1)e^{-\int_{-1}^{x}(\sigma(z)-\lambda)\ dz}, \quad R(x) = R(-1)e^{\int_{-1}^{x}(\sigma(z)-\lambda)\ dz}.$$

The boundary conditions yield (1.2.3). Then λ is an eigenvalue if and only if

$$\exp\Big(-\int_{-1}^{1}(\sigma(z)-\lambda)\ dz\Big) = \exp\Big(\int_{-1}^{1}(\sigma(z)-\lambda)\ dz\Big). \tag{1.2.4}$$

Hence the result (1.2.1).

Remark 1.2.2. Note that the eigenvalues are totally explicit for all damping potentials σ. This is not the case for the damped wave equation (1.1.6), which, when written as a first order system, corresponds to adding the damping potential only in one of the equations of the system. In that case, (1.2.1) only holds asymptotically for high frequencies and this under the assumption that $\sigma \in BV(-1,1)$ (see [19]).

1.2.3 Analysis of the spectrum : Eigenvectors

Define the function θ by

$$\theta(x) = \int_{-1}^{x} \Big(\sigma(z) - \frac{I}{2}\Big)\, dz. \tag{1.2.5}$$

This function can be seen as a measure of the difference between the dissipative term σ and the average dissipation $I/2$. Note also that $\theta(-1) = \theta(1) = 0$.
We remark that for all eigenvectors P_k, V_k, the functions Q_k, R_k as in the previous section satisfy (taking $Q(-1) = R(-1) = 1$) :

$$Q_k(x)\exp(\theta(x)) = e^{\frac{ik\pi}{2}(x+1)}, \quad R_k(x)\exp(-\theta(x)) = e^{-\frac{ik\pi}{2}(x+1)}.$$

Our purpose now is to check that the family (P_k, V_k) constitutes a Riesz basis in $L^2(-1,1)^2$ (see [38] for an introduction to that subject). This means in particular that any pair of functions $(f,g) \in L^2(-1,1)^2$ can be written in an unique way as follows:

$$(f,g) = \sum a_k (P_k, V_k), \tag{1.2.6}$$

with

$$\sum |a_k|^2 \simeq \|(f,g)\|^2. \tag{1.2.7}$$

To prove this, we observe that (1.2.6) is equivalent to:

$$\begin{cases} (f+g)(x)e^{\theta(x)} = \sum a_k Q_k(x) e^{\theta(x)} = \sum a_k e^{\frac{ik\pi}{2}(x+1)} \\ (g-f)(x)e^{-\theta(x)} = \sum a_k R_k(x) e^{-\theta(x)} = \sum a_k e^{-\frac{ik\pi}{2}(x+1)}. \end{cases}$$

Then, the coefficients $\{a_k\}$ of the decomposition (1.2.6) of (f,g) on the basis $\{(P_k, V_k)\}$ can be identified as the Fourier coefficients of the function W defined in $(-3,1)$ by

$$W(x) = \begin{cases} (f+g)(x)\exp(\theta(x)), & -1 < x < 1, \\ (g-f)(-2-x)\exp(-\theta(-2-x)), & -3 < x < -1. \end{cases} \tag{1.2.8}$$

In other words (1.2.6) holds if and only if

$$W(x) = \sum_k a_k \exp\Big(\frac{ik\pi}{2}(x+1)\Big), \ x \in (-3,1). \tag{1.2.9}$$

Obviously W is in $L^2(-3,1)$ if and only if (f,g) is in $L^2(-1,1)^2$, and therefore (1.2.7) holds.
This construction defines an isomorphism $\mathcal{I}$, which maps the eigenvectors $\psi_k = (P_k, V_k)$ to the classical Fourier basis of $L^2(-3,1)$:

$$\mathcal{I}(f,g) = W, \tag{1.2.10}$$

where W is the function given in (1.2.8). Note that this implies that any function $\psi \in (L^2(-1,1))^2$ can be expanded as $\sum a_k \psi_k$, where the coefficients a_k satisfies:

$$\|\mathcal{I}\psi\|^2_{L^2(-3,1)} = 4\sum |a_k|^2.$$

Remark 1.2.3. In [19], it was proved (see Theorem 5.5) that the solution $y_2(x,\lambda)$ of the Cauchy-Lipschitz system

$$\begin{cases} -\partial^2_{xx} u + \lambda^2 u + 2a(x)\lambda u = 0, \ x \in (-1,1), \\ u(-1,\lambda) = 0, \quad \partial_x u(-1,\lambda) = 1, \end{cases}$$

which naturally arises when dealing with the spectral problem associated to a damped string, satisfies the following properties:

$$\begin{cases} y_2(x,\lambda_n) = 2\dfrac{\sinh(\xi(x)+in\pi(x+1)/2)}{in\pi - \int_{-1}^{1} a(x)\ dx} + O(1/n^2), \\ \\ \partial_x y_2(x,\lambda_n) = \cosh(\xi(x)+in\pi(x+1)/2) + O(1/|n|), \end{cases}$$

where λ_n is the n-th root of $\lambda \mapsto y_2(1,\lambda)$ and ξ is

$$\xi(x) = \int_{-1}^{x} a(s)\ ds - (x+1)\frac{1}{2}\int_{-1}^{1} a(x')\ dx'.$$

As indicated in the introduction, the dissipative potential $\sigma(x)$ of the PML method plays the same role as $a(x)$ in the dissipative wave equation (1.1.6). Obviously, the function $\xi(x)$ plays the same role as $\theta(x)$ in (1.2.5). We conclude that the eigenvectors of the damped wave equation are asymptotically close to the ones of the PML system.

1.3 On the decay of the energy

1.3.1 On the decay rate

Theorem 1.3.1. *The energy of the continuous PML system* (1.1.3) *is exponentially decaying. More precisely,*

$$\exists C > 0,\ s.t.\ \forall t > 0,\ E(t) \leq C\, E_0 \exp(-\omega(\sigma)t), \tag{1.3.1}$$

for all solution of (1.1.3) *with* $\omega(\sigma)$ *as in* (1.1.9)*. Moreover,* $\omega(\sigma) = I = 2S(\sigma)$*, with* I *and* $S(\sigma)$ *as in* (1.1.13) *and* (1.1.12)*, and the best constant* $C(\omega(\sigma))$ *in* (1.3.1) *as defined in* (1.1.10) *satisfies:*

$$C(\omega(\sigma)) \leq \exp(4\,\|\theta\|_\infty), \tag{1.3.2}$$

where $\theta = \theta(x)$ *is as in* (1.2.5)*.*

Proof. Equality $I = 2S(\sigma)$ was actually proved in the last section. From the previous section, we also know that the family of eigenvectors $\psi_k = (P_k, V_k)$ constitutes a Riesz basis of $L^2(-1,1)^2$ and this is sufficient to characterize the exponential decay rate as the spectral abscissa, i.e. $\omega(\sigma) = 2S(\sigma)$.
We now give further estimates on the decay rate in order to obtain (1.3.2), using the explicit isomorphism $\mathcal{I}$ given in (1.2.8).
Given $U_0 = (P_0, V_0) \in L^2(-1,1)^2$, we expand U_0 in the basis ψ_k : $U_0 = \sum a_k \psi_k$. We have :

$$2E_0 = \|U_0\|^2_{L^2(-1,1)^2} \geq \|\mathcal{I}\|^{-2} \|\mathcal{I}U_0\|^2_{L^2(-3,1)} \geq 4\,\|\mathcal{I}\|^{-2} \sum |a_k|^2.$$

It is easy to check that

$$U(t) = \sum a_k \exp(-\lambda_k t)\psi_k,$$

and then

$$\|\mathcal{I}U(t)\|^2_{L^2(-3,1)^2} = \exp(-tI) \sum |a_k|^2.$$

But

$$2E(t) = \|U(t)\|^2_{L^2(-1,1)^2} \leq \left\|\mathcal{I}^{-1}\right\|^2 \|\mathcal{I}U(t)\|^2_{L^2(-3,1)^2}\,.$$

Combining these inequalities, we get

$$E(t) \leq \|\mathcal{I}\|^2 \left\|\mathcal{I}^{-1}\right\|^2 \exp(-tI)E_0. \tag{1.3.3}$$

On the other hand, obviously, the exponential decay rate I is optimal as one can see by analyzing the solutions in separated variables.

According to (1.3.3) we have $C(\omega(\sigma)) \leq \kappa(\mathcal{I})^2$, where $\kappa(\mathcal{I})$ is the conditioning number $\kappa(\mathcal{I}) = \|\mathcal{I}\| \cdot \left\|\mathcal{I}^{-1}\right\|$, but we would like to derive a more explicit expression in terms of the damping potential σ. By Parseval's identity applied to (1.2.9), for f and g in $L^2(-1,1)$ we get:

$$\|\mathcal{I}(f,g)\|^2_{L^2(-3,1)} = 4\sum |a_k|^2 = \int_{-1}^{1} |f(x)+g(x)|^2 \exp(2\theta(x))\ dx + \int_{-1}^{1} |f(x)-g(x)|^2 \exp(-2\theta(x))\ dx. \tag{1.3.4}$$

As a consequence,

$$2\exp(-2\|\theta\|_\infty)\,\|(f,g)\|^2_{L^2(-1,1)^2} = 2\exp(-2\|\theta\|_\infty)\int_{-1}^{1}\left(|f(x)|^2+|g(x)|^2\right)dx \leq \|\mathcal{I}(f,g)\|^2_{L^2(-3,1)} \leq 2\exp(2\|\theta\|_\infty)\,\|(f,g)\|^2_{L^2(-1,1)^2}.$$

Accordingly,

$$\|\mathcal{I}\|^2 \leq 2\exp(2\|\theta\|_\infty), \quad \left\|\mathcal{I}^{-1}\right\|^2 \leq \frac{1}{2}\exp(2\|\theta\|_\infty),$$

and (1.3.2) holds. □

In order to discuss the efficiency of the PML method and, more precisely, that of system (1.1.3), we recall that it has been designed to provide an approximation of the solution of (1.1.2) in $(-1,0)$ for initial data with support in $(-1,0)$. Accordingly, we define E_l and E_r as the energy on the left and right subdomains respectively:

$$\begin{aligned} E_l(P,V) &= \frac{1}{2}\int_{-1}^{0}\left(|P(x)|^2+|V(x)|^2\right)dx, \\ E_r(P,V) &= \frac{1}{2}\int_{0}^{1}\left(|P(x)|^2+|V(x)|^2\right)dx. \end{aligned} \tag{1.3.5}$$

Theorem 1.3.2. *Let P_0 and V_0 be the initial data for the PML equations* (1.1.3) *with support in* $(-1,0)$. *Then,*

$$\begin{aligned} &E_l(P(t),V(t)) \leq E_0\ \exp(I(2-t)), \\ &E_r(P(t),V(t)) \leq E_0\ \exp(I+2\|\theta\|_\infty - It). \end{aligned} \tag{1.3.6}$$

Proof. The result follows from careful upper bounds in the previous proof, using (1.3.4), the conditions on the support of initial data, and the fact that the $L^\infty(-1,0)$ norm of θ is precisely $I/2$. This leads us to

$$E_0\exp(I) \geq \sum |a_k|^2 \geq E_l(P(t),V(t))\exp((t-1)I).$$

This establishes the first inequality. The second one is left to the reader. □

1.3.2 Comments

As a consequence of (1.3.6), if we fix a shape σ for the damping potential, and if we define the sequence of amplified potentials $\sigma_n(x) = n\sigma(x)$, then the corresponding solutions (P_n, V_n) to the PML system with initial data (P_0, V_0) supported in $(-1, 0)$ damped by σ_n tend to zero in $L^2((-1, 0))^2$ for $t > 2$ as $n \to \infty$.

Theorem 1.3.1 also confirms the results in [11, 12, 13, 14], where it was proved by a plane wave analysis that the reflection coefficient on $x = 0$ is of order $\exp(-I)$ and that, taking a function $\sigma \notin L^1(0, 1)$, makes the PML method very efficient. In [11, 13] numerical computations were done for different choices of σ : $\sigma_1(x) = (1 - x)^{-1} - 1$, $\sigma_2(x) = (1 - x)^{-2} - 1$ and $\sigma_3(x) = (1 - x)^2$. Numerical evidences in [11] show that the Helmholtz PML system is clearly more accurate for σ_1 and σ_2 than for σ_3. A precise proof was also given in [14] through the analysis of the Dirichlet-to-Neumann operator associated to the PML. Unfortunately, this kind of proof does not seem to hold anymore at the discrete level. Our result (1.3.1) on the decay rate of the energy also justifies these numerical evidences, since σ_1 and σ_2 do not belong to L^1 and have infinite average. As we shall see in the sequel, the methods we present here are more robust and will allow us to study the semi-discrete equations as well.

Let us now analyze the function θ entering in (1.3.2), which is obviously continuous on $(-1, 1)$. It is easy to see that the L^∞ norm of θ is exactly $I/2$ on $(-1, 0)$. On $(0, 1)$, the situation is more complex: θ is differentiable on $(0, 1)$, its derivative is $\theta'(x) = \sigma(x) - I/2$, and $\theta(0) = -I/2$, and $\theta(1) = 0$. We can also remark that $\|\theta\|_\infty = -\inf \theta \leq I$.
A natural question is trying to minimize the quantity $\|\theta\|_\infty$ on the positive potentials σ which have a given integral I_0. Easy considerations indicate that there are many different σ which satisfy $\|\theta\|_\infty = I_0/2$, the most natural one being the choice $\sigma = I_0$. However, in view of (1.3.6), this discussion is irrelevant if we are only considering the energy E_l concentrated in $(-1, 0)$.

1.3.3 Optimality of the decay rate

We complete this section with some results on the optimality of the decay rates we observed.

Theorem 1.3.3. *The estimates given in* (1.3.2) *and in Theorem 1.3.2 are sharp.*

Proof. We rewrite the system (1.1.3) in the following way :

$$\begin{cases} \partial_t(P+V) + \partial_x(P+V) + \sigma(P+V) = 0 & \text{in } (-1,1)\times(0,T), \\ \partial_t(P-V) - \partial_x(P-V) + \sigma(P-V) = 0 & \text{in } (-1,1)\times(0,T), \\ P(-1,t) = P(1,t) = 0. \end{cases}$$

Using characteristics leads to :

$$(P-V)(x,t) = (P_0 - V_0)(x+t)\, \exp\Big(-\int_x^{x+t} \sigma(y)\, dy\Big), \quad x \leq 1-t,$$

$$(P-V)(x,t) = (P-V)(1, x+t-1)\, \exp\Big(-\int_x^{1} \sigma(y)\, dy\Big), \quad x > 1-t,$$

$$(P+V)(x,t) = (P+V)(-1, t-x-1)\, \exp\Big(-\int_{-1}^{x} \sigma(y)\, dy\Big), \quad x < t-1,$$

$$(P+V)(x,t) = (P_0 + V_0)(x-t)\, \exp\Big(-\int_{x-t}^{x} \sigma(y)\, dy\Big), \quad x \geq t-1.$$

Using boundary conditions, we easily deduce that :

$$\forall n \in \mathbb{N},\, \forall x \in (-1,1), \quad (P(x,4n),V(x,4n)) = (P_0(x),V_0(x))\, \exp(-2nI),$$

which directly provides the good value for the decay rate, namely I.
To compute the optimal constant in (1.3.2), we need to be more precise:

$$\begin{aligned} E(t) &= \frac{1}{4}\int_{-1}^{1} \big(|(P+V)(x,t)|^2 + |(P-V)(x,t)|^2\big)\, dx \\ &\leq \frac{1}{4}\exp\Big(-2\inf_{\gamma\in\mathcal{R}_t}\int_\gamma \sigma(y)\, dy\Big)\int_{-1}^{1}\big(|(P_0+V_0)(x)|^2 + |(P_0-V_0)(x)|^2\big)\, dx \\ &= \exp\Big(-2\inf_{\gamma\in\mathcal{R}_t}\int_\gamma \sigma(y)\, dy\Big)E_0, \end{aligned}$$

where $\mathcal{R}_t$ is the set of characteristic rays of length t, that is the set of all continuous broken lines with slopes ± 1 in $(\tilde{t},x) \in [0,t]\times[-1,1]$. Besides, by these formulas it is easy to see that this estimate is sharp since we can concentrate waves around these rays (see subsection 1.4.1 where this analysis is carried out on the semi-discrete model).
Then, the best constant $C(\omega(\sigma))$ in (1.3.2) is precisely

$$C(\omega(\sigma)) = \sup_{t>0}\Big\{\frac{E(t)}{E_0}\exp(It)\Big\} = \sup_{t>0}\exp\Big(It - 2\inf_{\gamma\in\mathcal{R}_t}\int_\gamma \sigma(y)\, dy\Big).$$

It is then enough to compute

$$M = \sup_{t>0}\sup_{\gamma\in\mathcal{R}_t}\int_\gamma \Big(\frac{I}{2}-\sigma(y)\Big)\, dy.$$

Then, looking at rays γ_a^t starting at $a\in[-1,1]$ and traveling toward the left we get

$$\begin{aligned} M &\geq \sup_{t>0}\sup_a \int_{\gamma_a^t}\Big(\frac{I}{2}-\sigma(y)\Big)\, dy \\ &\geq \sup_a \sup_{t\in[1+a,3+a]}\Big(\int_{-1}^{a}\Big(\frac{I}{2}-\sigma(y)\Big)\, dy + \int_{-1}^{t-2-a}\Big(\frac{I}{2}-\sigma(y)\Big)\, dy\Big) \\ &\geq \sup_a\Big\{\int_{-1}^{a}\Big(\frac{I}{2}-\sigma(y)\Big)\, dy\Big\} + \sup_b\Big\{\int_{-1}^{b}\Big(\frac{I}{2}-\sigma(y)\Big)\, dy\Big\} \\ &\geq -2\inf_a\theta(a) = 2\,\|\theta\|_\infty . \end{aligned}$$

This implies that $C(\omega(\sigma)) \geq \exp(4\,\|\theta\|_\infty)$. The optimality of (1.3.2) follows.
The method of proof carries over to the other two estimates given in Theorem 1.3.2. The details are left to the reader. □

Note that all the results on the continuous model could have been obtained using this explicit representation formula along characteristics without using spectral analysis.

1.4 On the semi-discrete PML equations

In this section, we analyze the space semi-discrete PML system (1.1.14). For this purpose, we need to define a discrete space operator L_h, the discretization of L, defined in (1.1.11).
System (1.1.14) can be written as

$$\partial_t(P,V) + L_h(P,V) = 0,$$

where L_h is the discretization of L derived from (1.1.14). If we use a matrix representation, writing (P, V) as the vector

$$(V_{-N+1/2}, P_{-N+1}, V_{-N+3/2}, \cdots, P_{N-1}, V_{N-1/2}),$$

L_h is the matrix defined by

$$\begin{cases} L_h(j,j) = \sigma_{j/2-N}, \quad \forall j \in \{1, \ldots, 4N-1\}, \\ L_h(j,j+1) = \dfrac{1}{h}, \quad \forall j \in \{1, \cdots, 4N-2\}, \\ L_h(j+1,j) = -\dfrac{1}{h}, \quad \forall j \in \{1, \cdots, 4N-2\}, \\ L_h(i,j) = 0, \quad \text{if } |i-j| > 1. \end{cases} \tag{1.4.1}$$

If $\sigma_{j-1/2} = \sigma_j = \sigma_{j+1/2} = \sigma_{j+1}$, then both P_j and $V_{j+1/2}$ satisfy

$$\partial_{tt}^2 U_j - \frac{1}{h^2}(U_{j+1} + U_{j-1} - 2U_j) + 2\sigma_j \partial_t U_j + \sigma_j^2 U_j = 0, \tag{1.4.2}$$

which is a discretization of (1.1.8).
The energy E_h in (1.1.15) of the semi-discrete PML satisfies the dissipation law:

$$\frac{dE_h}{dt}(t) = -h \sum_{j=-N}^{N-1} \Big(\sigma_j |P_j|^2 + \sigma_{j+1/2} |V_{j+1/2}|^2 \Big). \tag{1.4.3}$$

It is then natural to investigate the decay rate of this discrete energy E_h when $h \to 0$. Our first result is of negative nature and states the lack of uniform exponential decay due to high frequency spurious oscillations:

Theorem 1.4.1. *There are no positive constants C and μ such that for all h small enough*

$$E_h(t) \leq C\, E_h(0) \exp(-\mu t), \tag{1.4.4}$$

for all solutions of (1.1.14).

One could have expected this behavior: indeed, it is well known since [34] that the group velocity for numerical schemes differs from the continuous case, because of the numerical dispersion relations. This indeed produces wave packets captured in the undamped subinterval $(-1, 0)$ and it is natural to expect them to have a very low exponential decay.
We will propose two proofs in the sequel. The first one is based on a very general construction of waves concentrated along the rays of Geometric Optics for system (1.1.14). More precisely, we construct non propagating waves concentrated in $(-1, 0)$, whose exponential decay rate tends to zero as $h \to 0$. In the second approach, we do a precise description of the spectrum of the operator L_h in (1.4.1) in the particular case where σ is constant. In particular, we prove that the real part of the high frequency eigenvalues can be small of order $o(1)$, which provides another proof of Theorem 1.4.1.

1.4.1 Construction of non propagating waves

We only sketch this construction, whose details can be done similarly as in [29, 30]. To simplify the presentation, we immediately focus on the behavior of the waves in $(-1, 0)$, that is in the domain where the damping is not effective. According to (1.4.2), system (1.1.14) reduces to the conservative space semi-discrete 1-d wave equation.

Let us therefore consider the semi-discrete 1-d wave equation in an infinite lattice $h\mathbb{Z}$, where h is the mesh size:

$$\begin{cases} \partial_{tt}^2 u_j - \Delta_h u_j = 0, \ (t,j) \in (0,\infty) \times \mathbb{Z}, \\ u_j(0) = u_j^0, \quad \partial_t u_j(0) = u^1(0). \end{cases} \tag{1.4.5}$$

We claim that this is sufficient to exhibit non propagating waves for system (1.4.5) to prove Theorem 1.4.1. Indeed, the system (1.1.14) coincides with system (1.4.5) for $j < 0$, $t \in [0,T]$, up to the boundary conditions, which can be easily handled. Namely, we will construct waves for system (1.4.5), whose energy is concentrated, for instance in $[-3/4,-1/4]$, in the sense that the energy outside $[-3/4,-1/4]$ is arbitrary small on $(0,T)$. Therefore, to obtain a true solution of (1.1.14), one needs to add arbitrary small corrections and hence the energy of (1.1.14), which satisfies the law (1.4.3), cannot decay exponentially.

To properly define the rays of Geometric Optics, we need to use the space discrete Fourier transform defined for $\xi h \in (-\pi,\pi]$ by:

$$\begin{aligned} \hat{\phi}(\xi) &= h \sum_j \phi_j \exp(-i\xi j h), \ \xi h \in (-\pi,\pi], \\ \phi^h(x) &= \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} \hat{\phi}(\xi) \exp(i\xi x) \ d\xi, \ x \in \mathbb{R}. \end{aligned} \tag{1.4.6}$$

Note that the inverse Fourier transform provides a natural extension of ϕ_j as a continuous function, denoted ϕ^h in the sequel.
The symbol of the operator (1.4.5) is given by

$$\tau^2 - \omega_h(\xi)^2, \quad \omega_h(\xi) = \frac{2}{h} \sin\Big(\frac{\xi h}{2}\Big). \tag{1.4.7}$$

Thus, taking $\zeta_0 \in (-\pi,\pi]$, the rays of Geometric Optics for frequencies $\xi_0^h = \zeta_0/h$ are the trajectories ([39]):

$$X_\pm^{\zeta_0} : (x_0,t) \to x_0 \pm t\cos(\zeta_0/2). \tag{1.4.8}$$

We then look for solutions concentrated along the trajectory $t \to X_+^{\zeta_0}(0,t)$. Note that we can take $x_0 = 0$ without loss of generality because of the translation invariance of system (1.4.5).

For we consider initial data of the form

$$u_j^{0,h} = \phi(jh)\exp(i\zeta_0 j), \quad u_j^{1,h} = i\omega_h(\xi_0^h)\phi(jh)\exp(i\zeta_0 j), \tag{1.4.9}$$

where ϕ is a smooth positive function of compact support in $(-a,a)$. Then, from the smoothness assumption on ϕ, one can prove that $\hat{u}^0$ and $\hat{u}(t)$ are concentrated in the region $\xi h \in [\zeta_0 - \epsilon_0, \zeta_0 + \epsilon_0]$, where ϵ_0 is a small parameter:

$$\begin{aligned} &\Big| u^{0,h}(x) - \frac{h}{2\pi} \int_{|\xi - \xi_0^h| < \epsilon_0/h} \hat{u}^0(\xi) \exp(i\xi x) \ d\xi \Big| \le \frac{C}{\omega_h(\epsilon_0/h)^2} \\ &\Big| u^h(t,x) - \frac{h}{2\pi} \int_{|\xi - \xi_0^h| < \epsilon_0/h} \hat{u}(t,\xi) \exp(i\xi x) \ d\xi \Big| \le C \frac{(1 + T\omega_h(\xi_0^h))}{\omega_h(\epsilon_0/h)^2}. \end{aligned} \tag{1.4.10}$$

On the other hand,

$$\hat{u}(t,\xi) = \hat{u}^0(\xi)\Big(\cos(t\omega_h(\xi)) + it\omega_h(\xi_0^h)\mathrm{sinc}(t\omega_h(\xi))\Big), \tag{1.4.11}$$

where $\mathrm{sinc}(y) = \sin(y)/y$. But, for ξ such that $|\xi - \xi_0^h| < \epsilon_0/h$, it is easy to see that this behaves as $\hat{u}^0(\xi)\exp(it\omega_h(\xi))$, and then the analysis of the oscillating integral in (1.4.10) gives that, when $h \to 0$,

$$\Big||u(t, x + t\cos(\zeta_0/2))| - |u^0(x)|\Big| \leq C\epsilon_0. \tag{1.4.12}$$

Choosing $\zeta_0 = \pi$ gives a sequence of solutions of (1.1.14) of unit energy such that the energy outside $\{(t,x) \in (0,T) \times \mathbb{R}, x \in X_+(t,[-a,a])\}$ tends to zero.
Note that the construction given above proves that the lack of uniform exponential decay of the energy actually takes its origin from the discretization scheme employed rather than from the PML method in itself.

1.4.2 Spectral analysis for constant σ

From now, we make the assumption that the damping function σ is a piecewise constant function vanishing in $(-1,0)$ and taking the value σ in $(0,1)$. This leads to set $\sigma_j = \sigma_{j-1/2} = \sigma$ if $j \geq 1$, $\sigma_j = \sigma_{j+1/2} = 0$ for $j \leq -1$ and $\sigma_0 = \sigma/2$.

In the sequel, as we did for the operator L, we perform a spectral analysis of the operator L_h. As we shall see, some numerical pathologies appear at high frequencies. More precisely, for frequencies of the order $2/h$ there appear eigenvalues whose real part is close to zero. This makes the exponential decay rate of the corresponding semigroups not uniform in h.

Accordingly, we analyze the asymptotic properties of the spectrum. We fix σ, and analyze the behavior of the eigenvalues of L_h when h goes to zero.

Proposition 1.4.2. *For $\sigma > 0$, we consider the spectral problem :*

$$\begin{cases} \dfrac{V_{j+1/2} - V_{j-1/2}}{h} + \sigma\,\chi_{j\geq 1} P_j = \lambda P_j,\ j \in \{-N+1, \cdots, N-1\}\backslash\{0\}, \\ \dfrac{P_{j+1} - P_j}{h} + \sigma\,\chi_{j\geq 1} V_{j+1/2} = \lambda V_{j+1/2},\ j \in \{-N, \cdots, N-1\}, \\ \dfrac{V_{1/2} - V_{-1/2}}{h} + \dfrac{\sigma}{2} P_0 = \lambda P_0, \\ P_{-N} = P_N = 0. \end{cases} \tag{1.4.13}$$

The following properties hold :

- *For any eigenvalue λ, its conjugate $\bar{\lambda}$ is also an eigenvalue.*
- *All the eigenvalues are simple.*
- *All the eigenvalues satisfy $0 < \mathcal{R}e(\lambda) < \sigma$ and $|\mathcal{I}m(\lambda)| \leq 2/h$.*
- *If λ is an eigenvalue, $\sigma - \lambda$ is also an eigenvalue.*

Proof. The first statement is obvious since the coefficients of system (1.4.13) are real. The second one is classical and follows from easy algebraic considerations. The third one is a consequence of the energy dissipation law (1.4.3):

$$0 \geq \frac{dE_h}{dt}(t) \geq -2\sigma E_h(t).$$

To analyze the imaginary part of the eigenvalues, we use the matrix representation of L_h given in (1.4.1): if $|\mathcal{I}m(\lambda)| > 2/h$, then the matrix $L_h - \lambda I$ is invertible, since it is diagonally dominant.
The last statement follows from this remark: If (P, V) is an eigenvector corresponding to λ, then $(\tilde{P}, \tilde{V})$ defined by $\tilde{P}_j = P_{-j}$ and $\tilde{V}_j = V_{-j+1}$ is an eigenvector for the eigenvalue $\sigma - \lambda$. □

From the previous proposition, we can assume that λ has a positive imaginary part, since the other eigenvalues can be obtained by reflection. Setting $\mu = \lambda - \sigma$, P satisfies

$$\begin{cases} \dfrac{P_{j+1} + P_{j-1} - 2P_j}{h^2} = \lambda^2 P_j, \quad j \leq -1, \\ \dfrac{P_{j+1} + P_{j-1} - 2P_j}{h^2} = \mu^2 P_j, \quad j \geq 1, \\ P_{-N} = P_N = 0. \end{cases}$$

As for the classical discrete Laplace operator, we define α and β, two complex numbers with imaginary parts in $(-\pi/h, \pi/h]$ and satisfying the numerical dispersion relations :

$$\sinh\Big(\frac{\alpha h}{2}\Big) = \frac{\lambda h}{2} \quad ; \quad \sinh\Big(\frac{\beta h}{2}\Big) = \frac{\mu h}{2}. \tag{1.4.14}$$

Then, we can express P for $j \leq -1$ and for $j \geq 1$ as

$$P_j = A \sinh(\alpha(jh+1)),\ j \leq -1, \quad P_j = B \sinh(\beta(jh-1)),\ j \geq 1.$$

These two quantities have to coincide at $j = 0$ and therefore:

$$A \sinh(\alpha) = -B \sinh(\beta).$$

We can then compute the corresponding value for V:

$$\begin{aligned} V_{j-1/2} &= A \cosh(\alpha((j-1/2)h+1)),\ j \leq 0 \\ V_{j-1/2} &= B \cosh(\beta((j-1/2)h-1)),\ j \geq 1. \end{aligned}$$

The transmission conditions are given by the equation on P_0:

$$V_{1/2} - \sinh\Big(\frac{\beta h}{2}\Big) P_0 = V_{-1/2} + \sinh\Big(\frac{\alpha h}{2}\Big) P_0.$$

Then if λ is an eigenvalue, there exists a non trivial solution (A, B) to the system:

$$\begin{cases} 0 = A \sinh(\alpha) + B \sinh(\beta) \\ 0 = A \cosh(\alpha) \cosh\Big(\dfrac{\alpha h}{2}\Big) - B \cosh(\beta) \cosh\Big(\dfrac{\beta h}{2}\Big), \end{cases}$$

where (α, β) are given by (1.4.14), μ being $\lambda - \sigma$. It is well-known that this system has non trivial solutions if and only if its determinant vanishes, that is to say:

$$\sinh(\alpha)\cosh(\beta)\cosh\Big(\frac{\beta h}{2}\Big) + \cosh(\alpha)\sinh(\beta)\cosh\Big(\frac{\alpha h}{2}\Big) = 0. \tag{1.4.15}$$

This equation actually is a polynomial in λ. Indeed, using Tchebychev polynomials P_{2k} and Q_{2k} defined by

$$\forall a \in \mathbb{C},\ \sinh(2ka) = \cosh(a) P_{2k}(\sinh(a)), \quad \cosh(2ka) = Q_{2k}(\sinh(a)),$$

the condition (1.4.15) is equivalent to

$$\cosh\Big(\frac{\alpha h}{2}\Big)\cosh\Big(\frac{\beta h}{2}\Big)\Big(P_{2N}\Big(\sinh\Big(\frac{\alpha h}{2}\Big)\Big)Q_{2N}\Big(\sinh\Big(\frac{\beta h}{2}\Big)\Big) + P_{2N}\Big(\sinh\Big(\frac{\beta h}{2}\Big)\Big)Q_{2N}\Big(\sinh\Big(\frac{\alpha h}{2}\Big)\Big)\Big) = 0. \quad (1.4.16)$$

This equation has two particular solutions corresponding to $\alpha h = i\pi$ and $\beta h = i\pi$. Nevertheless, although these two solutions allow a non-trivial choice (A, B), the corresponding solutions are identically zero, and therefore they do not correspond to eigenvalues. Since the degree of this polynomial in (1.4.16) is exactly $4N - 1$ and since all the eigenvalues are simple, the roots of (1.4.15) are exactly the eigenvalues of the problem, except the special solutions $\lambda = 2i/h$ and $\lambda = \sigma + 2i/h$.

Our interest now is to compute the eigenvalues, or at least to give their asymptotic form. We present

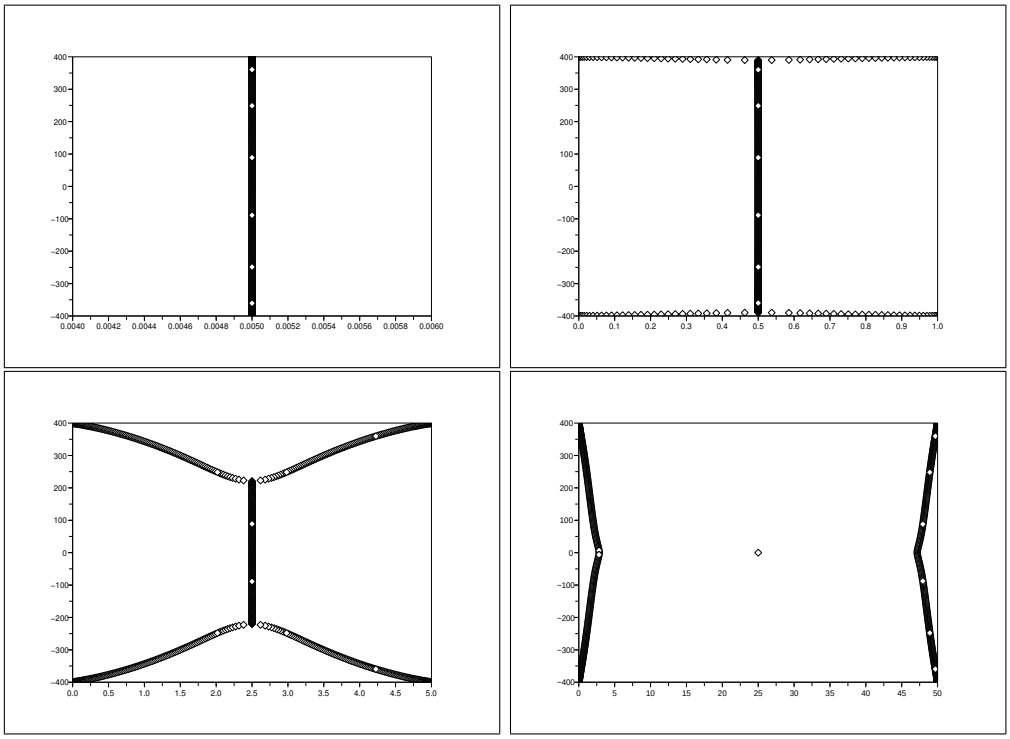

Figure 1.1: Eigenvalues for $N = 200$ and various values of σ : $\sigma = 0.01$ on the upper left, $\sigma = 1$ on the upper right, $\sigma = 5$ on the bottom left, $\sigma = 50$ on the bottom right.

in Figure 1.1 numerical computations of the distribution of eigenvalues for different values of σ. Three different cases occur. When σ is very small (of order h or less), then the real parts of the eigenvalues are very close to $\sigma/2$ at all frequencies. When σ is such that $h << \sigma << 1/h$, two branches appear at the high frequencies, their abscissa having two accumulation points, namely 0 and σ. Finally, Figure 1.1 illustrates the well-known fact ([17]) that, on the numerical approximation of PML equations, taking σ too large deteriorates the decay rate, in opposition to the continuous case. In the sequel, we will prove that these numerical evidences are indeed true.

To study the asymptotic behavior of the spectrum, we will need a number of notations.

We rewrite (1.4.15) as $f(\alpha,\beta,h)=0$, where f is defined by

$$f(\alpha,\beta,h) := \sinh(\alpha+\beta)\Big(\cosh\Big(\frac{\alpha h}{2}\Big)+\cosh\Big(\frac{\beta h}{2}\Big)\Big) + \sinh(\alpha-\beta)\Big(\cosh\Big(\frac{\beta h}{2}\Big)-\cosh\Big(\frac{\alpha h}{2}\Big)\Big). \quad (1.4.17)$$

In the sequel, we use the function Argsh defined as the inverse function of sinh, which coincides with $\log(z+\sqrt{1+z^2})$, which is holomorphic on the set $\Omega=\mathbb{C}\backslash\{z:\ \mathcal{R}e(z)=0,\ |\mathcal{I}m(z)|\geq 1\}$ and continuous at the points $z=\pm i$:

$$\forall z\in\Omega,\ \sinh(\mathrm{Argsh}(z))=z$$
$$\forall z\in\mathbb{C},\ s.t.\ \mathcal{I}m(z)\in(-\pi/2,\pi/2),\ \mathrm{Argsh}(\sinh(z))=z.$$

Then, β given by the relation (1.4.14) is an holomorphic function of α:

$$\beta(\alpha,h)=\frac{2}{h}\mathrm{Argsh}\Big(\sinh\Big(\frac{\alpha h}{2}\Big)-\frac{\sigma h}{2}\Big). \quad (1.4.18)$$

Hence the solutions of (1.4.15) correspond precisely to the roots α of the holomorphic function g

$$g(\alpha,h)=\cosh\Big(\frac{\alpha h}{2}\Big)\sinh(\alpha+\beta)+\Big(\cosh\Big(\frac{\beta h}{2}\Big)-\cosh\Big(\frac{\alpha h}{2}\Big)\Big)\sinh(\alpha)\cosh(\beta), \quad (1.4.19)$$

where $\beta=\beta(\alpha)$ as in (1.4.18). Of course, α given by (1.4.14) is a holomorphic function of λ and we can also define $\tilde{g}$ as a holomorphic function of λ by

$$\tilde{g}(\lambda,h):=g(\alpha(\lambda),h).$$

The analysis of the roots of (1.4.15) can be carried out using tools from complex analysis, as for instance Rouché's theorem.

The low frequencies We choose a number $\delta<1$ and study the eigenvalues λ of the operator L_h such that $|\mathcal{I}m(\lambda)h|\leq 2\delta$ when $h\to 0$.

Theorem 1.4.3. *Assume $\delta<1$. There exists C_δ such that for h small enough, the set of the eigenvalues λ_k^h of the operator L_h such that $|\mathcal{I}m(\lambda)h|\leq 2\delta$ is composed by one point in each disk D_k^h*

$$|\lambda-\hat{\lambda}_k^h|\leq C_\delta h,\quad \hat{\lambda}_k^h=\frac{2i}{h}\sin\Big(\frac{k\pi h}{4}\Big)+\frac{\sigma}{2}, \quad (1.4.20)$$

k being an integer satisfying $\Big|\sin\Big(\frac{k\pi h}{4}\Big)\Big|\leq\delta$.

Let us first remark that these disks D_k^h are disconnected for h small enough since the distance between two consecutive eigenvalues λ_k^h and λ_j^h is bounded from below by $\cos(\arcsin(\delta))=\sqrt{1-\delta^2}>0$. This implies that for h small enough, the number of eigenvalues in the range $|\mathcal{I}m(\lambda)h|\leq 2\delta$ is exactly $\lfloor\frac{8}{\pi h}\arcsin(\delta)\rfloor$ ($\lfloor\cdot\rfloor$ denotes the integer part).
Moreover, their real part being essentially $\sigma/2$, the energy of the solutions $\exp(-\lambda_k t)(P^{k,h},V^{k,h})$, where $(P^{k,h},V^{k,h})$ is an eigenvector associated to λ_k, is decreasing exponentially, the decay rate being $\sigma+o(h)$.

Proof. The proof is divided into two steps. First we derive some basic estimates on the parameters entering in (1.4.19). Second we approximate the function g by another holomorphic function $\hat{g}$ in order to apply Rouché's theorem.

We first need to derive some basic estimates on $\alpha(\lambda)$ given in (1.4.14), mainly by using the previous theorem. In the strip $|\mathcal{I}m(z)| \leq \delta$ and $|\mathcal{R}e(z)| \leq \sigma h$, if $z = a + ib$, we have that

$$z + \sqrt{1+z^2} = a + \sqrt{1-b^2} + ib\Big(1 + \frac{a}{\sqrt{1-b^2}}\Big) + O(h).$$

Then, we can check that the (complex) logarithm of that quantity satisfies:

$$|\mathcal{R}e(\mathrm{Argsh}(z))| \leq Ch \quad ; \quad |\tan(\mathcal{I}m(\mathrm{Argsh}(z)))| \leq \frac{\delta}{\sqrt{1-\delta^2}} + o(1),$$

where the constant C depends on δ. Then, using (1.4.14), we obtain the following estimates :

$$|\mathcal{R}e(\alpha)| \leq C \quad ; \quad |\mathcal{I}m(\alpha)| \leq \gamma = \arctan\Big(\frac{\delta}{\sqrt{1-\delta^2}}\Big). \tag{1.4.21}$$

Using (1.4.18) and the Taylor's formula applied to the function Argsh in $\sinh(\alpha h/2)$, we get that

$$\Big|\beta - \Big(\alpha - \frac{\sigma}{\cosh(\frac{\alpha h}{2})}\Big)\Big| \leq Ch. \tag{1.4.22}$$

Again using the estimates (1.4.21), we get

$$\left|\cosh\Big(\frac{\alpha h}{2}\Big)\sinh(\alpha+\beta) - \cosh\Big(\frac{\alpha h}{2}\Big)\sinh\left(2\alpha - \frac{\sigma}{\cosh\Big(\frac{\alpha h}{2}\Big)}\right)\right| \leq Ch.$$

The well-known formula $\cosh^2(x) = 1 + \sinh^2(x)$ and the estimates (1.4.21), (1.4.22) give

$$\Big|\cosh\Big(\frac{\beta h}{2}\Big) - \cosh\Big(\frac{\alpha h}{2}\Big)\Big| \leq Ch. \tag{1.4.23}$$

Combining all these inequalities and (1.4.19), we get that

$$|g(\alpha, h) - \hat{g}(\alpha, h)| \leq C_1 h, \tag{1.4.24}$$

where $\hat{g}$ is the function defined by :

$$\hat{g}(\alpha, h) = \cosh\Big(\frac{\alpha h}{2}\Big)\sinh\left(2\alpha - \frac{\sigma}{\cosh\Big(\frac{\alpha h}{2}\Big)}\right). \tag{1.4.25}$$

The roots of $\hat{g}$ satisfy

$$\hat{\alpha}_k^h = \frac{1}{2}\left(ik\pi + \frac{\sigma}{\cosh\Big(\frac{\hat{\alpha}_k^h h}{2}\Big)}\right).$$

From the estimate (1.4.21) on α, we can give the following approximation

$$\left|\hat{\alpha}_k^h - \frac{1}{2}\left(ik\pi + \frac{\sigma}{\cos\Big(\frac{k\pi h}{4}\Big)}\right)\right| \leq Ch.$$

For each h, we define $K_h = \lfloor \frac{4}{h\pi} \arcsin(\delta) \rfloor$. We consider the rectangle R_h delimited by the lines $|\mathcal{R}e(\alpha)| = M$ and $|2\mathcal{I}m(\alpha)| = \pi((K_h - 1) + \epsilon)$, where $\epsilon < 1$ is a positive number. On its boundary, we can check that

$$|\hat{g}(\alpha, h)| \geq |\sin(\pi\epsilon)| - Ch.$$

Using (1.4.24), there exists h_0 such that for all $h < h_0$, on the boundary of R_h,

$$|g(\alpha, h) - \hat{g}(\alpha, h)| < |\hat{g}(\alpha, h)|.$$

Then for all $h < h_0$, the number of roots in R_h is precisely $2K_h - 1$.
We can go further in the description of the zeros of $g(., h)$. We define

$$\tilde{\alpha}_k^h = \frac{1}{2}\left(ik\pi + \frac{\sigma}{\cos\left(\frac{k\pi h}{4}\right)} \right).$$

Now we fix the rectangle R_k^h by $|2\mathcal{I}m(\alpha - \tilde{\alpha}_k^h)| = \pi\epsilon_1$ and $|\mathcal{R}e(\alpha - \tilde{\alpha}_k^h)| = \epsilon_2$. On the boundary of R_k^h, again we can check that

$$|\hat{g}(\alpha, h)| \geq \inf\{|\sin(\pi\epsilon_1)|, |\sinh(\epsilon_2)|\} - Ch.$$

Then it exists a constant C_2 independent of k such that the conditions $|\epsilon_1| \geq C_2 h$ and $|\epsilon_2| \geq C_2 h$ are enough to prove that the following inequality holds on the boundary R_k^h :

$$|g(\alpha, h) - \hat{g}(\alpha, h)| < |\hat{g}(\alpha, h)|.$$

By Rouché's theorem, this establishes that $g(., h)$ has only one root α_k^h in R_k^h satisfying

$$|\alpha_k^h - \tilde{\alpha}_k^h| \leq Ch. \tag{1.4.26}$$

Back in the variable λ, it gives that for h small enough, each eigenvalue λ such that $|\mathcal{I}m(\lambda)h| \leq 2\delta$ is in one of the disks defined by

$$|\lambda - \hat{\lambda}_k^h| \leq Ch, \quad \hat{\lambda}_k^h = \frac{2i}{h}\sin(\frac{k\pi h}{4}) + \frac{\sigma}{2}.$$

□

The high frequencies Here we will deal with the limit case $\delta = 1$.

Theorem 1.4.4. *For any $\epsilon > 0$, there exists h_ϵ such that for all $h < h_\epsilon$, the set of eigenvalues satisfying $|h\mathcal{I}m(\lambda_h) - 2| \leq \epsilon$ is non empty. The set of accumulation points of the abscissa $\mathcal{R}e(\lambda_h)$ for sequences λ_h satisfying $\lambda_h h \to 2i$ when $h \to 0$ is exactly $\{0, \sigma\}$.*

Proof. The first point comes from the fact that a set of accumulation points is closed. Indeed, from the previous theorem, taking $\epsilon > 0$ and setting $\delta = 1 - \epsilon/4$, there exists a sequence of eigenvalues λ_h such that $\mathcal{I}m(\lambda_h)h \to 2\delta > 2 - \epsilon$.

Now we assume we have a sequence of eigenvalues λ_h for the operator L_h, such that $\lambda_h h \to 2i$, and we analyze the behavior of their real parts a_h. For that purpose, we need to know precisely how $\lambda_h h$ is converging to $2i$. We assume that

$$\frac{\mathcal{I}m(\lambda_h)h}{2} = 1 - \epsilon(h) \tag{1.4.27}$$

with $\epsilon(h)$ a positive function of h continuous at zero, such that $\epsilon(0) = 0$. To simplify notations, we will skip the index h in the sequel.
Remark that the difficulty comes from the fact that $\lambda h/2 \to i$, which is precisely a point where Argsh is not holomorphic anymore. However, from the explicit form of Argsh, we may derive some estimates on α and β. Indeed, recall that:

$$\mathrm{Argsh}(z) = \log(z + \sqrt{1+z^2}) \quad ; \quad \cosh(z) = \sqrt{1+\sinh(z)^2}.$$

Actually, it is sufficient to estimate these functions. Since

$$1 + \Big(\frac{\lambda h}{2}\Big)^2 = 2\epsilon(h) - \epsilon(h)^2 + \Big(\frac{ah}{2}\Big)^2 + i(1-\epsilon(h))\frac{ah}{2},$$

we will need to distinguish several cases depending on which is the dominant term.

The case $h = o(\epsilon(h))$: In that case, we get that

$$\cosh\Big(\frac{\alpha h}{2}\Big) = \sqrt{2\epsilon(h)} + o(\sqrt{\epsilon(h)}).$$

This also implies that

$$\mathcal{R}e\Big(\frac{\alpha h}{2}\Big) = \frac{1}{2}\log\Big|z + \sqrt{1+z^2}\Big|^2 = \epsilon(h) + o(\epsilon(h)).$$

And the same estimates hold true for β.
It follows that $f(\alpha, \beta, h)$ defined in (1.4.17) cannot vanish. Indeed, our estimates imply that the real parts of both α and β blow up, which implies that

$$\begin{aligned}
&|\sinh(\alpha+\beta)| = \exp\Big(4\frac{\epsilon(h)}{h} + o\Big(\frac{\epsilon(h)}{h}\Big)\Big),\\
&|\sinh(\alpha-\beta)| \le \exp\Big(o\Big(\frac{\epsilon(h)}{h}\Big)\Big),\\
&\Big|\cosh\Big(\frac{\alpha h}{2}\Big) + \cosh\Big(\frac{\beta h}{2}\Big)\Big| = \sqrt{2\epsilon(h)} + o\Big(\sqrt{\epsilon(h)}\Big),\\
&\Big|\cosh\Big(\frac{\alpha h}{2}\Big) - \cosh\Big(\frac{\beta h}{2}\Big)\Big| \le o\Big(\sqrt{\epsilon(h)}\Big).
\end{aligned}$$

The case $\epsilon(h) = o(h)$: Under this assumption, we get

$$\cosh\Big(\frac{\alpha h}{2}\Big) = \sqrt{i\frac{ah}{2}} + o(\sqrt{h}), \quad \cosh\Big(\frac{\beta h}{2}\Big) = \sqrt{-i\frac{(\sigma-a)h}{2}} + o(\sqrt{h}).$$

Besides, using the explicit formula of the function Argsh, we obtain :

$$\mathcal{R}e\Big(\frac{\alpha h}{2}\Big) = \frac{\sqrt{ah}}{2} + o(\sqrt{h}), \quad \mathcal{R}e\Big(\frac{\beta h}{2}\Big) = -\frac{\sqrt{(\sigma-a)h}}{2} + o(\sqrt{h}).$$

But these estimates lead to

$$\begin{aligned}
&\Big|\cosh\Big(\frac{\alpha h}{2}\Big) + \cosh\Big(\frac{\beta h}{2}\Big)\Big| = \sqrt{\sigma h} + o(\sqrt{h}),\\
&\Big|\cosh\Big(\frac{\alpha h}{2}\Big) - \cosh\Big(\frac{\beta h}{2}\Big)\Big| = \sqrt{\sigma h} + o(\sqrt{h})
\end{aligned}$$

and

$$|\sinh(\alpha+\beta)| \simeq \exp(\frac{1}{2}|\sqrt{ah} - \sqrt{(\sigma-a)h}|),$$
$$|\sinh(\alpha-\beta)| \simeq \exp(\frac{1}{2}\sqrt{ah} + \sqrt{(\sigma-a)h}).$$

Thus, if $f(\alpha,\beta,h)=0$, f being as in (1.4.17), we need that $|\,|\sqrt{\sigma-a}-\sqrt{a}| - (\sqrt{\sigma-a}+\sqrt{a})| \to 0$, which implies $a \to 0$ or $a \to \sigma$.

The case where $\epsilon(h) = Kh$ follows from similar considerations and is left to the reader.

Summarizing, we deduce the existence of a sequence of eigenvalues such that $\lambda h \to 2i$, and hence whose real part is converging to zero or σ. To finish the analysis, we only have to prove that both 0 and σ are accumulation points. This assertion is obvious since the spectrum is symmetric around $\sigma/2$. □

Theorems 1.4.3 and 1.4.4 fully explain Figure 1.1 for $h << \sigma << 1/h$, since they state, roughly speaking, that the eigenvalues λ are close to the line $\mathcal{R}e(\lambda) = \sigma/2$ except when their imaginary part is close to $\pm 2/h$, in which case, their real parts tend to 0 or σ.

To describe the behavior of the eigenvectors, we define the energies in the left and right intervals $(-1,0)$ and $(0,1)$, respectively :

$$\begin{cases} E_h^l = \dfrac{h}{4}|P_0|^2 + \dfrac{h}{2}\displaystyle\sum_{j=1}^{N}(|P_j|^2 + |V_{j-1/2}|^2), \\ \\ E_h^r = \dfrac{h}{4}|P_0|^2 + \dfrac{h}{2}\displaystyle\sum_{j=-N}^{-1}(|V_{j+1/2}|^2 + |P_j|^2). \end{cases} \tag{1.4.28}$$

Proposition 1.4.5 (Distribution of the energy)**.** *Let $(\lambda_k^h)_h$ be a sequence of eigenvectors of L_h such that $h\mathcal{I}m(\lambda_k^h) \to 2$, and that $a_k^h = \mathcal{R}e(\lambda_k^h)$ converges to a. Then*

$$\frac{E_h^r(P_k^h, V_k^h)}{E_h^l(P_k^h, V_k^h)} \underset{h\to 0}{\longrightarrow} \frac{a}{\sigma - a}. \tag{1.4.29}$$

In particular, there exists a sequence of high frequency eigenvectors whose energy is concentrated on the left interval $(-1,0)$.

Proof. In view of (1.4.3), the solution $\exp(-\lambda_k^h t)(P_k^h, V_k^h)$ corresponding to the eigenvector (P_k^h, V_k^h) satisfies

$$\frac{dE_h}{dt}(t) = -2\mathcal{R}e(\lambda_k^h)E_h(t) = -2\sigma E_h^r(t).$$

The result follows. □

Remark 1.4.6. According to this result we have a new evidence of the lack of uniform exponential decay, as stated in Theorem 1.4.1. There this was proved by means of a gaussian beam construction, whereas here we have built concentrated eigenvectors.

1.4.3 Connections with the theory of stabilization

In this subsection, we discuss the links between our analysis and the existing controllability and stabilization theory and reread our results in this context.

Let us consider the 1-d damped wave equation (1.1.6) on $(-1,1)$. The decay rate of the solutions of this damped wave equation has been analyzed in several articles: see [19, 18, 20] and [27] for the multi-dimensional case. The exponential decay rate was characterized as the minimum of the spectral abscissa and the minimal value of the damping potential along the rays of geometric optics (In 1-d, these two quantities coincide as shown in [19]). One of the main features of system (1.1.6) is that an overdamping phenomenon occurs, in the sense that increasing the damping potential does not necessarily increase the decay rate. This is not the case for the PML system since, as observed in Theorem 1.2.1 and 1.3.1 the decay rate is $I = \int_0^1 \sigma(x)\, dx$, and this is precisely what makes PML so efficient.

We may now investigate the same questions in the semi-discrete 1-d case on a regular mesh of size $h = 1/N$. Then the finite difference approximation of (1.1.6) gives :

$$\begin{cases} \partial_{tt}^2 u_j - \Delta_h u_j + 2a_j \partial_t u_j = 0, \quad j \in \{-N+1, \dots, N-1\}, \\ u_{-N} = u_N = 0. \end{cases} \tag{1.4.30}$$

It was proved in [25, 30, 33] that the energy of solutions of (1.4.30) does not decay exponentially uniformly with respect to the mesh size h. Actually, this lack of uniform exponential decay can be deduced from the construction given in Subsection 1.4.1. As pointed out in [23], this has also interesting consequences when analyzing the optimal choice of dampers in which one observes also a different behavior from the continuous to the discrete case.
We claim that this lack of uniform exponential decay can also be seen at the level of the spectrum. If we set $v_j = u_j'$, the system takes the form:

$$\frac{d}{dt}(u_{-N+1}, \cdots, u_{N-1}, v_{-N+1} \cdots, v_{N-1})^* + A(u_{-N+1}, \cdots, v_{N-1})^* = 0,$$

where A is the following matrix:

$$A = \begin{pmatrix} 0 & -I_{2N-1} \\ -\Delta_h & 2\text{diag}\,(a_{-N+1}, \cdots, a_{N-1}) \end{pmatrix}.$$

We have performed the spectral computation of this matrix for piecewise constant damping potentials vanishing in $(-1,0)$ and taking a constant value a on $(0,1)$. The spectrum exhibits a behavior which is very close to the one we have observed for the PML system (see Figure 1.2), except at the low frequencies, where we observe the so-called overdamping phenomenon, which is reminiscent of the continuous system.

1.5 A semi-discrete viscous PML

The goal of this section is to propose a remedy to the defect of exponential decay proved in the previous section (see Theorem 1.4.1) for the semi-discrete approximation (1.1.14) of the PML system.
Along this section, we assume that $\sigma \in L^\infty(-1,1)$ is a positive function strictly positive on a subinterval (r_1, r_2) of $(0,1)$. To be more precise :

$$0 \le \sigma(x) \le M, \ x \text{ a.e.} \in (-1,1), \qquad \sigma(x) \ge m > 0, \ x \text{ a.e.} \in (r_1, r_2). \tag{1.5.1}$$

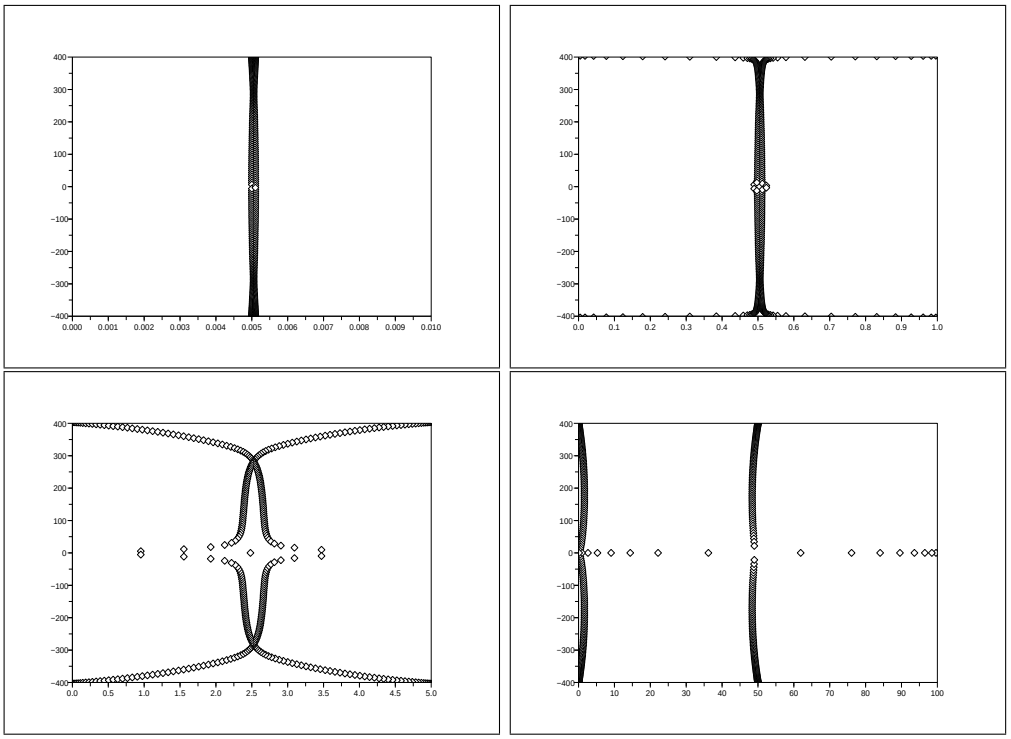

Figure 1.2: Eigenvalues of the semi-discrete damped wave equation (1.4.30) for $N = 200$ and various values of the damping potentials a : $a = 0.01$ in the upper left, $a = 1$ in the upper right, $a = 5$ on the bottom left, $a = 50$ on the bottom right.

For each h, we define σ_j^h as an approximation of σ in the points $x_j = jh$ satisfying

$$0 \leq \sigma_j^h \leq M, \ \forall j, \quad \sigma_j^h \geq m, \ \forall j \ s.t. \ jh \in (r_1, r_2). \tag{1.5.2}$$

To simplify the notations, we will write σ_j in the sequel, the dependence in h being clear within the context.
We propose to analyze system (1.1.16), which is a variant of the semi-discrete scheme (1.1.14), where a numerical viscosity term damping out the high frequencies has been added. Recall that, for system (1.1.16), the energy dissipation law (1.1.17) holds. In this way, the new semi-discrete problem satisfies the required property of uniform exponential decay:

Theorem 1.5.1. *Under the hypothesis* (1.5.2)*, there exist two positive constants C and μ such that for all $h > 0$, for all initial data (P_0^h, V_0^h), the energy of the solution (P, V) of* (1.1.16) *satisfies*

$$E_h(t) \leq C \, E_h(0) \exp(-\mu t), \ t > 0. \tag{1.5.3}$$

Furthermore, we will see in Theorem 1.5.3 that one can choose the numerical viscosity such that this decay rate coincides with the continuous one I.

Proof. The method of proof we will use is classical in the theory of stabilization.
We claim that the energy of this viscous numerical approximation scheme (1.1.16) is exponentially decaying, uniformly in h, if and only if the following observability inequality holds for some time T

and a constant C uniformly in h for all the solutions of (1.1.16):

$$E_h(0) \leq C\Bigg(h\sum_j \int_0^T \left(\sigma_j |P_j|^2 + \sigma_{j+1/2}|V_{j+1/2}|^2\right)\, dt \\ + h^3 \sum_j \int_0^T \Big[\Big(\frac{P_{j+1}-P_j}{h}\Big)^2 + \Big(\frac{V_{j+1/2}-V_{j-1/2}}{h}\Big)^2\Big]\, dt\Bigg). \quad (1.5.4)$$

Indeed, according to the energy dissipation law (1.1.17), we easily deduce that the two statements are equivalent.
On the other hand, to prove (1.5.4) for solutions of (1.1.16), it is sufficient to prove the existence of a time T and a constant C such that for all $h>0$, any solution (p,v) of the conservative system (1.1.14) with $\sigma = 0$ satisfies

$$E_h(0) \leq C\Bigg(h\sum_{jh\in(r_1,r_2)} \int_0^T m(|p_j|^2 + |v_{j+1/2}|^2)\, dt \\ + h^3 \sum_j \int_0^T \Big[\Big(\frac{p_{j+1}-p_j}{h}\Big)^2 + \Big(\frac{v_{j+1/2}-v_{j-1/2}}{h}\Big)^2\Big]\, dt\Bigg). \quad (1.5.5)$$

Indeed, since the two systems (1.1.16) and (1.1.14) with $\sigma = 0$ coincide up to a term which can be bounded by the right hand-side quantity in (1.5.4), it can be shown that inequality (1.5.4) follows from inequality (1.5.5). The details of this process are classical and can be found for instance in [33].
From now, we focus on the observability inequality (1.5.5) for the conservative system (1.1.14), that we prove using a multiplier method. Given $K > \sup\{1+r_1,\ 1-r_2\}$, where r_1 and r_2 are given by (1.5.1) and (1.5.2), we define a discrete function η^h satisfying the following properties:

$$\begin{cases} \eta^h_{-N} = \eta^h_N = 0, \quad |\eta^h_j| \leq K, \quad \forall j, \\ \dfrac{\eta^h_{j+1}-\eta^h_j}{h} = 1, \quad \forall j \ s.t.\ jh \in [-1,1]\backslash(r_1,r_2), \\ \Big|\dfrac{\eta^h_{j+1}-\eta^h_j}{h}\Big| \leq \dfrac{3}{r_2-r_1}, \quad \forall j. \end{cases} \quad (1.5.6)$$

Actually, we can choose η^h as a discrete approximation of a continuous piecewise affine function η. In the sequel we therefore write η instead of η^h to simplify the notations. For convenience, we also denote $(\eta_j + \eta_{j+1})/2$ by $\eta_{j+1/2}$.
Multiplying the first line of the conservative system (1.1.14) by $\eta_j(v_{j-1/2}+v_{j+1/2})$ and the second by $\eta_{j+1/2}(p_j+p_{j+1})$, after tedious computations mainly involving discrete integration by parts, we get :

$$h\sum_{j=-N}^{N-1}\Big[v_{j+1/2}(T)\Big(\eta_j p_j(T) + \eta_{j+1}p_{j+1}(T)\Big) - v_{j+1/2}(0)\Big(\eta_j p_j(0) + \eta_{j+1}p_{j+1}(0)\Big)\Big] \\ - h\sum_{j=-N}^{N-1}\int_0^T \Big(\frac{\eta_{j+1}-\eta_j}{h}\Big)|v_{j+1/2}|^2\, dt - h\sum_{j=-N+1}^{N-1}\int_0^T \Big(\frac{\eta_{j+1/2}-\eta_{j-1/2}}{h}\Big)|p_j|^2\, dt \\ - \frac{h^3}{2}\int_0^T \sum_{j=-N}^{N-1} \partial_t v_{j+1/2}\Big(\frac{\eta_{j+1}-\eta_j}{h}\Big)\Big(\frac{p_{j+1}-p_j}{h}\Big)\, dt = 0. \quad (1.5.7)$$

The conservation of the energy allows us to bound the time boundary term by $4K\,E_h(0)$ thanks to the following inequality:

$$\Big|v_{j+1/2}(\eta_j p_j + \eta_{j+1}p_{j+1})\Big| \leq K|v_j|^2 + \frac{K}{2}(|p_j|^2 + |p_{j+1}|^2).$$

The only term in which numerical viscosity is needed is the last one:

$$A = -\frac{h^3}{2}\int_0^T \sum_{j=-N}^{N-1} \partial_t v_{j+1/2}\Big(\frac{\eta_{j+1}-\eta_j}{h}\Big)\Big(\frac{p_{j+1}-p_j}{h}\Big)\,dt.$$

Since (p,v) is a solution of the conservative system (1.1.14), we get

$$\begin{aligned} A &= \frac{h^3}{2}\int_0^T \sum_{j=-N}^{N-1} \Big(\frac{\eta_{j+1}-\eta_j}{h}\Big)\Big(\frac{p_{j+1}-p_j}{h}\Big)^2\,dt \\ &\leq \frac{3}{r_2-r_1}\frac{h^3}{2}\int_0^T \sum_{j=-N}^{N-1} \Big(\frac{p_{j+1}-p_j}{h}\Big)^2\,dt. \end{aligned}$$

On the other hand, due to the assumptions (1.5.6) on η, we have

$$\begin{aligned} h\sum_{j=-N}^{N-1}\int_0^T \Big(\frac{\eta_{j+1}-\eta_j}{h}\Big)|v_{j+1/2}|^2\,dt + h\sum_{j=-N+1}^{N-1}\int_0^T \Big(\frac{\eta_{j+1/2}-\eta_{j-1/2}}{h}\Big)|p_j|^2\,dt \\ \geq 2TE_h(0) - \Big(1+\frac{3}{r_2-r_1}\Big)\,h\sum_{jh\in(r_1,r_2)}\int_0^T (|p_j|^2+|v_{j+1/2}|^2)\,dt. \end{aligned}$$

Combining these inequalities we get

$$\begin{aligned} (2T-4K)E_h(0) \leq \frac{1}{m}\Big(1+\frac{3}{r_2-r_1}\Big)\int_0^T h\sum_{jh\in(r_1,r_2)} m(|p_j|^2+|v_{j+1/2}|^2)\,dt \\ + \frac{3}{r_2-r_1}\int_0^T h^3\sum_j\Big(\frac{p_{j+1}-p_j}{h}\Big)^2\,dt. \quad (1.5.8) \end{aligned}$$

This completes the proof of Theorem 1.5.1. Note that, by this method, we find that the observability inequality (1.5.5) actually holds for any $T > 2\sup\{1+r_1,\ 1-r_2\}$ (r_1 and r_2 as in (1.5.1) and (1.5.2)), which corresponds precisely to the optimal characteristic time in the continuous setting. □

Remark 1.5.2. We emphasize that Theorem 1.5.1 is false if we do not add viscosity everywhere in the domain. Indeed, the construction given in Subsection 1.4.1 proves that if the viscosity is not everywhere in the domain, there exist non-propagating waves which are not damped.
Also note that the proof above actually yields a stronger result than the one stated in Theorem 1.5.1. Indeed, following the previous proof, inequality (1.5.8) shows that this is actually enough to add the viscosity into only one of the two equations (1.1.16) to obtain a uniform exponential decay of the energy.

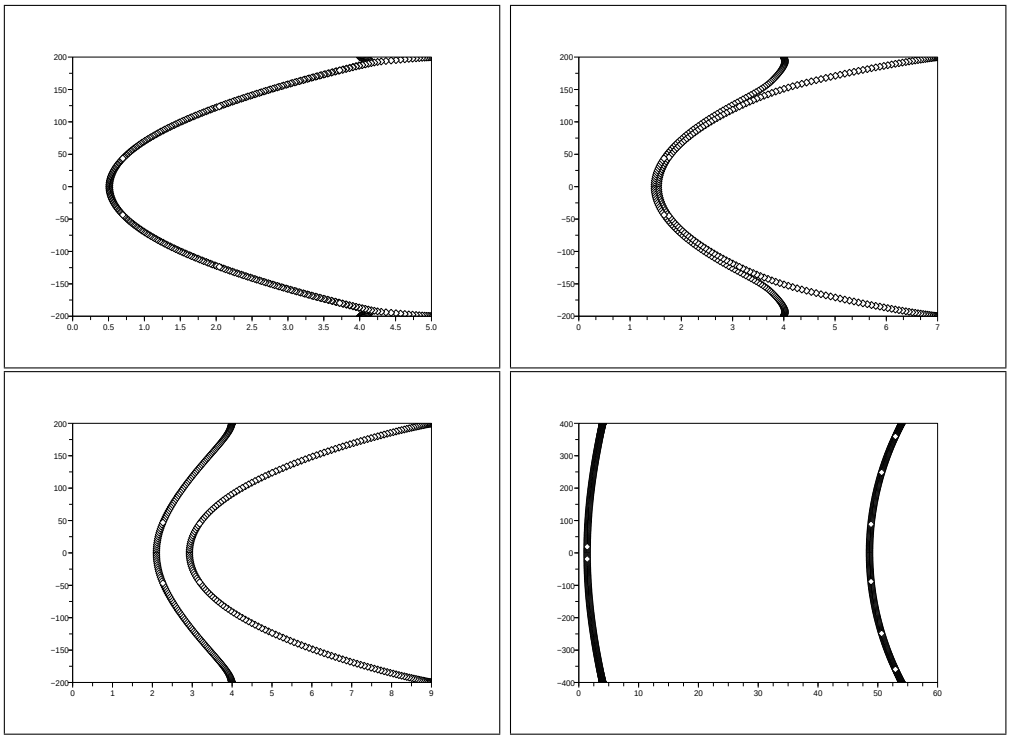

Figure 1.3: Eigenvalues of the viscous scheme (1.1.16) for $N = 100$ and various values of σ: $\sigma = 1$ on the upper left, $\sigma = 3$ on the upper right, $\sigma = 5$ on the bottom left and $\sigma = 50$ on the bottom right.

Unfortunately, the method of proof of Theorem 1.5.1 does not give a good estimate on the decay rate in terms of the parameters entering in the system. Since the system under consideration is finite dimensional, the decay rate of the energy is obviously given by the spectral abscissa. Therefore we have computed the eigenvalues of the system (1.1.16) in Figure 1.3 for damping potentials vanishing in $(-1, 0)$ and taking the value σ in $(0, 1)$. We observe that, first, at low frequencies, the numerical viscosity does not seem to change the spectrum, as one can check by comparing the figures with the ones obtained without the viscosity term (see Figure 1.1). This indicates that, as expected, the numerical viscosity does not modify the system at low frequencies. Second, at intermediate and high frequencies, one can see that the spectrum has a parabolic shape. Actually, one can easily check that, when $\sigma = 0$, the spectrum of (1.1.16) is exactly a parabolic curve $\mathcal{C}$. It is surprising to check that the spectrum given in Figure 1.3 fits quite well with the curve $\sigma/2 + \mathcal{C}$. Third, looking more closely at the high frequencies, the same phenomenon as before occurs, that is, two branches appear, corresponding to eigenvectors concentrated either in $(-1, 0)$, either in $(0, 1)$. But, thanks to the numerical viscosity, which efficiently damps them out, these two branches are away from zero. Moreover, it appears that the abscissa of the lowest branch is always 4. This precisely corresponds to the abscissa of the high frequency eigenvectors when $\sigma = 0$ in (1.1.16). In other words, this corresponds to waves concentrated in the undamped part $(-1, 0)$, which are only dissipated by the additional viscosity.

In view of these spectral properties and with the purpose of recovering at the semi-discrete level the properties of the continuous PML system, it is natural to ask whether one can choose numerical viscosity coefficients α such that the decay rate μ_h of (1.1.16) as $h \to 0$ converges to I.
In the sequel, we address this issue. System (1.1.16) can be read as:

$$\partial_t(P, V) + (A_h + B_h)(P, V) = \alpha h^2 A_h^2(P, V), \tag{1.5.9}$$

where $A_h + B_h = L_h$, and

$$\begin{aligned}(A_h(P,V))_j &= \Big(\frac{V_{j+1/2}-V_{j-1/2}}{h}, \frac{P_{j+1}-P_j}{h}\Big),\\ (B_h(P,V))_j &= (\sigma_j^h P_j, \sigma_{j+1/2}^h V_{j+1/2}).\end{aligned}$$

We need the following assumption:
There exists $\delta > 0$, such that for h small enough, the eigenvalues $\lambda_h = a_h + ib_h$ of $L_h = A_h + B_h$ with $|b_h| \le \delta/h$ satisfy

$$a_h \ge I/2 + o_{h\to 0}(1). \tag{1.5.10}$$

Note that in the particular case where σ is constant, (1.5.10) holds for any $\delta < 2$ (see Theorem 1.4.3). We expect this property to hold for non constant σ as well, but this issue will be addressed elsewhere.

Theorem 1.5.3. *Fix $\alpha = \alpha_\delta = I/\delta$ in (1.5.9), with δ as in (1.5.10). Then, for all h small enough, there exists C_h such that the solutions (P,V) of (1.5.9) satisfy:*

$$E_h(t) \le C_h E_h(0) \exp(-(I - o_{h\to 0}(1))t),\ t > 0. \tag{1.5.11}$$

Note that the constant C_h in (1.5.11) depends on h. In particular, we cannot guarantee C_h to be bounded.

Proof. Let us first consider the following modification of (1.5.9):

$$\partial_t(P,V) + (A_h + B_h)(P,V) = \alpha h^2 (A_h + B_h)^2 (P,V), \tag{1.5.12}$$

It is straightforward to show that the eigenvalues $\mu(\alpha)$ of system (1.5.12) can be expressed in terms of $\mu(0)$, which coincide with the eigenvalues $\lambda = a + ib$ of system (1.4.13):

$$\mu(\alpha) = \lambda - \alpha h^2 \lambda^2, \quad \mathcal{R}e(\mu(\alpha)) = a + \alpha h^2 (b^2 - a^2).$$

Under assumption (1.5.10), with the choice $\alpha = \alpha_\delta$, each eigenvalue $\mu(\alpha_\delta)$ satisfies

$$\mathcal{R}e(\mu(\alpha_\delta)) \ge I/2 - o_{h\to 0}(1). \tag{1.5.13}$$

Then, since the system is finite dimensional, there exists a constant C_h such that the solutions (P,V) of (1.5.12) satisfy

$$E_h(t) \le C_h E_h(0) \exp(-(I - o_{h\to 0}(1))t),\ t > 0.$$

Now, we estimate the norm of the matrix $D_h = (A_h + B_h)^2 - A_h^2$:

$$\begin{aligned}D_h(P,V)_j = \Bigg(& 2\sigma_j \Big(\frac{V_{j+1/2}-V_{j-1/2}}{h}\Big) + \sigma_j^2 P_j + \Big(\frac{V_{j+1/2}+V_{j-1/2}}{2}\Big)\Big(\frac{\sigma_{j+1/2}-\sigma_{j-1/2}}{h}\Big),\\ & \Big(\frac{P_{j-1}+P_j}{2}\Big)\Big(\frac{\sigma_{j+1}-\sigma_{j-1}}{h}\Big) + \sigma_{j+1/2}^2 V_{j+1/2} + \Big(\sigma_{j+1/2} + \frac{\sigma_j + \sigma_{j+1}}{2}\Big)\Big(\frac{P_{j+1}-P_j}{h}\Big)\Bigg).\end{aligned}$$

Note that systems (1.5.9) and (1.5.12) differ precisely by the term associated with $\alpha h^2 D_h$. Then, since

$$\big\|\alpha h^2 D_h\big\|_{L^{2,h}\to L^{2,h}} \le Ch, \tag{1.5.14}$$

where $L^{2,h}$ denotes the discrete $L^2(-1,1)$ norm, a simple perturbation argument gives the result. Indeed, setting $L_h(\alpha) = L_h - \alpha h^2 L_h^2$, the solution $\psi = (P, V)$ of (1.5.9) is given by

$$\exp(tL_h(\alpha_\delta))\psi(t) = \psi(0) - \int_0^t \exp(sL_h(\alpha_\delta))\alpha_\delta h^2 D_h \psi(s)\ ds.$$

Setting

$$f(t) = \exp(tI/2)\left\|\psi(t)\right\|,$$

this gives the equation

$$f(t) \leq f(0) + Ch\int_0^t f(s)\ ds,$$

and then Gronwall's lemma gives the result. □

1.6 Discussion and remarks

In this paper we have presented a complete analysis of the decay of the energy of the 1-d PML system both at the continuous and semi-discrete settings.

1. Analyzing the continuous system, we have shown that the two relevant parameters to describe the dissipation of the energy are $I = \int_0^1 \sigma(x)\ dx$ and $\|\theta\|_\infty$ as in (1.2.5). The exponential decay rate is exactly I while θ enters in the estimate of the multiplicative constant $C(\omega(\sigma))$ (see Theorem 1.3.1). This also confirms the interest in taking singular $\sigma \notin L^1$ as in [11, 13, 14].
2. An interesting question would be to investigate the decay of the energy in higher dimensions and to make precise which are the relevant parameters entering in it. According to [27], one could expect that the abscissa of the high frequency eigenvalues is related to the mean value of the damping along the rays of Geometric Optics. But the analysis of the low frequencies could be more complex, because of the possible overdamping phenomena, that could arise in the multi-dimensional case, although they have been excluded in 1-d.
3. At the semi-discrete level, we have studied in detail 1-d finite-difference approximation schemes. However, our analysis holds in a much more general setting. For instance, the same results holds for a finite element method. Besides, the construction we did in subsection 1.4.1 can also be done for semi-discrete multi-dimensional problems. Especially, the discrete energy will not decay uniformly on the mesh size, and a numerical viscosity will be needed to recover the property of exponential decay of the energy.
4. To the best of our knowledge, Theorem 1.5.3 is the first one where the uniform decay rate of the energy for an approximation scheme is proved to coincide with the decay rate of the energy of the continuous equation. This subject requires further investigation, for instance in the context of the damped wave equation. Moreover, this could be of significant importance in optimal design problems (see [23]), the goal being to design numerical schemes for which the optimal dampers converge to those of the continuous model. In view of Theorem 1.5.3 it is very likely that for a suitable viscous semi-discretization of the damped wave equation (1.1.6) this convergence property will hold.

Bibliography

[1] S. Abarbanel and D. Gottlieb. A mathematical analysis of the PML method. *J. Comput. Phys.*, 134(2):357–363, 1997.

[2] S. Abarbanel and D. Gottlieb. On the construction and analysis of absorbing layers in CEM. *Appl. Numer. Math.*, 27(4):331–340, 1998. Absorbing boundary conditions.

[3] S. Abarbanel, D. Gottlieb, and J. S. Hesthaven. Well-posed perfectly matched layers for advective acoustics. *J. Comput. Phys.*, 154(2):266–283, 1999.

[4] D. Appelö, T. Hagstrom, and G. Kreiss. Perfectly matched layers for hyperbolic systems: general formulation, well-posedness, and stability. *SIAM J. Appl. Math.*, 67(1):1–23 (electronic), 2006.

[5] H. T. Banks, K. Ito, and C. Wang. Exponentially stable approximations of weakly damped wave equations. In *Estimation and control of distributed parameter systems (Vorau, 1990)*, volume 100 of *Internat. Ser. Numer. Math.*, pages 1–33. Birkhäuser, Basel, 1991.

[6] A. Bayliss, M. Gunzburger, and E. Turkel. Boundary conditions for the numerical solution of elliptic equations in exterior regions. *SIAM J. Appl. Math.*, 42(2):430–451, 1982.

[7] E. Bécache, S. Fauqueux, and P. Joly. Stability of perfectly matched layers, group velocities and anisotropic waves. *J. Comput. Phys.*, 188(2):399–433, 2003.

[8] E. Bécache and P. Joly. On the analysis of Bérenger's perfectly matched layers for Maxwell's equations. *M2AN Math. Model. Numer. Anal.*, 36(1):87–119, 2002.

[9] E. Bécache, P. G. Petropoulos, and S. D. Gedney. On the long-time behavior of unsplit perfectly matched layers. *IEEE Trans. Antennas and Propagation*, 52(5):1335–1342, 2004.

[10] J.-P. Bérenger. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.*, 114(2):185–200, 1994.

[11] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. An exact bounded PML for the Helmholtz equation. *C. R. Math. Acad. Sci. Paris*, 339(11):803–808, 2004.

[12] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. Numerical simulation of time-harmonic scattering problems with an optimal PML. *Sci. Ser. A Math. Sci. (N.S.)*, 13:58–71, 2006.

[13] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. An optimal perfectly matched layer with unbounded absorbing function for time-harmonic acoustic scattering problems. *J. Comput. Phys.*, 223(2):469–488, 2007.

[14] A. Bermúdez, L.M. Hervella-Nieto, A. Prieto, and R. Rodríguez. An exact bounded perfectly matched layer for time-harmonic scattering problems. *SIAM Journal on Scientific Computing*, to be published, 2007.

[15] J. H. Bramble and J. E. Pasciak. Analysis of a finite element PML approximation for the three dimensional time-harmonic Maxwell problem. *Math. Comp.*, 77(261):1–10 (electronic), 2008.

[16] C. Castro and S. Micu. Boundary controllability of a linear semi-discrete 1-d wave equation derived from a mixed finite element method. *Numer. Math.*, 102(3):413–462, 2006.

[17] F. Collino and P. Monk. The perfectly matched layer in curvilinear coordinates. *SIAM J. Sci. Comput.*, 19(6):2061–2090 (electronic), 1998.

[18] S. Cox and C. Castro. Achieving arbitrarily large decay in the damped wave equation. *SIAM J. Control Optim.*, 39(6):1748–1755, 2001.

[19] S. Cox and E. Zuazua. The rate at which energy decays in a damped string. *Comm. Partial Differential Equations*, 19(1-2):213–243, 1994.

[20] S. Cox and E. Zuazua. The rate at which energy decays in a string damped at one end. *Indiana Univ. Math. J.*, 44(2):545–573, 1995.

[21] B. Engquist and A. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Math. Comp.*, 31(139):629–651, 1977.

[22] R. Glowinski. Ensuring well-posedness by analogy: Stokes problem and boundary control for the wave equation. *J. Comput. Phys.*, 103(2):189–221, 1992.

[23] P. Hébrard and A. Henrot. A spillover phenomenon in the optimal location of actuators. *SIAM J. Control Optim.*, 44(1):349–366 (electronic), 2005.

[24] L. I. Ignat and E. Zuazua. A two-grid approximation scheme for nonlinear Schrödinger equations: dispersive properties and convergence. *C. R. Math. Acad. Sci. Paris*, 341(6):381–386, 2005.

[25] J.A. Infante and E. Zuazua. Boundary observability for the space semi discretizations of the 1-d wave equation. *Math. Model. Num. Ann.*, 33:407–438, 1999.

[26] M. Lassas and E. Somersalo. On the existence and convergence of the solution of PML equations. *Computing*, 60(3):229–241, 1998.

[27] G. Lebeau. Équations des ondes amorties. *Séminaire sur les Équations aux Dérivées Partielles, 1993–1994,École Polytech.*, 1994.

[28] J.-L. Lions. *Contrôlabilité exacte, Stabilisation et Perturbations de Systèmes Distribués. Tome 1. Contrôlabilité exacte*, volume RMA 8. Masson, 1988.

[29] F. Macià. *Propagación y control de vibraciones en medios discretos y continuos.* PhD thesis, Universidad Complutense de Madrid, 2001.

[30] F. Macià. The effect of group velocity in the numerical analysis of control problems for the wave equation. In *Mathematical and numerical aspects of wave propagation—WAVES 2003*, pages 195–200. Springer, Berlin, 2003.

[31] P. G. Petropoulos. Reflectionless sponge layers as absorbing boundary conditions for the numerical solution of Maxwell equations in rectangular, cylindrical, and spherical coordinates. *SIAM J. Appl. Math.*, 60(3):1037–1058 (electronic), 2000.

[32] I. Singer and E. Turkel. A perfectly matched layer for the Helmholtz equation in a semi-infinite strip. *J. Comput. Phys.*, 201(2):439–465, 2004.

[33] L.R. Tcheugoué Tébou and E. Zuazua. Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity. *Numer. Math.*, 95(3):563–598, 2003.

[34] L. N. Trefethen. Group velocity in finite difference schemes. *SIAM Rev.*, 24(2):113–136, 1982.

[35] S. V. Tsynkov. Numerical solution of problems on unbounded domains. A review. *Appl. Numer. Math.*, 27(4):465–532, 1998. Absorbing boundary conditions.

[36] S. V. Tsynkov and E. Turkel. A Cartesian perfectly matched layer for the Helmholtz equation. In *Absorbing boundaries and layers, domain decomposition methods*, pages 279–309. Nova Sci. Publ., Huntington, NY, 2001.

[37] E. Turkel and A. Yefet. Absorbing PML boundary layers for wave-like equations. *Appl. Numer. Math.*, 27(4):533–557, 1998. Absorbing boundary conditions.

[38] R. M. Young. *An introduction to nonharmonic Fourier series.* Academic Press Inc., San Diego, CA, first edition, 2001.

[39] E. Zuazua. Boundary observability for the finite-difference space semi-discretizations of the 2-D wave equation in the square. *J. Math. Pures Appl. (9)*, 78(5):523–563, 1999.

Chapter 2

Observability properties of a semi-discrete 1d wave equation derived from a mixed finite element method on nonuniform meshes

Abstract: The goal of this article is to analyze the observability properties for a space semi-discrete approximation scheme derived from a mixed finite element method of the 1d wave equation on nonuniform meshes. More precisely, we prove that observability properties hold uniformly with respect to the mesh-size under some assumptions, which, roughly, measures the lack of uniformity of the meshes, thus extending the work [5] to nonuniform meshes. Our results are based on a precise description of the spectrum of the discrete approximation schemes on nonuniform meshes, and the use of Ingham's inequality. We also mention applications to the boundary null controllability of the 1d wave equation, and to stabilization properties for the 1d wave equation.

2.1 Introduction

The goal of this article is to address the observability properties for a semi-discrete 1d wave equation.

We consider the following 1d wave equation:

$$\begin{cases} \partial_{tt}^2 u - \partial_{xx}^2 u = 0, & (x,t) \in (0,1) \times \mathbb{R}, \\ u(0,t) = u(1,t) = 0, & t \in \mathbb{R}, \\ u(x,0) = u^0(x),\ \partial_t u(x,0) = u^1(x), & x \in (0,1), \end{cases} \tag{2.1.1}$$

where $u^0 \in H_0^1(0,1)$ and $u^1(x) \in L^2(0,1)$. The energy of solutions of (2.1.1), given by

$$E(t) = \frac{1}{2} \int_0^1 |\partial_t u(t,x)|^2 + |\partial_x u(t,x)|^2 \, dx, \tag{2.1.2}$$

is constant.

It is well-known (see [21]) that for all $T > 0$, there exists a constant K_T such that the admissibility inequality

$$\int_0^T |\partial_x u(0,t)|^2 \, dt \leq K_T E(0) \tag{2.1.3}$$

holds for any solution of (2.1.1) with $(u^0, u^1) \in H_0^1(0,1) \times L^2(0,1)$.

Besides, for any time $T > 2$, there exists a positive constant k_T such that the boundary observability inequality

$$k_T E(0) \leq \int_0^T |\partial_x u(0,t)|^2 \, dt \tag{2.1.4}$$

holds for any solution of (2.1.1) with $(u^0, u^1) \in H_0^1(0,1) \times L^2(0,1)$.

Inequalities (2.1.3)-(2.1.4) arise naturally when dealing with boundary controllability properties of the 1d wave equation, see [21]. Indeed, the observability and controllability properties are dual notions. We will clarify this relation in Section 2.3.

Let us also present another relevant observability inequality, which is useful when dealing with distributed controls or stabilization properties of damped wave equations (see [16, 21]). If (a, b) denotes a non empty subinterval of $(0,1)$, the following distributed observability property holds: for any time $T > 2\max\{a, 1-b\}$, there exists a constant C_1 such that any solution of (2.1.1) with initial data $(u^0, u^1) \in H_0^1(0,1) \times L^2(0,1)$ satisfies:

$$E(0) \leq C_1 \int_0^T \int_a^b |\partial_t u(x,t)|^2 \, dx \, dt. \tag{2.1.5}$$

In the sequel, we will consider observability properties for the 1d space semi-discrete wave equation derived from a mixed finite element method on a nonuniform mesh.

For any integer $n \in \mathbb{N}^*$, let us consider a mesh $\mathcal{S}_n$ given by $n+2$ points as:

$$0 = x_{0,n} < x_{1,n} < \cdots < x_{n,n} < x_{n+1,n} = 1, \qquad h_{j+1/2,n} = x_{j+1,n} - x_{j,n}, \quad j \in \{0, \cdots, n\}. \tag{2.1.6}$$

On $\mathcal{S}_n$, the mixed finite element approximation scheme for system (2.1.1) reads as (see [7], [15] or [5]):

$$\begin{cases} \dfrac{h_{j-1/2,n}}{4}(u''_{j-1,n} + u''_{j,n}) + \dfrac{h_{j+1/2,n}}{4}(u''_{j,n} + u''_{j+1,n}) \\ \qquad\qquad = \dfrac{u_{j+1,n} - u_{j,n}}{h_{j+1/2,n}} - \dfrac{u_{j,n} - u_{j-1,n}}{h_{j-1/2,n}}, \quad j = 1, \cdots n, \ t \in \mathbb{R}, \\ u_{0,n}(t) = u_{n+1,n}(t) = 0, \quad t \in \mathbb{R}, \\ u_j(0) = u^0_{j,n}, \qquad u'_j(0) = u^1_{j,n}, \qquad j = 1, \cdots, n. \end{cases} \tag{2.1.7}$$

The notations we use are the standard ones: A prime denotes differentiation with respect to time, and $u_{j,n}(t)$ is an approximation of the solution u of (2.1.1) at the point $x_{j,n}$ at time t.

System (2.1.7) is conservative. The energy of solutions u_n of (2.1.7), given by

$$E_n(t) = \frac{1}{2}\sum_{j=0}^{n} h_{j+1/2,n}\left(\frac{u_{j+1,n}(t) - u_{j,n}(t)}{h_{j+1/2,n}}\right)^2 + \frac{1}{2}\sum_{j=0}^{n} h_{j+1/2,n}\left(\frac{u'_{j+1,n}(t) + u'_{j,n}(t)}{2}\right)^2, \quad t \in \mathbb{R}, \tag{2.1.8}$$

is constant.

In this semi-discrete setting, we will investigate the observability properties corresponding to (2.1.4) and (2.1.5), and especially under which assumptions on the meshes $\mathcal{S}_n$ we can guarantee discrete observability inequalities to be uniform with respect to n.

For this purpose, we introduce the notion of regularity of a mesh:

Definition 2.1.1. For a mesh $\mathcal{S}_n$ given by $n+2$ points as in (2.1.6), we define the regularity of the mesh $\mathcal{S}_n$ by

$$\mathrm{Reg}(\mathcal{S}_n) = \frac{\max_j\{h_{j+1/2,n}\}}{\min_j\{h_{j+1/2,n}\}}. \tag{2.1.9}$$

Given $M \geq 1$, we say that a mesh $\mathcal{S}_n$ given by $n+2$ points as in (2.1.6) is M-regular if

$$\mathrm{Reg}(\mathcal{S}_n) = \frac{\max_j\{h_{j+1/2,n}\}}{\min_j\{h_{j+1/2,n}\}} \leq M. \tag{2.1.10}$$

Obviously, a 1-regular mesh is uniform. In other words, the regularity of the mesh $\mathrm{Reg}(\mathcal{S}_n)$ measures the lack of uniformity of the mesh.

Within this class, we will prove the following observability properties:

Theorem 2.1.2. *Let M be a real number greater than one, and consider a sequence $(\mathcal{S}_n)_n$ of M-regular meshes.*

Then for any time $T > 2$, there exist positive constants k_T and K_T such that for all integer n, any solution u_n of (2.1.7) *satisfies*

$$k_T E_n(0) \leq \int_0^T \Big(\Big| \frac{u_{1,n}(t)}{h_{1/2,n}} \Big|^2 + |u'_{1,n}(t)|^2 \Big) dt \leq K_T E_n(0). \tag{2.1.11}$$

Besides, if $J = (a,b) \subset (0,1)$ denotes a subinterval of $(0,1)$, then, for any time $T > 2$, there exists a constant C_1 such that for all integer n, any solution u_n of (2.1.7) *satisfies*

$$E_n(0) \leq C_1 \int_0^T \sum_{x_{j,n} \in J} h_{j+1/2,n} \Big(\frac{u'_{j,n}(t) + u'_{j+1,n}(t)}{2} \Big)^2 dt. \tag{2.1.12}$$

Obviously, these properties are discrete versions of inequalities (2.1.3),(2.1.4) and (2.1.5). Also note that the right hand-side inequality in (2.1.11) holds, as (2.1.3), for all time $T > 0$, taking $K_T = K_3$ for $T \leq 3$.

Theorem 2.1.2 is based on an explicit spectral analysis of (2.1.7) in the discrete setting, that proves the existence of a gap between the eigenvalues of the space discrete operator in (2.1.7). Thanks to Ingham's inequality [18], this reduces the analysis to the study of the observability properties of the eigenvectors of (2.1.7), which will again be deduced from the explicit form of the spectrum of (2.1.7).

Besides, we emphasize that Theorem 2.1.2 provides uniform (with respect to n) observability results. Therefore, as in the continuous setting, Theorem 2.1.2 has several applications to controllability and stabilization properties for the space semi-discrete 1d wave equations (2.1.7). In Section 2.3, similarly as in [5], using precisely the same duality as in the continuous case, we present an application to the boundary null controllability of the space semi-discrete approximation scheme of the 1d wave equation. Later, in Section 2.4, following [1], we study the decay properties of the energy for semi-discrete approximation schemes of 1d damped wave equations.

Let us briefly comment some relative works. Similar problems have been extensively studied in the last decade for various space semi-discrete approximation schemes of the 1d wave equation, see for instance the review article [32]. The numerical schemes on uniform meshes provided by finite difference and finite element methods do not have uniform observability properties, whatever the time T is (see [17]). This is due to high frequency waves that do not propagate, see [29, 22]. To be more precise, these numerical schemes create some spurious high-frequency wave solutions that are localized.

However some remedies exist. The most natural one consists in filtering the initial data and thus removing these spurious waves, as in [17, 31]. Another way to filter is to use the bi-grid method as introduced and developed in [14] and analyzed in [25]. A new approach was proposed recently in [24] based on wavelet filtering. Let us also mention the results [28, 27, 26, 11] that amounts to adding an extra term in (2.1.12) which is non-negligible only for the high frequencies. A last possible cure was proposed in [1, 15] and later analyzed in [5]: a 1d semi-discrete scheme derived from a mixed finite element method was proposed, which has the property that the group velocity of the waves is bounded from below. Also note that an extension of [5] to the 2d case in the square was proposed in [6].

To the best of our knowledge, there is no result at all for the space semi-discrete wave equation on nonuniform meshes, although most of the domains used in practice are recovered by non periodic triangulations. A first step in this direction can be found in [26], in which a study of a non homogeneous string equation on a uniform mesh was proposed. This can indeed be seen, up to a change of variable, as a discretization of a wave equation with constant velocity on a slightly nonuniform mesh.

Let us also mention that some results are available in the context of the heat equation for space semi-discrete approximation schemes on nonuniform meshes in [19], even in dimension greater than 1.

The outline of this paper is as follows. In Section 2.2, we precisely describe the spectrum of the space semi-discrete operator and prove Theorem 2.1.2. Sections 2.3 and 2.4 respectively aim at presenting precise applications of Theorem 2.1.2 to controllability and stabilization properties.

2.2 Spectral Theory

In this Section, we first study the spectrum of the space semi-discrete operator in (2.1.7) on a general mesh $\mathcal{S}_n$ given by $n+2$ points as in (2.1.6). Second, we derive more precise estimates on the spectrum when $\mathcal{S}_n$ is an M-regular mesh. Third, we derive Theorem 2.1.2 from our analysis. Finally, we discuss the assumption on the regularity of the meshes, and show that, in some sense, the M-regularity assumption is sharp with respect to the observability properties given in Theorem 2.1.2.

Given a mesh $\mathcal{S}_n$ of $n+2$ points as in (2.1.6), since the system (2.1.7) is conservative, the spectral problem for (2.1.7) reads as: Find $\lambda_n \in \mathbb{R}$ and a non-trivial solution ϕ_n such that

$$\begin{cases} -\dfrac{\lambda_n^2}{4}(h_{j-1/2,n}(\phi_{j,n}+\phi_{j-1,n})+h_{j+1/2,n}(\phi_{j,n}+\phi_{j+1,n})) \\ \qquad\qquad = \dfrac{\phi_{j+1,n}-\phi_{j,n}}{h_{j+1/2,n}} - \dfrac{\phi_{j,n}-\phi_{j-1,n}}{h_{j-1/2,n}}, \qquad j=1,\cdots,n, \\ \phi_{0,n}=\phi_{n+1,n}=0. \end{cases} \tag{2.2.1}$$

2.2.1 Computations of the eigenvalues for a general mesh

In this Subsection, we consider a general mesh $\mathcal{S}_n$ given by $n+2$ points as in (2.1.6).

Theorem 2.2.1. *The spectrum of system* (2.1.7) *is precisely the set of* $\pm\lambda_n^k$ *with* $k \in \{1, \cdots, n\}$, *where* λ_n^k *is defined by the implicit formula*

$$\sum_{j=0}^{n} \arctan\Big(\frac{\lambda_n^k h_{j+1/2,n}}{2}\Big) = \frac{k\pi}{2}. \tag{2.2.2}$$

The gap between two eigenvalues is bounded from below:

$$\min_{k\in\{1,\cdots,n-1\}} \{\lambda_n^{k+1} - \lambda_n^k\} \geq \pi. \tag{2.2.3}$$

Besides, for each $k \in \{1, \cdots, n\}$, *the following estimate holds:*

$$\lambda_n^k \geq \lambda_{*n}^k = 2(n+1)\tan\Big(\frac{k}{n+1}\frac{\pi}{2}\Big) \geq k\pi. \tag{2.2.4}$$

Remark 2.2.2. Note that λ_{*n}^k coincides with the k-th eigenvalue of system (2.1.7) for a uniform mesh constituted by $n+2$ points. Also note that $k\pi$ is the k-th eigenvalue of system (2.1.1). In other words, inequality (2.2.4) implies that the dispersion diagrams corresponding to the spectrum of (2.1.7) for a general nonuniform mesh, for a uniform mesh, and for the continuous system (2.1.1) are sorted.

Proof. To simplify notation, we drop the subscript n.

Let us introduce functions p and q corresponding to $\partial_x\phi$ and $i\lambda\phi$ in the continuous case:

$$p_{j+1/2} = \frac{\phi_{j+1} - \phi_j}{h_{j+1/2}}, \qquad q_{j+1/2} = \frac{i\lambda}{2}(\phi_j + \phi_{j+1}), \qquad j \in \{0, \cdots, n\}. \tag{2.2.5}$$

The spectral system (2.2.1) then becomes :

$$\begin{cases} \dfrac{i\lambda}{2}(h_{j-1/2}\, q_{j-1/2} + h_{j+1/2}\, q_{j+1/2}) = p_{j+1/2} - p_{j-1/2}, & j = 1, \cdots, n, \\ \dfrac{i\lambda}{2}(h_{j-1/2}\, p_{j-1/2} + h_{j+1/2}\, p_{j+1/2}) = q_{j+1/2} - q_{j-1/2}, & j = 1, \cdots, n, \end{cases} \tag{2.2.6}$$

with boundary conditions

$$\frac{i\lambda h_{n+1/2}}{2} p_{n+1/2} + q_{n+1/2} = 0, \qquad \frac{i\lambda h_{1/2}}{2} p_{1/2} - q_{1/2} = 0.$$

Equations (2.2.6) rewrite, for $j \in \{1, \cdots, n\}$, as:

$$\begin{cases} \Big(\dfrac{i\lambda h_{j-1/2}}{2} q_{j-1/2} + p_{j-1/2}\Big) + \Big(\dfrac{i\lambda h_{j+1/2}}{2} q_{j+1/2} - p_{j+1/2}\Big) = 0, \\ \Big(\dfrac{i\lambda h_{j-1/2}}{2} p_{j-1/2} + q_{j-1/2}\Big) + \Big(\dfrac{i\lambda h_{j+1/2}}{2} p_{j+1/2} - q_{j+1/2}\Big) = 0, \end{cases} \tag{2.2.7}$$

For $j \in \{1, \cdots, n\}$, this leads to:

$$\begin{aligned} \Big(1 + \frac{i\lambda h_{j-1/2}}{2}\Big)(p_{j-1/2} + q_{j-1/2}) &= \Big(1 - \frac{i\lambda h_{j+1/2}}{2}\Big)(p_{j+1/2} + q_{j+1/2}) \\ \Big(1 - \frac{i\lambda h_{j-1/2}}{2}\Big)(p_{j-1/2} - q_{j-1/2}) &= \Big(1 + \frac{i\lambda h_{j+1/2}}{2}\Big)(p_{j+1/2} - q_{j+1/2}). \end{aligned}$$

These two equations can be seen as propagation formulas, each term corresponding to $\partial_t w \pm \partial_x w$. Especially, they imply:

$$p_{j+1/2} + q_{j+1/2} = (p_{1/2} + q_{1/2})\Big(\frac{2 + i\lambda h_{1/2}}{2 - i\lambda h_{j+1/2}}\Big) \prod_{k=1}^{j-1} \Big(\frac{2 + i\lambda h_{k+1/2}}{2 - i\lambda h_{k+1/2}}\Big), \tag{2.2.8}$$

$$p_{j+1/2} - q_{j+1/2} = (p_{1/2} - q_{1/2})\Big(\frac{2 - i\lambda h_{1/2}}{2 + i\lambda h_{j+1/2}}\Big) \prod_{k=1}^{j-1} \Big(\frac{2 - i\lambda h_{k+1/2}}{2 + i\lambda h_{k+1/2}}\Big). \tag{2.2.9}$$

We remark that each term in the product has modulus 1, and therefore there exists $\alpha_{j+1/2} \in (-\pi, \pi]$, given by $\tan(\alpha_{j+1/2}/2) = \lambda h_{j+1/2}/2$, such that :

$$\frac{2 + i\lambda h_{j+1/2}}{2 - i\lambda h_{j+1/2}} = \exp(i\alpha_{j+1/2}).$$

We also denote by β_j the coefficient

$$\beta_j = \frac{2 + i\lambda h_{1/2}}{2 - i\lambda h_{j+1/2}},$$

which satisfies

$$\frac{\beta_j}{\bar{\beta}_j} = \exp(i\alpha_{j+1/2}) \exp(i\alpha_{1/2}).$$

Combined with the boundary conditions, identities (2.2.8)-(2.2.9) give:

$$\begin{aligned} p_{n+1/2}\Big(1 - \frac{i\lambda h_{n+1/2}}{2}\Big) &= \beta_n \exp\Big(i \sum_{k=1}^{n-1} \alpha_{k+1/2}\Big) p_{1/2}\Big(1 + \frac{i\lambda h_{1/2}}{2}\Big) \\ p_{n+1/2}\Big(1 + \frac{i\lambda h_{n+1/2}}{2}\Big) &= \bar{\beta}_n \exp\Big(-i \sum_{k=1}^{n-1} \alpha_{k+1/2}\Big) p_{1/2}\Big(1 - \frac{i\lambda h_{1/2}}{2}\Big). \end{aligned}$$

Then, if λ is an eigenvalue, λ satisfies:

$$\Big(\frac{\beta_n}{\bar{\beta}_n}\Big)^2 \exp\Big(2i \sum_{k=1}^{n-1} \alpha_{k+1/2}\Big) = \exp\Big(2i \sum_{k=0}^{n} \alpha_{k+1/2}\Big) = 1. \tag{2.2.10}$$

To simplify notation, we define:

$$f(\lambda) = 4 \sum_{k=0}^{n} \arctan\Big(\frac{\lambda h_{k+1/2}}{2}\Big).$$

Due to (2.2.10), if λ is an eigenvalue, there exists an integer k such that:

$$f(\lambda) = 2k\pi.$$

The image of f is exactly $(-2(n+1)\pi, 2(n+1)\pi)$, and therefore k must belong to $\{-n, \cdots, n\}$.

Conversely, if λ is a solution of $f(\lambda) = 2k\pi$ for an integer $k \in \{-n, \cdots, n\}$, then λ is an eigenvalue, except if $k = 0$, which corresponds to $p_{j+1/2} = q_{j+1/2} = 0$ for all $j \in \{0, \cdots, n\}$. This gives us exactly $2n$ eigenvalues $\pm\lambda^k$, $k \in \{1, \cdots, n\}$.

Moreover, the derivative of f is explicit:

$$f'(\lambda) = 8 \sum_{k=0}^{n} \frac{1}{4 + (\lambda h_{k+1/2})^2} h_{k+1/2}.$$

It follows that

$$0 \leq f'(\lambda) \leq 2 \sum_{k=0}^{n} h_{k+1/2} = 2.$$

Since all the eigenvalues are simple and $f(\lambda^{k+1}) - f(\lambda^k) = 2\pi$ for all $k \in \{1, \cdots, n-1\}$, this implies that the gap between the eigenvalues is bounded from below by π, and therefore (2.2.3) holds.

Using the concavity of arctan gives the following estimate:

$$\arctan\left(\frac{\lambda^k}{2(n+1)}\right) = \arctan\left(\frac{1}{2(n+1)} \sum_{j=0}^{n} \lambda^k h_{j+1/2}\right) \geq \frac{1}{n+1} \sum_{j=0}^{n} \arctan\left(\frac{\lambda^k h_{j+1/2}}{2}\right) = \frac{k}{n+1} \frac{\pi}{2}.$$

In other words,

$$\lambda^k \geq 2(n+1) \tan\left(\frac{k}{n+1} \frac{\pi}{2}\right),$$

and (2.2.4) follows. Indeed, the right hand-side inequality in (2.2.4) simply follows from the standard inequality $\tan(\eta) \geq \eta$ for $\eta \in [0, \pi/2)$. □

We illustrate this result on Figures 2.1-2.2 by computing dispersion diagrams for various nonuniform meshes $\mathcal{S}_n$, that we characterize by their regularity $\mathrm{Reg}(\mathcal{S}_n)$, as defined in (2.1.9).

Let us briefly explain the two ways we have chosen for generating them.

- **Method 1.** In Figure 2.1, we create a random vector h of length $n+1$ whose values are chosen according to a uniform law on $(0, 1)$. This vector is then normalized such that the sum of its components is one, so that h corresponds to the vector $(h_{1/2,n}, \cdots, h_{n+1/2,n})$, which describes the mesh in a unique way.
- **Method 2.** In Figure 2.2, we create a random vector x of length n whose components are chosen according to a uniform law on $(0, 1)$. Then we sort its components in an increasing way to obtain a vector $(x_{1,n}, \cdots, x_{n,n})$, which represents the mesh points.

In both cases, the dispersion diagrams look the same. It is particularly striking that the shape of the dispersion diagrams does not seem to depend significantly on the meshes.

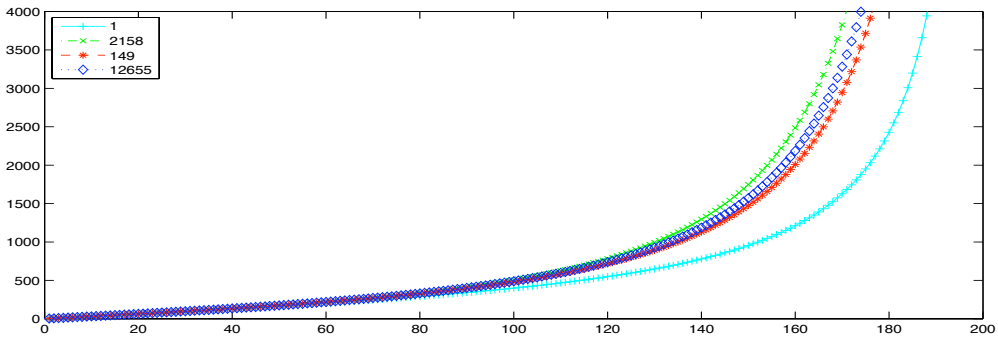

Figure 2.1: Dispersion diagrams for various meshes constituted by 200 points generated by Method 1 for different values of Reg.

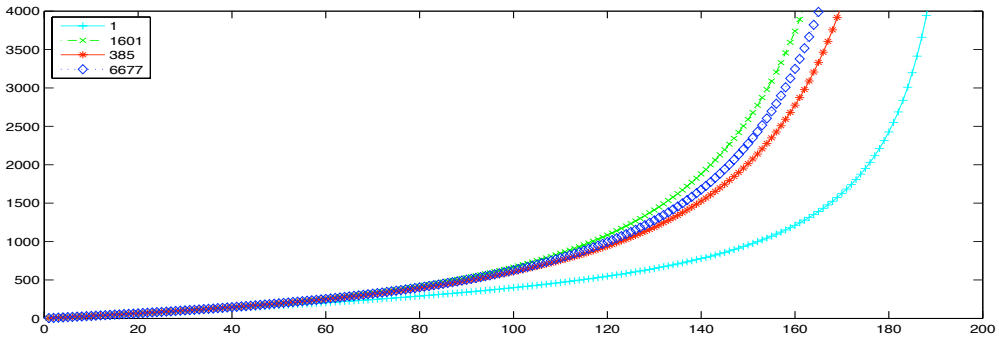

Figure 2.2: Dispersion diagrams for various meshes constituted by 200 points generated by Method 2 for different values of Reg.

2.2.2 Spectral properties on M-regular meshes

This subsection is devoted to prove additional properties for the spectrum of (2.1.7) when the mesh $\mathcal{S}_n$ is M-regular for some $M \geq 1$.

Theorem 2.2.3. *Let $M \geq 1$.*

Then, for any M-regular mesh $\mathcal{S}_n$, the eigenvalue λ_n^n of (2.2.1) on $\mathcal{S}_n$ satisfies

$$\lambda_n^n \leq \frac{4M}{\pi}(n+1)^2. \tag{2.2.11}$$

Besides, for any M-regular mesh $\mathcal{S}_n$, if ϕ_n^k denotes the eigenvector corresponding to λ_n^k in (2.2.1), then its energy

$$E_n^k = \frac{1}{2}\sum_{j=0}^{n} h_{j+1/2,n}\Big(\Big|\frac{\phi_{j+1,n}^k - \phi_{j,n}^k}{h_{j+1/2,n}}\Big|^2 + |\lambda_n^k|^2\Big|\frac{\phi_{j,n}^k + \phi_{j+1,n}^k}{2}\Big|^2\Big) \tag{2.2.12}$$

satisfies

$$\frac{1}{1+M^2}\Big(\Big|\frac{\phi_{1,n}^k}{h_{1/2,n}}\Big|^2 + \frac{h_{1/2,n}^2}{4}\Big|\frac{\lambda_n^k\phi_{1,n}^k}{h_{1/2,n}}\Big|^2\Big) \leq E_n^k \leq (1+M^2)\Big(\Big|\frac{\phi_{1,n}^k}{h_{1/2,n}}\Big|^2 + \frac{h_{1/2,n}^2}{4}\Big|\frac{\lambda_n^k\phi_{1,n}^k}{h_{1/2,n}}\Big|^2\Big), \tag{2.2.13}$$

Moreover, if $\omega = (a,b)$ is some subinterval of $(0,1)$, then the energy of the k-th eigenvector ϕ_n^k in ω, defined by

$$E_{\omega,n}^k = \frac{1}{2}\sum_{x_{j,n}\in\omega} h_{j+1/2,n}\Big(\Big|\frac{\phi_{j+1,n}^k - \phi_{j,n}^k}{h_{j+1/2,n}}\Big|^2 + |\lambda_n^k|^2\Big|\frac{\phi_{j,n}^k + \phi_{j+1,n}^k}{2}\Big|^2\Big), \tag{2.2.14}$$

satisfies

$$E_n^k \leq \frac{M^2}{|\omega|}E_{\omega,n}^k. \tag{2.2.15}$$

Remark 2.2.4. These inequalities roughly say that the eigenvectors cannot concentrate in some part of an M-regular mesh. These properties are indeed the one needed for control and stabilization purposes, as we will see in next Sections.

Remark 2.2.5. Note that Theorem 2.2.1 gives the estimate

$$\lambda_n^n \geq 2(n+1)\tan\Big(\Big(1-\frac{1}{n+1}\Big)\frac{\pi}{2}\Big) \underset{n\to\infty}{\simeq} \frac{4}{\pi}(n+1)^2.$$

Combined with estimate (2.2.11), this indicates that, when considering sequences of M-regular meshes, the eigenvalues λ_n^n really grow as n^2 when $n \to \infty$.

Proof. Along the proof, we fix an integer n, a real number $M \geq 1$ and an M-regular mesh $\mathcal{S}_n$, so that we can remove the index n without confusion.

Inequality (2.2.11) is a consequence of (2.2.2). Indeed, if we set $h = \min\{h_{j+1/2}\}$ and $H = \max\{h_{j+1/2}\}$, then we have

$$1 \leq (n+1)H \leq (n+1)Mh. \tag{2.2.16}$$

Besides, using (2.2.2), we get

$$\sum_{j=0}^{n} \arctan\Big(\frac{\lambda^n h_{j+1/2}}{2}\Big) = \frac{n\pi}{2} \geq (n+1)\arctan\Big(\frac{\lambda^n h}{2}\Big),$$

which provides

$$\frac{\lambda^n}{(n+1)^2} \leq \frac{2}{h(n+1)^2}\tan\Big(\frac{\pi}{2}\Big(1-\frac{1}{n+1}\Big)\Big) \leq M \sup_{\eta\in[0,1]}\Big\{2\eta\tan\Big(\frac{\pi}{2}(1-\eta)\Big)\Big\},$$

from which (2.2.13) follows.

To derive the properties (2.2.13) and (2.2.15) of the eigenvectors, we use the computations and notations (2.2.5) introduced in the proof of Theorem 2.2.1. Namely, we introduce:

$$p_{j+1/2}^k = \frac{\phi_{j+1}^k - \phi_j^k}{h_{j+1/2}}, \qquad q_{j+1/2}^k = \frac{i\lambda^k}{2}(\phi_j^k + \phi_{j+1}^k), \qquad j \in \{0,\cdots,n\}.$$

Then the previous computations, and in particular identities (2.2.8)-(2.2.9), give:

$$\begin{aligned}
E^k &= \frac{1}{2}\sum_{j=0}^{n} h_{j+1/2}\Big(|p_{j+1/2}^k|^2 + |q_{j+1/2}^k|^2\Big)\\
&= \frac{1}{4}\sum_{j=0}^{n} h_{j+1/2}\Big(|p_{j+1/2}^k - q_{j+1/2}^k|^2 + |p_{j+1/2}^k + q_{j+1/2}^k|^2\Big)\\
&= \frac{1}{4}\sum_{j=0}^{n} h_{j+1/2}\Big(|\bar\beta_j|^2|p_{1/2}^k - q_{1/2}^k|^2 + |\beta_j|^2|p_{1/2}^k + q_{1/2}^k|^2\Big)\\
&= \frac{1}{4}\sum_{j=0}^{n} h_{j+1/2}\frac{4+(\lambda h_{1/2})^2}{4+(\lambda h_{j+1/2})^2}\Big(|p_{1/2}^k - q_{1/2}^k|^2 + |p_{1/2}^k + q_{1/2}^k|^2\Big).
\end{aligned}$$

Using the definition (2.2.5) of $(p_{1/2}^k, q_{1/2}^k)$, this leads to

$$E^k = \frac{1}{2}\left(\sum_{j=0}^{n}\frac{h_{j+1/2}}{4+(\lambda^k h_{j+1/2})^2}\right)\Big(4+(\lambda^k h_{1/2})^2\Big)\Big(\Big|\frac{\phi_1^k}{h_{1/2}}\Big|^2 + \frac{h_{1/2}^2}{4}\Big|\frac{\lambda^k\phi_1^k}{h_{1/2}}\Big|^2\Big). \tag{2.2.17}$$

Given an interval ω, the same computations give for E_ω^k :

$$E_\omega^k = \frac{1}{2}\left(\sum_{x_j\in\omega}\frac{h_{j+1/2}}{4+(\lambda^k h_{j+1/2})^2}\right)\left(4+(\lambda^k h_{1/2})^2\right)\left(\Big|\frac{\phi_1^k}{h_{1/2}}\Big|^2+\frac{h_{1/2}^2}{4}\Big|\frac{\lambda^k\phi_1^k}{h_{1/2}}\Big|^2\right). \tag{2.2.18}$$

Inequalities (2.2.13) and (2.2.15) easily follow from (2.2.17)-(2.2.18) and the M-regularity assumption. □

2.2.3 Proof of Theorem 2.1.2

Our strategy is based on Ingham's Lemma on non-harmonic Fourier series, which we recall hereafter (see [18, 30]):

Lemma 2.2.6 (Ingham's Lemma)**.** *Let $(\lambda_k)_{k\in\mathbb{N}}$ be an increasing sequence of real numbers and $\gamma>0$ be such that*

$$\lambda_{k+1}-\lambda_k\geq\gamma>0,\ \forall k\in\mathbb{N}. \tag{2.2.19}$$

Then, for any $T>2\pi/\gamma$, there exist two positive constants $c=c(T,\gamma)>0$ and $C=C(T,\gamma)>0$ such that, for any sequence $(a_k)_{k\in\mathbb{N}}$,

$$c\sum_{k\in\mathbb{N}}|a_k|^2\leq\int_0^T\Big|\sum_{k\in\mathbb{N}}a_k e^{i\lambda_k t}\Big|^2\,dt\leq C\sum_{k\in\mathbb{N}}|a_k|^2. \tag{2.2.20}$$

Proof of Theorem 2.1.2. Let us consider a sequence $(\mathcal{S}_n)_n$ of M-regular meshes.

According to inequality (2.2.3), the gap condition (2.2.19) holds with $\gamma=\pi$. Thus, due to Lemma 2.2.6, we only need to prove the observability inequalities (2.1.11)-(2.1.12) for the stationnary solutions

$$u_n^k(t)=\exp(i\lambda_n^k t)\phi_n^k$$

of (2.1.7) corresponding to the eigenvectors ϕ_n^k of system (2.2.1) on $\mathcal{S}_n$.

Since each mesh $\mathcal{S}_n$ is M-regular, we can apply Theorem 2.2.3. Especially, inequality (2.2.13) holds, and therefore Ingham's inequality (2.2.20) directly implies (2.1.11).

To prove (2.1.12), we fix $J=(a,b)\subset(0,1)$ a subinterval of $(0,1)$. According to Ingham's Lemma and (2.2.3), it is sufficient to prove that there exists a constant C independent of n and k such that, for any eigenvector ϕ_n^k solution of (2.2.1) on $\mathcal{S}_n$ corresponding to the eigenvalue λ_n^k, the quantity

$$I_{J,n}^k=\sum_{x_{j,n}\in J}h_{j+1/2,n}|\lambda_n^k|^2\Big(\frac{\phi_{j,n}^k+\phi_{j+1,n}^k}{2}\Big)^2 \tag{2.2.21}$$

satisfies

$$E_n^k\leq CI_{J,n}^k. \tag{2.2.22}$$

We thus investigate inequality (2.2.22) on a mesh $\mathcal{S}_n$ by using a multiplier technique.

Let ω be a strict subinterval of J and let us denote by η a function of $x\in[0,1]$ such that:

$$\begin{cases}\eta(x)=0, & \forall x\in(0,1)\backslash J,\\ \eta(x)=1, & \forall x\in\omega,\end{cases}\qquad\begin{cases}\|\eta\|_\infty\leq 1,\\ \|\eta'\|_\infty\leq C_{J,\omega}.\end{cases} \tag{2.2.23}$$

To simplify notation, we drop the exponent k and the index n hereafter. Below, we denote by η_j the value of η in the mesh point x_j.

We consider system (2.2.1) and multiply each equation by $\eta_j^2\phi_j$. Discrete integrations by parts yield:

$$\lambda^2\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)\Big(\frac{\eta_j^2\phi_j+\eta_{j+1}^2\phi_{j+1}}{2}\Big)=\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\phi_{j+1}-\phi_j}{h_{j+1/2}}\Big)\Big(\frac{\eta_{j+1}^2\phi_{j+1}-\eta_j^2\phi_j}{h_{j+1/2}}\Big).$$

Then we deduce that

$$\lambda^2\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\eta_j^2+\eta_{j+1}^2}{2}\Big)\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)^2-\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\eta_j^2+\eta_{j+1}^2}{2}\Big)\Big(\frac{\phi_{j+1}-\phi_j}{h_{j+1/2}}\Big)^2=A_1+A_2,\quad(2.2.24)$$

where A_1 and A_2 are defined by

$$\begin{aligned}A_1&=-\frac{\lambda^2}{2}\sum_{j=0}^{n}h_{j+1/2}^3\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)\Big(\frac{\phi_{j+1}-\phi_j}{h_{j+1/2}}\Big)\Big(\frac{\eta_{j+1}-\eta_j}{h_{j+1/2}}\Big)\Big(\frac{\eta_j+\eta_{j+1}}{2}\Big),\\A_2&=2\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)\Big(\frac{\phi_{j+1}-\phi_j}{h_{j+1/2}}\Big)\Big(\frac{\eta_{j+1}-\eta_j}{h_{j+1/2}}\Big)\Big(\frac{\eta_j+\eta_{j+1}}{2}\Big).\end{aligned}$$

Then, for any choices of positive parameters δ_1 and δ_2, we get:

$$\begin{aligned}|A_1|&\le\frac{1}{4\delta_1}\sum_{j=0}^{n}h_{j+1/2}\lambda^2\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)^2\Big(\frac{\eta_{j+1}-\eta_j}{h_{j+1/2}}\Big)^2\\&\qquad+\frac{\delta_1}{4}\sum_{j=0}^{n}h_{j+1/2}(\lambda^2h_{j+1/2}^4)\Big(\frac{\phi_{j+1}-\phi_j}{h_{j+1/2}}\Big)^2\Big(\frac{\eta_j+\eta_{j+1}}{2}\Big)^2,\\|A_2|&\le\frac{1}{\delta_2}\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)^2\Big(\frac{\eta_{j+1}-\eta_j}{h_{j+1/2}}\Big)^2+\delta_2\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\phi_{j+1}-\phi_j}{h_{j+1/2}}\Big)^2\Big(\frac{\eta_j+\eta_{j+1}}{2}\Big)^2.\end{aligned}$$

Using that

$$\Big(\frac{n+1}{M}\Big)\sup h_{j+1/2}\le(n+1)\inf h_{j+1/2}\le1$$

estimate (2.2.11) gives

$$\lambda^2h_{j+1/2}^4\le\Big(\frac{4M}{\pi}(n+1)^2\Big)^2\Big(\frac{M}{(n+1)}\Big)^4\le\Big(\frac{4}{\pi}\Big)^2M^4.$$

Therefore, if we set

$$\delta_1=\frac{\pi^2}{16M^4}\quad;\quad\delta_2=\frac{1}{4},$$

using the classical inequality

$$\Big(\frac{\eta_j+\eta_{j+1}}{2}\Big)^2\le\frac{\eta_j^2+\eta_{j+1}^2}{2}.$$

we deduce from (2.2.24) the existence of two constants independent of k and n such that

$$\begin{aligned}\frac{1}{2}\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\eta_j^2+\eta_{j+1}^2}{2}\Big)\Big(\frac{\phi_{j+1}-\phi_j}{h_{j+1/2}}\Big)^2&\le\lambda^2\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\eta_j^2+\eta_{j+1}^2}{2}\Big)\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)^2\\&+C_1\sum_{j=0}^{n}h_{j+1/2}\lambda^2\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)^2\Big(\frac{\eta_{j+1}-\eta_j}{h_{j+1/2}}\Big)^2+C_2\sum_{j=0}^{n}h_{j+1/2}\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)^2\Big(\frac{\eta_{j+1}-\eta_j}{h_{j+1/2}}\Big)^2.\end{aligned}$$

But $|\lambda|$ is also uniformly bounded from below (see (2.2.4)), and therefore we obtain that

$$\sum_{j=0}^{n} h_{j+1/2}\Big(\frac{\eta_j^2+\eta_{j+1}^2}{2}\Big)\Big(\frac{\phi_{j+1}-\phi_j}{h_{j+1/2}}\Big)^2 \leq \lambda^2 \sum_{j=0}^{n} h_{j+1/2}\Big(\frac{\eta_j^2+\eta_{j+1}^2}{2}\Big)\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)^2 + C\sum_{j=0}^{n} h_{j+1/2}\lambda^2\Big(\frac{\phi_j+\phi_{j+1}}{2}\Big)^2\Big(\frac{\eta_{j+1}-\eta_j}{h_{j+1/2}}\Big)^2.$$

Using the properties (2.2.23) of the function η leads us to the following result:

$$E_{\omega,n}^k \leq C I_{J,n}^k.$$

Therefore inequality (2.2.22) can be deduced from inequality (2.2.15) applied to ω. □

2.2.4 The regularity assumption

Let us discuss the assumption on the regularity on the meshes.

Concentration effects without the M-regularity assumption

Here, we design a sequence of meshes $\mathcal{S}_n$ such that:

- The sequence $\mathrm{Reg}(\mathcal{S}_n)$ goes to infinity arbitrarily slowly when $n \to \infty$.
- There exists an interval $J = [a, b]$ for which there is no constant C such that for all n, for all eigenvectors ϕ_n^k of (2.2.1) on $\mathcal{S}_n$,
$$E_n^k \leq C E_{J,n}^k, \tag{2.2.25}$$
where E_n^k and $E_{J,n}^k$ are, respectively, as in (2.2.12) and (2.2.14).

Note that (2.2.25) constitutes an obstruction for (2.1.12) to hold.

Choose a strict non-empty closed subinterval J of $(0,1)$, and a sequence K_n going to infinity when $n \to \infty$. Introduce a sequence of meshes $(\mathcal{S}_n)$, each one constituted by $n+2$ points such that

$$x_{0,n} = 0, \quad x_{n+1,n} = 1, \quad \begin{cases} x_{j+1,n} - x_{j,n} = H_n, & \text{if } [x_{j,n}, x_{j+1,n}] \subset J, \\ x_{j+1,n} - x_{j,n} = h_n, & \text{if } [x_{j,n}, x_{j+1,n}] \subset [0,1]\backslash J, \end{cases}$$

where $H_n = K_n h_n$. Remark that the mesh $\mathcal{S}_n$ is then totally described by the quantity K_n. From identities (2.2.17)-(2.2.18), we get:

$$\frac{E_n^k}{E_{J,n}^k} = 1 + \frac{E_{(0,1)\backslash J,n}^k}{E_{J,n}^k} = 1 + \frac{1-|J|}{|J|}\ \frac{4+(\lambda_n^k H_n)^2}{4+(\lambda_n^k h_n)^2}.$$

But

$$\frac{|J|}{H_n} + \frac{1-|J|}{h_n} = n+1,$$

and so $(n+1)h_n = (1-|J|) + |J|/K_n$ converges to $1-|J|$ when $n \to \infty$. But inequality (2.2.4) gives

$$\frac{\lambda_n^n h_n}{2} \geq (n+1)h_n\ \tan\Big(\frac{n}{n+1}\ \frac{\pi}{2}\Big),$$

and then $(\lambda_n^n h_n)_n$ goes to infinity when $n \to \infty$. Especially, this implies that

$$\frac{E_n^n}{E_{J,n}^n} \underset{n\to\infty}{\simeq} \frac{1-|J|}{|J|}\,\frac{H_n^2}{h_n^2} = \frac{1-|J|}{|J|}\,K_n^2 \to \infty,$$

and therefore there is no constant such that (2.2.25) holds uniformly with respect to $n \in \mathbb{N}$ and $k \in \{1, \cdots, n\}$.

Partial regularity assumptions

Without the M-regularity assumption, one can derive partial results, due to the explicit form (2.2.17) of the energy.

For instance, identity (2.2.17) on the energy of the k-th eigenvector ϕ_n^k on $\mathcal{S}_n$ gives:

$$E_n^k \le \frac{4+(\lambda_n^k h_{1/2,n})^2}{4+\inf_j(\lambda_n^k h_{j+1/2,n})^2}\Big(\Big|\frac{\phi_{1,n}^k}{h_{1/2,n}}\Big|^2 + \frac{h_{1/2,n}^2}{4}\Big|\frac{\lambda_n^k\phi_{1,n}^k}{h_{1/2,n}}\Big|^2\Big).$$

In particular, if there exists a constant $M_1 > 0$ such that for all n,

$$h_{1/2,n} \le M_1 \inf_j h_{j+1/2,n}, \tag{2.2.26}$$

then for all n and k,

$$E_n^k \le (1+M_1^2)\Big(\Big|\frac{\phi_{1,n}^k}{h_{1/2,n}}\Big|^2 + \frac{h_{1/2,n}^2}{4}\Big|\frac{\lambda_n^k\phi_{1,n}^k}{h_{1/2,n}}\Big|^2\Big).$$

Now, consider the reverse equality. From (2.2.17), we get

$$E_n^k \ge \frac{4+(\lambda_n^k h_{1/2,n})^2}{4+\sup_j(\lambda_n^k h_{j+1/2,n})^2}\Big(\Big|\frac{\phi_{1,n}^k}{h_{1/2,n}}\Big|^2 + \frac{h_{1/2,n}^2}{4}\Big|\frac{\lambda_n^k\phi_{1,n}^k}{h_{1/2,n}}\Big|^2\Big).$$

In particular, if there exists a constant $M_2 > 0$ such that for all n,

$$\sup_j h_{j+1/2,n} \le M_2 h_{1/2,n}, \tag{2.2.27}$$

then, for all n and k, we get

$$E_n^k \ge \frac{1}{1+M_2^2}\Big(\Big|\frac{\phi_{1,n}^k}{h_{1/2,n}}\Big|^2 + \frac{h_{1/2,n}^2}{4}\Big|\frac{\lambda_n^k\phi_{1,n}^k}{h_{1/2,n}}\Big|^2\Big).$$

Besides, as in Subsubsection 2.2.4, for each integer n, we can consider sequences of meshes $\mathcal{S}_n$ given as in (2.1.6) defined by

$$x_{1,n} - x_{0,n} = h_{1/2,n}, \qquad x_{j+1,n} - x_{j,n} = h_n, \quad \forall j \in \{1, \cdots, n\},$$

where $h_{1/2,n}$ and h_n are two sequences going to zero. It is then easy to check that if condition (2.2.27) is not satisfied, that is if $h_n/h_{1/2,n} \to \infty$ when $n \to \infty$, then there is no positive constant c such that

$$E_n^k \ge c\Big(\Big|\frac{\phi_{1,n}^k}{h_{1/2,n}}\Big|^2 + \frac{h_{1/2,n}^2}{4}\Big|\frac{\lambda_n^k\phi_{1,n}^k}{h_{1/2,n}}\Big|^2\Big)$$

uniformly in k and n.

On the contrary, if $h_n/h_{1/2,n} \to 0$ when $n \to \infty$, then there is no constant C such that

$$E_n^k \leq C\Big(\Big|\frac{\phi_{1,n}^k}{h_{1/2,n}}\Big|^2 + \frac{h_{1/2,n}^2}{4}\Big|\frac{\lambda_n^k \phi_{1,n}^k}{h_{1/2,n}}\Big|^2\Big)$$

uniformly in k and n.

Therefore, if we consider a sequence of meshes $\mathcal{S}_n$ such that $\mathrm{Reg}(\mathcal{S}_n)$ is unbounded, we cannot expect in general to have both observability and admissibility properties (2.1.11) uniformly with respect to n.

Remark 2.2.7. If we are interested in the observability inequality (2.1.12) for a particular subinterval $(a,b) \subset (0,1)$, the situation is more intricate. As above, due to the explicit description of the energies (2.2.17) and (2.2.18), one easily check that if there exists a constant M_3 such that for all $n \in \mathbb{N}$,

$$\sup_{x_{j,n} \in (a,b)} \{h_{j+1/2,n}\} \leq M_3 \inf_{x_{j,n} \notin (a,b)} \{h_{j+1/2,n}\}, \tag{2.2.28}$$

then for all $n \in \mathbb{N}$ and for all $k \in \{1, \cdots, n\}$,

$$E_{(a,b),n}^k \leq \frac{M_3^2}{(b-a)} E_n^k.$$

However, under the only condition (2.2.28), the estimates (2.2.11) on the eigenvalues might be false, and therefore the proof presented above of inequality (2.2.22) (with $J = (a,b)$) fails. We do not know if assumption (2.2.28) suffices to guarantee (2.2.22) to hold uniformly with respect to $n \in \mathbb{N}$ and $k \in \{1, \cdots, n\}$.

Also remark that if assumption (2.2.28) holds for a sequence of meshes $\mathcal{S}_n$ for any subinterval $(a,b) \subset (0,1)$, then there exists a real number M such that all the meshes $\mathcal{S}_n$ are M-regular.

2.3 Application to the null controllability of the wave equation

2.3.1 The continuous setting

Let us first present the problem. It is well-known that for any time $T > 2$, given any initial data $(y^0, y^1) \in L^2(0,1) \times H^{-1}(0,1)$, we can find a control function $v(t) \in L^2(0,T)$ such that the solution of

$$\begin{cases} \partial_{tt}^2 y - \partial_{xx}^2 y = 0, & (x,t) \in (0,1) \times (0,T), \\ y(0,t) = v(t),\ y(1,t) = 0, & t \in (0,T), \\ y(x,0) = y^0(x),\ \partial_t y(x,0) = y^1(x), & x \in (0,1), \end{cases} \tag{2.3.1}$$

satisfies

$$y(T) = 0, \quad \partial_t y(T) = 0. \tag{2.3.2}$$

By duality (namely the Hilbert Uniqueness Method, or HUM in short), this property is equivalent to the observability inequality (2.1.4), see [21].

Note that there might be several controls $v \in L^2(0,T)$ such that (2.3.2) holds for solutions of (2.3.1). In the sequel, we will say that such a v is an admissible control for (2.3.1).

Besides, there is an explicit method to compute the so-called HUM control v_{HUM}, which is the one of minimal $L^2(0,T)$-norm among all admissible controls for (2.3.1). Indeed, set $T > 2$ and consider the functional

$$\begin{aligned} &\mathcal{J} : H_0^1(0,1) \times L^2(0,1) \to \mathbb{R} \\ &\mathcal{J}(z^0, z^1) = \frac{1}{2}\int_0^T (\partial_x z)^2(0,t)\ dt - \int_0^1 y^0(x)\partial_t z(x,0)\ dx + < y^1, z(.,0) >_{H^{-1}\times H_0^1}, \end{aligned} \tag{2.3.3}$$

where z is the solution of the backward conservative wave equation

$$\begin{cases} \partial_{tt}^2 z - \partial_{xx}^2 z = 0, & (x,t) \in (0,1)\times(0,T), \\ z(0,t) = z(1,t) = 0, & t \in (0,T), \\ z(x,T) = z^0(x),\ \partial_t z(x,T) = z^1(x), & x \in (0,1). \end{cases} \tag{2.3.4}$$

Then $\mathcal{J}$ is strictly convex, coercive (see (2.1.4)), and therefore has a unique minimizer $(Z^0, Z^1) \in H_0^1(0,1)\times L^2(0,1)$. The HUM control is then given by $v_{HUM}(t) = \partial_x Z(0,t)$, where Z is the solution of (2.3.4) with initial data (Z^0, Z^1).

Note also that the HUM control is the only admissible control v for (2.3.1) that can be written as $v(t) = \partial_x z(0,t)$ for some z solution of (2.3.4) with initial data in $H_0^1(0,1)\times L^2(0,1)$.

It is then natural to try to compute this control numerically. This question will be investigated in the sequel.

2.3.2 The semi-discrete setting

This part is inspired in [5, 6] where similar results have been derived for uniform meshes.

We consider a mesh $\mathcal{S}_n$ as in (2.1.6) and derive an approximation scheme for (2.3.1) from a mixed finite element method. The problem reads as follows: Given y_n^0 and y_n^1 defined on $\mathcal{S}_n$, find a discrete control $v_n \in L^2(0,T)$ such that the solution y_n of

$$\begin{cases} \dfrac{h_{j-1/2,n}}{4}(y''_{j-1,n} + y''_{j,n}) + \dfrac{h_{j+1/2,n}}{4}(y''_{j,n} + y''_{j+1,n}) \\ \qquad\qquad = \dfrac{y_{j+1,n} - y_{j,n}}{h_{j+1/2,n}} - \dfrac{y_{j,n} - y_{j-1,n}}{h_{j-1/2,n}}, \quad j = 1,\cdots,n,\ t \in [0,T], \\ y_{0,n}(t) = v_n(t), \quad y_{n+1,n}(t) = 0, \quad t \in (0,T), \\ y_{j,n}(0) = y_{j,n}^0 \quad y'_{j,n}(0) = y_{j,n}^1, \quad j = 1,\cdots,n, \end{cases} \tag{2.3.5}$$

satisfies

$$y_{j,n}(T) = 0, \quad y'_{j,n}(T) = 0, \quad j = 1,\cdots,n. \tag{2.3.6}$$

Again, the study of this problem is based on a duality principle. Given any $T > 2$, we choose $\epsilon > 0$ such that $T - 4\epsilon > 2$ and a smooth function ρ satisfying

$$\begin{cases} \rho(t) = 1, & \text{if } t \in [2\epsilon, T-2\epsilon], \\ \rho(t) = 0, & \text{if } t \in [0,\epsilon] \cup [T-\epsilon, T], \end{cases} \quad \text{and} \quad 0 \le \rho(t) \le 1, \quad \forall t. \tag{2.3.7}$$

We then introduce the functional $\mathcal{J}_n$ defined by:

$$\begin{aligned}\mathcal{J}_n(z_n^0, z_n^1) = \frac{1}{8}\int_0^T \rho(t)|z'_{1,n}|^2(t)\ dt + \frac{1}{2}\int_0^T \Big(\frac{z_{1,n}(t)}{h_{1/2,n}}\Big)^2\ dt \\ + \left(\frac{h_{1/2,n}}{4}y^1_{1,n}z_{1,n}(0) + \sum_{j=1}^n \frac{h_{j+1/2,n}}{4}(y^1_{j,n}+y^1_{j+1,n})(z_{j,n}(0)+z_{j+1,n}(0))\right) \\ - \left(\frac{h_{1/2,n}}{4}y^0_{1,n}z'_{1,n}(0) + \sum_{j=1}^n \frac{h_{j+1/2,n}}{4}(y^0_{j,n}+y^0_{j+1,n})(z'_{j,n}(0)+z'_{j+1,n}(0))\right),\end{aligned} \tag{2.3.8}$$

where z_n is the solution of

$$\begin{cases} \dfrac{h_{j-1/2,n}}{4}(z''_{j-1,n}+z''_{j,n}) + \dfrac{h_{j+1/2,n}}{4}(z''_{j,n}+z''_{j+1,n}) \\ \qquad\qquad = \dfrac{z_{j+1,n}-z_{j,n}}{h_{j+1/2,n}} - \dfrac{z_{j,n}-z_{j-1,n}}{h_{j-1/2,n}}, \quad j=1,\cdots,n,\ t\in[0,T], \\ z_{0,n}(t) = z_{n+1,n}(t) = 0, \quad t\in(0,T), \\ z_{j,n}(T) = z^0_{j,n}, \quad z'_{j,n}(T) = z^1_{j,n}, \quad j=1,\cdots,n. \end{cases} \tag{2.3.9}$$

Then the following Lemma holds:

Lemma 2.3.1. *For any integer n, the functional $\mathcal{J}_n$ is strictly convex and coercive, and then has a unique minimizer (Z_n^0, Z_n^1). Besides, for all n, if v_n is the solution of*

$$\begin{cases} -\dfrac{h_{1/2,n}}{4}v''_n + \dfrac{1}{h_{1/2,n}}v_n = -\dfrac{1}{4}(\rho Z'_{1,n})' + \dfrac{1}{h^2_{1/2,n}}Z_{1,n}, \quad t\in[0,T], \\ v'_n(0) = v'_n(T) = 0, \end{cases} \tag{2.3.10}$$

where Z_n is the solution of (2.3.9) *with initial data (Z_n^0, Z_n^1), then $v_n(t)$ is a control of* (2.3.5) *in time T.*

The proof of Lemma 2.3.1 is the same as in [5]. For completeness, we will give a sketch of the proof hereafter.

For convenience, we introduce the operators $\mathbb{P}_{\mathcal{S}_n}$, $\mathbb{Q}_{\mathcal{S}_n}$ and $\mathbb{R}_{\mathcal{S}_n}$ which map discrete data $a_n = (a_{j,n})_{j\in\{1,\cdots,n\}}$ given on a mesh $\mathcal{S}_n$ as in (2.1.6) to functions defined on $(0,1)$ by:

$$\begin{aligned} &\mathbb{P}_{\mathcal{S}_n}a_n(x) = a_{j,n} + (a_{j+1,n}-a_{j,n})\Big(\frac{x-x_{j,n}}{h_{j+1/2,n}}\Big), \\ &\mathbb{Q}_{\mathcal{S}_n}a_n(x) = \frac{a_{j,n}+a_{j+1,n}}{2}, \qquad\qquad \text{on } [x_{j,n}, x_{j+1,n}], \\ &\mathbb{R}_{\mathcal{S}_n}a_n(x) = \frac{h_{j+1/2,n}}{4}(a_{j,n}+a_{j+1,n}) + \sum_{k=j+1}^n h_{k+1/2,n}\Big(\frac{a_{k,n}+a_{k+1,n}}{2}\Big), \end{aligned}$$

with the convention $a_{0,n} = a_{n+1,n} = 0$. With these definitions, $\mathbb{P}_{\mathcal{S}_n}$ and $\mathbb{Q}_{\mathcal{S}_n}$ are extension operators, and $\mathbb{R}_{\mathcal{S}_n}$ corresponds to a piecewise continuous approximation operator of the discrete integrals $x \mapsto \int_x^1 \mathbb{Q}_{\mathcal{S}_n}a_n(s)\ ds$.

Let us rewrite all discrete computations in terms of the operators $\mathbb{P}_{\mathcal{S}_n}, \mathbb{Q}_{\mathcal{S}_n}, \mathbb{R}_{\mathcal{S}_n}$. First, for any solution z_n of (2.3.9), the energy (2.1.8) writes

$$E_n(t) = \frac{1}{2}\ \|\mathbb{Q}_{\mathcal{S}_n}z_n(t)\|^2_{L^2(0,1)} + \frac{1}{2}\ \|\partial_x(\mathbb{P}_{\mathcal{S}_n}z_n(t))\|^2_{L^2(0,1)}\,. \tag{2.3.11}$$

Second, the functional $\mathcal{J}_n$ reads as

$$\mathcal{J}_n(z_n^0, z_n^1) = \frac{1}{8}\int_0^T \rho(t)|z'_{1,n}|^2(t)\ dt + \frac{1}{2}\int_0^T \Big(\frac{z_{1,n}(t)}{h_{1/2,n}}\Big)^2\ dt \\ + \int_0^1 (\mathbb{R}_{\mathcal{S}_n} y_n^1)(\partial_x \mathbb{P}_{\mathcal{S}_n} z_n(0))\ dx - \int_0^1 (\mathbb{Q}_{\mathcal{S}_n} y_n^0)(\mathbb{Q}_{\mathcal{S}_n} z'_n(0))\ dx. \quad (2.3.12)$$

We are now in position to sketch the proof of Lemma 2.3.1.

Sketch of the proof of Lemma 2.3.1. Fix an integer $n \in \mathbb{N}$. The functional $\mathcal{J}_n$ is strictly convex, and its coercivity is obvious since we are working in a finite dimensional setting. It follows that $\mathcal{J}_n$ has a unique minimizer (Z_n^0, Z_n^1).

Let us compute the Fréchet derivative of $\mathcal{J}_n$ in the minimizer (Z_n^0, Z_n^1): For any (z_n^0, z_n^1), the solution z_n of (2.3.9) on $\mathcal{S}_n$ satisfies (Recall the definition (2.3.7) of ρ):

$$0 = \int_0^T \Big(-\frac{1}{4}(\rho(t) Z'_{1,n}(t))' + \frac{1}{h_{1/2,n}^2} Z_{1,n}(t)\Big) z_{1,n}(t)\ dt \\ + \int_0^1 (\mathbb{R}_{\mathcal{S}_n} y_n^1)(\partial_x \mathbb{P}_{\mathcal{S}_n} z_n(0))\ dx - \int_0^1 (\mathbb{Q}_{\mathcal{S}_n} y_n^0)(\mathbb{Q}_{\mathcal{S}_n} z'_n(0))\ dx,$$

which rewrites, in terms of v_n defined in (2.3.10), as

$$0 = \frac{1}{4}\int_0^T h_{1/2,n} v'_n z'_{1,n}\ dt + \int_0^T v_n \frac{z_{1,n}}{h_{1/2,n}}\ dt \\ + \int_0^1 (\mathbb{R}_{\mathcal{S}_n} y_n^1)(\partial_x \mathbb{P}_{\mathcal{S}_n} z_n(0))\ dx - \int_0^1 (\mathbb{Q}_{\mathcal{S}_n} y_n^0)(\mathbb{Q}_{\mathcal{S}_n} z'_n(0))\ dx. \quad (2.3.13)$$

Now, consider y_n the solution of (2.3.5) with boundary control v_n. Multiplying (2.3.5) by z_n solution of (2.3.9) with initial data (z_n^0, z_n^1), we get, after tedious computations that are left to the reader, that

$$0 = \frac{1}{4}\int_0^T h_{1/2,n} v'_n z'_{1,n}\ dt + \int_0^T v_n \frac{z_{1,n}}{h_{1/2,n}}\ dt \\ + \int_0^1 (\mathbb{R}_{\mathcal{S}_n} y_n^1)(\partial_x \mathbb{P}_{\mathcal{S}_n} z_n(0))\ dx - \int_0^1 (\mathbb{Q}_{\mathcal{S}_n} y_n^0)(\mathbb{Q}_{\mathcal{S}_n} z'_n(0))\ dx \\ - \int_0^1 (\mathbb{R}_{\mathcal{S}_n} y'_n(T))(\partial_x \mathbb{P}_{\mathcal{S}_n} z_n^0)\ dx + \int_0^1 (\mathbb{Q}_{\mathcal{S}_n} y_n(T))(\mathbb{Q}_{\mathcal{S}_n} z_n^1)\ dx. \quad (2.3.14)$$

Combined with (2.3.13), this yields that the solution y_n of (2.3.5) satisfies the following property: For any (z_n^0, z_n^1),

$$-\int_0^1 (\mathbb{R}_{\mathcal{S}_n} y'_n(T))(\partial_x \mathbb{P}_{\mathcal{S}_n} z_n^0)\ dx + \int_0^1 (\mathbb{Q}_{\mathcal{S}_n} y_n(T))(\mathbb{Q}_{\mathcal{S}_n} z_n^1)\ dx = 0.$$

This obviously implies (2.3.6). □

It is natural to ask if the discrete controls v_n constructed in Lemma 2.3.1 converge to an admissible control for (2.3.1) under some assumptions on the convergence of (y_n^0, y_n^1). We will prove that this is indeed the case.

Given a sequence of meshes $(\mathcal{S}_n)_n$, we say that the sequence of discrete data $(a_n, b_n)_n$ defined on the meshes $\mathcal{S}_n$ strongly converges to (a, b) in $L^2(0,1) \times H^{-1}(0,1)$ if:

$$\mathbb{Q}_{\mathcal{S}_n} a_n \to a \quad \text{in } L^2(0,1), \quad \text{and} \quad \mathbb{R}_{\mathcal{S}_n} b_n \to \Big(x \mapsto \int_x^1 b(s)\, ds\Big) \quad \text{in } L^2(0,1). \tag{2.3.15}$$

Remark that this definition makes sense, since for $b \in H^{-1}(0,1)$, classical arguments allow to define the function $x \mapsto \int_x^1 b(s)\, ds$ in $L^2(0,1)$.

Theorem 2.3.2. *Let $(y^0, y^1) \in L^2(0,1) \times H^{-1}(0,1)$ and $T > 2$.*

Given $M \geq 1$, we consider a sequence $(\mathcal{S}_n)$ of M-regular meshes, and a sequence of initial data (y_n^0, y_n^1) which strongly converges to (y^0, y^1) in $L^2(0,1) \times H^{-1}(0,1)$ in the sense of (2.3.15).

Then the sequence of discrete controls $(v_n)_n$ given by Lemma 2.3.1 strongly converges in $L^2(0,T)$ to the HUM control v_{HUM} for (2.3.1) with initial data (y^0, y^1).

First of all, let us mention that, given $(y^0, y^1) \in L^2(0,1) \times H^{-1}(0,1)$, it is possible to find a sequence of initial data (y_n^0, y_n^1) which strongly converges to (y^0, y^1) in $L^2(0,1) \times H^{-1}(0,1)$ in the sense of (2.3.15). We will briefly explain later (Remark 2.3.5 below) how this can be done.

The proof of Theorem 2.3.2 is mainly based on inequality (2.1.11), that implies that the discrete controls v_n are bounded in $L^2(0,T)$. Once this is proved, the result can be deduced from classical convergence properties of the scheme.

Proof. The proof is divided into several steps. First, we prove uniform bounds on the sequence v_n. Second, we prove that any weak limit of v_n is an admissible control for (2.3.1). Third, we prove that there is only one weak limit, which coincides with the HUM-control v_{HUM} of (2.3.1). We finally prove the strong convergence of the controls v_n in $L^2(0,T)$.

Uniform bounds. Since $\mathcal{J}_n(Z_n^0, Z_n^1) \leq \mathcal{J}_n(0,0) = 0$, we have that

$$\frac{1}{8}\int_0^T \rho(t)|Z'_{1,n}|^2(t)\, dt + \frac{1}{2}\int_0^T \Big(\frac{Z_{1,n}(t)}{h_{1/2,n}}\Big)^2\, dt \;\leq \sqrt{2E_{*n}(0)}\sqrt{\|\mathbb{R}_{\mathcal{S}_n} y_n^1\|^2_{L^2(0,1)} + \|\mathbb{Q}_{\mathcal{S}_n} y_n^0\|^2_{L^2(0,1)}},$$

where $E_{*n}(t)$ denotes the energy of $Z_n(t)$, which is constant. In view of the definition of ρ, since we assume that the meshes $\mathcal{S}_n$ are M-regular, inequality (2.1.11) holds. This, combined with the fact that $(\mathbb{Q}_{\mathcal{S}_n} y_n^0)$ and $(\mathbb{R}_{\mathcal{S}_n} y_n^1)$ are convergent in $L^2(0,1)$ and therefore bounded, leads us to

$$k_T E_{*n}(T) \leq \frac{1}{8}\int_0^T \rho(t)|Z'_{1,n}|^2(t)\, dt + \frac{1}{2}\int_0^T \Big(\frac{Z_{1,n}(t)}{h_{1/2,n}}\Big)^2\, dt \leq C. \tag{2.3.16}$$

Besides, multiplying (2.3.10) by $h_{1/2,n} v_n$ and integrating in time gives

$$\begin{aligned}\int_0^T \frac{h_{1/2,n}^2}{4}|v_n'(t)|^2 + |v_n(t)|^2\, dt &= \int_0^T \Big(\frac{h_{1/2,n}}{4}\rho(t) Z'_{1,n}(t) v_n'(t) + \frac{Z_{1,n}(t)}{h_{1/2,n}} v_n(t)\Big)\, dt \\ &\leq \Big(\int_0^T \frac{h_{1/2,n}^2}{4}|v_n'(t)|^2 + |v_n(t)|^2\, dt\Big)^{1/2}\Big(\int_0^T \frac{\rho(t)}{8}|Z'_{1,n}|^2(t)\, dt + \frac{1}{2}\int_0^T \Big(\frac{Z_{1,n}(t)}{h_{1/2,n}}\Big)^2\, dt\Big)^{1/2}, \end{aligned} \tag{2.3.17}$$

and therefore we obtain

$$\int_0^T \frac{h_{1/2,n}^2}{4}|v_n'(t)|^2 + |v_n(t)|^2 \, dt \leq C. \tag{2.3.18}$$

We have thus proved, using the M-regularity assumption, that the sequence of discrete controls v_n is bounded in $L^2(0,T)$. Therefore there exists a function $v \in L^2(0,T)$ such that

$$v_n \rightharpoonup v, \quad \text{in } L^2(0,T) \text{ } weak, \quad \text{and} \quad h_{1/2,n} v_n' \rightharpoonup 0, \quad \text{in } L^2(0,T) \text{ } weak. \tag{2.3.19}$$

The second statement in (2.3.19) comes from the continuity of the derivation in the sense of distributions.

The function v is an admissible control for (2.3.1). We need the following classical Lemma on the convergence of the numerical schemes (which can be found for instance in [7]):

Lemma 2.3.3. *Consider two smooth functions (u^0, u^1) on $(0,1)$ such that $u^0(0) = u^0(1) = 0$ and $u(x,t)$ the solution of the conservative system (2.1.1) with initial data (u^0, u^1).*

Given a sequence $(\mathcal{S}_n)_n$ of M-regular meshes, for all $n \in \mathbb{N}$, we denote by $u_n(t)$ the solution of the conservative semi-discrete scheme (2.1.7) with initial data

$$u_{j,n}^0 = u^0(x_{j,n}), \quad u_{j,n}^1 = u^1(x_{j,n}), \quad j \in \{1, \cdots, n\}.$$

Then $(\mathbb{P}_{\mathcal{S}_n} u_{j,n}, \mathbb{Q}_{\mathcal{S}_n} u_{j,n}')$ strongly converges to (u, u') in $C([0,T]; H_0^1(0,1) \times L^2(0,1))$ and

$$\frac{u_{1,n}(t)}{h_{1/2,n}} \to \partial_x u(0,t) \text{ in } L^2(0,T), \quad \text{and} \quad u_{1,n}'(t) \to 0 \text{ in } L^2(0,T). \tag{2.3.20}$$

This result is of course still true for the backward system (2.3.4) and its semi-discrete approximations (2.3.9).

Now, consider two smooth functions (z^0, z^1), and define, as in Lemma 2.3.3, the solution z of the backward wave equation (2.3.4) with initial data (z^0, z^1), and the solution z_n of the semi-discrete systems (2.3.9), with initial data $(z^0(x_{j,n}), z^1(x_{j,n}))$.

Using (2.3.19) and Lemma 2.3.3, we can pass to the limit in (2.3.13) and obtain that the solution z of (2.3.4) satisfies:

$$0 = \int_0^T v(t)\partial_x z(0,t) \, dt + < y^1, z(.,0) >_{H^{-1}(0,1)\times H_0^1(0,1)} - \int_0^1 y^0(x)\partial_t z(x,0) \, dx. \tag{2.3.21}$$

By a density argument, this identity can be extended to any $(z^0, z^1) \in H_0^1(0,1) \times L^2(0,T)$.

Besides, for any $(z^0, z^1) \in H_0^1(0,1) \times L^2(0,1)$, as in (2.3.14), multiplying the solution of (2.3.1) with boundary condition $y(0,t) = v(t)$ and initial data (y^0, y^1) by z solution of (2.3.4) with initial data (z_0, z_1), we obtain that

$$0 = \int_0^T v(t)\partial_x z(0,t) \, dt + < y^1, z(.,0) >_{H^{-1}(0,1)\times H_0^1(0,1)} - \int_0^1 y^0(x)\partial_t z(x,0) \, dx$$
$$- < \partial_t y(T), z^0 >_{H^{-1}(0,1)\times H_0^1(0,1)} + \int_0^1 y(T,x) z^1(x) \, dx.$$

Hence we deduce from (2.3.21) that

$$< \partial_t y(T), z^0 >_{H^{-1}(0,1)\times H_0^1(0,1)} - \int_0^1 y(T,x) z^1(x)\ dx = 0.$$

Therefore y satisfies (2.3.2). This precisely means that v is an admissible control for (2.3.1).

The limit v is the HUM control v_{HUM}. It is sufficient to prove that $v(t)$ coincides with some $\partial_x z(t,0)$, where z is the solution of (2.3.4) for some initial data $(z^0, z^1) \in H_0^1(0,1) \times L^2(0,1)$, see for instance [21].

From (2.3.16), there exist two functions $Z^0 \in H_0^1(0,1)$ and $Z^1 \in L^2(0,1)$ such that

$$\mathbb{P}_{\mathcal{S}_n} Z_n^0 \rightharpoonup Z^0,\ H_0^1(0,1)\ weak, \quad \text{and} \quad \mathbb{Q}_{\mathcal{S}_n} Z_n^1 \rightharpoonup Z^1,\ L^2(0,1)\ weak.$$

Using the weak formulations of (2.3.9) and the conservation of the energy, we can prove (the proof can be adapted in a standard way from the arguments in [7], in particular Lemma 2.3.3, and is left to the reader) that:

$$\begin{array}{l} (\mathbb{P}_{\mathcal{S}_n} Z_n, \mathbb{Q}_{\mathcal{S}_n} Z_n) \rightharpoonup (Z, Z') \quad \text{in } L^\infty(0,T; H_0^1(0,1)\times L^2(0,1))\ * weak, \\ \forall t \in [0,T], \quad (\mathbb{P}_{\mathcal{S}_n} Z_n(t), \mathbb{Q}_{\mathcal{S}_n} Z_n(t)) \rightharpoonup (Z(t), Z'(t)) \quad \text{in } H_0^1(0,1)\times L^2(0,1)\ weak, \end{array} \tag{2.3.22}$$

where Z is the solution of (2.3.4) with initial data (Z^0, Z^1). Besides, one easily shows that

$$\frac{Z_{1,n}}{h_{1/2,n}} - \frac{h_{1/2,n}}{4} Z_{1,n}'' \rightharpoonup \partial_x Z(0,t), \quad \text{in} \quad \mathcal{D}'(0,T). \tag{2.3.23}$$

But $Z_{1,n}/h_{1/2,n}$ is bounded in $L^2(0,T)$ from (2.3.16), and therefore $h_{1/2,n} Z_{1,n}'' \rightharpoonup 0$ in $\mathcal{D}'(0,T)$. This also gives that

$$\frac{Z_{1,n}}{h_{1/2,n}} \rightharpoonup \partial_x Z \quad \text{in } \mathcal{D}'(0,T), \qquad Z_{1,n} \rightharpoonup 0 \quad \text{in } \mathcal{D}'(0,T), \qquad h_{1/2,n}(\rho Z_{1,n}')' \rightharpoonup 0 \quad \text{in } \mathcal{D}'(0,T). \tag{2.3.24}$$

Combined with the definition of v_n in Lemma 2.3.1, it follows that

$$-\frac{h_{1/2,n}^2}{4} v_n'' + v_n \rightharpoonup \partial_x Z(0,t),\ \text{in}\ \ \mathcal{D}'(0,T).$$

But, since v_n is bounded in $L^2(0,T)$ by (2.3.18),

$$h_{1/2,n}^2 v_n'' \rightharpoonup 0 \text{ in } \mathcal{D}'(0,T),$$

and therefore $v(t) = \partial_x Z(0,t)$ in $\mathcal{D}'(0,T)$.

Since we have already proved that v is an admissible control for (2.1.1), this proves that v is the HUM control v_{HUM}.

Strong convergence. Since the weak convergence is already proven, it is sufficient to prove the convergence of the $L^2(0,T)$-norms.
Since $v(t) = \partial_x Z(0,t)$ for a solution Z of (2.3.4) with initial data (Z^0, Z^1), we get from (2.3.21) that :

$$0 = \int_0^T (\partial_x Z(0,t))^2\ dt - < y^1, Z(.,0) >_{H^{-1}(0,1)\times H_0^1(0,1)} - \int_0^1 y^0(x) \partial_t Z(x,0)\ dx. \tag{2.3.25}$$

But (2.3.13) gives:

$$0 = \frac{1}{4}\int_0^T \rho(t)|Z'_{1,n}(t)|^2\ dt + \int_0^T \left|\frac{Z_{1,n}(t)}{h_{1/2,n}}\right|^2 dt + \int_0^1 (\mathbb{R}_{\mathcal{S}_n} y_n^1)(x)\partial_x(\mathbb{P}_{\mathcal{S}_n} Z_n)(x,0)\ dx - \int_0^1 (\mathbb{Q}_{\mathcal{S}_n} y_n^0)(x)(\mathbb{Q}_{\mathcal{S}_n} Z'_n)(x,0)\ dx.$$

Convergences (2.3.22) and (2.3.15) imply that we can pass to the limit in the linear term, and therefore, by (2.3.25), we get:

$$\frac{1}{4}\int_0^T \rho(t)|Z'_{1,n}(t)|^2\ dt + \int_0^T \left|\frac{Z_{1,n}(t)}{h_{1/2,n}}\right|^2 dt \to \int_0^T |\partial_x Z(0,t)|^2\ dt.$$

Combined with the weak convergences (2.3.24), this proves the following strong convergences:

$$\begin{cases} \sqrt{\rho}\, Z'_{1,n} \to 0, \\ \dfrac{Z_{1,n}}{h_{1/2,n}}(t) \to \partial_x Z(0,t), \end{cases} \qquad \text{in } L^2(0,T).$$

But, from the definition (2.3.10) of v_n, the convergence (2.3.19) implies that:

$$\int_0^T \frac{h_{1/2,n}^2}{4}|v'_n(t)|^2 + |v_n(t)|^2\ dt = \int_0^T \frac{h_{1/2,n}}{4}\rho(t)Z'_{1,n}(t)v'_n(t) + \frac{Z_{1,n}(t)}{h_{1/2,n}}v_n(t)\ dt \longrightarrow \int_0^T \partial_x Z(0,t)v(t)\ dt = \int_0^T v(t)^2\ dt.$$

Hence we deduce from (2.3.19) that:

$$h_{1/2,n}v'_n \to 0 \quad \text{in } L^2(0,T), \quad \text{and} \quad v_n \to v = v_{HUM} \quad \text{in } L^2(0,T),$$

which concludes the proof of Theorem 2.3.2. □

Remark 2.3.4. The proof of Theorem 2.3.2 slightly differs from the one in [5], which presented an approach based on the spectral decomposition of the solutions. This technique, in our context, seems more technically involved than the one presented above, since the spectrum is not as explicit as in the case of a uniform mesh.

Remark 2.3.5. Let us briefly comment the hypothesis (2.3.15), and prove that, given $(a,b) \in L^2(0,1) \times H^{-1}(0,1)$ and a sequence $\mathcal{S}_n$ of M-regular meshes, there exists a sequence of discrete data (a_n, b_n) defined on the mesh $\mathcal{S}_n$ which strongly converges to (a,b) in $L^2(0,1) \times H^{-1}(0,1)$ in the sense of (2.3.15).

Indeed, for $a \in L^2(0,1)$, define $a_n = \mathbb{A}_{\mathcal{S}_n}(a)$ as follows (recall the convention $a_{n+1,n} = 0$):

$$a_{j,n} + a_{j+1,n} = \frac{2}{h_{j+1/2,n}}\int_{x_{j,n}}^{x_{j+1,n}} a(x)\ dx, \quad 1 \le j \le n.$$

If a is continuous on $[0,1]$, one easily checks that

$$\|\mathbb{Q}_{\mathcal{S}_n}(\mathbb{A}_{\mathcal{S}_n}(a)) - a\|_{L^2(0,1)} \to 0.$$

Besides, if a is in L^2, we have that

$$\|\mathbb{Q}_{\mathcal{S}_n}(\mathbb{A}_{\mathcal{S}_n}(a)) - a\|_{L^2(0,1)} \le C\,\|a\|_{L^2(0,1)}.$$

This, using the density of the continuous functions in $L^2(0,1)$, is sufficient to prove that the sequence of discrete data $a_n = \mathbb{Q}_{\mathcal{S}_n}(\mathbb{A}_{\mathcal{S}_n}(a))$ converges to a in $L^2(0,1)$ for all $a \in L^2(0,1)$.

For the approximation of $b \in H^{-1}(0,1)$, we look for an approximation of

$$B(x) = \int_x^1 b(s)\ ds,$$

which lies in $L^2(0,1)$. Thus, the sequence $B_n = \mathbb{A}_{\mathcal{S}_n} B$ provides discrete data which satisfy $\mathbb{Q}_{\mathcal{S}_n}(B_n) \to B$ in $L^2(0,1)$ when $n \to \infty$. It is then sufficient to find discrete data b_n such that $\mathbb{R}_{\mathcal{S}_n} b_n = \mathbb{Q}_{\mathcal{S}_n} B_n$, and this can be done explicitly.

2.4 Application to the damped wave equation

2.4.1 The continuous setting

We consider the continuous damped wave equation on the interval $(0,1)$:

$$\begin{cases} \partial_{tt}^2 w - \partial_{xx}^2 w + 2\sigma\partial_t w = 0, & (x,t) \in (0,1)\times(0,\infty), \\ w(0,t) = w(1,t) = 0, & t \in (0,\infty), \\ w(x,0) = w^0(x), \quad \partial_t w(x,0) = w^1(x), & x \in (0,1), \end{cases} \tag{2.4.1}$$

with $w^0 \in H_0^1(0,1)$ and $w^1 \in L^2(0,1)$.

We assume that the damping function $\sigma = \sigma(x)$ is bounded, non-negative and bounded from below by a positive number on a subinterval J, that is there exists $\alpha > 0$, such that

$$\sigma(x) \geq \alpha, \quad \forall x \in J, \quad \text{and} \quad \|\sigma\|_\infty = K. \tag{2.4.2}$$

Then the energy, defined by (2.1.2), satisfies the dissipation law

$$\frac{dE}{dt}(t) = -2\int_0^1 \sigma(x)|\partial_t w(t,x)|^2\ dx, \qquad t \geq 0. \tag{2.4.3}$$

It is well-known that, under the assumption (2.4.2), the energy is exponentially decaying: There exist positive constants C and μ such that

$$E(t) \leq C\,E(0)\exp(-\mu t), \qquad t \geq 0. \tag{2.4.4}$$

Using classical arguments in stabilization theory (see [16]), the energy of (2.4.1) is exponentially decaying if and only if the observability inequality (2.1.5) holds for solutions of the conservative system (2.1.1).

2.4.2 The semi-discrete setting

We consider a mesh $\mathcal{S}_n$ as in (2.1.6), and discretize equation (2.4.1) according to the mixed finite element method:

$$\begin{cases} \dfrac{h_{j-1/2,n}}{4}(w''_{j-1,n}+w''_{j,n})+\dfrac{h_{j+1/2,n}}{4}(w''_{j,n}+w''_{j+1,n})= \\ \qquad -\dfrac{h_{j-1/2,n}\sigma_{j-1/2,n}}{2}(w'_{j-1,n}+w'_{j,n})-\dfrac{h_{j+1/2,n}\sigma_{j+1/2,n}}{2}(w'_{j,n}+w'_{j+1,n}) \\ \qquad +\dfrac{w_{j+1,n}-w_{j,n}}{h_{j+1/2,n}}-\dfrac{w_{j,n}-w_{j-1,n}}{h_{j-1/2,n}}, \qquad j=1,\cdots,n,\ t\in[0,\infty), \\ w_0(t)=w_{n+1}(t)=0, \qquad t\in[0,\infty), \\ w_j(0)=w^0_{j,n}, \quad w'_j(0)=w^1_{j,n}, \qquad j=1,\cdots,n, \end{cases} \tag{2.4.5}$$

where $\sigma_{j+1/2,n}$ is an approximation on $[x_{j,n},x_{j+1,n}]$ of the damping function σ in (2.4.1) which is assumed to satisfy the following properties:

$$\sigma_{j+1/2,n}\geq\alpha, \quad \forall[x_{j,n},x_{j+1,n}]\subset J, \quad \text{and} \quad 0\leq\sigma_{j+1/2,n}\leq K, \quad \forall j\in\{0,\cdots,n\}, \tag{2.4.6}$$

where α, K and J are as in (2.4.2).

The energy (2.1.8) of solutions of (2.4.5) satisfies

$$\frac{dE_n}{dt}(t)=-2\sum_{j=0}^{n}h_{j+1/2,n}\sigma_{j+1/2,n}\left(\frac{w'_{j,n}(t)+w'_{j+1,n}(t)}{2}\right)^2. \tag{2.4.7}$$

Obviously, this dissipation law corresponds to a discrete version of (2.4.3).

The question we investigate is the following: Given a sequence $(\mathcal{S}_n)_n$ of meshes, can we find positive constants C and μ independent of n such that

$$E_n(t)\leq C\,E_n(0)\exp(-\mu t), \quad t\geq 0, \tag{2.4.8}$$

for any solution of (2.4.5) on $\mathcal{S}_n$?

Similarly as in the continuous setting, this property is equivalent to the uniform observability inequality (2.1.12) for solutions of the conservative system (2.1.7) (see for instance [28]). Therefore Theorem 2.1.2 leads to the following result:

Theorem 2.4.1. *Let $M\geq 1$, and consider a sequence $(\mathcal{S}_n)_n$ of M-regular meshes and a sequence of damping functions σ_n satisfying* (2.4.6).

Then there exist positive constants C and μ such that for all n, inequality (2.4.8) *holds for any solution of* (2.4.5) *on $\mathcal{S}_n$.*

The proof of Theorem 2.4.1, which can be adapted in a standard way from [16] or [28], is left to the reader.

Remark 2.4.2. Note that this method yields an estimate on the decay rate μ appearing in (2.4.8), which is far from being optimal in general. This is a drawback of the method, which is based on a perturbation argument of the conservative system. Even in the continuous setting, the decay rate parameter obtained through this method is not in general the sharp one, which is known to coincide (at least in the one dimensional case) with the spectral abscissa (see [8]).

Remark 2.4.3. The analysis proposed here can be applied as well to the 1d Perfectly Matched Layers equations (see [2, 11]), which, roughly, consists in a damped wave equation written in hyperbolic form:

$$\begin{cases} \partial_t p + \partial_x q + \sigma p = 0, & (x,t) \in (0,1) \times (0,\infty), \\ \partial_t q + \partial_x p + \sigma q = 0, & (x,t) \in (0,1) \times (0,\infty), \\ q(0,t) = p(1,t) = 0, & t \in (0,\infty), \\ q(x,0) = q_0(x), \quad p(x,0) = p_0(x), & x \in (0,1), \end{cases} \tag{2.4.9}$$

where σ satisfies the assumptions (2.4.2).

In [11], it is proven that the 1d PML system is exponentially stable: The energy of solutions of (2.4.9), defined as

$$E(t) = \frac{1}{2} \int_0^1 |p(t,x)|^2 + |q(t,x)|^2 \, dx,$$

is exponentially decaying.

Besides, stabilization properties for space semi-discrete approximation schemes on uniform meshes are studied in [11]: It is proved that finite difference approximation schemes are not uniformly exponentially stable, but adding a viscosity term in space makes the schemes uniformly exponentially stable.
We claim that the so-called Box scheme (see for instance [13, 4]) on M-regular meshes for the 1d PML equations also are exponentially stable. To be more precise, for $\mathcal{S}_n$ is a M-regular mesh, we consider the space approximation scheme of (2.4.9) given by:

$$\begin{cases} \Big(\frac{p'_{j,n} + p'_{j+1,n}}{2}\Big) + \Big(\frac{q_{j+1,n} - q_{j,n}}{h_{j+1/2,n}}\Big) = 0, \quad 0 \le j \le n, \ t \ge 0, \\ \Big(\frac{q'_{j,n} + q'_{j+1,n}}{2}\Big) + \Big(\frac{p_{j+1,n} - p_{j,n}}{h_{j+1/2,n}}\Big) = 0, \quad 0 \le j \le n, \ t \ge 0, \\ q_{0,n}(t) = p_{n+1,n}(t) = 0, \quad t \ge 0. \end{cases} \tag{2.4.10}$$

Then the energy of solutions (p_n, q_n) of (2.4.10), defined by

$$E_n(t) = \frac{1}{2} \sum_{j=0}^{n} h_{j+1/2,n} \Big(\Big(\frac{p_{j,n} + p_{j+1,n}}{2}\Big)^2 + \Big(\frac{q_{j,n} + q_{j+1,n}}{2}\Big)^2 \Big) + \frac{1}{8} \Big(\frac{1}{n+1}\Big)^2 \sum_{j=0}^{n} h_{j+1/2,n} \Big(\Big(\frac{p'_{j,n} + p'_{j+1,n}}{2}\Big)^2 + \Big(\frac{q'_{j,n} + q'_{j+1,n}}{2}\Big)^2 \Big), \tag{2.4.11}$$

is exponentially decaying, uniformly with respect to n.

2.5 Further comments

In this paper, we have analyzed a space semi-discrete scheme derived from a mixed finite element method for a 1d wave equation, which has a good behavior with respect to both stabilization and controllability properties for a large class of nonuniform meshes.

1. The key point of our analysis is the description of the spectrum of the space discrete operator given in Theorems 2.2.1-2.2.3. It is particularly surprising that the spectrum can be described in a

rather explicit way for any mesh. This does not seem to be the case for other classical schemes, as the ones provided by finite difference or finite element methods. To our knowledge, in these cases, only asymptotic distributions of the eigenvalues are available, see for instance [3] and the literature therein.

2. It would be particularly challenging to understand the behavior of the discrete waves in higher dimension on nonuniform meshes. To our knowledge, this question has not been addressed so far. We expect this question to be difficult to address with the tools used until now, which require either a good knowledge of the eigenvalues (see [17, 25, 23, 26, 24, 31, 5, 6] and our own approach) or the existence of multipliers that behave well (see [28, 27, 11]) on the discrete systems.

3. Let us mention the recent work [10], which studied observability properties for time-discrete approximation schemes of linear conservative systems in a very general abstract setting. The approach developed in [10] allows to derive uniform observability inequalities for time-discrete approximation schemes in a systematic way. One of the interesting features of this technique is that it can be applied to fully discrete schemes as soon as the space semi-discrete approximation schemes satisfy uniform observability properties (see [10, Section 5]). Note that the study presented here fits in this abstract setting. Therefore, combining Theorem 2.1.2 and the results in [10], one can derive uniform (with respect to the time and space discretization parameters) observability properties for time-discrete approximation schemes of the space semi-discrete approximation scheme (2.1.7).

4. It would be interesting to estimate the (asymptotic) decay rate for the semi-discrete damped equation as in the continuous case, see [8]. In the continuous case, the computation of the decay rate of the energy is technically involved and requires to work directly on the damped system. We refer to the works [8, 9, 20] that deal with these questions for damped wave equations.

To our knowledge, even in the case of uniform meshes, this question is still open. Only some partial results in this direction are available in [11] for the space semi-discrete Perfectly Matched Layers equations (see [2]).

5. Let us also mention the recent work [12], which analyzes stabilization properties for time-discrete approximation schemes of abstract damped systems. In particular, in [12], several time-discrete approximation schemes have been designed to guarantee uniform (with respect to the *time* discretization parameter) stabilization properties, by adding a numerical viscosity term in time which efficiently damps out the high frequency components. Besides, this can also be applied to families of uniformly exponentially stable systems, and in particular to families of space semi-discrete approximation schemes that fit into the abstract setting of [12], which is the case for discrete approximations of damped wave equations. Thus, one can combine Theorem 2.4.1 and the results in [12] to derive uniformly (with respect to both time and space discretization parameters) exponentially stable fully discrete approximation schemes.

Acknowledgments. The author is grateful to E. Zuazua and J.-P. Puel for several suggestions and remarks related to this work.

Bibliography

[1] H. T. Banks, K. Ito, and C. Wang. Exponentially stable approximations of weakly damped wave equations. In *Estimation and control of distributed parameter systems (Vorau, 1990)*, volume 100 of *Internat. Ser. Numer. Math.*, pages 1–33. Birkhäuser, Basel, 1991.

[2] J.-P. Bérenger. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.*, 114(2):185–200, 1994.

[3] E. Bogomolny, O. Bohigas, and C. Schmit. Spectral properties of distance matrices. *J. Phys. A*, 36(12):3595–3616, 2003. Random matrix theory.

[4] T. J. Bridges and S. Reich. Numerical methods for Hamiltonian PDEs. *J. Phys. A*, 39(19):5287–5320, 2006.

[5] C. Castro and S. Micu. Boundary controllability of a linear semi-discrete 1-d wave equation derived from a mixed finite element method. *Numer. Math.*, 102(3):413–462, 2006.

[6] C. Castro, S. Micu, and A. Münch. Numerical approximation of the boundary control for the wave equation with mixed finite elements in a square. *IMA J. Numer. Anal.*, 28(1):186–214, 2008.

[7] L.C. Cowsar, T.F. Dupont, and M.F. Wheeler. A priori estimates for mixed finite element methods for the wave equations. *Comput. Methods Appl. Mech. Engrg.*, 82:205–222, 1990.

[8] S. Cox and E. Zuazua. The rate at which energy decays in a damped string. *Comm. Partial Differential Equations*, 19(1-2):213–243, 1994.

[9] S. Cox and E. Zuazua. The rate at which energy decays in a string damped at one end. *Indiana Univ. Math. J.*, 44(2):545–573, 1995.

[10] S. Ervedoza, C. Zheng, and E. Zuazua. On the observability of time-discrete conservative linear systems. *J. Funct. Anal.*, 254(12):3037–3078, June 2008. *Cf Chapitre 3.*

[11] S. Ervedoza and E. Zuazua. Perfectly matched layers in 1-d: Energy decay for continuous and semi-discrete waves. *Numer. Math.*, 109(4):597–634, 2008. *Cf Chapitre 1.*

[12] S. Ervedoza and E. Zuazua. Uniformly exponentially stable approximations for a class of damped systems. *To appear in J. Math. Pures Appl.*, 2008. *Cf Chapitre 5.*

[13] J. Frank, B. E. Moore, and S. Reich. Linear PDEs and numerical methods that preserve a multisymplectic conservation law. *SIAM J. Sci. Comput.*, 28(1):260–277 (electronic), 2006.

[14] R. Glowinski. Ensuring well-posedness by analogy: Stokes problem and boundary control for the wave equation. *J. Comput. Phys.*, 103(2):189–221, 1992.

[15] R. Glowinski, W. Kinton, and M. F. Wheeler. A mixed finite element formulation for the boundary controllability of the wave equation. *Internat. J. Numer. Methods Engrg.*, 27(3):623–635, 1989.

[16] A. Haraux. Une remarque sur la stabilisation de certains systèmes du deuxième ordre en temps. *Portugal. Math.*, 46(3):245–258, 1989.

[17] J.A. Infante and E. Zuazua. Boundary observability for the space semi discretizations of the 1-d wave equation. *Math. Model. Num. Ann.*, 33:407–438, 1999.

[18] A. E. Ingham. Some trigonometrical inequalities with applications to the theory of series. *Math. Z.*, 41(1):367–379, 1936.

[19] S. Labbé and E. Trélat. Uniform controllability of semidiscrete approximations of parabolic control systems. *Systems Control Lett.*, 55(7):597–609, 2006.

[20] G. Lebeau. Équations des ondes amorties. *Séminaire sur les Équations aux Dérivées Partielles, 1993–1994,École Polytech.*, 1994.

[21] J.-L. Lions. *Contrôlabilité exacte, Stabilisation et Perturbations de Systèmes Distribués. Tome 1. Contrôlabilité exacte*, volume RMA 8. Masson, 1988.

[22] F. Macià. The effect of group velocity in the numerical analysis of control problems for the wave equation. In *Mathematical and numerical aspects of wave propagation—WAVES 2003*, pages 195–200. Springer, Berlin, 2003.

[23] A. Münch. A uniformly controllable and implicit scheme for the 1-D wave equation. *M2AN Math. Model. Numer. Anal.*, 39(2):377–418, 2005.

[24] M. Negreanu, A.-M. Matache, and C. Schwab. Wavelet filtering for exact controllability of the wave equation. *SIAM J. Sci. Comput.*, 28(5):1851–1885 (electronic), 2006.

[25] M. Negreanu and E. Zuazua. Convergence of a multigrid method for the controllability of a 1-d wave equation. *C. R. Math. Acad. Sci. Paris*, 338(5):413–418, 2004.

[26] K. Ramdani, T. Takahashi, and M. Tucsnak. Uniformly exponentially stable approximations for a class of second order evolution equations—application to LQR problems. *ESAIM Control Optim. Calc. Var.*, 13(3):503–527, 2007.

[27] L. R. Tcheugoué Tebou and E. Zuazua. Uniform boundary stabilization of the finite difference space discretization of the $1-d$ wave equation. *Adv. Comput. Math.*, 26(1-3):337–365, 2007.

[28] L.R. Tcheugoué Tébou and E. Zuazua. Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity. *Numer. Math.*, 95(3):563–598, 2003.

[29] L. N. Trefethen. Group velocity in finite difference schemes. *SIAM Rev.*, 24(2):113–136, 1982.

[30] R. M. Young. *An introduction to nonharmonic Fourier series*. Academic Press Inc., San Diego, CA, first edition, 2001.

[31] E. Zuazua. Boundary observability for the finite-difference space semi-discretizations of the 2-D wave equation in the square. *J. Math. Pures Appl. (9)*, 78(5):523–563, 1999.

[32] E. Zuazua. Propagation, observation, and control of waves approximated by finite difference methods. *SIAM Rev.*, 47(2):197–243 (electronic), 2005.

Part II

Observability and stabilization properties for time-discrete approximation schemes

Chapter 3

On the observability of time-discrete conservative linear systems

Joint work with Chuang Zheng and Enrique Zuazua.

Abstract: We consider various time discretization schemes of abstract conservative evolution equations of the form $\dot{z} = Az$, where A is a skew-adjoint operator. We analyze the problem of observability through an operator B. More precisely, we assume that the pair (A, B) is exactly observable for the continuous model, and we derive uniform observability inequalities for suitable time-discretization schemes within the class of conveniently filtered initial data. The method we use is mainly based on the resolvent estimate given by Burq & Zworski in [2]. We present some applications of our results to time-discrete schemes for wave, Schrödinger and KdV equations and fully discrete approximation schemes for wave equations.

3.1 Introduction

Let X be a Hilbert space endowed with the norm $\|\cdot\|_X$ and let $A : \mathcal{D}(A) \to X$ be a skew-adjoint operator with compact resolvent. Let us consider the following abstract system:

$$\dot{z}(t) = Az(t), \quad z(0) = z_0. \tag{3.1.1}$$

Here and henceforth, a dot ($\dot{}$) denotes differentiation with respect to the time t. The element $z_0 \in X$ is called the *initial state*, and $z = z(t)$ is the *state* of the system. Such systems are often used as models of vibrating systems (e.g., the wave equation), electromagnetic phenomena (Maxwell's equations) or in quantum mechanics (Schrödinger's equation).

Assume that Y is another Hilbert space equipped with the norm $\|\cdot\|_Y$. We denote by $\mathfrak{L}(X, Y)$ the space of bounded linear operators from X to Y, endowed with the classical operator norm. Let $B \in \mathfrak{L}(\mathcal{D}(A), Y)$ be an observation operator and define the output function

$$y(t) = Bz(t). \tag{3.1.2}$$

In order to give a sense to (3.1.2), we make the assumption that B is an admissible observation operator in the following sense (see [27]):

Definition 3.1.1. The operator B is an admissible observation operator for system (3.1.1)-(3.1.2) if for every $T > 0$ there exists a constant $K_T > 0$ such that

$$\int_0^T \|y(t)\|_Y^2 \, dt \leq K_T \, \|z_0\|_X^2 \,, \qquad \forall \, z_0 \in \mathcal{D}(A). \tag{3.1.3}$$

Note that if B is *bounded* in X, i.e. if it can be extended such that $B \in \mathfrak{L}(X, Y)$, then B is obviously an admissible observation operator. However, in applications, this is often not the case, and the admissibility condition is then a consequence of a suitable "hidden regularity" property of the solutions of the evolution equation (3.1.1).

The exact observability property of system (3.1.1)-(3.1.2) can be formulated as follows:

Definition 3.1.2. System (3.1.1)-(3.1.2) is exactly observable in time T if there exists $k_T > 0$ such that

$$k_T \, \|z_0\|_X^2 \leq \int_0^T \|y(t)\|_Y^2 \, dt, \qquad \forall \, z_0 \in \mathcal{D}(A). \tag{3.1.4}$$

Moreover, (3.1.1)-(3.1.2) is said to be exactly observable if it is exactly observable in some time $T > 0$.

Note that observability issues arise naturally when dealing with controllability and stabilization properties of linear systems (see for instance the textbook [16]). Indeed, controllability and observability are dual notions, and therefore each statement concerning observability has its counterpart in controllability. In the sequel, we mainly focus on the observability properties of (3.1.1)-(3.1.2).

It was proved in [2, 18] that system (3.1.1)-(3.1.2) is exactly observable if and only if the following assertion holds:

$$\begin{cases} \text{There exist constants } \ M, m > 0 \ \text{ such that} \\ M^2 \, \|(i\omega I - A)z\|^2 + m^2 \, \|Bz\|_Y^2 \geq \|z\|^2 \,, \qquad \forall \ \omega \in \mathbb{R}, \ \ z \in \mathcal{D}(A). \end{cases} \tag{3.1.5}$$

This spectral condition can be viewed as a Hautus-type test, and generalizes the classical Kalman rank condition, see for instance [18, 26]. To be more precise, if (3.1.5) holds, then system (3.1.1)-(3.1.2) is exactly observable in any time $T > T_0 = \pi M$ (see [18]).

There is an extensive literature providing observability results for wave, plate, Schrödinger and elasticity equations, among other models and by various methods including microlocal analysis, multipliers and Fourier series, etc. Our goal in this paper is to develop a theory allowing to get results for time-discrete systems as a direct consequence of those corresponding to the time-continuous ones.

Let us first present a natural discretization of the continuous system. For any $\triangle t > 0$, we denote by z^k and y^k respectively the approximations of the solution z and the output function y of system (3.1.1)–(3.1.2) at time $t_k = k\triangle t$ for $k \in \mathbb{Z}$. Consider the following *implicit midpoint* time discretization of system (3.1.1):

$$\begin{cases} \dfrac{z^{k+1} - z^k}{\triangle t} = A\Big(\dfrac{z^{k+1} + z^k}{2}\Big), \qquad \text{in} \ \ X, \quad k \in \mathbb{Z}, \\ z^0 \text{ given.} \end{cases} \tag{3.1.6}$$

The output function of (3.1.6) is given by

$$y^k = Bz^k, \qquad k \in \mathbb{Z}. \tag{3.1.7}$$

Note that (3.1.6)–(3.1.7) is a discrete version of (3.1.1)–(3.1.2).

Taking into account that the spectrum of A is purely imaginary, it is easy to show that $\left\|z^k\right\|_X$ is conserved in the discrete time variable $k \in \mathbb{Z}$, i.e. $\left\|z^k\right\|_X = \left\|z^0\right\|_X$. Consequently the scheme under consideration is stable and its convergence (in the classical sense of numerical analysis) is guaranteed in an appropriate functional setting.

The uniform exact observability problem for system (3.1.6) is formulated as follows: *To find a positive constant $\tilde{k}_T$, independent of $\triangle t$, such that the solutions z^k of system* (3.1.6) *satisfy:*

$$\tilde{k}_T \left\|z^0\right\|_X^2 \leq \triangle t \sum_{k\in(0,T/\triangle t)} \left\|y^k\right\|_Y^2, \tag{3.1.8}$$

for all initial data z^0 in an appropriate class.

Clearly, (3.1.8) is a discrete version of (3.1.4).

Note that this type of observability inequalities appears naturally when dealing with stabilization and controllability problems (see, for instance, [16, 26, 31]). For numerical approximation processes, it is important that these inequalities hold uniformly with respect to the discretization parameter(s) (here $\triangle t$ only) to recover uniform stabilization properties or the convergence of discrete controls to the continuous ones. We refer to the survey [31] and the references therein for more precise statements. To our knowledge, there are very few results addressing the observability issues for time semi-discrete schemes. We refer to [19], where the uniform controllability of a fully discrete approximation scheme of the 1-d wave equation is analyzed, and to [28], where a time discretization of the wave equation is analyzed using multiplier techniques. Especially, the results in [28] may be viewed as a particular instance of the abstract models we address here.

In the sequel, we are interested in understanding under which assumptions inequality (3.1.8) holds uniformly on $\triangle t$. One expects to do it so that, when letting $\triangle t \to 0$, one recovers the observability property of the continuous model.

It can be done by means of a spectral filtering mechanism. More precisely, since A is skew-adjoint with compact resolvent, its spectrum is discrete and $\sigma(A) = \{i\mu_j : j \in \mathbb{N}\}$, where $(\mu_j)_{j\in\mathbb{N}}$ is a sequence of real numbers. Set $(\Phi_j)_{j\in\mathbb{N}}$ an orthonormal basis of eigenvectors of A associated to the eigenvalues $(i\mu_j)_{j\in\mathbb{N}}$, that is:

$$A\Phi_j = i\mu_j\Phi_j. \tag{3.1.9}$$

Moreover, we define

$$\mathcal{C}_s = \text{span}\,\{\Phi_j : \text{the corresponding } i\mu_j \text{ satisfies } |\mu_j| \leq s\}. \tag{3.1.10}$$

We will prove that inequality (3.1.8) holds uniformly (with respect to $\triangle t > 0$) in the class $\mathcal{C}_{\delta/\triangle t}$ for any $\delta > 0$ and for T_δ large enough, depending on the filtering parameter δ.

This result will be obtained as a consequence of the following theorem:

Theorem 3.1.3. *Let $\delta > 0$.*

Assume that we have a family of vector spaces $X_{\delta,\triangle t} \subset X$ and a family of unbounded operators $(A_{\triangle t}, B_{\triangle t})$ depending on the parameter $\triangle t > 0$ such that

(H1) For each $\triangle t > 0$, the operator $A_{\triangle t}$ is skew-adjoint on $X_{\delta,\triangle t}$, and the vector space $X_{\delta,\triangle t}$ is globally invariant by $A_{\triangle t}$. Moreover,

$$\|A_{\triangle t} z\|_X \leq \frac{\delta}{\triangle t} \|z\|_X , \qquad \forall z \in X_{\delta,\triangle t},\ \forall \triangle t > 0. \tag{3.1.11}$$

(H2) There exists a positive constant C_B such that

$$\|B_{\triangle t} z\|_Y \leq C_B \|A_{\triangle t} z\|_X , \qquad \forall z \in X_{\delta,\triangle t},\ \forall \triangle t > 0. \tag{3.1.12}$$

(H3) There exist two positive constants M and m such that

$$\begin{aligned} M^2 \|(A_{\triangle t} - i\omega I) z\|_X^2 + m^2 \|B_{\triangle t} z\|_Y^2 \geq \|z\|_X^2 , \\ \forall z \in X_{\delta,\triangle t} \cup \mathcal{D}(A_{\triangle t}), \forall \omega \in \mathbb{R},\ \forall \triangle t > 0. \end{aligned} \tag{3.1.13}$$

Then there exists a time T_δ such that for all time $T > T_\delta$, there exists a positive constant $k_{T,\delta}$ such that for $\triangle t > 0$ small enough, the solution of

$$\frac{z^{k+1} - z^k}{\triangle t} = A_{\triangle t}\Big(\frac{z^{k+1} + z^k}{2}\Big), \qquad in \ \ X_{\delta,\triangle t}, \quad k \in \mathbb{Z}, . \tag{3.1.14}$$

with initial data $z^0 \in X_{\delta,\triangle t}$ satisfies

$$k_{T,\delta} \left\|z^0\right\|_X^2 \leq \triangle t \sum_{k \in (0, T/\triangle t)} \left\|B_{\triangle t} z^k\right\|_Y^2 , \qquad \forall\ z^0 \in X_{\delta,\triangle t}. \tag{3.1.15}$$

Moreover, T_δ can be taken to be such that

$$T_\delta = \pi \Big[\Big(1 + \frac{\delta^2}{4}\Big)^2 M^2 + m^2 C_B^2 \frac{\delta^4}{16}\Big]^{1/2}, \tag{3.1.16}$$

where C_B is as in (3.2.1).

As we shall see in Theorem 3.2.1, taking $A_{\triangle t} = A$, $B_{\triangle t} = B$ and $X_{\delta/\triangle t} = \mathcal{C}_{\delta/\triangle t}$, Theorem 3.1.3 provides an observability result within the class $\mathcal{C}_{\delta/\triangle t}$ for system (3.1.6)-(3.1.7), as a consequence of assumption (3.1.5) and $B \in \mathfrak{L}(\mathcal{D}(A), Y)$.

Theorem 3.1.3 is also useful to address observability issues for more general time-discretization schemes of (3.1.1)-(3.1.2) than (3.1.6). For instance, one can consider time semi-discrete schemes of the form

$$z^{k+1} = \mathbb{T}_{\triangle t} z^k, \qquad y^k = B z^k, \tag{3.1.17}$$

where $\mathbb{T}_{\triangle t}$ is a linear operator with the same eigenvectors as the operator A. We will prove that, under some general assumptions on $\mathbb{T}_{\triangle t}$, inequality (3.1.8) holds uniformly on $\triangle t$ for solutions of (3.1.17) when the initial data are taken in the class $\mathcal{C}_{\delta/\triangle t}$, as we shall see in Theorem 3.3.1.

We can also consider second order in time systems such as

$$\ddot{u}(t) + A_0 u(t) = 0; \qquad u(0) = u_0, \ \ \dot{u}(0) = v_0, \tag{3.1.18}$$

where A_0 is a positive self-adjoint operator. Of course, such systems can be written in the same first-order form as (3.1.1). However, there are time-discretization schemes such as the Newmark method

which cannot be put in the form (3.1.17). Hence we present a specific analysis of the Newmark method for (3.1.18), still based on Theorem 3.1.3.

One of the interesting applications of our results, and, in particular of Theorem 3.1.3, is that they allow us to develop a two-step strategy to study the observability of fully discrete approximation schemes of (3.1.1)-(3.1.2). Roughly speaking, first, one needs to derive observability properties for *space* semi-discrete approximation schemes, uniformly with respect to the space mesh-size parameter, as it has already been done in many cases (see [4, 6, 7, 10, 20, 21, 30] and [31] for more references). Second, applying the results of this paper on time discretizations, the uniform observability (with respect to both the time and space mesh-sizes) for the *fully* discrete approximation schemes is derived. This procedure will be described in detail in Section 3.5. To our knowledge, the observability properties of fully discrete approximation schemes have been studied only in [19], in the very particular case of the 1-d wave equation. The results we present here can be applied to a much wider class of systems, time-discretization schemes, in one and several space dimensions, etc.

To complete our analysis of the discretizations of system (3.1.1)-(3.1.2), we also analyze admissibility properties for the time semi-discrete systems introduced throughout this paper. They are useful when deriving controllability results out of the observability ones. More precisely, it allows proving controllability results by means of duality arguments combined with observability and admissibility results (see for instance the textbook [16] and the survey article [31]). In particular, we prove that the admissibility inequality (3.1.3) can be interpreted in terms of the behavior of wave packets. From this wave packet estimate, we will deduce admissibility inequalities for the time semi-discrete schemes. This part can be read independently from the rest of the article.

The outline of this paper is as follows.

In Section 3.2 we prove Theorem 3.1.3, from which we deduce the uniform observability property (3.1.8) for system (3.1.6)-(3.1.7), assuming that the initial data are taken in some subspace of filtered data $\mathcal{C}_{\delta/\triangle t}$ for arbitrary $\delta > 0$. Our proof of Theorem 3.1.3 is mainly based on the resolvent estimate (3.1.13), combined with standard Fourier arguments adapted to the time-discrete setting. In Section 3.3, we show how to apply Theorem 3.1.3 to obtain similar results for time semi-discrete approximation schemes such as (3.1.17) and the Newmark approximation schemes, for which we prove that a uniform observability inequality holds as well, provided the initial data belong to $\mathcal{C}_{\delta/\triangle t}$. In Section 3.4, we give some applications to the observability of some classical conservative equations, such as the Schrödinger equation or the linearized KdV equation, etc. In Section 3.5, we give some applications of our main results to fully discrete schemes for skew-adjoint systems as (3.1.1). In Section 3.6, we present admissibility results similar to (3.1.3) for the time semi-discrete schemes used along the article. We end the paper by stating some further comments and open problems.

3.2 The implicit mid-point scheme

In this section we show the uniform observability of system (3.1.6)-(3.1.7), which can be seen as a direct consequence of Theorem 3.1.3. In other words, its proof is a simplified version of the one of Theorem 3.1.3. To avoid the duplication of the process, we only give the proof of the latter one, which is more general.

Let us first introduce some notations and definitions.

The Hilbert space $\mathcal{D}(A)$ is endowed with the norm of the graph of A, which is equivalent to $\|A\cdot\|$

since A has a compact resolvent. It follows that $B \in \mathfrak{L}(\mathcal{D}(A), Y)$ implies

$$\|Bz\|_Y \leq C_B \|Az\|_X, \qquad \forall z \in \mathcal{D}(A). \tag{3.2.1}$$

We are now in position to claim the following theorem based on the resolvent estimate (3.1.5):

Theorem 3.2.1. *Assume that* (A, B) *satisfy* (3.1.5) *and that* $B \in \mathfrak{L}(\mathcal{D}(A), Y)$.

Then, for any $\delta > 0$, *there exists* T_δ *such that for any* $T > T_\delta$, *there exists a positive constant* $k_{T,\delta}$, *independent of* $\triangle t$, *such that for* $\triangle t > 0$ *small enough, the solution* z^k *of* (3.1.6) *satisfies*

$$k_{T,\delta} \left\|z^0\right\|_X^2 \leq \triangle t \sum_{k \in (0, T/\triangle t)} \left\|Bz^k\right\|_Y^2, \qquad \forall\ z^0 \in \mathcal{C}_{\delta/\triangle t}. \tag{3.2.2}$$

Moreover, T_δ *can be taken to be such that*

$$T_\delta = \pi \Big[M^2 \Big(1 + \frac{\delta^2}{4}\Big)^2 + m^2 C_B^2 \frac{\delta^4}{16} \Big]^{1/2}, \tag{3.2.3}$$

where C_B *is as in* (3.2.1).

Remark 3.2.2. If we filter at a scale smaller than $\triangle t$, for instance in the class $\mathcal{C}_{\delta/(\triangle t)^\alpha}$, with $\alpha < 1$, then δ in (3.2.3) vanishes as $\triangle t$ tends to zero. In that case the uniform observability time T_0 we obtain is $T_0 = \pi M$, which coincides with the time obtained by the resolvent estimate (3.1.5) in the continuous setting (see [18]). Note that, however, even in the continuous setting, in general πM is not the optimal observability time.

Proof of Theorem 3.2.1. Theorem 3.2.1 can be seen as a direct consequence of Theorem 3.1.3, which will be proved below. Indeed, one can easily verify that (H1)–(H3) hold by taking $A_{\triangle t} = A$, $B_{\triangle t} = B$ and $X_{\delta, \triangle t} = \mathcal{C}_{\delta/\triangle t}$. □

Before getting into the proof of Theorem 3.1.3, let us first introduce the discrete Fourier transform at scale $\triangle t$, which is one of the main ingredients of the proof of Theorem 3.1.3.

Definition 3.2.3. Given any sequence $(u^k) \in l^2(\triangle t \mathbb{Z})$, we define its Fourier transform as:

$$\hat{u}(\tau) = \triangle t \sum_{k \in \mathbb{Z}} u^k \exp(-i\tau k \triangle t), \quad \tau \triangle t \in (-\pi, \pi]. \tag{3.2.4}$$

For any function $v \in L^2(-\pi/\triangle t, \pi/\triangle t)$, we define the inverse Fourier transform at scale $\triangle t > 0$:

$$\tilde{v}^k = \frac{1}{2\pi} \int_{-\pi/\triangle t}^{\pi/\triangle t} v(\tau) \exp(i\tau k \triangle t)\ d\tau, \quad k \in \mathbb{Z}. \tag{3.2.5}$$

According to Definition 3.2.3,

$$\tilde{\hat{u}} = u, \quad \hat{\tilde{v}} = v, \tag{3.2.6}$$

and the Parseval identity holds

$$\frac{1}{2\pi} \int_{-\pi/\triangle t}^{\pi/\triangle t} |\hat{u}(\tau)|^2\ d\tau = \triangle t \sum_{k \in \mathbb{Z}} |u^k|^2. \tag{3.2.7}$$

These properties will be used in the sequel.

Proof of Theorem 3.1.3. The proof is split into three parts.

Step 1: Estimates in the class $X_{\delta,\triangle t}$. Let us take $z^0 \in X_{\delta,\triangle t}$. Then the solution of (3.1.14) has constant norm since $A_{\triangle t}$ is skew-adjoint (see (H1)). Indeed,

$$z^{k+1} = \Big(\frac{I + \frac{\triangle t}{2}A_{\triangle t}}{I - \frac{\triangle t}{2}A_{\triangle t}}\Big)z^k := \mathbb{T}_{\triangle t}z^k,$$

where the operator $\mathbb{T}_{\triangle t}$ is obviously unitary.

Further, since

$$\frac{z^k + z^{k+1}}{2} = \frac{1}{2}\Big(I + \mathbb{T}_{\triangle t}\Big)z^k = \Big(\frac{I}{I - \frac{\triangle t}{2}A_{\triangle t}}\Big)z^k,$$

we get that for any k,

$$\Big\|\frac{z^0 + z^1}{2}\Big\|_X^2 = \Big\|\frac{z^k + z^{k+1}}{2}\Big\|_X^2 \geq \frac{1}{1 + \Big(\frac{\delta}{2}\Big)^2}\left\|z^0\right\|_X^2, \tag{3.2.8}$$

as a consequence of (3.1.11) and the skew-adjointness assumption (H1) of $A_{\triangle t}$.

Step 2: The resolvent estimate. Set $\chi \in H^1(\mathbb{R})$ and $\chi^k = \chi(k\triangle t)$. Let $g^k = \chi^k z^k$, and

$$f^k = \frac{g^{k+1} - g^k}{\triangle t} - A_{\triangle t}\Big(\frac{g^{k+1} + g^k}{2}\Big). \tag{3.2.9}$$

One can easily check that

$$\begin{aligned} f^k &= \frac{\chi^{k+1} - \chi^k}{\triangle t}\frac{z^{k+1} + z^k}{2} + \frac{\chi^{k+1} + \chi^k}{2}\frac{z^{k+1} - z^k}{\triangle t} \\ &\qquad -A_{\triangle t}\Big(\frac{\chi^{k+1} + \chi^k}{2}\frac{z^{k+1} + z^k}{2} + \frac{\chi^{k+1} - \chi^k}{2}\frac{z^{k+1} - z^k}{2}\Big) \\ &= \Big(\frac{\chi^{k+1} - \chi^k}{\triangle t}\Big)\Big(\frac{z^k + z^{k+1}}{2} - \frac{(\triangle t)^2}{4}A_{\triangle t}\Big(\frac{z^{k+1} - z^k}{\triangle t}\Big)\Big) \\ &= \Big(\frac{\chi^{k+1} - \chi^k}{\triangle t}\Big)\Big(I - \frac{(\triangle t)^2}{4}A_{\triangle t}^2\Big)\Big(\frac{z^k + z^{k+1}}{2}\Big). \end{aligned} \tag{3.2.10}$$

Especially, recalling (3.2.8) and (3.1.11), (3.2.10) implies

$$\Big\|f^k\Big\|_X^2 \leq \Big(\frac{\chi^{k+1} - \chi^k}{\triangle t}\Big)^2\Big\|\frac{z^0 + z^1}{2}\Big\|_X^2\Big(1 + \frac{\delta^2}{4}\Big). \tag{3.2.11}$$

In particular, $f^k \in l^2(\triangle t\mathbb{Z}; X)$.

Taking the Fourier transform of (3.2.9), for all $\tau \in (-\pi/\triangle t, \pi/\triangle t)$, we get

$$\begin{aligned} \hat{f}(\tau) &= \triangle t\sum_{k\in\mathbb{Z}} f^k \exp(-ik\triangle t\tau) \\ &= \triangle t\sum_{k\in\mathbb{Z}}\Big(\frac{g^{k+1} - g^k}{\triangle t} - A_{\triangle t}\Big(\frac{g^{k+1} + g^k}{2}\Big)\Big)\exp(-ik\triangle t\tau) \\ &= \triangle t\sum_{k\in\mathbb{Z}}\Big(\frac{\exp(i\triangle t\tau) - 1}{\triangle t} - A_{\triangle t}\Big(\frac{\exp(i\triangle t\tau) + 1}{2}\Big)\Big)g^k\exp(-ik\triangle t\tau) \\ &= \Big(i\frac{2}{\triangle t}\tan\Big(\frac{\tau\triangle t}{2}\Big)I - A_{\triangle t}\Big)\hat{g}(\tau)\exp\Big(i\frac{\tau\triangle t}{2}\Big)\cos\Big(\frac{\tau\triangle t}{2}\Big). \end{aligned} \tag{3.2.12}$$

We claim the following Lemma:

Lemma 3.2.4. *l The solution* (z^k) *in* (3.1.14) *satisfies*

$$(1+\alpha)m^2\triangle t\sum_{k\in\mathbb{Z}}\Big(\frac{\chi^k+\chi^{k+1}}{2}\Big)^2\Big\|B_{\triangle t}\Big(\frac{z^k+z^{k+1}}{2}\Big)\Big\|_Y^2$$
$$\geq\Big\|\frac{z^0+z^1}{2}\Big\|_X^2\left[a_1\triangle t\sum_{k\in\mathbb{Z}}\Big(\frac{\chi^k+\chi^{k+1}}{2}\Big)^2-a_2\triangle t\sum_{k\in\mathbb{Z}}\Big(\frac{\chi^{k+1}-\chi^k}{\triangle t}\Big)^2\right], \quad (3.2.13)$$

with

$$a_1=\Big(1-\frac{1}{\beta}\Big),\qquad a_2=M^2\Big(1+\frac{\delta^2}{4}\Big)^2+m^2C_B^2\Big(1+\frac{1}{\alpha}\Big)\frac{\delta^4}{16}+\frac{(\triangle t)^2}{16}\delta^2(\beta-1), \quad (3.2.14)$$

for any $\alpha>0$ *and* $\beta>1$, *where* C_B, M, m *are as in* (3.1.12)-(3.1.13).

Proof of Lemma 3.2.4. Let

$$G(\tau)=\hat{g}(\tau)\exp(i\frac{\tau\triangle t}{2})\cos(\frac{\tau\triangle t}{2}). \quad (3.2.15)$$

By its definition and the fact that $z^k\in X_{\delta,\triangle t}$, it is obvious that $G(\tau)\in X_{\delta,\triangle t}$.

In view of (3.2.12), applying the resolvent estimate (3.1.13) to $G(\tau)$, integrating on τ from $-\pi/\triangle t$ to $\pi/\triangle t$, it holds

$$M^2\int_{-\pi/\triangle t}^{\pi/\triangle t}\Big\|\hat{f}(\tau)\Big\|_X^2\,d\tau+m^2\int_{-\pi/\triangle t}^{\pi/\triangle t}\|B_{\triangle t}G(\tau)\|_Y^2\,d\tau\geq\int_{-\pi/\triangle t}^{\pi/\triangle t}\|G(\tau)\|_X^2\,d\tau. \quad (3.2.16)$$

Applying Parseval's identity (3.2.7) to (3.2.16), and noticing that

$$\tilde{G}^k=\frac{g^k+g^{k+1}}{2},\quad i.e.\ \ G(\tau)=\Big(\widehat{\frac{g^k+g^{k+1}}{2}}\Big)(\tau),$$

we get

$$M^2\triangle t\sum_{k\in\mathbb{Z}}\Big\|f^k\Big\|_X^2+m^2\triangle t\sum_{k\in\mathbb{Z}}\Big\|B_{\triangle t}\Big(\frac{g^k+g^{k+1}}{2}\Big)\Big\|_Y^2\geq\triangle t\sum_{k\in\mathbb{Z}}\Big\|\frac{g^k+g^{k+1}}{2}\Big\|_X^2. \quad (3.2.17)$$

Now we estimate the three terms in (3.2.17). The first term can be bounded above in view of (3.2.11).

Second, since

$$\frac{g^{k+1}+g^k}{2}=\Big(\frac{\chi^{k+1}+\chi^k}{2}\Big)\Big(\frac{z^{k+1}+z^k}{2}\Big)+\frac{\triangle t}{2}\Big(\frac{\chi^{k+1}-\chi^k}{\triangle t}\Big)\Big(\frac{z^{k+1}-z^k}{2}\Big), \quad (3.2.18)$$

using

$$\|a+b\|^2\leq(1+\alpha)\|a\|^2+\Big(1+\frac{1}{\alpha}\Big)\|b\|^2,$$

we deduce that

$$\begin{aligned}\Big\|B_{\triangle t}\Big(\frac{g^{k+1}+g^k}{2}\Big)\Big\|_Y^2 &\le (1+\alpha)\Big(\frac{\chi^{k+1}+\chi^k}{2}\Big)^2\Big\|B_{\triangle t}\Big(\frac{z^{k+1}+z^k}{2}\Big)\Big\|_Y^2\\ &\quad+\Big(1+\frac{1}{\alpha}\Big)\frac{(\triangle t)^4}{16}\Big(\frac{\chi^{k+1}-\chi^k}{\triangle t}\Big)^2\Big\|B_{\triangle t}\Big(\frac{z^{k+1}-z^k}{\triangle t}\Big)\Big\|_Y^2\\ &\le (1+\alpha)\Big(\frac{\chi^{k+1}+\chi^k}{2}\Big)^2\Big\|B_{\triangle t}\Big(\frac{z^{k+1}+z^k}{2}\Big)\Big\|_Y^2\\ &\quad+\Big(1+\frac{1}{\alpha}\Big)\frac{\delta^4}{16}C_B^2\Big(\frac{\chi^{k+1}-\chi^k}{\triangle t}\Big)^2\Big\|\frac{z^0+z^1}{2}\Big\|_X^2.\end{aligned}\tag{3.2.19}$$

In (3.2.19) we use the fact that (recalling (3.1.11) and (3.1.12))

$$\Big\|B_{\triangle t}A_{\triangle t}\Big(\frac{z^k+z^{k+1}}{2}\Big)\Big\|_Y\le C_B\Big\|A_{\triangle t}^2\Big(\frac{z^k+z^{k+1}}{2}\Big)\Big\|_X\le\frac{\delta^2C_B}{(\triangle t)^2}\Big\|\frac{z^0+z^1}{2}\Big\|_X.$$

Finally, for any $\beta>1$, recalling (3.2.8), (3.1.11) and (3.2.18), we get

$$\begin{aligned}\Big\|\frac{g^{k+1}+g^k}{2}\Big\|_X^2 &\ge \Big(1-\frac{1}{\beta}\Big)\Big(\frac{\chi^{k+1}+\chi^k}{2}\Big)^2\Big\|\frac{z^{k+1}+z^k}{2}\Big\|_X^2\\ &\quad-(\beta-1)\Big(\frac{\triangle t}{2}\Big)^2\Big(\frac{\chi^{k+1}-\chi^k}{\triangle t}\Big)^2\Big\|\frac{z^{k+1}-z^k}{2}\Big\|_X^2\\ &\ge \Big(1-\frac{1}{\beta}\Big)\Big(\frac{\chi^{k+1}+\chi^k}{2}\Big)^2\Big\|\frac{z^0+z^1}{2}\Big\|_X^2\\ &\quad-(\beta-1)\Big(\frac{\triangle t}{2}\Big)^4\Big(\frac{\chi^{k+1}-\chi^k}{\triangle t}\Big)^2\Big\|A_{\triangle t}\Big(\frac{z^0+z^1}{2}\Big)\Big\|_X^2\\ &\ge \Big(1-\frac{1}{\beta}\Big)\Big(\frac{\chi^{k+1}+\chi^k}{2}\Big)^2\Big\|\frac{z^0+z^1}{2}\Big\|_X^2\\ &\quad-(\beta-1)\Big(\frac{\delta\triangle t}{4}\Big)^2\Big(\frac{\chi^{k+1}-\chi^k}{\triangle t}\Big)^2\Big\|\Big(\frac{z^0+z^1}{2}\Big)\Big\|_X^2,\end{aligned}\tag{3.2.20}$$

where we used

$$\|a+b\|^2\ge\Big(1-\frac{1}{\beta}\Big)\|a\|^2-\Big(\beta-1\Big)\|b\|^2.$$

Applying (3.2.11), (3.2.19) and (3.2.20) to (3.2.17), we complete the proof of Lemma 3.2.4. □

Step 3: The observability estimate. This step is aimed to derive the observability estimate (3.1.15) stated in Theorem 3.1.3 from Lemma 3.2.4 with explicit estimates on the optimal time T_δ.

First of all, let us recall the following classical Lemma on Riemann sums:

Lemma 3.2.5. *Let $\chi(t)=\phi(t/T)$ with $\phi\in H^2\cap H_0^1(0,1)$, extended by zero outside $(0,T)$. Recalling that $\chi^k=\chi(k\triangle t)$, the following estimates hold:*

$$\begin{aligned}&\Big|\triangle t\sum_{k\in\mathbb{Z}}\Big(\frac{\chi^k+\chi^{k+1}}{2}\Big)^2-T\|\phi\|_{L^2(0,1)}^2\Big|\le 2T\triangle t\|\phi\|_{L^2(0,1)}\Big\|\dot\phi\Big\|_{L^2(0,1)},\\ &\Big|\triangle t\sum_{k\in\mathbb{Z}}\Big(\frac{\chi^{k+1}-\chi^k}{\triangle t}\Big)^2-\frac{1}{T}\Big\|\dot\phi\Big\|_{L^2(0,1)}^2\Big|\le\frac{2}{T}\triangle t\Big\|\dot\phi\Big\|_{L^2(0,1)}\Big\|\ddot\phi\Big\|_{L^2(0,1)}.\end{aligned}\tag{3.2.21}$$

Sketch of the proof of Lemma 3.2.5. It is easy to show that for all $f = f(t) \in C^1(0,T)$ and sequence $\tau_k \in [k\triangle t, (k+1)\triangle t]$, it holds

$$\begin{aligned}\Big| \int_0^T f(t)dt - \triangle t \sum_{k\in(0,T/\triangle t)} f(\tau_k)\Big| &\leq \sum_{k\in(0,T/\triangle t)} \int\int_{[k\triangle t,(k+1)\triangle t]^2} |\dot f(s)|\ ds\ dt \\ &\leq \triangle t \int_0^T |\dot f|\ dt. \end{aligned} \tag{3.2.22}$$

Replacing f by ϕ^2 we get the first inequality (3.2.21). Similarly, replacing f by $\dot\phi^2$, the second one can be proved too. □

Taking Lemma 3.2.4 and 3.2.5 into account, the coefficient of $\big\|(z^0+z^1)/2\big\|_X^2$ in (3.2.13) tends to

$$\begin{aligned}k_{T,\delta,\alpha,\beta,\phi} = \frac{1}{m^2(1+\alpha)}\Big[\Big(1-\frac{1}{\beta}\Big)T\,\|\phi\|^2_{L^2(0,1)} \\ - \Big(M^2\Big(1+\frac{\delta^2}{4}\Big)^2 + m^2C_B^2\Big(1+\frac{1}{\alpha}\Big)\frac{\delta^4}{16}\Big)\frac{1}{T}\left\|\dot\phi\right\|^2_{L^2(0,1)}\Big],\end{aligned} \tag{3.2.23}$$

when $\triangle t \to 0$.

Note that $k_{T,\delta,\alpha,\beta,\phi}$ is an increasing function of T tending to $-\infty$ when $T \to 0^+$ and to $+\infty$ when $T \to \infty$. Let $T_{\delta,\alpha,\beta,\phi}$ be the unique positive solution of $k_{T,\delta,\alpha,\beta,\phi} = 0$. Then, for any time $T > T_{\delta,\alpha,\beta,\phi}$, choosing a positive $k_{T,\delta}$ such that

$$0 < k_{T,\delta} < k_{T,\delta,\alpha,\beta,\phi},$$

there exists $\triangle t_0 > 0$ such that for any $\triangle t < \triangle t_0$, the following holds:

$$k_{T,\delta}\left\|\frac{z^0+z^1}{2}\right\|_X^2 \leq \triangle t \sum_{k\in(0,T/\triangle t)} \left\|B_{\triangle t}\Big(\frac{z^k+z^{k+1}}{2}\Big)\right\|_Y^2. \tag{3.2.24}$$

This combined with (3.2.8) yields (3.1.15).

This construction yields the following estimate on the time T_δ in Theorem 3.1.3. Namely, for any $\alpha > 0$, $\beta > 1$ and smooth function ϕ, compactly supported in $[0,1]$:

$$T_\delta \leq \frac{\left\|\dot\phi\right\|_{L^2}}{\|\phi\|_{L^2}}\Big[\frac{\beta}{\beta-1}\Big]^{1/2}\Big[M^2\Big(1+\frac{\delta^2}{4}\Big)^2 + m^2C_B^2\Big(1+\frac{1}{\alpha}\Big)\frac{\delta^4}{16}\Big]^{1/2}.$$

We optimize in α, β and ϕ by choosing $\alpha = \infty$, $\beta = \infty$ and

$$\phi(t) = \begin{cases} \sin(\pi t), & t \in (0,1) \\ 0, & \text{elsewhere,}\end{cases} \tag{3.2.25}$$

which is well-known to minimize the ratio

$$\frac{\left\|\dot\phi\right\|_{L^2}}{\|\phi\|_{L^2}}.$$

For this choice of ϕ, this quotient equals π, and thus we recover the estimate (3.1.16). This completes the proof of Theorem 3.1.3. □

Theorem 3.2.1 has many applications. Indeed, it roughly says that, for any continuous conservative system, which is observable in finite time, there exists a time semi-discretization which uniformly preserves the observability property in finite time, provided the initial data are filtered at a scale $1/\triangle t$. Later, using formally some microlocal tools, we will explain why this filtering scale is the optimal one. Note that in Theorem 7.1 of [28] this scale was proved to be optimal for a particular time-discretization scheme on the wave equation.

Besides, as we will see in Section 3.3, Theorem 3.1.3 is a key ingredient to address observability issues.

3.3 General time-discrete schemes

3.3.1 General time-discrete schemes for first order systems

In this section, we deal with more general time-discretization schemes of the form (3.1.17). We will show that, under some appropriate assumptions on the operator $\mathbb{T}_{\triangle t}$, inequality (3.1.8) holds uniformly on $\triangle t$ for solutions of (3.1.17) when the initial data are taken in the class $\mathcal{C}_{\delta/\triangle t}$.

More precisely, we assume that (3.1.17) is conservative in the sense that there exist real numbers $\lambda_{j,\triangle t}$ such that

$$\mathbb{T}_{\triangle t}\Phi_j = \exp(i\lambda_{j,\triangle t}\triangle t)\Phi_j. \tag{3.3.1}$$

Moreover, we assume that there is an explicit relation between $\lambda_{j,\triangle t}$ and μ_j (as in (3.1.9)) of the following form:

$$\lambda_{j,\triangle t} = \frac{1}{\triangle t}\, h(\mu_j \triangle t), \tag{3.3.2}$$

where $h : (-R, R) \mapsto [-\pi, \pi]$ is a smooth strictly increasing function, with $R \in (0, \infty]$, i.e.

$$|h(\eta)| \leq \pi, \qquad \inf\{h'(\eta),\ |\eta| \leq \delta\} > 0; \qquad \forall \delta < R. \tag{3.3.3}$$

The parameter R corresponds to a frequency limit $R/\triangle t$ imposed by the discretization scheme, see for instance the example given in Subsection 3.4.2. Roughly speaking, the first part of (3.3.3) reflects the fact that one cannot measure frequencies higher than $\pi/\triangle t$ in a mesh of size $\triangle t$. The second part is a non-degeneracy condition on the group velocity (see [25]) of solutions of (3.1.17) which is necessary to guarantee the propagation of solutions that is required for observability to hold.

We also assume

$$\frac{h(\eta)}{\eta} \longrightarrow 1 \quad \text{as} \quad \eta \to 0. \tag{3.3.4}$$

This guarantees the consistency of the time-discrete scheme with the continuous model (3.1.1).

We have the following Theorem:

Theorem 3.3.1. *Assume that (A, B) satisfy (3.1.5) and that $B \in \mathfrak{L}(\mathcal{D}(A), Y)$.*

Under assumptions (3.3.1), (3.3.2), (3.3.3) and (3.3.4), for any $\delta \in (0, R)$, there exists a time T_δ such that for all $T > T_\delta$, there exists a constant $k_{T,\delta} > 0$ such that for all $\triangle t > 0$ small enough, any solution of (3.1.17) with initial value $z^0 \in \mathcal{C}_{\delta/\triangle t}$ satisfies

$$k_{T,\delta} \left\|z^0\right\|_X^2 \leq \triangle t \sum_{k \in (0, T/\triangle t)} \left\| B\Big(\frac{z^k + z^{k+1}}{2}\Big)\right\|_Y^2. \tag{3.3.5}$$

Besides, we have the following estimate on T_δ:

$$T_\delta \leq \pi \Bigg[M^2 \Big(1 + \tan^2\Big(\frac{h(\delta)}{2}\Big)\Big)^2 \sup_{|\eta|\leq\delta} \Big\{ \frac{\cos^4(h(\eta)/2)}{h'(\eta)^2} \Big\} + m^2 C_B^2 \sup_{|\eta|\leq\delta} \Big\{ \frac{2}{\eta} \tan\Big(\frac{h(\eta)}{2}\Big) \Big\}^2 \tan^4\Big(\frac{h(\delta)}{2}\Big) \Bigg]^{1/2}, \quad (3.3.6)$$

where C_B is as in (3.2.1).

Proof. The main idea is to use Theorem 3.1.3. Hence we introduce an operator $A_{\triangle t}$ such that the solution of (3.1.17) with $z_0 \in \mathcal{C}_{R/\triangle t}$ coincides with the solution of the linear system

$$\frac{z^{k+1} - z^k}{\triangle t} = A_{\triangle t}\Big(\frac{z^k + z^{k+1}}{2}\Big), \quad z^0 = z_0. \quad (3.3.7)$$

This can be done defining the action of the operator $A_{\triangle t}$ on each eigenfunction:

$$A_{\triangle t}\Phi_j = ik_{\triangle t}(\mu_j)\Phi_j, \quad (3.3.8)$$

where

$$k_{\triangle t}(\omega) = \frac{2}{\triangle t} \tan\Big(\frac{h(\omega\triangle t)}{2}\Big). \quad (3.3.9)$$

Indeed, if

$$z_0 = \sum a_j \Phi_j,$$

then the solution of (3.1.17) can be written as

$$z^k = \sum a_j \phi_j \exp(i\lambda_j k \triangle t) = \sum a_j \phi_j \exp(ih(\mu_j \triangle t)k)$$

and the definition of $A_{\triangle t}$ follows naturally.

Obviously, when the scheme (3.1.17) under consideration is the one of Section 3.2, that is (3.1.6), the operator $A_{\triangle t}$ is precisely the operator A.

Then (3.3.5) would be a straightforward consequence of Theorem 3.1.3, if we could prove the resolvent estimate for $A_{\triangle t}$. We will see in the sequel that a weak form of the resolvent estimate holds, and that this is actually sufficient to get the desired observability inequality. In the sequel, δ is a given positive number, determining the class of filtered data under consideration.

Step 1: A weak form of the resolvent estimate. By hypothesis (3.1.5),

$$M^2 \|(A - i\omega)z\|_X^2 + m^2 \|Bz\|_Y^2 \geq \|z\|_X^2, \quad z \in \mathcal{D}(A),\ \omega \in \mathbb{R}. \quad (3.3.10)$$

For $z \in \mathcal{C}_{\delta/\triangle t}$, that is

$$z = \sum_{|\mu_j| \leq \delta/\triangle t} a_j \phi_j, \quad (3.3.11)$$

one can easily check that

$$\|(A - i\omega)z\|_X^2 = \sum |a_j|^2 \Big(\mu_j - \omega\Big)^2$$

and

$$\|(A_{\triangle t}-i\omega)z\|_X^2=\sum|a_j|^2\Big(k_{\triangle t}(\mu_j)-\omega\Big)^2.$$

Especially, for any $\omega\in\mathbb{R}$, this last estimate takes the form

$$\|(A_{\triangle t}-ik_{\triangle t}(\omega))z\|_X^2=\sum|a_j|^2\Big(k_{\triangle t}(\mu_j)-k_{\triangle t}(\omega)\Big)^2$$

with $k_{\triangle t}$ as in (3.3.9). Thus, taking $\varepsilon>0$, it follows that for any $\omega<(\delta+\varepsilon)/\triangle t$,

$$\|(A_{\triangle t}-ik_{\triangle t}(\omega))z\|_X^2\geq\Big(\inf_{|\omega|\triangle t\leq\delta+\varepsilon}\Big\{|k'_{\triangle t}(\omega)|\Big\}\Big)^2\|(A-i\omega)z\|_X^2\,.$$

Hence, setting

$$\alpha_{\triangle t,\varepsilon}=k_{\triangle t}\Big(\frac{\delta+\varepsilon}{\triangle t}\Big),\quad C_{\delta,\varepsilon}=\Big(\inf\{k'_{\triangle t}(\omega)\ :\ |\omega|\triangle t\leq\delta+\varepsilon\}\Big)^{-1},\tag{3.3.12}$$

which is finite in view of (3.3.3), we get the following weak resolvent estimate:

$$C_{\delta,\varepsilon}^2M^2\left\|\Big(A_{\triangle t}-i\omega\Big)z\right\|_X^2+m^2\|Bz\|_Y^2\geq\|z\|_X^2\,,\quad z\in\mathcal{C}_{\delta/\triangle t},\ |\omega|\leq\alpha_{\triangle t,\varepsilon}.\tag{3.3.13}$$

Our purpose is now to show that this is enough to get the time-discrete observability estimate. We emphasize that the main difference between (3.3.13) and (3.1.13) is that (3.1.13) is assumed to hold for all $\omega\in\mathbb{R}$ while (3.3.13) only holds for $|\omega|\leq\alpha_{\triangle t,\varepsilon}$.

Step 2: Improving the resolvent estimate (3.3.13)**.** Here we prove that (3.3.13) can be extended to all $\omega\in\mathbb{R}$. Indeed, consider ω such that $|\omega|\geq\alpha_{\triangle t,\varepsilon}$ and $z\in\mathcal{C}_{\delta/\triangle t}$ as in (3.3.11). Then

$$\begin{aligned}\|(A_{\triangle t}-i\omega)z\|_X^2&\geq\sum_{|\mu_j|\leq\delta/\triangle t}\Big(k_{\triangle t}(\mu_j)-k_{\triangle t}\Big(\frac{\delta+\varepsilon}{\triangle t}\Big)\Big)^2a_j^2\\&\geq\sum_{|\mu_j|\leq\delta/\triangle t}\Big(k_{\triangle t}\Big(\frac{\delta}{\triangle t}\Big)-k_{\triangle t}\Big(\frac{\delta+\varepsilon}{\triangle t}\Big)\Big)^2a_j^2\\&\geq\Big(\frac{\varepsilon}{\triangle t}\Big)^2\Big(\inf_{\omega\triangle t\in[\delta,\delta+\varepsilon]}k'_{\triangle t}(\omega)\Big)^2\|z\|^2\,.\end{aligned}$$

Using the explicit expression (3.3.9) of $k_{\triangle t}$, we get

$$\|(A_{\triangle t}-i\omega)z\|_X^2\geq\Big(\frac{\varepsilon}{\triangle t}\Big)^2\inf_{\eta\in[\delta,\delta+\varepsilon]}\{h'(\eta)\}^2\|z\|^2\,.\tag{3.3.14}$$

Therefore, for each $\varepsilon>0$, in view of (3.3.3) and (3.3.12), there exists $(\triangle t)_\varepsilon>0$ such that, for $\triangle t\leq(\triangle t)_\varepsilon$,

$$C_{\delta,\varepsilon}^2M^2\left\|\Big(A_{\triangle t}-i\omega\Big)z\right\|_X^2+m^2\|Bz\|_Y^2\geq\|z\|_X^2\,,\quad z\in\mathcal{C}_{\delta/\triangle t},\ \omega\in\mathbb{R}.\tag{3.3.15}$$

Step 3: Application of Theorem 3.1.3. First, one easily checks from (3.3.8)-(3.3.9) that

$$\triangle t\,\|A_{\triangle t}z\|_X\leq\tilde{\delta}\,\|z\|_X\,,\qquad z\in\mathcal{C}_{\delta/\triangle t},\tag{3.3.16}$$

with $\tilde{\delta}=2\tan(h(\delta)/2)$.

Second, we check that there exists a constant $C_{B,\delta}$ such that

$$\|Bz\|_Y \leq C_{B,\delta} \|A_{\triangle t} z\|_X , \qquad z \in \mathcal{C}_{\delta/\triangle t}, \tag{3.3.17}$$

where C_B is as in (3.2.1). Indeed, for $z \in \mathcal{C}_{\delta/\triangle t}$,

$$\|Az\|_X \leq \sup_{|\omega|\triangle t \leq \delta} \Big\{ \Big| \frac{k_{\triangle t}(\omega)}{\omega} \Big| \Big\} \|A_{\triangle t} z\|_X ,$$

and therefore one can take

$$C_{B,\delta} = \beta_\delta C_B, \tag{3.3.18}$$

where

$$\beta_\delta = \sup_{|\eta| \leq \delta} \Big\{ \frac{2}{\eta} \tan\Big(\frac{h(\eta)}{2} \Big) \Big\},$$

which is finite from hypothesis (3.3.3) and (3.3.4).

Third, the resolvent estimate (3.3.15) holds.

Then Theorem 3.1.3 can be applied and proves the observability inequality (3.3.5) for the solutions of (3.1.17) with initial data in $\mathcal{C}_{\delta/\triangle t}$. Besides, we have the following estimate on the observability time $T_{\delta,\varepsilon}$:

$$T_{\delta,\varepsilon} = \pi \Big[\Big(1 + \frac{\tilde{\delta}^2}{4} \Big)^2 M^2 C_{\delta,\varepsilon}^2 + m^2 C_B^2 \beta_\delta^2 \frac{\tilde{\delta}^4}{16} \Big]^{1/2}.$$

In the limit $\varepsilon \to 0$, $T_{\delta,\varepsilon}$ converges to an admissible observability time $T_{\delta,0}$. Besides, using the explicit form of the constants $C_{\delta,\varepsilon}, \tilde{\delta}$ and β_δ one gets (3.3.6). □

3.3.2 The Newmark method for second order in time systems

In this subsection we investigate observability properties for time-discrete schemes for the second order in time evolution equation (3.1.18).

Let H be a Hilbert space endowed with the norm $\|\cdot\|_H$ and let $A_0 : \mathcal{D}(A_0) \to H$ be a self-adjoint positive operator with compact resolvent. We consider the initial value problem (3.1.18), which can be seen as a generic model for the free vibrations of elastic structures such as strings, beams, membranes, plates or three-dimensional elastic bodies.

The energy of (3.1.18) is given by

$$E(t) = \|\dot{u}(t)\|_H^2 + \left\| A_0^{1/2} u(t) \right\|_H^2 , \tag{3.3.19}$$

which is constant in time.

We consider the output function

$$y(t) = B_1 u(t) + B_2 \dot{u}(t), \tag{3.3.20}$$

where B_1 and B_2 are two observation operators satisfying $B_1 \in \mathfrak{L}(\mathcal{D}(A_0), Y)$ and $B_2 \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$. In other words, we assume that there exist two constants $C_{B,1}$ and $C_{B,2}$, such that

$$\|B_1 u\|_Y \leq C_{B,1} \|A_0 u\|_H , \qquad \|B_2 v\|_Y \leq C_{B,2} \left\| A_0^{1/2} v \right\| . \tag{3.3.21}$$

In the sequel, we assume either $B_1 = 0$ or $B_2 = 0$. This assumption is needed for technical reasons, as we shall see in Remark 3.3.3 and in the proof of Theorem 3.3.2.

System (3.1.18)–(3.3.20) can be put in the form (3.1.1)–(3.1.2). Indeed, setting

$$z_1(t) = \dot{u} + iA_0^{1/2}u, \qquad z_2(t) = \dot{u} - iA_0^{1/2}u, \tag{3.3.22}$$

equation (3.1.18) is equivalent to

$$\dot{z} = Az, \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad A = \begin{pmatrix} iA_0^{1/2} & 0 \\ 0 & -iA_0^{1/2} \end{pmatrix}, \tag{3.3.23}$$

for which the energy space is $X = H \times H$ with the domain $\mathcal{D}(A) = \mathcal{D}(A_0^{1/2}) \times \mathcal{D}(A_0^{1/2})$. Moreover, the energy $E(t)$ given in (3.3.19) coincides with half of the norm of z in X.

Note that the spectrum of A is explicitly given by the spectrum of A_0. Indeed, if $(\mu_j^2)_{j\in\mathbb{N}^*}$ $(\mu_j > 0)$ is the sequence of eigenvalues of A_0, i.e.

$$A_0\phi_j = \mu_j^2\phi_j, \qquad j \in \mathbb{N}^*,$$

with corresponding eigenvectors ϕ_j, then the eigenvalues of A are $\pm i\mu_j$, with corresponding eigenvectors

$$\Phi_j = \begin{pmatrix} \phi_j \\ 0 \end{pmatrix}, \qquad \Phi_{-j} = \begin{pmatrix} 0 \\ \phi_j \end{pmatrix}, \qquad j \in \mathbb{N}^*. \tag{3.3.24}$$

Besides, in the new variables (3.3.22), the output function is given by

$$y(t) = Bz(t) = B_1A_0^{-1/2}\Big(\frac{iz_2(t) - iz_1(t)}{2}\Big) + B_2\Big(\frac{z_1(t) + z_2(t)}{2}\Big). \tag{3.3.25}$$

Recalling the assumptions on B_1 and B_2 in (3.3.21), the admissible observation B belongs to $\mathfrak{L}(\mathcal{D}(A), Y)$.

In the sequel, we assume that the system (3.1.18)–(3.3.20) is exactly observable. As a consequence of this, we obtain that system (3.3.23)–(3.3.25) is exactly observable and therefore the resolvent estimate (3.1.5) holds.

We now introduce the time-discrete schemes we are interested in. For any $\triangle t > 0$ and $\beta > 0$, we consider the following Newmark time-discrete scheme for system (3.1.18):

$$\begin{cases} \dfrac{u^{k+1} + u^{k-1} - 2u^k}{(\triangle t)^2} + A_0\Big(\beta u^{k+1} + (1-2\beta)u^k + \beta u^{k-1}\Big) = 0, \\ \Big(\dfrac{u^0 + u^1}{2}, \dfrac{u^1 - u^0}{\triangle t}\Big) = (u_0, v_0) \in \mathcal{D}(A_0^{1/2}) \times H. \end{cases} \tag{3.3.26}$$

The energy of (3.3.26) is given by

$$E^{k+1/2} = \Big\|A_0^{1/2}\Big(\frac{u^k + u^{k+1}}{2}\Big)\Big\|^2 + \Big\|\frac{u^{k+1} - u^k}{\triangle t}\Big\|^2 + (4\beta - 1)\frac{(\triangle t)^2}{4}\Big\|A_0^{1/2}\Big(\frac{u^{k+1} - u^k}{\triangle t}\Big)\Big\|^2, \ k \in \mathbb{Z}, \tag{3.3.27}$$

which is a discrete counterpart of the continuous energy (3.3.19). Multiplying the first equation of (3.3.26) by $(u^{k+1} - u^{k-1})/2$ and using integration by parts, it is easy to show that (3.3.27) remains

constant with respect to k. Furthermore, we assume in the sequel that $\beta \geq 1/4$ to guarantee that system (3.3.26) is unconditionally stable.

The output function is given by the following discretization of (3.3.20):

$$y^{k+1/2} = B_1\Big(\frac{u^k + u^{k+1}}{2}\Big) + B_2\Big(\frac{u^{k+1} - u^k}{\triangle t}\Big), \tag{3.3.28}$$

where, as in (3.3.20), we assume that either B_1 or B_2 vanishes.

For any $s > 0$, we define $\mathcal{C}_s$ as in (3.1.10). Note that this space is invariant under the actions of the discrete semi-groups associated to the Newmark time-discrete schemes (3.3.26).

We have the following theorem:

Theorem 3.3.2. *Let $\beta \geq 1/4$ and $\delta > 0$. We assume that either $B_1 \equiv 0$ or $B_2 \equiv 0$.*

Then there exists a time T_δ such that for all $T > T_\delta$, there exists a positive constant $k_{T,\delta}$, such that for $\triangle t > 0$ small enough, the solution of (3.3.26) *with initial data $(u_0, v_0) \in \mathcal{C}_{\delta/\triangle t}$ satisfies*

$$k_{T,\delta} E^{1/2} \leq \triangle t \sum_{k\triangle t \in (0,T)} \Big\| y^{k+1/2} \Big\|_Y^2, \tag{3.3.29}$$

where $y^{k+1/2}$ is defined in (3.3.28) *and B_1, B_2 satisfy* (3.3.21).

Besides, T_δ can be chosen as

$$T_{\delta,1} = \pi\Big[(1+\beta\delta^2)^2\Big(1 + \Big(\beta - \frac{1}{4})\delta^2\Big)^2 M^2 + m^2 C_{B,1}^2 \frac{\delta}{16}^4\Big]^{1/2}, \tag{3.3.30}$$

if $B_2 = 0$ and as

$$T_{\delta,2} = \pi\Big[(1+\beta\delta^2)^2\Big(1 + \Big(\beta - \frac{1}{4}\Big)\delta^2\Big) M^2 + m^2 C_{B,2}^2 \frac{\delta^4}{16}\Big]^{1/2}, \tag{3.3.31}$$

if $B_1 = 0$.

Remark 3.3.3. This result and especially the time estimates (3.3.30) and (3.3.31) on the observability time need further comments.

As in Theorem 3.2.1, we see that, if we filter at a scale smaller than $\triangle t$, for instance in the class $\mathcal{C}_{\delta/(\triangle t)^\alpha}$, with $\alpha < 1$, then the uniform observability time T_0 is given by $T_0 = \pi M$, which coincides with the value obtained by the resolvent estimate (3.1.5) in the continuous setting.

Note that the estimates (3.3.30) and (3.3.31) do not have the same growth in δ when δ goes to ∞. This fact does not seem to be natural because the observability time is expected to depend on the group velocity (see [25]) and not on the form of the observation operator.

By now we could not avoid the assumption that either B_1 or B_2 vanishes, the special case $\beta = 1/4$ being excepted. However, we can deal with an observable of the form

$$y^{k+1/2} = B_1\Big(I + (\beta - 1/4)(\triangle t)^2 A_0\Big)^{1/2}\Big(\frac{u^k + u^{k+1}}{2}\Big) + B_2\Big(\frac{u^{k+1} - u^k}{\triangle t}\Big), \tag{3.3.32}$$

with both non-trivial B_1 and B_2. Indeed, in this case, the operator $B_{\triangle t}$ arising in the proof of Theorem 3.3.2 does not depend on $\triangle t$ and therefore the proof works as in the case $B_1 = 0$, and yields the time

estimate (3.3.31). However, this observation operator, which compares to the continuous one (3.3.20) when $\delta \to 0$, does not seem to be the most natural discretization of (3.3.25).

When $\beta = 1/4$, both (3.3.30) and (3.3.31) have the same form. Besides, one can easily adapt the proof to show that when $\beta = 1/4$, we can deal with a general observation operator B as in (3.3.20). Actually, the Newmark scheme (3.3.26) with $\beta = 1/4$ is equivalent to a midpoint scheme, and therefore Theorem 3.2.1 applies.

Proof. **Step** 1. We first transform system (3.3.26) into a first order time-discrete scheme similar to (3.3.23). For this, we define

$$A_{0,\triangle t} = A_0[I + (\beta - 1/4)(\triangle t)^2 A_0]^{-1}. \tag{3.3.33}$$

Then (3.3.26) can be rewritten as

$$\frac{u^{k+1} + u^{k-1} - 2u^k}{(\triangle t)^2} + A_{0,\triangle t}\Big(\frac{u^{k-1} + 2u^k + u^{k+1}}{4}\Big) = 0. \tag{3.3.34}$$

As in (3.3.22), using the following change of variables

$$\left\{ \begin{array}{l} z_1^{k+1/2} = \dfrac{u^{k+1} - u^k}{\triangle t} + iA_{0,\triangle t}^{1/2}\Big(\dfrac{u^k + u^{k+1}}{2}\Big), \\ z_2^{k+1/2} = \dfrac{u^{k+1} - u^k}{\triangle t} - iA_{0,\triangle t}^{1/2}\Big(\dfrac{u^k + u^{k+1}}{2}\Big), \end{array} \right. \tag{3.3.35}$$

system (3.3.26) (and also system (3.3.34)) is equivalent to

$$\frac{z^{k+1/2} - z^{k-1/2}}{\triangle t} = A_{\triangle t}\Big(\frac{z^{k-1/2} + z^{k+1/2}}{2}\Big), \tag{3.3.36}$$

with

$$A_{\triangle t} = \begin{pmatrix} iA_{0,\triangle t}^{1/2} & 0 \\ 0 & -iA_{0,\triangle t}^{1/2} \end{pmatrix}, \quad z^{k+1/2} = \begin{pmatrix} z_1^{k+1/2} \\ z_2^{k+1/2} \end{pmatrix}. \tag{3.3.37}$$

Consequently, the observation operator $y^{k+1/2}$ in (3.3.28) is given by

$$\begin{aligned} y^{k+1/2} &= B_1 A_{0,\triangle t}^{-1/2}\Big(\frac{iz_2^{k+1/2} - iz_1^{k+1/2}}{2}\Big) + B_2\Big(\frac{z_1^{k+1/2} + z_2^{k+1/2}}{2}\Big) \\ &\stackrel{\triangle}{=} B_{\triangle t} z^{k+1/2}. \end{aligned} \tag{3.3.38}$$

Step 2. We now verify that system (3.3.36)–(3.3.38) satisfies the hypothesis of Theorem 3.1.3.

We first check (H1). It is obvious that the eigenvectors of $A_{\triangle t}$ are the same as those of A (see (3.3.24)). Moreover, for any Φ_j we compute

$$A_{\triangle t}\Phi_j = i\ell_j \Phi_j, \quad \text{with} \quad \ell_j = \frac{\mu_j}{\sqrt{1 + (\beta - 1/4)(\triangle t)^2 \mu_j^2}}. \tag{3.3.39}$$

In other words, we are close to the situation considered in Subsection 3.3.1, and the time semi-discrete approximation scheme (3.3.36) satisfies the hypotheses (3.3.1), (3.3.2), (3.3.3), (3.3.3) and (3.3.4) with the function h defined by

$$h(\eta) = 2\arctan\Big(\frac{\eta}{2}\frac{1}{\sqrt{1+(\beta-1/4)\eta^2}}\Big). \tag{3.3.40}$$

In particular, this implies that (3.3.16) holds in the class $\mathcal{C}_{\delta/\triangle t}$, and takes the form

$$\triangle t\,\|A_{\triangle t}z\|_X \le \frac{\delta}{\sqrt{1+(\beta-1/4)\delta^2}}\,\|z\|_X\,, \qquad z\in\mathcal{C}_{\delta/\triangle t}. \tag{3.3.41}$$

Second, we check hypothesis (H2):

$$\begin{aligned}\|B_{\triangle t}z\|_Y &\le \|A_{\triangle t}z\|_H\left(C_{B,1}\left\|A_0A_{0,\triangle t}^{-1}\right\|_{\mathfrak{L}(\mathcal{C}_{\delta/\triangle t},H)} + C_{B,2}\left\|A_0^{1/2}A_{0,\triangle t}^{-1/2}\right\|_{\mathfrak{L}(\mathcal{C}_{\delta/\triangle t},H)}\right)\\ &\le \|A_{\triangle t}z\|_H\left((1+(\beta-1/4)\delta^2)C_{B,1} + \sqrt{1+(\beta-1/4)\delta^2}C_{B,2}\right)\\ &\le C_{B,\delta}\,\|A_{\triangle t}z\|_H\,.\end{aligned} \tag{3.3.42}$$

The third point is more technical. Following the proof of Theorem 3.3.1, for any $\varepsilon>0$, we obtain the following resolvent estimate:

$$C_{\delta,\varepsilon}^2M^2\left\|\Big(A_{\triangle t}-i\omega\Big)z\right\|_X^2 + m^2\,\|Bz\|_Y^2 \ge \|z\|_X^2\,, \quad z\in\mathcal{C}_{\delta/\triangle t},\ \omega\in\mathbb{R}, \tag{3.3.43}$$

where $C_{\delta,\varepsilon}$ is given by (3.3.12), with

$$k_{\triangle t}(\omega) = \frac{\omega}{\sqrt{1+(\beta-1/4)(\omega\triangle t)^2}}.$$

Straightforward computations show that, actually,

$$C_{\delta,\varepsilon} = \Big(1+(\beta-1/4)(\delta+\varepsilon)^2\Big)^{3/2}. \tag{3.3.44}$$

Our goal now is to derive from (3.3.43) the resolvent estimate (H3) given in (3.1.13). Here, we will handle separately the two cases $B_1=0$ and $B_2=0$.

The case $B_1=0$. Under this assumption, $B_{\triangle t}=B$, and therefore, (3.3.43) is the resolvent estimate (H3) we need.

The case $B_2=0$. In this case, we observe that

$$B_{\triangle t}z = BR_{\triangle t}z, \quad \text{where } R_{\triangle t} = \begin{pmatrix} A_0^{1/2}A_{0,\triangle t}^{-1/2} & 0\\ 0 & A_0^{1/2}A_{0,\triangle t}^{-1/2}\end{pmatrix} = AA_{\triangle t}^{-1}.$$

Note that the operator $R_{\triangle t}$ commutes with $A_{\triangle t}$, maps $\mathcal{C}_{\delta/\triangle t}$ into itself, and is invertible. Then, applying (3.3.43) to $R_{\triangle t}z$, we obtain that

$$C_{\delta,\varepsilon}^2M^2\left\|R_{\triangle t}\Big(A_{\triangle t}-i\omega\Big)z\right\|_X^2 + m^2\,\|B_{\triangle t}z\|_Y^2 \ge \|R_{\triangle t}z\|_X^2\,, \quad \forall z\in\mathcal{C}_{\delta/\triangle t},\ \forall\omega\in\mathbb{R}. \tag{3.3.45}$$

We now compute explicitly the norm of $R_{\triangle t}$ and $R_{\triangle t}^{-1}$ in the class $\mathcal{C}_{\delta/\triangle t}$. Since

$$A_0A_{0,\triangle t}^{-1} = 1+(\beta-1/4)(\triangle t)^2A_0,$$

one easily checks that

$$\|R_{\triangle t}\|_\delta^2 = 1 + (\beta - 1/4)\delta^2, \qquad \left\|R_{\triangle t}^{-1}\right\|_\delta^2 = 1, \tag{3.3.46}$$

where $\|\cdot\|_\delta$ denotes the operator norm from $\mathcal{C}_{\delta/\triangle t}$ into itself. Applying (3.3.46) into (3.3.45), we obtain

$$C_{\delta,\varepsilon}^2 M^2\Big(1 + (\beta - 1/4)\delta^2\Big)\left\|\Big(A_{\triangle t} - i\omega\Big)z\right\|_X^2 + m^2\left\|B_{\triangle t}z\right\|_Y^2 \geq \|z\|_X^2\,, \quad \forall z \in \mathcal{C}_{\delta/\triangle t}, \forall \omega \in \mathbb{R}. \tag{3.3.47}$$

Thus, in both cases, we can apply Theorem 3.1.3, which gives the existence of a time $T_{\delta,\varepsilon}$ such that for $T > T_{\delta,\varepsilon}$, there exists a positive $k_{T,\delta}$ such that any solution of (3.3.36) with initial data $z^{1/2} \in \mathcal{C}_{\delta/\triangle t}$ satisfies

$$k_{T,\delta}\left\|z^{1/2}\right\|_X^2 \leq \sum_{k=0}^{T/\triangle t}\left\|B_{\triangle t}z^{k+1/2}\right\|_Y^2.$$

Besides, the estimates of Theorem 3.1.3 allow to estimate the observability time $T_{\delta,\varepsilon}$:

$$T_{\delta,\varepsilon} = \begin{cases} \pi\Big[(1+\beta\delta^2)^2\dfrac{(1+(\beta-1/4)(\delta+\varepsilon)^2)^3}{1+(\beta-1/4)\delta^2}M^2 + m^2C_{B,1}^2\dfrac{\delta}{16}^4\Big]^{1/2}, & if\ B_2 = 0, \\ \pi\Big[(1+\beta\delta^2)^2\dfrac{(1+(\beta-1/4)(\delta+\varepsilon)^2)^3}{(1+(\beta-1/4)\delta^2)^2}M^2 + m^2C_{B,2}^2\dfrac{\delta^4}{16}\Big]^{1/2}, & if\ B_1 = 0. \end{cases}$$

Letting $\varepsilon \to 0$, we obtain the estimates (3.3.30)-(3.3.31).

To complete the proof we check that if the initial data $z^{1/2}$ is taken within the class $\mathcal{C}_{\delta/\triangle t}$, the solution of (3.3.26) satisfies

$$\left\|z^{1/2}\right\|_X^2 = \left\|z^{k+1/2}\right\|_X^2 \geq \frac{2}{1+(\beta-1/4)\delta^2}E^{k+1/2},$$

which can be deduced from the explicit expression of the energy (3.3.27) and the formula (3.3.35). □

3.4 Applications

3.4.1 Application of Theorem 3.2.1

Boundary observation of the Schrödinger equation

The goal of this subsection is to present a straightforward application of Theorem 3.2.1 to the observability properties of the Schrödinger equation based on the results in [14].

Let $\Omega \subset \mathbb{R}^n$ be a smooth bounded domain. Consider the equation

$$\begin{cases} iu_t = \Delta_x u, & (t,x) \in (0,T)\times\Omega, \\ u(0) = u_0,\ x \in \Omega, & \dfrac{\partial u}{\partial \nu}(t,x) = 0,\ (t,x) \in (0,T)\times\partial\Omega. \end{cases} \tag{3.4.1}$$

where $u_0 \in L^2(\Omega)$ is the initial data. Equation (3.4.1) obviously has the form (3.1.1) with $A = -i\Delta_x$ of domain

$$\mathcal{D}(A) = \Big\{\varphi \in H^2(\Omega) \text{ such that } \frac{\partial\varphi}{\partial\nu} = 0\Big\}.$$

Let $\Gamma_0 \subset \partial\Omega$ be an open subset of $\partial\Omega$ and define the output

$$y(t) = u(t)_{|\Gamma_0}.$$

Using Sobolev's embedding theorems, one can easily check that this defines a continuous observation operator B from $\mathcal{D}(A)$ to $L^2(\Gamma_0)$.

Let us assume that Γ_0 satisfies in some time T_0 the *Geometric Control Condition* (GCC) introduced in [1], which asserts that all the rays of Geometric Optics in Ω touch the sub-boundary Γ_0 in a time smaller than T_0. In this case, the following observability result is known ([14]) :

Theorem 3.4.1. *For any $T > 0$, there exist positive constants $k_T > 0$ and $K_T > 0$ such that for any $u_0 \in L^2(\Omega)$, the solution of* (3.4.1) *satisfies*

$$k_T \left\|u_0\right\|^2_{L^2(\Omega)} \leq \int_0^T \int_{\Gamma_0} |u(t)|^2 \, d\Gamma_0 dt \leq K_T \left\|u_0\right\|^2_{L^2(\Omega)}. \tag{3.4.2}$$

We introduce the following time semi-discretization of system (3.4.1):

$$\begin{cases} i\dfrac{u^{k+1}-u^k}{\triangle t} = \Delta_x\Big(\dfrac{u^{k+1}+u^k}{2}\Big), & x \in \Omega,\ k \in \mathbb{N}, \\ \dfrac{\partial u^k}{\partial \nu}(x) \ = 0, & x \in \partial\Omega,\ k \in \mathbb{N}, \\ u^0(x) \ = u_0(x), & x \in \Omega, \end{cases} \tag{3.4.3}$$

that we observe through

$$y^k = u^k_{|\Gamma_0}.$$

Then Theorem 3.2.1 implies the following result:

Theorem 3.4.2. *For any $\delta > 0$, there exists a time T_δ such that for any time $T > T_\delta$, there exists a positive constant $k_{T,\delta} > 0$ such that for $\triangle t$ small enough, the solution of* (3.4.3) *satisfies*

$$k_{T,\delta} \left\|u_0\right\|^2_{L^2(\Omega)} \leq \triangle t \sum_{k \in (0, T/\triangle t)} \int_{\Gamma_0} \left|u^k\right|^2 d\Gamma_0 \tag{3.4.4}$$

for any $u_0 \in \mathcal{C}_{\delta/\triangle t}$.

Note that we do not know if inequality (3.4.4) holds in any time $T > 0$ as in the continuous case (see (3.4.2)). This question is still open.

Remark 3.4.3. Note that in the present section, we do not state any admissibility result for the time-discrete systems under consideration. However, uniform (with respect to $\triangle t > 0$) admissibility results hold for all the examples presented in this article. These results will be derived in Section 3.6 using the admissibility property of the continuous system (3.1.1)-(3.1.2).

Boundary observation of the linearized KdV equation

We now present an application of Theorem 3.2.1 to the boundary observability of the linear KdV equation.

We consider the following initial-value boundary problem for the KdV equation:

$$\left\{\begin{array}{rcll} u_t + u_{xxx} &=& 0, & (t,x) \in (0,T)\times(0,2\pi), \\ u(t,0) &=& u(t,2\pi), & t \in (0,T), \\ u_x(t,0) &=& u_x(t,2\pi), & t \in (0,T), \\ u_{xx}(t,0) &=& u_{xx}(t,2\pi), & t \in (0,T), \\ u(0,x) &=& u_0(x), & x \in (0,2\pi). \end{array}\right. \tag{3.4.5}$$

For any integer k we set

$$H_p^k \stackrel{\triangle}{=} \left\{u \in H^k(0,2\pi);\quad \partial_x^j u(0) = \partial_x^j u(2\pi) \ \text{ for } \ 0 \le j \le k-1\right\}, \tag{3.4.6}$$

where $H^k(0,2\pi)$ denotes the classical Sobolev spaces on the interval $(0,2\pi)$. The initial data of (3.4.5) are taken in the space $X \stackrel{\triangle}{=} H_p^2(0,2\pi)$, endowed with the classical $H^2(0,2\pi)$-norm.

Let A denote the operator $Au = -\partial_x^3 u$ with domain $\mathcal{D}(A) = H_p^5$. As shown in [24], A is a skew-adjoint operator with compact resolvent. Moreover, its spectrum is given by $\sigma(A) = \{i\mu_j \quad \text{with } \mu_j = j^3,\ j \in \mathbb{Z}\}$. The output function $y(t)$ and the corresponding operator $B : \mathcal{D}(A) \longrightarrow Y = \mathbb{R}^3$ is given by

$$y(t) \stackrel{\triangle}{=} Bu(t) = \left(\begin{array}{c} u(t,0) \\ u_x(t,0) \\ u_{xx}(t,0) \end{array}\right),$$

with the norm $\|Bu\|_Y^2 = |u(0)|^2 + |u_x(0)|^2 + |u_{xx}(0)|^2$. Note that $B \in \mathfrak{L}(H_p^5, \mathbb{R}^3)$.

The following observability inequality for system (3.4.5) is well-known (Prop. 2.2 of [23]):

Lemma 3.4.4. *Let $T > 0$. Then there exist positive numbers k_T and K_T such that for every $u_0 \in H_p^2(0,2\pi)$,*

$$k_T \|u_0\|_{H_p^2}^2 \le \int_0^T \Big(|u(t,0)|^2 + |u_x(t,0)|^2 + |u_{xx}(t,0)|^2\Big)dt \le K_T \|u_0\|_{H_p^2}^2 . \tag{3.4.7}$$

We now introduce the following time semi-discretization of system (3.4.5):

$$\left\{\begin{array}{ll} \dfrac{u^{k+1}-u^k}{\triangle t} + \dfrac{u_{xxx}^{k+1}+u_{xxx}^k}{2} = 0, & x \in (0,2\pi),\ k \in \mathbb{N}, \\ u^k(0) = u^k(2\pi), & k \in \mathbb{N}, \\ u_x^k(0) = u_x^k(2\pi), & k \in \mathbb{N}, \\ u_{xx}^k(0) = u_{xx}^k(2\pi), & k \in \mathbb{N}, \\ u^0(x) = u_0(x), & x \in (0,2\pi). \end{array}\right. \tag{3.4.8}$$

It is easy to show that the eigenfunctions of A are given by $\{\Phi_j = e^{ijx}\}_{j\in\mathbb{Z}}$ with the corresponding eigenvalues $\{ij^3\}_{j\in\mathbb{Z}}$. Hence, for any $\delta > 0$, we have

$$\mathcal{C}_{\delta/\triangle t} = \text{span}\,\{\Phi_j, j^3 \leq \delta/\triangle t\}. \tag{3.4.9}$$

As a direct consequence of Theorem 3.2.1 we have the following uniform observability result for system (3.4.8):

Theorem 3.4.5. *For any $\delta > 0$, there exists a time T_δ such that for any $T > T_\delta$, there exists a positive constant $k_{T,\delta} > 0$ such that for $\triangle t > 0$ small enough, the solution u^k of* (3.4.8) *satisfies*

$$k_{T,\delta}\,\|u_0\|^2_{H_p^2} \leq \triangle t \sum_{k\triangle t\in(0,T)} \Big(|u^k(0)|^2 + |u_x^k(0)|^2 + |u_{xx}^k(0)|^2\Big), \tag{3.4.10}$$

for any initial data $u^0 \in \mathcal{C}_{\delta/\triangle t}$.

As in Theorem 3.4.2, we do not know if the observability estimate (3.4.10) holds in any time $T > 0$ as in the continuous case (see Lemma 3.4.4).

3.4.2 Application of Theorem 3.3.1

Let us present an application of Theorem 3.3.1 to the so-called fourth order Gauss method discretization of equation (3.1.1) (see for instance [8, 9]). This fourth order Gauss method is a special case of the Runge-Kutta time approximation schemes, which corresponds to the only conservative scheme within this class.

Consider the following discrete system:

$$\begin{cases} \kappa_i = A\Big(z^k + \triangle t \sum_{j=1}^{2} \alpha_{ij}\kappa_j\Big), \qquad i = 1, 2, \\ z^{k+1} = z^k + \dfrac{\triangle t}{2}(\kappa_1 + \kappa_2), \quad (\alpha_{ij}) = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \end{pmatrix}. \\ z^0 \in \mathcal{C}_{\delta/\triangle t} \text{ given}, \end{cases} \tag{3.4.11}$$

The scheme is unstable for the eigenfunctions corresponding to the eigenvalues μ_j such that $\mu_j\triangle t \geq 2\sqrt{3}$ ([8, 9]). Thus we immediately impose the following restriction on the filtering parameter :

$$\delta < 2\sqrt{3}.$$

To use Theorem 3.3.1, we only need to check the behavior of the semi-discrete scheme (3.4.11) on the eigenvectors. If $z^0 = \Phi_j$, an easy computation shows that

$$z^1 = \exp(i\ell_j\triangle t)z^0,$$

where

$$\ell_j = \frac{2}{\triangle t}\arctan\Big(\frac{\mu_j\triangle t}{2 - (\mu_j\triangle t)^2/6}\Big). \tag{3.4.12}$$

In other words, $\ell_j\triangle t = h(\mu_j\triangle t)$, where $h : (-2\sqrt{3}, 2\sqrt{3}) \longrightarrow [-\pi, \pi]$ is given by

$$h(\eta) = 2\arctan\Big(\frac{\eta}{2 - \eta^2/6}\Big).$$

Then, a simple application of Theorem 3.3.1 gives :

Theorem 3.4.6. *Assume that B is an observation operator such that (A, B) satisfy* (3.1.5) *and $B \in \mathfrak{L}(\mathcal{D}(A), Y)$.*

For any $\delta \in (0, 2\sqrt{3})$, there exists a time $T_\delta > 0$ such that for any $T > T_\delta$, there exists a constant $k_{T,\delta} > 0$, independent of $\triangle t$, such that for $\triangle t > 0$ small enough, the solutions of system (3.4.11) *satisfy*

$$k_{T,\delta} \left\| z^0 \right\|_X^2 \leq \triangle t \sum_{k \in (0, T/\triangle t)} \left\| B z^k \right\|_Y^2, \quad \forall\ z^0 \in \mathcal{C}_{\delta/\triangle t}. \tag{3.4.13}$$

Note that Theorem 3.3.1 also provides an estimate on T_δ by using (3.3.6).

In particular, this provides another possible time-discretization of (3.4.5), for which the observability inequality holds uniformly in $\triangle t$ provided the initial data are taken in $\mathcal{C}_{\delta/\triangle t}$, with $\delta < 2\sqrt{3}$, where $\mathcal{C}_{\delta/\triangle t}$ is as in (3.4.9).

3.4.3 Application of Theorem 3.3.2

There are plenty of applications of Theorem 3.3.2. We present here an application to the boundary observability of the wave equation.

Consider a smooth nonempty open bounded domain $\Omega \subset \mathbb{R}^d$ and let Γ_0 be an open subset of $\partial\Omega$. We consider the following initial boundary value problem:

$$\begin{cases} u_{tt} - \Delta_x u = 0, & x \in \Omega, \quad t \geq 0, \\ u(x,t) = 0, & x \in \partial\Omega, \ \ t \geq 0, \\ u(x,0) = u_0, \quad u_t(x,0) = v_0, & x \in \Omega \end{cases} \tag{3.4.14}$$

with the output

$$y(t) = \frac{\partial u}{\partial \nu}\Big|_{\Gamma_0}. \tag{3.4.15}$$

This system is conservative and the energy of (3.4.14)

$$E(t) = \frac{1}{2} \int_\Omega \Big[|u_t(t,x)|^2 + |\nabla u(t,x)|^2 \Big] dx, \tag{3.4.16}$$

remains constant, i.e.

$$E(t) = E(0), \qquad \forall\ t \in [0, T]. \tag{3.4.17}$$

The boundary observability property for system (3.4.14) is as follows: *For some constant $C = C(T, \Omega, \Gamma_0) > 0$, solutions of* (3.4.14) *satisfy*

$$E(0) \leq C \int_0^T \int_{\Gamma_0} \Big| \frac{\partial u}{\partial \nu} \Big|^2 d\Gamma_0 dt, \qquad \forall\ (u_0, v_0) \in H_0^1(\Omega) \times L^2(\Omega). \tag{3.4.18}$$

Note that this inequality holds true for all triplets (T, Ω, Γ_0) satisfying the *Geometric Control Condition* (GCC) introduced in [1], see Subsection 3.4.1. In this case, (3.4.18) is established by means of microlocal analysis tools (see [1]). From now, we assume this condition to hold.

We then introduce the following time semi-discretization of (3.4.14):

$$\begin{cases} \dfrac{u^{k+1}+u^{k-1}-2u^k}{(\triangle t)^2} = \Delta_x\Big(\beta u^{k+1}+(1-2\beta)u^k+\beta u^{k-1}\Big), \text{ in } \Omega\times\mathbb{Z}, \\ u^k = 0, \qquad\qquad \text{in } \partial\Omega\times\mathbb{Z}, \\ \Big(\dfrac{u^0+u^1}{2}, \dfrac{u^1-u^0}{\triangle t}\Big) = (u_0,\ v_0)\ \in H_0^1(\Omega)\times L^2(\Omega), \end{cases} \tag{3.4.19}$$

where β is a given parameter satisfying $\beta \geq \frac{1}{4}$.

The output functions y^k are given by

$$y^k = \frac{\partial u^k}{\partial \nu}\Big|_{\Gamma_0}. \tag{3.4.20}$$

System (3.4.14)–(3.4.15) (or system (3.4.19)–(3.4.20)) can be written in the form (3.1.18) (or (3.3.26)) with observation operator (3.3.20) by setting:

$$H = L^2(\Omega),\quad \mathcal{D}(A_0) = H^2(\Omega)\cap H_0^1(\Omega),\quad Y = L^2(\Gamma_0),$$
$$A_0\varphi = -\Delta_x\varphi \quad \forall\varphi\in\mathcal{D}(A_0), \qquad B_1\varphi = \frac{\partial\varphi}{\partial\nu}\Big|_{\Gamma_0}, \varphi\in\mathcal{D}(A_0).$$

One can easily check that A_0 is self-adjoint in H, positive and boundedly invertible and

$$\mathcal{D}(A_0^{1/2}) = H_0^1(\Omega), \qquad \mathcal{D}(A_0^{1/2})^* = H^{-1}(\Omega).$$

Proposition 3.4.7. *With the above notation, $B_1\in\mathfrak{L}(\mathcal{D}(A_0),Y)$ is an admissible observation operator, i.e. for all $T>0$ there exists a constant $K_T>0$ such that: If u satisfies* (3.4.14) *then*

$$\int_0^T\int_{\Gamma_0}\Big|\frac{\partial u}{\partial\nu}\Big|^2 d\Gamma_0 dt \leq K_T\Big(\|u_0\|^2_{H_0^1(\Omega)}+\|v_0\|^2_{L^2(\Omega)}\Big)$$

for all $(u_0,v_0)\in H_0^1(\Omega)\times L^2(\Omega)$.

The above proposition is classical (see, for instance, p. 44 of [16]), so we skip the proof.

Hence we are in the position to give the following theorem:

Theorem 3.4.8. *Set $\beta\geq 1/4$.*

For any $\delta>0$, system (3.4.19) *is uniformly observable with $(u_0,v_0)\in\mathcal{C}_{\delta/\triangle t}$. More precisely, there exists T_δ, such that for any $T>T_\delta$, there exists a positive constant $k_{T,\delta}$ independent of $\triangle t$, such that for $\triangle t>0$ small enough, the solutions of system* (3.4.19) *satisfy*

$$k_{T,\delta}\Big(\|\nabla u_0\|^2+\|v_0\|^2\Big)\leq\triangle t\sum_{k\in(0,T/\triangle t)}\int_{\Gamma_0}\Big|\frac{\partial u^k}{\partial\nu}\Big|^2 d\Gamma_0, \tag{3.4.21}$$

for any $(u_0,v_0)\in\mathcal{C}_{\delta/\triangle t}$.

Proof. The scheme proposed here is a Newmark discretization. Hence this result is a direct consequence of Theorem 3.3.2. □

Remark 3.4.9. One can use Fourier analysis and microlocal tools to discuss the optimality of the filtering condition as in [28]. The symbol of the operator in (3.4.19), that can be obtained by taking the Fourier transform of the differential operator in space-time is of the form (see for instance [17])

$$\frac{4}{\triangle t^2}\sin^2\Big(\frac{\tau\triangle t}{2}\Big)-\Big|\xi\Big|^2\Big(1-4\beta\sin^2\Big(\frac{\tau\triangle t}{2}\Big)\Big).$$

Note that this symbol is not hyperbolic in the whole range $(\tau,\xi)\in(-\pi/\triangle t,\pi/\triangle t)\times\mathbb{R}^n$. However, the Fourier transform of any solution of (3.4.19) is supported in the set of (τ,ξ) satisfying $1-4\beta\sin^2(\tau\triangle t/2)>0$, where the symbol is hyperbolic.

As in the continuous case, one expects the optimal observability time to be the time needed by all the rays to meet Γ_0. Along the bicharacteristic rays associated to this hamiltonian the following identity holds

$$|\tau|=\frac{2}{\triangle t}\arctan\left(\frac{|\xi|\triangle t}{2}\ \frac{1}{\sqrt{1+(\beta-1/4)|\xi|^2(\triangle t)^2}}\right).$$

These rays are straight lines as in the continuous case, but their velocity is not 1 anymore. Indeed, one can prove that along the rays corresponding to $|\xi|<\delta/\triangle t$, the velocity of propagation is given by

$$\Big|\frac{dx}{dt}\Big|=\frac{1}{1+\beta(|\xi|\triangle t)^2}\ \frac{1}{\sqrt{1+(\beta-1/4)(\xi\triangle t)^2}}\geq\frac{1}{(1+\beta\delta^2)\sqrt{1+(\beta-1/4)\delta^2}}.$$

In other words, in the class $\mathcal{C}_{\delta/\triangle t}$, the velocity of propagation of the rays concentrated in frequency around $\delta/\triangle t$ is $(1+\beta\delta^2)^{-1}(1+(\beta-1/4)\delta^2)^{-1/2}$ times that of the continuous wave equation. Therefore we expect the optimal observability time T^*_δ in the class $\mathcal{C}_{\delta/\triangle t}$ to be

$$T^*_\delta=T^*_0(1+\beta\delta^2)\sqrt{1+\Big(\beta-\frac{1}{4}\Big)\delta^2},\tag{3.4.22}$$

where T^*_0 is the optimal observability time for the continuous system. According to this, the estimate $T_{\delta,2}$ in (3.3.31) on the time of observability has the good growth rate when $\delta\to\infty$. Besides, when δ goes to ∞, we have that

$$T_{\delta,2}\simeq\pi M(1+\beta\delta^2)\sqrt{1+\Big(\beta-\frac{1}{4}\Big)\delta^2}.\tag{3.4.23}$$

Recall that $\pi M=T_0$ is the time of observability that the resolvent estimate (3.1.5) in the continuous setting yields (see [18]). The similarity between (3.4.22) and (3.4.23) indicates that the resolvent method accurately measures the group velocity.

Note however that πM is not the expected sharp observability time T^*_0 in (3.4.22) in the continuous setting. This is one of the drawbacks of the method based on the resolvent estimates we use in this paper. Even at the continuous level the observability time one gets this way is far from being the optimal one that Geometric Optics yields.

3.5 Fully discrete schemes

3.5.1 Main statement

In this section, we deal with the observability properties for time-discretization systems such as (3.1.1)-(3.1.2) depending on an extra parameter, for instance the *space* mesh-size, or the size of the microstructure in homogenization.

To this end, it is convenient to introduce the following class of operators:

Definition 3.5.1. For any $(m, M, C_B) \in (\mathbb{R}_+^*)^3$, we define $\mathfrak{C}(m, M, C_B)$ as the class of operators (A, B) satisfying:

(A1) The operator A is skew-adjoint on some Hilbert space X, and has a compact resolvent.

(A2) The operator B is defined from $\mathcal{D}(A)$ with values in a Hilbert space Y, and satisfies (3.2.1) with C_B.

(A3) The pair of operators (A, B) satisfies the resolvent estimate (3.1.5) with constants m and M.

In this class, Theorems 3.2.1-3.3.1-3.3.2 apply and provide uniform observability results for any of the time semi-discrete approximation schemes (3.1.6)-(3.1.7), (3.1.17), and (3.1.18). Indeed, this can be deduced by the explicit form of the constants T_δ and $k_{T,\delta}$ which only depend on m, M and C_B. Note that this definition does not depend on the spaces X and Y. For instance, the following holds:

Theorem 3.5.2 (Corollary of Theorem 3.2.1)**.** *For any $(m, M, C_B) \in (\mathbb{R}_+^*)^3$, for any $\delta > 0$, there exists T_δ^{m,M,C_B} such that for any $T > T_\delta^{m,M,C_B}$, there exists a positive constant k_{T,δ,m,M,C_B}, independent of $\triangle t$, such that for $\triangle t$ small enough, for any $(A, B) \in \mathfrak{C}(m, M, C_B)$, the solution z^k of* (3.1.6) *with $z^0 \in \mathcal{C}_{\delta/\triangle t}$ satisfies* (3.2.2)*. Moreover, T_δ^{m,M,C_B} can be taken as in* (3.2.3)*.*

When considering families of pairs of operators (A, B), it is not easy, in general, to show that they belong to the same class $\mathfrak{C}(m, M, C_B)$ for some choice of the constants (m, M, C_B). Indeed, item (A3) is not obvious in general. Therefore, in the sequel, we define another class included in some $\mathfrak{C}(m, M, C_B)$ and which is easier to handle in practice.

Definition 3.5.3. For any $(C_B, T, k_T, K_T) \in (\mathbb{R}_+^*)^4$, we define $\mathfrak{D}(C_B, T, k_T, K_T)$ as the class of operators (A, B) satisfying (A1), (A2) and:

(B1) The admissibility inequality

$$\int_0^T \left\| B \exp(tA) z^0 \right\|_Y^2 \, dt \leq K_T \left\| z^0 \right\|_X^2, \tag{3.5.1}$$

where $\exp(tA)$ stands for the semigroup associated to the equation

$$\dot{z} = Az, \quad z(0) = z^0 \in X. \tag{3.5.2}$$

(B2) The observability inequality

$$k_T \left\| z^0 \right\|_X^2 \leq \int_0^T \left\| B \exp(tA) z^0 \right\|_Y^2 \, dt. \tag{3.5.3}$$

As we will see below, assumptions (B1)-(B2) imply (A3):

Lemma 3.5.4. *If the pair (A, B) belongs to $\mathfrak{D}(C_B, T, k_T, K_t)$, then there exist m and M such that $(A, B) \in \mathfrak{C}(m, M, C_B)$.*

Besides m and M can be chosen as

$$m = \sqrt{\frac{2T}{k_T}}, \qquad M = T\sqrt{\frac{K_T}{2k_T}}. \tag{3.5.4}$$

In fact, we only need to prove (A3). This is actually already done in [18] or in [26]. Indeed, it was proved that once the admissibility inequality (3.1.3) and the observability inequality (3.1.4) hold for some time T, then the resolvent estimate (3.1.5) hold with m and M as in (3.5.4).

Note that assumptions (B1)-(B2) are related to the *continuous* systems (3.5.2).

Now we consider a sequence of operators (A_p, B_p) depending on a parameter $p \in P$, which are in some $\mathfrak{L}(X_p) \times \mathfrak{L}(\mathcal{D}(A_p), Y_p)$ for each p, where X_p and Y_p are Hilbert spaces. We want to address the observability problem for a time-discretization scheme of

$$\dot{z} = A_p z, \quad z(0) = z^0 \in X_p, \qquad y(t) = B_p z(t) \in Y_p. \tag{3.5.5}$$

In applications, we need the observability to be uniform in both $p \in P$ and $\triangle t > 0$ small enough. The previous analysis and the properties of the class $\mathfrak{D}(C_B, T, k_T, K_T)$ suggest the following two-steps strategy:

1. Study the continuous system (3.5.5) for every parameter p and prove the uniform admissibility (3.5.1) and observability (3.5.3).

2. Apply one of the Theorems 3.2.1, 3.3.1 and 3.3.2 to obtain uniform observability estimates (3.1.8) for the corresponding time-discrete approximation schemes.

This allows dealing with fully discrete approximation schemes. In that setting the parameter p is actually the standard parameter $h > 0$ associated with the space mesh-size. In this way one can use automatically the existing results for space semi-discretizations as, for instance, [4, 6, 7, 10, 20, 21, 30, 31].

Remark 3.5.5. We emphasize that this approach is based on the systematic use of existing results for space semi-discretizations. One could proceed all the way around, first, applying the results in this paper to derive uniform observability results for time-discrete schemes and then discretizing the space variables. For doing this, however, due to the more complex dependence of the PDE and its space discretizations on the space variable, there is no systematic way of transfering results from the continuous to the discrete setting. In this sense, the method we propose here of using the existing results for space semi-discretizations to later apply the results in this paper about time discretizations is much more easier to be implemented and yields better results.

3.5.2 Applications

The fully discrete wave equation

Let us consider the wave equation (3.4.14) in a 2-d square. More precisely, let $\Omega = (0,\pi)\times(0,\pi) \subset \mathbb{R}^2$ and Γ_0 be a subset of the boundary of Ω constituted by two consecutive sides, for instance,

$$\Gamma_0 = \{(x_1,\pi) : x_1 \in (0,\pi)\} \cup \{(\pi,x_2) : x_2 \in (0,\pi)\} \stackrel{\triangle}{=} \Gamma_1 \cup \Gamma_2.$$

As in (3.4.15), the output function $y(t) = Bu(t)$ is given by

$$Bu = \frac{\partial u}{\partial \nu}\Big|_{\Gamma_0} = \frac{\partial}{\partial x_2}u(x_1,\pi)\Big|_{\Gamma_1} + \frac{\partial}{\partial x_1}u(\pi,x_2)\Big|_{\Gamma_2}.$$

Let us first consider the finite-difference semi-discretization of (3.4.14). The following can be found in [30]. Given $J, K \in \mathbb{N}$ we set

$$h_1 = \frac{\pi}{J+1}, \qquad h_2 = \frac{\pi}{K+1}. \tag{3.5.6}$$

We denote by $u_{jk}(t)$ the approximation of the solution u of (3.4.14) at the point $x_{jk} = (jh_1, kh_2)$. The space semi-discrete approximation scheme of (3.4.14) is as follows:

$$\begin{cases} \ddot{u}_{jk} - \dfrac{u_{j+1k} + u_{j-1k} - 2u_{jk}}{h_1^2} - \dfrac{u_{jk+1} + u_{jk-1} - 2u_{jk}}{h_2^2} = 0, \\ \qquad\qquad\qquad\qquad 0 < t < T,\ j = 1,\cdots,J;\ k = 1,\cdots,K; \\ u_{jk} = 0, \qquad 0 < t < T,\ j = 0, J+1;\ k = 0, K+1, \\ u_{jk}(0) = u_{jk,0}, \quad \dot{u}_{jk}(0) = u_{jk,1},\ j = 1,\cdots,J;\ k = 1,\cdots,K. \end{cases} \tag{3.5.7}$$

System (3.5.7) is a system of JK linear differential equations. Moreover, if we denote the unknown

$$U(t) = (u_{11}(t), u_{21}(t), \cdots, u_{J1}(t), \cdots, u_{1K}(t), u_{2K}(t), \cdots, u_{JK}(t))^T,$$

then system (3.5.7) can be rewritten in vector form as follows

$$\begin{cases} \ddot{U}(t) + A_{0,h}U(t) = 0, \qquad 0 < t < T. \\ U(0) = U_{h,0}, \quad \dot{U}(0) = U_{h,1}, \end{cases} \tag{3.5.8}$$

where $(U_{h,0}, U_{h,1}) = (u_{jk,0}, u_{jk,1})_{1\le j\le J, 1\le k\le K} \in \mathbb{R}^{2JK}$ are the initial data. The corresponding solution of (3.5.7) is given by $(U_h, \dot{U}_h) = (u_{jk}, \dot{u}_{jk})_{1\le j\le J, 1\le k\le K}$. Note that the entries of $A_{0,h}$ belonging to $\mathcal{M}_{JK}(\mathbb{R})$ may be easily deduced from (3.5.7).

As a discretization of the output, we choose

$$B_hU = \Big(\Big(\frac{u_{jK}}{h_2}\Big)_{j\in\{1,\cdots,J\}}, \Big(\frac{u_{Jk}}{h_1}\Big)_{k\in\{1,\cdots K\}}\Big). \tag{3.5.9}$$

The corresponding norm for the observation operator B_h is given by

$$\|B_hU(t)\|_{Y_h}^2 = h_1\sum_{j=1}^{J}\Big|\frac{u_{jK}(t)}{h_2}\Big|^2 + h_2\sum_{k=1}^{K}\Big|\frac{u_{Jk}(t)}{h_1}\Big|^2.$$

Besides, the energy of the system (3.5.8) is given by

$$E_h(t) = \frac{h_1 h_2}{2} \sum_{j=0}^{J} \sum_{k=0}^{K} \Big(|\dot{u}_{jk}(t)|^2 + \Big| \frac{u_{j+1k}(t) - u_{jk}(t)}{h_1} \Big|^2 + \Big| \frac{u_{jk+1}(t) - u_{jk}(t)}{h_2} \Big|^2 \Big). \tag{3.5.10}$$

As in the continuous case, this quantity is constant.

$$E_h(t) = E_h(0), \qquad \forall \ \ 0 < t < T.$$

In order to prove the uniform observability of (3.5.8), we have to filter the high frequencies. To do that we consider the eigenvalue problem associated with (3.5.8):

$$A_{0,h}\varphi = \lambda^2 \varphi. \tag{3.5.11}$$

As in the continuous case, it is easy to show that the eigenvalues λ^{j,k,h_1,h_2} are positive numbers. Let us denote by φ^{j,k,h_1,h_2} the corresponding eigenvectors.

Let us now introduce the following classes of solutions of (3.5.8) for any $0 < \gamma < 1$:

$$\widehat{\mathcal{C}_\gamma}(h) = \text{span}\,\{\varphi^{j,k,h_1,h_2} \text{ such that } \ |\lambda^{j,k,h_1,h_2}| \max(h_1, h_2) \leq 2\sqrt{\gamma}\}.$$

The following Lemma holds (*see [30]*):

Lemma 3.5.6. *Let $0 < \gamma < 1$. Then there exist T_γ such that for all $T > T_\gamma$ there exist $k_{T,\gamma} > 0$ and $K_{T,\gamma} > 0$ such that*

$$k_{T,\gamma} E_h(0) \leq \int_0^T \|B_h U(t)\|^2_{Y_h}\, dt \leq K_{T,\gamma} E_h(0) \tag{3.5.12}$$

holds for every solution of (3.5.8) in the class $\widehat{\mathcal{C}_\gamma}(h)$ and every h_1, h_2 small enough satisfying

$$\sup \Big| \frac{h_1}{h_2} \Big| < \sqrt{\frac{\gamma}{4-\gamma}}.$$

Now we present the time discrete schemes we are interested in. For any $\triangle t > 0$, we consider the following time Newmark approximation scheme of system (3.5.8):

$$\begin{cases} \dfrac{U^{k+1} + U^{k-1} - 2U^k}{(\triangle t)^2} + A_{0,h}\Big(\beta U^{k+1} + (1-2\beta)U^k + \beta U^{k-1}\Big) = 0, \\ \Big(\dfrac{U^0 + U^1}{2}, \dfrac{U^1 - U^0}{\triangle t}\Big) = (U_{h,0}, U_{h,1}), \end{cases} \tag{3.5.13}$$

with $\beta \geq 1/4$.

The energy of (3.5.13) given by

$$\begin{aligned} E^k = \frac{1}{2} \Big\| A_{0,h}^{1/2} \Big(\frac{U^k + U^{k+1}}{2} \Big) \Big\|^2 + \frac{1}{2} \Big\| \frac{U^{k+1} - U^k}{\triangle t} \Big\|^2 \\ + (4\beta - 1) \frac{(\triangle t)^2}{8} \Big\| A_{0,h}^{1/2} \Big(\frac{U^{k+1} - U^k}{\triangle t} \Big) \Big\|^2, \end{aligned} \tag{3.5.14}$$

which is a discrete counterpart of the time continuous energy (3.3.19) and remains constant (see (3.3.27) as well).

In view of (3.5.12), conditions (B1) and (B2) are satisfied. Besides, conditions (A1) and (A2) are straightforward. Therefore the following theorem can be obtained as a direct consequence of Theorem 3.3.2:

Theorem 3.5.7. *Set $\beta \geq 1/4$. Set $0 < \gamma < 1$. Assume that the mesh sizes h_1, h_2 and $\triangle t$ tend to zero and*

$$\sup \left|\frac{h_1}{h_2}\right| < \sqrt{\frac{\gamma}{4-\gamma}}, \qquad \frac{\max\{h_1, h_2\}}{\triangle t} \leq \tau, \tag{3.5.15}$$

where τ is a positive constant.

Then, for any $0 < \delta \leq 2\sqrt{\gamma}/\tau$, there exist $T_\delta > 0$ such that for any $T > T_\delta$, there exists $k_{T,\delta,\gamma} > 0$ such that the observability inequality

$$k_{T,\delta,\gamma} E^k \leq \triangle t \sum_{k\triangle t \in (0,T)} \left\| B_h U^k \right\|^2_{Y_h}$$

holds for every solution of (3.5.13) *with initial data in the class*

$$\mathcal{C}^h_{\delta/\triangle t} = \text{span}\,\{\varphi^{j,k,h_1,h_2} \text{ such that } |\lambda^{j,k,h_1,h_2}| \leq \delta/\triangle t\}$$

for $h_1, h_2, \triangle t$ small enough satisfying (3.5.15).

Proof. We are in the setting given before and thus Lemma 3.5.4 applies. Hence, to apply Theorem 3.3.1, we only need to verify that $\mathcal{C}^h_{\delta/\triangle t} \subset \widehat{\mathcal{C}_\gamma}(h)$. But

$$|\lambda| < \frac{\delta}{\triangle t} \Rightarrow |\lambda| \leq 2\frac{\sqrt{\gamma}}{\tau \triangle t} \leq 2\frac{\sqrt{\gamma}}{\max\{h_1, h_2\}}.$$

and this completes the proof. □

The 1-d string with rapidly oscillating density

In this paragraph, we consider a one-dimensional wave equation with rapidly oscillating density, which provides another example where the model under consideration depends on an extra parameter.

Let us state the problem. Let $\rho \in L^\infty(\mathbb{R})$ be a periodic function such that $0 < \rho_m \leq \rho(x) \leq \rho_M < \infty$, *a.e.* $x \in \mathbb{R}$. Given $\varepsilon > 0$, set $\rho^\varepsilon(x) = \rho(x/\varepsilon)$ and consider the one-dimensional wave equation

$$\begin{cases} \rho^\varepsilon(x)\ddot{u}^\varepsilon - \partial^2_{xx} u^\varepsilon = 0, & (x,t) \in (0,1) \times (0,T), \\ u^\varepsilon(0,t) = u^\varepsilon(1,t) = 0, & t \in (0,T), \\ u^\varepsilon(x,0) = u_0(x), \quad \dot{u}^\varepsilon(x,0) = v_0(x), & x \in (0,1). \end{cases} \tag{3.5.16}$$

We consider the observation operator

$$Bu^\varepsilon(t) = \partial_x u^\varepsilon(1,t). \tag{3.5.17}$$

The mathematical setting is the same as in Subsection 3.4.3 and therefore we do not recall it.

The eigenvalue problem for (3.5.16) reads

$$\rho^\varepsilon(x)\lambda^2\Phi + \partial^2_{xx}\Phi = 0, \quad x \in (0,1); \qquad \Phi(0) = \Phi(1) = 0. \tag{3.5.18}$$

For each $\varepsilon > 0$, there exists a sequence of eigenvalues

$$0 < \lambda_1^\varepsilon < \lambda_2^\varepsilon < \cdots < \lambda_n^\varepsilon < \cdots \to \infty$$

and a sequence of associated eigenfunctions $(\Phi_n^\varepsilon)_n$ which can be chosen to constitute an orthonormal basis in $L^2(0,1)$ with respect to the norm

$$\|\phi\|^2_{L^2} = \int_0^1 \rho^\varepsilon(x)|\phi(x)|^2\, dx.$$

In [3], the following is proved:

Theorem 3.5.8 ([3]). *There exists a positive number $D > 0$, such that the following holds:*

Let $T > 2\sqrt{\bar\rho}$, where $\bar\rho$ denotes the mean value of ρ. Then there exist two positive constants k_T and K_T such that for any initial data (u_0, v_0) in

$$\widetilde{\mathcal{C}}_{D/\varepsilon} = \operatorname{span}\{\Phi_n^\varepsilon \ : \ n < D/\varepsilon\},$$

the solution u^ε of (3.5.16) verifies

$$k_T \left\|(u_0,v_0)\right\|^2_{H_0^1(0,1)\times L^2(0,1)} \le \int_0^T |u_x^\varepsilon(1,t)|^2 dt \le K_T \left\|(u_0,v_0)\right\|^2_{H_0^1(0,1)\times L^2(0,1)}.$$

Given $\beta \ge 1/4$, let us consider the following time semi-discretization of (3.5.16)

$$\rho^\varepsilon(x)\Big(\frac{u^{\varepsilon,k+1} - 2u^{\varepsilon,k} + u^{\varepsilon,k-1}}{(\triangle t)^2}\Big) - \partial^2_{xx}\Big((1-2\beta)u^{\varepsilon,k} + \beta(u^{\varepsilon,k-1} + u^{\varepsilon,k+1})\Big) = 0, \qquad (x,k) \in (0,1)\times\mathbb{N}, \tag{3.5.19}$$

completed with the following boundary conditions and initial data

$$\begin{cases} u^{\varepsilon,k}(0) = u^{\varepsilon,k}(1) = 0, & k \in \mathbb{N}, \\ \Big(\dfrac{u^{\varepsilon,0} + u^{\varepsilon,1}}{2}\Big)(x) = u_0(x), \quad \Big(\dfrac{u^{\varepsilon,1} - u^{\varepsilon,0}}{\triangle t}\Big)(x) = v_0(x), & x \in (0,1). \end{cases} \tag{3.5.20}$$

Since conditions (A1)-(A2)-(B1)-(B2) hold, we get the following result as a consequence of Theorem 3.3.2:

Theorem 3.5.9. *Let $\delta > 0$ and $\beta \ge 1/4$. Assume that the parameters $\triangle t$ and ε tend to zero.*

Then there exists a time T_δ such that for any $T > T_\delta$, there exists a positive constant $k_{T,\delta}$ such that the observability inequality

$$k_{T,\delta} \left\|(u_0,v_0)\right\|^2_{H_0^1(0,1)\times L^2(0,1)} \le \triangle t \sum_{k\triangle t \in (0,T)} |u_x^{\varepsilon,k}(1)|^2 \tag{3.5.21}$$

holds for every solution of (3.5.19)-(3.5.20) with initial data (u_0, v_0) in the class

$$\mathcal{C}^\varepsilon_{\delta/\triangle t} = \operatorname{span}\{\Phi_n^\varepsilon : \ \lambda_n^\varepsilon \le \delta/\triangle t\} \cap \widetilde{\mathcal{C}}_{D/\varepsilon}$$

independently of $\triangle t$ and ε.

3.6 On the admissibility condition

The goal of this section is to provide admissibility results for the time-discrete schemes used throughout the paper. These results are complementary to the observability results proved in Theorems 3.2.1, 3.3.1 and 3.3.2 when dealing with controllability problems (see [16]).

3.6.1 The time-continuous setting

Let us assume that system (3.1.1)-(3.1.2) is admissible. By definition, there exists a positive constant K_T such that:

$$\int_0^T \|y(t)\|_Y^2 \, dt \leq K_T \|z_0\|_X^2 \qquad \forall \, z_0 \in \mathcal{D}(A). \tag{3.6.1}$$

The goal of this section is to prove that this property can be read on the wave packets setting as well.

Proposition 3.6.1. *System* (3.1.1)-(3.1.2) *is admissible if and only if*

$$\begin{cases} \text{There exist } \; r > 0 \; \text{ and } \; D > 0 \; \text{ such that} \\ \text{for all } n \in \Lambda \text{ and for all } \; z = \displaystyle\sum_{l \in J_r(\mu_n)} c_l \Phi_l : \quad \|Bz\|_Y \leq D \|z\|_X \,, \end{cases} \tag{3.6.2}$$

where

$$J_r(\mu) = \{l \in \mathbb{N}, \text{ such that } |\mu_l - \mu| \leq r\}. \tag{3.6.3}$$

Proof. We will prove separately the two implications.

First let us assume that system (3.1.1)-(3.1.2) is admissible.
Denote by

$$V(\omega, \varepsilon) = \text{span}\{\Phi_j \text{ such that } |\mu_j - \omega| \leq \varepsilon\}.$$

Then the following lemma holds:

Lemma 3.6.2. *Let us define* $K(\omega, \varepsilon)$ *as*

$$K(\omega, \varepsilon) = \left\| B(A - i\omega I)^{-1} \right\|_{\mathfrak{L}(V(\omega,\varepsilon)^*, Y)}.$$

Then for any $\varepsilon > 0$, $K(\omega, \varepsilon)$ *is uniformly bounded in* ω, *that is*

$$K(\varepsilon) = \sup_{\omega \in \mathbb{R}} K(\omega, \varepsilon) < \infty. \tag{3.6.4}$$

Besides, the following estimate holds

$$K(\varepsilon) \leq \sqrt{\frac{K_1}{1 - \exp(-1)}} \Big(1 + \frac{1}{\varepsilon}\Big), \tag{3.6.5}$$

where K_1 *is the admissibility constant in* (3.1.3).

Proof of Lemma 3.6.2. Let us first notice these resolvent identities:

$$\begin{aligned}(A - i\omega I) - I &= A - (1 + i\omega)I,\\ (A - (1 + i\omega)I)^{-1}(I - (A - i\omega I)^{-1}) &= (A - i\omega I)^{-1}.\end{aligned}$$

Hence

$$K(\omega, \varepsilon) \leq \left\|B(A - (1 + i\omega)I)^{-1}\right\|_{\mathfrak{L}(X,Y)} \ \left\|(I - (A - i\omega I)^{-1})\right\|_{\mathfrak{L}(V(\omega,\varepsilon)^*,X)}.$$

Obviously

$$\left\|(I - (A - i\omega I)^{-1})\right\|_{\mathfrak{L}(V(\omega,\varepsilon)^*,X)} \leq 1 + \frac{1}{\varepsilon}$$

Hence we restrict ourselves to the study of

$$\left\|B(A - (1 + i\omega)I)^{-1}\right\|_{\mathfrak{L}(X,Y)}.$$

Let us remark that for all $z = \sum a_j \Phi_j \in X$,

$$(A - (1 + i\omega)I)^{-1} z = \sum \frac{1}{i(\mu_j - \omega) - 1} a_j \Phi_j = \int_0^\infty \exp(-(1 + i\omega)t) z(t)\ dt, \tag{3.6.6}$$

where $z(t)$ is the solution of (3.1.1) with initial value z. This implies that

$$\begin{aligned}\left\|B(A - (1 + i\omega)I)^{-1} z\right\|_Y^2 &= \left\|\int_0^\infty \exp(-(1 + i\omega)t) Bz(t)\ dt\right\|_Y^2\\ &\leq \Big(\int_0^\infty \Big|\exp(-(1 + 2i\omega)t)\Big|\ dt\Big)\ \Big(\int_0^\infty \exp(-t)\left\|Bz(t)\right\|_Y^2\ dt\Big) \leq \int_0^\infty \exp(-t)\left\|Bz(t)\right\|_Y^2\ dt.\end{aligned}$$

But using the admissibility property of the operator B, we obtain

$$\begin{aligned}\int_0^\infty \exp(-t)\left\|Bz(t)\right\|_Y^2\ dt &\leq \sum_{k\in\mathbb{N}} \exp(-k) \int_k^{k+1} \left\|Bz(t)\right\|_Y^2\ dt\\ &\leq \Big(\sum_{k\in\mathbb{N}} \exp(-k)\Big) K_1 \left\|z\right\|_X^2 \leq \frac{K_1}{1 - \exp(-1)} \left\|z\right\|_X^2.\end{aligned}$$

The estimate (3.6.5) follows. □

Let us now consider a wave packet $z_0 = \sum_{l\in J_1(\mu_n)} c_l \Phi_l$. Then taking $\varepsilon = 1$ in Lemma 3.6.2, one gets that

$$\begin{aligned}\left\|Bz\right\|_Y &\leq \left\|B(A - i(\mu_n - 2)I)^{-1}\right\|_{\mathfrak{L}(V(\mu_n - 2,1)^*,Y)} \left\|(A - i(\mu_n - 2)I)z\right\|\\ &\leq K(1)\Big(\max_{l\in J_1(\mu_n)} |\mu_l - \mu_n| + 2\Big) \left\|z\right\| \leq 3K(1) \left\|z\right\|.\end{aligned}$$

Now we assume that estimate (3.6.2) holds for some $r > 0$ and $D > 0$. Set $z_0 \in \mathcal{D}(A)$, and expand z_0 as

$$z_0 = \sum_{k\in\mathbb{Z}} z_k, \qquad z_k = \sum_{l\in J_r(2kr)} c_l \Phi_l.$$

We need a special test function whose existence is established in the following Lemma:

Lemma 3.6.3. *There exists a time T and a function M satisfying*

$$\begin{cases} M(t) \geq 0, & |t| \geq T/2, \\ M(t) \geq 1, & |t| \leq T/2, \\ \text{Supp } \hat{M} \subseteq (-2r, 2r). \end{cases} \tag{3.6.7}$$

The proof is postponed to the end of this section. Note that functions satisfying similar properties appear naturally in the proofs of various Ingham's type inequalities, see [11, 26].

Taking Lemma 3.6.3 into account, we estimate

$$\begin{aligned} \int_0^T \|Bz(t)\|_Y^2 &\leq \int_{\mathbb{R}} M(t-T/2)\, \|Bz(t)\|_Y^2 \ dt \\ &\leq \sum_{k_1,k_2} \int_{\mathbb{R}} M(t-T/2) < Bz_{k_1}(t), Bz_{k_2}(t) >_{Y\times Y} \ dt. \end{aligned}$$

But these scalar products vanish most of the time. Indeed, if $|k_1 - k_2| \geq 2$, from (3.6.7), we get

$$\begin{aligned} &\int_{\mathbb{R}} M(t-T/2) < Bz_{k_1}(t), Bz_{k_2}(t) >_Y \ dt \\ &\qquad = \sum_{(l_1,l_2)\in J_r(2k_1\, r)\times J_r(2k_2\, r)} \hat{M}(\mu_{l_1} - \mu_{l_2}) < a_{l_1} B\Phi_{l_1}, a_{l_2} B\Phi_{l_2} >_Y = 0. \end{aligned}$$

This implies that

$$\begin{aligned} \int_0^T \|Bz(t)\|_Y^2 &\leq \int_{\mathbb{R}} M(t-T/2) \sum_k \Big(\|Bz_k(t)\|_Y^2 + 2\mathcal{R}e\langle Bz_k(t), Bz_{k+1}(t)\rangle_{Y\times Y} \Big) dt \\ &\leq 3\int_{\mathbb{R}} M(t-T/2) \sum_k \|Bz_k(t)\|_Y^2 \ dt \leq 3D \int_{\mathbb{R}} M(t-T/2) \sum_k \|z_k(t)\|_X^2 \ dt \\ &\leq 3D\hat{M}(0)\, \|z_0\|_X^2 \,. \end{aligned}$$

This completes the proof, since admissibility at time T is obviously equivalent to admissibility in any time. □

Proof of Lemma 3.6.3. In this proof, we do not care about the value of the parameters r and T that can be handled through a scaling argument.
Let us consider the function

$$f(t) = \frac{1}{\pi}\text{sinc}(t) = \frac{\sin(t)}{\pi t}.$$

It is well-known that its Fourier transform is $\hat{f}(\tau) = \chi_{(-1,1)}(\tau)$, where $\chi_{(-1,1)}$ denotes the characteristic function of $(-1, 1)$.

Hence, the function

$$M(t) = f(t)^2 = \frac{\text{sinc}^2(t)}{\pi^2}$$

satisfies the following properties

$$M(t) \geq \frac{2}{\pi^3}, \quad |t| < \frac{\pi}{4}; \quad M(t) \geq 0, \quad t \in \mathbb{R}; \quad \hat{M}(\tau) = (2 - |\tau|)_+, \quad \tau \in \mathbb{R}$$

and the proof is complete. For instance, for $r > 0$, one can take the function $M_r(t)$ as

$$M_r(t) = \frac{\pi^2}{8}\text{sinc}^2(rt) \tag{3.6.8}$$

which satisfies (3.6.7) with $T = \pi/2r$. □

Remark 3.6.4. In the context of families of pairs (A, B), according to Proposition 3.6.1, the uniform admissibility condition (3.5.1) is equivalent to a uniform wave packet estimate similar to (3.6.2). To be more precise, if $(\Phi_j^p)_{j\in\mathbb{N}}$ denotes the eigenvectors of A_p associated to the eigenvalues $(\lambda_j^p)_{j\in\mathbb{N}}$, that is $A_p\Phi_j^p = \lambda_j^p\Phi_j^p$, the uniform admissibility condition is equivalent to:

$$\begin{cases} \text{There exist} \quad r > 0 \quad \text{and} \quad D > 0 \quad \text{such that for all} \quad p,\ n \in \mathbb{N} \\ \text{and for all} \quad z = \displaystyle\sum_{l\in J_r(\lambda_n^p)} c_l\Phi_l^p : \qquad \|B_p z\|_{Y_p} \leq D\,\|z\|_{X_p}\,. \end{cases}$$

3.6.2 The time-discrete setting

This subsection is aimed to prove that if the continuous system (3.1.1)-(3.1.2) is admissible, in the sense of Definition 3.1.1, then its time semi-discrete approximation will be admissible as well under suitable assumptions. In this part, we will focus on the particular discretization given in Subsection 3.3.1, but everything works as well in all the time semi-discretization schemes considered in the article.

More precisely, we assume that the continuous system (3.1.1)-(3.1.2) is admissible, that is, from Proposition 3.6.1, the wave packet estimate (3.6.2) holds.

Then we claim that, under the assumptions (3.1.17), (3.3.1), (3.3.2), (3.3.3) and (3.3.4), the following discrete admissibility inequality holds:

Theorem 3.6.5. *Assume that system* (3.1.1)-(3.1.2) *is admissible. Set* $\delta > 0$. *For any* $T > 0$, *there exists a constant* $K_{T,\delta} > 0$ *such that for all* $\triangle t$ *small enough, the solution of equation* (3.1.17) *with initial data in* $\mathcal{C}_{\delta/\triangle t}$ *satisfies*

$$\triangle t \sum_{k=0}^{T/\triangle t} \left\|Bz^k\right\|_Y^2 \leq K_{T,\delta}\left\|z^0\right\|_X^2. \tag{3.6.9}$$

Proof. The proof follows the one given in the continuous case. First of all, let us remark the following straightforward fact: There exists $r_\delta > 0$ such that for all $n \in \mathbb{Z}$ satisfying $\triangle t|\lambda_{n,\triangle t}| \leq \delta$, for all $\triangle t > 0$, the set

$$\tilde{J}_{r_\delta}(\lambda_{n,\triangle t}) = \{l \in \mathbb{Z}, \text{ such that } |\lambda_{l,\triangle t} - \lambda_{n,\triangle t}| \leq r_d\},$$

where $\lambda_{l,\triangle t}$ is as in (3.3.2), is a subset of $J_r(\mu_n)$ (recall (3.6.3)). Besides, one can take:

$$r_\delta = r\inf\{|h'(\eta)|,\ |\eta| \leq \delta\}.$$

Note that condition (3.3.3) implies the positivity of the right hand side.

Given $\triangle t > 0$, assume that there is a time T and a function $M^{\triangle t} \in l^2(\triangle t\mathbb{Z})$ such that

$$\begin{cases} M^{\triangle t,k} \geq 0, & |k\triangle t| \geq T/2, \\ M^{\triangle t,k} \geq 1, & |t| \leq T/2, \\ \text{Supp } \hat{M}^{\triangle t} \subseteq (-2r_\delta, 2r_\delta), \end{cases} \tag{3.6.10}$$

where this time $\hat{M}^{\triangle t}$ denotes the discrete Fourier transform at scale $\triangle t$ defined in Definition 3.2.3. One can easily check that we can take $M^{\triangle t} = M_{r_\delta}$ for all $\triangle t > 0$ where M_{r_δ} is as in (3.6.8).

With this definition, the proof of inequality (3.6.9) consists in rewriting the one of Proposition 3.6.1 by replacing the continuous integrals and the Fourier transform by their discrete versions. Since all the steps are independent of $\triangle t$, the admissibility inequality holds uniformly. □

Note that this proof can be applied to derive uniform admissibility results for families of operators (A, B) within the class $\mathfrak{D}(C_B, T, k_T, K_T)$ for the fully discrete schemes. Indeed, in the setting of Section 3.5, according to Remark 3.6.4, the proof presented above directly implies uniform admissibility properties for operators in the class $\mathfrak{D}(C_B, T, k_T, K_T)$ when the initial data are taken in the filtered class $\mathcal{C}_{\delta/\triangle t}$.

3.7 Further comments and open problems

1. The resolvent estimate is a useful tool to analyze time-discrete approximation schemes, as we have seen in this paper. However, although this method is quite robust, it does not allow to deal with observability inequalities with loss, arising, for instance, when dealing with networks of vibrating strings (see [5, Chapter 4]) or for the wave equation in the absence of the Geometric Control Conditions (see [13, 15]). In those cases one only needs a weaker version of the observability inequality (3.1.4), in which the observed norm is weaker than $\|\cdot\|_X$. Actually, this question is also open at the continuous level.

2. As said in Remark 3.4.9, we are not able to recover the optimal value of the time of observability for systems (3.1.1)–(3.1.2) and their time-discrete approximation schemes. This is a drawback of the method based on the resolvent estimate. Indeed, even in the continuous setting, to our knowledge, this method does not allow to recover the optimal time of observability.

3. There are several different methods to derive uniform observability inequalities for systems (3.4.19). In [28], a discrete multiplier technique is developed to derive the uniform observability of the time semi-discrete wave equation in a bounded domain. There, the same order of filtering parameter $\delta/(\triangle t)$ is attained but a smallness condition on δ is imposed. Theorem 3.3.2 generalizes this result to any $\delta > 0$, as the dispersion diagram analysis in [28] suggests.

4. Along the paper, we derived uniform observability inequalities and admissibility results for time-discretization schemes of abstract first order and second order (in time) systems. As it is well-known in controllability theory, they imply uniform controllability results as well. For instance, in the context of the time-discrete wave equation analyzed in [28], combining the duality arguments in it and the results of this paper, one can immediately deduce the uniform (with respect to $\triangle t > 0$) controllability of projections on the classes of filtered space $\mathcal{C}_{\delta/\triangle t}$, for $T > T_\delta$ large enough and $\delta > 0$ arbitrary. This improves the results in [28] that required the filtering parameter $\delta > 0$ to be small enough.

The same duality arguments combined with the uniform observability and admissibility results we have presented in this paper allow proving uniform controllability results in a number of other cases including the time-discrete KdV and Schrödinger equations, the fully discrete wave equation, the time-discretization of wave equations with rapidly oscillating coefficients, etc.

5. In this paper, we have only dealt with observability properties of time-discrete conservative systems, but the same questions arise for dissipative systems. However the situation is completely different for unbounded dissipative perturbations. One such example is the heat equation for which, as far as we know, there is no resolvent characterization of the well-known properties of observability from an arbitrarily small observation set and time. The observability of time-discrete heat equations has been analyzed in [29] for the heat operator. But as far as we know, there is no systematic way of transferring the known results on space semi-discretizations (see [32]) to observability properties of full discretization schemes. At this respect the article [12] is also worth mentioning in which the existing results on the control of continuous parabolic equations are transformed into approximate controllability results for space semi-discretizations, with an explicit estimate of the error term.

Bibliography

[1] C. Bardos, G. Lebeau, J. Rauch, Sharp sufficient conditions for the observation, control, and stabilization of waves from the boundary, SIAM J. Control Optim. 30 (5) (1992) 1024–1065.

[2] N. Burq, M. Zworski, Geometric control in the presence of a black box, J. Amer. Math. Soc. 17 (2) (2004) 443–471 (electronic).

[3] C. Castro, Boundary controllability of the one-dimensional wave equation with rapidly oscillating density, Asymptot. Anal. 20 (3-4) (1999) 317–350.

[4] C. Castro, S. Micu, Boundary controllability of a linear semi-discrete 1-D wave equation derived from a mixed finite element method, Numer. Math. 102 (3) (2006) 413–462.

[5] R. Dáger, E. Zuazua, Wave propagation, observation and control in $1-d$ flexible multi-structures, Vol. 50 of Mathématiques & Applications (Berlin), Springer-Verlag, Berlin, 2006.

[6] S. Ervedoza, Observability of the mixed finite element method for the 1d wave equation on non-uniform meshes, *To appear in ESAIM: COCV*, 2008. *Cf Chapitre 2.*

[7] S. Ervedoza, E. Zuazua, Perfectly matched layers in 1-d: Energy decay for continuous and semi-discrete waves, Numer. Math., 109(4):597–634, 2008. *Cf Chapitre 1.*

[8] E. Hairer, S. P. Nørsett, G. Wanner, Solving ordinary differential equations. I, 2nd Edition, Vol. 8 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1993.

[9] E. Hairer, G. Wanner, Solving ordinary differential equations. II, Vol. 14 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1991.

[10] J.-A. Infante, E. Zuazua, Boundary observability for the space-discretizations of the 1-d wave equation, C. R. Acad. Sci. Paris Sér. I Math. 326 (6) (1998) 713–718.

[11] A. E. Ingham, Some trigonometrical inequalities with applications to the theory of series, Math. Z. 41 (1) (1936) 367–379.

[12] S. Labbé, E. Trélat, Uniform controllability of semi-discrete approximations of parabolic control systems, Systems Control Lett., 55 (7)(2006) 597–609.

[13] G. Lebeau, Contrôle analytique. I. Estimations a priori, Duke Math. J. 68 (1) (1992) 1–30.

[14] G. Lebeau, Contrôle de l'équation de Schrödinger, J. Math. Pures Appl. (9) 71 (3) (1992) 267–291.

[15] G. Lebeau, L. Robbiano, Stabilization of the wave equation by the boundary, in: Partial differential equations and mathematical physics (Copenhagen, 1995; Lund, 1995), Vol. 21 of Progr. Nonlinear Differential Equations Appl., Birkhäuser Boston, Boston, MA, 1996, pp. 207–210.

[16] J.-L. Lions, Contrôlabilité exacte, perturbations et stabilisation de systèmes distribués. Tome 2, Vol. 9 of Recherches en Mathématiques Appliquées, Masson, Paris, 1988.

[17] F. Macià, E. Zuazua, On the lack of observability for wave equations: a Gaussian beam approach, Asymptot. Anal. 32 (1) (2002) 1–26.

[18] L. Miller, Controllability cost of conservative systems: resolvent condition and transmutation, J. Funct. Anal. 218 (2) (2005) 425–444.

[19] A. Münch, A uniformly controllable and implicit scheme for the 1-D wave equation, M2AN Math. Model. Numer. Anal. 39 (2) (2005) 377–418.

[20] M. Negreanu, E. Zuazua, Uniform boundary controllability of a discrete 1-D wave equation, Systems Control Lett. 48 (3-4) (2003) 261–279, optimization and control of distributed systems.

[21] M. Negreanu, E. Zuazua, Convergence of a multigrid method for the controllability of a 1-d wave equation, C. R. Math. Acad. Sci. Paris 338 (5) (2004) 413–418.

[22] K. Ramdani, T. Takahashi, G. Tenenbaum, M. Tucsnak, A spectral approach for the exact observability of infinite-dimensional systems with skew-adjoint generator, J. Funct. Anal. 226 (1) (2005) 193–229.

[23] L. Rosier, Exact boundary controllability for the Korteweg-de Vries equation on a bounded domain, ESAIM Control Optim. Calc. Var. 2 (1997) 33–55 (electronic).

[24] D. L. Russell, B. Y. Zhang, Controllability and stabilizability of the third-order linear dispersion equation on a periodic domain, SIAM J. Control Optim. 31 (3) (1993) 659–676.

[25] L. N. Trefethen, Group velocity in finite difference schemes, SIAM Rev. 24 (2) (1982) 113–136.

[26] M. Tucsnak, G. Weiss, Passive and conservative linear systems, Preprint.

[27] G. Weiss, Admissible observation operators for linear semigroups, Israel J. Math. 65 (1) (1989) 17–43.

[28] X. Zhang, C. Zheng, E. Zuazua, Exact controllability of the time discrete wave equation, Preprint on Discrete Contin. Dyn. Syst.

[29] C. Zheng, Controllability of the time discrete heat equation, Preprint.

[30] E. Zuazua, Boundary observability for the finite-difference space semi-discretizations of the 2-D wave equation in the square, J. Math. Pures Appl. (9) 78 (5) (1999) 523–563.

[31] E. Zuazua, Propagation, observation, and control of waves approximated by finite difference methods, SIAM Rev. 47 (2) (2005) 197–243 (electronic).

[32] E. Zuazua, Control and numerical approximation of the wave and heat equations, Proceedings of the ICM Madrid 2006, Vol. III, "Invited Lectures", Eur. Math. Soc., Zürich, (2006) 1389-1417.

Chapter 4

Uniform exponential decay for viscous damped systems

Joint work with Enrique Zuazua.

Abstract: We consider a class of viscous damped vibrating systems. We prove that, under the assumption that the damping term ensures the exponential decay for the corresponding inviscid system, then the exponential decay rate is uniform for the viscous one, regardless what the value of the viscosity parameter is. Our method is mainly based on a decoupling argument of low and high frequencies. Low frequencies can be dealt with because of the effectiveness of the damping term in the inviscid case while the dissipativity of the viscous term guarantees the decay of the high frequency components. This method is inspired in previous work by the authors on time-discretization schemes for damped systems in which a numerical viscosity term needs to be added to ensure the uniform exponential decay with respect to the time-step parameter.

4.1 Introduction

Let X and Y be Hilbert spaces endowed with the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ respectively. Let $A : \mathcal{D}(A) \subset X \to X$ be a skew-adjoint operator with compact resolvent and $B \in \mathfrak{L}(X, Y)$.

We consider the system described by

$$\dot{z} = Az + \varepsilon A^2 z - B^* B z, \quad t \geq 0, \qquad z(0) = z_0 \in X. \tag{4.1.1}$$

Here and henceforth, a dot ($\dot{}$) denotes differentiation with respect to time t. The element $z_0 \in X$ is the initial state, and $z(t)$ is the state of the system. Most of the linear equations modeling the damped viscous vibrations of elastic structures (strings, beams, plates,...) can be written in the form (4.1.1) or some variants that we shall also discuss, in which the viscosity term has a more general form, namely,

$$\dot{z} = Az + \varepsilon \mathcal{V}_\varepsilon z - B^* B z, \quad t \geq 0, \qquad z(0) = z_0 \in X, \tag{4.1.2}$$

for a suitable viscosity operator $\mathcal{V}_\varepsilon$, which might depend on ε.

We define the energy of the solutions of system (4.1.1) by

$$E(t) = \frac{1}{2} \|z(t)\|_X^2, \quad t \geq 0, \tag{4.1.3}$$

which satisfies

$$\frac{dE}{dt}(t) = - \|Bz(t)\|_Y^2 - \varepsilon ||Az||_X^2, \quad t \geq 0. \tag{4.1.4}$$

In this paper, we assume that system (4.1.1) is exponentially stable when $\varepsilon = 0$. For the sake of completeness and clarity we distinguish the case in which the viscosity parameter vanishes

$$\dot{z} = Az - B^*Bz, \quad t \geq 0, \qquad z(0) = z_0 \in X. \tag{4.1.5}$$

This model corresponds to a conservative system in which a bounded damping term has been added. The damped wave and Schrödinger equations enter in this class, for instance.

Thus, we assume that there exist positive constants μ and ν such that any solution of (4.1.5) satisfies

$$E(t) \leq \mu \; E(0) \exp(-\nu t), \quad t \geq 0. \tag{4.1.6}$$

Our goal is to prove that the exponential decay property (4.1.6) for (4.1.5) implies the uniform exponential decay of solutions of (4.1.1) with respect to the viscosity parameter $\varepsilon > 0$.

This result might seem immediate a priori since the viscous term that (4.1.1) adds to (4.1.5) should in principle increase the decay rate of the solutions of the later. But, this is far from being trivial because of the possible presence of overdamping phenomena. Indeed, in the context of the damped wave equation, for instance, it is well known that the decay rate does not necessarily behave monotonically with respect to the size of the damping operator (see, for instance, [6, 7, 15]). In our case, however, the viscous damping operator is such that the decay rate is kept uniformly on ε. This is so because it adds dissipativity to the high frequency components, while it does not deteriorate the low frequency damping that the bounded feedback operator $-B^*B$ introduces.

The main result of this paper is that system (4.1.1) enjoys a uniform stabilization property. It reads as follows:

Theorem 4.1.1. *Assume that system* (4.1.5) *is exponentially stable and satisfies* (4.1.6) *for some positive constants* μ *and* ν*, and that* $B \in \mathfrak{L}(X, Y)$*.*

Then there exist two positive constants μ_0 *and* ν_0 *depending only on* $\|B\|_{\mathfrak{L}(X,Y)}$*,* ν *and* μ *such that any solution of* (4.1.1) *satisfies* (4.1.6) *with constants* μ_0 *and* ν_0 *uniformly with respect to the viscosity parameter* $\varepsilon > 0$*.*

Our strategy is based on the fact that the uniform exponential decay properties of the energy for systems (4.1.5) and (4.1.1), respectively, are equivalent to observability properties for the conservative system

$$\dot{y} = Ay, \quad t \in \mathbb{R}, \qquad y(0) = y_0 \in X, \tag{4.1.7}$$

and its viscous counterpart

$$\dot{u} = Au + \varepsilon A^2 u, \quad t \in \mathbb{R}, \qquad u(0) = u_0 \in X. \tag{4.1.8}$$

For (4.1.7) the observability property consists in the existence of a time $T^* > 0$ and a positive constant $k_* > 0$ such that

$$k_* \|y_0\|_X^2 \leq \int_0^{T^*} \|By(t)\|_Y^2 \; dt, \tag{4.1.9}$$

for every solution of (4.1.7) (see [11]).

A similar argument can be applied to the viscous system (4.1.8). In this case the relevant inequality is the following: There exist a time $T > 0$ and a constant $k_T > 0$ such that any solution of (4.1.8) satisfies

$$k_T \|u_0\|_X^2 \leq \int_0^T \|Bu(t)\|_Y^2 \ dt + \varepsilon \int_0^T \|Au(t)\|_X^2 \ dt. \tag{4.1.10}$$

Note however that, for the uniform exponential decay property of the solutions of (4.1.1) to be independent of ε, we also need the time T and the observability constant k_T in (4.1.10) to be uniform. Actually we will prove the observability property (4.1.10) for the time $T = T^*$ given in (4.1.9).

The observability inequality (4.1.10) can not be obtained directly from (4.1.9) since the viscosity operator εA^2 is an unbounded perturbation of the dynamics associated to the conservative system (4.1.7). Therefore, we decompose the solution u of (4.1.8) into its low and high frequency parts, that we handle separately. We first use the observability of (4.1.7) to prove (4.1.10), uniformly on ε, for the low frequency components. Second, we use the dissipativity of (4.1.8) to obtain a similar estimate for the high-frequency components.

In this way, we derive observability properties of the low and high frequency components separately, that, together, yield the needed observability property (4.1.10) leading to the uniform exponential decay result.

Our arguments do not apply when the damping operator B is not bounded, as it happens when the damping is concentrated on the boundary for the wave equation, see for instance [7]. Dealing with unbounded damping operators B needs further work.

As we mentioned above, the results in this paper are related with the literature on the uniform stabilization of numerical approximation schemes for damped equations of the form (4.1.5) and in particular with [21, 20, 18, 19, 9]. Similar techniques have also been employed to obtain uniform dispersive estimates for numerical approximation schemes to Schrödinger equations in [12].

The recent work [8] is also worth mentioning. There, observability issues were discussed for time and fully discrete approximation schemes of (4.1.7) and was one of the sources of motivation for this work.

The outline of this paper is as follows.
In Section 4.2, we recall the results of [8] and prove Theorem 4.1.1. In Section 4.3, we present a generalization of Theorem 4.1.1 to other viscosity operators. We also specify an application of our technique for viscous second order in time evolution equations which fit (4.1.2). In Section 4.4, we present some applications to viscous approximations of damped Schrödinger and wave equations. Finally, some further comments and open problems are collected in Section 4.5.

4.2 Proof of Theorem 4.1.1

We first need to introduce some notations.

Since A is a skew-adjoint operator with compact resolvent, its spectrum is discrete and $\sigma(A) = \{i\mu_j \ : \ j \in \mathbb{N}\}$, where $(\mu_j)_{j\in\mathbb{N}}$ is a sequence of real numbers such that $|\mu_j| \to \infty$ when $j \to \infty$. Set $(\Phi_j)_{j\in\mathbb{N}}$ an orthonormal basis of eigenvectors of A associated to the eigenvalues $(i\mu_j)_{j\in\mathbb{N}}$, that is

$$A\Phi_j = i\mu_j\Phi_j. \tag{4.2.1}$$

Moreover, define

$$\mathcal{C}_s = \text{span}\,\{\Phi_j : \text{ the corresponding } i\mu_j \text{ satisfies } |\mu_j| \leq s\}. \tag{4.2.2}$$

In the sequel, we assume that system (4.1.5) is exponentially stable and that $B \in \mathfrak{L}(X, Y)$, i.e. there exists a constant K_B such that

$$\|Bz\|_Y \leq K_B \|z\|_X , \qquad \forall z \in X. \tag{4.2.3}$$

The proof is divided into several steps.

First, we write carefully the energy identity for z solution of (4.1.1).

Consider z a solution of (4.1.1). Its energy $\|z(t)\|_X^2$ satisfies

$$\|z(T)\|_X^2 + 2\int_0^T \|Bz(t)\|_Y^2 \; dt + 2\int_0^T \varepsilon \|Az(t)\|_X^2 \; dt = \|z(0)\|_X^2 . \tag{4.2.4}$$

Therefore our goal is to prove that, with T^* as in (4.1.9), there exists a constant $c > 0$ such that any solution of (4.1.1) satisfies

$$c\,\|z(0)\|_X^2 \leq \int_0^{T^*} \|Bz(t)\|_Y^2 \; dt + \varepsilon \int_0^{T^*} \|Az(t)\|_X^2 \; dt. \tag{4.2.5}$$

It is easy to see that, combining (4.2.4) and (4.2.5), the semigroup S_ε generated by (4.1.1) satisfies

$$\|S_\varepsilon(T^*)\| \leq \gamma = 1 - c, \tag{4.2.6}$$

for a constant $0 < \gamma < 1$ independent of $\varepsilon > 0$. This, by the semigroup property, yields the uniform exponential decay result.

We also claim that, for (4.2.5) to hold for the solutions of (4.1.1), it is sufficient to show (4.1.10) for solutions of (4.1.8). To do that, it is sufficient to follow the argument in [11] developed in the context of system (4.1.5).

We decompose z as $z = u + w$ where u is the solution of the system (4.1.8) with initial data $u(0) = z_0$ and w satisfies

$$\dot{w} = Aw + \varepsilon A^2 w - B^* Bz, \quad t \geq 0, \qquad w(0) = 0. \tag{4.2.7}$$

Indeed, multiplying (4.2.7) by w and integrating in time, we get

$$\|w(t)\|_X^2 + 2\varepsilon \int_0^t \|Aw(s)\|_X^2 \; ds + 2\int_0^t < Bz(s), Bw(s) >_Y \; ds = 0.$$

Using that B is bounded, this gives

$$\|w(t)\|_X^2 + 2\varepsilon \int_0^t \|Aw(s)\|_X^2 \; ds \leq \int_0^t \|Bz(s)\|_Y^2 + K_B^2 \int_0^t \|w(s)\|_X^2 \; ds. \tag{4.2.8}$$

Grönwall's inequality then gives a constant G, that depends only on K_B and T^*, such that

$$\sup_{t\in[0,T^*]} \Big\{ \|w(t)\|_X^2 \Big\} + \varepsilon \int_0^{T^*} \|Aw(s)\|_X^2 \; ds \leq G \int_0^{T^*} \|Bz(s)\|_Y^2 \; ds. \tag{4.2.9}$$

Therefore in the sequel we deal with solutions u of (4.1.8), for which we prove (4.1.10) for $T = T^*$.

As said in the introduction, we decompose the solution u of (4.1.8) into its low and high frequency parts. To be more precise, we consider

$$u_l = \pi_{1/\sqrt{\varepsilon}} u, \quad u_h = (I - \pi_{1/\sqrt{\varepsilon}})u, \tag{4.2.10}$$

where $\pi_{1/\sqrt{\varepsilon}}$ is the orthogonal projection on $\mathcal{C}_{1/\sqrt{\varepsilon}}$ defined in (4.2.2). Here the notation u_l and u_h stands for the low and high frequency components, respectively.

Note that both u_l and u_h are solutions of (4.1.8) since the projection $\pi_{1/\sqrt{\varepsilon}}$ and the viscosity operator A^2 commute.

Besides, u_h lies in the space $\mathcal{C}^{\perp}_{1/\sqrt{\varepsilon}}$, in which the following property holds:

$$\sqrt{\varepsilon}\,\|Ay\|_X \geq \|y\|_X\,, \quad \forall y \in \mathcal{C}^{\perp}_{1/\sqrt{\varepsilon}}. \tag{4.2.11}$$

In a first step, we compare u_l with y_l solution of (4.1.7) with initial data $y_l(0) = u_l(0)$. Now, set $w_l = u_l - y_l$. From (4.1.9), which is valid for solutions of (4.1.7), we get

$$k_* \|u_l(0)\|_X^2 = k_* \|y_l(0)\|_X^2 \leq 2\int_0^{T^*} \|Bu_l(t)\|_Y^2 \; dt + 2\int_0^{T^*} \|Bw_l(t)\|_Y^2 \; dt. \tag{4.2.12}$$

In the sequel, to simplify the notation, $c > 0$ will denote a positive constant that may change from line to line, but which does not depend on ε.

Let us therefore estimate the last term in the right hand side of (4.2.12). To this end, we write the equation satisfied by w_l, which can be deduced from (4.1.7) and (4.1.8):

$$\dot{w}_l = Aw_l + \varepsilon A^2 u_l, \quad t \geq 0, \qquad w_l(0) = 0.$$

Note that $w_l \in \mathcal{C}_{1/\sqrt{\varepsilon}}$, since u_l and y_l both belong to $\mathcal{C}_{1/\sqrt{\varepsilon}}$. Therefore, the energy estimate for w_l leads, for $t \geq 0$, to

$$\|w_l(t)\|_X^2 = -2\varepsilon \int_0^t < Au_l(s), Aw_l(s) >_X \, ds \leq \varepsilon \int_0^t \|Au_l(s)\|_X^2 \, ds + \int_0^t \|w_l(s)\|_X^2 \, ds.$$

Grönwall's Lemma applies and allows to deduce from (4.2.12) and the fact that the operator B is bounded, the existence of a positive c independent of ε, such that

$$c\,\|u_l(0)\|_X^2 \leq \int_0^{T^*} \|Bu_l(t)\|_Y^2 \; dt + \varepsilon \int_0^{T^*} \|Au_l(s)\|_X^2 \, ds.$$

Besides,

$$\int_0^{T^*} \|Bu_l(t)\|_Y^2 \; dt \leq 2\int_0^{T^*} \|Bu(t)\|_Y^2 \; dt + 2\int_0^{T^*} \|Bu_h(t)\|_Y^2 \; dt$$

and, since $u_h(t) \in \mathcal{C}^{\perp}_{1/\sqrt{\varepsilon}}$ for all t,

$$\int_0^{T^*} \|Bu_h(t)\|_Y^2 \; dt \leq K_B^2 \int_0^{T^*} \|u_h(t)\|_X^2 \; dt \leq K_B \varepsilon \int_0^{T^*} \|Au_h(t)\|_X^2 \; dt.$$

It follows that there exists $c > 0$ independent of ε such that

$$c\,\|u_l(0)\|_X^2 \leq \int_0^{T^*} \|Bu(t)\|_Y^2 \; dt + \varepsilon \int_0^{T^*} \|Au(s)\|_X^2 \, ds. \tag{4.2.13}$$

Let us now consider the high frequency component u_h. Since $u_h(t)$ is a solution of (4.1.8) and belongs to $\mathcal{C}^{\perp}_{1/\sqrt{\varepsilon}}$ for all time $t \geq 0$, the energy dissipation law for u_h solution of (4.1.8) reads

$$\|u_h(t)\|_X^2 + 2\varepsilon \int_0^t \|Au_h(s)\|_X^2 \ ds = \|u_h(0)\|_X^2 \,, \quad t \geq 0, \tag{4.2.14}$$

and

$$\|u_h(t)\|_X^2 \leq \exp(-2t) \, \|u_h(0)\|_X^2 \,, \quad \forall t \geq 0.$$

In particular, these two last inequalities imply the existence of a constant $c > 0$ independent of ε such that any solution u_h of (4.1.8) with initial data $u_h(0) \in \mathcal{C}^{\perp}_{1/\sqrt{\varepsilon}}$ satisfies

$$c \, \|u_h(0)\|_X^2 \leq \varepsilon \int_0^{T^*} \|Au_h(s)\|_X^2 \ ds. \tag{4.2.15}$$

Combining (4.2.13) and (4.2.15) leads to the observability inequality (4.1.10). This, combined with the arguments of [11] and (4.2.9), allows to prove that any solution z of (4.1.1) satisfies (4.2.5), and proves (4.2.6), from which Theorem 4.1.1 follows.

4.3 Variants of Theorem 4.1.1

4.3.1 General viscosity operators

Other viscosity operators could have been chosen. In our approach, we used the viscosity operator εA^2, which is unbounded, but we could have considered the viscosity operator

$$\varepsilon \mathcal{V}_\varepsilon = \frac{\varepsilon A^2}{I - \varepsilon A^2}, \tag{4.3.1}$$

which is well defined, since A^2 is a definite negative operator, and commutes with A. This choice presents the advantage that the viscosity operator now is bounded, keeping the properties of being small at frequencies of order less than $1/\sqrt{\varepsilon}$ and of order 1 on frequencies of order $1/\sqrt{\varepsilon}$ and more. Again, the same proof as the one presented above works.

The following result constitutes a generalization of Theorem 4.1.1, which applies to a wide range of viscosity operators, and, in particular, to (4.3.1).

Theorem 4.3.1. *Assume that system* (4.1.5) *is exponentially stable and satisfies* (4.1.6)*, and that* $B \in \mathfrak{L}(X, Y)$.

Consider a viscosity operator $\mathcal{V}_\varepsilon$ *such that*

1. $\mathcal{V}_\varepsilon$ *defines a self-adjoint definite negative operator.*
2. *The projection* $\pi_{1/\sqrt{\varepsilon}}$ *and the viscosity operator* $\mathcal{V}_\varepsilon$ *commute.*
3. *There exist positive constants* c *and* C *such that for all* $\varepsilon > 0$,

$$\begin{cases} \sqrt{\varepsilon} \left\| \left(\sqrt{-\mathcal{V}_\varepsilon} \right) z \right\|_X \leq C \, \|z\|_X \,, \ \forall z \in \mathcal{C}_{1/\sqrt{\varepsilon}}, \\ \sqrt{\varepsilon} \left\| \left(\sqrt{-\mathcal{V}_\varepsilon} \right) z \right\|_X \geq c \, \|z\|_X \,, \ \forall z \in \mathcal{C}^{\perp}_{1/\sqrt{\varepsilon}}. \end{cases}$$

Then the solutions of (4.1.2) *are exponentially decaying in the sense of* (4.1.6)*, uniformly with respect to the viscosity parameter* $\varepsilon \geq 0$.

The proof of Theorem 4.3.1 can be easily deduced from the one of Theorem 4.1.1 and is left to the reader.

Especially, note that the second item implies that both spaces $\mathcal{C}_{1/\sqrt{\varepsilon}}$ and $\mathcal{C}^{\perp}_{1/\sqrt{\varepsilon}}$ are left globally invariant by the viscosity operator $\mathcal{V}_\varepsilon$. Therefore, if $u_l \in \mathcal{C}_{1/\sqrt{\varepsilon}}$ and $u_h \in \mathcal{C}^{\perp}_{1/\sqrt{\varepsilon}}$, we have

$$< \mathcal{V}_\varepsilon(u_l + u_h), (u_l + u_h) >_X = < \mathcal{V}_\varepsilon u_l, u_l >_X + < \mathcal{V}_\varepsilon u_h, u_h >_X .$$

Also remark that the second item is always satisfied when the operators $\mathcal{V}_\varepsilon$ and A commute.

4.3.2 Wave type systems

In this subsection we investigate the exponential decay properties for viscous approximations of second order in time evolution equation.

Let H be a Hilbert space endowed with the norm $\|\cdot\|_H$. Let $A_0 : \mathcal{D}(A_0) \to H$ be a self-adjoint positive operator with compact resolvent and $C \in \mathfrak{L}(H, Y)$.

We then consider the initial value problem

$$\begin{cases} \ddot{v} + A_0 v + \varepsilon A_0 \dot{v} + C^* C \dot{v} = 0, \quad t \geq 0, \\ v(0) = v_0 \in \mathcal{D}(A_0^{1/2}), \quad \dot{v}(0) = v_1 \in H. \end{cases} \tag{4.3.2}$$

System (4.3.2) can be seen as a particular instance of (4.1.2) modeling wave and beams equations.

The energy of solutions of (4.3.2) is given by

$$E(t) = \frac{1}{2} \|\dot{v}(t)\|_H^2 + \frac{1}{2} \left\| A_0^{1/2} v(t) \right\|_H^2, \tag{4.3.3}$$

and satisfies

$$\frac{dE}{dt}(t) = - \|C\dot{v}(t)\|_Y^2 - \varepsilon \left\| A_0^{1/2} \dot{v}(t) \right\|_H^2. \tag{4.3.4}$$

As before, we assume that, for $\varepsilon = 0$, the system

$$\ddot{v} + A_0 v + C^* C \dot{v} = 0, \quad t \geq 0, \qquad v(0) = v_0 \in \mathcal{D}(A_0^{1/2}), \quad \dot{v}(0) = v_1 \in H, \tag{4.3.5}$$

is exponentially stable, i.e. (4.1.6) holds.

We are indeed in the setting of (4.1.2), since (4.3.2) can be written as

$$\dot{Z} = AZ + \varepsilon \mathcal{V}_\varepsilon Z - B^* B Z, \tag{4.3.6}$$

with

$$Z = \begin{pmatrix} v \\ \dot{v} \end{pmatrix}, \quad A = \begin{pmatrix} 0 & I \\ -A_0 & 0 \end{pmatrix}, \quad \mathcal{V}_\varepsilon = \begin{pmatrix} 0 & 0 \\ 0 & -A_0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & C \end{pmatrix}. \tag{4.3.7}$$

Note that the viscosity operator $\mathcal{V}_\varepsilon$ introduced in (4.3.7) does not satisfy Condition 1 in Theorem 4.3.1. Though, we can prove the following theorem:

Theorem 4.3.2. *Assume that system* (4.3.5) *is exponentially stable and satisfies* (4.1.6) *for some positive constants* μ *and* ν*, and that* $C \in \mathfrak{L}(H,Y)$*. Set* $K < \infty$*.*

Then there exist two positive constants μ_K *and* ν_K *depending only on* $\|C\|_{\mathfrak{L}(H,Y)}$*,* K*,* ν *and* μ *such that any solution of* (4.3.2) *satisfies* (4.1.6) *with constants* μ_0 *and* ν_0 *uniformly with respect to the viscosity parameter* $\varepsilon \in [0,K]$*.*

Before going into the proof, we introduce the spectrum of A_0. Since A_0 is self-adjoint positive definite with compact resolvent, its spectrum is discrete and $\sigma(A_0) = \{\lambda_j^2 : \ j \in \mathbb{N}\}$, where λ_j is an increasing sequence of real positive numbers such that $\lambda_j \to \infty$ when $j \to \infty$. Set $(\Psi_j)_{j\in\mathbb{N}}$ an orthonormal basis of eigenvectors of A_0 associated to the eigenvalues $(\lambda_j^2)_{j\in\mathbb{N}}$.

These notations are consistent with the ones introduced in Section 4.2, by setting A as in (4.3.7), and

$$\mu_{\pm j} = \pm\lambda_j, \qquad \Phi_j = \begin{pmatrix} \dfrac{1}{i\mu_j}\Psi_j \\ \Psi_j \end{pmatrix}.$$

For convenience, similarly as in (4.2.2), we define

$$\mathfrak{C}_s = \operatorname{span}\{\Psi_j : \text{ the corresponding } \lambda_j \text{ satisfies } |\lambda_j| \le s\}, \tag{4.3.8}$$

which satisfies $\mathcal{C}_s = (\mathfrak{C}_s)^2$.

Sketch of the proof. The proof of Theorem 4.3.2 closely follows the one of Theorem 4.1.1.

As before, we read the exponential stability of (4.3.5) into the following observability inequality: There exist a time T^* and a positive constant k_* such that any solution of

$$\ddot{y} + A_0 y = 0, \quad t \ge 0, \qquad y(0) = y_0 \in \mathcal{D}(A_0^{1/2}), \quad \dot{y}(0) = y_1 \in H, \tag{4.3.9}$$

satisfies

$$k_*\Big(\|y_1\|_H^2 + \left\|A_0^{1/2} y_0\right\|_H^2 \Big) \le \int_0^{T^*} \|C\dot{y}(t)\|_Y^2 \ dt. \tag{4.3.10}$$

Due to (4.3.4), as in (4.2.5), the exponential decay of the energy for solutions of (4.3.2) is equivalent to the following observability inequality: There exist a time $\tilde{T}$ and a positive constant c such that for any $\varepsilon \in [0,K]$,

$$c\Big(\|v_1\|_H^2 + \left\|A_0^{1/2} v_0\right\|_H^2 \Big) \le \int_0^{\tilde{T}} \|C\dot{v}(t)\|_Y^2 \ dt + \varepsilon \int_0^{\tilde{T}} \left\|A_0^{1/2}\dot{v}(t)\right\|_H^2 \ dt \tag{4.3.11}$$

holds for any solution v of (4.3.2).

Using the same perturbative arguments as in [11] or (4.2.7)-(4.2.9), the observability inequality (4.3.11) holds if and only if there exist a time T and a positive constant $k_T > 0$ such that, for any $\varepsilon \in [0,K]$, the observability inequality

$$k_T\Big(\|u_1\|_H^2 + \left\|A_0^{1/2} u_0\right\|_H^2 \Big) \le \int_0^{T} \|C\dot{u}(t)\|_Y^2 \ dt + \varepsilon \int_0^{T} \left\|A_0^{1/2}\dot{u}(t)\right\|_H^2 \ dt \tag{4.3.12}$$

holds for any solution u of

$$\ddot{u} + A_0 u + \varepsilon A_0 \dot{u} = 0, \quad t \ge 0, \qquad u(0) = u_0 \in \mathcal{D}(A_0^{1/2}), \quad \dot{u}(0) = u_1 \in H. \tag{4.3.13}$$

As before, we then focus on the observability inequality (4.3.12) for solutions of (4.3.13). As in the proof of Theorem 4.1.1, we now decompose the solution of (4.3.13) into its low and high frequency parts, that we handle separately. To be more precise, we consider

$$u_l = P_{1/\sqrt{\varepsilon}}\, u, \qquad u_h = (I - P_{1/\sqrt{\varepsilon}})u.,$$

where $P_{1/\sqrt{\varepsilon}}$ is the orthogonal projection in H on $\mathfrak{C}_{1/\sqrt{\varepsilon}}$ as defined in (4.3.8). Again, both u_l and u_h are solutions of (4.3.13) since $P_{1/\sqrt{\varepsilon}}$ commute with A_0.

Arguing as before, the low frequency component u_l can be compared to y_l solution of (4.3.9) with initial data $(y_0, y_1) = (P_{1/\sqrt{\varepsilon}}u_0, P_{1/\sqrt{\varepsilon}}u_1)$, and using (4.3.10) for solutions of (4.3.9), we obtain the existence of a positive constant c_1 such that

$$c_1\Big(\left\|P_{1/\sqrt{\varepsilon}}u_1\right\|_H^2 + \left\|A_0^{1/2}P_{1/\sqrt{\varepsilon}}u_0\right\|_H^2 \Big) \le \int_0^{T^*} \|C\dot{u}(t)\|_Y^2 \ dt + \varepsilon \int_0^{T^*} \left\|A_0^{1/2}\dot{u}(t)\right\|_H^2 \ dt. \tag{4.3.14}$$

For the high frequency component u_h, the situation is slightly more intricate than in Theorem 4.1.1. The energy of the solution u_h satisfies the dissipation law

$$\frac{1}{2}\frac{d}{dt}\Big(\|\dot{u}_h(t)\|_H^2 + \left\|A_0^{1/2}u_h(t)\right\|_H^2 \Big) = -\varepsilon \left\|A_0^{1/2}\dot{u}_h\right\|_H^2 \le -\|\dot{u}_h\|_H^2\,, \tag{4.3.15}$$

where the last inequality comes from $\dot{u}_h \in \mathfrak{C}^{\perp}_{1/\sqrt{\varepsilon}}$.

Setting

$$E_h(t) = \frac{1}{2}\|\dot{u}_h(t)\|_H^2 + \frac{1}{2}\left\|A_0^{1/2}u_h(t)\right\|_H^2,$$

we thus obtain that

$$E_h(t) + \int_0^t \|\dot{u}_h(s)\|_H^2 \ ds \le E_h(0). \tag{4.3.16}$$

We now prove the so-called equirepartition of the energy for the solutions u of (4.3.13). Multiplying (4.3.13) by u and integrating by parts between 0 and t, we obtain

$$< \dot{u}(t), u(t) >_H - < \dot{u}(0), u(0) >_H - \int_0^t \|\dot{u}(s)\|_H^2 \ ds + \int_0^t \left\|A_0^{1/2}u(s)\right\|_H^2 \ ds \\ + \varepsilon \int_0^t < A_0^{1/2}\dot{u}(s), A_0^{1/2}u(s) >_H \ ds = 0.$$

In particular,

$$\int_0^t \|\dot{u}(s)\|_H^2 \ ds = \int_0^t \left\|A_0^{1/2}u(s)\right\|_H^2 \ ds + \frac{\varepsilon}{2}\Big(\left\|A_0^{1/2}u(t)\right\|_H^2 - \left\|A_0^{1/2}u_0\right\|_H^2 \Big) \\ + < \dot{u}(t), u(t) >_H - < \dot{u}(0), u(0) >_H\,. \tag{4.3.17}$$

Now, for u_h, which is a solution of (4.3.13), for all $t \ge 0$, $u_h(t) \in \mathfrak{C}^{\perp}_{1/\sqrt{\varepsilon}}$. In particular, for all $t \ge 0$, we have

$$\Big| < \dot{u}_h(t), u_h(t) >_H \Big| \le \frac{\sqrt{\varepsilon}}{2}\|\dot{u}_h\|_H^2 + \frac{1}{2\sqrt{\varepsilon}}\|u_h(t)\|_H^2 \le \sqrt{\varepsilon}E_h(t), \tag{4.3.18}$$

where we used that for $\phi \in \mathfrak{C}^{\perp}_{1/\sqrt{\varepsilon}}$,

$$\|\phi\|_H^2 \le \varepsilon \left\|A_0^{1/2}\phi\right\|_H^2.$$

Combining (4.3.18) with identity (4.3.17) for u_h, we obtain

$$\int_0^t \|\dot{u}_h(s)\|_H^2 \ ds \geq \int_0^t \left\|A_0^{1/2} u_h(s)\right\|_H^2 \ ds - \Big(\sqrt{\varepsilon}+\varepsilon\Big)(E_h(t)+E_h(0)). \tag{4.3.19}$$

This yields

$$\int_0^t \|\dot{u}_h(s)\|_H^2 \ ds \geq \int_0^t E_h(s) \ ds - \frac{1}{2}\Big(\sqrt{\varepsilon}+\varepsilon\Big)(E_h(t)+E_h(0)). \tag{4.3.20}$$

Combined with (4.3.16), we obtain

$$\Big(1-\frac{1}{2}(\sqrt{\varepsilon}+\varepsilon)\Big)E_h(t) + \int_0^t E_h(s) \ ds \leq E_h(0)\Big(1+\frac{1}{2}(\sqrt{\varepsilon}+\varepsilon)\Big) \tag{4.3.21}$$

Assuming that $K \geq 1$, which can always be assumed, for $\varepsilon \in [0,K]$, we thus have

$$(1-K)E_h(t) + \int_0^t E_h(s) \ ds \leq (1+K)E_h(0).$$

The decay of $E_h(t)$, guaranteed by the dissipation law (4.3.15), then proves that

$$(t+1-K)E_h(t) \leq (1+K)E_h(0).$$

For $t = 1+3K$, we thus have $E_h(1+3K) \leq E_h(0)/2$. We then deduce from the dissipation law (4.3.15) the existence of a positive constant c_K such that

$$c_K E_h(0) \leq \varepsilon \int_0^{1+3K} \left\|A_0^{1/2}\dot{u}_h(s)\right\|_H^2 \ ds. \tag{4.3.22}$$

We finally conclude Theorem 4.3.2 by combining (4.3.14) and (4.3.22) as before. □

Remark 4.3.3. One cannot expect the results of Theorem 4.3.2 to hold uniformly with respect to $\varepsilon \in [0,\infty]$. Indeed, an overdamping phenomenon appears when $\varepsilon \to \infty$. This can indeed be deduced from the existence of the following solutions of (4.3.13):

$$u_j(t) = \exp(t\tau_j^\varepsilon)\Psi_j, \quad t \geq 0, \quad \text{where } \tau_j^\varepsilon = \frac{\varepsilon\lambda_j^2}{2}\left(\sqrt{1-\frac{4}{(\varepsilon\lambda_j)^2}}-1\right) \underset{\varepsilon\lambda_j\to\infty}{\sim} -\frac{1}{\varepsilon}.$$

Plugging these solutions in (4.3.12), one can check that the observability inequality (4.3.12) cannot hold uniformly with respect to $\varepsilon \in [0,\infty)$. Finally, using the equivalence between the observability inequality (4.3.12) for solutions of (4.3.13) and the observability inequality (4.3.11) for solutions of (4.3.2), this proves that the results of Theorem 4.3.2 do not hold uniformly with respect to $\varepsilon \in [0,\infty]$.

Remark 4.3.4. To avoid the overdamping phenomenon when $\varepsilon \to \infty$, one can for instance add a dispersive term in (4.3.2), and consider the initial value problem

$$\begin{cases} \ddot{v} + A_0 v + \varepsilon A_0 \dot{v} + \varepsilon A_0 v + C^*C\dot{v} = 0, \quad t \geq 0, \\ v(0) = v_0 \in \mathcal{D}(A_0^{1/2}), \quad \dot{v}(0) = v_1 \in H. \end{cases} \tag{4.3.23}$$

The energy of solutions of (4.3.23) is now given by

$$E_\varepsilon(t) = \frac{1}{2}\|\dot{v}(t)\|_H^2 + \Big(\frac{1+\varepsilon}{2}\Big)\left\|A_0^{1/2}v(t)\right\|_H^2. \tag{4.3.24}$$

One can then prove that, if system (4.3.5) is exponentially stable, then the energy E_ε of solutions of systems (4.3.23) is exponentially stable, uniformly with respect to the viscosity parameter $\varepsilon \in [0, \infty)$. The proof can be done similarly as the one of Theorem 4.3.2 and is left to the reader. The main difference that the dispersive term introduces is that the high frequency solutions u_h of

$$\ddot{u}_h + A_0 u_h + \varepsilon A_0 \dot{u}_h + \varepsilon A_0 u_h = 0, \quad t \geq 0, \tag{4.3.25}$$

with initial data $(u_h(0), \dot{u}_h(0)) \in (\mathfrak{C}^\perp_{1/\sqrt{\varepsilon}})^2 \cap (\mathcal{D}(A_0^{1/2}) \times H)$ now satisfy, instead of (4.3.19), which deteriorates when $\varepsilon \to \infty$, the following property of equirepartition of the energy

$$\left| \int_0^t \|\dot{u}_h\|_H^2 \ ds - (1+\varepsilon) \int_0^t \left\| A_0^{1/2} u(s) \right\|_H^2 \ ds \right| \leq 2E_{h,\varepsilon}(t) + 2E_{h,\varepsilon}(0), \tag{4.3.26}$$

where $E_{h,\varepsilon}$ is the energy of the solutions u_h of (4.3.25).

4.4 Applications

This section is devoted to present some precise examples.

4.4.1 The viscous Schrödinger equation

Let Ω be a smooth bounded domain of $\mathbb{R}^N$.

Let us now consider the following damped Schrödinger equation:

$$\begin{cases} i\dot{z} + \Delta_x z + ia(x)\, z = 0, & \text{in } \Omega \times (0, \infty), \\ z = 0, & \text{on } \partial\Omega \times (0, \infty), \\ z(0) = z_0, & \text{in } \Omega, \end{cases} \tag{4.4.1}$$

where $a = a(x)$ is a nonnegative damping function in $L^\infty(\Omega)$, that we assume to be positive in some open subdomain ω of Ω, that is there exists $a_0 > 0$ such that

$$a(x) \geq a_0, \quad \forall x \in \omega. \tag{4.4.2}$$

The energy of solutions of (4.4.1), given by

$$E(t) = \frac{1}{2} \|z(t)\|^2_{L^2(\Omega)}, \tag{4.4.3}$$

satisfies

$$\frac{dE}{dt}(t) = -\int_\Omega a(x)|z(t,x)|^2 \ dx. \tag{4.4.4}$$

The stabilization problem for (4.4.1) has already been studied in the recent years. Let us briefly present some known results. Some of them concern the problem of exact controllability but, as explained for instance in [16], it is equivalent to the observability and the stabilization ones addressed in this article in the case where the damping operator B is bounded.

For instance, in [14], it is proved that the Geometric Control Condition (GCC) is sufficient to guarantee the stabilization property (4.1.6) for the damped Schrödinger equation (4.4.1). The GCC

can be, roughly, formulated as follows (see [2] for the precise setting): The subdomain ω of Ω is said to satisfy the GCC if there exists a time $T > 0$ such that all rays of Geometric Optics that propagate inside the domain Ω at velocity one reach the set ω in time less than T. This condition is necessary and sufficient for the stabilization property to hold for the wave equation.

But, in fact, the Schrödinger equation behaves slightly better than a wave equation from the stabilization point of view because of the infinite velocity of propagation and, in this case, the GCC is sufficient but not always necessary. For instance, in [13], it has been proved that when the domain Ω is a square, for any non-empty bounded open subset ω, the stabilization property (4.1.6) holds for system (4.4.1). Other geometries have been also dealt with: We refer to the articles [4, 1].

Now, we assume that ω satisfies the GCC and, consequently, that we are in a situation where the stabilization property (4.1.6) for (4.4.1) holds, and we consider the viscous approximations

$$\begin{cases} i\dot{z} + \Delta_x z + ia(x)\, z - i\sqrt{\varepsilon}\Delta_x z = 0, & \text{in } \Omega \times (0,\infty), \\ z = 0, & \text{on } \partial\Omega \times (0,\infty), \\ z(0) = z_0, & \text{in } \Omega, \end{cases} \tag{4.4.5}$$

where $\varepsilon \geq 0$.

System (4.4.1) can be seen as a Ginzburg-Landau type approximation. More precisely, system (4.4.1) is the inviscid limit of (4.4.5). We refer to the works [17, 3] where inviscid limits were analyzed in a nonlinear context.

For the stabilization problem, Theorem 4.3.1 applies and provides the following result:

Theorem 4.4.1. *Assume that system* (4.4.1) *is exponentially stable, i.e. it satisfies* (4.1.6).

Then the solutions of (4.4.5) *are exponentially decaying in the sense of* (4.1.6)*, uniformly with respect to the viscosity parameter* $\varepsilon \geq 0$.

Proof. Let us check the hypothesis of Theorem 4.3.1.

This example enters in the abstract setting given in the introduction: The operator $A = i\Delta_x$ with the Dirichlet boundary conditions is indeed skew-adjoint in $L^2(\Omega)$ with compact resolvent and domain $\mathcal{D}(A) = H^2 \cap H_0^1(\Omega) \subset L^2(\Omega)$. Since a is a nonnegative function, the damping term in (4.4.1) takes the form B^*Bz where B is defined as the multiplication by $\sqrt{a(x)}$, which is obviously bounded from $L^2(\Omega)$ to $L^2(\Omega)$.

The viscosity operator is

$$\varepsilon\mathcal{V}_\varepsilon = \sqrt{\varepsilon}\Delta_x = -i\sqrt{\varepsilon}A = -\sqrt{\varepsilon}|A|.$$

Obviously, this viscosity operator $\mathcal{V}_\varepsilon$ satisfies the assumptions 1, 2 and 3, and therefore Theorem 4.3.1 applies. □

4.4.2 The viscous damped wave equation

Again, let Ω be a smooth bounded domain of $\mathbb{R}^N$.

We now consider the damped wave equation

$$\begin{cases} \ddot{v} - \Delta_x v + a(x)\,\dot{v} = 0, & \text{in } \Omega \times (0,\infty), \\ v = 0, & \text{on } \partial\Omega \times (0,\infty), \\ v(0) = v_0, \quad \dot{v}(0) = v_1 & \text{in } \Omega, \end{cases} \tag{4.4.6}$$

where a is a nonnegative function as before, and satisfies (4.4.2) for some non-empty open subset ω of Ω.

The energy of solutions of (4.4.6), given by

$$E(t) = \frac{1}{2} \|\dot{v}\|^2_{L^2(\Omega)} + \frac{1}{2} \|\nabla v\|^2_{L^2(\Omega)}, \tag{4.4.7}$$

satisfies the dissipation law

$$\frac{dE}{dt}(t) = -\int_\Omega a(x)|\dot{v}|^2 \, dx. \tag{4.4.8}$$

We assume that system (4.4.6) is exponentially stable. From the works [2, 5], this is the case if and only if ω satisfies the Geometric Control Condition given above.

We now consider viscous approximations of (4.4.6) given, for $\varepsilon > 0$, by

$$\begin{cases} \ddot{v} - \Delta_x v + a(x)\dot{v} - \varepsilon \Delta_x \dot{v} = 0, & \text{in } \Omega \times (0,\infty), \\ v = 0, & \text{on } \partial\Omega \times (0,\infty), \\ v(0) = v_0 \in H^1_0(\Omega), \quad \dot{v}(0) = v_1 \in L^2(\Omega). \end{cases} \tag{4.4.9}$$

Setting $A_0 = -\Delta_x$ with Dirichlet boundary conditions and $C = \sqrt{a(x)}$, Theorem 4.3.2 applies:

Theorem 4.4.2. *Assume that ω satisfies the Geometric Control Condition.*

Then the solutions of (4.4.9) *decay exponentially, i.e. satisfy* (4.1.6) *uniformly with respect to the viscosity parameter $\varepsilon \in [0,1]$. To be more precise, there exist positive constants μ_0 and ν_0 such that for all $\varepsilon \in [0,1]$, for any initial data in $H^1_0(\Omega) \times L^2(\Omega)$, the solution of* (4.4.9) *satisfies*

$$E(t) \leq \mu_0 \, E(0) \exp(-\nu_0 t), \quad t \geq 0. \tag{4.4.10}$$

4.5 Further comments

1. In this article, we have identified a class of damped systems, with added viscosity term, in which overdamping does not occur. This is to be compared with the existing literature on the overdamping phenomenon for the damped wave equation ([6, 7]).

2. As we mentioned in the introduction, our methods and results require the assumption that the damping operator B is bounded. This is due to the method we employ, which is based on the equivalence between the exponential decay of the energy and the observability properties of the conservative system, that requires the damping operator to be bounded. However, in several relevant applications, as for instance when dealing with the problem of boundary stabilization of the wave equation (see [16]), the feedback law is unbounded, and our method does not apply. This issue requires further work.

3. The same methods allow obtaining numerical approximation schemes with uniform decay properties.

The discrete analogue of the viscosity term added above for the stabilization of the wave equation has already been discussed in the works [21, 20, 18, 9] for space semi-discrete approximation schemes of damped wave equations. In those articles, though, the viscosity term is needed due to the presence

of high-frequency spurious solutions that do not propagate and therefore are not efficiently damped by the damping operator B^*B when it is localized in space as in the examples considered above.

Following the same ideas as in [21, 20, 18, 9], if observability properties such as (4.1.9) hold for fully discrete approximation schemes of the conservative linear system (4.1.7) in a filtered space (see [8]), then adding a suitable viscosity term to the corresponding fully discrete version of the dissipative system (4.1.5) suffices to obtain uniform (with respect to space time discretization parameters) stabilization properties. This issue is currently investigated by the authors and will be published in [10].

Bibliography

[1] B. Allibert. Contrôle analytique de l'équation des ondes et de l'équation de Schrödinger sur des surfaces de revolution. *Comm. Partial Differential Equations*, 23(9-10):1493–1556, 1998.

[2] C. Bardos, G. Lebeau, and J. Rauch. Sharp sufficient conditions for the observation, control and stabilization of waves from the boundary. *SIAM J. Control and Optimization*, 30(5):1024–1065, 1992.

[3] P. Bechouche and A. Jüngel. Inviscid limits of the complex Ginzburg-Landau equation. *Comm. Math. Phys.*, 214(1):201–226, 2000.

[4] N. Burq. Contrôle de l'équation des plaques en présence d'obstacles strictement convexes. *Mém. Soc. Math. France (N.S.)*, (55):126, 1993.

[5] N. Burq and P. Gérard. Condition nécessaire et suffisante pour la contrôlabilité exacte des ondes. *C. R. Acad. Sci. Paris Sér. I Math.*, 325(7):749–752, 1997.

[6] S. Cox and E. Zuazua. The rate at which energy decays in a damped string. *Comm. Partial Differential Equations*, 19(1-2):213–243, 1994.

[7] S. Cox and E. Zuazua. The rate at which energy decays in a string damped at one end. *Indiana Univ. Math. J.*, 44(2):545–573, 1995.

[8] S. Ervedoza, C. Zheng, and E. Zuazua. On the observability of time-discrete conservative linear systems. *J. Funct. Anal.*, 254(12):3037–3078, June 2008. *Cf Chapitre 3.*

[9] S. Ervedoza and E. Zuazua. Perfectly matched layers in 1-d: Energy decay for continuous and semi-discrete waves. *Numer. Math.*, 109(4):597–634, 2008. *Cf Chapitre 1.*

[10] S. Ervedoza and E. Zuazua. Uniformly exponentially stable approximations for a class of damped systems. *To appear in J. Math. Pures Appl.*, 2008. *Cf Chapitre 5.*

[11] A. Haraux. Une remarque sur la stabilisation de certains systèmes du deuxième ordre en temps. *Portugal. Math.*, 46(3):245–258, 1989.

[12] L. I. Ignat and E. Zuazua. Dispersive properties of a viscous numerical scheme for the Schrödinger equation. *C. R. Math. Acad. Sci. Paris*, 340(7):529–534, 2005.

[13] S. Jaffard. Contrôle interne exact des vibrations d'une plaque rectangulaire. *Portugal. Math.*, 47(4):423–429, 1990.

[14] G. Lebeau. Contrôle de l'équation de Schrödinger. *J. Math. Pures Appl. (9)*, 71(3):267–291, 1992.

[15] G. Lebeau. Équations des ondes amorties. *Séminaire sur les Équations aux Dérivées Partielles, 1993–1994,École Polytech.*, 1994.

[16] J.-L. Lions. *Contrôlabilité exacte, Stabilisation et Perturbations de Systèmes Distribués. Tome 1. Contrôlabilité exacte*, volume RMA 8. Masson, 1988.

[17] S. Machihara and Y. Nakamura. The inviscid limit for the complex Ginzburg-Landau equation. *J. Math. Anal. Appl.*, 281(2):552–564, 2003.

[18] A. Münch and A. F. Pazoto. Uniform stabilization of a viscous numerical approximation for a locally damped wave equation. *ESAIM Control Optim. Calc. Var.*, 13(2):265–293 (electronic), 2007.

[19] K. Ramdani, T. Takahashi, and M. Tucsnak. Uniformly exponentially stable approximations for a class of second order evolution equations—application to LQR problems. *ESAIM Control Optim. Calc. Var.*, 13(3):503–527, 2007.

[20] L. R. Tcheugoué Tebou and E. Zuazua. Uniform boundary stabilization of the finite difference space discretization of the $1 - d$ wave equation. *Adv. Comput. Math.*, 26(1-3):337–365, 2007.

[21] L.R. Tcheugoué Tébou and E. Zuazua. Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity. *Numer. Math.*, 95(3):563–598, 2003.

Chapter 5

Uniformly exponentially stable approximations for a class of damped systems

Joint work with Enrique Zuazua.

Abstract: We consider time semi-discrete approximations of a class of exponentially stable infinite dimensional systems modeling, for instance, damped vibrations. It has recently been proved that for time semi-discrete systems, due to high frequency spurious components, the exponential decay property may be lost as the time step tends to zero. We prove that adding a suitable numerical viscosity term in the numerical scheme, one obtains approximations that are uniformly exponentially stable. This result is then combined with previous ones on space semi-discretizations to derive similar results on fully-discrete approximation schemes. Our method is mainly based on a decoupling argument of low and high frequencies, the low frequency observability property for time semi-discrete approximations of conservative linear systems and the dissipativity of the numerical viscosity on the high frequency components. Our methods also allow to deal directly with stabilization properties of fully discrete approximation schemes without numerical viscosity, under a suitable CFL type condition on the time and space discretization parameters.

5.1 Introduction

Let X and Y be Hilbert spaces endowed with the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ respectively. Let $A : \mathcal{D}(A) \subset X \to X$ be a skew-adjoint operator with compact resolvent and $B \in \mathfrak{L}(X, Y)$.

We consider the system described by

$$\dot{z} = Az - B^*Bz, \quad t \geq 0, \qquad z(0) = z_0 \in X. \tag{5.1.1}$$

Here and henceforth, a dot $(\dot{})$ denotes differentiation with respect to time t. The element $z_0 \in X$ is the initial state, and $z(t)$ is the state of the system.

Most of the linear equations modeling the damped vibrations of elastic structures can be written

in the form (5.1.1). Some other relevant models, as the damped Schrödinger equations, fit in this setting as well.

We define the energy of the solutions of system (5.1.1) by

$$E(t) = \frac{1}{2} \|z(t)\|_X^2, \quad t \geq 0, \tag{5.1.2}$$

which satisfies

$$\frac{dE}{dt}(t) = - \|Bz(t)\|_Y^2, \quad t \geq 0. \tag{5.1.3}$$

In this paper, we assume that system (5.1.1) is exponentially stable, that is there exist positive constants μ and ν such that any solution of (5.1.1) satisfies

$$E(t) \leq \mu \; E(0) \exp(-\nu t), \quad t \geq 0. \tag{5.1.4}$$

Our goal is to develop a theory allowing to get, as a consequence of (5.1.4), exponential stability results for time-discrete systems.

We start considering the following natural time-discretization scheme for the continuous system (5.1.1). For any $\triangle t > 0$, we denote by z^k the approximation of the solution z of system (5.1.1) at time $t_k = k\triangle t$, for $k \in \mathbb{N}$, and introduce the following *implicit midpoint* time discretization of system (5.1.1):

$$\begin{cases} \dfrac{z^{k+1} - z^k}{\triangle t} = A\Big(\dfrac{z^k + z^{k+1}}{2}\Big) - B^*B\Big(\dfrac{z^k + z^{k+1}}{2}\Big), \quad k \in \mathbb{N}, \\ z^0 = z_0. \end{cases} \tag{5.1.5}$$

As in (5.1.2), we can define the discrete energy by

$$E^k = \frac{1}{2} \left\| z^k \right\|_X^2, \quad k \in \mathbb{N}, \tag{5.1.6}$$

which satisfies the dissipation law

$$\frac{E^{k+1} - E^k}{\triangle t} = - \left\| B\Big(\frac{z^k + z^{k+1}}{2}\Big) \right\|_Y^2, \quad k \in \mathbb{N}. \tag{5.1.7}$$

The results in [28], in the context of the conservative wave equation, which is a particular instance of (5.1.1) with $B = 0$, show that we cannot expect in general to find positive constants μ_0 and ν_0 such that

$$E^k \leq \mu_0 \; E^0 \exp(-\nu_0 k \triangle t), \quad k \in \mathbb{N}, \tag{5.1.8}$$

holds for any solution of (5.1.9) uniformly with respect to $\triangle t > 0$. Indeed, it was proved in [28] that spurious high-frequency modes may arise when discretizing in time the wave equation, which propagate with an arbitrarily small velocity and that, when the operator B is localized somewhere in the domain where waves propagate, cannot be observed uniformly with respect to $\triangle t$. This constitutes an obstruction to the stabilization property (5.1.8) as well.

Therefore, in order to get a uniform decay, it seems natural to add in system (5.1.5) a suitable extra numerical viscosity term to damp these high-frequency spurious components. When doing it at the right scale, the new system we obtain is as follows:

$$\begin{cases} \dfrac{\tilde{z}^{k+1} - z^k}{\triangle t} &= A\Big(\dfrac{z^k + \tilde{z}^{k+1}}{2}\Big) - B^*B\Big(\dfrac{z^k + \tilde{z}^{k+1}}{2}\Big), \quad k \in \mathbb{N}, \\ \dfrac{z^{k+1} - \tilde{z}^{k+1}}{\triangle t} &= (\triangle t)^2 A^2 z^{k+1}, \quad k \in \mathbb{N}, \\ z^0 &= z_0. \end{cases} \tag{5.1.9}$$

This system introduces, indeed, numerical viscosity at the right scale since the spurious high-frequency modes arising in [28] precisely correspond to solutions for which $(\triangle t)A$ is of unit order or more.

Let us also remark that system (5.1.9) can be rewritten as

$$\frac{z^{k+1}-z^k}{\triangle t} = A\Big(\frac{z^k+z^{k+1}}{2}\Big) - B^*B\Big(\frac{z^k+z^{k+1}}{2}\Big) + (\triangle t)^2A^2z^{k+1} - \frac{(\triangle t)^3}{2}A^3z^{k+1} + \frac{(\triangle t)^3}{2}B^*BA^2z^{k+1}, \quad (5.1.10)$$

which is consistent with system (5.1.1).

To motivate system (5.1.9), one can compare it with the time continuous system

$$\dot{z} = Az - B^*Bz + (\triangle t)^2A^2z, \quad (5.1.11)$$

which generates the semigroup $S(t) = \exp(t(A - B^*B + (\triangle t)^2A^2))$. In (5.1.9), $\tilde{z}^{k+1}$ corresponds to an approximation of $\exp(\triangle t(A - B^*B))z^k$ and z^{k+1} to an approximation of $\exp((\triangle t)^3A^2)\tilde{z}^{k+1}$. Doing this, z^{k+1} is an approximation of $S(\triangle t)z^k \simeq \exp((\triangle t)^3A^2)\exp(\triangle t(A-B^*B))z^k$. Thus, system (5.1.9) can be viewed as an alternating direction time-discrete approximation of (5.1.11), for which dissipation properties have been derived in the recent article [14].

Note that this numerical scheme is based on the decomposition of the operator $A-B^*B+(\triangle t)^2A^2$ into its conservative and dissipative parts, that we treat differently. Indeed, the midpoint scheme is appropriate for conservative systems since it preserves the norm conservation property. This is not the case for dissipative systems, since midpoint schemes do not preserve the dissipative properties of high frequency solutions. Therefore, we rather use an implicit Euler scheme, which efficiently preserves these dissipative properties.

In Subsection 5.2.3, we will consider other possible discretization schemes, variants of (5.1.9), which still preserve the conservative properties of $\exp(tA)$ and the dissipative effects of $\exp(t(\triangle t)^2A^2)$. We will also present other possible choices for the numerical viscosity term.

The energy of (5.1.9), still defined by (5.1.6), now satisfies

$$\begin{cases} \tilde{E}^{k+1} = E^k - \triangle t\left\|B\Big(\frac{z^k+\tilde{z}^{k+1}}{2}\Big)\right\|_Y^2, \quad k\in\mathbb{N}, \\ E^{k+1} + (\triangle t)^3\left\|Az^{k+1}\right\|_X^2 + \frac{(\triangle t)^6}{2}\left\|A^2z^{k+1}\right\|_X^2 = \tilde{E}^{k+1}, \quad k\in\mathbb{N}. \end{cases} \quad (5.1.12)$$

Putting these identities together, we get

$$E^{k+1} + (\triangle t)^3\left\|Az^{k+1}\right\|_X^2 + \frac{(\triangle t)^6}{2}\left\|A^2z^{k+1}\right\|_X^2 + \triangle t\left\|B\Big(\frac{z^k+\tilde{z}^{k+1}}{2}\Big)\right\|_Y^2 = E^k. \quad (5.1.13)$$

The convergence of the solutions of (5.1.9) towards those of the original system (5.1.1) when $\triangle t\to 0$ holds in a suitable topology. Indeed, the scheme is stable in view of (5.1.12), and its consistency is obvious. Therefore its convergence (in the classical sense of numerical analysis) is guaranteed: When $\triangle t\to 0$, the solutions $z_{\triangle t}$ of (5.1.9), extended in a standard way as piecewise affine functions on $\mathbb{R}_+$, converge to the solution z of (5.1.1) in $L^2((0,T);X)$.

The main result of this paper is that system (5.1.9) enjoys a uniform stabilization property. It reads as follows:

Theorem 5.1.1. *Assume that system* (5.1.1) *is exponentially stable, i.e. satisfies* (5.1.4) *with constants* μ *and* ν*, and that* $B \in \mathfrak{L}(X,Y)$*.*

Then there exist two positive constants μ_0 *and* ν_0 *depending only on* μ*,* ν *and* $\|B\|_{\mathfrak{L}(X,Y)}$ *such that any solution of* (5.1.9) *satisfies* (5.1.8) *with constants* μ_0 *and* ν_0 *uniformly with respect to the discretization parameter* $\triangle t > 0$*.*

Our strategy is based on the fact that the uniform exponential decay properties of the energy for systems (5.1.1) and (5.1.9) respectively are equivalent to uniform observability properties for the conservative system

$$\dot{y} = Ay, \quad t \in \mathbb{R}, \qquad y(0) = y_0 \in X, \tag{5.1.14}$$

and its time semi-discrete viscous version

$$\begin{cases} \dfrac{\tilde{u}^{k+1} - u^k}{\triangle t} = A\Big(\dfrac{u^k + \tilde{u}^{k+1}}{2}\Big), \quad k \in \mathbb{N}, \\ \dfrac{u^{k+1} - \tilde{u}^{k+1}}{\triangle t} = (\triangle t)^2 A^2 u^{k+1}, \quad k \in \mathbb{N}, \\ u^0 = u_0, \end{cases} \tag{5.1.15}$$

At the continuous level the observability property consists in the existence of a time $T > 0$ and a positive constant $k_T > 0$ such that

$$k_T \|y_0\|_X^2 \leq \int_0^T \|By(t)\|_Y^2 \, dt, \tag{5.1.16}$$

for every solution of (5.1.14) (see [16] and Lemma 5.2.3 below).

A similar argument can be applied to the semi-discrete system (5.1.9). Namely, the uniform exponential decay (5.1.8) of the energy of solutions of (5.1.9) is equivalent to the following observability inequality: there exist positive constants T and c such that, for any $\triangle t > 0$, every solution u of (5.1.15) satisfies

$$c \|u_0\|_X^2 \leq \triangle t \sum_{k\triangle t \in [0,T]} \left\|Bu^k\right\|_Y^2 + \triangle t \sum_{k\triangle t \in [0,T]} (\triangle t)^2 \left\|Au^{k+1}\right\|_X^2 + \triangle t \sum_{k\triangle t \in [0,T]} (\triangle t)^5 \left\|A^2 u^{k+1}\right\|_X^2. \tag{5.1.17}$$

Note that, since the operator $(\triangle t)^2 A^2$ is unbounded, we cannot use the standard arguments in [16], which state the equivalence between the uniform exponential decay of the energy for (5.1.9) and uniform observability properties such as (5.1.17) for solutions of the conservative system

$$\frac{y^{k+1} - y^k}{\triangle t} = A\Big(\frac{y^k + y^{k+1}}{2}\Big), \quad k \in \mathbb{N}, \qquad y^0 = y_0, \tag{5.1.18}$$

or, equivalently,

$$\frac{\tilde{y}^{k+1} - y^k}{\triangle t} = A\Big(\frac{y^k + \tilde{y}^{k+1}}{2}\Big), \quad y^{k+1} = \tilde{y}^{k+1} \quad k \in \mathbb{N}, \qquad y^0 = y_0. \tag{5.1.19}$$

Let us now give some insights of the proof of (5.1.17) for solutions of (5.1.15). The main idea is to decompose the solution u of (5.1.15) into its low and high frequency parts, that we handle separately. We first use a uniform observability inequality proven in [12] for solutions of (5.1.18) in a filtered space, which yields a partial observability inequality for the low frequency components of solutions of (5.1.15). Second, using the explicit dissipativity of (5.1.15) at high frequencies, we deduce a partial observability inequality for the high frequency components. Together, these two partial observability inequalities yield the needed observability property (5.1.17) leading to the uniform exponential decay result.

Our results yield also uniform exponential decay rates for families of equations of the form (5.1.1), with pairs of operators (A, B), within a class in which the exponential decay rate of the continuous system (5.1.1) is known to be uniform.

One of the interesting applications of this fact is that our results can be combined with the existing ones derived for *space* semi-discrete approximation schemes of various PDE models entering in the abstract frame (5.1.1) as [5, 6, 13, 11, 24, 27, 23] (see [32] for more references). Indeed, knowing that some *space* semi-discrete approximation schemes of (5.1.1) are exponentially stable, uniformly with respect to the space mesh size, this fact, combined with Theorem 5.1.1, allows deducing uniform exponential decay properties for the corresponding *fully* discrete approximation schemes.

Our methods can also be applied directly to fully discrete approximation schemes under a suitable CFL type condition on the time and space discretization parameters. This can be done without adding a numerical viscosity term since the CFL condition by itself rules out the high frequency components. As we will see in the examples, this CFL condition might be very strong and yield severe restrictions, which do not appear when adding numerical viscosity as in (5.1.9) (see Theorem 5.1.1).

As said above, these approaches require observability properties such as (5.1.16) to hold uniformly (with respect to the space discretization parameter) for solutions of the space semi-discrete schemes for *any* initial data. However, it often occurs in applications that the space semi-discrete schemes are uniformly observable only for *filtered* initial data corresponding to low frequencies (see [18, 31, 13, 32]). We therefore adapt our methods to this case, and prove that adding a numerical viscosity term which is strong enough to efficiently damp out the high frequency components, one obtains uniformly exponentially stable fully discrete approximation schemes. When doing this, we also prove that, when considering space semi-discrete approximation schemes that are uniformly observable in filtered low-frequency subspaces, adding a suitable numerical viscosity term makes the space semi-discrete approximation schemes uniformly (with respect to the space discretization parameter) exponentially stable. This generalizes the results [27, 25, 13], where particular instances of viscosity terms have been used. This also generalizes [14], where it was proven that if (5.1.1) is exponentially stable, then adding a suitable viscosity term does not deteriorate the exponential stability of solutions.

In this sense, the approaches presented in this article are complementary.

Note however that we cannot apply these methods when the damped operator B is not bounded, as in [26], where the wave equation is damped by a feedback law on the boundary. Dealing with unbounded damping operators B needs further work.

The results in this paper on the uniform stabilization of time-discrete approximation schemes with numerical viscosity term are related to several previous ones. The following ones are worth mentioning. In [27, 26, 23, 13] numerical viscosity is added to guarantee the uniform exponential decay for finite-difference space semi-discrete approximation schemes of the wave equation. Similar results, in an

abstract setting, with a stronger viscous damping term, have been proved in [25]. Similar techniques have also been employed to obtain uniform dispersive estimates for numerical approximation schemes to Schrödinger equations in [17].

Let us also mention the recent work [12], where observability issues were discussed for time and fully discrete approximation schemes of (5.1.18). The results of [12] will be used in the present work to derive observability properties for system (5.1.18) within the class of conveniently filtered low frequency data. Since they constitute a key point of our proofs, we recall them in Section 5.2.

Despite all the existing literature, this article seems to be the first one to provide a systematic way of transferring exponential decay properties from the continuous to the time-discrete setting.

The outline of this paper is as follows.
In Section 5.2, we recall the results of [12] and prove Theorem 5.1.1. Section 5.3 is devoted to explain how we can deduce uniform stabilization results for the fully discrete approximation schemes combining Theorem 5.1.1 and known results on uniform stabilization for space semi-discrete approximations. We also present an abstract setting specifically designed to address stabilization issues for fully discrete approximation schemes without viscosity. In Section 5.4, we present some concrete applications in the context of the wave equation for which several uniformly exponentially stable schemes are derived. Finally, some further comments and open problems are collected in Section 5.5.

5.2 Stabilization of time-discrete systems

This section is organized as follows. We first recall the results of [12] on the observability of the time-discrete conservative system (5.1.18). Second, we prove Theorem 5.1.1. Third, we present several variants of the numerical scheme (5.1.9) that lead to uniform exponential decay results similar to Theorem 5.1.1.

5.2.1 Observability of time-discrete conservative systems

We first need to introduce some notations.

Since A is a skew-adjoint operator with compact resolvent, its spectrum is discrete and $\sigma(A) = \{i\mu_j \ : \ j \in \mathbb{N}\}$, where $(\mu_j)_{j\in\mathbb{N}}$ is a sequence of real numbers such that $|\mu_j| \to \infty$ when $j \to \infty$. Set $(\Phi_j)_{j\in\mathbb{N}}$ an orthonormal basis of eigenvectors of A associated to the eigenvalues $(i\mu_j)_{j\in\mathbb{N}}$, that is

$$A\Phi_j = i\mu_j\Phi_j. \tag{5.2.1}$$

Moreover, define

$$\mathcal{C}_s(A) = \text{span}\,\{\Phi_j : \text{ the corresponding } i\mu_j \text{ satisfies } |\mu_j| \le s\}. \tag{5.2.2}$$

The following was proved in [12]:

Theorem 5.2.1. *Assume that $B \in \mathfrak{L}(\mathcal{D}(A), Y)$, that is*

$$\|Bz\|_Y^2 \le C_B^2\Big(\|Az\|_X^2 + \|z\|_X^2 \Big), \qquad \forall z \in \mathcal{D}(A), \tag{5.2.3}$$

and that A and B satisfy the following hypothesis:

$$\begin{cases} \text{There exist constants } M, m > 0 \text{ such that} \\ M^2 \|(i\omega I - A)y\|_X^2 + m^2 \|By\|_Y^2 \geq \|y\|_X^2, \quad \forall\ \omega \in \mathbb{R},\ \ y \in \mathcal{D}(A). \end{cases} \tag{5.2.4}$$

Then, for any $\delta > 0$, there exists T_δ such that for any $T > T_\delta$, there exists a positive constant $k_{T,\delta}$, independent of $\triangle t$, that depends only on m, M, C_B, T and δ, such that for $\triangle t > 0$ small enough, the solution y^k of (5.1.18) satisfies

$$k_{T,\delta} \left\|y^0\right\|_X^2 \leq \triangle t \sum_{k\triangle t \in [0,T]} \left\| B\Big(\frac{y^k + y^{k+1}}{2}\Big)\right\|_Y^2, \qquad \forall\ y^0 \in \mathcal{C}_{\delta/\triangle t}(A). \tag{5.2.5}$$

Moreover, T_δ can be taken to be such that

$$T_\delta = \pi\Big[M^2\Big(1 + \frac{\delta^2}{4}\Big)^2 + m^2 C_B^2 \frac{\delta^4}{16}\Big]^{1/2}, \tag{5.2.6}$$

where C_B is as in (5.2.3).

In the sequel, when there is no ambiguity, we will use the simplified notation $\mathcal{C}_{\delta/\triangle t}$ instead of $\mathcal{C}_{\delta/\triangle t}(A)$.

Note that if $B \in \mathfrak{L}(X, Y)$, then the operator B is also in $\mathfrak{L}(\mathcal{D}(A), Y)$, and (5.2.3) holds. Thus the assumption (5.2.3) is satisfied in the abstract setting we are working on.

Hypothesis (5.2.4) is the so-called resolvent estimate, which has been proved in [4, 22] to be equivalent to the continuous observability inequality (5.1.16) for the conservative system (5.1.14) for suitable positive constants T and k_T, which turns out to be equivalent to the exponential decay property (5.1.4) for the continuous damped system (5.1.1).

To be more precise, it was proved in [22] that if the operator B is bounded, then the observability property (5.1.16) implies hypothesis (5.2.4) with

$$m = \sqrt{\frac{2T}{k_T}}, \quad M = T\|B\|_{\mathfrak{L}(X,Y)} \sqrt{\frac{T}{2k_T}}, \tag{5.2.7}$$

where k_T is as in (5.1.16).

Observe that Theorem 5.2.1 guarantees that, as soon as the observability inequality (5.1.16) holds for the continuous system (5.1.14), then its time-discrete counterpart holds uniformly for the solutions of the time discrete systems (5.1.18) within the class of filtered solutions $\mathcal{C}_{\delta/\triangle t}(A)$ involving only the low-frequency components corresponding to the eigenvalues $|\mu_i| \leq \delta/\triangle t$. This fact will play a key role in the proof of Theorem 5.1.1.

5.2.2 Proof of Theorem 5.1.1

In this Subsection, we assume that system (5.1.1) is exponentially stable and that $B \in \mathfrak{L}(X, Y)$, i.e. there exists a constant K_B such that

$$\|Bz\|_Y \leq K_B \|z\|_X, \qquad \forall z \in X. \tag{5.2.8}$$

The proof is divided into several steps. First, we write carefully the energy identity for z solution of (5.1.9). Second, we observe that the resolvent estimate (5.2.4) holds, from which we deduce that (5.2.5) holds as well for solutions of system (5.1.18) in the filtered space $\mathcal{C}_{\delta/\triangle t}$. Third, we derive the observability inequality (5.1.17) for solutions of (5.1.15). Finally, we deduce that the time-discrete systems (5.1.9) are uniformly exponentially stable.

The energy identity

Lemma 5.2.2. *For any $\triangle t > 0$ and $z^0 \in X$, the solution z of* (5.1.9) *satisfies*

$$\left\|z^{k_2}\right\|_X^2 + 2\triangle t \sum_{j=k_1}^{k_2-1} \left\|B\Big(\frac{z^j + \tilde{z}^{j+1}}{2}\Big)\right\|_Y^2 + 2\triangle t \sum_{j=k_1}^{k_2-1} (\triangle t)^2 \left\|Az^{j+1}\right\|_X^2 + \triangle t \sum_{j=k_1}^{k_2-1} (\triangle t)^5 \left\|A^2 z^{j+1}\right\|_X^2 = \left\|z^{k_1}\right\|_X^2, \quad \forall k_1 < k_2. \quad (5.2.9)$$

The proof simply consists in summing the identities in (5.1.13) from $k = l_1$ to $k = l_2 - 1$. Especially, it implies that $\left\|z^k\right\|_X^2$ is decreasing, which confirms the dissipativity of the time-discrete system.

The resolvent estimate

Lemma 5.2.3. *Under the assumptions of Theorem 5.1.1, the resolvent estimate* (5.2.4) *holds, with constants m and M that depend only on μ and ν given by* (5.1.4).

Proof. The proof is based on [16].

Since system (5.1.1) is exponentially stable, inequality (5.1.4) holds. In particular, there exists a positive constant $T > 0$ such that $2E(T) \leq E(0)$. But equality (5.1.3) implies that any solution z of (5.1.1) satisfies

$$E(T) + \int_0^T \|Bz(t)\|_Y^2 \ dt = E(0),$$

and therefore that

$$\int_0^T \|Bz(t)\|_Y^2 \ dt \geq \frac{1}{4} \|z_0\|_X^2 . \quad (5.2.10)$$

Let us now show that, as a consequence of this, (5.1.16) holds for the solution of (5.1.14) as well.

Given $y_0 \in X$, let y and z be the solutions of (5.1.14) and (5.1.1) with initial data y_0. Then $w = z - y$ satisfies

$$\dot{w} = Aw - B^*Bw - B^*By, \quad t \in \mathbb{R}, \qquad w(0) = 0.$$

Multiplying by w and integrating in time, we obtain that

$$\begin{aligned} \frac{1}{2} \|w(T)\|_X^2 + \int_0^T \|Bw(t)\|_Y^2 \ dt &\leq \int_0^T | < Bw(t), By(t) >_Y | \ dt \\ &\leq \frac{1}{2} \int_0^T \Big(\|Bw(t)\|_Y^2 + \|By(t)\|_Y^2 \Big) \ dt. \end{aligned}$$

In particular,

$$\int_0^T \|Bw(t)\|_Y^2 \ dt \leq \int_0^T \|By(t)\|_Y^2 \ dt.$$

This inequality, combined with (5.2.10), leads to

$$\frac{1}{4}\|y_0\|_X^2 \leq \int_0^T \|Bz(t)\|_Y^2 \ dt \leq 2\int_0^T \Big(\|Bw(t)\|_Y^2 + \|By(t)\|_Y^2\Big) \ dt \leq 3\int_0^T \|By(t)\|_Y^2 \ dt.$$

It follows that (5.1.16) holds, and the resolvent estimate (5.2.4) holds with m and M as in (5.2.7), according to the results in [22]. □

Applying Theorem 5.2.1, for any $\delta > 0$, choosing a time $T^* > T_\delta$ (where T_δ is defined in (5.2.6)) there exists a positive constant $k_{T^*,\delta}$ such that inequality (5.2.5) holds for any solution y of (5.1.18) with $y^0 \in \mathcal{C}_{\delta/\triangle t}$. In the sequel, we fix a positive number $\delta > 0$ (for instance $\delta = 1$), and $T^* = 2T_\delta$.

Uniform observability inequalities

Lemma 5.2.4. *There exists a constant $c > 0$ such that* (5.1.17) *holds with $T = T^*$ for all solutions u of* (5.1.15) *uniformly with respect to $\triangle t$.*

Proof. In the sequel we deal with the solutions u of (5.1.15), for which we prove (5.1.17) for $T = T^* = 2T_\delta$. The proof presented below is inspired in previous work [14] from the authors, where similar arguments have been used in the continuous setting.

As said in the introduction, we decompose the solution u of (5.1.15) into its low and high frequency parts. To be more precise, we consider

$$u_l = \pi_{\delta/\triangle t} u, \quad u_h = (I - \pi_{\delta/\triangle t})u, \tag{5.2.11}$$

where $\delta > 0$ is the positive number that have been chosen above, and $\pi_{\delta/\triangle t}$ is the orthogonal projection on $\mathcal{C}_{\delta/\triangle t}$ defined in (5.2.2). Here the notations u_l and u_h stand for the low and high frequency components, respectively.

Note that both u_l and u_h are solutions of (5.1.15).

Besides, u_h lies in the space $\mathcal{C}^\perp_{\delta/\triangle t}$, in which the following property holds:

$$\triangle t \|Ay\|_X \geq \delta \|y\|_X, \quad \forall y \in \mathcal{C}^\perp_\delta. \tag{5.2.12}$$

The low frequencies. In a first step, we compare u_l with y_l solution of (5.1.18) with initial data $y_l(0) = u_l(0)$. Now, set $w_l = u_l - y_l$. From (5.2.5), which is valid for solutions of (5.1.18) with initial data in $\mathcal{C}_{\delta/\triangle t}$, we get

$$k_{T^*,\delta}\left\|u_l^0\right\|_X^2 = k_{T^*,\delta}\left\|y_l^0\right\|_X^2 \leq 2\triangle t \sum_{k\triangle t\in[0,T^*]} \left\|B\Big(\frac{u_l^k + \tilde{u}_l^{k+1}}{2}\Big)\right\|_Y^2 + 2\triangle t \sum_{k\triangle t\in[0,T^*]} \left\|B\Big(\frac{w_l^k + \tilde{w}_l^{k+1}}{2}\Big)\right\|_Y^2. \tag{5.2.13}$$

In the sequel, to simplify the notation, $c > 0$ will denote a positive constant that may change from line to line, but which does not depend on $\triangle t$.

Let us then estimate the last term in the right hand side of (5.2.13). To this end, we write the equation satisfied by w_l, which can be deduced from (5.1.18) and (5.1.15):

$$\left\{\begin{array}{l} \dfrac{\tilde{w}_l^{k+1}-w_l^k}{\triangle t} = A\Big(\dfrac{w_l^k+\tilde{w}_l^{k+1}}{2}\Big), k\in\mathbb{N}, \\ \dfrac{w_l^{k+1}-\tilde{w}_l^{k+1}}{\triangle t} = (\triangle t)^2 A^2 u_l^{k+1}, \quad k\in\mathbb{N}, \\ w_l^0 = 0. \end{array}\right. \tag{5.2.14}$$

The energy estimates for w_l give

$$\left\{\begin{array}{l} \Big\|\tilde{w}_l^{k+1}\Big\|_X^2 = \Big\|w_l^k\Big\|_X^2, \\ \Big\|w_l^{k+1}\Big\|_X^2 = \Big\|\tilde{w}_l^{k+1}\Big\|_X^2 - 2(\triangle t)^3 < Au_l^{k+1}, A\Big(\dfrac{\tilde{w}_l^{k+1}+w_l^{k+1}}{2}\Big) >_X . \end{array}\right. \tag{5.2.15}$$

Note that w_l^k and $\tilde{w}_l^{k+1}$ belong to $\mathcal{C}_{\delta/\triangle t}$ for all $k\in\mathbb{N}$, since u_l and y_l both belong to $\mathcal{C}_{\delta/\triangle t}$. Therefore, the energy estimates for w_l lead, for $k\in\mathbb{N}$, to

$$\begin{aligned} \Big\|w_l^k\Big\|_X^2 &= -2\triangle t\sum_{j=1}^k (\triangle t)^2 < Au_l^j, A\Big(\frac{w_l^j+\tilde{w}_l^{j+1}}{2}\Big) >_X \\ &\le \triangle t\sum_{j=1}^k (\triangle t)^2 \Big\|Au_l^j\Big\|_X^2 + \delta^2\triangle t\sum_{j=1}^k \Big\|\frac{w_l^j+\tilde{w}_l^{j+1}}{2}\Big\|_X^2 \\ &\le \triangle t\sum_{j=1}^k (\triangle t)^2 \Big\|Au_l^j\Big\|_X^2 + \delta^2\triangle t\sum_{j=0}^k \Big\|w_l^j\Big\|_X^2, \end{aligned}$$

where we used the first line of (5.2.15).

Grönwall's Lemma applies and allows to deduce from (5.2.13) and the fact that the operator B is bounded, the existence of a positive c independent of $\triangle t$, such that

$$c\left\|u_l^0\right\|_X^2 \le \triangle t\sum_{k\triangle t\in[0,T^*]} \Big\|B\Big(\frac{u_l^k+\tilde{u}_l^{k+1}}{2}\Big)\Big\|_Y^2 + \triangle t\sum_{k\triangle t\in]0,T^*]} (\triangle t)^2\Big\|Au_l^k\Big\|_X^2 .$$

Besides,

$$\begin{aligned} \triangle t\sum_{k\triangle t\in[0,T^*]} \Big\|B\Big(\frac{u_l^k+\tilde{u}_l^{k+1}}{2}\Big)\Big\|_Y^2 \le 2\triangle t\sum_{k\triangle t\in[0,T^*]} \Big\|B\Big(\frac{u^k+\tilde{u}^{k+1}}{2}\Big)\Big\|_Y^2 \\ + 2\triangle t\sum_{k\triangle t\in[0,T^*]} \Big\|B\Big(\frac{u_h^k+\tilde{u}_h^{k+1}}{2}\Big)\Big\|_Y^2 \end{aligned}$$

and, since u_h^k and $\tilde{u}_h^{k+1}$ belong to $\mathcal{C}_{\delta/\triangle t}^{\perp}$ for all k, we get from (5.2.12) that

$$\triangle t \sum_{k\triangle t\in[0,T^*]} \left\| B\Big(\frac{u_h^k+\tilde{u}_h^{k+1}}{2}\Big)\right\|^2 \leq K_B^2 \triangle t \sum_{k\triangle t\in[0,T^*]} \left\|\frac{u_h^k+\tilde{u}_h^{k+1}}{2}\right\|_X^2$$
$$\leq K_B^2\triangle t \sum_{k\triangle t\in]0,T^*]} \left\|u_h^k\right\|_X^2 \leq \frac{K_B^2}{\delta^2}\triangle t \sum_{k\triangle t\in]0,T^*]} (\triangle t)^2 \left\|Au_h^k\right\|_X^2 + K_B^2\triangle t \left\|u_h^0\right\|_X^2,$$

since, from the first line of (5.1.15),

$$\left\|\tilde{u}_h^{k+1}\right\|_X^2 = \left\|u_h^k\right\|_X^2, \quad \forall k\in\mathbb{N}.$$

It follows that there exists $c>0$ independent of $\triangle t$ such that

$$c\left\|u_l^0\right\|_X^2 \leq \triangle t \sum_{k\triangle t\in[0,T^*]} \left\| B\Big(\frac{u^k+\tilde{u}^{k+1}}{2}\Big)\right\|_Y^2 + \triangle t \sum_{k\triangle t\in]0,T^*]} (\triangle t)^2 \left\|Au_l^k\right\|_X^2 + \triangle t\left\|u_h^0\right\|_X^2. \tag{5.2.16}$$

The high frequencies. We now discuss briefly the decay properties of solutions u_h of (5.1.15) with initial data $u_h^0\in\mathcal{C}_{\delta/\triangle t}^{\perp}$. In this case, we easily check that for all $k\in\mathbb{N}$, $u_h^k\in\mathcal{C}_{\delta/\triangle t}^{\perp}$. But, as in (5.1.13), we have

$$\left\|(I-(\triangle t)^3A^2)u_h^{k+1}\right\|_X^2 = \left\|u_h^{k+1}\right\|_X^2 + 2(\triangle t)^3\left\|Au_h^{k+1}\right\|_X^2$$
$$+ (\triangle t)^6\left\|A^2u_h^{k+1}\right\|_X^2 = \left\|\tilde{u}_h^{k+1}\right\|_X^2 = \left\|u_h^k\right\|_X^2, \quad k\in\mathbb{N}. \tag{5.2.17}$$

Due to the property (5.2.12), we get

$$(1+2(\triangle t)\delta^2)\left\|u_h^{k+1}\right\|_X^2 \leq \left\|u_h^k\right\|_X^2.$$

We deduce that

$$\left\|u_h^{k+1}\right\|_X^2 \leq \frac{1}{1+2(\triangle t)\delta^2}\left\|u_h^k\right\|_X^2, \quad k\in\mathbb{N},$$

which implies

$$\left\|u_h^k\right\|_X^2 \leq \Big(\frac{1}{1+2(\triangle t)\delta^2}\Big)^k\left\|u_h^0\right\|_X^2, \quad k\in\mathbb{N}. \tag{5.2.18}$$

Especially, taking $k^*=\lceil T^*/\triangle t\rceil$, we get a constant $\gamma<1$ independent of $\triangle t>0$ such that

$$\left\|u_h^{k^*}\right\|_X^2 \leq \gamma\left\|u_h^0\right\|_X^2.$$

Since we also have from (5.2.17) that, for $k\in\mathbb{N}$,

$$\left\|u_h^k\right\|_X^2 + 2\triangle t\sum_{j=0}^{k-1}(\triangle t)^2\left\|Au_h^{j+1}\right\|_X^2 + \triangle t\sum_{j=0}^{k-1}(\triangle t)^5\left\|A^2u_h^{j+1}\right\|_X^2 = \left\|u_h^0\right\|_X^2,$$

taking $k=k^*=\lceil T^*/\triangle t\rceil$, we deduce the existence of a positive constant C, which depends only on T^* and δ (namely $C=(1-\gamma)/2$), such that

$$C\left\|u_h^0\right\|_X^2 \leq \triangle t\sum_{j=0}^{k^*-1}(\triangle t)^2\left\|Au_h^{j+1}\right\|_X^2 + \triangle t\sum_{j=0}^{k^*-1}(\triangle t)^5\left\|A^2u_h^{j+1}\right\|_X^2, \tag{5.2.19}$$

holds uniformly with respect to $\triangle t > 0$ for any solution of (5.1.15) with initial data $u^0 \in \mathcal{C}^{\perp}_{\delta/\triangle t}$.

Combining (5.2.16) and (5.2.19) yields Lemma 5.2.4, since u_h and u_l lie in orthogonal spaces with respect to the scalar products $< \cdot, \cdot >_X$ and $< A\cdot, A\cdot >_X$. □

Proof of Theorem 5.1.1

Proof of Theorem 5.1.1. Here we follow the argument in [16, 14].

We decompose z solution of (5.1.9) as $z = u + w$ where u is the solution of the system (5.1.15) with initial data $u^0 = z^0$. Applying Lemma 5.2.4 to $u = z - w$, we get

$$
\begin{aligned}
c\left\|z^0\right\|_X^2 \leq 2\Big(\triangle t \sum_{k\triangle t\in[0,T^*]} \Big\|B\Big(\frac{z^k+\tilde z^{k+1}}{2}\Big)\Big\|_Y^2 + \triangle t \sum_{k\triangle t\in[0,T^*[} (\triangle t)^2 \left\|Az^{k+1}\right\|_X^2 \\
+ \triangle t \sum_{k\triangle t\in[0,T^*[} (\triangle t)^5 \left\|A^2 z^{k+1}\right\|_X^2\Big) + 2\Big(\triangle t \sum_{k\triangle t\in[0,T^*]} \Big\|B\Big(\frac{w^k+\tilde w^{k+1}}{2}\Big)\Big\|_Y^2 \\
+ \triangle t \sum_{k\triangle t\in[0,T^*[} (\triangle t)^2 \left\|Aw^{k+1}\right\|_X^2 + \triangle t \sum_{k\triangle t\in[0,T^*[} (\triangle t)^5 \left\|A^2 w^{k+1}\right\|_X^2\Big). \quad (5.2.20)
\end{aligned}
$$

Below, we bound the terms in the right hand-side of (5.2.20) involving w by the ones involving z.

The function w satisfies

$$
\begin{cases}
\dfrac{\tilde w^{k+1}-w^k}{\triangle t} = A\Big(\dfrac{w^k+\tilde w^{k+1}}{2}\Big) - B^*B\Big(\dfrac{z^k+\tilde z^{k+1}}{2}\Big), \quad k\in\mathbb{N}, \\
\dfrac{w^{k+1}-\tilde w^{k+1}}{\triangle t} = (\triangle t)^2 A^2 w^{k+1}, \quad k\in\mathbb{N}, \\
w^0 = 0.
\end{cases} \quad (5.2.21)
$$

Multiplying the first line of (5.2.21) by $w^k + \tilde w^{k+1}$ and taking the norm of each member in the second one, we get the following energy identities for $k \in \mathbb{N}$:

$$
\begin{aligned}
\left\|\tilde w^{k+1}\right\|_X^2 &= \left\|w^k\right\|_X^2 - 2\triangle t < B\Big(\frac{z^k+\tilde z^{k+1}}{2}\Big), B\Big(\frac{w^k+\tilde w^{k+1}}{2}\Big) >_Y, \\
\left\|w^{k+1}\right\|_X^2 + 2(\triangle t)^3 \left\|Aw^{k+1}\right\|_X^2 + (\triangle t)^6 \left\|A^2 w^{k+1}\right\|_X^2 &= \left\|\tilde w^{k+1}\right\|_X^2.
\end{aligned} \quad (5.2.22)
$$

In particular, this gives

$$
\begin{aligned}
\left\|w^{k+1}\right\|_X^2 + 2(\triangle t)^3 \left\|Aw^{k+1}\right\|_X^2 + (\triangle t)^6 \left\|A^2 w^{k+1}\right\|_X^2 \\
+ 2\triangle t < B\Big(\frac{z^k+\tilde z^{k+1}}{2}\Big), B\Big(\frac{w^k+\tilde w^{k+1}}{2}\Big) >_Y = \left\|w^k\right\|_X^2.
\end{aligned}
$$

Using that B is bounded, we get

$$
\begin{aligned}
\left\|w^k\right\|_X^2 + 2(\triangle t)\sum_{j=0}^{k-1}(\triangle t)^2 \left\|Aw^{j+1}\right\|_X^2 + (\triangle t)\sum_{j=0}^{k-1}(\triangle t)^5 \left\|A^2 w^{j+1}\right\|_X^2 \\
\leq \triangle t \sum_{j=0}^{k-1} \Big\|B\Big(\frac{z^j+\tilde z^{j+1}}{2}\Big)\Big\|_Y^2 + \frac{K_B^2}{2}(\triangle t)\sum_{j=0}^{k-1}\Big(\left\|w^j\right\|_X^2 + \left\|\tilde w^{j+1}\right\|_X^2\Big). \quad (5.2.23)
\end{aligned}
$$

But the second line in (5.2.22) gives that

$$\triangle t\sum_{j=0}^{k-1}\left\|\tilde{w}^{j+1}\right\|_X^2=\triangle t\sum_{j=0}^{k-1}\left\|w^{j+1}\right\|_X^2+2(\triangle t)^2\sum_{j=0}^{k-1}(\triangle t)^2\left\|Aw^{j+1}\right\|_X^2 + (\triangle t)^2\sum_{j=0}^{k-1}(\triangle t)^5\left\|A^2w^{j+1}\right\|_X^2. \quad (5.2.24)$$

Therefore, for $\triangle t$ small enough, (5.2.23) gives

$$\left\|w^k\right\|_X^2+\triangle t\sum_{j=0}^{k-1}(\triangle t)^2\left\|Aw^{j+1}\right\|_X^2+\frac{\triangle t}{2}\sum_{j=0}^{k-1}(\triangle t)^5\left\|A^2w^{j+1}\right\|_X^2 \leq \triangle t\sum_{j=0}^{k-1}\left\|B\Big(\frac{z^j+\tilde{z}^{j+1}}{2}\Big)\right\|_Y^2+K_B^2\triangle t\sum_{j=0}^{k-1}\left\|w^j\right\|_X^2. \quad (5.2.25)$$

Grönwall's inequality then gives a constant G, that depends only on K_B and T^*, such that

$$\sup_{k\triangle t\in[0,T^*]}\Big\{\left\|w^k\right\|_X^2\Big\}+\triangle t\sum_{k\triangle t\in]0,T^*]}(\triangle t)^2\left\|Aw^{k+1}\right\|_X^2 + \triangle t\sum_{k\triangle t\in]0,T^*]}(\triangle t)^5\left\|A^2w^{k+1}\right\|_X^2\leq G\triangle t\sum_{j\triangle t\in[0,T^*]}\left\|B\Big(\frac{z^j+\tilde{z}^{j+1}}{2}\Big)\right\|_Y^2.$$

Combined with (5.2.24), we get that

$$\triangle t\sum_{k\triangle t\in]0,T^*]}\Big(\left\|w^k\right\|_X^2+\left\|\tilde{w}^{k+1}\right\|_X^2\Big)+\triangle t\sum_{k\triangle t\in]0,T^*]}(\triangle t)^2\left\|Aw^{k+1}\right\|_X^2 + \triangle t\sum_{k\triangle t\in]0,T^*]}(\triangle t)^5\left\|A^2w^{k+1}\right\|_X^2\leq G\triangle t\sum_{j\triangle t\in[0,T^*]}\left\|B\Big(\frac{z^j+\tilde{z}^{j+1}}{2}\Big)\right\|_Y^2. \quad (5.2.26)$$

Combining (5.2.20), (5.2.26) and the fact that B is bounded, we get the existence of a constant c such that

$$c\left\|z^0\right\|_X^2\leq\triangle t\sum_{k\triangle t\in[0,T^*]}\left\|B\Big(\frac{z^k+\tilde{z}^{k+1}}{2}\Big)\right\|_Y^2+\triangle t\sum_{k\triangle t\in[0,T^*[}(\triangle t)^2\left\|Az^{k+1}\right\|_X^2 + \triangle t\sum_{k\triangle t\in[0,T^*[}(\triangle t)^5\left\|A^2z^{k+1}\right\|_X^2. \quad (5.2.27)$$

Finally, using the energy identity (5.2.9), we get that

$$\left\|z^{T^*/\triangle t}\right\|_X^2\leq(1-c)\left\|z^0\right\|_X^2. \quad (5.2.28)$$

The semi-group property then implies Theorem 5.1.1. □

Remark 5.2.5. Our proof of Theorem 5.1.1 needs to introduce a parameter $\delta>0$, that we can choose arbitrarily. It would be natural to look for the choice of $\delta>0$ yielding the best estimate in the decay rate of the energy. However, our method, based on the arguments of [16], does not give a good approximation of the decay rate of the energy. This is a drawback of this method, which also appears in the continuous setting.

5.2.3 Some variants

Other discretization schemes. Other discretization schemes for system (5.1.1) are possible. For instance, we can consider the following one:

$$\left\{\begin{array}{rcl} \dfrac{z_1^{k+1}-z^k}{\triangle t} &=& A\Big(\dfrac{z^k+z_1^{k+1}}{2}\Big), \quad k\in\mathbb{N},\\ \dfrac{z^{k+1}-z_1^{k+1}}{\triangle t} &=& -B^*Bz^{k+1}, \quad k\in\mathbb{N},\\ z^0 &=& z_0. \end{array}\right. \tag{5.2.29}$$

As for system (5.1.5), the results of [28], in the context of the conservative wave equation, allow proving the existence of spurious high-frequency waves, which do not propagate. This suffices to show the lack of uniform exponential decay for (5.2.29).

Therefore, we need to add a numerical viscosity term. We have at least two choices to introduce this numerical viscosity: Either we consider

$$\left\{\begin{array}{rcl} \dfrac{z_1^{k+1}-z^k}{\triangle t} &=& A\Big(\dfrac{z^k+z_1^{k+1}}{2}\Big), \quad k\in\mathbb{N},\\ \dfrac{z^{k+1}-z_1^{k+1}}{\triangle t} &=& -B^*Bz^{k+1}+(\triangle t)^2A^2z^{k+1}, \quad k\in\mathbb{N},\\ z^0 &=& z_0, \end{array}\right. \tag{5.2.30}$$

or

$$\left\{\begin{array}{rcl} \dfrac{z_1^{k+1}-z^k}{\triangle t} &=& A\Big(\dfrac{z^k+z_1^{k+1}}{2}\Big), \quad k\in\mathbb{N},\\ \dfrac{z_2^{k+1}-z_1^{k+1}}{\triangle t} &=& -B^*Bz_2^{k+1}, \quad k\in\mathbb{N},\\ \dfrac{z^{k+1}-z_2^{k+1}}{\triangle t} &=& (\triangle t)^2A^2z^{k+1}, \quad k\in\mathbb{N},\\ z^0 &=& z_0. \end{array}\right. \tag{5.2.31}$$

The proof above of the uniform exponential decay rate can be adapted to both systems. The low frequency components can be observed similarly. The same decoupling argument between low and high frequencies can be applied as well. Indeed, putting $B=0$ into systems (5.2.30) and (5.2.31) yields again system (5.1.15). Therefore we can get the same results as for system (5.1.9).

Theorem 5.2.6. *Assume that system* (5.1.1) *is exponentially stable, i.e. satisfies* (5.1.4) *with constants* μ *and* ν *and that* $B\in\mathfrak{L}(X,Y)$.

Then there exist two positive constants μ_0 *and* ν_0 *depending only on* μ, ν *and* $\|B\|_{\mathfrak{L}(X,Y)}$, *such that any solution of* (5.2.30) *or of* (5.2.31) *satisfies* (5.1.8) *with constants* μ_0 *and* ν_0 *uniformly with respect to the discretization parameter* $\triangle t>0$.

We skip the proof since it is similar to the previous one.

Other viscosity operators. Other viscosity operators could have been chosen. In our approach, we used the viscosity term $(\triangle t)^2A^2$, which is unbounded, but we could have considered the viscosity operator

$$(\triangle t)\mathcal{V}_{\triangle t}=\frac{(\triangle t)^2A^2}{I-(\triangle t)^2A^2}, \tag{5.2.32}$$

which is well defined, since A^2 is a definite negative operator, and commutes with A. This choice presents the advantage that the viscosity operator now is bounded, keeping the properties of being small at frequencies of order less than $1/\triangle t$ and of order 1 on frequencies of order $1/\triangle t$ and more. Again, the same proof as the one presented above works.

The following result constitutes a generalization of Theorem 5.1.1, and applies to a wide range of viscosity operators, and, in particular, to (5.2.32).

Theorem 5.2.7. *Assume that system* (5.1.1) *is exponentially stable, and that* $B \in \mathfrak{L}(X, Y)$.

Consider a viscosity operator $\mathcal{V}_{\triangle t}$ *such that there exists* $\delta > 0$ *such that:*

1. $\mathcal{V}_{\triangle t}$ *defines a self-adjoint negative definite operator.*
2. *The operators* $\pi_{\delta/\triangle t}$ *and* $\mathcal{V}_{\triangle t}$ *commute.*
3. *There exist two positive constants* $c > 0$ *and* $C > 0$ *such that*

$$\begin{cases} \sqrt{\triangle t}\left\|\left(\sqrt{-\mathcal{V}_{\triangle t}}\right)z\right\|_X \leq C\left\|z\right\|_X, \ \forall z \in \mathcal{C}_{\delta/\triangle t}, \\ \sqrt{\triangle t}\left\|\left(\sqrt{-\mathcal{V}_{\triangle t}}\right)z\right\|_X \geq c\left\|z\right\|_X, \ \forall z \in \mathcal{C}^{\perp}_{\delta/\triangle t}, \end{cases}$$

uniformly with respect to $\triangle t > 0$.

Then the solutions of

$$\begin{cases} \dfrac{\tilde{z}^{k+1} - z^k}{\triangle t} &= A\Big(\dfrac{z^k + \tilde{z}^{k+1}}{2}\Big) - B^*B\Big(\dfrac{z^k + \tilde{z}^{k+1}}{2}\Big), \quad k \in \mathbb{N}, \\ \dfrac{z^{k+1} - \tilde{z}^{k+1}}{\triangle t} &= (\triangle t)\mathcal{V}_{\triangle t} z^{k+1}, \quad k \in \mathbb{N}, \\ z^0 &= z_0. \end{cases} \tag{5.2.33}$$

are exponentially uniformly decaying in the sense of (5.1.8).

A similar result holds for the corresponding variants of systems (5.2.30) and (5.2.31).

5.3 Stabilization of time-discrete systems depending on a parameter

This section is devoted to study time-discrete approximation schemes of abstract systems of the form (5.1.1) depending on a parameter, that can be for instance the space-mesh size when dealing with fully discrete approximation schemes, in which case A is a space discretization of a partial differential operator. As we shall see, the results of the previous section apply.

Furthermore, in the context of fully discrete systems, we shall also show that introducing a suitable CFL type condition, it is unnecessary to add a numerical viscosity term to obtain the uniform exponential decay of the energy. This is so, roughly, because the CFL condition itself rules out the high frequency components without the need of numerical viscosity.

As said in the introduction, this approach requires observability properties to hold uniformly with respect to the space discretization parameter for solutions of the space semi-discrete schemes for *any*

initial data. However, in numerous applications, the space semi-discrete approximation schemes are only observable at low frequencies. We therefore develop our arguments to deal with this case adding a stronger numerical viscosity operator to efficiently damp out the high-frequencies which are not ruled out in the time continuous setting. Simultaneously, we prove a result for space semi-discrete approximation schemes which, to our knowledge, had not been stated so far in such a general setting, even if some instances can be found in [27, 25, 13].

Again, the strategy we propose is strongly based on the methods and results in [12], especially Theorem 5.2.1 given above. Applications to the stabilization of numerical approximation schemes for the damped wave equation are given in Section 5.4.

5.3.1 The general case

To state our results, it is convenient to introduce the following class of pairs of operators (A, B):

Definition 5.3.1. For any $(K_B, \mu, \nu) \in (\mathbb{R}_+^*)^3$, we define $\mathfrak{D}(K_B, \mu, \nu)$ as the class of operators (A, B) satisfying:

(A1) The operator A is skew-adjoint on some Hilbert space X, and has a compact resolvent.

(A2) The operator B is in $\mathfrak{L}(X, Y)$, where Y is a Hilbert space, and satisfies (5.2.8) with constant K_B.

(A3) System (5.1.1) is exponentially stable, and solutions of (5.1.1) satisfy (5.1.4) with constants μ and ν.

Note that this definition does not depend on the Hilbert spaces X and Y.

In this class, Theorems 5.1.1-5.2.6-5.2.7 apply and provide uniform exponential decay properties for the time semi-discrete approximation scheme (5.1.9). This can be deduced from the explicit dependence of the constants entering in Theorems 5.1.1-5.2.6-5.2.7, which only depend on K_B, μ and ν. At this point, the fact that the class $\mathcal{D}(K_B, \mu, \nu)$ is independent of the spaces X and Y plays a key role.

Also note that Definition 5.3.1 only refers to the behavior of the *continuous* system (5.1.1), although, as we have seen, and in particular in view of Theorem 5.2.1, it also has applications in what concerns time-discrete systems.

This method allows dealing with fully discrete approximation schemes. In that setting, we consider a family of operators $(A_{\triangle x}, B_{\triangle x})$, where $\triangle x > 0$ is the standard parameter associated with the space mesh-size. In this way one can use automatically the existing results for space semi-discretizations as, for instance, [1, 5, 6, 13, 11, 23, 24, 27].

Note that the work [24] is not dealing with stabilization properties, but rather with controllability properties of space semi-discrete schemes. However, it is standard that these two properties (controllability and stabilization) are very close, since both are equivalent to observability properties. Therefore, these works can be adapted to study the stabilization properties as well. We refer to the survey article [32] for more details and more references.

Remark 5.3.2. We emphasize that this approach is based on the systematic use of existing results for space semi-discretizations. One could proceed all the way around, first applying the results in

this paper to derive uniform stabilization results for time discrete approximation schemes and then discretizing the space variables. For doing this, however, due to the more complex dependence of the PDE and its space semi-discretizations on the space variables, there is no systematic way of transferring results from the continuous to the discrete setting. In this sense, the method we propose here of using the existing results for space semi-discretizations to later apply the results in this paper about time discretizations is much more easier to be implemented and yields better results.

5.3.2 Stabilization of fully discrete approximation schemes without viscosity

This subsection is devoted to prove a particular result for fully discrete approximation schemes under a CFL type assumption on the space and time discretization parameters, which does not require adding numerical viscosity terms. We observe, however, that this approach requires, often, restrictions on $\triangle t$ that can be avoided by adding numerical viscosity terms.

Theorem 5.3.3. *Let $(A_{\triangle x}, B_{\triangle x})_{\triangle x>0}$ be a family of operators defined on Hilbert spaces $X_{\triangle x}$ endowed with a norm $\|\cdot\|_{\triangle x}$. Assume that there exist positive constants K_B, μ and ν such that, for all $\triangle x > 0$, $(A_{\triangle x}, B_{\triangle x}) \in \mathfrak{D}(K_B, \mu, \nu)$.*

Then, for any $\eta > 0$, there exist positive constants μ_η and ν_η such that the solutions of

$$\left\{ \begin{array}{rcl} \dfrac{z^{k+1}_{\triangle x} - z^k_{\triangle x}}{\triangle t} & = & A_{\triangle x}\Big(\dfrac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big) - B^*_{\triangle x} B_{\triangle x}\Big(\dfrac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big),\ k \in \mathbb{N}, \\ z^0_{\triangle x} & = & z_{0,\triangle x} \in X_{\triangle x}, \end{array} \right. \tag{5.3.1}$$

satisfy

$$\Big\| z^k_{\triangle x} \Big\|^2_{\triangle x} \leq \mu_\eta \left\| z^0_{\triangle x} \right\|^2_{\triangle x} \exp(-\nu_\eta k \triangle t), \quad k \geq 0, \tag{5.3.2}$$

uniformly with respect to $\triangle t > 0$ and $\triangle x > 0$ provided that

$$\|A_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x}, X_{\triangle x})} \leq \frac{\eta}{\triangle t}. \tag{5.3.3}$$

Remark 5.3.4. In practical applications, the operator $A_{\triangle x}$ is often a space discretization of an unbounded operator A, for which we typically have a bound of the form $\|A_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x}, X_{\triangle x})} \simeq C(\triangle x)^{-\sigma}$ for some positive exponent σ. In this case, condition (5.3.3) is guaranteed as soon as

$$\frac{C}{(\triangle x)^\sigma} \leq \frac{\eta}{\triangle t}.$$

The CFL condition (5.3.3) therefore imposes the ratio $\triangle t/(\triangle x)^\sigma$ to be uniformly bounded when $\triangle x$ and $\triangle t$ go to 0.

Remark 5.3.5. This theorem implies that we do not need to add a numerical viscosity term on the time-discrete approximation schemes to get a uniform exponential decay of the energies if we impose a CFL type condition on the discretization parameters $\triangle x$ and $\triangle t$.

Proof. The proof of Theorem 5.3.3 is actually easier than the one of Theorem 5.1.1, since we do not need the decomposition (5.2.11) into low and high frequency components. In some sense, the CFL rules out the high frequency components.

First, we derive the energy identity for solutions of (5.3.1):

$$\Big\| z^l_{\triangle x} \Big\|^2_{\triangle x} = \left\| z^0_{\triangle x} \right\|^2_{\triangle x} - 2\triangle t \sum_{k=0}^{l-1} \left\| B_{\triangle x}\Big(\frac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big) \right\|^2_{Y_{\triangle x}}, \quad l \in \mathbb{N}. \tag{5.3.4}$$

Second, since $(A_{\triangle x}, B_{\triangle x}) \in \mathfrak{D}(K_B, \mu, \nu)$, the resolvent estimates (5.2.4) involving $A_{\triangle x}$ and $B_{\triangle x}$ hold uniformly with respect to $\triangle x > 0$, due to Lemma 5.2.3.

Then, applying Theorem 5.2.1 with $\delta = \eta$, because of assumption (5.3.3) that implies that $\mathcal{C}_{\eta/\triangle t}(A_{\triangle x}) = X_{\triangle x}$, we get a time $T^* > 0$ and a positive constant k_{T^*} independent of $\triangle x > 0$ such that any solution $y_{\triangle x}$ of

$$\begin{cases} \dfrac{y^{k+1}_{\triangle x} - y^k_{\triangle x}}{\triangle t} = A_{\triangle x}\Big(\dfrac{y^k_{\triangle x} + y^{k+1}_{\triangle x}}{2}\Big), \quad k \in \mathbb{N}, \\ y^0_{\triangle x} = y_{0,\triangle x} \in X_{\triangle x}, \end{cases} \tag{5.3.5}$$

satisfies

$$k_{T^*} \left\|y^0_{\triangle x}\right\|^2_{\triangle x} \leq \triangle t \sum_{k\triangle t \in [0,T^*]} \left\| B_{\triangle x}\Big(\frac{y^k_{\triangle x} + y^{k+1}_{\triangle x}}{2}\Big)\right\|^2_{Y_{\triangle x}}. \tag{5.3.6}$$

Now, let $z_{0,\triangle x} \in X_{\triangle x}$, and consider the solutions $z_{\triangle x}$ of (5.3.1) and $y_{\triangle x}$ of (5.3.5) with initial data $y_{0,\triangle x} = z_{0,\triangle x}$. Set $w_{\triangle x} = z_{\triangle x} - y_{\triangle x}$. Then

$$k_{T^*} \left\|z^0_{\triangle x}\right\|^2_{\triangle x} \leq 2\triangle t \sum_{k\triangle t \in [0,T^*]} \left(\left\| B_{\triangle x}\Big(\frac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big)\right\|^2_{Y_{\triangle x}} + \left\| B_{\triangle x}\Big(\frac{w^k_{\triangle x} + w^{k+1}_{\triangle x}}{2}\Big)\right\|^2_{Y_{\triangle x}} \right). \tag{5.3.7}$$

Therefore, we only need to bound the last term. This is easier than in (5.2.20). Indeed, $w_{\triangle x}$ satisfies

$$\frac{w^{k+1}_{\triangle x} - w^k_{\triangle x}}{\triangle t} = A_{\triangle x}\Big(\frac{w^k_{\triangle x} + w^{k+1}_{\triangle x}}{2}\Big) - B^*_{\triangle x} B_{\triangle x}\Big(\frac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big), \ k \in \mathbb{N}, \tag{5.3.8}$$

with $w^0_{\triangle x} = 0$.

The energy estimates on $w_{\triangle x}$ now give, for $l \in \mathbb{N}$

$$\left\|w^l_{\triangle x}\right\|^2_{\triangle x} = -2\triangle t \sum_{k=0}^{l-1} < B_{\triangle x}\Big(\frac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big), B_{\triangle x}\Big(\frac{w^k_{\triangle x} + w^{k+1}_{\triangle x}}{2}\Big) >_{Y_{\triangle x}},$$

and then

$$\left\|w^l_{\triangle x}\right\|^2_{\triangle x} \leq \triangle t \left\|B_{\triangle x}\right\|^2_{\mathfrak{L}(X_{\triangle x} Y_{\triangle x})} \sum_{k=0}^{l-1} \left\| \frac{w^k_{\triangle x} + w^{k+1}_{\triangle x}}{2}\right\|^2_{\triangle x} + \triangle t \sum_{k=0}^{l-1} \left\| B_{\triangle x}\Big(\frac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big)\right\|^2_{Y_{\triangle x}}.$$

Since $\left\|B_{\triangle x}\right\|_{\mathfrak{L}(X_{\triangle x}, Y_{\triangle x})} \leq K_B$, applying Grönwall's Lemma, we obtain a constant G independent of $\triangle x > 0$ such that

$$\triangle t \sum_{k\triangle t \in [0,T^*]} \left\|w^k_{\triangle x}\right\|^2_{\triangle x} \leq G\triangle t \sum_{k\triangle t \in [0,T^*]} \left\| B_{\triangle x}\Big(\frac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big)\right\|^2_{Y_{\triangle x}}.$$

This last inequality implies with (5.3.7) that

$$k_{T^*} \left\|z^0_{\triangle x}\right\|^2_{\triangle x} \leq 2(1 + K_B^2 G)\triangle t \sum_{k\triangle t \in [0,T^*]} \left\| B_{\triangle x}\Big(\frac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2}\Big)\right\|^2_{Y_{\triangle x}}.$$

Plugging this inequality in (5.3.4) for $l^* = \lceil T^*/\triangle t \rceil$ gives

$$\left\| z^{l^*}_{\triangle x} \right\|^2_{\triangle x} \leq \left\| z^0_{\triangle x} \right\|^2_{\triangle x} \Big(1 - \frac{k_{T^*}}{1 + K_B^2 G} \Big).$$

As previously, setting

$$\alpha = \Big(1 - \frac{k_{T^*}}{1 + K_B^2 G} \Big),$$

which is independent of $\triangle t$, we obtain that

$$\left\| z^{l}_{\triangle x} \right\|^2_{\triangle x} \leq \left\| z^0_{\triangle x} \right\|^2_{\triangle x} \exp\Big(\Big(\frac{l \triangle t}{T^*} - 1 \Big) \ln(\alpha) \Big), \quad \forall l \in \mathbb{N},$$

which proves the result. □

Remark 5.3.6. As before, the proof of Theorem 5.3.3 can also be carried out for the time-discrete scheme

$$\left\{ \begin{array}{rcl} \dfrac{z^{k+1}_{\triangle x} - z^k_{\triangle x}}{\triangle t} & = & A_{\triangle x} \Big(\dfrac{z^k_{\triangle x} + z^{k+1}_{\triangle x}}{2} \Big), \quad k \in \mathbb{N}, \\ \dfrac{z^{k+1}_{\triangle x} - z^{k+1}_{\triangle x}}{\triangle t} & = & -B^*_{\triangle x} B_{\triangle x} z^{k+1}_{\triangle x}, \quad k \in \mathbb{N}, \\ z^0_{\triangle x} & = & z_{0,\triangle x} \in X_{\triangle x}, \end{array} \right. \tag{5.3.9}$$

under the CFL condition (5.3.3).

5.3.3 Stabilization of fully discrete approximation schemes with viscosity

In this Subsection, we consider the case in which the space semi-discrete systems are uniformly observable for initial data lying in filtered subspaces, as it occurs often, see [18, 31, 13, 32].

Theorem 5.3.7. *Let $(A_{\triangle x}, B_{\triangle x})_{\triangle x > 0}$ be a family of operators defined on Hilbert spaces $X_{\triangle x}$ endowed with the norms $\|\cdot\|_{\triangle x}$.*

Assume that there exists a constant K_B such that for all $\triangle x > 0$, the operator norm $\|B_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x}, Y_{\triangle x})}$ is bounded by K_B.

Assume that there exist positive constants η, σ, T and k_T such that for all initial data $y_0 \in \mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x})$, the solution y of

$$\dot{y} = A_{\triangle x} y, \quad t \in \mathbb{R}, \qquad y(0) = y_0 \in \mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x}), \tag{5.3.10}$$

satisfies

$$k_T \left\| y_0 \right\|^2_{\triangle x} \leq \int_0^T \left\| B_{\triangle x} y(t) \right\|^2_{Y_{\triangle x}} \, dt. \tag{5.3.11}$$

Set $\varepsilon = \max\{\triangle t, (\triangle x)^\sigma\}$.

Consider a viscosity operator $\mathcal{V}_\varepsilon$ such that:

1. *$\mathcal{V}_\varepsilon$ defines a self-adjoint negative definite operator.*
2. *The operators $\pi_{1/\varepsilon}$ and $\mathcal{V}_\varepsilon$ commute.*

3. There exist two positive constants $c>0$ and $C>0$ such that

$$\begin{cases} \sqrt{\varepsilon}\left\|\left(\sqrt{-\mathcal{V}_\varepsilon}\right)z\right\|_{\triangle x} \leq C\left\|z\right\|_{\triangle x},\ \forall z\in\mathcal{C}_{1/\varepsilon}(A_{\triangle x}),\\ \sqrt{\varepsilon}\left\|\left(\sqrt{-\mathcal{V}_\varepsilon}\right)z\right\|_{\triangle x} \geq c\left\|z\right\|_{\triangle x},\ \forall z\in\mathcal{C}_{1/\varepsilon}(A_{\triangle x})^{\perp}, \end{cases}$$

uniformly with respect to $\varepsilon>0$.

Then the solutions of

$$\begin{cases} \dfrac{\tilde{z}^{k+1}-z^k}{\triangle t} &= A_{\triangle x}\Big(\dfrac{z^k+\tilde{z}^{k+1}}{2}\Big)-B_{\triangle x}^*B_{\triangle x}\Big(\dfrac{z^k+\tilde{z}^{k+1}}{2}\Big),\quad k\in\mathbb{N},\\ \dfrac{z^{k+1}-\tilde{z}^{k+1}}{\triangle t} &= \varepsilon\mathcal{V}_\varepsilon z^{k+1},\quad k\in\mathbb{N},\\ z^0 &= z_0. \end{cases} \tag{5.3.12}$$

are exponentially uniformly decaying in the sense of (5.3.2).

Sketch of the proof. The proof can be done similarly as the one of Theorems 5.1.1-5.2.7. The main difference in the proof is that the low and high-frequency components are separated by the frequency $1/\varepsilon$ instead of $1/\triangle t$.

As explained in [12], the observability inequalities (5.3.11) in the filtered spaces $\mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x})$ imply observability inequalities (5.2.5) for solutions of (5.1.18) with initial data lying in $\mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x})\cap\mathcal{C}_{1/\triangle t}(A_{\triangle x})=\mathcal{C}_{1/\varepsilon}(A_{\triangle x})$. The proof of this fact simply consists in the following remark: the uniform observability inequalities (5.3.11) in the filtered spaces $\mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x})$ imply uniform resolvent estimates (5.2.4) for data in $\mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x})$, and Theorem 5.2.1, due to the explicit dependence of the constants in (5.2.5) on the constants m and M appearing in (5.2.4), yields the result.

Then, we replace system (5.1.15) by

$$\begin{cases} \dfrac{\tilde{u}^{k+1}-u^k}{\triangle t}=A_{\triangle x}\Big(\dfrac{u^k+\tilde{u}^{k+1}}{2}\Big),\quad k\in\mathbb{N},\\ \dfrac{u^{k+1}-\tilde{u}^{k+1}}{\triangle t}=\varepsilon\mathcal{V}_\varepsilon u^{k+1},\quad k\in\mathbb{N},\\ u^0=u_0, \end{cases} \tag{5.3.13}$$

and consider u_l and u_h defined by

$$u_l=\pi_{1/\varepsilon}u,\qquad u_h=(I-\pi_{1/\varepsilon})u,$$

instead of (5.2.11).

The rest of the proof follows line to line that of Lemma 5.2.4 and is left to the reader. □

Theorem 5.3.7 also yields an interesting corollary for time-continuous systems:

Corollary 5.3.8. *Let $(A_{\triangle x},B_{\triangle x})_{\triangle x>0}$ be a family of operators defined on Hilbert spaces $X_{\triangle x}$ endowed with the norms $\|\cdot\|_{\triangle x}$.*

Assume that there exists a constant K_B such that for all $\triangle x>0$, the operator norm $\|B_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x},Y_{\triangle x})}$ is bounded by K_B.

Assume that there exist positive constants η, σ, T and k_T such that for all initial data $y_0 \in \mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x})$, the solution y of (5.3.10) *satisfies* (5.3.11).

Consider a viscosity operator $\mathcal{V}_{\triangle x}$ such that:

1. *$\mathcal{V}_{\triangle x}$ defines a self-adjoint negative definite operator.*
2. *The operators $\pi_{\eta/(\triangle x)^\sigma}$ and $\mathcal{V}_{\triangle x}$ commute.*
3. *There exist two positive constants $c > 0$ and $C > 0$ such that*

$$\begin{aligned}(\triangle x)^{\sigma/2} \left\| \left(\sqrt{-\mathcal{V}_{\triangle x}}\right) z \right\|_{\triangle x} &\leq C \|z\|_{\triangle x}, \quad \forall z \in \mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x}),\\ (\triangle x)^{\sigma/2} \left\| \left(\sqrt{-\mathcal{V}_{\triangle x}}\right) z \right\|_{\triangle x} &\geq c \|z\|_{\triangle x}, \quad \forall z \in \mathcal{C}_{\eta/(\triangle x)^\sigma}(A_{\triangle x})^\perp,\end{aligned}$$

uniformly with respect to $\triangle x > 0$.

Then the solutions of

$$\begin{cases} \dot{z} = A_{\triangle x} z - B_{\triangle x}^* B_{\triangle x} z + (\triangle x)^\sigma \mathcal{V}_{\triangle x} z, \quad t \in \mathbb{R}_+, \\ z(0) = z_0. \end{cases} \tag{5.3.14}$$

are exponentially uniformly decaying in the sense of (5.1.4).

Indeed, this can be deduced from Theorem 5.3.7 by letting $\triangle t \to 0$.

Corollay 5.3.8 can be seen as a generalization of [14], where similar results have been derived for viscous approximations of (5.1.1). In [14], the same result is obtained but the assumptions differ in one essential point: The observability inequality (5.1.16) for solutions of (5.1.14) is assumed to hold for *any* initial data, and not only in a filtered space as in Corollary 5.3.8. Thus, in [14], no assumption is required on the viscosity parameter.

Though, the proof in [14] can be easily adapted to prove Corollay 5.3.8 directly for time continuous systems.

Also remark that some instances of applications of variants of Corollary 5.3.8 can be found in several different articles dealing with space semi-discrete damped systems [27, 25, 23, 13].

In Subsection 5.4.3, we will indicate without proof how one can deduce the results in [27, 23] from the results in [18] and the methods developped in [14] and here.

Remark 5.3.9. Corollary 5.3.8 yields optimal results in the following sense: If system (5.3.14) is exponentially decaying for $\mathcal{V}_{\triangle x} = -|A_{\triangle x}|$, which always satisfies the assumptions of Corollary 5.3.8, uniformly with respect to the space discretization parameter, then there exists $\varepsilon > 0$ such that any solution y of (5.3.10) with initial data in $\mathcal{C}_{\varepsilon/(\triangle x)^\sigma}(A_{\triangle x})$ satisfies (5.3.11). Indeed, in this case, following the proof of Lemma 5.2.3, one can prove that there exist a time $T > 0$ and a constant $k_T > 0$ such that, for any $\triangle x > 0$, any solution y of (5.3.10) satisfies

$$k_T \|y_0\|_{\triangle x}^2 \leq \int_0^T \|B_{\triangle x} y(t)\|_{Y_{\triangle x}}^2 \, dt + \int_0^T (\triangle x)^\sigma \left\| \left(\sqrt{|A_{\triangle x}|}\right) y(t) \right\|_{\triangle x}^2 dt.$$

In particular, if the initial data lies in $\mathcal{C}_{\varepsilon/(\triangle x)^\sigma}(A_{\triangle x})$, we have that

$$k_T \|y_0\|_{\triangle x}^2 \leq \int_0^T \|B_{\triangle x} y(t)\|_{Y_{\triangle x}}^2 \, dt + \varepsilon T \|y_0\|_{\triangle x}^2,$$

and then, taking $\varepsilon = k_T/2T$, we recover (5.3.11).

5.4 Applications

The goal of this section is to present several applications of Theorems 5.1.1-5.3.3 to the damped wave equation. Of course, the Schrödinger and plate equations, and the system of elasticity, among others, enter in this frame too, but the applications to these other models will be presented elsewhere.

5.4.1 The time-discrete damped wave equation

Consider a smooth non-empty open bounded domain $\Omega \subset \mathbb{R}^d$.

We consider the following initial boundary value problem:

$$\begin{cases} u_{tt} - \Delta_x u + \sigma(x)^2 u_t = 0, & x \in \Omega, \quad t \geq 0, \\ u(x,t) = 0, & x \in \partial\Omega, \ \ t \geq 0, \\ u(x,0) = u_0 \in H_0^1(\Omega), \quad u_t(x,0) = v_0 \in L^2(\Omega), & x \in \Omega, \end{cases} \tag{5.4.1}$$

where $\sigma : \Omega \to \mathbb{R}_+$ is a non-negative bounded function which is strictly positive in some open non-empty subset $\omega \subset \Omega$: There exists $\alpha > 0$ such that

$$\sigma^2(x) \geq \alpha, \quad \forall x \in \omega. \tag{5.4.2}$$

The energy of solutions of (5.4.1)

$$E(t) = \frac{1}{2} \int_\Omega \Big[|\partial_t u(t,x)|^2 + |\nabla u(t,x)|^2 \Big] dx, \tag{5.4.3}$$

satisfies the dissipation law

$$\frac{dE}{dt}(t) = -\int_\omega \sigma(x)^2 |\partial_t u(t,x)|^2 \, dx, \qquad \forall\, t \in [0,T]. \tag{5.4.4}$$

It is well-known that the energy (5.4.3) decays exponentially if the set ω satisfies a geometric condition, namely the so-called *Geometric Control Condition*, introduced in [2, 3]: there exists a time $T > 0$ such that all the rays of Geometric Optics in Ω enter the set ω in a time smaller than T.

To show that system (5.4.1) enters in the abstract setting of this paper, let us recall that it is equivalent to

$$\dot{Z} = AZ - B^*BZ, \text{ with } Z = \begin{pmatrix} u \\ v \end{pmatrix}, \ A = \begin{pmatrix} 0 & Id \\ \Delta_x & 0 \end{pmatrix}, \ B = \begin{pmatrix} 0 & \sigma \end{pmatrix}. \tag{5.4.5}$$

In this setting, A is a skew-adjoint unbounded operator on the Hilbert space $X = H_0^1(\Omega) \times L^2(\Omega)$, with domain $\mathcal{D}(A) = H^2 \cap H_0^1(\Omega) \times H_0^1(\Omega)$. From the assumptions (5.4.2) on σ, the operator B is obviously continuous on X.

Besides, the energy (5.4.3) of (5.4.1) reads as $\|Z(t)\|_X^2 /2$.

Then, we introduce the following time semi-discrete approximation scheme:

$$\begin{cases} \dfrac{\tilde{Z}^{k+1}-Z^k}{\triangle t} = \begin{pmatrix} 0 & Id \\ \Delta_x & 0 \end{pmatrix} \Big(\dfrac{Z^k+\tilde{Z}^{k+1}}{2}\Big) - \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix} \Big(\dfrac{Z^k+\tilde{Z}^{k+1}}{2}\Big), \quad k \in \mathbb{N}^*, \\ \dfrac{Z^{k+1}-\tilde{Z}^{k+1}}{\triangle t} = (\triangle t)^2 \begin{pmatrix} \Delta_x & 0 \\ 0 & \Delta_x \end{pmatrix} Z^{k+1}, \quad k \in \mathbb{N}^*, \\ Z^0 = \begin{pmatrix} u_0 \\ \\ v_0 \end{pmatrix}. \end{cases} \tag{5.4.6}$$

We then define the energy as in (5.1.6).

According to Theorem 5.1.1, we get:

Theorem 5.4.1. *Assume that the damping function σ satisfies* (5.4.2) *for a non-empty open set $\omega \subset \Omega$, that satisfies the Geometric Control Condition.*

Then there exist positive constants ν_0 and μ_0 such that any solution of (5.4.6) *satisfies* (5.1.8) *uniformly with respect to the discretization parameter $\triangle t > 0$.*

5.4.2 A fully discrete damped wave equation: The mixed finite element method

Here we present an application to a fully discrete approximation scheme. To present our results properly, we first need to recall some properties of the *space* semi-discrete wave equation.

We now consider the damped wave equation (5.4.1) in 1d, that is with $\Omega = (0,1)$. We still assume that the damping function σ is non-negative, bounded, and satisfies (5.4.2). Note that in this case the *Geometric Control Condition* is automatically satisfied, and therefore the decay of the energy of (5.4.1) is exponential.

When semi-discretizing equation (5.4.1) in *space*, it may happen that the *space* semi-discrete approximations are *not* exponentially stable uniformly with respect to the space discretization parameter. This has been observed in many cases, for instance in [15, 18, 21, 13]. We refer to the review article [32] for more references.

A possible cure has been proposed in [1] and analyzed in [5, 6, 11] based on a mixed finite element method, on which we will focus now.

Let N be a nonnegative integer. Set $\triangle x = 1/(N+1)$ and consider the subdivision of $(0,1)$ given by

$$0 = x_0 < x_1 < \cdots < x_j = j\triangle x < \cdots x_{N+1} = 1.$$

Let us present the space semi-discrete approximation scheme of (5.4.1) in 1d, on $(0,1)$, derived from the mixed finite element method (see [1, 5, 6, 11])

$$\begin{cases} \dfrac{\ddot{u}_{j-1}+2\ddot{u}_j+\ddot{u}_{j+1}}{4} - \dfrac{u_{j+1}-2u_j+u_{j-1}}{(\triangle x)^2} + \dfrac{1}{4}\Big(\sigma^2_{j-1/2}(\dot{u}_{j-1}+\dot{u}_j) \\ \qquad\qquad +\sigma^2_{j+1/2}(\dot{u}_j+\dot{u}_{j+1})\Big) = 0, \qquad (t,j) \in \mathbb{R}_+ \times \{1,\cdots,N\}, \\ u_0(t) = u_{N+1}(t) = 0, \quad t \in \mathbb{R}_+, \\ u_j(0) = u_{j,0}, \quad \dot{u}_j(0) = v_{j,0}, \quad j \in \{1,\cdots,N\}, \end{cases} \tag{5.4.7}$$

where $\sigma^2_{j+1/2}$ is an approximation of σ^2 on $[j\triangle x, (j+1)\triangle x]$.

The energy of solutions of (5.4.7) is defined by

$$E_{\triangle x}(t) = \frac{\triangle x}{2}\sum_{j=0}^{N}\Big(\Big|\frac{\dot u_j + \dot u_{j+1}}{2}\Big|^2 + \Big|\frac{u_{j+1}-u_j}{\triangle x}\Big|^2\Big). \tag{5.4.8}$$

Following [1, 5, 6, 11], one can prove that the energy $E_{\triangle x}$ is exponentially stable, uniformly with respect to $\triangle x > 0$, when σ satisfies (5.4.2).

Let us check that system (5.4.7) is a particular instance of the abstract setting we provided.

Define the $N \times N$ matrix $M_{\triangle x}$ by

$$M_{\triangle x}(i,j) = \begin{cases} 1/2 & \text{if } i = j, \\ 1/4 & \text{if } |i-j| = 1, \\ 0 & \text{else,} \end{cases}$$

which is invertible, self-adjoint and positive definite.

The space semi-discrete approximation scheme (5.4.7) can be written as

$$M_{\triangle x}\ddot U_{\triangle x} + A_{0,\triangle x}U_{\triangle x} + C_{1,\triangle x}\dot U_{\triangle x} = 0, \quad t \in \mathbb{R}_+,$$

where $A_{0,\triangle x}$ is a positive definite matrix $N \times N$, which represents the Laplace discrete operator, and $C_{1,\triangle x}$ is the $N \times N$ matrix

$$C_{1,\triangle x}(i,j) = \begin{cases} (\sigma^2_{j+1/2} + \sigma^2_{j-1/2})/4 & \text{if } i = j, \\ \sigma^2_{i+1/2}/4 & \text{if } i+1 = j, \\ \sigma^2_{i-1/2}/4 & \text{if } i-1 = j, \\ 0 & \text{else.} \end{cases}$$

System (5.4.7) can be rewritten as

$$\dot Z_{\triangle x} = A_{\triangle x}Z_{\triangle x} - C_{\triangle x}Z_{\triangle x}, \quad t \in \mathbb{R}_+, \tag{5.4.9}$$

where $Z_{\triangle x}$, $A_{\triangle x}$ and $C_{\triangle x}$ denote

$$\begin{gathered} Z_{\triangle x} = \begin{pmatrix} U_{\triangle x} \\ V_{\triangle x} \end{pmatrix}, \quad A_{\triangle x} = \begin{pmatrix} 0 & Id \\ -M^{-1}_{\triangle x}A_{0,\triangle x} & 0 \end{pmatrix}, \\ C_{\triangle x} = \begin{pmatrix} 0 & 0 \\ 0 & M^{-1}_{\triangle x}C_{1,\triangle x} \end{pmatrix}. \end{gathered} \tag{5.4.10}$$

Remark that the matrix $A_{\triangle x}$ is skew-adjoint on the energy space $X_{\triangle x} = \mathbb{R}^{2N}$ endowed with the norm

$$\begin{aligned} \left\|\begin{pmatrix} U_{\triangle x} \\ V_{\triangle x} \end{pmatrix}\right\|^2_{\triangle x} &= \triangle x\sum_{j=0}^{N}\Big(\Big|\frac{V_{\triangle x,j} + V_{\triangle x,j+1}}{2}\Big|^2 + \Big|\frac{U_{\triangle x,j+1} - U_{\triangle x,j}}{\triangle x}\Big|^2\Big) \\ &= < M_{\triangle x}V_{\triangle x}, V_{\triangle x} >_{*\triangle x} + < A_{0,\triangle x}U_{\triangle x}, U_{\triangle x} >_{*\triangle x}, \end{aligned}$$

where the scalar product $< \cdot, \cdot >_{*\triangle x}$ is the classical discrete L^2 scalar product, corresponding to the discrete L^2 norm

$$\|V_{\triangle x}\|^2_{*\triangle x} = \triangle x \sum_{j=1}^{N} |V_{\triangle x,j}|^2. \tag{5.4.11}$$

Note that, in this setting, the energy (5.4.8) of solutions of (5.4.7) coincides with the energy $\|Z_{\triangle x}(t)\|^2_{\triangle x}/2$ of solutions of (5.4.9).

Let us check that $C_{\triangle x}$ has the form $B^*_{\triangle x} B_{\triangle x}$ for some $N \times N$ matrix $B_{\triangle x}$. According to Choleski's decomposition, we only have to check that $C_{\triangle x}$ is a selfadjoint positive matrix on $X_{\triangle x}$. For generic vectors $Z_{1\triangle x}$ and $Z_{2\triangle x}$ as in (5.4.10), we have:

$$\begin{aligned} < C_{\triangle x} Z_{1\triangle x}, Z_{2\triangle x} >_{\triangle x} &= < M_{\triangle x} M^{-1}_{\triangle x} C_{1\triangle x} V_{1\triangle x}, V_{2\triangle x} >_{*\triangle x} \\ &= < C_{1\triangle x} V_{1\triangle x}, V_{2\triangle x} >_{*\triangle x} \\ = \triangle x \sum_{j=0}^{N} \sigma^2_{j+1/2} \Big(\frac{V_{1\triangle x,j} + V_{1\triangle x,j+1}}{2} \Big) \Big(\frac{V_{2\triangle x,j} + V_{2\triangle x,j+1}}{2} \Big). \end{aligned} \tag{5.4.12}$$

This last expression shows that $C_{\triangle x}$ is a selfadjoint positive operator on $X_{\triangle x}$. Therefore there exists $B_{\triangle x}$ such that $B^*_{\triangle x} B_{\triangle x} = C_{\triangle x}$. Besides, classical linear algebra implies that

$$\|C_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x}, X_{\triangle x})} = \|B_{\triangle x}\|^2_{\mathfrak{L}(X_{\triangle x}, X_{\triangle x})}.$$

From the computations above, and especially (5.4.12), we have

$$\|C_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x}, X_{\triangle x})} = \sup_{\substack{\|Z_{1\triangle x}\|_{\triangle x} \le 1, \\ \|Z_{2\triangle x}\|_{\triangle x} \le 1}} \{< C_{\triangle x} Z_{1\triangle x}, Z_{2\triangle x} >_{\triangle x}\} \le \left\|\sigma^2\right\|_{L^\infty}. \tag{5.4.13}$$

We are then in the abstract setting given in Section 5.3: Hypothesis (A1) and (A2) of Definition 5.3.1 have been checked above, and (A3) has been proved in [5] (see [1, 6, 11] for related results).

Method I: Adding a numerical viscosity term in time

We add a numerical viscosity term to the scheme above, corresponding to (5.1.9). In this case, the fully discrete approximation scheme reads:

$$\begin{cases} \dfrac{\tilde{u}_j^{k+1}-u_j^k}{\triangle t} = \dfrac{v_j^k+\tilde{v}_j^{k+1}}{2}, \\ \dfrac{1}{4\triangle t}\Big((\tilde{v}_{j-1}^{k+1}+2\tilde{v}_j^{k+1}+\tilde{v}_{j+1}^{k+1})-(v_{j-1}^k+2v_j^k+v_{j+1}^k)\Big) \\ \qquad = \dfrac{1}{2(\triangle x)^2}(\tilde{u}_{j+1}^{k+1}+u_{j+1}^k-2\tilde{u}_j^{k+1}-2u_j^k+\tilde{u}_{j-1}^{k+1}+u_{j-1}^k) \\ \qquad -\dfrac{1}{8}\sigma_{j+1/2}^2\Big((v_j^k+v_{j+1}^k)+(\tilde{v}_j^{k+1}+\tilde{v}_{j+1}^{k+1})\Big) \\ \qquad -\dfrac{1}{8}\sigma_{j-1/2}^2\Big((v_{j-1}^k+v_j^k)+(\tilde{v}_{j-1}^{k+1}+\tilde{v}_j^{k+1})\Big), \\ \dfrac{1}{4\triangle t}\Big((u_{j-1}^{k+1}+2u_j^{k+1}+u_{j+1}^{k+1})-(\tilde{u}_{j-1}^{k+1}+2\tilde{u}_j^{k+1}+\tilde{u}_{j+1}^{k+1})\Big) \\ \qquad = \Big(\dfrac{\triangle t}{\triangle x}\Big)^2(u_{j+1}^{k+1}-2u_j^{k+1}+u_{j-1}^{k+1}), \\ \dfrac{1}{4\triangle t}\Big((v_{j-1}^{k+1}+2v_j^{k+1}+v_{j+1}^{k+1})-(\tilde{v}_{j-1}^{k+1}+2\tilde{v}_j^{k+1}+\tilde{v}_{j+1}^{k+1})\Big) \\ \qquad = \Big(\dfrac{\triangle t}{\triangle x}\Big)^2(v_{j+1}^{k+1}-2v_j^{k+1}+v_{j-1}^{k+1}), \end{cases} \tag{5.4.14}$$

which holds for $(k,j)\in\mathbb{N}\times\{1,\cdots,N\}$, with the boundary conditions

$$u_0^k=u_{N+1}^k=v_0^k=v_{N+1}^k=0,\quad \forall k\in\mathbb{N}, \tag{5.4.15}$$

and the initial data

$$u_j^0=u_{j,0},\quad v_j^0=v_{j,0},\quad \forall j\in\{1,\cdots,N\}. \tag{5.4.16}$$

Here u_j^k and v_j^k respectively denote approximations of the functions u and $\dot{u}$ in $x_j=j\triangle x$ at time $k\triangle t$.

As an application of Theorem 5.1.1, we get:

Theorem 5.4.2. *The energy*

$$E_{\triangle x}^k=\frac{\triangle x}{2}\sum_{j=0}^{N}\Big(\Big|\frac{v_j^k+v_{j+1}^k}{2}\Big|^2+\Big|\frac{u_{j+1}^k-u_j^k}{\triangle x}\Big|^2\Big),\quad k\in\mathbb{N},$$

of solutions of (5.4.14) *is exponentially decaying, uniformly with respect to* $\triangle t>0$ *and* $\triangle x>0$, *in the sense of* (5.3.2).

Method II: Imposing a CFL condition

Here we want to use Theorem 5.3.3 to derive uniform properties on the following fully discrete system, obtained by discretizing in time system (5.4.9) using (5.3.1):

$$\left\{\begin{array}{l} \dfrac{u_j^{k+1}-u_j^k}{\triangle t} = \dfrac{v_j^k+v_j^{k+1}}{2}, \\ \dfrac{1}{4\triangle t}\Big((v_{j-1}^{k+1}+2v_j^{k+1}+v_{j+1}^{k+1})-(v_{j-1}^k+2v_j^k+v_{j+1}^k)\Big) \\ \qquad = \dfrac{1}{2(\triangle x)^2}(u_{j+1}^{k+1}+u_{j+1}^k-2u_j^{k+1}-2u_j^k+u_{j-1}^{k+1}+u_{j-1}^k) \\ \qquad -\dfrac{1}{8}\sigma_{j+1/2}^2\Big((v_j^k+v_{j+1}^k)+(v_j^{k+1}+v_{j+1}^{k+1})\Big) \\ \qquad -\dfrac{1}{8}\sigma_{j-1/2}^2\Big((v_{j-1}^k+v_j^k)+(v_{j-1}^{k+1}+v_j^{k+1})\Big), \end{array}\right. \tag{5.4.17}$$

which holds for $(k,j)\in\mathbb{N}\times\{1,\cdots,N\}$, with the boundary conditions (5.4.15) and initial data (5.4.16).

To apply Theorem 5.3.3, we need to estimate the norm of the matrix $A_{\triangle x}$ defined in (5.4.10). Actually, its spectrum is given in [5]: The eigenvalues of $A_{\triangle x}$ are

$$\lambda_{\pm l,\triangle x}=\pm\frac{2i}{\triangle x}\tan\Big(l\triangle x\frac{\pi}{2}\Big),\quad l\in\{1,\cdots,N\}.$$

Since $A_{\triangle x}$ is skew-adjoint on $X_{\triangle x}$, its operator norm is given by its highest eigenvalue:

$$\|A_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x},X_{\triangle x})}=\frac{2}{\triangle x}\tan\Big((1-\triangle x)\frac{\pi}{2}\Big)\underset{\triangle x\to 0}{\simeq}\frac{4}{\pi(\triangle x)^2}.$$

As a consequence of Theorem 5.3.3, we get:

Theorem 5.4.3. *The energy*

$$E_{\triangle x}^k=\frac{\triangle x}{2}\sum_{j=0}^{N}\Big(\Big|\frac{v_j^k+v_{j+1}^k}{2}\Big|^2+\Big|\frac{u_{j+1}^k-u_j^k}{\triangle x}\Big|^2\Big),\quad k\in\mathbb{N},$$

of solutions of (5.4.17) *is exponentially decaying, uniformly with respect to* $\triangle t>0$ *and* $\triangle x>0$, *in the sense of* (5.3.2) *provided there exists a constant* η *such that*

$$\triangle t\le\eta(\triangle x)^2. \tag{5.4.18}$$

Remark 5.4.4. In this case, the CFL condition (5.4.18) is very restrictive for practical computations. Therefore, in practice, the fully discrete scheme (5.4.14) that involves a numerical viscosity term, for which no CFL condition is needed, seems preferable.

5.4.3 A fully discrete damped wave equation: A viscous finite difference approximation

We now describe how our results may be combined with those of [27, 23], which add numerical viscosity in the discretization with respect to the space-variable, to derive a uniformly exponentially stable fully discrete scheme.

The finite difference space semi-discrete approximation scheme of system (5.4.1) is as follows

$$\begin{cases} \ddot{u}_j - \dfrac{u_{j+1} - 2u_j + u_{j-1}}{(\triangle x)^2} + \sigma_j^2 \dot{u}_j = 0, & t \in \mathbb{R}_+,\ j \in \{1, \cdots, N\}, \\ u_0(t) = u_{N+1}(t) = 0, & t \in \mathbb{R}_+, \\ u_j(0) = u_{j,0}, \quad \dot{u}_j(0) = v_{j,0}, & j \in \{1, \cdots, N\}, \end{cases} \tag{5.4.19}$$

where σ_j, $u_{j,0}$, $v_{j,0}$ and u_j are, respectively, approximations of the functions σ, u_0, v_0 at the point x_j.

The energy of system (5.4.19), given by

$$E_{\triangle x}(t) = \frac{\triangle x}{2} \sum_{j=0}^{N} \Big(|\dot{u}_j(t)|^2 + \Big| \frac{\dot{u}_{j+1}(t) - \dot{u}_j(t)}{\triangle x} \Big|^2 \Big), \tag{5.4.20}$$

is dissipated according to the law

$$\frac{dE_{\triangle x}}{dt}(t) = -\triangle x \sum_{j=1}^{N} \sigma_j^2 |\dot{u}_j(t)|^2.$$

However, due to spurious high frequency solutions that are created by the numerical scheme, the energies $E_{\triangle x}$ do not decay exponentially uniformly with respect to $\triangle x$ (see [18, 27]), except in the particular case $\omega = (0, 1)$: If $\omega \neq (0, 1)$, there are no positive constants μ and ν such that the inequality

$$E_{\triangle x}(t) \leq \mu\, E_{\triangle x}(0) \exp(-\nu t), \quad t \geq 0, \tag{5.4.21}$$

holds for any $\triangle x > 0$ and for any solution of (5.4.19).

Therefore, to get a uniform decay rate of the energies $E_{\triangle x}$ (with respect to $\triangle x > 0$), an extra numerical viscosity term was added in [27]:

$$\begin{cases} \ddot{u}_j - \dfrac{u_{j+1} - 2u_j + u_{j-1}}{(\triangle x)^2} + \sigma_j^2 \partial_t u_j \\ \qquad\qquad -(\triangle x)^2 \Big(\dfrac{\dot{u}_{j+1} - 2\dot{u}_j + \dot{u}_{j-1}}{(\triangle x)^2} \Big) = 0, & t \in \mathbb{R}_+,\ j \in \{1, \cdots, N\}, \\ u_0(t) = u_{N+1}(t) = 0, & t \in \mathbb{R}_+, \\ u_j(0) = u_{j,0}, \quad u_j'(0) = v_{j,0}, & j \in \{1, \cdots, N\}. \end{cases} \tag{5.4.22}$$

For this system, the energy, still defined by (5.4.20), is now dissipated according to the law:

$$\frac{dE_{\triangle x}}{dt}(t) = -\triangle x \sum_{j=1}^{N} \sigma_j^2 |\dot{u}_j(t)|^2 - (\triangle x)^3 \sum_{j=0}^{N} \Big(\frac{u_{j+1}(t) - u_j(t)}{\triangle x} \Big)^2.$$

It was proved in [27] that, if σ satisfies (5.4.2), the energy of the solutions of (5.4.22) is exponentially stable uniformly with respect to the mesh size $\triangle x > 0$, in the sense that there exist positive constants μ and ν such that (5.4.21) holds for any $\triangle x > 0$ and for any solution of (5.4.22).

Besides, one can check that system (5.4.22) can be written as

$$\ddot{U}_{\triangle x} + A_{0,\triangle x} U_{\triangle x} + B_{0,\triangle x}^* B_{0,\triangle x} \dot{U}_{\triangle x} + (\triangle x)^2 A_{0,\triangle x} \dot{U}_{\triangle x} = 0, \quad t \in \mathbb{R}_+, \tag{5.4.23}$$

where $U_{\triangle x} = (u_1, \cdots, u_j, \cdots, u_N)^*$, $A_{0,\triangle x}$ is a positive definite matrix, which represents the discrete Laplace operator, and $B_{0,\triangle x}$ is the $N \times N$ matrix defined by:

$$B_{0,\triangle x} = \begin{pmatrix} & \operatorname{diag}(\sigma_j) & \end{pmatrix}.$$

Exponential decay for the time continuous system (5.4.23)

In this Subsection, we indicate how one can prove the uniform exponential decay result for solutions of (5.4.23) using the combination of the results in [18] and the methods introduced in [14] and further developed in Corollary 5.3.8.

Let us first recall the results in [14]. Let H be a Hilbert space endowed with the norm $\|\cdot\|_H$. Let $A_0 : \mathcal{D}(A_0) \to H$ be a self-adjoint positive operator with compact resolvent and $B \in \mathfrak{L}(H, Y)$.

We then consider the initial value problem

$$\begin{cases} \ddot{u} + A_0 u + \varepsilon A_0 \dot{u} + B^* B \dot{u} = 0, \quad t \geq 0, \\ u(0) = u_0 \in \mathcal{D}(A_0^{1/2}), \quad \dot{u}(0) = u_1 \in H. \end{cases} \tag{5.4.24}$$

The energy of solutions of (5.4.24) is given by

$$E(t) = \frac{1}{2} \|\dot{u}(t)\|_H^2 + \frac{1}{2} \left\| A_0^{1/2} u(t) \right\|_H^2, \tag{5.4.25}$$

and satisfies

$$\frac{dE}{dt}(t) = - \|B\dot{u}(t)\|_Y^2 - \varepsilon \left\| A_0^{1/2} \dot{u}(t) \right\|_H^2. \tag{5.4.26}$$

Theorem 5.4.5. *Assume that system* (5.4.24) *with* $\varepsilon = 0$ *is exponentially stable and satisfies* (5.1.4) *for some positive constants* μ *and* ν*, and that* $B \in \mathfrak{L}(H, Y)$.

Then there exist two positive constants μ_0 *and* ν_0 *depending only on* $\|B\|_{\mathfrak{L}(H,Y)}$, ν *and* μ *such that any solution of* (5.4.24) *satisfies* (5.1.4) *with constants* μ_0 *and* ν_0 *uniformly with respect to the viscosity parameter* $\varepsilon \in [0, 1]$.

We now introduce the spectrum of A_0. Since A_0 is self-adjoint positive definite with compact resolvent, its spectrum is discrete and $\sigma(A_0) = \{\lambda_j^2 : \ j \in \mathbb{N}\}$, where λ_j is an increasing sequence of real positive numbers such that $\lambda_j \to \infty$ when $j \to \infty$. Set $(\Psi_j)_{j\in\mathbb{N}}$ an orthonormal basis of eigenvectors of A_0 associated to the eigenvalues $(\lambda_j^2)_{j\in\mathbb{N}}$.

For convenience, similarly as in (5.2.2), we define

$$\mathfrak{C}_s = \text{span}\,\{\Psi_j : \text{ the corresponding } \lambda_j \text{ satisfies } |\lambda_j| \leq s\}. \tag{5.4.27}$$

We claim that the proof of Theorem 5.4.5 in [14] also proves the following Theorem:

Theorem 5.4.6. *Let* $\varepsilon \in (0, 1]$. *Assume that system*

$$\ddot{u} + A_0 u = 0, \quad t \geq 0, \qquad u(0) = u_0 \in \mathcal{D}(A_0^{1/2}), \quad \dot{u}(0) = u_1 \in H. \tag{5.4.28}$$

is exactly observable within the class $\mathfrak{C}_{1/\sqrt{\varepsilon}}$ *in the following sense: there exist a time* $T^* > 0$ *and a positive constant* $k_* > 0$ *such that any solution* u *of* (5.4.28) *with initial data* $(u_0, u_1) \in \mathfrak{C}^2_{1/\sqrt{\varepsilon}}$ *satisfies*

$$k_* \Big(\left\| A_0^{1/2} u_0 \right\|_H^2 + \|u_1\|_H^2 \Big) \leq \int_0^{T^*} \|B\dot{u}(t)\|_Y^2 \ dt.$$

Then there exist two positive constants μ *and* ν *depending only on* $\|B\|_{\mathfrak{L}(H,Y)}$, T^* *and* k_* *such that any solution of* (5.4.24) *satisfies* (5.1.4).

In [18], it has been proved that there exist positive constants T^* and k_* such that for all $\triangle x > 0$, the solution of

$$\ddot{U}_{\triangle x} + A_{0,\triangle x} U_{\triangle x} = 0, \quad t \geq 0, \tag{5.4.29}$$

with initial data $(U_{0,\triangle x}, U_{1,\triangle x}) \in \mathfrak{C}_{1/\triangle x}(A_{\triangle x})^2$ satisfies

$$k_* \Big(\left\| A_{0,\triangle x}^{1/2} U_{0,\triangle x} \right\|_{*\triangle x}^2 + \|U_{1,\triangle x}\|_{*\triangle x}^2 \Big) \leq \int_0^{T^*} \left\| B_{\triangle x} \dot{U}_{\triangle x}(t) \right\|_{*\triangle x}^2 \, dt.$$

Setting $X_{*\triangle x} = \mathbb{R}^N$ endowed with the norm $\|\cdot\|_{*\triangle x}$, one easily checks that $\|B_{\triangle x}\|_{\mathfrak{L}(X_{*\triangle x}, X_{*\triangle x})}$ is bounded uniformly in $\triangle x > 0$.

Theorem 5.4.6 then applies, and proves that systems (5.4.23) are exponentially stable uniformly with respect to $\triangle x > 0$.

Remark 5.4.7. Note that this method also applies in higher dimension, using for instance the results in [31] which state uniform observability properties for finite difference approximation schemes of a 2d wave equation. Doing this, we recover the results in [27] in 2d.

We now go on analyzing (5.4.22). We rewrite system (5.4.22) as

$$\dot{Z}_{\triangle x} = A_{\triangle x} Z_{\triangle x} - B_{\triangle x}^* B_{\triangle x} Z_{\triangle x}, \quad t \in \mathbb{R}_+, \tag{5.4.30}$$

where

$$\begin{aligned} Z_{\triangle x} &= \begin{pmatrix} U_{\triangle x} \\ V_{\triangle x} \end{pmatrix}, \quad A_{\triangle x} = \begin{pmatrix} 0 & Id \\ -A_{0,\triangle x} & 0 \end{pmatrix}, \\ B_{\triangle x} &= \begin{pmatrix} 0 & \sqrt{B_{0,\triangle x}^* B_{0,\triangle x} + (\triangle x)^2 A_{0,\triangle x}} \end{pmatrix}. \end{aligned} \tag{5.4.31}$$

One can check that the operator $A_{\triangle x}$ is skew-adjoint on the vector space $X_{\triangle x} = \mathbb{R}^{2N}$ endowed with the norm $\|\cdot\|_{\triangle x}$:

$$\left\| \begin{pmatrix} U_{\triangle x} \\ V_{\triangle x} \end{pmatrix} \right\|_{\triangle x}^2 = \triangle x \sum_{j=0}^{N} \Big(|v_j|^2 + \Big| \frac{u_{j+1} - u_j}{\triangle x} \Big|^2 \Big), \tag{5.4.32}$$

where $U_{\triangle x} = (u_1, \cdots, u_j, \cdots, u_N)^*$ and $V_{\triangle x} = (v_1, \cdots, v_j, \cdots, v_N)^*$, with the convention $u_0 = u_{N+1} = 0$.

Note that the original energy (5.4.20) of system (5.4.22) coincides with the quantity $\|Z_{\triangle x}\|_{\triangle x}^2 / 2$ of solutions of (5.4.30), with the notation above.

We then need to check that the operator $B_{\triangle x}$ is a bounded map from $X_{\triangle x}$ to $X_{*\triangle x}^2 = \mathbb{R}^{2N}$, where $X_{*\triangle x} = \mathbb{R}^N$ is endowed with the classical discrete L^2 norm $\|\cdot\|_{*\triangle x}$ given in (5.4.11). Since σ is assumed to be in $L^\infty(0,1)$, we obviously have

$$\|\mathrm{diag}(\sigma_j) V_{\triangle x}\|_{*\triangle x} \leq \|\sigma\|_{L^\infty} \|V_{\triangle x}\|_{*\triangle x}.$$

Besides,

$$\left\| (\triangle x)^2 A_{0,\triangle x} V_{\triangle x} \right\|_{*\triangle x} \leq 4 \|V_{\triangle x}\|_{*\triangle x},$$

since

$$(\triangle x)^2 A_{0,\triangle x} V_{\triangle x} = W_{\triangle x}, \quad \text{with } w_j = v_{j+1} - 2v_j + v_{j-1}, \quad \forall j \in \{1, \cdots, N\}.$$

Combining these last inequalities, we get the uniform bound

$$\|B_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x},X^2_{*\triangle x})} \leq 2 + \|\sigma\|_{L^\infty}\,.$$

We are therefore in the setting of Section 5.3: We checked hypothesis (A1) and (A2) of Definition 5.3.1 for the operators $A_{\triangle x}$ and $B_{\triangle x}$, and (A3) comes from the results of [27].

We now present the applications of the abstract methods in Section 5.3 to this particular setting.

Method I: Adding a numerical viscosity term in time

We introduce the fully discrete approximation scheme, corresponding to (5.1.9), given by

$$\begin{cases} \dfrac{\tilde u_j^{k+1}-u_j^k}{\triangle t} = \dfrac{v_j^k+\tilde v_j^{k+1}}{2}, \\ \dfrac{\tilde v_j^{k+1}-v_j^k}{\triangle t} = \dfrac{1}{2(\triangle x)^2}\big(\tilde u_{j+1}^{k+1}+u_{j+1}^k-2\tilde u_j^{k+1}-2u_j^k+\tilde u_{j-1}^{k+1}+u_{j-1}^k\big) \\ \qquad\qquad -\dfrac{1}{2}\sigma_j^2(v_j^k+\tilde v_j^{k+1})+\dfrac{1}{2}\big(\tilde v_{j+1}^{k+1}+v_{j+1}^k-2\tilde v_j^{k+1}-2v_j^k+\tilde v_{j-1}^{k+1}+v_{j-1}^k\big), \\ \dfrac{u_j^{k+1}-\tilde u_j^{k+1}}{\triangle t} = \Big(\dfrac{\triangle t}{\triangle x}\Big)^2\big(u_{j+1}^{k+1}-2u_j^{k+1}+u_{j-1}^{k+1}\big), \\ \dfrac{v_j^{k+1}-\tilde v_j^{k+1}}{\triangle t} = \Big(\dfrac{\triangle t}{\triangle x}\Big)^2\big(v_{j+1}^{k+1}-2v_j^{k+1}+v_{j-1}^{k+1}\big), \end{cases} \tag{5.4.33}$$

which holds for $(k,j)\in\mathbb{N}\times\{1,\cdots,N\}$, with the boundary conditions (5.4.15) and the initial data (5.4.16). Here again, u_j^k and v_j^k respectively denote approximations of the functions u and $\dot u$ in $x_j = j\triangle x$ at time $k\triangle t$.

This fully discrete approximation scheme coincides with the system (5.1.9) with $A = A_{\triangle x}$ and $B = B_{\triangle x}$.

Applying Theorem 5.1.1, we get:

Theorem 5.4.8. *The energy*

$$E^k_{\triangle x} = \frac{\triangle x}{2}\sum_{j=0}^{N}\Big(|v_j^k|^2+\Big|\frac{u_{j+1}^k-u_j^k}{\triangle x}\Big|^2\Big) \tag{5.4.34}$$

of solutions of system (5.4.33) *is exponentially decaying, uniformly with respect to both parameters* $\triangle x>0$ *and* $\triangle t>0$. *To be more precise, there exist positive constants* ν_0 *and* μ_0 *such that the energies of solutions* (5.4.33) *satisfy* (5.3.2).

Note that in Theorem 5.4.8, no CFL condition is required.

Method II: Imposing a CFL condition

Again, we consider the space semi-discrete approximation (5.4.22) (or equivalently (5.4.30)) of (5.4.1), that we now discretize in time using the midpoint scheme (5.1.5): For all $(k,j) \in \mathbb{N} \times \{1, \cdots, N\}$,

$$\left\{ \begin{array}{ll} \dfrac{u_j^{k+1}-u_j^k}{\triangle t} & = \dfrac{v_j^k+v_j^{k+1}}{2}, \\ \dfrac{v_j^{k+1}-v_j^k}{\triangle t} & = \dfrac{1}{2(\triangle x)^2}\big(u_{j+1}^{k+1}+u_{j+1}^k-2u_j^{k+1}-2u_j^k+u_{j-1}^{k+1}+u_{j-1}^k\big) \\ & \quad -\dfrac{1}{2}\sigma_j^2(v_j^k+v_j^{k+1})+\dfrac{1}{2}\big(v_{j+1}^{k+1}+v_{j+1}^k-2v_j^{k+1}-2v_j^k+v_{j-1}^{k+1}+v_{j-1}^k\big), \end{array} \right. \tag{5.4.35}$$

with the boundary conditions (5.4.15), and initial data (5.4.16).

The discrete energies are defined by (5.4.34) as before. Note that this scheme is simpler than (5.4.33), since it does not contain numerical viscosity terms in time.

To use Theorem 5.3.3, we need to estimate the norm $\|A_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x},X_{\triangle x})}$.

Actually, if

$$Z_{1\triangle x}=\begin{pmatrix} U_{1\triangle x} \\ V_{1\triangle x}\end{pmatrix}, \quad Z_{2\triangle x}=\begin{pmatrix} U_{2\triangle x} \\ V_{2\triangle x}\end{pmatrix},$$

then

$$< Z_{1\triangle x}, A_{\triangle x}Z_{2\triangle x} >_{\triangle x} = \triangle x \sum_{j=0}^{N}\Big(\frac{u_{1\triangle x,j+1}-u_{1\triangle x,j}}{\triangle x}\Big)\Big(\frac{v_{2\triangle x,j+1}-v_{2\triangle x,j}}{\triangle x}\Big) - \triangle x\sum_{j=1}^{N} v_{1\triangle x,j}\Big(\frac{u_{2\triangle x,j+1}-2u_{2\triangle x,j}+u_{2\triangle x,j-1}}{(\triangle x)^2}\Big).$$

In particular,

$$\begin{aligned}(\triangle x)^2\Big| < Z_{1\triangle x}, A_{\triangle x}Z_{2\triangle x} >_{\triangle x}\Big|^2 &\le \Big(\triangle x\sum_{j=0}^{N}\Big(\frac{u_{1\triangle x,j+1}-u_{1\triangle x,j}}{\triangle x}\Big)^2\Big)\Big(\triangle x\sum_{j=0}^{N}\Big(v_{2\triangle x,j+1}-v_{2\triangle x,j}\Big)^2\Big) \\ &\quad + \Big(\triangle x\sum_{j=1}^{N}|v_{1\triangle x,j}|^2\Big)\Big(\triangle x\sum_{j=0}^{N}\Big(\frac{u_{2\triangle x,j+1}-u_{2\triangle x,j}}{\triangle x}-\frac{u_{2\triangle x,j}-u_{2\triangle x,j-1}}{\triangle x}\Big)^2\Big),\end{aligned}$$

that gives

$$\Big| < Z_{1\triangle x}, A_{\triangle x}Z_{2\triangle x} >_{\triangle x}\Big| \le \frac{2}{\triangle x}\|Z_{1\triangle x}\|_{\triangle x}\|Z_{2\triangle x}\|_{\triangle x}.$$

This proves that $\|A_{\triangle x}\|_{\mathfrak{L}(X_{\triangle x},X_{\triangle x})} \le 2/\triangle x$. Actually, in this case, we know the eigenvalues and eigenvectors explicitly (see for instance [18]), and therefore this norm can be computed explicitly to be $2\sin((1-\triangle x)\pi/2)/\triangle x$.

As a corollary of Theorem 5.3.3, we get:

Theorem 5.4.9. *Given $\eta > 0$, if we impose the CFL type condition*

$$\triangle t \leq \eta \triangle x, \tag{5.4.36}$$

then there exist positive constants ν_η and μ_η such that the energy of solutions of (5.4.35) *satisfies* (5.3.2)*, uniformly with respect to the discretization parameters $\triangle x > 0$ and $\triangle t > 0$.*

Remark 5.4.10. Here it seems more natural to use the discretization (5.4.35) than (5.4.33) since the CFL condition (5.4.36) is not very restrictive.

Note that the results we presented here for the 1d wave equation can be adapted to deal with 2d wave equations in a square as in [27] or more general domains as in [23].

Method III: Discretizing with only one viscosity term

We are in the setting of Theorem 5.3.7, and therefore we can use only one viscosity term: Set $\varepsilon = \max\{\triangle t, \triangle x\}$ and consider

$$\begin{cases} \dfrac{\tilde{u}_j^{k+1} - u_j^k}{\triangle t} = \dfrac{v_j^k + \tilde{v}_j^{k+1}}{2}, \\ \dfrac{\tilde{v}_j^{k+1} - v_j^k}{\triangle t} = \dfrac{1}{2(\triangle x)^2}\big(\tilde{u}_{j+1}^{k+1} + u_{j+1}^k - 2\tilde{u}_j^{k+1} - 2u_j^k + \tilde{u}_{j-1}^{k+1} + u_{j-1}^k\big) - \dfrac{1}{2}\sigma_j^2(v_j^k + \tilde{v}_j^{k+1}), \\ \dfrac{u_j^{k+1} - \tilde{u}_j^{k+1}}{\triangle t} = \Big(\dfrac{\varepsilon}{\triangle x}\Big)^2\big(u_{j+1}^{k+1} - 2u_j^{k+1} + u_{j-1}^{k+1}\big), \\ \dfrac{v_j^{k+1} - \tilde{v}_j^{k+1}}{\triangle t} = \Big(\dfrac{\varepsilon}{\triangle x}\Big)^2\big(v_{j+1}^{k+1} - 2v_j^{k+1} + v_{j-1}^{k+1}\big), \end{cases} \tag{5.4.37}$$

which holds for $(k, j) \in \mathbb{N} \times \{1, \cdots, N\}$, with the boundary conditions (5.4.15) and initial data (5.4.16).

Theorem 5.4.11. *Setting $\varepsilon = \max\{\triangle t, \triangle x\}$, the energy $E_{\triangle x}^k$ defined in* (5.4.34) *of solutions of system* (5.4.37) *is exponentially decaying, uniformly with respect to both parameters $\triangle x > 0$ and $\triangle t > 0$. To be more precise, there exist positive constants ν_0 and μ_0 such that the energy of solutions* (5.4.33) *satisfies* (5.3.2)*.*

Remark 5.4.12. The main advantage of (5.4.37) over (5.4.33) is the presence of only one viscosity operator. In other words, (5.4.33) dissipates too much.

The advantage of (5.4.37) over (5.4.35) consists in the absence of CFL condition, which makes (5.4.37) more robust in practice.

5.5 Further comments

1. As we mentioned in the introduction, our methods and results require the assumption that the damping operator B is bounded. This is due to the method we employ, which is based on the equivalence between the exponential decay of the energy and the observability properties of the conservative system, that requires the damping operator to be bounded. That is the case, even in the continuous setting. However, in several relevant applications, as for instance when dealing with the

problem of boundary stabilization of the wave equation (see [20]), the feedback law is unbounded, and our method does not apply. This issue requires further work.

2. Another drawback of our method is that it provides an explicit estimate of the exponential decay rate of the energy of the time semi-discrete approximation systems, which is far from sharp in general. Again, this also happens in the continuous case, since we deduce stabilization properties from the study of the observability properties of the corresponding conservative systems. In the continuous case, the computation of the decay rate of the energy is technically involved and requires to work directly on the damped system. We refer to the works [7, 8, 19] that deal with these questions for damped wave equations.

In our context, it would be also relevant to ask if one can choose the numerical viscosity term such that the time-discrete damped systems are exponentially stable, uniformly with respect to the time discretization parameter, and such that the decay rate of the energy of these time discrete systems coincides with the one of the continuous system. To our knowledge, this issue is still open. Let us mention the work [13], which gives a partial answer to this question for space semi-discrete approximation schemes of the 1d Perfectly Matched Layers equations, which correspond to a particular instance of damped wave equations.

3. In this article, we assumed exponential decay properties for the continuous damped systems under consideration. However, there are several important models of vibrations where the energy decay rate is polynomial or even logarithmic within the class of solutions with initial data in $\mathcal{D}(A)$ instead of X. That is the case for instance for networks of vibrating strings [9] or damped wave equations, when the damping operator is effective on a subdomain where the *Geometric Control Condition* is not fulfilled [2, 19]. One could ask if there is a systematic discretization method for these systems that preserves these decay properties. To our knowledge, this issue is widely open. The time semi-discrete schemes provided here are good candidates to preserve these decay properties.

4. The same questions arise when discretizing in time semilinear wave equations. For instance, in [10] (see also [29, 30]), the exponential decay property of solutions of semilinear wave equations in $\mathbb{R}^3$ with a damping term which is effective on the exterior of a ball are analyzed. Under suitable properties of the nonlinearity, it is proved that the exponential decay of the energy holds locally uniformly for finite energy solutions. It would be interesting to analyze whether the same exponential decay property holds, uniformly with respect to the time-step, for the numerical schemes analyzed in this article in this semilinear setting.

Bibliography

[1] H. T. Banks, K. Ito, and C. Wang. Exponentially stable approximations of weakly damped wave equations. In *Estimation and control of distributed parameter systems (Vorau, 1990)*, volume 100 of *Internat. Ser. Numer. Math.*, pages 1–33. Birkhäuser, Basel, 1991.

[2] C. Bardos, G. Lebeau, and J. Rauch. Sharp sufficient conditions for the observation, control and stabilization of waves from the boundary. *SIAM J. Control and Optimization*, 30(5):1024–1065, 1992.

[3] N. Burq and P. Gérard. Condition nécessaire et suffisante pour la contrôlabilité exacte des ondes. *C. R. Acad. Sci. Paris Sér. I Math.*, 325(7):749–752, 1997.

[4] N. Burq and M. Zworski. Geometric control in the presence of a black box. *J. Amer. Math. Soc.*, 17(2):443–471 (electronic), 2004.

[5] C. Castro and S. Micu. Boundary controllability of a linear semi-discrete 1-d wave equation derived from a mixed finite element method. *Numer. Math.*, 102(3):413–462, 2006.

[6] C. Castro, S. Micu, and A. Münch. Numerical approximation of the boundary control for the wave equation with mixed finite elements in a square. *IMA J. Numer. Anal.*, 28(1):186–214, 2008.

[7] S. Cox and E. Zuazua. The rate at which energy decays in a damped string. *Comm. Partial Differential Equations*, 19(1-2):213–243, 1994.

[8] S. Cox and E. Zuazua. The rate at which energy decays in a string damped at one end. *Indiana Univ. Math. J.*, 44(2):545–573, 1995.

[9] R. Dáger and E. Zuazua. *Wave propagation, observation and control in* 1*-d flexible multi-structures*, volume 50 of *Mathématiques & Applications (Berlin)*. Springer-Verlag, Berlin, 2006.

[10] B. Dehman, G. Lebeau, and E. Zuazua. Stabilization and control for the subcritical semilinear wave equation. *Ann. Sci. École Norm. Sup. (4)*, 36(4):525–551, 2003.

[11] S. Ervedoza. Observability of the mixed finite element method for the 1d wave equation on non-uniform meshes. *To appear in ESAIM: COCV*, 2008. *Cf Chapitre 2.*

[12] S. Ervedoza, C. Zheng, and E. Zuazua. On the observability of time-discrete conservative linear systems. *J. Funct. Anal.*, 254(12):3037–3078, June 2008. *Cf Chapitre 3.*

[13] S. Ervedoza and E. Zuazua. Perfectly matched layers in 1-d: Energy decay for continuous and semi-discrete waves. *Numer. Math.*, 109(4):597–634, 2008. *Cf Chapitre 1.*

[14] S. Ervedoza and E. Zuazua. Uniform exponential decay for viscous damped systems. *To appear in Proc. of Siena "Phase Space Analysis of PDEs 2007", Special issue in honor of Ferrucio Colombini*, 2008. *Cf Chapitre 4.*

[15] R. Glowinski. Ensuring well-posedness by analogy: Stokes problem and boundary control for the wave equation. *J. Comput. Phys.*, 103(2):189–221, 1992.

[16] A. Haraux. Une remarque sur la stabilisation de certains systèmes du deuxième ordre en temps. *Portugal. Math.*, 46(3):245–258, 1989.

[17] L. I. Ignat and E. Zuazua. Dispersive properties of a viscous numerical scheme for the Schrödinger equation. *C. R. Math. Acad. Sci. Paris*, 340(7):529–534, 2005.

[18] J.A. Infante and E. Zuazua. Boundary observability for the space semi discretizations of the 1-d wave equation. *Math. Model. Num. Ann.*, 33:407–438, 1999.

[19] G. Lebeau. Équations des ondes amorties. *Séminaire sur les Équations aux Dérivées Partielles, 1993–1994,École Polytech.*, 1994.

[20] J.-L. Lions. *Contrôlabilité exacte, Stabilisation et Perturbations de Systèmes Distribués. Tome 1. Contrôlabilité exacte*, volume RMA 8. Masson, 1988.

[21] F. Macià. The effect of group velocity in the numerical analysis of control problems for the wave equation. In *Mathematical and numerical aspects of wave propagation—WAVES 2003*, pages 195–200. Springer, Berlin, 2003.

[22] L. Miller. Controllability cost of conservative systems: resolvent condition and transmutation. *J. Funct. Anal.*, 218(2):425–444, 2005.

[23] A. Münch and A. F. Pazoto. Uniform stabilization of a viscous numerical approximation for a locally damped wave equation. *ESAIM Control Optim. Calc. Var.*, 13(2):265–293 (electronic), 2007.

[24] M. Negreanu and E. Zuazua. Convergence of a multigrid method for the controllability of a 1-d wave equation. *C. R. Math. Acad. Sci. Paris*, 338(5):413–418, 2004.

[25] K. Ramdani, T. Takahashi, and M. Tucsnak. Uniformly exponentially stable approximations for a class of second order evolution equations—application to LQR problems. *ESAIM Control Optim. Calc. Var.*, 13(3):503–527, 2007.

[26] L. R. Tcheugoué Tebou and E. Zuazua. Uniform boundary stabilization of the finite difference space discretization of the $1-d$ wave equation. *Adv. Comput. Math.*, 26(1-3):337–365, 2007.

[27] L.R. Tcheugoué Tébou and E. Zuazua. Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity. *Numer. Math.*, 95(3):563–598, 2003.

[28] X. Zhang, C. Zheng, and E. Zuazua. Exact controllability of the time discrete wave equation. *Discrete and Continuous Dynamical Systems*, 2007.

[29] E. Zuazua. Exponential decay for the semilinear wave equation with locally distributed damping. *Comm. Partial Differential Equations*, 15(2):205–235, 1990.

[30] E. Zuazua. Exponential decay for the semilinear wave equation with localized damping in unbounded domains. *J. Math. Pures Appl. (9)*, 70(4):513–529, 1991.

[31] E. Zuazua. Boundary observability for the finite-difference space semi-discretizations of the 2-D wave equation in the square. *J. Math. Pures Appl. (9)*, 78(5):523–563, 1999.

[32] E. Zuazua. Propagation, observation, and control of waves approximated by finite difference methods. *SIAM Rev.*, 47(2):197–243 (electronic), 2005.

Part III

Admissibility and Observability for finite element discretizations of conservative systems

Chapter 6

Schrödinger equations

Abstract: In this article, we derive uniform admissibility and observability properties for the finite element space semi-discretizations of $i\dot{z} = A_0 z$, where A_0 is an unbounded self-adjoint positive definite operator with compact resolvent. In order to address this problem, we present several spectral criteria for admissibility and observability of such systems, which will be used to derive several results for space semi-discretizations of $i\dot{z} = A_0 z$. Our approach provides very general results, which stand in any dimension and for any regular mesh (in the sense of finite elements). We also present applications to admissibility and observability for fully discrete approximation schemes, and to controllability and stabilization issues.

6.1 Introduction

Let X be a Hilbert space endowed with the norm $\|\cdot\|_X$ and let $A_0 : \mathcal{D}(A_0) \subset X \to X$ be an unbounded self-adjoint positive definite operator with compact resolvent. Let us consider the following abstract system:

$$i\dot{z}(t) = A_0 z(t), \quad t \in \mathbb{R}, \qquad z(0) = z_0 \in X. \tag{6.1.1}$$

Here and henceforth, a dot ($\dot{}$) denotes differentiation with respect to the time t. The element $z_0 \in X$ is called the *initial state*, and $z = z(t)$ is the *state* of the system. Such systems are often used as models for quantum dynamics (Schrödinger's equation).

Note that the system (6.1.1) is conservative: The energy $\|z(t)\|_X^2$ of solutions of (6.1.1) is constant.

Assume that Y is another Hilbert space endowed with the norm $\|\cdot\|_Y$. We denote by $\mathfrak{L}(X, Y)$ the space of bounded linear operators from X to Y, endowed with the classical operator norm. Let $B \in \mathfrak{L}(\mathcal{D}(A_0), Y)$ be an observation operator and define the output function

$$y(t) = Bz(t). \tag{6.1.2}$$

We assume that the operator $B \in \mathfrak{L}(\mathcal{D}(A_0), Y)$ is admissible for system (6.1.1) in the following sense:

Definition 6.1.1. The operator B is an admissible observation operator for system (6.1.1) if for every

$T > 0$ there exists a constant $K_T > 0$ such that

$$\int_0^T \|Bz(t)\|_Y^2 \, dt \leq K_T \|z_0\|_X^2 , \qquad \forall \, z_0 \in \mathcal{D}(A_0), \tag{6.1.3}$$

for every solutions of (6.1.1).

Note that if B is *bounded* in X, i.e. if it can be extended in such a way that $B \in \mathfrak{L}(X, Y)$, then B is obviously an admissible observation operator, and K_T can be chosen as $K_T = T \|B\|_{\mathfrak{L}(X,Y)}^2$. However, in applications, this is often not the case, and the admissibility condition is then a consequence of a suitable "hidden regularity" property of the solutions of the evolution equation (6.1.1).

The exact observability property for system (6.1.1)-(6.1.2) can be formulated as follows:

Definition 6.1.2. System (6.1.1)-(6.1.2) is exactly observable in time T if there exists $k_T > 0$ such that

$$k_T \|z_0\|_X^2 \leq \int_0^T \|Bz(t)\|_Y^2 \, dt, \qquad \forall \, z_0 \in \mathcal{D}(A_0). \tag{6.1.4}$$

for every solution of (6.1.1).

Moreover, system (6.1.1)-(6.1.2) is said to be exactly observable if it is exactly observable in some time $T > 0$.

Note that observability and admissibility issues arise naturally when dealing with controllability and stabilization properties of linear systems (see for instance the textbook [28]). These links will be made precise later.

There is an extensive literature providing observability results for Schrödinger equations, by several different methods including microlocal analysis [3, 26], multipliers and Fourier series [30], etc. Our goal in this paper is to develop a theory allowing to get admissibility and observability results for space semi-discrete systems as a direct consequence of those corresponding to the continuous ones, thus avoiding technical developments in the discrete settings.

Let us now introduce the finite element method for (6.1.1).

Consider $(V_h)_{h>0}$ a sequence of vector spaces of finite dimension n_h which embed into X via a linear injective map $\pi_h : V_h \to X$. For each $h > 0$, the inner product $< \cdot, \cdot >_X$ in X induces a structure of Hilbert space for V_h endowed by the scalar product $< \cdot, \cdot >_h = < \pi_h \cdot, \pi_h \cdot >_X$.

We assume that, for each $h > 0$, the vector space $\pi_h(V_h)$ is a subspace of $\mathcal{D}(A_0^{1/2})$. We thus define the linear operator $A_{0h} : V_h \to V_h$ by

$$< A_{0h}\phi_h, \psi_h >_h = < A_0^{1/2}\pi_h\phi_h, A_0^{1/2}\pi_h\psi_h >_X, \quad \forall (\phi_h, \psi_h) \in V_h^2. \tag{6.1.5}$$

The operator A_{0h} defined in (6.1.5) obviously is self-adjoint and positive definite. If we introduce the adjoint π_h^* of π_h, definition (6.1.5) reads as:

$$A_{0h} = \pi_h^* A_0 \pi_h. \tag{6.1.6}$$

This operator A_{0h} corresponds to the finite element discretization of the operator A_0. We thus consider the following space semi-discretisation of (6.1.1):

$$i\dot{z}_h = A_{0h} z_h, \quad t \in \mathbb{R}, \qquad z_h(0) = z_{0h} \in V_h. \tag{6.1.7}$$

In this context, for all $h > 0$, the observation operator naturally becomes $B_h = B\pi_h$. Note that, when $B \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$, this definition always make sense. We are thus lead to impose $B \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$.

We now make precise the assumptions we have, usually, on π_h, and which will be needed in our analysis. One easily checks that

$$\pi_h^* \pi_h = Id_{V_h}. \tag{6.1.8}$$

The injection π_h describes the finite element approximation we have chosen. Especially, the vector space $\pi_h(V_h)$ approximates, in the sense given hereafter, the space $\mathcal{D}(A_0^{1/2})$: There exist $\theta > 0$ and $C_0 > 0$, such that for all $h > 0$,

$$\begin{cases} \left\| A_0^{1/2}(\pi_h \pi_h^* - I)\phi \right\|_X \leq C_0 \left\| A_0^{1/2}\phi \right\|_X, \quad \forall \phi \in \mathcal{D}(A_0^{1/2}), \\ \left\| A_0^{1/2}(\pi_h \pi_h^* - I)\phi \right\|_X \leq C_0 h^\theta \left\| A_0 \phi \right\|_X, \quad \forall \phi \in \mathcal{D}(A_0). \end{cases} \tag{6.1.9}$$

Note that in many applications, and in particular for A_0 the Laplace operator on a bounded domain with Dirichlet boundary conditions, estimates (6.1.9) are satisfied for $\theta = 1$.

We will not discuss convergence results for the numerical approximation schemes presented here, which are classical under assumption (6.1.9), and which can be found for instance in the textbook [39].

In the sequel, our goal is to obtain uniform observability properties for (6.1.7) similar to (6.1.4).

Let us mention that similar questions have already been investigated in [27] for the finite difference approximation schemes of the beam equation, for which we expect the same admissibility and observability properties as for (6.1.7) to hold. To be more precise, in [27], the authors considered the finite-difference approximation scheme of the 1d beam equation on a uniform mesh, observed through the boundary value. They proved that, in this case, the observability properties do not hold uniformly in the space discretization parameter for any initial data. Though, they proved, similarly as in [23] which dealt with 1d finite difference schemes of the wave equation, that one can recover uniform observability results when filtering the data. Actually, as pointed out by Otared Kavian in [46], it may even happen that unique continuation properties do not hold anymore in the discrete setting due to the existence of localized high frequency solutions.

Therefore, it is natural to restrict ourselves to classes of suitable filtered initial data. For all $h > 0$, since A_{0h} is a self-adjoint positive definite matrix, the spectrum of A_{0h} is given by a sequence of positive eigenvalues

$$0 < \lambda_1^h \leq \lambda_2^h \leq \cdots \leq \lambda_{n_h}^h, \tag{6.1.10}$$

and normalized (in V_h) eigenvectors $(\Phi_j^h)_{1 \leq j \leq n_h}$. For any $s > 0$, we can now define, for any $h > 0$, the filtered space

$$\mathcal{C}_h(s) = \text{span}\left\{ \Phi_j^h \text{ such that the corresponding eigenvalue satisfies } |\lambda_j^h| \leq s \right\}.$$

We are now in position to state the main results of this article:

Theorem 6.1.3. *Let A_0 be a self-adjoint positive definite operator with compact resolvent, and $B \in \mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$, with $\kappa < 1/2$. Assume that the maps $(\pi_h)_{h>0}$ satisfy property* (6.1.9)*. Set*

$$\sigma = \theta \min\left\{ 2(1 - 2\kappa), \frac{2}{5} \right\}. \tag{6.1.11}$$

Admissibility: *Assume that system* (6.1.1)-(6.1.2) *is admissible.*

Then, for any $\eta > 0$ *and* $T > 0$*, there exists a positive constant* $K_{T,\eta} > 0$ *such that, for any* $h > 0$*, any solution of* (6.1.7) *with initial data*

$$z_{0h} \in \mathcal{C}_h(\eta/h^\sigma) \tag{6.1.12}$$

satisfies

$$\int_0^T \|B_h z_h(t)\|_Y^2 \ dt \leq K_{T,\eta} \|z_{0h}\|_h^2 . \tag{6.1.13}$$

Observability: *Assume that system* (6.1.1)-(6.1.2) *is admissible and exactly observable.*

Then there exist $\epsilon > 0$*, a time* T^* *and a positive constant* $k_* > 0$ *such that, for any* $h > 0$*, any solution of* (6.1.7) *with initial data*

$$z_{0h} \in \mathcal{C}_h(\epsilon/h^\sigma) \tag{6.1.14}$$

satisfies

$$k_* \|z_{0h}\|_h^2 \leq \int_0^{T^*} \|B_h z_h(t)\|_Y^2 \ dt. \tag{6.1.15}$$

This theorem is based on new spectral characterizations of admissibility and exact observability for (6.1.1)-(6.1.2).

For characterizing the admissibility property, we use the results in [12] to obtain a characterization based on a resolvent estimate and, later, on an interpolation property.

Our characterization of the exact observability property uses the resolvent estimates in [6, 32]. Again, we prove that these estimates can be interpreted as interpolation properties.

The main idea, then, consists in proving uniform (in h) interpolation properties for the operators A_{0h} and B_h, in order to recover uniform (in h) admissibility and observability estimates. This idea is completely natural since the operators A_{0h} and B_h correspond to discrete versions of A_0 and B, respectively.

Theorem 6.1.3 has several important applications. As a straightforward corollary of the results in [12], one can thus derive observability properties for general fully discrete approximation schemes based on (6.1.7). Precise statements will be given in Section 6.5.

Besides, it also has relevant applications in control theory. Indeed, it implies that the Hilbert Uniqueness Method (see [28]) can be adapted in the discrete setting to provide efficient algorithms to compute approximations of exact controls for the continuous systems. This will be clarified in Section 6.6.

We will also present consequences of Theorem 6.1.3 to stabilization issues for space semi-discrete and fully discrete models based on (6.1.7), using the results [15]. Indeed, in [15], this problem has been addressed in a very general setting which includes our models.

Let us briefly comment some relative works. Similar problems have been extensively studied in the last decade for various space semi-discretizations of the 1d wave equation, see for instance the review article [46] and the references therein. The numerical schemes on uniform meshes provided by finite difference and finite element methods do not have uniform observability properties, whatever the time

T is, see [23] (see also [27] for the beam equation). This is due to high frequency waves which do not propagate, see [43, 31]. In other words, these numerical schemes create some spurious high-frequency wave solutions which do not travel.

In this context, filtering techniques have been extensively developed. It has been proved in [23, 44] (or [27] for the beam equation) that filtering the initial data removes these spurious waves, and make possible uniform observability properties to hold. Other ways to filter these spurious waves exist, for instance using wavelet filtering approaches as in [35] or bi-grids techniques [16, 36]. However, to the best of our knowledge, these methods have been analyzed only for uniform grids in small dimensions (namely in 1d or 2d). Also note that these results prove uniform observability properties for larger classes of initial data than the ones stated here, but in more particular cases. Especially, we emphasize that Theorem 6.1.3 holds in any dimension and for any regular mesh.

Let us also mention that observability properties are equivalent to stabilization properties (see [19]), at least when the observation operator is bounded. Therefore, observability properties can be deduced from the literature in stabilization theory. Especially, we refer to the works [41, 40, 34, 13], which prove uniform exponential decay results for damped space semi-discrete wave equations in 1d and 2d, discretized on uniform meshes using finite difference methods, in which a numerical viscosity term has been added. Again, these results are better than the ones derived here, but apply in the more restrictive context of 1d or 2d wave equations on uniform meshes. Similar results have also been proved in [38] in a general context close to ours, but for bounded observation operators. Besides, in [38], a non trivial spectral condition on A_0 is needed, which reduces the scope of applications mainly to 1d equations.

To the best of our knowledge, there are very few papers dealing with nonuniform meshes. A first step in this direction can be found in the context of the stabilization of the 1d wave equation in [38]: Indeed, stabilization properties are equivalent (see [19]) to observability properties for the corresponding conservative systems. The results in [38] can therefore be applied to 1d wave equation on nonuniform meshes to derive uniform observability results within the class of data filtered at the scale $h^{-\theta}$. Though, they strongly use a spectral gap condition on the eigenvalues of the operator, which do not hold for the wave equation in higher dimension. Another result in this direction is presented in [11], again in the context of the 1d wave equation, but discretized using a mixed finite element method as in [2, 7, 8]. In [11], it is proved that observability properties for schemes derived from a mixed finite element method hold uniformly with respect to the mesh size for a large class of meshes, and, in particular, no filtering condition is required on the data.

We shall also mention recent works on spectral characterizations of the exact observability properties for abstract conservative systems. We refer to [6, 32] for a very general approach for linear conservative systems, which yields a necessary and sufficient spectral condition for exact observability to hold. Let us also mention the article [37], in which a spectral characterization of observability properties based on wave packets is given. We also point out the recent article [4], which considers several (weak) observability properties given as interpolation properties, which are close to the ones that we will prove in the present work.

We also mention the recent work [12] which proved admissibility and observability estimates for general time semi-discrete conservative linear systems. In [12], a very general approach is given, which allows to deal with a large class of time-discrete approximation schemes. This approach is based, as here, on a spectral characterization of exact observability for conservative linear systems (namely the one in [6, 32]). Later on in [15] (see also [14]), the stabilization properties of time discrete approximation schemes of damped systems were studied. In particular, [15] introduces time-discretizations which are guaranteed to enjoy uniform stabilization properties.

Let us finally notice that the results in Theorem 6.1.3 may not be sharp, in view of the results in [27], which can be adapted to the finite element space semi-discretization of the 1d Schrödinger equation to prove that the sharp filtering scale, in 1d and on uniform meshes, is h^{-2}. In the very general setting presented here, we do not have any conjecture on the sharp filtering scale. This question deserves further work.

This article is organized as follows:

In Section 6.2, we present several spectral conditions which are equivalent to admissibility and exact observability properties for abstract systems taking the form (6.1.1)-(6.1.2). In Section 6.3, we prove Theorem 6.1.3. In Section 6.4, we provide some examples of applications of Theorem 6.1.3. In Section 6.5, we consider admissibility and exact observability properties for fully discrete approximation schemes of (6.1.7). In Section 6.6, some applications of Theorem 6.1.3 in controllability theory are indicated. In Section 6.7, we also present applications to stabilization theory. We finally present some further comments and open questions.

6.2 Spectral methods

This section recalls and presents various spectral characterizations of admissibility and observability for abstract systems such as (6.1.1)-(6.1.2). Here, we do not deal with the discrete approximation schemes (6.1.7).

To state our results properly, we introduce some notations.

When dealing with the abstract system (6.1.1)-(6.1.2), it is convenient to introduce the spectrum of the operator A_0. Since A_0 is self-adjoint and positive definite, its spectrum is given by a sequence of positive eigenvalues

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq \cdots \to \infty, \tag{6.2.1}$$

and normalized (in X) eigenvectors $(\Phi_j)_{j\in\mathbb{N}^*}$.

Since some of the results below extend to a larger class of systems than (6.1.1), we introduce the following abstract system

$$\begin{cases} \dot{z} = Az, \quad t \geq 0, \\ z(0) = z_0 \in X, \end{cases} \qquad y(t) = Bz(t), \tag{6.2.2}$$

where $A : \mathcal{D}(A) \subset X \to X$ is an unbounded skew-adjoint operator with compact resolvent. In particular, its spectrum is given by a sequence $(i\mu_j)_j$, where the constants μ_j are real and $|\mu_j| \to \infty$ when $j \to \infty$, and the corresponding eigenvectors $(\Psi_j)_j$ (normalized in X) constitute an orthonormal basis of X. Note that systems of the form (6.1.1)-(6.1.2) indeed are particular instances of (6.2.2).

This section is organized as follows.

First, we present spectral characterizations for the admissibility of systems (6.1.1)-(6.1.2), based on the results in [12], which we recall. Then, we present spectral characterizations for the exact observability of systems (6.1.1)-(6.1.2), based on the articles [6, 32].

6.2.1 Characterizations of admissibility

Wave packet characterization

First, we consider the general abstract conservative equation (6.2.2), and recall the results in [12, Section 6]. Note that the admissibility inequality for (6.2.2) consists in the existence, for any $T > 0$, of a positive constant K_T such that any solution z of (6.2.2) satisfies

$$\int_0^T \|Bz(t)\|_Y^2 \ dt \leq K_T \|z_0\|_X^2, \quad \forall z_0 \in \mathcal{D}(A). \tag{6.2.3}$$

Theorem 6.2.1 ([12])**.** *Let A be a skew-adjoint unbounded operator on X with compact resolvent, and B be in $\mathfrak{L}(\mathcal{D}(A), Y)$.*

System (6.2.2) *is admissible in the sense of* (6.2.3) *if and only if*

$$\left\{ \begin{array}{l} \text{There exist } \ r > 0 \ \text{ and } \ D > 0 \ \text{ such that} \\ \text{for all } n \in \mathbb{N} \text{ and for all } \ z = \displaystyle\sum_{l \in J_r(\mu_n)} c_l \Psi_l : \quad \|Bz\|_Y \leq D \|z\|_X , \end{array} \right. \tag{6.2.4}$$

where

$$J_r(\mu) = \{l \in \mathbb{N}, \ \text{such that } |\mu_l - \mu| \leq r\}. \tag{6.2.5}$$

Besides, if (6.2.4) *holds, then the constant K_T in* (6.2.3) *can be chosen as follows:*

$$K_T = K_{\pi/2r} \Big\lceil \frac{2rT}{\pi} \Big\rceil, \quad \text{with } K_{\pi/2r} = \frac{3\pi^4 D}{4r}. \tag{6.2.6}$$

To be more precise, in [12, Section 6], the estimates (6.2.6) are not given explicitly, but directly come from the proof of Theorem 6.1 in [12], which yields the constant

$$K_{\pi/2r} = 3D\hat{M}_r(0),$$

where $\hat{M}_r(0)$ is the Fourier transform at 0 of the function

$$M_r(t) = \frac{\pi^2}{8} \Big(\frac{\sin(rt)}{rt} \Big)^2.$$

This makes precise the constant $K_{\pi/2r}$, and the constant K_T for $T > 0$ can be obtained as a simple consequence of the semi-group property and the conservation of the energy for solutions of (6.2.2).

Resolvent characterization

In practice, when dealing with sequences of operators, whose eigenvectors may change, Theorem 6.2.1 is not easy to use. We therefore introduce other characterizations of admissibility of (6.2.2), which yield more convenient criteria.

Theorem 6.2.2. *Let A be a skew-adjoint unbounded operator on X with compact resolvent, and B be in $\mathfrak{L}(\mathcal{D}(A), Y)$.*

System (6.2.2) *is admissible in the sense of* (6.2.3) *if and only if there exist positive constants m and M such that*

$$M^2 \|(A - i\omega I)z\|_X^2 + m^2 \|z\|_X^2 \geq \|Bz\|_Y^2, \quad \forall z \in \mathcal{D}(A), \forall \omega \in \mathbb{R}. \tag{6.2.7}$$

Besides, if (6.2.7) *holds, then the constant* K_T *in* (6.2.3) *can be chosen as follows:*

$$K_T = K_1 \lceil T \rceil \quad \textit{with } K_1 = \frac{3\pi^3}{2}\sqrt{m^2 + M^2\frac{\pi^2}{4}}. \tag{6.2.8}$$

Proof. Assume that system (6.2.2) is admissible in the sense of (6.2.3). Then Theorem 6.2.1 proves the existence of constants r and D such that (6.2.4) holds.

We now recall the following result, which is inspired by [37], and precisely stated in [12, Lemma 6.2]:

Lemma 6.2.3. *Under the hypotheses of Theorem 6.2.2, assume that system* (6.2.2) *is admissible. For* $\varepsilon > 0$*, define*

$$V(\omega, \varepsilon) = \operatorname{span}\{\Psi_j \textit{ such that } |\mu_j - \omega| \leq \varepsilon\}.$$

Let us define $K(\omega, \varepsilon)$ *as*

$$K(\omega, \varepsilon) = \left\| B(A - i\omega I)^{-1} \right\|_{\mathfrak{L}(V(\omega,\varepsilon)^\perp, Y)}.$$

Then, for any $\varepsilon > 0$*,* $K(\omega, \varepsilon)$ *is uniformly bounded in* ω*, that is*

$$K(\varepsilon) = \sup_{\omega \in \mathbb{R}} K(\omega, \varepsilon) < \infty. \tag{6.2.9}$$

Besides, the following estimate holds

$$K(\varepsilon) \leq \sqrt{\frac{K_1}{1 - \exp(-1)}}\Big(1 + \frac{1}{\varepsilon}\Big), \tag{6.2.10}$$

where K_1 *is the admissibility constant in* (6.2.3) *for* $T = 1$*.*

Let $z \in \mathcal{D}(A)$ and $\omega \in \mathbb{R}$. Write $z = z_\omega + z_{\omega^\perp}$, with $z_\omega \in V(\omega, r)$ and $z_{\omega^\perp} \in V(\omega, r)^\perp$. Note that this decomposition is unique and that z_ω and $z_{\omega^\perp}$ are orthogonal in X, and with respect to the scalar product $< (A - i\omega I)\cdot, (A - i\omega I)\cdot >_X$. Then we have

$$\begin{aligned} \|Bz\|_Y^2 &\leq 2\|Bz_\omega\|_Y^2 + 2\|Bz_{\omega^\perp}\|_Y^2 \\ &\leq 2D^2\|z_\omega\|_X^2 + 2K(r)^2\|(A - i\omega I)z_{\omega^\perp}\|_X^2 \\ &\leq 2D^2\|z\|_X^2 + 2K(r)^2\|(A - i\omega I)z\|_X^2, \end{aligned}$$

and (6.2.7) is proved.

Conversely, assume that (6.2.7) holds. Let ε be a positive constant. Then, for all $\omega \in \mathbb{R}$, for all $z \in V(\omega, \varepsilon)$,

$$\|(A - i\omega I)z\|_X^2 \leq \varepsilon^2\|z\|_X^2,$$

and thus we get

$$\|Bz\|_Y^2 \leq (m^2 + M^2\varepsilon^2)\|z\|_X^2.$$

Estimate (6.2.4) follows with $r = \varepsilon$ and $D = \sqrt{m^2 + M^2\varepsilon^2}$, and, by Theorem 6.2.1, this implies the admissibility of system (6.2.2). Taking $\varepsilon = \pi/2$, we obtain the estimate (6.2.8). □

Applications to System (6.1.1)-(6.1.2)

Let us consider the abstract setting of (6.1.1)-(6.1.2), which is a particular instance of (6.2.2), with $A = -iA_0$.

In this case, one can obtain a more convenient spectral characterization of the admissibility of (6.1.1)-(6.1.2) by removing the dependence in the extra parameter $\omega \in \mathbb{R}$:

Theorem 6.2.4. *Let A_0 be an unbounded self-adjoint positive definite operator on X with compact resolvent, and B be in $\mathfrak{L}(\mathcal{D}(A_0), Y)$.*

System (6.1.1)-(6.1.2) *is admissible in the sense of* (6.1.3) *if and only if there exist positive constants α and β such that*

$$\left\| A_0^{1/2} z \right\|_X^4 \leq \|z\|_X^2 \left(\|A_0 z\|_X^2 + \alpha^2 \|z\|_X^2 - \beta^2 \|Bz\|_Y^2 \right), \quad \forall z \in \mathcal{D}(A_0). \tag{6.2.11}$$

Besides, if (6.2.11) *holds, then system* (6.1.1) *is admissible, and the constant K_T in* (6.1.3) *can be chosen as follows:*

$$K_T = K_1 \lceil T \rceil, \quad \text{with } K_1 = \frac{3\pi^3}{2\beta} \sqrt{\alpha^2 + \frac{\pi^2}{4}}. \tag{6.2.12}$$

Proof. The idea is very simple. Thanks to Theorem 6.2.2, we only need to prove the equivalence between (6.2.7) and (6.2.11).

Now, remark that condition (6.2.7) for (6.1.1)-(6.1.2) reads as follows: There exist positive constants m and M such that

$$M^2 \|(A_0 - \omega I)z\|_X^2 + m^2 \|z\|_X^2 \geq \|Bz\|_Y^2, \quad \forall z \in \mathcal{D}(A_0), \forall \omega \in \mathbb{R}.$$

This is equivalent to say that the quadratic form in ω

$$\omega^2 \|z\|_X^2 - 2\omega \left\| A_0^{1/2} z \right\|_X^2 + \|A_0 z\|_X^2 + \frac{m^2}{M^2} \|z\|_X^2 - \frac{1}{M^2} \|Bz\|_Y^2$$

is nonnegative for all $z \in \mathcal{D}(A_0)$, or, equivalently, that its determinant is nonpositive, i.e.

$$\left\| A_0^{1/2} z \right\|_X^4 \leq \|z\|_X^2 \left(\|A_0 z\|_X^2 + \frac{m^2}{M^2} \|z\|_X^2 - \frac{1}{M^2} \|Bz\|_Y^2 \right).$$

This coincides with (6.2.11) by the identification

$$\alpha = \frac{m}{M}, \qquad \beta = \frac{1}{M}. \tag{6.2.13}$$

The equivalence is then straightforward and estimate (6.2.12) follows from (6.2.8), and identity (6.2.13). □

6.2.2 Characterizations of observability

We first recall the results in [6, 32] concerning the observability properties for (6.2.2), which consist in the existence of a time T^* and a constant k_{T^*} such that any solution of (6.2.2) with initial date $z_0 \in \mathcal{D}(A)$ satisfies

$$k_{T^*} \|z_0\|_X^2 \leq \int_0^{T^*} \|Bz(t)\|_Y^2 \, dt. \tag{6.2.14}$$

Theorem 6.2.5 ([6, 32])**.** *Let A be a skew-adjoint unbounded operator on X with compact resolvent, and $B \in \mathfrak{L}(\mathcal{D}(A), Y)$.*

If system (6.2.2) *is admissible and exactly observable in time T^*, then there exist positive constants m and M such that*

$$M^2 \|(A - i\omega I)z\|_X^2 + m^2 \|Bz\|_Y^2 \geq \|z\|_X^2, \quad \forall z \in \mathcal{D}(A),\ \forall \omega \in \mathbb{R}. \tag{6.2.15}$$

Besides, in (6.2.15)*, one can choose $m = \sqrt{2T^*/k_{T^*}}$ and $M = T^*\sqrt{K_{T^*}/k_{T^*}}$ where the constants k_{T^*} and K_{T^*} are the ones in* (6.2.14) *and* (6.2.3) *respectively.*

Conversely, if (6.2.15) *holds, then for any time $T > \pi M$, system* (6.2.2) *is exactly observable, and the constant k_T in* (6.1.4) *can be chosen as*

$$k_T = \frac{1}{2m^2 T}(T^2 - \pi^2 M^2). \tag{6.2.16}$$

Theorem 6.2.5, when specified to system (6.1.1)-(6.1.2), yields the following result:

Theorem 6.2.6. *Assume that $A_0 : \mathcal{D}(A_0) \subset X \to X$ is an unbounded self-adjoint positive definite operator with compact resolvent, and that $B \in \mathfrak{L}(\mathcal{D}(A_0), Y)$ for some Hilbert space Y.*

If system (6.1.1)-(6.1.2) *is admissible and exactly observable, then there exist positive constants α and β such that*

$$\left\|A_0^{1/2} z\right\|_X^4 \leq \|z\|_X^2 \left(\|A_0 z\|_X^2 + \alpha^2 \|Bz\|_Y^2 - \beta^2 \|z\|_X^2 \right), \quad \forall z \in \mathcal{D}(A_0). \tag{6.2.17}$$

Conversely, if (6.2.17) *holds, then system* (6.1.1)-(6.1.2) *is exactly observable in any time $T > \pi/\beta$, and the constant k_T in* (6.1.4) *can be chosen as*

$$k_T = \frac{\beta^2}{2\alpha^2 T}\Big(T^2 - \frac{\pi^2}{\beta^2}\Big). \tag{6.2.18}$$

Proof. This result is based on Theorem 6.2.5. Indeed, we only prove that conditions (6.2.17) and (6.2.15) are equivalent. Note that condition (6.2.15) for (6.1.1)-(6.1.2) simply takes the form

$$M^2 \|(A_0 - \omega I)z\|_X^2 + m^2 \|Bz\|_Y^2 \geq \|z\|_X^2, \quad \forall z \in \mathcal{D}(A_0),\ \forall \omega \in \mathbb{R}. \tag{6.2.19}$$

Remark that (6.2.19) can be rewritten as

$$\omega^2 \|z\|_X^2 - 2\omega \left\|A_0^{1/2} z\right\|_X^2 + \left(\|A_0 z\|_X^2 + \frac{m^2}{M^2} \|Bz\|_Y^2 - \frac{1}{M^2} \|z\|_X^2 \right) \geq 0,$$
$$\forall z \in \mathcal{D}(A_0),\ \forall \omega \in \mathbb{R}. \tag{6.2.20}$$

Since this last expression simply is a quadratic expression in $\omega \in \mathbb{R}$, then the nonnegativity of (6.2.20) is equivalent to the nonpositivity of the discriminant of (6.2.20), i.e.

$$\left\|A_0^{1/2} z\right\|_X^4 \leq \|z\|_X^2 \left(\|A_0 z\|_X^2 + \frac{m^2}{M^2} \|Bz\|_Y^2 - \frac{1}{M^2} \|z\|_X^2 \right), \quad \forall z \in \mathcal{D}(A_0). \tag{6.2.21}$$

which is obviously equivalent to (6.2.17), with $\alpha = m/M$ and $\beta = 1/M$.

Conversely, if (6.2.17) holds, inequality (6.2.19) holds for any $z \in \mathcal{D}(A_0)$ and $\omega \in \mathbb{R}$ by taking $m = \alpha/\beta$ and $M = 1/\beta$. Therefore, using the estimates in Theorem 6.2.5, it follows that if (6.2.17) holds, system (6.1.1)-(6.1.2) is exactly observable for any time $T > \pi/\beta$, and estimate (6.2.18) follows from (6.2.16). □

6.3 Proof of Theorem 6.1.3

In this section, we present the proof of Theorem 6.1.3. To this end, we consider an unbounded self-adjoint positive definite operator A_0 with compact resolvent, and $B \in \mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$, with $\kappa < 1/2$, and we work under the assumptions of Theorem 6.1.3.

For convenience, since B is assumed to belong to $\mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$, we introduce a constant K_B such that

$$\|B\phi\|_Y \leq K_B \|A_0^\kappa \phi\|_X , \quad \forall \phi \in \mathcal{D}(A_0^\kappa).$$

The proof is divided into two major parts, one analyzing the admissibility properties (6.1.13), and the other one the observability properties (6.1.15).

6.3.1 Admissibility

Proof of Theorem 6.1.3: Admissibility. Assume that system (6.1.1)-(6.1.2) is admissible. Then, from Theorem 6.2.4, (6.2.11) holds for some positive constants α and β.

In view of Theorem 6.2.4, the admissibility properties (6.1.13) is equivalent to the existence of two positive constants α_* and β_* such that, for all $h > 0$,

$$\left\| A_{0h}^{1/2} z_h \right\|_h^4 \leq \|z_h\|_h^2 \left(\|A_{0h} z_h\|_h^2 + \alpha_*^2 \|z_h\|_h^2 - \beta_*^2 \|B_h z_h\|_Y^2 \right), \quad \forall z_h \in \mathcal{C}_h(\eta/h^\sigma). \tag{6.3.1}$$

To prove inequality (6.3.1), a natural idea would have been to choose $z = \pi_h z_h$ in (6.2.11). However, since we did not assume that $\pi_h(V_h) \subset \mathcal{D}(A_0)$, this cannot be done. For instance, in the case of P1 finite elements for A_0 the Laplace operator (say on $(0,1)$) with Dirichlet boundary conditions, we have that $\pi_h(V_h) \cap \mathcal{D}(A_0) = \{0\}$. Actually, even if we assume $\pi_h(V_h) \subset \mathcal{D}(A_0)$, for z_h lying in a filtered class, it is not clear that the quantities $\|A_{0h} z_h\|_h$ and $\|A_0 \pi_h z_h\|_X$ are close.

Therefore, in the sequel, we fix $h > 0$, and, for $z_h \in \mathcal{C}_h(\eta/h^\sigma)$, where η is an arbitrary positive number independent of $h > 0$, we consider $Z_h \in X$ defined by

$$A_0 Z_h = \pi_h A_{0h} z_h = \pi_h \pi_h^* A_0 \pi_h z_h. \tag{6.3.2}$$

Note that (6.3.2) defines Z_h properly, since A_0 is invertible.

Besides, $Z_h \in \mathcal{D}(A_0)$, since $A_0 Z_h$ belongs to X by (6.3.2). It follows that (6.2.11) applies and gives

$$\left\| A_0^{1/2} Z_h \right\|_X^4 \leq \|Z_h\|_X^2 \left(\|A_0 Z_h\|_X^2 + \alpha^2 \|Z_h\|_X^2 - \beta^2 \|B Z_h\|_X^2 \right). \tag{6.3.3}$$

Below, we will deduce estimate (6.3.1) from (6.3.3), by comparing each term carefully.

From the definition (6.3.2) of Z_h, we have

$$\|A_{0h} z_h\|_h = \|\pi_h A_{0h} z_h\|_X = \|A_0 Z_h\|_X . \tag{6.3.4}$$

We now estimate $Z_h - \pi_h z_h$. Using (6.1.6) and (6.3.2), for all $\phi \in \mathcal{D}(A_0)$, we have:

$$\begin{aligned} < Z_h, A_0 \phi >_X = < A_0 Z_h, \phi >_X = < \pi_h A_{0h} z_h, \phi >_X \\ = < \pi_h \pi_h^* A_0 \pi_h z_h, \phi >_X = < A_0^{1/2} \pi_h z_h, A_0^{1/2} \pi_h \pi_h^* \phi >_X . \end{aligned} \tag{6.3.5}$$

In particular, this implies that

$$\begin{aligned} < (Z_h - \pi_h z_h), A_0\phi >_X &= < Z_h, A_0\phi >_X - < A_0^{1/2}\pi_h z_h, A_0^{1/2}\phi > \\ &= < A_0^{1/2}\pi_h z_h, A_0^{1/2}(\pi_h\pi_h^* - I)\phi >_X . \end{aligned}$$

Using (6.1.9) and the invertibility of A_0, we obtain

$$\begin{aligned} \|Z_h - \pi_h z_h\|_X &= \sup_{\substack{\phi\in\mathcal{D}(A_0),\\ \|A_0\phi\|_X=1}} \left\{ < (Z_h - \pi_h z_h), A_0\phi >_X \right\} \\ &\le \left\|A_0^{1/2}\pi_h z_h\right\|_X \sup_{\substack{\phi\in\mathcal{D}(A_0),\\ \|A_0\phi\|_X=1}} \left\|A_0^{1/2}(\pi_h\pi_h^* - I)\phi\right\|_X \\ &\le C_0 h^\theta \left\|A_0^{1/2}\pi_h z_h\right\|_X . \end{aligned}$$

Besides, for any $\gamma \in [0,1]$, in view of (6.1.9), interpolation properties yield

$$\left\|A_0^{1/2}(\pi_h\pi_h^* - I)\phi\right\|_X \le C_0 h^{\theta(1-\gamma)} \left\|A_0^{1-\gamma/2}\phi\right\|_X, \quad \forall\phi \in \mathcal{D}(A_0^{1-\gamma/2}),$$

and thus, as above,

$$\begin{aligned} \left\|A_0^{\gamma/2}(Z_h - \pi_h z_h)\right\|_X &= \sup_{\substack{\phi\in\mathcal{D}(A_0^{1-\gamma/2}),\\ \left\|A_0^{1-\gamma/2}\phi\right\|_X=1}} \{< A_0^{\gamma/2}(Z_h - \pi_h z_h), A_0^{1-\gamma/2}\phi >_X\} \\ &\le \left\|A_0^{1/2}\pi_h z_h\right\|_X \sup_{\substack{\phi\in\mathcal{D}(A_0^{1-\gamma/2}),\\ \left\|A_0^{1-\gamma/2}\phi\right\|_X=1}} \left\|A_0^{1/2}(\pi_h\pi_h^* - I)\phi\right\|_X \\ &\le C_0 h^{\theta(1-\gamma)} \left\|A_0^{1/2}\pi_h z_h\right\|_X . \end{aligned}$$

Especially, for $\gamma = 2\kappa$, we obtain

$$\|A_0^\kappa(Z_h - \pi_h z_h)\|_X \le C_0 h^{\theta(1-2\kappa)} \left\|A_0^{1/2}\pi_h z_h\right\|_X .$$

Besides, using the definition (6.1.5) of A_{0h}, one easily gets that

$$\left\|A_{0h}^{1/2}\phi_h\right\|_h = \left\|A_0^{1/2}\pi_h\phi_h\right\|_X, \quad \forall\phi_h \in V_h. \tag{6.3.6}$$

It follows that

$$\begin{cases} \|Z_h - \pi_h z_h\|_X \le C_0 h^\theta \left\|A_{0h}^{1/2} z_h\right\|_h , \\ \|A_0^\kappa(Z_h - \pi_h z_h)\|_X \le C_0 h^{\theta(1-2\kappa)} \left\|A_{0h}^{1/2} z_h\right\|_h . \end{cases} \tag{6.3.7}$$

In particular, this implies, by the definition of the norm $\|\cdot\|_h$, that

$$\|z_h\|_h - C_0 h^\theta \left\|A_{0h}^{1/2} z_h\right\|_h \le \|Z_h\|_X \le \|z_h\|_h + C_0 h^\theta \left\|A_{0h}^{1/2} z_h\right\|_h , \tag{6.3.8}$$

and that

$$\|Z_h\|_X^2 \le 2\|z_h\|_h^2 + 2C_0^2 h^{2\theta} \left\|A_{0h}^{1/2} z_h\right\|_h^2 . \tag{6.3.9}$$

Using $B \in \mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$ and the estimate (6.3.7), we obtain

$$\|BZ_h\|_Y \geq \|B_h z_h\|_Y \ - K_B C_0 h^{\theta(1-2\kappa)} \left\|A_{0h}^{1/2} z_h\right\|_h . \tag{6.3.10}$$

Then we obtain

$$\|BZ_h\|_Y^2 \geq \frac{1}{2} \|B_h z_h\|_Y^2 \ - K_B^2 C_0^2 h^{2\theta(1-2\kappa)} \left\|A_{0h}^{1/2} z_h\right\|_h^2 . \tag{6.3.11}$$

We now estimate $\left\|A_0^{1/2} Z_h\right\|_X^2 - \left\|A_{0h}^{1/2} z_h\right\|_h^2$. On one hand, we have

$$\left\|A_0^{1/2} Z_h\right\|_X^2 =< A_0 Z_h, Z_h >_X =< \pi_h A_{0h} z_h, Z_h >_X =< A_{0h} z_h, \pi_h^* Z_h >_h .$$

On the other hand, we have

$$\left\|A_{0h}^{1/2} z_h\right\|_h^2 =< A_{0h} z_h, z_h >_h =< A_{0h} z_h, \pi_h^* \pi_h z_h >_h .$$

Subtracting these two identities, we get

$$\left\|A_0^{1/2} Z_h\right\|_X^2 - \left\|A_{0h}^{1/2} z_h\right\|_h^2 =< A_{0h} z_h, \pi_h^* (Z_h - \pi_h z_h) >_h,$$

and therefore, using (6.3.7), that

$$\left| \left\|A_0^{1/2} Z_h\right\|_X^2 - \left\|A_{0h}^{1/2} z_h\right\|_h^2 \right| \leq C_0 h^\theta \|A_{0h} z_h\|_h \left\|A_{0h}^{1/2} z_h\right\|_h . \tag{6.3.12}$$

Plugging (6.3.4), (6.3.8), (6.3.9), (6.3.10) and (6.3.12) into (6.3.3), we get

$$\begin{aligned} \Big(\left\|A_{0h}^{1/2} z_h\right\|_h^2 - C_0 h^\theta \|A_{0h} z_h\|_h \ \left\|A_{0h}^{1/2} z_h\right\|_h \Big)^2 \leq \Big(\|z_h\|_h + C_0 h^\theta \left\|A_{0h}^{1/2} z_h\right\|_h \Big)^2 \\ \Big[\|A_{0h} z_h\|_h^2 + \alpha^2 \Big(2 \|z_h\|_h^2 + 2 C_0^2 h^{2\theta} \left\|A_{0h}^{1/2} z_h\right\|_h^2 \Big) \\ - \frac{\beta^2}{2} \|B_h z_h\|_Y^2 + \beta^2 K_B^2 C_0^2 h^{2\theta(1-2\kappa)} \left\|A_{0h}^{1/2} z_h\right\|_h^2 \Big]. \end{aligned}$$

Since z_h is assumed to belong to $\mathcal{C}_h(\eta/h^\sigma)$, we get

$$\begin{aligned} \left\|A_{0h}^{1/2} z_h\right\|_h^4 (1 - C_0 h^{\theta-\sigma/2} \sqrt{\eta})^2 \leq \|z_h\|_h^2 (1 + C_0 h^{\theta-\sigma/2} \sqrt{\eta})^2 \Big[\|A_{0h} z_h\|_h^2 \\ + \Big(2\alpha^2 + 2\alpha^2 C_0^2 h^{2\theta-\sigma} \eta + \beta^2 K_B^2 C_0^2 h^{2\theta(1-2\kappa)-\sigma} \eta \Big) \|z_h\|_h^2 - \frac{\beta^2}{2} \|B_h z_h\|_Y^2 \Big]. \end{aligned}$$

Using $\sigma < 2\theta$, one gets that, for h small enough,

$$1 \leq \Big(\frac{1 + C_0 h^{\theta-\sigma/2} \sqrt{\eta}}{1 - C_0 h^{\theta-\sigma/2} \sqrt{\eta}} \Big)^2 \leq 1 + 5 C_0 h^{\theta-\sigma/2} \sqrt{\eta} \leq 2, \tag{6.3.13}$$

and thus,

$$\begin{aligned} \left\|A_{0h}^{1/2} z_h\right\|_h^4 \leq \|z_h\|_h^2 \Big[\|A_{0h} z_h\|_h^2 + 5 C_0 h^{\theta-\sigma/2} \sqrt{\eta} \|A_{0h} z_h\|_h^2 \\ + 2 \Big(2\alpha^2 + 2\alpha^2 C_0^2 h^{2\theta-\sigma} \eta + \beta^2 K_B^2 C_0^2 h^{2\theta(1-2\kappa)-\sigma} \eta \Big) \|z_h\|_h^2 - \frac{\beta^2}{2} \|B_h z_h\|_Y^2 \Big]. \end{aligned}$$

Since z_h belongs to $\mathcal{C}_h(\eta/h^\sigma)$, this yields

$$\left\|A_{0h}^{1/2}z_h\right\|_h^4 \leq \|z_h\|_h^2 \Big[\|A_{0h}z_h\|_h^2 + \Big(5C_0h^{\theta-\sigma/2-2\sigma}\eta^{5/2} + 4\alpha^2 + 4\alpha^2C_0^2h^{2\theta-\sigma}\eta + 2\beta^2K_B^2C_0^2h^{2\theta(1-2\kappa)-\sigma}\eta\Big)\|z_h\|_h^2 - \frac{\beta^2}{2}\|B_hz_h\|_Y^2\Big].$$

Thus, with σ as in (6.1.11), we obtain (6.3.1) with

$$\alpha_*^2 = 5C_0\eta^{5/2} + 4\alpha^2(1+C_0^2\eta) + 2\beta^2K_B^2C_0^2\eta, \qquad \beta_*^2 = \frac{1}{2}\beta^2.$$

This completes the proof of the first statement in Theorem 6.1.3. Also note that, using Theorem 6.2.4, one can get explicit estimates on the constant $K_{T,\eta}$ in (6.1.13). □

6.3.2 Observability

Proof of Theorem 6.1.3: Observability. Assume that system (6.1.1)-(6.1.2) is admissible and exactly observable. Then, from Theorem 6.2.6, there exist positive constants α and β such that (6.2.17) holds.

In view of Theorem 6.2.6, our goal is to prove that there exist positive constants α_* and β_* such that for any $h>0$, the following inequality holds:

$$\left\|A_{0h}^{1/2}z_h\right\|_h^4 \leq \|z_h\|_h^2\Big(\ \|A_{0h}z_h\|_h^2 + \alpha_*^2\|B_hz_h\|_Y^2 - \beta_*^2\|z_h\|_h^2\Big), \quad \forall z_h\in\mathcal{C}_h(\epsilon/h^\sigma). \tag{6.3.14}$$

To prove inequality (6.3.14), as before, we fix $z_h\in\mathcal{C}_h(\epsilon/h^\sigma)$, where ϵ is a positive parameter independent of $h>0$ that we will choose later on, and we introduce the element $Z_h\in X$ defined by (6.3.2). Again, since A_0Z_h belongs to X by (6.3.2), $Z_h\in\mathcal{D}(A_0)$. Then (6.2.17) applies and yields

$$\left\|A_0^{1/2}Z_h\right\|_X^4 \leq \|Z_h\|_X^2\Big(\ \|A_0Z_h\|_X^2 + \alpha^2\|BZ_h\|_Y^2 - \beta^2\|Z_h\|_X^2\Big). \tag{6.3.15}$$

Using (6.3.8), we get

$$\frac{1}{2}\|z_h\|_h^2 - C_0^2h^{2\theta}\left\|A_{0h}^{1/2}z_h\right\|_h^2 \leq \|Z_h\|_X^2. \tag{6.3.16}$$

Using $B\in\mathfrak{L}(\mathcal{D}(A_0^\kappa),Y)$ and the estimate (6.3.7), we obtain

$$\|BZ_h\|_Y \leq \|B_hz_h\|_Y\ + K_BC_0h^{\theta(1-2\kappa)}\left\|A_{0h}^{1/2}z_h\right\|_h,$$

and then

$$\|BZ_h\|_Y^2 \leq 2\|B_hz_h\|_Y^2\ + 2K_B^2C_0^2h^{2\theta(1-2\kappa)}\left\|A_{0h}^{1/2}z_h\right\|_h^2. \tag{6.3.17}$$

Now, plugging estimates (6.3.4), (6.3.9), (6.3.12), (6.3.16) and (6.3.17) into (6.3.15), we obtain

$$\Big(\left\|A_{0h}^{1/2}z_h\right\|_h^2 - C_0h^\theta\|A_{0h}z_h\|_h\left\|A_{0h}^{1/2}z_h\right\|_h\Big)^2 \leq \Big(\|z_h\|_h + C_0h^\theta\left\|A_{0h}^{1/2}z_h\right\|_h\Big)^2 \Bigg[\|A_{0h}z_h\|_h^2 + \alpha^2\Big(2\|B_hz_h\|_Y^2\ + 2K_B^2C_0^2h^{2\theta(1-2\kappa)}\left\|A_{0h}^{1/2}z_h\right\|_h^2\Big) - \frac{\beta^2}{2}\|z_h\|_h^2 + \beta^2C_0^2h^{2\theta}\left\|A_{0h}^{1/2}z_h\right\|_h^2\Bigg].$$

Using that $z_h \in \mathcal{C}_h(\epsilon/h^\sigma)$, we get that

$$(1 - C_0 h^{\theta-\sigma/2}\sqrt{\epsilon})^2 \left\|A_{0h}^{1/2} z_h\right\|_h^4 \leq \|z_h\|_h^2 (1 + C_0 h^{\theta-\sigma/2}\sqrt{\epsilon})^2 \left[\|A_{0h} z_h\|_h^2 + 2\alpha^2 \|B_h z_h\|_Y^2 \ + 2\alpha^2 K_B^2 C_0^2 h^{2\theta(1-2\kappa)-\sigma} \epsilon \|z_h\|_h^2 - \frac{\beta^2}{2} \|z_h\|_h^2 + \beta^2 C_0^2 h^{2\theta-\sigma} \epsilon \|z_h\|_h^2 \right].$$

For h small enough, estimate (6.3.13) holds, and then it follows that

$$\left\|A_{0h}^{1/2} z_h\right\|_h^4 \leq \|z_h\|_h^2 \left[\|A_{0h} z_h\|_h^2 + 5C_0 h^{\theta-\sigma/2-2\sigma} \epsilon^{5/2} \|z_h\|_h^2 + 4\alpha^2 \|B_h z_h\|_Y^2 + 4\alpha^2 K_B^2 C_0^2 h^{2\theta(1-2\kappa)-\sigma} \epsilon \|z_h\|_h^2 - \frac{\beta^2}{2} \|z_h\|_h^2 + 2\beta^2 C_0^2 h^{2\theta-\sigma} \epsilon \|z_h\|_h^2 \right].$$

According to the choice (6.1.11) of σ, this yields

$$\left\|A_{0h}^{1/2} z_h\right\|_h^4 \leq \|z_h\|_h^2 \left[\|A_{0h} z_h\|_h^2 + 4\alpha^2 \|B_h z_h\|_Y^2 + \left(5C_0 \epsilon^{5/2} + \ +4\alpha^2 K_B^2 C_0^2 \epsilon + 2\beta^2 C_0^2 \epsilon - \frac{\beta^2}{2}\right) \|z_h\|_h^2 \right].$$

Choosing $\epsilon > 0$ such that

$$5C_0 \epsilon^{5/2} + \ 4\alpha^2 K_B^2 C_0^2 \epsilon + 2\beta^2 C_0^2 \epsilon = \frac{\beta^2}{4},$$

we finally obtain (6.3.14) with

$$\alpha_* = 2\alpha, \qquad \beta_* = \frac{1}{2}\beta,$$

which completes the proof of Theorem 6.1.3.

Also remark that Theorem 6.2.6 provides explicit estimates on the constants T and k_* in (6.1.15). □

Remark 6.3.1. Similar results hold when the operator A_0 only is nonnegative. This can be done without restriction with the following argument.

The function z is solution of (6.1.1) if and only if $z_* = z\exp(-it)$ is the solution of

$$\begin{cases} i\dot{z}_* = (A_0 + Id) z_*, \quad t \geq 0, \\ z_*(0) = z_0. \end{cases} \tag{6.3.18}$$

The observation y in (6.1.2) now reads on (6.3.18) as $y(t) = \exp(it) B z_*(t)$.

Thus the admissibility and observability properties for (6.1.1)-(6.1.2) are equivalent to the corresponding ones for (6.3.18). Also remark that $A_* = A_0 + Id$ has exactly the same domain as A_0, with equivalent norms, but now, A_* is positive definite.

Besides, when discretizing (6.3.18) using a finite element method, the discretized version of A_* simply is $A_{*h} = A_{0h} + Id_{V_h}$, and again, the admissibility and observability properties for (6.1.7) and for

$$\begin{cases} \dot{z}_{*h} = A_{0h} z_{*h} + z_{*h}, & t \geq 0, \\ z_{*h}(0) = z_{0h} \in V_h, \end{cases} \qquad y_h(t) = e^{it} B_h z_{*h}(t), \quad t \geq 0,$$

are equivalent.

Note that this argument can also be applied to deal with self-adjoint operators A_0 that are only bounded from below in the sense of quadratic forms.

6.4 Examples of applications

This section is dedicated to present some applications to Theorem 6.1.3, and to confront our results with the existing ones in the literature.

6.4.1 The 1-d case

Let us consider the classical 1d Schrödinger equation:

$$\begin{cases} i\partial_t z + \partial_{xx}^2 z = 0, & (t,x) \in \mathbb{R} \times (0,1), \\ z(t,0) = z(t,1) = 0, & t \in \mathbb{R}, \\ z(0,x) = z_0(x), & x \in (0,1). \end{cases} \tag{6.4.1}$$

For (a,b) a subset of $(0,1)$, we observe system (6.4.1) through

$$y(t,x) = z(t,x)\chi_{(a,b)}(x), \tag{6.4.2}$$

where $\chi_{(a,b)}$ is the characteristic function of (a,b).

This models indeed enters in the abstract framework considered in this article, by setting $A_0 = -\partial_{xx}^2$ with Dirichlet boundary conditions, and $B = \chi_{(a,b)}$. Indeed, A_0 is a self-adjoint positive definite operator with compact resolvent in $L^2(0,1)$ and of domain $H^2(0,1) \cap H_0^1(0,1)$. The operator B obviously is continuous on $L^2(0,1)$ with values in $L^2(0,1)$. The admissibility property for (6.4.1)-(6.4.2) is then straightforward.

The observability property for (6.4.1)-(6.4.2) is well-known to hold in any time $T > 0$ when the *Geometric Control Condition* is satisfied, see [26, 3]. This condition, roughly speaking, asserts the existence of a time T^* such that all the rays of Geometric Optics enters in the observation domain in a time smaller than T^*. In 1d, this condition is always satisfied, and thus system (6.4.1)-(6.4.2) is exactly observable in any time $T > 0$. This can also be seen using multipliers techniques [30].

To construct the space V_h, we use P1 finite elements. More precisely, for $n_h \in \mathbb{N}$, set $h = 1/(n_h+1) > 0$ and define the points $x_j = jh$ for $j \in \{0, \cdots, n_h+1\}$. We define the basis functions

$$e_j(x) = \Big[1 - \frac{|x - x_j|}{h}\Big]^+, \quad \forall j \in \{1, \cdots, n_h\}.$$

Now, $V_h = \mathbb{C}^{n_h}$, and the injection π_h simply is

$$\begin{array}{rcl} \pi_h : V_h = \mathbb{C}^{n_h} & \to & L^2(0,1) \\ z_h = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n_h} \end{pmatrix} & \mapsto & \pi_h z_h(x) = \displaystyle\sum_{j=1}^{n_h} z_j e_j(x). \end{array}$$

Usually, the resulting schemes are written as

$$\left\{ \begin{array}{l} iM_h \dot{z}_h(t) + K_h z_h(t) = 0, \quad t \in \mathbb{R}, \\ z_h(0) = z_{0h}, \end{array} \right. \qquad y_h(t) = B\pi_h z_h(t), \quad t \in \mathbb{R}, \tag{6.4.3}$$

where M_h and K_h are $n_h \times n_h$ matrices defined by $(M_h)_{i,j} = \int_0^1 e_i(x) e_j(x)\ dx$ and $(K_h)_{i,j} = \int_0^1 \partial_x e_i(x) \partial_x e_j(x)\ dx$. Note that, since M_h is a Gram matrix associated to a basis, it is invertible, self-adjoint and positive definite, and thus the following defines a scalar product:

$$< \phi_h, \psi_h >_h = \phi_h^* M_h \psi_h, \quad (\phi_h, \psi_h) \in V_h^2. \tag{6.4.4}$$

Besides, from the definition of M_h, one easily checks that

$$< \phi_h, \psi_h >_h = \int_0^1 \overline{\pi_h(\phi_h)(x)} \pi_h(\psi_h)(x)\ dx, \quad \forall (\phi_h, \psi_h) \in V_h^2,$$

as presented in the introduction.

Similarly, one obtains that, for all $(\phi_h, \psi_h) \in V_h^2$,

$$\begin{aligned} \phi_h^* K_h \psi_h = \phi_h^* M_h M_h^{-1} K_h \psi_h =< \phi_h, M_h^{-1} K_h \psi_h >_h = \phi_h^* K_h M_h^{-1} M_h \psi_h \\ =< M_h^{-1} K_h \phi_h, \psi_h >_h = \int_0^1 \overline{\partial_x(\pi_h \phi_h)(x)} \partial_x(\pi_h \psi_h)(x)\ dx, \end{aligned}$$

This proves that the operator $M_h^{-1} K_h$ coincides with the operator A_{0h} of our framework. Note that this operator indeed is self-adjoint, as expected, but with respect to the scalar product (6.4.4) and not with the usual hilbertian norm of $\mathbb{C}^{n_h}$.

It is by now a common feature of finite element techniques (see for instance [39]) that, in this case, estimates (6.1.9) hold for $\theta = 1$. We can thus apply Theorem 6.1.3 to systems (6.4.3):

Theorem 6.4.1. *There exist $\epsilon > 0$, a time T^* and a constant k_* such that for any $h > 0$, any solution z_h of* (6.4.3) *with initial data $z_{0h} \in \mathcal{C}_h(\epsilon/h^{2/5})$ satisfies* (6.1.15).

This result is to be compared with the ones in [27]: In [27], it is proved that, for finite difference approximation schemes of the 1d beam equation, observability properties hold uniformly within the larger class $\mathcal{C}_h(\alpha/h^2)$ for $\alpha < 4$. Though not stated in [27], the same results hold for Schrödinger equation, thus leading better results than our approach.

Though, as we will see hereafter, we can tackle more general cases, even in 1d, for instance taking sequence of meshes $\mathcal{S}_n$ given by $n+2$ points as

$$x_{0,n} = 0 < x_{1,n} < \cdots < x_{n,n} < x_{n+1,n} = 1, \qquad h_{j+1/2,n} = x_{j+1,n} - x_{j,n},$$

for which we assume $h_n = \sup_j \{h_{j+1/2,n}\}$ to go to zero when $n \to \infty$.

6.4.2 More general cases

Let us mention that our results also apply in more intricate cases. Let Ω be a smooth bounded domain of $\mathbb{R}^N$ for $N \in \mathbb{N}^*$, and consider

$$\begin{cases} i\partial_t z + \mathrm{div}_x(\sigma(x)\nabla_x z) = V(x)z, & (t,x) \in \mathbb{R} \times \Omega, \\ z(t,x) = 0, & (t,x) \in \mathbb{R} \times \partial\Omega, \\ z(0,x) = z_0(x), & x \in \Omega, \end{cases} \tag{6.4.5}$$

where σ is a C^1 positive real valued function on $\bar{\Omega}$, and V is a real-valued nonnegative bounded function in Ω. This indeed enters in the abstract setting of (6.1.1) by setting $A_0 = -\mathrm{div}_x(\sigma(x)\nabla_x \cdot) + V(x)$ with Dirichlet boundary condition, which is a self-adjoint positive definite operator with compact resolvent in $L^2(\Omega)$ and of domain $H^2(\Omega) \cap H_0^1(\Omega)$.

Let ω be an open subdomain of Ω and consider the observation operator

$$y(t,x) = \chi_\omega(x) z(t,x), \quad t \in \mathbb{R}. \tag{6.4.6}$$

Assume that system (6.4.5)-(6.4.6) is exactly observable.

To guarantee this property to hold, one can assume for instance that the *Geometric Control Condition* (see [3] and above) is satisfied. But, in fact, the Schrödinger equation behaves slightly better than a wave equation from the observability point of view because of the infinite velocity of propagation. The *Geometric Control Condition* is sufficient but not always necessary. For instance, in [24], it has been proved that when the domain Ω is a square, for any non-empty bounded open subset ω, the observability property (6.1.4) holds for system (6.1.1). Other geometries have been also dealt with, see for instance [5, 1, 6, 42].

We consider P1 finite elements on meshes $\mathcal{T}_h$. We furthermore assume that the meshes $\mathcal{T}_h$ of the domain Ω are regular in the sense of [39, Section 5]. Roughly speaking, this assumption imposes that the polyhedra in $(\mathcal{T}_h)$ are not too flat:

Definition 6.4.2. Let $\mathcal{T} = \cup_{K \in \mathcal{T}} K$ be a mesh of a bounded domain Ω. For each polyhedron $K \in \mathcal{T}$, we define h_K as the diameter of K and ρ_K as the maximum diameter of the spheres $S \subset K$. We then define the regularity of $\mathcal{T}$ as

$$\mathrm{Reg}(\mathcal{T}) = \sup_{K \in \mathcal{T}} \Big\{ \frac{h_K}{\rho_K} \Big\}.$$

A sequence of meshes $(\mathcal{T}_h)_{h>0}$ is said to be uniformly regular if

$$\sup_h \mathrm{Reg}(\mathcal{T}_h) < \infty.$$

In this case, see [39, Section 5], estimates (6.1.9) again hold for $\theta = 1$, and Theorem 6.1.3 implies:

Theorem 6.4.3. *Assume that system* (6.4.5)-(6.4.6) *is exactly observable. Given a sequence of meshes* $(\mathcal{T}_h)_{h>0}$ *which is uniformly regular, there exist* $\epsilon > 0$*, a time* T^* *and a constant* k_* *such that for any* $h > 0$*, any solution* z_h *of the P1 finite element approximation scheme of* (6.4.5) *corresponding to the mesh* $\mathcal{T}_h$ *with initial data* $z_{0h} \in \mathcal{C}_h(\epsilon/h^{2/5})$ *satisfies* (6.1.15)*.*

To our knowledge, this is the first time that observability properties for space semi-discretizations of (6.4.5) are derived in such generality. In particular, we emphasize that the only non-trivial assumption we used is (6.1.9), which is needed anyway to guarantee the convergence of the numerical schemes under consideration.

6.5 Fully discrete approximation schemes

This section is based on the article [12], which studied observability properties of time discrete conservative linear systems. As said in [12, Section 5], this study can be combined with observability results on space semi-discrete systems to deduce observability properties for fully discrete systems. Below, we present some applications of the results in [12].

Let us consider time discretizations of (6.1.7) which takes the form

$$z_h^{k+1} = \mathbb{T}_{\triangle t,h} z_h^k, \quad k \in \mathbb{N}, \qquad z_h^0 = z_{0h} \in V_h. \tag{6.5.1}$$

Here $\triangle t > 0$ denotes the time discretization parameter, and z_h^k corresponds to an approximation of the solution z_h of (6.1.7) at time $t_k = k\triangle t$. The operator $\mathbb{T}_{\triangle t,h} : V_h \to V_h$ is an approximation of $\exp(-i(\triangle t)A_{0h})$.

To be more precise, we assume that there exists a smooth strictly increasing function ζ defined on an interval $[-R, R]$ (with $R \in (0, \infty]$) with values in $(-\pi, \pi)$, and such that

$$\mathbb{T}_{\triangle t,h} = \exp(-i\zeta((\triangle t)A_{0h})). \tag{6.5.2}$$

In particular, this assumption implies that the operator $\mathbb{T}_{\triangle t,h}$ is unitary, and then the solutions of (6.5.1) have constant norms. The parameter R corresponds to a frequency limit $R/\triangle t$ imposed by the time discretization method we consider. The fact that the range of ζ is included in $(-\pi, \pi)$ reflects that one cannot measure frequencies higher than $\pi/\triangle t$ in a mesh of size $\triangle t$. The hypothesis on the strict monotonicity of ζ is a non-degeneracy condition on the group velocity (see for instance [43] and [12, Remark 4.9]) for solutions of (6.5.1) which is necessary to guarantee the propagation of solutions required for observability properties to hold.

We also assume

$$\frac{\zeta(\eta)}{\eta} \to 1 \quad \text{as} \quad \eta \to 0,$$

which guarantees the consistency of the time discrete schemes (6.5.1) with the time continuous models (6.1.7).

Remark that these hypotheses are usually satisfied for conservative time-discrete approximation schemes such as the midpoint discretization or the so-called fourth order Gauss method (see for instance [18] or [12, Subsection 4.2]).

Then, from [12], we get:

Theorem 6.5.1. *Let A_0 be an unbounded self-adjoint positive definite operator with compact resolvent on X, and $B \in \mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$, with $\kappa < 1/2$.*

Assume that the maps $(\pi_h)_{h>0}$ satisfy property (6.1.9). *Set σ as in* (6.1.11).

Consider a time discrete approximation scheme characterized by a function ζ as above, and let $\delta \in (0, R)$.

Admissibility: *Assume that system* (6.1.1)-(6.1.2) *is admissible.*

Then, for any $\eta > 0$ and $T > 0$, there exists a positive constant $K_{T,\eta,\delta} > 0$ such that, for any $h > 0$ and $\triangle t > 0$, any solution of (6.5.1) *with initial data*

$$z_{0h} \in \mathcal{C}_h(\eta/h^\sigma) \cap \mathcal{C}_h(\delta/\triangle t) \tag{6.5.3}$$

satisfies

$$\triangle t \sum_{k\triangle t\in[0,T]} \left\| B_h z_h^k \right\|_Y^2 \leq K_{T,\eta,\delta} \left\| z_{0h} \right\|_h^2 . \tag{6.5.4}$$

Observability: *Assume that system* (6.1.1)-(6.1.2) *is admissible and exactly observable.*

Then there exist $\epsilon > 0$*, a time* T^* *and a positive constant* $k_* > 0$ *such that, for any* $h > 0$ *and* $\triangle t > 0$*, any solution of* (6.5.1) *with initial data*

$$z_{0h} \in \mathcal{C}_h(\epsilon/h^\sigma) \cap \mathcal{C}_h(\delta/\triangle t) \tag{6.5.5}$$

satisfies

$$k_* \left\| z_{0h} \right\|_h^2 \leq \triangle t \sum_{k\triangle t\in[0,T^*]} \left\| B_h z_h^k \right\|_Y^2 . \tag{6.5.6}$$

Obviously, inequalities (6.5.4)-(6.5.6) are time discrete counterparts of (6.1.13)-(6.1.15). Remark that, as in Theorem 6.1.3, a filtering condition is needed, but which now depends on both time and space discretization parameters.

Also remark that if $(\triangle t)h^{-\sigma}$ is small enough, then $\mathcal{C}_h(\epsilon/h^\sigma) \cap \mathcal{C}_h(\delta/\triangle t) = \mathcal{C}_h(\epsilon/h^\sigma)$. Roughly speaking, this indicates that under the CFL type condition $(\triangle t)h^{-\sigma} \leq \epsilon/\delta$, then system (6.5.1) behaves, with respect to the admissibility and observability properties, similarly as the space semi-discrete equations (6.1.7).

6.6 Controllability properties

In this section, we present applications of Theorem 6.1.3 to controllability properties. In the sequel, we thus assume the hypotheses of Theorem 6.1.3.

6.6.1 The continuous setting

We consider the following control problem: Given $T > 0$, for any $y_0 \in X$, find a control $v \in L^2(0,T;Y)$ such that the solution y of

$$\dot{y} = -iA_0 y + B^* v(t), \quad t \in [0,T], \qquad y(0) = y_0, \tag{6.6.1}$$

satisfies

$$y(T) = 0. \tag{6.6.2}$$

It is well-known (see for instance [28]) that the controllability issue in time T for (6.6.1) is equivalent to the exact observability property for (6.1.1)-(6.1.2) in time T. Indeed, these two properties are dual, and this duality can be made precise using the Hilbert Uniqueness Method (HUM in short), see [28].

Roughly speaking, the idea of HUM is to consider the set of all functions $v \in L^2(0,T;Y)$ such that the corresponding solution of (6.6.1) satisfies (6.6.2), which we will call in the sequel admissible controls for (6.6.1), and to select the one of minimal $L^2(0,T;Y)$ norm.

This control of minimal $L^2(0,T;Y)$ norm for (6.6.1), which we will denote by v_{HUM}, is characterized through the minimizer of the functional $\mathcal{J}$ defined on X by

$$\mathcal{J}(z_T) = \frac{1}{2}\int_0^T \|Bz(t)\|_Y^2 \ dt + \mathcal{R}e(< y_0, z(0) >_X), \tag{6.6.3}$$

where $\mathcal{R}e$ denotes the real part application and z is the solution of

$$\dot{z} = -iA_0 z, \quad t \in [0,T], \qquad z(T) = z_T. \tag{6.6.4}$$

Indeed, if z_T^* is the minimizer of $\mathcal{J}$, then $v_{HUM}(t) = Bz^*(t)$, where z^* is the solution of (6.6.4) with initial data z_T^*.

Besides, the only admissible control v for (6.6.1) that can be written as $v = Bz$ for a solution z of (6.6.4) is the HUM control v_{HUM}. This characterization will be used in the sequel.

Note that the observability property for (6.1.1)-(6.1.2) implies the strict convexity and the coercivity of $\mathcal{J}$ and therefore guarantees the existence of a unique minimizer for $\mathcal{J}$.

6.6.2 The space semi-discrete setting

We are in the setting of Theorem 6.1.3. Therefore there exists a time T^* such that (6.1.15) holds for any solution of (6.1.7) with initial data in the filtered space $\mathcal{C}_h(\epsilon/h^\sigma)$.

Now, if we try to compute an approximation of the control v_{HUM}, a natural idea consists in computing the discrete HUM controls for discrete versions of (6.6.1), which provides a sequence of controls that shall converge to the HUM control v_{HUM} for (6.6.1). However, this method may fail due to high-frequency spurious waves created by the discretization process. We refer for instance to [46] for a detailed presentation of this fact in the context of the 1d wave equation. It is then natural to develop filtering techniques which overcome this difficulty. This is precisely the object of several articles, see for instance [36, 45, 46, 35, 17], and the methods presented below follow and adapt their approach.

We now fix $T \geq T^*$.

Following the strategy of HUM, we will introduce the adjoint problem:

$$\dot{z}_h = -iA_{0h} z_h, \quad t \in [0,T], \qquad z_h(T) = z_{Th}. \tag{6.6.5}$$

Method I

For any $h > 0$, we consider the following control problem: For any $y_{0h} \in V_h$ find $v_h \in L^2(0,T;Y)$ of minimal $L^2(0,T;Y)$ such that the solution y_h of

$$\dot{y}_h = -iA_{0h} y_h + B_h^* v_h(t), \quad t \in [0,T], \qquad y_h(0) = y_{0h}, \tag{6.6.6}$$

satisfies

$$P_h y_h(T) = 0, \tag{6.6.7}$$

where P_h is the orthogonal projection in V_h on $\mathcal{C}_h(\epsilon/h^\sigma)$.

To deal with this problem, we introduce the functional $\mathcal{J}_h$ defined for $z_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma)$ by

$$\mathcal{J}_h(z_{Th}) = \frac{1}{2}\int_0^T \|B_h z_h(t)\|_Y^2 \ dt + \mathcal{R}e(< y_{0h}, z_h(0) >_h), \tag{6.6.8}$$

where z_h is the solution of (6.6.5) with initial data $z_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma)$.

For each $h > 0$, the functional $\mathcal{J}_h$ is strictly convex and coercive (see (6.1.15)), and thus has a unique minimizer $z^*_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma)$. Besides, we have:

Lemma 6.6.1. *For all $h > 0$, let $z^*_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma)$ be the unique minimizer of $\mathcal{J}_h$, and denote by z^*_h the corresponding solution of* (6.6.5).

Then the solution of (6.6.6) *with* $v_h = B_h z^*_h$ *satisfies* (6.6.7).

Sketch of the proof. We present briefly the proof, which is standard (see for instance [28]).

On one hand, multiplying (6.6.6) by z_h solution of (6.6.5) with initial data z_{Th}, we get that, for all $z_{Th} \in V_h$,

$$\int_0^T < v_h(t), B_h z_h(t) >_Y \ dt + < y_{0h}, z_h(0) >_h - < y_h(T), z_h(T) >_h = 0. \tag{6.6.9}$$

On the other hand, the Fréchet derivative of the functional $\mathcal{J}_h$ at z^*_{Th} yields:

$$\mathcal{R}e\Big(\int_0^T < B_h z^*_h(t), B_h z_h(t) >_Y \ dt\Big) + \mathcal{R}e(< y_{0h}, z_h(0) >_h) = 0, \ \forall z_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma). \tag{6.6.10}$$

Therefore, setting $v_h = B_h z^*_h$, taking the real part of (6.6.9) and subtracting it to (6.6.10), we obtain

$$\mathcal{R}e(< y_h(T), z_{Th} >_h) = 0, \quad \forall z_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma),$$

or, equivalently, (6.6.7). □

We then investigate the convergence of the discrete controls v_h obtained in Lemma 6.6.1.

Theorem 6.6.2. *Assume that the hypotheses of Theorem 6.1.3 are satisfied. Also assume that*

$$Y_X = \Big\{ v \in Y, \ \textit{such that } B^* v \in X \Big\} \tag{6.6.11}$$

is dense in Y.

Let $y_0 \in X$, and consider a sequence $(y_{0h})_{h>0}$ such that y_{0h} belongs to V_h for any $h > 0$ and

$$\pi_h y_{0h} \to y_0 \quad \text{in } X. \tag{6.6.12}$$

Then the sequence $(v_h)_{h>0}$ of discrete controls given by Lemma 6.6.1 converges in $L^2(0,T;Y)$ to the HUM control v_{HUM} of (6.6.1).

Remark that, for $y_0 \in \mathcal{D}(A_0)$, in view of (6.1.9), the sequence $(y_{0h})_h = (\pi^*_h y_0)$ converges to y_0 in X in the sense of (6.6.12). For $y_0 \in X$, one can then find a sequence $(y_{0h})_{h>0}$ satisfying (6.6.12) and $y_{0h} \in V_h$ for any $h > 0$ by using the density of $\mathcal{D}(A_0)$ into X.

The technical assumption (6.6.11) on B is usually satisfied, and thus does not limit the range of applications of Theorem 6.6.2. Also note that when B is bounded from X to Y, the space Y_X coincides with Y and (6.6.11) is then automatically satisfied.

Proof. The proof is divided into several parts: First, we prove that the sequence $(v_h)_{h>0}$ is bounded in $L^2(0,T;Y)$. Then, we show that any weak accumulation point v of $(v_h)_{h>0}$ is an admissible control for (6.6.1). We then prove that v coincides with the HUM control v_{HUM} of (6.6.1), which also proves that there is only one accumulation point for the sequence (v_h). Finally, we prove the strong convergence of the sequence (v_h) to $v = v_{HUM}$ in $L^2(0,T;Y)$.

The discrete controls are bounded Using that z^*_{Th} minimizes $\mathcal{J}_h$, we obviously have that $\mathcal{J}_h(z^*_{Th}) \leq J_h(0) = 0$, and therefore

$$\int_0^T \|B_h z_h^*(t)\|_Y^2 \ dt \leq -2\mathcal{R}e(< y_{0h}, z_h^*(0) >_h) \leq 2\,\|\pi_h y_{0h}\|_X\,\|z_h^*(0)\|_h\,.$$

Since T has been chosen such that the observability inequality (6.1.15) holds for any solution of (6.1.7) -or equivalently (6.6.5)- with initial data in $\mathcal{C}_h(\epsilon/h^\sigma)$ with a constant k_* independent of h, we get the following both inequalities:

$$k_*\,\|z_h^*(0)\|_h \leq 2\,\|\pi_h y_{0h}\|_X\,, \qquad \int_0^T \|B_h z_h^*(t)\|_Y^2 \ dt \leq \frac{4}{k_*}\,\|\pi_h y_{0h}\|_X^2\,. \tag{6.6.13}$$

Since $v_h = B_h z_h^*$ and the sequence $(\pi_h y_{0h})$ is convergent in X, we deduce from (6.6.13) that the sequence $(v_h)_{h>0}$ is bounded in $L^2(0,T;Y)$. Therefore we can extract subsequences such that the sequence $(v_h)_{h>0}$ weakly converges in $L^2(0,T;Y)$. From now on, we assume that

$$v_h \rightharpoonup v \quad \text{in } L^2(0,T;Y). \tag{6.6.14}$$

The weak accumulation point v is an admissible control for (6.6.1) Using the same duality as in (6.6.9), v is an admissible control for (6.6.1) if and only if for any solution z of (6.6.4), we have

$$\mathcal{R}e\Big(\int_0^T < v(t), Bz(t) >_Y \ dt\Big) + \mathcal{R}e(< y_0, z(0) >_X) = 0. \tag{6.6.15}$$

Since we already get from (6.6.10) that any solution of (6.6.5) with initial data $z_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma)$ satisfies

$$\mathcal{R}e\Big(\int_0^T < v_h(t), B_h z_h(t) >_Y \ dt\Big) + \mathcal{R}e(< y_{0h}, z_h(0) >_h) = 0, \tag{6.6.16}$$

the proof of (6.6.15) is based on the convergence of the solutions of (6.6.5) to the solutions of (6.6.4):

Lemma 6.6.3. *[39, Section 8] Assume that $z_T \in \mathcal{D}(A_0)$, and consider a sequence $(\pi_h z_{Th})_{h>0}$ which weakly converges to z_T in $\mathcal{D}(A_0^{1/2})$.*

Then the sequence of solutions $(z_h)_{h>0}$ of (6.6.5) with initial data z_{Th} converges to the solution z of (6.6.4) with initial data z_T in the following sense:

$$\begin{aligned} &\pi_h z_h \to z \quad \text{in } \ C([0,T];X), \\ &\pi_h z_h \to z \quad \text{in } L^\infty(0,T;\mathcal{D}(A_0^{1/2}))\ w-*. \end{aligned} \tag{6.6.17}$$

Strictly speaking, the proof in [39] is dealing with the convergence of wave type equations, but it can be easily adapted to our case.

Therefore, taking $z_T \in \mathcal{D}(A_0)$, we only have to choose $z_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma)$ such that $(\pi_h z_{Th}) \to z_T$ in $\mathcal{D}(A_0^{1/2})$. This can be done by choosing

$$z_{Th} = P_h \pi_h^* z_T.$$

Indeed, with this choice, we have

$$\begin{aligned}
\|\pi_h z_{Th} - z_T\|_X &\leq \|(P_h - I)\pi_h^* z_T\|_h + \|(\pi_h \pi_h^* - I) z_T\|_X \\
&\leq \frac{h^{\sigma/2}}{\sqrt{\epsilon}} \left\| A_{0h}^{1/2} \pi_h^* z_T \right\|_h + \|(\pi_h \pi_h^* - I) z_T\|_X \\
&\leq \frac{h^{\sigma/2}}{\sqrt{\epsilon}} \left\| A_0^{1/2} \pi_h \pi_h^* z_T \right\|_X + \|(\pi_h \pi_h^* - I) z_T\|_X \\
&\leq \frac{h^{\sigma/2}}{\sqrt{\epsilon}} \Big(\left\| A_0^{1/2} z_T \right\|_X + \left\| A_0^{1/2} (\pi_h \pi_h^* - I) z_T \right\|_X \Big) + \|(\pi_h \pi_h^* - I) z_T\|_X ,
\end{aligned}$$

and therefore the strong convergence of $(\pi_h z_{Th})_{h>0}$ to z_T in X follows from (6.1.9). Besides, using (6.3.6), we have that

$$\begin{aligned}
\left\| A_0^{1/2} (\pi_h z_{Th} - \pi_h \pi_h^* z_T) \right\|_X &= \left\| A_0^{1/2} \pi_h (P_h - I d_{V_h}) \pi_h^* z_T \right\|_X \\
&= \left\| A_{0h}^{1/2} (P_h - I d_{V_h}) \pi_h^* z_T \right\|_h \leq \left\| A_{0h}^{1/2} \pi_h^* z_T \right\|_h \leq \left\| A_0^{1/2} \pi_h \pi_h^* z_T \right\|_X .
\end{aligned}$$

Combined with (6.1.9), this indicates that the sequence $(\pi_h z_{Th})_{h>0}$ is bounded in $\mathcal{D}(A_0^{1/2})$. Since it converges strongly to z_T in X, the sequence $(\pi_h z_{Th})_{h>0}$ converges weakly to z_T in $\mathcal{D}(A_0^{1/2})$.

Applying Lemma 6.6.3 to this particular sequence $(z_{Th})_{h>0}$, the corresponding sequence $(z_h)_{h>0}$ of solutions of (6.6.5) satisfies (6.6.17), and for all $h > 0$, $z_{Th} \in \mathcal{C}_h(\epsilon/h^\sigma)$. In particular, the convergences (6.6.17) imply that the sequence $(\pi_h z_h)_{h>0}$ converges strongly to z in $C([0,T]; \mathcal{D}(A_0^\kappa))$.

Thus, for $z_T \in \mathcal{D}(A_0)$, passing to the limit when $h \to 0$ in (6.6.16), we obtain that (6.6.15) holds for solutions of (6.6.4) for any initial data $z_T \in \mathcal{D}(A_0)$. By density of $\mathcal{D}(A_0)$ in X, we obtain that (6.6.15) actually holds for any solutions of (6.6.4) with any initial data $z_T \in X$, and thus v is an admissible control for (6.6.1).

The weak limit v is the HUM control of (6.6.1) Here we use that the HUM control v_{HUM} is the only admissible control that can be written as $Bz(t)$ for a solution z of (6.6.4). Since for all $h > 0$, $v_h(t) = B\pi_h z_h^*(t)$, a natural candidate for z is the limit (in a sense that will be made precise below) of the sequence z_h^*.

Here again, we will use a classical Lemma on the convergence of the finite element approximation schemes:

Lemma 6.6.4. *[39, Section 8] Let z_T be in X, and consider a sequence $(z_{Th})_{h>0}$ of elements of V_h which weakly converges to z_T in X, in the sense that $(\pi_h z_{Th}) \rightharpoonup z_T$ in X.*

Then the sequence of solutions z_h of (6.6.5) *with initial data z_{Th} weakly converges in $L^2(0,T;X)$ to the solution z of* (6.6.4) *with initial data z_T. Besides, for all time $t \in [0,T]$, the sequence $(\pi_h z_h(t))_{h>0}$ weakly converges in X to $z(t)$.*

Lemma 6.6.4 obviously is a refined version of Lemma 6.6.3. Actually, it can be deduced directly from Lemma 6.6.3 by a duality argument.

We now apply Lemma 6.6.4 to z^*_{Th}: Indeed, since system (6.6.5) is conservative, estimate (6.6.13) implies that

$$\|\pi_h z^*_{Th}\|_X = \|z^*_{Th}\|_h = \|z^*_h(0)\|_h$$

is bounded, and thus, up to an extracting process, that the sequence $(\pi_h z^*_{Th})_{h>0}$ weakly converges to some $\tilde{z}^*_T$ in X.

It follows that

$$\pi_h z^*_h \rightharpoonup \tilde{z}^* \quad \text{in } L^2(0,T;X),$$

where $\tilde{z}^*$ denotes the solution of (6.6.4) with initial data $\tilde{z}^*_T$. Using (6.6.11), we thus obtain that

$$v_h = B\pi_h z^*_h \rightharpoonup B\tilde{z}^* \quad \text{in } \ L^2(0,T;Y).$$

Therefore we obtain that

$$v_h \rightharpoonup v = v_{HUM} \quad \text{in } L^2(0,T;Y), \qquad \pi_h z_h \rightharpoonup \tilde{z}^* = z^* \quad \text{in } \ L^2(0,T;X), \tag{6.6.18}$$

where z^* is the solution of (6.6.4) with initial data z^*_T defined as the unique minimizer of the functional $\mathcal{J}$ defined in (6.6.3).

Strong convergence Since the sequence $(v_h)_{h>0}$ weakly converges to $v = v_{HUM}$ in $L^2(0,T;Y)$, we only have to check the convergence of the $L^2(0,T;Y)$ norms.

On one hand, applying (6.6.15) to z^*, and recalling that $v = v_{HUM} = Bz^*$, we obtain

$$\int_0^T \|v(t)\|^2_Y \ dt + \mathcal{R}e(< y_0, z^*(0) >_X) = 0.$$

On the other hand, applying (6.6.16) to z^*_{Th}, and recalling that $v_h = B_h z^*_h$, we obtain

$$\int_0^T \|v_h(t)\|^2_Y \ dt + \mathcal{R}e(< \pi_h y_{0h}, \pi_h z^*_h(0) >_X) = 0.$$

From Lemma 6.6.4, the sequence $(\pi_h z^*_h(0))$ weakly converges in X to $z^*(0)$. Since the sequence $(\pi_h y_{0h})_{h>0}$ is assumed to be strongly convergent in X to y_0, we get that

$$\int_0^T \|v_h(t)\|^2_Y \ dt \longrightarrow \int_0^T \|v(t)\|^2_Y \ dt,$$

and the strong convergence $v_h \to v = v_{HUM}$ in $L^2(0,T;Y)$ is proved. □

Method II

It might seem hard to implement in practice an efficient algorithm to filter the data. We therefore remind the works [17, 46] where an alternate process is given, which uses a Tychonoff regularization of the functionals $\mathcal{J}_h$. Roughly speaking, it consists in the addition of an extra term in the functionals $\mathcal{J}_h$ which makes the functionals coercive on the whole space V_h, uniformly with respect to h. However, for the proofs, we will require the more restrictive condition $B \in \mathfrak{L}(X,Y)$.

Let us introduce, for $h>0$, the functional $\mathcal{J}_h^*$, defined for $z_{Th}\in V_h$ by

$$\mathcal{J}_h^*(z_{Th})=\frac{1}{2}\int_0^T \|B_h z_h(t)\|_Y^2\ dt+\frac{h^\sigma}{2}<A_{0h}\tilde{z}_{Th},z_{Th}>_h+\mathcal{R}e(<y_{0h},z_h(0)>_h), \tag{6.6.19}$$

where z_h is the solution of (6.6.5) and $\tilde{z}_{Th}$ is the solution of

$$(Id_{V_h}+h^\sigma A_{0h})\tilde{z}_{Th}=z_{Th}. \tag{6.6.20}$$

This equation simply consists in an elliptic regularization of z_{Th}. The variational formulation of (6.6.20) is given by

$$<\pi_h\tilde{z}_{Th},\pi_h\phi_h>_X+h^\sigma<A_0^{1/2}\pi_h\tilde{z}_{Th},A_0^{1/2}\pi_h\phi_h>_X=<\pi_h z_{Th},\pi_h\phi_h>_X,\qquad\forall\phi_h\in V_h,$$

and thus $\tilde{z}_{Th}$ can be computed directly. To simplify the presentation, it is convenient to introduce the operator

$$\tilde{A}_{0h}=A_{0h}\Big(Id_{V_h}+h^\sigma A_{0h}\Big)^{-1}, \tag{6.6.21}$$

which satisfies

$$<\tilde{A}_{0h}z_{Th},z_{Th}>=<A_{0h}\tilde{z}_{Th},z_{Th}>_h=\left\|\tilde{A}_{0h}^{1/2}z_{Th}\right\|_h^2,$$

and the following two properties:

$$\begin{aligned}&\left\|h^{\sigma/2}\tilde{A}_{0h}^{1/2}\psi_h\right\|_h^2\le\|\psi_h\|_h^2,\quad\forall\psi_h\in V_h,\\&\left\|h^{\sigma/2}\tilde{A}_{0h}^{1/2}\psi_h\right\|_h^2\ge\frac{\delta}{1+\delta}\|\psi_h\|_h^2,\quad\forall\psi_h\in\mathcal{C}_h(\delta/h^\sigma)^\perp,\quad\forall\delta\ge0.\end{aligned} \tag{6.6.22}$$

Note in particular, that the operator $h^\sigma\tilde{A}_{0h}$ is bounded on V_h uniformly with respect to $h>0$. This guarantees uniform continuity properties for $\mathcal{J}_h^*$.

We now check that, for $B\in\mathfrak{L}(X,Y)$, the functionals $\mathcal{J}_h^*$ are strictly convex and uniformly coercive on V_h: Indeed, for $z_{Th}\in V_h$, Theorem 6.1.3 implies that any solution of (6.6.5) satisfies

$$k_T\|P_h z_{Th}\|_h^2\le\int_0^T\|B_hP_hz_h(t)\|_Y^2\ dt.$$

It follows that

$$\begin{aligned}\int_0^T\|B_hz_h(t)\|_Y^2\ dt&\ge\frac{1}{2}\int_0^T\|B_hP_hz_h(t)\|_Y^2\ dt-\int_0^T\left\|B_h\Big(P_h-Id_{V_h}\Big)z_h(t)\right\|_Y^2\ dt\\&\ge\frac{1}{2}\int_0^T\|B_hP_hz_h(t)\|_Y^2\ dt-T\|B\|_{\mathfrak{L}(X,Y)}^2\|(P_h-Id_{V_h})z_{Th}\|_h^2\\&\ge\frac{k_T}{2}\|P_hz_{Th}\|_h^2-T\|B\|_{\mathfrak{L}(X,Y)}^2\|(P_h-Id_{V_h})z_{Th}\|_h^2\\&\ge\frac{k_T}{2}\|z_{Th}\|_h^2-\Big(T\|B\|_{\mathfrak{L}(X,Y)}^2+\frac{k_T}{2}\Big)\|(P_h-Id_{V_h})z_{Th}\|_h^2\\&\ge\frac{k_T}{2}\|z_{Th}\|_h^2-\Big(T\|B\|_{\mathfrak{L}(X,Y)}^2+\frac{k_T}{2}\Big)\Big(\frac{1+\epsilon}{\epsilon}\Big)\left\|h^{\sigma/2}\tilde{A}_{0h}^{1/2}\Big(Id_{V_h}-P_h\Big)z_{Th}\right\|_h^2\\&\ge\frac{k_T}{2}\|z_{Th}\|_h^2-\Big(T\|B\|_{\mathfrak{L}(X,Y)}^2+\frac{k_T}{2}\Big)\Big(\frac{1+\epsilon}{\epsilon}\Big)\left\|h^{\sigma/2}\tilde{A}_{0h}^{1/2}z_{Th}\right\|_h^2.\end{aligned}$$

This proves the uniform coercivity of the functionals $\mathcal{J}_h^*$.

Thus, for each $h > 0$, $\mathcal{J}_h^*$ has a unique minimizer $Z_{Th} \in V_h$, and the uniform coercivity implies the existence of two constants C_1 and C_2 independent of $h > 0$ such that, setting Z_h the solution of (6.6.5) with initial data Z_{Th},

$$\|Z_h(0)\|_h^2 \leq C_1\Big(\int_0^T \|B_h Z_h(t)\|_Y^2 \ dt + h^\sigma \left\|\tilde{A}_{0h}^{1/2} Z_{Th}\right\|_h^2\Big) \leq C_2 \|y_{0h}\|_h^2 .$$

Besides, setting $v_h = B_h Z_h$, the solution y_h of (6.6.1) satisfies

$$y_h(T) = -h^\sigma A_{0h} \tilde{Z}_{Th} = -h^\sigma \tilde{A}_{0h} Z_{Th}.$$

In particular, if the sequence $(\pi_h y_{0h})_{h>0}$ strongly converges to $y_0 \in X$, the same arguments as before, combined with the uniform coercivity of the functional $\mathcal{J}_h^*$, prove that the sequence (v_h) converges to v_{HUM} strongly in $L^2(0,T;Y)$.

To sum up, the following statement holds:

Theorem 6.6.5. *Assume that the hypotheses of Theorem 6.1.3 are satisfied, and that $B \in \mathfrak{L}(X,Y)$.*

Let $y_0 \in X$, and consider a sequence $(y_{0h})_{h>0}$ such that y_{0h} belongs to V_h for any $h > 0$ and $(\pi_h y_{0h}) \to y_0$ in X.

Then the sequence $(v_h)_{h>0}$ of discrete controls given by $v_h = B_h Z_h$, where Z_h is the solution of (6.6.5) associated to the minimizer Z_{Th} of $\mathcal{J}_h^*$ (defined in (6.6.19)), converges in $L^2(0,T;Y)$ to the HUM control v_{HUM} of (6.6.1).

Remark 6.6.6. Similar results can be obtained for fully discrete approximation schemes obtained by discretizing equations (6.1.7) in time. In this case, the proof is based on the observability inequality (6.5.6) and on convergence results for the fully discrete approximation schemes, which can be found for instance in [39]. We deliberately choose to present the proof in the simpler case of the time continuous setting for simplifying the presentation.

6.7 Stabilization properties

This section is mainly based on the articles [15, 14], in which stabilization properties are derived for abstract linear damped systems. In this section, we assume $B \in \mathfrak{L}(X,Y)$.

6.7.1 The continuous setting

Consider the following damped Schr̈odinger type equations:

$$i\dot{z} = A_0 z - iB^*Bz, \quad t \geq 0, \qquad z(0) = z_0 \in X. \tag{6.7.1}$$

The energy of solutions of (6.7.1), defined by $E(t) = \|z(t)\|_X^2/2$, satisfies the dissipation law

$$\frac{dE}{dt}(t) = -\|Bz(t)\|_Y^2 , \quad t \geq 0. \tag{6.7.2}$$

System (6.7.1) is said to be exponentially stable if there exist two positive constants μ and ν such that

$$E(t) \leq \mu\, E(0)\, \exp(-\nu t), \quad t \geq 0. \tag{6.7.3}$$

It is by now classical (see [30, 19]) that the exponential decay of the energy of solutions of (6.7.1) is equivalent (here the operator B is bounded on X) to the observability inequality (6.1.4) for solutions of (6.1.1)-(6.1.2).

6.7.2 The space semi-discrete setting

We now assume that system (6.1.1)-(6.1.2) is exactly observable in the sense of (6.1.4), or, equivalently (see [30, 19]), that system (6.7.1) is exponentially stable.

Then, combining Theorem 6.1.3 and [15], we get:

Theorem 6.7.1. *Let A_0 be a unbounded self-adjoint with compact resolvent in X, and B be a bounded operator in $\mathfrak{L}(X,Y)$. Assume that system* (6.7.1) *is exponentially stable in the sense of* (6.7.3)*. Also assume that the hypotheses of Theorem 6.1.3 are satisfied, and set σ as in* (6.1.11)*.*

Consider a sequence of operators $(\mathcal{V}_h)_{h>0}$ defined on V_h such that for all $h>0$, $\mathcal{V}_h$ is self-adjoint and positive definite. Also assume that for all $h>0$, the operators $\mathcal{V}_h$ and P_h (recall that P_h is the orthogonal projection in V_h on $\mathcal{C}_h(\epsilon/h^\sigma)$) commute, and that there exist two positive constants c and C independent of $h>0$ such that

$$\begin{cases} h^{\sigma/2} \left\| \sqrt{\mathcal{V}_h} z_h \right\|_h \leq C \left\| z_h \right\|_h, \quad \forall z_h \in \mathcal{C}_h(\epsilon/h^\sigma), \\ h^{\sigma/2} \left\| \sqrt{\mathcal{V}_h} z_h \right\|_h \geq c \left\| z_h \right\|_h, \quad \forall z_h \in \mathcal{C}_h(\epsilon/h^\sigma)^\perp. \end{cases} \tag{6.7.4}$$

Then the space semi-discrete systems

$$i\dot{z}_h = A_{0h} z_h - i B_h^* B_h z_h - i h^\sigma \mathcal{V}_h z_h, \quad t \geq 0, \qquad z_h(0) = z_{0h} \in V_h, \tag{6.7.5}$$

are exponentially stable, uniformly with respect to the space discretization parameter $h>0$: there exist two positive constants μ_0 and ν_0 independent of $h>0$ such that for any $h>0$, any solution z_h of (6.7.5) *satisfies*

$$\left\| z_h(t) \right\|_h \leq \mu_0 \left\| z_h(0) \right\|_h \exp(-\nu_0 t), \quad t \geq 0. \tag{6.7.6}$$

Note that, since we assumed B bounded on X, $\kappa = 0$ in Theorem 6.1.3, and then σ coincides with $2\theta/5$.

The conditions (6.7.4) on the viscosity operator, roughly speaking, say that the operator $h^\sigma \mathcal{V}_h$ is negligible for frequencies in the range $\mathcal{C}_h(\epsilon/h^\sigma)$ and is dominant in the range $\mathcal{C}_h(\epsilon/h^\sigma)$. In other words, the viscosity operator $h^\sigma \mathcal{V}_h$ modifies significantly the dynamical properties of system (6.7.5) only at high frequencies.

In general, the viscosity operator is chosen as a function of A_{0h}, for instance as:

$$\mathcal{V}_{1h} = A_{0h}, \qquad \mathcal{V}_{2h} = \frac{A_{0h}}{I + h^\sigma A_{0h}}, \qquad \mathcal{V}_{3h} = h^\sigma A_{0h}^2.$$

Here, the choice $\mathcal{V}_{2h}$ has the advantage that the operator $h^\sigma \mathcal{V}_{2h}$ is bounded. Remark that the viscosity operator $\mathcal{V}_{2h}$ also coincides with the elliptic regularization operator $\tilde{A}_{0h}$ introduced in (6.6.20).

Remark 6.7.2. In [15], several time discrete approximation schemes are proposed to guarantee uniform exponential decay properties for the energy of the time semi-discrete schemes as a consequence of the exponential decay of the energy of the time continuous system. Since the results of [15] also apply to families of uniformly exponentially stable systems, one can apply them to fully discrete approximation schemes of (6.7.1).

6.8 Further comments

1. One of the interesting features of our approach is that it works in any dimension and in a very general setting. To our knowledge, this is the first work which proves in such a systematic way admissibility and observability properties for space semi-discrete approximation schemes as a consequence of the ones of the continuous setting.

2. A widely open question consists in finding the sharp filtering scale. We think that the results in [9, 10], which prove the lack of observability for the 1d wave equation in highly heterogeneous media, might give some insights on the best results we can expect on the filtering scale.

3. Our methods and results require the observation operator B to be continuous on $\mathcal{D}(A_0^\kappa)$, with $\kappa < 1/2$. However, in several relevant applications, as for instance when dealing with the boundary observation of the Schrödinger equation (see for instance [29]), this is not the case. This question deserves further work.

4. An interesting issue for Schrödinger type equations concerns their dispersive properties. To our knowledge, this question, which has been extensively studied in the last decades (see for instance [25] and the references therein), has been successfully addressed for numerical approximation schemes discretized using finite difference (or finite elements) on uniform meshes in dimension 1 and 2, see [21, 20, 22]. We think that, similarly as for the observability properties, one could use spectral conditions to derive uniform dispersive properties for space semi-discretizations of Schrödinger equations in a very general setting, for instance by adapting Morawetz's estimates (see [33]).

5. Following the same ideas as the ones presented here, one can derive admissibility and observability results for space semi-discretizations of wave type equations derived from the finite element method. This issue is currently investigated by the author and will be published elsewhere.

Bibliography

[1] B. Allibert. Contrôle analytique de l'équation des ondes et de l'équation de Schrödinger sur des surfaces de revolution. *Comm. Partial Differential Equations*, 23(9-10):1493–1556, 1998.

[2] H. T. Banks, K. Ito, and C. Wang. Exponentially stable approximations of weakly damped wave equations. In *Estimation and control of distributed parameter systems (Vorau, 1990)*, volume 100 of *Internat. Ser. Numer. Math.*, pages 1–33. Birkhäuser, Basel, 1991.

[3] C. Bardos, G. Lebeau, and J. Rauch. Sharp sufficient conditions for the observation, control and stabilization of waves from the boundary. *SIAM J. Control and Optimization*, 30(5):1024–1065, 1992.

[4] P. Bégout and F. Soria. A generalized interpolation inequality and its application to the stabilization of damped equations. *J. Differential Equations*, 240(2):324–356, 2007.

[5] N. Burq. Contrôle de l'équation des plaques en présence d'obstacles strictement convexes. *Mém. Soc. Math. France (N.S.)*, (55):126, 1993.

[6] N. Burq and M. Zworski. Geometric control in the presence of a black box. *J. Amer. Math. Soc.*, 17(2):443–471 (electronic), 2004.

[7] C. Castro and S. Micu. Boundary controllability of a linear semi-discrete 1-d wave equation derived from a mixed finite element method. *Numer. Math.*, 102(3):413–462, 2006.

[8] C. Castro, S. Micu, and A. Münch. Numerical approximation of the boundary control for the wave equation with mixed finite elements in a square. *IMA J. Numer. Anal.*, 28(1):186–214, 2008.

[9] C. Castro and E. Zuazua. Low frequency asymptotic analysis of a string with rapidly oscillating density. *SIAM J. Appl. Math.*, 60(4):1205–1233 (electronic), 2000.

[10] C. Castro and E. Zuazua. Concentration and lack of observability of waves in highly heterogeneous media. *Arch. Ration. Mech. Anal.*, 164(1):39–72, 2002.

[11] S. Ervedoza. Observability of the mixed finite element method for the 1d wave equation on non-uniform meshes. *To appear in ESAIM: COCV*, 2008. *Cf Chapitre 2.*

[12] S. Ervedoza, C. Zheng, and E. Zuazua. On the observability of time-discrete conservative linear systems. *J. Funct. Anal.*, 254(12):3037–3078, June 2008. *Cf Chapitre 3.*

[13] S. Ervedoza and E. Zuazua. Perfectly matched layers in 1-d: Energy decay for continuous and semi-discrete waves. *Numer. Math.*, 109(4):597–634, 2008. *Cf Chapitre 1.*

[14] S. Ervedoza and E. Zuazua. Uniform exponential decay for viscous damped systems. *To appear in Proc. of Siena "Phase Space Analysis of PDEs 2007", Special issue in honor of Ferrucio Colombini*, 2008. *Cf Chapitre 4.*

[15] S. Ervedoza and E. Zuazua. Uniformly exponentially stable approximations for a class of damped systems. *To appear in J. Math. Pures Appl.*, 2008. *Cf Chapitre 5.*

[16] R. Glowinski. Ensuring well-posedness by analogy: Stokes problem and boundary control for the wave equation. *J. Comput. Phys.*, 103(2):189–221, 1992.

[17] R. Glowinski, C. H. Li, and J.-L. Lions. A numerical approach to the exact boundary controllability of the wave equation. I. Dirichlet controls: description of the numerical methods. *Japan J. Appl. Math.*, 7(1):1–76, 1990.

[18] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving ordinary differential equations. I*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1993. Nonstiff problems.

[19] A. Haraux. Une remarque sur la stabilisation de certains systèmes du deuxième ordre en temps. *Portugal. Math.*, 46(3):245–258, 1989.

[20] L. I. Ignat and E. Zuazua. Dispersive properties of a viscous numerical scheme for the Schrödinger equation. *C. R. Math. Acad. Sci. Paris*, 340(7):529–534, 2005.

[21] L. I. Ignat and E. Zuazua. A two-grid approximation scheme for nonlinear Schrödinger equations: dispersive properties and convergence. *C. R. Math. Acad. Sci. Paris*, 341(6):381–386, 2005.

[22] L. I. Ignat and E. Zuazua. Numerical schemes for the nonlinear Schrödinger equation. *Preprint*, 2008.

[23] J.A. Infante and E. Zuazua. Boundary observability for the space semi discretizations of the 1-d wave equation. *Math. Model. Num. Ann.*, 33:407–438, 1999.

[24] S. Jaffard. Contrôle interne exact des vibrations d'une plaque rectangulaire. *Portugal. Math.*, 47(4):423–429, 1990.

[25] M. Keel and T. Tao. Endpoint Strichartz estimates. *Amer. J. Math.*, 120(5):955–980, 1998.

[26] G. Lebeau. Contrôle de l'équation de Schrödinger. *J. Math. Pures Appl. (9)*, 71(3):267–291, 1992.

[27] L. León and E. Zuazua. Boundary controllability of the finite-difference space semi-discretizations of the beam equation. *ESAIM Control Optim. Calc. Var.*, 8:827–862 (electronic), 2002. A tribute to J. L. Lions.

[28] J.-L. Lions. *Contrôlabilité exacte, Stabilisation et Perturbations de Systèmes Distribués. Tome 1. Contrôlabilité exacte*, volume RMA 8. Masson, 1988.

[29] E. Machtyngier. Contrôlabilité exacte et stabilisation frontière de l'équation de Schrödinger. *C. R. Acad. Sci. Paris Sér. I Math.*, 310(12):801–806, 1990.

[30] E. Machtyngier and E. Zuazua. Stabilization of the Schrödinger equation. *Portugal. Math.*, 51(2):243–256, 1994.

[31] F. Macià. The effect of group velocity in the numerical analysis of control problems for the wave equation. In *Mathematical and numerical aspects of wave propagation—WAVES 2003*, pages 195–200. Springer, Berlin, 2003.

[32] L. Miller. Controllability cost of conservative systems: resolvent condition and transmutation. *J. Funct. Anal.*, 218(2):425–444, 2005.

[33] C. S. Morawetz. Decay for solutions of the exterior problem for the wave equation. *Comm. Pure Appl. Math.*, 28:229–264, 1975.

[34] A. Münch and A. F. Pazoto. Uniform stabilization of a viscous numerical approximation for a locally damped wave equation. *ESAIM Control Optim. Calc. Var.*, 13(2):265–293 (electronic), 2007.

[35] M. Negreanu, A.-M. Matache, and C. Schwab. Wavelet filtering for exact controllability of the wave equation. *SIAM J. Sci. Comput.*, 28(5):1851–1885 (electronic), 2006.

[36] M. Negreanu and E. Zuazua. Convergence of a multigrid method for the controllability of a 1-d wave equation. *C. R. Math. Acad. Sci. Paris*, 338(5):413–418, 2004.

[37] K. Ramdani, T. Takahashi, G. Tenenbaum, and M. Tucsnak. A spectral approach for the exact observability of infinite-dimensional systems with skew-adjoint generator. *J. Funct. Anal.*, 226(1):193–229, 2005.

[38] K. Ramdani, T. Takahashi, and M. Tucsnak. Uniformly exponentially stable approximations for a class of second order evolution equations—application to LQR problems. *ESAIM Control Optim. Calc. Var.*, 13(3):503–527, 2007.

[39] P.-A. Raviart and J.-M. Thomas. *Introduction à l'analyse numérique des équations aux dérivées partielles*. Collection Mathématiques Appliquées pour la Maitrise. Masson, Paris, 1983.

[40] L. R. Tcheugoué Tebou and E. Zuazua. Uniform boundary stabilization of the finite difference space discretization of the $1 - d$ wave equation. *Adv. Comput. Math.*, 26(1-3):337–365, 2007.

[41] L.R. Tcheugoué Tébou and E. Zuazua. Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity. *Numer. Math.*, 95(3):563–598, 2003.

[42] G. Tenenbaum and M. Tucsnak. Fast and strongly localized observation for the schrödinger equation. *Trans. Amer. Math. Soc.*, To appear.

[43] L. N. Trefethen. Group velocity in finite difference schemes. *SIAM Rev.*, 24(2):113–136, 1982.

[44] E. Zuazua. Boundary observability for the finite-difference space semi-discretizations of the 2-D wave equation in the square. *J. Math. Pures Appl. (9)*, 78(5):523–563, 1999.

[45] E. Zuazua. Optimal and approximate control of finite-difference approximation schemes for the 1D wave equation. *Rend. Mat. Appl. (7)*, 24(2):201–237, 2004.

[46] E. Zuazua. Propagation, observation, and control of waves approximated by finite difference methods. *SIAM Rev.*, 47(2):197–243 (electronic), 2005.

Chapter 7

Wave equations

Abstract: In this article, we derive uniform admissibility and observability properties for the finite element space semi-discretizations of $\ddot{u} + A_0 u = 0$, where A_0 is an unbounded self-adjoint positive definite operator with compact resolvent. To address this problem, we present a new spectral approach based on several spectral criteria for admissibility and observability of such systems. Our approach provides very general admissibility and observability results for finite element approximation schemes of $\ddot{u}+A_0 u = 0$, which stand in any dimension and for any *regular* mesh (in the sense of finite elements). Our results can be combined with previous works to derive admissibility and observability properties for fully discretizations of $\ddot{u} + A_0 u = 0$. We also present applications of our results to controllability and stabilization problems. We finally give applications of our results to space semi-discretizations of Schrödinger systems $i\dot{z} = A_0 z$, again based on spectral techniques.

7.1 Introduction

Let X be a Hilbert space endowed with the norm $\|\cdot\|_X$ and let $A_0 : \mathcal{D}(A_0) \subset X \to X$ be a self-adjoint positive definite operator with compact resolvent.

Let us consider the following abstract system:

$$\ddot{u}(t) + A_0 u(t) = 0, \quad t \in \mathbb{R}, \qquad u(0) = u_0, \quad \dot{u}(0) = u_1. \tag{7.1.1}$$

Here and henceforth, a dot ($\dot{}$) denotes differentiation with respect to the time t. In (7.1.1), the initial state (u_0, u_1) lies in $\mathfrak{X} = \mathcal{D}(A_0^{1/2}) \times X$.

Such systems are often used as models of vibrating systems (e.g., the wave and beams equations). Note that system (7.1.1) is conservative: the energy

$$E(t) = \frac{1}{2} \left\| A_0^{1/2} u(t) \right\|_X^2 + \frac{1}{2} \|\dot{u}(t)\|_X^2 \tag{7.1.2}$$

of solutions of (7.1.1) is constant.

Assume that Y is another Hilbert space equipped with the norm $\|\cdot\|_Y$. We denote by $\mathfrak{L}(X, Y)$ the space of bounded linear operators from X to Y, endowed with the classical operator norm. Let

$B \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$ be an observation operator and define the output function

$$y(t) = B\dot{u}(t). \tag{7.1.3}$$

We assume that the operator $B \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$ is admissible for system (7.1.1) in the following sense:

Definition 7.1.1. System (7.1.1)-(7.1.3) is admissible if for every $T > 0$ there exists a constant $K_T > 0$ such that any solution of (7.1.1) with initial data $(u_0, u_1) \in \mathcal{D}(A_0) \times \mathcal{D}(A_0^{1/2})$ satisfies:

$$\int_0^T \|B\dot{u}(t)\|_Y^2 \, dt \leq K_T \Big(\left\| A_0^{1/2} u_0 \right\|_X^2 + \|u_1\|_X^2 \Big). \tag{7.1.4}$$

Note that if B is *bounded* on X, i.e. if it can be extended in such a way that $B \in \mathfrak{L}(X, Y)$, then B is obviously an admissible observation operator, and K_T can be chosen as $K_T = T \|B\|_{\mathfrak{L}(X,Y)}^2$. However, in applications, this is often not the case, and the admissibility condition is then a consequence of a suitable "hidden regularity" property of the solutions of the evolution equation (7.1.1).

The exact observability property for system (7.1.1)-(7.1.3) can be formulated as follows:

Definition 7.1.2. System (7.1.1)-(7.1.3) is exactly observable in time T if there exists $k_T > 0$ such that any solution of (7.1.1) with initial data $(u_0, u_1) \in \mathcal{D}(A_0) \times \mathcal{D}(A_0^{1/2})$ satisfies:

$$k_T \Big(\left\| A_0^{1/2} u_0 \right\|_X^2 + \|u_1\|_X^2 \Big) \leq \int_0^T \|B\dot{u}(t)\|_Y^2 \, dt. \tag{7.1.5}$$

Moreover, system (7.1.1)-(7.1.3) is said to be exactly observable if it is exactly observable in some time $T > 0$.

Note that observability and admissibility issues arise naturally when dealing with controllability and stabilization properties of linear systems (see for instance the textbook [23]). These links will be clarified later on.

There is an extensive literature providing observability results for wave and plate equations, among other models, and by various methods including microlocal analysis [2, 3], multipliers techniques [21, 30] and Carleman estimates [18, 39], etc. Our goal in this paper is to develop a theory allowing to get observability results for space semi-discrete systems as a direct consequence of those corresponding to the continuous ones, thus avoiding technical developments in the discrete setting.

Let us now introduce the finite element method for (7.1.1).

Consider $(V_h)_{h>0}$ a sequence of vector spaces of finite dimension n_h which embed into X via a linear injective map $\pi_h : V_h \to X$. For each $h > 0$, the inner product $< \cdot, \cdot >_X$ in X induces a structure of Hilbert space for V_h endowed by the scalar product $< \cdot, \cdot >_h = < \pi_h \cdot, \pi_h \cdot >_X$.

We assume that for each $h > 0$, the vector space $\pi_h(V_h)$ is a subspace of $\mathcal{D}(A_0^{1/2})$. We thus define the linear operator $A_{0h} : V_h \to V_h$ by

$$< A_{0h}\phi_h, \psi_h >_h = < A_0^{1/2}\pi_h\phi_h, A_0^{1/2}\pi_h\psi_h >_X, \quad \forall (\phi_h, \psi_h) \in V_h^2. \tag{7.1.6}$$

The operator A_{0h} defined in (7.1.6) obviously is self-adjoint and positive definite. If we introduce the adjoint π_h^* of π_h, definition (7.1.6) implies that

$$A_{0h} = \pi_h^* A_0 \pi_h. \tag{7.1.7}$$

This operator A_{0h} corresponds to the finite element discretization of the operator A_0 (see [33]). We thus consider the following space semi-discretizations for (7.1.1):

$$\ddot{u}_h + A_{0h} u_h = 0, \quad t \geq 0, \qquad u_h(0) = u_{0h} \in V_h, \quad \dot{u}_h(0) = u_{1h} \in V_h. \tag{7.1.8}$$

In this context, for all $h > 0$, the observation operator naturally becomes

$$y_h(t) = B_h \dot{u}_h(t) = B\pi_h \dot{u}_h(t). \tag{7.1.9}$$

Note that, since $B \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$, this definition always makes sense since $\pi_h(V_h) \subset \mathcal{D}(A_0^{1/2})$.

We now make precise the assumptions we have, usually, on π_h, and which will be needed in our analysis. One easily checks that $\pi_h^* \pi_h = Id_{V_h}$. Besides, the injective map π_h describes the finite element approximation we have chosen. Especially, the vector space $\pi_h(V_h)$ approximates, in the sense given hereafter, the space $\mathcal{D}(A_0^{1/2})$: There exist $\theta > 0$ and $C_0 > 0$, such that for all $h > 0$,

$$\begin{cases} \left\| A_0^{1/2}(\pi_h \pi_h^* - I)\phi \right\|_X \leq C_0 \left\| A_0^{1/2} \phi \right\|_X, & \forall \phi \in \mathcal{D}(A_0^{1/2}), \\ \left\| A_0^{1/2}(\pi_h \pi_h^* - I)\phi \right\|_X \leq C_0 h^\theta \left\| A_0 \phi \right\|_X, & \forall \phi \in \mathcal{D}(A_0). \end{cases} \tag{7.1.10}$$

Note that in many applications, and in particular for A_0 the Laplace operator on a bounded domain with Dirichlet boundary conditions, estimates (7.1.10) are satisfied for $\theta = 1$.

We will not discuss convergence results for the numerical approximation schemes presented here, which are classical under assumption (7.1.10), and which can be found for instance in the textbook [33].

In the sequel, our goal is to obtain uniform admissibility and observability properties for (7.1.8)-(7.1.9) similar to (7.1.4) and (7.1.5) respectively.

Let us mention that similar questions have already been investigated in [19] for the 1d wave equation observed from the boundary on a 1d mesh. In [19], it has been proved that, for the space semi-discrete schemes derived from a finite element method for the 1d wave equation on uniform meshes (which is a particular instance of (7.1.1)), observability properties do not hold uniformly with respect to the discretization parameter, because of the presence of spurious high frequency solutions which do not travel. However, if the initial data are filtered in a suitable way, then observability inequalities hold uniformly with respect to the space discretization parameter. Actually, as pointed out by Otared Kavian in [41], it may even happen that unique continuation properties do not hold anymore in the discrete setting due to the existence of localized high-frequency solutions.

Therefore, it is natural to restrict ourselves to classes of suitable filtered initial data. For all $h > 0$, since A_{0h} is a self-adjoint positive definite matrix, the spectrum of A_{0h} is given by a sequence of positive eigenvalues

$$0 < \lambda_1^h \leq \lambda_2^h \leq \cdots \leq \lambda_{n_h}^h, \tag{7.1.11}$$

and normalized (in V_h) eigenvectors $(\Phi_j^h)_{1 \leq j \leq n_h}$. For any $s > 0$, we can now define, for each $h > 0$, the filtered space

$$\mathcal{C}_h(s) = \text{span}\left\{ \Phi_j^h \text{ such that the corresponding eigenvalue satisfies } |\lambda_j^h| \leq s \right\}.$$

We are now in position to state the main results of this article:

Theorem 7.1.3. *Let A_0 be a self-adjoint positive definite operator with compact resolvent and $B \in \mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$, with $\kappa < 1/2$. Assume that the maps $(\pi_h)_{h>0}$ satisfy property* (7.1.10)*. Set*

$$\sigma = \theta \min\Big\{2(1-2\kappa), \frac{2}{3}\Big\}. \tag{7.1.12}$$

Admissibility: *Assume that system* (7.1.1)-(7.1.3) *is admissible.*

Then, for any $\eta > 0$ and $T > 0$, there exists a positive constant $K_{T,\eta}$ such that, for any $h > 0$ small enough, any solution of (7.1.8) *with initial data*

$$(u_{0h}, u_{1h}) \in \mathcal{C}_h(\eta/h^\sigma)^2 \tag{7.1.13}$$

satisfies

$$\int_0^T \|B_h \dot{u}_h(t)\|_Y^2 \ dt \le K_{T,\eta}\Big(\Big\|A_{0h}^{1/2} u_{0h}\Big\|_h^2 + \|u_{1h}\|_h^2 \Big). \tag{7.1.14}$$

Observability: *Assume that system* (7.1.1)-(7.1.3) *is admissible and exactly observable.*

Then there exist $\epsilon > 0$, a time T^* and a positive constant $k_* > 0$ such that, for any $h > 0$ small enough, any solution of (7.1.8) *with initial data*

$$(u_{0h}, u_{1h}) \in \mathcal{C}_h(\epsilon/h^\sigma)^2 \tag{7.1.15}$$

satisfies

$$k_*\Big(\Big\|A_{0h}^{1/2} u_{0h}\Big\|_h^2 + \|u_{1h}\|_h^2 \Big) \le \int_0^{T^*} \|B_h \dot{u}_h(t)\|_Y^2 \ dt. \tag{7.1.16}$$

These two results are based on new spectral characterizations of admissibility and exact observability for (7.1.1)-(7.1.3).

To characterize the admissibility property, we use the results in [11, 10] to obtain a characterization based on a resolvent estimate and, later, on an interpolation property.

Our characterization of the exact observability property is deduced from the resolvent estimates in [24, 31, 37] and the wave packet characterization obtained in [31] and made more precise in [37]. However, our approach requires explicit estimates, which, to our knowledge, cannot be found in the literature. We thus propose a new proof of the wave packet spectral characterization in [31], which yields quantitative estimates. Again, we show that these criteria can be interpreted as interpolation properties.

The main idea, then, consists in proving uniform (in h) interpolation properties for the operators A_{0h} and B_h, in order to recover uniform (in h) admissibility and observability estimates. This idea is completely natural since the operators A_{0h} and B_h correspond to discrete versions of A_0 and B, respectively.

Theorem 7.1.3 has several important applications. As a straightforward corollary of the results in [11], one can derive observability properties for general fully discrete approximation schemes based on (7.1.8). Precise statements will be given in Section 7.5.

Besides, it also has relevant applications in control theory. Indeed, it implies that the Hilbert Uniqueness Method (see [23]) can be adapted in the discrete setting to provide efficient algorithms to compute approximations of exact controls for the continuous systems. This will be clarified in Section 7.6.

We will also present consequences of Theorem 7.1.3 to stabilization issues for space semi-discrete damped models. These will be deduced from [14], which addressed this problem in a very general setting which includes our models.

We finally investigate observability properties for space semi-discretizations of two other models, namely the wave equation (7.1.1) observed through $y(t) = Bu(t)$ instead of (7.1.3), for which we can adapt the method we have developed to prove Theorem 7.1.3, and the Schrödinger equation $i\dot{z} = A_0 z$, for which we can use Theorem 7.1.3 to derive observability properties, similarly as in [26].

Let us briefly comment some relative works. Similar problems have been extensively studied in the last decade for various space semi-discretizations of the 1d wave equation, see for instance the review article [41] and the references therein. The numerical schemes on uniform meshes provided by finite difference and finite element methods do not have uniform observability properties, whatever the time T is ([19]). This is due to high frequency waves which do not propagate, see [36, 25]. In other words, these numerical schemes create some spurious high-frequency wave solutions which are localized.

In this context, filtering techniques have been extensively developed. It has been proved in [19, 40] that filtering the initial data removes these spurious waves, and make possible uniform observability properties to hold. Other ways to filter these spurious waves exist, for instance using a wavelet filtering approach [28] or bi-grids techniques [15, 29]. However, to the best of our knowledge, these methods have been analyzed only for uniform grids in small dimensions (namely in 1d or 2d). Also note that these results prove uniform observability properties for larger classes of initial data than the ones stated here, but in more particular cases. Especially, Theorem 7.1.3 depends on neither the dimension nor the uniformity of the meshes.

Let us also mention that observability properties are equivalent to stabilization properties (see [17]), when the observation operator is bounded. Therefore, observability properties can be deduced from the literature in stabilization theory. Especially, we refer to the works [35, 34, 27, 12], which prove uniform stabilization results for damped space semi-discrete wave equations in 1d and 2d, discretized on uniform meshes using finite difference approximation schemes, in which a numerical viscosity term has been added. Again, these results are better than the ones derived here, but apply in the more restrictive context of 1d or 2d wave equations on uniform meshes. Similar results have also been proved in [32], but using a non trivial spectral condition on A_0, which reduces the scope of applications mainly to 1d equations.

To the best of our knowledge, there are very few paper dealing with nonuniform meshes. A first step in this direction can be found in the context of the stabilization of the 1d wave equation in [32]: Indeed, stabilization properties are equivalent (see [17]) to observability properties for the corresponding conservative systems. The results in [32] can therefore be applied to 1d wave equations on nonuniform meshes to derive uniform observability results within the class $\mathcal{C}_h(\epsilon/h^\theta)$ for $\epsilon > 0$ small enough. Though, they strongly use a spectral gap condition on the eigenvalues of the operator A_0, which does not hold for the wave operator in dimension higher than one.

Another result in this direction is presented in [9], in the context of the 1d wave equation discretized using a mixed finite element method as in [1, 5]. In [9], it is proved that observability properties for schemes derived from a mixed finite element method hold uniformly within a large class of nonuniform

meshes.

Also remark that observability and admissibility properties have been derived recently in [10] for Schrödinger type equations discretized using finite element methods. The results in [10] are strongly based on spectral characterizations of admissibility and observability properties for abstract systems. Actually, the present work follows the investigation in [10]. The main difference consists in the lack of simple spectral conditions for observability properties of wave type systems. This requires to design new spectral characterizations of admissibility and observability properties adapted to deal with systems (7.1.1)-(7.1.3).

We shall also mention recent works on spectral characterizations of exact observability for abstract conservative systems. We refer to [4, 26] for a very general approach of observability properties for conservative linear systems, which yields a necessary and sufficient resolvent condition for exact observability to hold. Let us also mention the articles [24, 31], which derived several spectral conditions for the exact observability of wave type equations. In [31], a spectral characterization of observability properties based on wave packets is also given. Our approach is inspired in all these works.

We also mention the recent article [11], which proved admissibility and observability estimates for general time semi-discrete conservative linear systems. In [11], a very general approach is given, which allows to deal with a large class of time discrete approximation schemes. This approach is based, as here, on a spectral characterization of exact observability for conservative linear systems (namely the one in [4, 26]). Later on in [14] (see also [13]), the stabilization properties of time discrete approximation schemes of damped systems were studied. In particular, [14] introduces time discrete schemes which are guaranteed to enjoy uniform (in the time discretization parameter) stabilization properties.

This article is organized as follows:

In Section 7.2, we present several spectral conditions for admissibility and exact observability properties of abstract systems (7.1.1)-(7.1.3). In Section 7.3, we prove Theorem 7.1.3. In Section 7.4, we give some precise examples of applications. In Section 7.5, we consider admissibility and exact observability properties for fully discrete approximation schemes of (7.1.8). In Section 7.6, we present applications of Theorem 7.1.3 to controllability issues. In Section 7.7, we also present applications to stabilization theory. In Section 7.8, we present similar results for two other different models, namely for the wave equation (7.1.1) observed through $y(t) = Bu(t)$ instead of (7.1.3), and for Schrödinger type systems. We finally present some further comments and open questions.

7.2 Spectral methods

This section recalls and presents various spectral characterizations of admissibility and observability for abstract systems (7.1.1)-(7.1.3). Here, we are not dealing with the discrete approximation schemes (7.1.8).

To state our results properly, we introduce some notations.

When dealing with the abstract system (7.1.1), it is convenient to introduce the spectrum of the operator A_0. Since A_0 is self-adjoint and positive definite, its spectrum is given by a sequence of positive eigenvalues

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq \cdots \to \infty, \tag{7.2.1}$$

and normalized (in X) eigenvectors $(\Phi_j)_{j\in\mathbb{N}^*}$.

Since some of the results below extend to a larger class of systems than (7.1.1)-(7.1.3), we also introduce the following abstract system

$$\begin{cases} \dot{z} = \mathcal{A}z, \quad t \geq 0, \\ z(0) = z_0 \in \mathfrak{X}, \end{cases} \qquad y(t) = Cz(t), \tag{7.2.2}$$

where $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset \mathfrak{X} \to \mathfrak{X}$ is an unbounded skew-adjoint operator with compact resolvent and $C \in \mathfrak{L}(\mathcal{D}(\mathcal{A}), Y)$. In particular, the spectrum of $\mathcal{A}$ is given by a sequence $(i\mu_j)_j$, where the constants μ_j are real and $|\mu_j| \to \infty$ when $j \to \infty$, and the corresponding eigenvectors (Ψ_j) (normalized in $\mathfrak{X}$) constitute an orthonormal basis of $\mathfrak{X}$. Note that systems of the form (7.1.1)-(7.1.3) indeed are particular instances of (7.2.2).

This section is organized as follows.

First, we present spectral characterizations for the admissibility of systems (7.2.2) and (7.1.1)-(7.1.3), based on the results in [10], which we will recall. Then we present spectral characterizations for the exact observability of systems (7.2.2) and (7.1.1)-(7.1.3), based on the articles [31, 24].

7.2.1 Characterizations of admissibility

Note that for (7.2.2), the admissibility inequality consists in the existence, for all $T > 0$, of a positive constant K_T such that any solution z of (7.2.2) with initial data $z_0 \in \mathcal{D}(\mathcal{A})$ satisfies

$$\int_0^T \|Cz(t)\|_Y^2 \; dt \leq K_T \|z_0\|_{\mathfrak{X}}^2 . \tag{7.2.3}$$

Resolvent characterization

The following result was proved in [10]:

Theorem 7.2.1. *Let $\mathcal{A}$ be a skew-adjoint operator on $\mathfrak{X}$ with compact resolvent and C be in $\mathfrak{L}(\mathcal{D}(\mathcal{A}), Y)$. The following statements are equivalent:*

1. *System (7.2.2) is admissible.*
2. *There exist $r > 0$ and $D > 0$ such that*

$$\forall \mu \in \mathbb{R}, \; \forall \; z = \sum_{l \in J_r(\mu)} c_l \Psi_l, \quad \|Cz\|_Y \leq D \|z\|_{\mathfrak{X}}, \tag{7.2.4}$$

where

$$J_r(\mu) = \{l \in \mathbb{N}, \; \textit{such that } |\mu_l - \mu| \leq r\}. \tag{7.2.5}$$

Besides, if (7.2.4) holds, then system (7.2.2) is admissible, and the constant K_T in (7.2.3) can be chosen as follows:

$$K_T = K_{\pi/2r} \left\lceil \frac{2rT}{\pi} \right\rceil, \quad \textit{with } K_{\pi/2r} = \frac{3\pi^4 D}{4r}. \tag{7.2.6}$$

3. *There exist positive constants m and M such that*

$$M^2 \|(\mathcal{A} - i\omega I)z\|_{\mathfrak{X}}^2 + m^2 \|z\|_{\mathfrak{X}}^2 \geq \|Cz\|_Y^2 , \quad \forall z \in \mathcal{D}(\mathcal{A}), \forall \omega \in \mathbb{R}. \tag{7.2.7}$$

Besides, if (7.2.7) *holds, then system* (7.2.2) *is admissible, and the constant* K_T *in* (7.2.3) *can be chosen as follows:*

$$K_T = K_1\lceil T\rceil, \quad \textit{with } K_1 = \frac{3\pi^3}{2}\sqrt{m^2 + M^2\frac{\pi^2}{4}}. \tag{7.2.8}$$

The proof of Theorem 7.2.1 in [10] is based on the previous work [11] which proves a wave packet characterization for the admissibility of systems (7.2.2).

Applications to Wave type equations

We now consider the abstract setting (7.1.1)-(7.1.3), which is a particular instance of (7.2.2) with $\mathfrak{X} = \mathcal{D}(A_0^{1/2}) \times X$, and

$$\mathcal{A} = \begin{pmatrix} 0 & Id \\ -A_0 & 0 \end{pmatrix}, \quad C = (\ 0\ ,\ B). \tag{7.2.9}$$

In particular, the domain of $\mathcal{A}$ simply is $\mathcal{D}(A_0) \times \mathcal{D}(A_0^{1/2})$ and the conditions $C \in \mathfrak{L}(\mathcal{D}(\mathcal{A}), Y)$ and $B \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$ are equivalent.

Theorem 7.2.2. *Let* A_0 *be a self-adjoint positive definite operator on* X *with compact resolvent and* B *be in* $\mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$. *The following statements are equivalent:*

1. *System* (7.1.1)-(7.1.3) *is admissible in the sense of* (7.1.4)*;*

2. *There exist positive constants* m *and* M *such that:*

$$\omega^2 \|B\phi\|_Y^2 \le M^2 \left\|(A_0 - \omega^2 I)\phi\right\|_X^2 + m^2\Big(|\omega|^2 \|\phi\|_X^2 + \left\|A_0^{1/2}\phi\right\|_X^2\Big), \quad \forall\omega \in \mathbb{R}, \forall\phi \in \mathcal{D}(A_0). \tag{7.2.10}$$

Besides, if (7.2.10) *holds, then system* (7.1.1)-(7.1.3) *is admissible, and the constant* K_T *in* (7.1.4) *can be chosen as follows:*

$$K_T = K_{\pi/2}\Big\lceil\frac{2T}{\pi}\Big\rceil, \quad \textit{with } K_{\pi/2} = \frac{3\pi^4}{4\sqrt{2}}\sqrt{9M^2 + 5m^2}. \tag{7.2.11}$$

3. *There exist positive constants* α, β *and* γ *such that*

$$\left\|A_0^{1/2}\phi\right\|_X^2 + \alpha^2 \|B\phi\|_Y^2 \le \|\phi\|_X \sqrt{\|A_0\phi\|_X^2 + \beta^2 \left\|A_0^{1/2}\phi\right\|_X^2} + \gamma^2 \|\phi\|_Y^2, \quad \forall\phi \in \mathcal{D}(A_0). \tag{7.2.12}$$

Besides, if (7.2.12) *holds, then system* (7.1.1)-(7.1.3) *is admissible, and the constant* K_T *in* (7.1.4) *can be chosen as follows:*

$$K_T = K_{\pi/2}\Big\lceil\frac{2T}{\pi}\Big\rceil, \quad \textit{with } K_{\pi/2} = \frac{9\pi^4}{8\alpha}\sqrt{1 + \frac{5}{9}\sup\{\beta^2, 2\gamma^2\}}. \tag{7.2.13}$$

Proof. Let us first prove that statements 1 and 2 are equivalent.

Assume that system (7.1.1)-(7.1.3) is admissible. Then, from Theorem 7.2.1, there exist positive constants m and M such that (7.2.7) holds:

$$\|Bv\|_Y^2 \le M^2\Big(\left\|A_0^{1/2}\Big(v - i\omega u\Big)\right\|_X^2 + \|A_0 u + i\omega v\|_X^2\Big) + m^2\Big(\left\|A_0^{1/2}u\right\|_X^2 + \|v\|_X^2\Big),$$
$$\forall\omega \in \mathbb{R},\ \forall(u, v) \in \mathcal{D}(A_0) \times \mathcal{D}(A_0^{1/2}).$$

Taking $\phi \in \mathcal{D}(A_0)$, setting $u = \phi$ and $v = i\omega\phi$ in this last expression, we obtain (7.2.10).

Assume now that (7.2.10) holds. To prove the admissibility of (7.1.1)-(7.1.3), we use the wave packet criterion (7.2.4). Before going into the proof, let us recall that the spectrum $(i\mu_j, \Psi_j)_{j\in\mathbb{Z}^*}$ of $\mathcal{A}$ can be deduced from the spectrum $(\lambda_j, \Phi_j)_{j\in\mathbb{N}^*}$ of A_0 as follows:

$$\mu_{\pm j} = \pm\sqrt{\lambda_j}, \quad j \in \mathbb{N}^*, \qquad \Psi_{\pm j} = \frac{1}{\sqrt{2}} \begin{pmatrix} \dfrac{\pm 1}{i\sqrt{\lambda_j}}\Phi_j \\ \Phi_j \end{pmatrix}, \quad j \in \mathbb{N}^*. \tag{7.2.14}$$

Now, let ω_0 be a real number, take $r = 1$ and consider a wave packet

$$z = \sum_{l\in J_1(\omega_0)} c_l \Psi_l = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}. \tag{7.2.15}$$

For $|\omega_0| \geq 1$, applying (7.2.10) to z_2 for $\omega = \omega_0$, we get

$$\|Cz\|_Y^2 = \|Bz_2\|_Y^2 \leq \frac{M^2}{\omega_0^2} \left\|(A_0 - \omega_0^2 I)z_2\right\|_X^2 + m^2 \|z_2\|_X^2 + \frac{m^2}{\omega_0^2} \left\|A_0^{1/2} z_2\right\|_X^2 .$$

But, using the explicit expansion of z_2, one easily checks that

$$\left\|(A_0 - \omega_0^2 I)z_2\right\|_X^2 = \frac{1}{2} \sum_{|\mu_j - \omega_0|\leq 1} |\mu_j + \omega_0|^2 |\mu_j - \omega|^2 c_j^2 \leq 2(|\omega_0| + 1)^2 \|z_2\|_X^2 \leq 8|\omega_0|^2 \|z_2\|_X^2 ,$$

and

$$\left\|A_0^{1/2} z_2\right\|_X^2 = \frac{1}{2} \sum_{|\mu_j - \omega_0|\leq 1} |c_j|^2 \mu_j^2 \leq 2\omega_0^2 \|z_2\|_X^2 ,$$

since $|\omega_0| \geq 1$.

Using $\|z\|_{\mathfrak{X}}^2 = 2\|z_2\|_X^2$, we then obtain

$$\|Cz\|_Y \leq \sqrt{8M^2 \|z_2\|_X^2 + 3m^2 \|z_2\|_X^2} \leq \left(\sqrt{4M^2 + \frac{3}{2}m^2}\right) \|z\|_X . \tag{7.2.16}$$

We now need to prove a similar estimate for z as in (7.2.15) with $|\omega_0| < 1$. In this case, we apply (7.2.10) for $\omega = 1$, and as before, we obtain

$$\begin{aligned} \|Cz\|_Y^2 &\leq M^2 \|(A_0 - I)z_2\|_X^2 + m^2 \Big(\|z_2\|_X^2 + \left\|A_0^{1/2} z_2\right\|_X^2 \Big) \\ &\leq 9M^2 \|z_2\|_X^2 + 5m^2 \|z_2\|_X^2 = \Big(\frac{9M^2 + 5m^2}{2}\Big) \|z\|_{\mathfrak{X}}^2 , \end{aligned} \tag{7.2.17}$$

where we used that for z as in (7.2.15), $\|z\|_{\mathfrak{X}}^2 = 2\|z_2\|_X^2$ and, when $|\omega_0| < 1$,

$$\|(A_0 - I)z_2\|_X^2 \leq 9\|z_2\|_X^2 , \qquad \left\|A_0^{1/2} z_2\right\|_X^2 \leq 4\|z_2\|_X^2 .$$

Combining (7.2.16) and (7.2.17), we get (7.2.4) for any wave packet z with $r = 1$ and

$$D = \sqrt{\frac{9M^2 + 5m^2}{2}}.$$

The estimate (7.2.11) then follows from (7.2.6).

We now prove that statements 2 and 3 are equivalent. As in [10], the idea consists in noticing that (7.2.10) is equivalent to the nonnegativity of the quadratic form (in ω^2)

$$\omega^4 \|\phi\|_X^2 - 2\omega^2\Big(\left\|A_0^{1/2}\phi\right\|_X^2 + \frac{1}{2M^2}\|B\phi\|_Y^2 - \frac{m^2}{2M^2}\|\phi\|_X^2 \Big) + \|A_0\phi\|_X^2 + \frac{m^2}{M^2}\left\|A_0^{1/2}\phi\right\|_X^2,$$

which is equivalent to (as one can easily check by studying the positivity of the quadratic form $x \mapsto ax^2 - 2bx + c$ on $\mathbb{R}_+$ for $a > 0$ and $c > 0$):

$$\left\|A_0^{1/2}\phi\right\|_X^2 + \frac{1}{2M^2}\|B\phi\|_Y^2 - \frac{m^2}{2M^2}\|\phi\|_X^2 \le \|\phi\|_X \sqrt{\|A_0\phi\|_X^2 + \frac{m^2}{M^2}\left\|A_0^{1/2}\phi\right\|_X^2},$$

or, equivalently, (7.2.12) with

$$\alpha = \frac{1}{\sqrt{2}M}, \quad \beta = \frac{m}{M}, \quad \gamma = \frac{m}{\sqrt{2}M}.$$

Conversely, if (7.2.12) holds, then we can take

$$M = \frac{1}{\sqrt{2}\alpha}, \quad m = \frac{\sup\{\beta, \sqrt{2}\gamma\}}{\sqrt{2}\alpha}$$

in (7.2.10), and this completes the proof of Theorem 7.2.2. □

7.2.2 Characterizations of observability

We first recall the following criterion for the observability of (7.1.1)-(7.1.3):

Theorem 7.2.3 ([31], see also [24]). *Let A_0 be a self-adjoint positive definite operator on X with compact resolvent and $B \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$. Assume that system* (7.1.1)-(7.1.3) *is admissible in the sense of* (7.1.4).

Then system (7.1.1)-(7.1.3) *is exactly observable if and only if there exist positive constants m and M such that*

$$M^2 \left\|(A_0 - \omega^2 I)u\right\|_X^2 + m^2\|\omega B u\|_Y^2 \ge \|\omega u\|_X^2, \quad \forall u \in \mathcal{D}(A_0),\ \forall \omega \in \mathbb{R}. \tag{7.2.18}$$

Note that Theorem 7.2.3 does not provide precise estimates on the constants in (7.1.5). This is due to the proof of this theorem, based on Theorem 7.2.4 below.

Before stating Theorem 7.2.4, note that for (7.2.2), the exact observability property consists in the existence of a time T and a positive constant k_T such that any solution of (7.2.2) with initial data $z_0 \in \mathcal{D}(\mathcal{A})$ satisfies

$$k_T \|z_0\|_{\mathfrak{X}}^2 \le \int_0^T \|Cz(t)\|_Y^2 \ dt. \tag{7.2.19}$$

Theorem 7.2.4 ([31]). *Let $\mathcal{A}$ be a skew-adjoint operator on $\mathfrak{X}$ with compact resolvent, and $C \in \mathfrak{L}(\mathcal{D}(\mathcal{A}), Y)$. Assume that system* (7.2.2) *is admissible in the sense of* (7.2.3).

Then system (7.2.2) *is exactly observable if and only if*

$$\left\{\begin{array}{l} \text{There exist } \alpha>0 \text{ and } \beta>0 \text{ such that} \\ \text{for all } \mu\in\mathbb{R} \text{ and for all } z=\displaystyle\sum_{l\in J_\alpha(\mu)} c_l\Psi_l: \quad \|Cz\|_Y \geq \beta\,\|z\|_{\mathfrak{X}}, \end{array}\right. \tag{7.2.20}$$

where $J_\alpha(\mu)$ *is as in* (7.2.5). *Besides, if system* (7.2.2) *is admissible and exactly observable in time* T^*, *then one can choose*

$$\alpha=\frac{1}{T^*}\sqrt{\frac{k_{T^*}}{(2K_{T^*})}}, \qquad \beta=\frac{2}{\sqrt{k_{T^*}}}.$$

Here again, no estimates on the constants entering in (7.2.19) are given. Though, a non-explicit constant is given in [37], but which makes the use of Theorems 7.2.3 and 7.2.4 delicate for the applications we have in mind, which involve sequences of operators.

Therefore, we present below a new proof of the fact that (7.2.20) implies the exact observability of system (7.2.2), which yields explicit estimates in Theorem 7.2.3 as well. These estimates are crucial in our setting.

A refined version of Theorem 7.2.4

Theorem 7.2.5. *Let* $\mathcal{A}$ *be a skew-adjoint operator on* $\mathfrak{X}$ *with compact resolvent, and* $C\in\mathfrak{L}(\mathcal{D}(\mathcal{A}),Y)$. *Assume that system* (7.2.2) *is admissible in the sense of* (7.2.3).

If (7.2.20) *holds, then system* (7.2.2) *is exactly observable in any time* $T>T^*$, *for*

$$T^*=\frac{2e}{\alpha}\Big(\frac{\pi}{4}\ln(L)+\frac{3\pi}{4}\Big)^{1+1/\ln(L)}, \tag{7.2.21}$$

where

$$L=\frac{2\pi}{3}\frac{K_{1/\alpha}\alpha}{\beta^2}. \tag{7.2.22}$$

Besides, the constant k_T *in* (7.2.19) *can be chosen as*

$$k_T=\frac{\pi\beta^2}{\alpha}\Big(1-\Big(\frac{T^*}{T}\Big)^{2n^*-1}\Big), \quad \text{where } n^*=\Big\lceil\frac{1}{2}\Big(\ln(L)+1\Big)\Big\rceil. \tag{7.2.23}$$

Remark 7.2.6. Note that the constant L is always greater than $2\pi/3$, and then $\ln(L)>0$. Indeed, one can consider the solution $z(t)=\exp(i\mu_1 t)\Psi_1$ of (7.2.2), for which we get

$$\int_0^{1/\alpha}\|Cz(t)\|_Y^2\ dt\leq K_{1/\alpha},$$

as a consequence of the admissibility of system (7.2.2), and

$$\int_0^{1/\alpha}\|Cz(t)\|_Y^2\ dt\geq\int_0^{1/\alpha}\beta^2\,\|z(t)\|_{\mathfrak{X}}^2\ dt\geq\frac{\beta^2}{\alpha},$$

which follows from (7.2.20).

Proof. Set $z_0 \in \mathfrak{X}$, and denote by $z(t)$ the solution of (7.2.2) with initial data z_0. Set

$$g(t) = \chi(t) z(t), \tag{7.2.24}$$

where $\chi : \mathbb{R} \to \mathbb{R}$ is a function whose Fourier transform is smooth and satisfies

$$\text{Supp } \hat{\chi} \subset (-\alpha, \alpha). \tag{7.2.25}$$

Note that these conditions imply that χ is in the Schwartz class $\mathcal{S}(\mathbb{R})$ and therefore g and $\hat{g}$ both are in $L^2(\mathbb{R}, \mathfrak{X})$.

We expand z_0 and $z(t)$ on the basis Ψ_j:

$$z_0 = \sum_j a_j \Psi_j, \qquad z(t) = \sum_j a_j \exp(i\mu_j t)\Psi_j. \tag{7.2.26}$$

One then easily check that

$$\hat{g}(\omega) = \sum_j a_j \hat{\chi}(\omega - \mu_j)\Psi_j. \tag{7.2.27}$$

Especially, due to the property (7.2.25), for all ω, $\hat{g}(\omega)$ is a wave packet and therefore (7.2.20) implies

$$\beta^2 \left\| \hat{g}(\omega) \right\|_{\mathfrak{X}}^2 \leq \left\| C\hat{g}(\omega) \right\|_Y^2 . \tag{7.2.28}$$

Note that, due to the explicit expansion (7.2.27), we have the identity

$$\left\| \hat{g}(\omega) \right\|_{\mathfrak{X}}^2 = \sum_j |a_j|^2 |\hat{\chi}(\omega - \mu_j)|^2.$$

Then, integrating (7.2.28) in ω, and using Parseval's identity on the right hand-side of (7.2.28), one easily obtains

$$\beta^2 \Big(\int \hat{\chi}^2(\omega) d\omega \Big) \Big(\sum_j |a_j|^2 \Big) \leq \int_{\mathbb{R}} \left\| Cg(t) \right\|_Y^2 \, dt = \int_{\mathbb{R}} \chi^2(t) \left\| Cz(t) \right\|_Y^2 \, dt, \tag{7.2.29}$$

where the last equality comes from the definition (7.2.24) of g.

Now, since $\chi \in \mathcal{S}(\mathbb{R})$, we know that for each $n \in \mathbb{N}^*$, there exists a constant c_n such that

$$|\chi(t)| \leq c_n \frac{1}{|t|^n}, \qquad \forall t \neq 0. \tag{7.2.30}$$

Hence, for any time $T > 0$, using the admissibility in time T, we obtain that

$$\begin{aligned} \int_{\mathbb{R}} \chi^2(t) \left\| Cz(t) \right\|_Y^2 \, dt &\leq \int_{-T}^{T} \chi^2(t) \left\| Cz(t) \right\|_Y^2 \, dt + 2\Big(\sum_{k=1}^{\infty} \frac{1}{(kT)^{2n}} \Big) c_n^2 K_T \left\| z_0 \right\|_{\mathfrak{X}}^2 \\ &\leq \int_{-T}^{T} \chi^2(t) \left\| Cz(t) \right\|_Y^2 \, dt + \frac{\pi^2}{3} c_n^2 \frac{1}{T^{2n}} K_T \left\| z_0 \right\|_{\mathfrak{X}}^2 . \end{aligned} \tag{7.2.31}$$

We therefore need to estimate c_n in (7.2.30). Of course, one cannot expect it to be uniform in the whole Schwartz class, and it will strongly depend on the choice of χ. By a scaling argument, we assume without loss of generality that

$$\chi(t) = \psi(t\alpha), \quad \hat{\chi}(\omega) = \frac{1}{\alpha} \hat{\psi}\Big(\frac{t}{\alpha} \Big), \tag{7.2.32}$$

where ψ belongs to the Schwartz class and satisfies

$$\text{Supp } \hat{\psi} \subset (-1,1). \tag{7.2.33}$$

Remark that integrations by parts then yield:

$$\psi(t) = \frac{1}{\sqrt{2\pi}} \int \hat{\psi}(\omega) \exp(i\omega t)\ d\omega = \frac{1}{\sqrt{2\pi}(it)^n} \int \hat{\psi}^{(n)} \exp(i\omega t)\ d\omega.$$

Thus we obtain the following decay estimate on ψ:

$$|\psi(t)| \leq \frac{1}{\sqrt{\pi}} \frac{1}{|t|^n} \Big(\int |\hat{\psi}^{(n)}|^2\ d\omega \Big)^{1/2}, \quad t \in \mathbb{R}^*.$$

Therefore χ satisfies

$$|\chi(t)| \leq \frac{1}{\sqrt{\pi}} \Big(\frac{1}{\alpha|t|} \Big)^n \Big(\int |\hat{\psi}^{(n)}|^2\ d\omega \Big)^{1/2}, \quad t \in \mathbb{R}^*. \tag{7.2.34}$$

Also note that the L^∞ norm of χ can be estimated by the L^2 norm of ψ:

$$|\chi(t)| = |\psi(t\alpha)| = \Big| \frac{1}{\sqrt{2\pi}} \int \hat{\psi}(\omega) \exp(i\omega t\alpha)\ d\omega \Big| \leq \frac{1}{\sqrt{\pi}} \Big(\int |\hat{\psi}|^2\ d\omega \Big)^{1/2}.$$

Besides, since one easily checks that

$$\int |\hat{\chi}|^2 d\omega = \frac{1}{\alpha} \int |\hat{\psi}|^2 d\omega,$$

we obtain from (7.2.29), (7.2.31) and (7.2.34) that

$$\begin{aligned} \Big(\frac{1}{\alpha} \beta^2 \int |\hat{\psi}|^2 d\omega - K_T \frac{\pi}{3} \Big(\frac{1}{\alpha T} \Big)^{2n} \int |\hat{\psi}^{(n)}|^2\ d\omega \Big) \|z_0\|_{\mathfrak{X}}^2 \\ \leq \int_{-T}^{T} \chi^2(t) \|Cz(t)\|_Y^2\ dt \leq \frac{1}{\pi} \Big(\int |\hat{\psi}|^2 d\omega \Big) \int_{-T}^{T} \|Cz(t)\|_Y^2\ dt. \end{aligned} \tag{7.2.35}$$

Let us now assume that $T\alpha$ is strictly greater than 1. In this case, we can estimate K_T by

$$K_T \leq K_{1/\alpha}(1 + T\alpha) \leq 2K_{1/\alpha} T\alpha. \tag{7.2.36}$$

Therefore, to guarantee that the left hand side of (7.2.35) is positive, we only need $T\alpha > 1$ and

$$T\alpha > \inf_n \left\{ \left(\frac{2\pi K_{1/\alpha}\alpha}{3\beta^2} \right)^{1/(2n-1)} \inf_{\hat{\psi} \in \mathcal{D}(-1,1)} \left\{ \frac{\left\| \hat{\psi}^{(n)} \right\|_{L^2}^2}{\left\| \hat{\psi} \right\|_{L^2}^2} \right\}^{1/(2n-1)} \right\}. \tag{7.2.37}$$

We now derive an estimate on the following coefficient:

$$\gamma_n = \left(\inf_{\phi \in \mathcal{D}(-1,1)} \frac{\left\| \phi^{(n)} \right\|_{L^2}^2}{\|\phi\|_{L^2}^2} \right)^{1/2n}. \tag{7.2.38}$$

Lemma 7.2.7. *We have the following estimate:*

$$\gamma_n \leq \frac{n\pi}{2}, \quad \forall n \in \mathbb{N}^*. \tag{7.2.39}$$

Proof of Lemma 7.2.7. Set $n \in \mathbb{N}^*$. Let us consider

$$\phi_n(x) = \sin\Big(\frac{\pi}{2}(x+1)\Big)^n,$$

which belongs to $H_0^n(-1,1)$, and which, by density, is admissible as a test function in the infimum (7.2.38).

Consider the Fourier development of ϕ_n, which takes the form

$$\phi_n(x) = \sum_{k=-n}^{n} a_k \exp\Big(\frac{ik\pi x}{2}\Big).$$

Then we have

$$\Big\|\phi_n^{(n)}\Big\|_{L^2}^2 = \sum_{k=-n}^{n} |a_k|^2\Big(\frac{k\pi}{2}\Big)^{2n} \le \Big(\frac{n\pi}{2}\Big)^{2n} \sum_{k=-n}^{n} |a_k|^2 \le \Big(\frac{n\pi}{2}\Big)^{2n} \|\phi_n\|_{L^2}^2 .$$

Lemma 7.2.7 follows. □

Therefore, using the constant L introduced in (7.2.22), we need to minimize on $\mathbb{N}$

$$f(n) = L^{1/(2n-1)}\Big(\frac{n\pi}{2}\Big)^{2n/(2n-1)}.$$

In $\mathbb{R}$, the infimum is attained in $\tilde{n}$ such that

$$2\tilde{n} - 1 = \ln(L) + \ln\Big(\frac{\tilde{n}\pi}{2}\Big).$$

Therefore, a good approximation of the minimizer of f on $\mathbb{N}$ is given by n^* as in (7.2.23), for which we have

$$f(n^*) \le e\Big(\frac{\pi}{4}\ln(L) + \frac{3\pi}{4}\Big)^{1+1/\ln(L)} = \frac{T^*\alpha}{2}.$$

Choosing $n = n^*$ in (7.2.35) and using (7.2.36), we obtain that

$$\int_{-T}^{T} \|Cz(t)\|_Y^2 \ dt \ge \frac{\pi\beta^2}{\alpha}\Big(1 - \frac{L}{(T\alpha)^{2n^*-1}}\Big(\frac{n^*\pi}{2}\Big)^{2n^*}\Big)\|z_0\|_{\mathfrak{X}}^2 \ge \frac{\pi\beta^2}{\alpha}\Big(1 - \Big(\frac{T^*}{2T}\Big)^{2n^*-1}\Big)\|z(-T)\|_{\mathfrak{X}}^2 .$$

Since the semi-group generated by (7.2.2) is a bijective isometry on $\mathfrak{X}$, this gives, for any $z_0 \in \mathfrak{X}$,

$$\int_{0}^{2T} \|Cz(t)\|_Y^2 \ dt \ge \frac{\pi\beta^2}{\alpha}\Big(1 - \Big(\frac{T^*}{2T}\Big)^{2n^*-1}\Big)\|z_0\|_{\mathfrak{X}}^2 .$$

This completes the proof of Theorem 7.2.5 by replacing $2T$ by T. □

Remark 7.2.8. The time estimate we obtain with this strategy strongly depends on the estimate (7.2.39) on γ_n defined in (7.2.38). To our knowledge, though this problem might seem classical, there is no precise bounds on γ_n. Especially, note that if we were able to prove that $\liminf_{n\to\infty} \gamma_n = \aleph < \infty$, then condition (7.2.37) would simply become $T\alpha > 2\aleph$, which would be very similar to the assumptions of Ingham's Lemma [20] (see also [38] on the completeness of non harmonic Fourier series in $L^2(0,T)$).

Application to Theorem 7.2.3 We can now make precise the estimates in Theorem 7.2.3.

Theorem 7.2.9. *Under the assumptions of Theorem 7.2.3, assume that* (7.2.18) *holds. Also assume that the first eigenvalue of A_0 satisfies $\lambda_1 \geq \gamma > 0$.*

Set

$$\alpha = \min\Big\{\frac{1}{3\sqrt{2}M}, \frac{\sqrt{\gamma}}{2}\Big\}, \quad \beta = \frac{1}{2m}. \tag{7.2.40}$$

Then system (7.1.1)-(7.1.3) *is exactly observable in any time $T > T^*$, for T^* as in* (7.2.21). *Besides, the constant k_T in* (7.1.5) *can be chosen as in* (7.2.23) *as an explicit expression of T, m, M, γ, and the admissibilty constant $K_{1/\alpha}$.*

Proof. The proof combines the estimates given in Theorem 7.2.5 with the following proposition:

Proposition 7.2.10. *Let $\mathcal{A}$, A_0, B and C be related as in* (7.2.9). *Under the assumptions of Theorem 7.2.9, setting α and β as in* (7.2.40), *the following wave packet estimates holds: For all $\omega \in \mathbb{R}$,*

$$\forall z = \sum_{l \in J_\alpha(\omega)} c_l \Psi_l, \quad \beta \|z\|_{\mathfrak{X}} \leq \|Cz\|_Y . \tag{7.2.41}$$

Proof. First, we remark that, since $\alpha \leq \sqrt{\gamma}/2$, when $|\omega| < \sqrt{\gamma}/2$, the set $J_\alpha(\omega)$ is empty. Therefore we only need to prove (7.2.41) for $|\omega| \geq \sqrt{\gamma}/2$, or, due to the explicit form of the spectrum and the relations (7.2.14), only for $\omega \geq \sqrt{\gamma}/2$.

Given $\omega \geq \sqrt{\gamma}/2$, let z be a wave packet

$$z = \sum_{l \in J_\alpha(\omega)} c_l \Psi_l = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

for which we have

$$z_2 = \frac{1}{\sqrt{2}} \sum_{l \in J_\alpha(\omega)} c_l \Phi_l, \quad \text{and} \quad \|z_2\|_X^2 = \frac{1}{2} \sum_{l \in J_\alpha(\omega)} |c_l|^2 = \frac{1}{2} \|z\|_{\mathfrak{X}}^2 .$$

Applying (7.2.18) to z_2, we obtain

$$\frac{1}{2}\|z\|_{\mathfrak{X}}^2 = \|z_2\|_X^2 \leq m^2 \|Bz_2\|_Y^2 + \frac{M^2}{\omega^2}\big\|(A_0 - \omega^2)z_2\big\|_X^2 = m^2\|Cz\|_Y^2 + \frac{M^2}{\omega^2}\big\|(A_0 - \omega^2)z_2\big\|_X^2 .$$

But the last term satisfies

$$\begin{aligned}
\big\|(A_0 - \omega^2)z_2\big\|_X^2 &= \frac{1}{2}\sum_{l \in J_\alpha(\omega)} |c_l|^2 \Big(\mu_l^2 - \omega^2\Big)^2 \\
&\leq 2 \sum_{l \in J_\alpha(\omega)} |c_l|^2 \Big(\frac{\mu_l + \omega}{2}\Big)^2 (\mu_l - \omega)^2 \\
&\leq 2\alpha^2 \sum_{l \in J_\alpha(\omega)} |c_l|^2 \Big(\omega + \frac{\alpha}{2}\Big)^2 \leq \frac{9}{2}\alpha^2\omega^2 \|z\|_{\mathfrak{X}}^2 ,
\end{aligned}$$

where we used that, for $l \in J_\alpha(\omega)$ with $\omega \geq \alpha > 0$, we have $\mu_l \leq \omega + \alpha \leq 2\omega$.

With the choice of α given in (7.2.40), we thus obtain

$$\|z\|_{\mathfrak{X}}^2 \leq 4m^2 \|Cz\|_Y^2 ,$$

and the result follows. □

Theorem 7.2.9 then directly follows from Theorem 7.2.5. □

An interpolation criterion We finally deduce another criterion for the observability of wave type equations (7.1.1)-(7.1.3).

Theorem 7.2.11. *Let $A_0 : \mathcal{D}(A_0) \subset X \to X$ be a self adjoint positive definite operator with compact resolvent, and let $B \in \mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$ be an admissible observation operator for* (7.1.1)-(7.1.3)*. Assume that there exists a positive constant γ such that the first eigenvalue of A_0 is greater than γ.*

If system (7.1.1)-(7.1.3) *is exactly observable, there exist positive constants α and β such that*

$$\left\|A_0^{1/2}u\right\|_X^2 \leq \|u\|_X \ \|A_0u\|_X + \alpha^2 \|Bu\|_Y^2 - \beta^2 \|u\|_X^2, \quad \forall u \in \mathcal{D}(A_0). \tag{7.2.42}$$

Conversely, if (7.2.42) *holds, then system* (7.1.1)-(7.1.3) *is exactly observable: There exists a time T^*, which only depends on α, β, γ and the admissibility constants, such that for any time $T > T^*$, there exists a positive constant $k_T > 0$, which only depends on T, α, β, γ and the admissibility constants, such that* (7.1.5) *holds for any solution of* (7.1.1)*.*

Proof. The proof is based on Theorem 7.2.9. In view of Theorem 7.2.9, it is sufficient to prove that conditions (7.2.42) and (7.2.18) are equivalent.

Remark that (7.2.18) can be rewritten as

$$\omega^4 \|u\|_X^2 - 2\omega^2\Big(\left\|A_0^{1/2}u\right\|_X^2 - \frac{m^2}{2M^2}\|Bu\|_Y^2 + \frac{1}{2M^2}\|u\|_X^2 \Big) + \|A_0u\|_X^2 \geq 0, \quad \forall u \in \mathcal{D}(A_0),\ \forall \omega \in \mathbb{R}. \tag{7.2.43}$$

Since this last expression simply is a quadratic expression in $\omega^2 \in \mathbb{R}_+$, then the nonnegativity of (7.2.43) is equivalent to (again, this follows from the study of the polynomial function $x \mapsto ax^2 - 2bx + c$ on $\mathbb{R}_+$):

$$\left\|A_0^{1/2}u\right\|_X^2 - \frac{m^2}{2M^2}\|Bu\|_Y^2 + \frac{1}{2M^2}\|u\|_X^2 \leq \|u\|_X \|A_0u\|_X, \quad \forall u \in \mathcal{D}(A_0). \tag{7.2.44}$$

This last inequality obviously is equivalent to (7.2.42), with $\alpha = m/\sqrt{2}M$ and $\beta = 1/\sqrt{2}M$.

Conversely, if (7.2.42) holds, inequality (7.2.18) holds for any $u \in \mathcal{D}(A_0)$ and $\omega \in \mathbb{R}$ by taking $m = \alpha/\beta$ and $M = 1/\sqrt{2}\beta$.

Theorem 7.2.11 then follows from Theorem 7.2.9. □

7.3 Proof of Theorem 7.1.3

In this Section, we prove Theorem 7.1.3. Below, we assume that the assumptions of Theorem 7.1.3 are satisfied.

For convenience, since B is assumed to belong to $\mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$, we introduce a constant K_B such that

$$\|B\phi\|_Y \leq K_B \|A_0^\kappa \phi\|_X, \quad \forall \phi \in \mathcal{D}(A_0^\kappa).$$

7.3.1 Admissibility

Proof of Theorem 7.1.3: Admissibility. Assume that system (7.1.1)-(7.1.3) is admissible. Then, from Theorem 7.2.2, there exist positive constants α, β and γ such that (7.2.12) holds.

Again using Theorem 7.2.2, it is sufficient to prove the existence of positive constants α_*, β_* and γ_* such that for any $h > 0$,

$$\left\|A_{0h}^{1/2}u_h\right\|_h^2 + \alpha_*^2 \left\|B_h u_h\right\|_Y^2 \leq \|u_h\|_h \sqrt{\|A_{0h}u_h\|_h^2 + \beta_*^2 \left\|A_{0h}^{1/2}u_h\right\|_h^2} + \gamma_*^2 \|u_h\|_h^2 , \quad \forall u_h \in \mathcal{C}_h(\eta/h^\sigma). \tag{7.3.1}$$

For $h > 0$, we fix $u_h \in \mathcal{C}_h(\eta/h^\sigma)$. Similarly as in [10], we introduce $U_h \in \mathcal{D}(A_0)$, defined by

$$A_0 U_h = \pi_h \pi_h^* A_0 \pi_h u_h = \pi_h A_{0h} u_h. \tag{7.3.2}$$

This defines an element $U_h \in \mathcal{D}(A_0)$, which we expect to be close to u_h.

Since $U_h \in \mathcal{D}(A_0)$, inequality (7.2.12) applies:

$$\left\|A_0^{1/2}U_h\right\|_X^2 + \alpha^2 \left\|BU_h\right\|_Y^2 \leq \|U_h\|_X \sqrt{\|A_0 U_h\|_X^2 + \beta^2 \left\|A_0^{1/2}U_h\right\|_X^2} + \gamma^2 \|U_h\|_X^2 . \tag{7.3.3}$$

The computations below are the same as in [10]. For convenience, we recall them.

From the definition (7.3.2) of U_h, we have

$$\|A_{0h}u_h\|_h = \|\pi_h A_{0h} u_h\|_X = \|A_0 U_h\|_X . \tag{7.3.4}$$

We now estimate $U_h - \pi_h u_h$. Using (7.1.7) and (7.3.2), for all $\phi \in \mathcal{D}(A_0)$, we have:

$$\begin{aligned} < U_h, A_0\phi >_X = < A_0 U_h, \phi >_X = < \pi_h A_{0h} u_h, \phi >_X \\ = < \pi_h \pi_h^* A_0 \pi_h u_h, \phi >_X = < A_0^{1/2}\pi_h u_h, A_0^{1/2}\pi_h \pi_h^* \phi >_X . \end{aligned} \tag{7.3.5}$$

In particular, this implies

$$\begin{aligned} < (u_h - \pi_h u_h), A_0\phi >_X &= < U_h, A_0\phi >_X - < A_0^{1/2}\pi_h u_h, A_0^{1/2}\phi >_X \\ &= < A_0^{1/2}\pi_h u_h, A_0^{1/2}(\pi_h\pi_h^* - I)\phi >_X . \end{aligned}$$

Using (7.1.10) and the invertibility of A_0, we obtain

$$\begin{aligned} \|U_h - \pi_h u_h\|_X &= \sup_{\substack{\phi\in\mathcal{D}(A_0),\\ \|A_0\phi\|_X=1}} \left\{ < (U_h - \pi_h u_h), A_0\phi >_X \right\} \\ &\leq \left\|A_0^{1/2}\pi_h u_h\right\|_X \sup_{\substack{\phi\in\mathcal{D}(A_0),\\ \|A_0\phi\|_X=1}} \left\|A_0^{1/2}(\pi_h\pi_h^* - I)\phi\right\|_X \\ &\leq C_0 h^\theta \left\|A_0^{1/2}\pi_h u_h\right\|_X . \end{aligned}$$

Besides, for any $\delta \in [0,1]$, in view of (7.1.10), interpolation properties yield

$$\left\|A_0^{1/2}(\pi_h\pi_h^* - I)\phi\right\|_X \leq C_0 h^{\theta(1-\delta)} \left\|A_0^{1-\delta/2}\phi\right\|_X , \quad \forall \phi \in \mathcal{D}(A_0^{1-\delta/2}),$$

and thus, as above,

$$\begin{aligned}\left\|A_0^{\delta/2}(U_h-\pi_h u_h)\right\|_X &= \sup_{\substack{\phi\in\mathcal{D}(A_0^{1-\delta/2}),\\ \left\|A_0^{1-\delta/2}\phi\right\|_X=1}} \left\{ < A_0^{\delta/2}(U_h-\pi_h u_h), A_0^{1-\delta/2}\phi >_X \right\}\\ &\le \left\|A_0^{1/2}\pi_h u_h\right\|_X \sup_{\substack{\phi\in\mathcal{D}(A_0^{1-\delta/2}),\\ \left\|A_0^{1-\delta/2}\phi\right\|_X=1}} \left\|A_0^{1/2}(\pi_h\pi_h^*-I)\phi\right\|_X\\ &\le C_0 h^{\theta(1-\delta)}\left\|A_0^{1/2}\pi_h u_h\right\|_X.\end{aligned}$$

Especially, for $\delta = 2\kappa$, we obtain

$$\left\|A_0^{\kappa}(U_h-\pi_h u_h)\right\|_X \le C_0 h^{\theta(1-2\kappa)}\left\|A_0^{1/2}\pi_h u_h\right\|_X.$$

Besides, using the definition (7.1.6) of A_{0h}, one easily gets

$$\left\|A_{0h}^{1/2}\phi_h\right\|_h = \left\|A_0^{1/2}\pi_h\phi_h\right\|_X, \quad \forall\phi_h\in V_h. \tag{7.3.6}$$

It follows that

$$\begin{cases} \left\|U_h-\pi_h u_h\right\|_X \le C_0 h^{\theta}\left\|A_{0h}^{1/2}u_h\right\|_h,\\ \left\|A_0^{\kappa}(U_h-\pi_h u_h)\right\|_X \le C_0 h^{\theta(1-2\kappa)}\left\|A_{0h}^{1/2}u_h\right\|_h. \end{cases} \tag{7.3.7}$$

In particular, this implies, by definition of $\|\cdot\|_h$, that

$$\|u_h\|_h - C_0 h^{\theta}\left\|A_{0h}^{1/2}u_h\right\|_h \le \|U_h\|_X \le \|u_h\|_h + C_0 h^{\theta}\left\|A_{0h}^{1/2}u_h\right\|_h, \tag{7.3.8}$$

and that

$$\|U_h\|_X^2 \le 2\|u_h\|_h^2 + 2C_0^2 h^{2\theta}\left\|A_{0h}^{1/2}u_h\right\|_h^2. \tag{7.3.9}$$

Using $B\in\mathfrak{L}(\mathcal{D}(A_0^{\kappa}),Y)$ and the estimates (7.3.7), we obtain

$$\left|\,\|BU_h\|_Y - \|B_h u_h\|_Y\,\right| \le K_B C_0 h^{\theta(1-2\kappa)}\left\|A_{0h}^{1/2}u_h\right\|_h. \tag{7.3.10}$$

In particular,

$$\|BU_h\|_Y \ge \|B_h u_h\|_Y - K_B C_0 h^{\theta(1-2\kappa)}\left\|A_{0h}^{1/2}u_h\right\|_h. \tag{7.3.11}$$

Then we obtain

$$\|BU_h\|_Y^2 \ge \frac{1}{2}\|B_h u_h\|_Y^2 - K_B^2 C_0^2 h^{2\theta(1-2\kappa)}\left\|A_{0h}^{1/2}u_h\right\|_h^2. \tag{7.3.12}$$

We now estimate $\left\|A_0^{1/2}U_h\right\|_X^2 - \left\|A_{0h}^{1/2}u_h\right\|_h^2$. On one hand, we have

$$\left\|A_0^{1/2}U_h\right\|_X^2 =< A_0U_h, U_h >_X =< \pi_h A_{0h}u_h, U_h >_X =< A_{0h}u_h, \pi_h^*U_h >_h .$$

On the other hand, we have

$$\left\|A_{0h}^{1/2}u_h\right\|_h^2 =< A_{0h}u_h, u_h >_h =< A_{0h}u_h, \pi_h^*\pi_h u_h >_h .$$

Subtracting these two identities, we get

$$\left\|A_0^{1/2}U_h\right\|_X^2 - \left\|A_{0h}^{1/2}u_h\right\|_h^2 =< A_{0h}u_h, \pi_h^*(U_h - \pi_h u_h) >_h,$$

and therefore, using (7.3.7),

$$\left|\left\|A_0^{1/2}U_h\right\|_X^2 - \left\|A_{0h}^{1/2}u_h\right\|_h^2\right| \le C_0 h^\theta \left\|A_{0h}u_h\right\|_h \left\|A_{0h}^{1/2}u_h\right\|_h. \tag{7.3.13}$$

Since $u_h \in \mathcal{C}_h(\eta/h^\sigma)$, estimates (7.3.4), (7.3.8), (7.3.9), (7.3.12) and (7.3.13) imply:

$$\begin{aligned}
&\|U_h\|_X \le \|u_h\|_h (1 + C_0 h^{\theta-\sigma/2}\sqrt{\eta}),\\
&\|U_h\|_X^2 \le 2\|u_h\|_h^2 (1 + C_0^2 h^{2\theta-\sigma}\eta),\\
&\|BU_h\|_Y^2 \ge \frac{1}{2}\|B_h u_h\|_Y^2 - K_B^2 C_0^2 h^{2\theta(1-2\kappa)-\sigma}\eta\|u_h\|_h^2,\\
&\left\|A_0^{1/2}U_h\right\|_X^2 \ge \left\|A_{0h}^{1/2}u_h\right\|_h^2 (1 - C_0 h^{\theta-\sigma/2}\sqrt{\eta}),\\
&\left\|A_0^{1/2}U_h\right\|_X^2 \le \left\|A_{0h}^{1/2}u_h\right\|_h^2 (1 + C_0 h^{\theta-\sigma/2}\sqrt{\eta}).
\end{aligned} \tag{7.3.14}$$

From (7.3.3) we then deduce

$$\begin{aligned}
(1 - C_0 h^{\theta-\sigma/2}\sqrt{\eta})\left\|A_{0h}^{1/2}u_h\right\|_h^2 + \frac{\alpha^2}{2}\|B_h u_h\|_Y^2 \le \|u_h\|_h (1 + C_0 h^{\theta-\sigma/2}\sqrt{\eta})\times\\
\left[\|A_{0h}u_h\|_h^2 + \beta^2\left\|A_{0h}^{1/2}u_h\right\|_h^2 (1 + C_0\sqrt{\eta}h^{\theta-\sigma/2})\right]^{1/2}\\
+ 2\gamma^2\|u_h\|_h^2 (1 + C_0^2\eta h^{2\theta-\sigma}) + \alpha^2 K_B^2 C_0^2 h^{2\theta(1-2\kappa)-\sigma}\eta\|u_h\|_h^2.
\end{aligned} \tag{7.3.15}$$

Using $\sigma < 2\theta$ and $\sigma \le 2\theta(1-2\kappa)$ (by definition (7.1.12)), we simplify this expression into

$$\begin{aligned}
(1 - C_0 h^{\theta-\sigma/2}\sqrt{\eta})\left\|A_{0h}^{1/2}u_h\right\|_h^2 + \frac{\alpha^2}{2}\|B_h u_h\|_Y^2 \le \|u_h\|_h (1 + C_0 h^{\theta-\sigma/2}\sqrt{\eta})\times\\
\left[\|A_{0h}u_h\|_h^2 + \beta^2\left\|A_{0h}^{1/2}u_h\right\|_h^2 (1 + C_0\sqrt{\eta})\right]^{1/2} + \left(2\gamma^2(1 + C_0^2\eta) + \alpha^2 K_B^2 C_0^2\eta\right)\|u_h\|_h^2.
\end{aligned}$$

Again using $\sigma < 2\theta$, we get, for h small enough,

$$1 \le \frac{1}{1 - C_0 h^{\theta-\sigma/2}\sqrt{\eta}} \le 2, \quad \text{and} \quad \frac{1 + C_0 h^{\theta-\sigma/2}\sqrt{\eta}}{1 - C_0 h^{\theta-\sigma/2}\sqrt{\eta}} \le 1 + 3C_0 h^{\theta-\sigma/2}\sqrt{\eta},$$

and thus

$$\begin{aligned}
\left\|A_{0h}^{1/2}u_h\right\|_h^2 + \frac{\alpha^2}{2}\|B_h u_h\|_Y^2 \le \|u_h\|_h (1 + 3C_0 h^{\theta-\sigma/2}\sqrt{\eta})\\
\left[\|A_{0h}u_h\|_h^2 + \beta^2\left\|A_{0h}^{1/2}u_h\right\|_h^2 (1 + C_0^2\eta)\right]^{1/2}\\
+ 2\left(2\gamma^2(1 + C_0^2\eta) + \alpha^2 K_B^2 C_0^2\eta\right)\|u_h\|_h^2.
\end{aligned} \tag{7.3.16}$$

Again using $\sigma < 2\theta$, we get, for h small enough,

$$(1 + 3C_0 h^{\theta-\sigma/2}\sqrt{\eta})^2 \le 1 + 7C_0 h^{\theta-\sigma/2}\sqrt{\eta} \le 2.$$

In particular,

$$(1+3C_0h^{\theta-\sigma/2}\sqrt{\eta})^2\Big(\|A_{0h}u_h\|_h^2+\beta^2\left\|A_{0h}^{1/2}u_h\right\|_h^2(1+C_0^2\eta)\Big)$$
$$\leq \|A_{0h}u_h\|_h^2+7C_0h^{\theta-\sigma/2}\sqrt{\eta}\,\|A_{0h}u_h\|_h^2+2\beta^2\left\|A_{0h}^{1/2}u_h\right\|_h^2(1+C_0^2\eta)$$
$$\leq \|A_{0h}u_h\|_h^2+\left\|A_{0h}^{1/2}u_h\right\|_h^2\Big(7C_0h^{\theta-3\sigma/2}\eta^{3/2}+2\beta^2(1+C_0^2\eta)\Big).$$

With σ as in (7.1.12), we thus obtain (7.3.1) for h small enough with

$$\alpha_*^2=\frac{\alpha^2}{2},\qquad \beta_*^2=7C_0\eta^{3/2}+2\beta^2(1+C_0^2\eta),$$
$$\gamma_*^2=4\gamma^2(1+C_0^2\eta)+2\alpha^2K_B^2C_0^2\eta.$$

Remark that applying Theorem 7.2.2, one can obtain explicit estimates on the constants in (7.1.14). □

7.3.2 Observability

Proof of Theorem 7.1.3: Observability. Assume that system (7.1.1)-(7.1.3) is admissible and exactly observable. Then, from Theorem 7.2.11, there exist positive constants α and β such that (7.2.42) holds.

Our proof is now based on the spectral criterion given in Theorem 7.2.11.

We first prove that there exist positive constants α_* and β_* such that for any $h>0$, the following inequality holds:

$$\left\|A_{0h}^{1/2}u_h\right\|_h^2\leq \|u_h\|_h\ \|A_{0h}u_h\|_h+\alpha_*^2\|B_hu_h\|_Y^2-\beta_*^2\|u_h\|_h^2,\quad \forall u_h\in\mathcal{C}_h(\epsilon/h^\sigma). \tag{7.3.17}$$

In the sequel, we fix $h>0$, $u_h\in\mathcal{C}_h(\epsilon/h^\sigma)$, where ϵ is a positive parameter independent of $h>0$ which we will choose later on, and, similarly as in (7.3.2), we introduce $U_h\in\mathcal{D}(A_0)$ defined by (7.3.2).

Since U_h belongs to $\mathcal{D}(A_0)$, (7.2.42) applies:

$$\left\|A_0^{1/2}U_h\right\|_X^2\leq \|U_h\|_X\ \|A_0U_h\|_X+\alpha^2\|BU_h\|_Y^2-\beta^2\|U_h\|_X^2. \tag{7.3.18}$$

We will then deduce estimate (7.3.17) from (7.3.18), by comparing each term carefully. Actually, we only need the estimates (7.3.14) used above, and the following estimates,

$$\begin{aligned}\|BU_h\|_Y^2&\leq 2\|B_hu_h\|_h^2+2K_B^2C_0^2h^{2\theta(1-2\kappa)}\left\|A_{0h}^{1/2}u_h\right\|_h^2,\\ \|U_h\|_h^2&\geq\frac{1}{2}\|u_h\|^2-C_0^2h^{2\theta}\left\|A_{0h}^{1/2}u_h\right\|_h^2,\end{aligned} \tag{7.3.19}$$

which follows easily from (7.3.10) and (7.3.8).

Now, plugging estimates (7.3.14) and (7.3.19) into (7.3.18), we get:

$$(1-C_0\sqrt{\epsilon}h^{\theta-\sigma/2})\left\|A_{0h}^{1/2}u_h\right\|_h^2\leq(1+C_0\sqrt{\epsilon}h^{\theta-\sigma/2})\|u_h\|_h\|A_{0h}u_h\|_h+2\alpha^2\|B_hu_h\|_Y^2$$
$$+2\alpha^2K_B^2C_0^2\epsilon h^{2\theta(1-2\kappa)-\sigma}\|u_h\|_h^2-\frac{\beta^2}{2}\|u_h\|_h^2+\beta^2C_0^2h^{2\theta-\sigma}\epsilon\|u_h\|_h^2. \tag{7.3.20}$$

But, for h small enough,

$$\frac{1+C_0\sqrt{\epsilon}h^{\theta-\sigma/2}}{1-C_0\sqrt{\epsilon}h^{\theta-\sigma/2}} \leq 1+3C_0\sqrt{\epsilon}h^{\theta-\sigma/2}, \qquad \text{and} \quad \frac{1}{1-C_0\sqrt{\epsilon}h^{\theta-\sigma/2}} \leq 2,$$

and thus we obtain

$$\left\|A_{0h}^{1/2}u_h\right\|_h^2 \leq \left(1+3C_0\sqrt{\epsilon}h^{\theta-\sigma/2}\right)\|u_h\|_h\|A_{0h}u_h\|_h + 4\alpha^2\|B_hu_h\|_Y^2 + 4\alpha^2K_B^2C_0^2\epsilon h^{2\theta(1-2\kappa)-\sigma}\|u_h\|_h^2 - \frac{\beta^2}{2}\|u_h\|_h^2 + 2\beta^2C_0^2h^{2\theta-\sigma}\epsilon\|u_h\|_h^2.$$

This yields

$$\left\|A_{0h}^{1/2}u_h\right\|_h^2 \leq \|u_h\|_h\|A_{0h}u_h\|_h + 4\alpha^2\|B_hu_h\|_Y^2 + \|u_h\|_h^2 \times \left(3C_0h^{\theta-3\sigma/2}\epsilon^{3/2} + 4\alpha^2K_B^2C_0^2\epsilon h^{2\theta(1-2\kappa)-\sigma} + 2\beta^2C_0^2h^{2\theta-\sigma}\epsilon - \frac{\beta^2}{2}\right). \quad (7.3.21)$$

Let us then check that we can choose $\epsilon > 0$ such that, for all $h > 0$ small enough,

$$3C_0\epsilon^{3/2}h^{\theta-3\sigma/2} + 4\alpha^2K_B^2C_0^2\epsilon h^{2\theta(1-2\kappa)-\sigma} + 2\beta^2C_0^2h^{2\theta-\sigma}\epsilon - \frac{\beta^2}{2} \leq -\frac{\beta^2}{4}. \quad (7.3.22)$$

This can indeed be done, due to the choice (7.1.12) of σ. Then, taking such an $\epsilon > 0$, we obtain (7.3.17) by setting

$$\alpha_* = 2\alpha, \quad \beta_* = \frac{\beta}{2}.$$

Now, we need to check that the first eigenvalues λ_1^h of the operators A_{0h} are uniformly bounded from below by a positive constant. This can be easily deduced from the Rayleigh characterization of the first eigenvalues of A_{0h} and A_0:

$$\lambda_1^h = \inf_{\phi_h\in V_h}\frac{\left\|A_{0h}^{1/2}\phi_h\right\|_h^2}{\|\phi_h\|_h^2}, \quad \lambda_1 = \inf_{\phi\in\mathcal{D}(A_0^{1/2})}\frac{\left\|A_0^{1/2}\phi\right\|_X^2}{\|\phi\|_X^2}. \quad (7.3.23)$$

Indeed, from (7.3.6), identities (7.3.23) imply

$$\lambda_1^h = \inf_{\phi_h\in V_h}\frac{\left\|A_{0h}^{1/2}\phi_h\right\|_h^2}{\|\phi_h\|_h^2} = \inf_{\phi_h\in V_h}\frac{\left\|A_0^{1/2}\pi_h\phi_h\right\|_X^2}{\|\pi_h\phi_h\|_X^2} \geq \lambda_1 > 0. \quad (7.3.24)$$

The observability property stated in Theorem 7.1.3 then follows from Theorem 7.2.11 and the uniform admissibility properties stated in Theorem 7.1.3, already obtained in the previous subsection. □

7.4 Examples

In this section, we present several applications of Theorem 7.1.3, and confront our results with the existing ones.

7.4.1 The 1d wave equation

Let us consider the classical 1d wave equation:

$$\begin{cases} \ddot{u} - \partial_{xx}^2 u = 0, & (t,x) \in \mathbb{R} \times (0,1), \\ u(t,0) = u(t,1) = 0, & t \in \mathbb{R}, \\ u(0,x) = u_0(x), \quad \ddot{u}(0,x) = u_1(x), & x \in (0,1). \end{cases} \tag{7.4.1}$$

For (a,b) a subset of $(0,1)$, we observe system (7.4.1) through

$$y(t,x) = \dot{u}(t,x)\chi_{(a,b)}(x), \tag{7.4.2}$$

where $\chi_{(a,b)}$ is the characteristic function of (a,b).

This model indeed enters in the abstract framework considered in this article, by setting $A_0 = -\partial_{xx}^2$ on $(0,1)$ with Dirichlet boundary conditions, and $B = \chi_{(a,b)}$. Indeed, A_0 is self-adjoint, positive definite with compact resolvent in $L^2(0,1)$. The operator B obviously is continuous on $L^2(0,1)$ with values in $L^2(0,1)$. The admissibility of (7.4.1)-(7.4.2) is then straightforward.

The observability property for (7.4.1)-(7.4.2) is well-known to hold if and only if the *Geometric Control Condition* is satisfied, see [2, 3]. This condition, roughly speaking, asserts the existence of a time T^* such that all the rays of Geometric Optics enters in the observation domain in a time smaller than T^*. In 1d, this condition is always satisfied, and thus system (7.4.1)-(7.4.2) is exactly observable. This can also be seen using multipliers techniques as in [21, 30].

To construct the space V_h, we use P1 finite elements. More precisely, for $n_h \in \mathbb{N}$, set $h = 1/(n_h+1) > 0$ and define the points $x_j = jh$ for $j \in \{0, \cdots, n_h+1\}$. We define the basis functions

$$e_j(x) = \Big[1 - \frac{|x - x_j|}{h}\Big]^+, \quad \forall j \in \{1, \cdots, n_h\}.$$

Now, $V_h = \mathbb{R}^{n_h}$, and the injection π_h simply is

$$\begin{array}{rcl} \pi_h : V_h = \mathbb{R}^{n_h} & \to & L^2(0,1) \\ u_h = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n_h} \end{pmatrix} & \mapsto & \pi_h u_h(x) = \displaystyle\sum_{j=1}^{n_h} u_j e_j(x). \end{array}$$

Usually, the resulting schemes are written as

$$\begin{cases} M_h \ddot{u}_h(t) + K_h u_h(t) = 0, \quad t \in \mathbb{R}, \\ u_h(0) = u_{0h}, \quad \dot{u}_h(0) = u_{1h}, \end{cases} \qquad y_h(t) = B\pi_h \dot{u}_h(t), \quad t \in \mathbb{R}, \tag{7.4.3}$$

where M_h and K_h are $n_h \times n_h$ matrices defined by $(M_h)_{i,j} = \int_0^1 e_i(x)e_j(x)\ dx$ and $(K_h)_{i,j} = \int_0^1 \partial_x e_i(x)\partial_x e_j(x)\ dx$. Note that, since M_h is a Gram matrix corresponding to a linearly independent family, it is invertible, self-adjoint and positive definite, and thus the following defines a scalar product:

$$< \phi_h, \psi_h >_h = \phi_h^* M_h \psi_h, \quad (\phi_h, \psi_h) \in V_h^2. \tag{7.4.4}$$

Besides, from the definition of M_h, one easily checks that

$$< \phi_h, \psi_h >_h = \int_0^1 \pi_h(\phi_h)(x)\pi_h(\psi_h)(x)\ dx, \quad \forall (\phi_h, \psi_h) \in V_h^2,$$

as presented in the introduction.

Similarly, one obtains that, for all $(\phi_h, \psi_h) \in V_h^2$,

$$\phi_h^* K_h \psi_h = \phi_h^* M_h M_h^{-1} K_h \psi_h =< \phi_h, M_h^{-1} K_h \psi_h >_h = \phi_h^* K_h M_h^{-1} M_h \psi_h$$
$$=< M_h^{-1} K_h \phi_h, \psi_h >_h = \int_0^1 \partial_x(\pi_h \phi_h)(x) \partial_x(\pi_h \psi_h)(x)\, dx,$$

which proves that the operator $M_h^{-1} K_h$ coincides with the operator A_{0h} of our framework. Note that this operator indeed is self-adjoint, but with respect to the scalar product (7.4.4) and not with the usual euclidean norm of $\mathbb{R}^{n_h}$.

It is by now a common feature of finite element techniques (see for instance [33]) that estimates (7.1.10) hold for $\theta = 1$. We can thus apply Theorem 7.1.3 to systems (7.4.3):

Theorem 7.4.1. *There exist $\epsilon > 0$, a time T^* and a positive constant k_* such that for any $h > 0$, any solution u_h of (7.4.3) with initial data $(u_{0h}, u_{1h}) \in \mathcal{C}_h(\epsilon/h^{2/3})^2$ satisfies (7.1.16).*

This result is to be compared with the better ones obtained in [19]: In [19], it is proved that, for finite element approximation schemes of the 1d wave equation, observability properties hold uniformly within the larger class $\mathcal{C}_h(\alpha/h^2)$ for $\alpha < 4$.

Though, as we will see hereafter, we can tackle more general cases, even in 1d, for instance taking sequence of meshes $\mathcal{S}_n$ given by $n+2$ points as

$$x_{0,n} = 0 < x_{1,n} < \cdots < x_{n,n} < x_{n+1,n} = 1, \qquad h_{j+1/2,n} = x_{j+1,n} - x_{j,n},$$

for which we only assume $h_n = \sup_j\{h_{j+1/2,n}\}$ to go to zero when $n \to \infty$.

7.4.2 More general cases

Let Ω be a bounded smooth domain of $\mathbb{R}^N$, with $N \geq 1$, and consider the following wave equation:

$$\begin{cases} \ddot{u} - \mathrm{div}(M(x)\nabla u) = 0, & (x,t) \in \Omega \times \mathbb{R}, \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times \mathbb{R}, \\ u(x,0) = u_0(x), \quad \dot{u}(x,0) = u_1(x), & x \in \Omega, \end{cases} \tag{7.4.5}$$

where $M(x)$ is a C^1 function on $\bar{\Omega}$ with values in the self-adjoint $N \times N$ matrices. We also assume that there exist positive constants α and β such that for all $\xi \in \mathbb{R}^N$,

$$\alpha|\xi|^2 \leq (M(x)\xi, \xi) \leq \beta|\xi|^2, \quad \forall x \in \Omega, \tag{7.4.6}$$

where $(\cdot,\cdot)$ is the canonical scalar product of $\mathbb{R}^N$ and $|\cdot|$ is the corresponding norm.

Under these assumptions, it is well-known that system (7.4.5) is well-posed for initial data $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$.

System (7.4.5) is a particular instance of (7.1.1) for $A_0 = -\mathrm{div}(M(x)\nabla\cdot)$ on Ω with Dirichlet boundary condition. This operator is indeed self-adjoint positive definite with compact resolvent, and its domain is $\mathcal{D}(A_0) = H^2(\Omega) \cap H_0^1(\Omega)$.

Now, set ω a non-empty open subset of Ω, which satisfies the *Geometric Control Condition* (see [2] and above), and consider the observation

$$y(x,t) = \chi_\omega(x)\dot{u}(x,t), \quad (x,t) \in \Omega \times (0,T). \tag{7.4.7}$$

This defines a bounded operator B on $L^2(\Omega)$. Therefore, the admissibility condition for (7.4.5)-(7.4.7) is obvious.

As said above, the *Geometric Control Condition* guarantees the exact observability property for (7.4.5)-(7.4.7). Note that, in our case, the rays are not necessarily straight lines, but correspond to the bicharacteristic rays of the pseudo-differential operator $\tau^2 - (M(x)\xi, \xi)$.

We consider P1 finite elements on meshes $\mathcal{T}_h$. We furthermore assume that the meshes $\mathcal{T}_h$ of the domain Ω are regular in the sense of finite elements [33, Section 5]. Roughly speaking, this assumption imposes that the polyhedra of $(\mathcal{T}_h)$ are not too flat.

Definition 7.4.2. Let $\mathcal{T} = \cup_{K\in\mathcal{T}} K$ be a mesh of a bounded domain Ω. For each polyhedron $K \in \mathcal{T}$, we define h_K as the diameter of K and ρ_K as the maximum diameter of the spheres $S \subset K$. We then define the regularity of $\mathcal{T}$ as

$$\mathrm{Reg}(\mathcal{T}) = \sup_{K\in\mathcal{T}} \Big\{\frac{h_K}{\rho_K}\Big\}.$$

A sequence of mesh $(\mathcal{T}_h)$ is said to be uniformly regular if

$$\sup_h \mathrm{Reg}(\mathcal{T}_h) < \infty.$$

In this case, see [33], setting $h = \sup_{K\in\mathcal{T}} h_K$, estimates (7.1.10) again hold for $\theta = 1$, and Theorem 7.1.3 implies:

Theorem 7.4.3. *Assume that system* (7.4.5)-(7.4.7) *is observable. Given a sequence of uniformly regular meshes* $(\mathcal{T}_h)_{h>0}$ *satisfying* $h = \sup_{K\in\mathcal{T}_h} h_K$*, there exist* $\epsilon > 0$*, a time* T^* *and a positive constant* k_* *such that for any* $h > 0$ *small enough, any solution* u_h *of the P1 finite element approximation scheme of* (7.4.5)-(7.4.7) *corresponding to the mesh* $\mathcal{T}_h$ *with initial data* $(u_{0h}, u_{1h}) \in \mathcal{C}_h(\epsilon/h^{2/3})^2$ *satisfies* (7.1.16).

To our knowledge, this is the first time that observability properties for space semi-discretizations of (7.4.5)-(7.4.7) are derived in such generality for the wave equation. In particular, we emphasize that the only non-trivial assumption we used is (7.1.10), which is needed anyway to guarantee the convergence of the numerical schemes.

7.5 Fully discrete approximation schemes

This section is based on the article [11], which studied observability properties of time discrete conservative linear systems. As said in [11, Section 5], this study can be combined with observability results on space semi-discrete systems to deduce observability properties for fully discrete systems. Below, we present an application of the results in [11].

Let $\beta \geq 1/4$ and consider the following time discrete approximation scheme - the so-called Newmark method, see for instance [33] - of (7.1.8):

$$\begin{cases} \dfrac{u_h^{k+1} + u_h^{k-1} - 2u_h^k}{(\triangle t)^2} + A_{0h}\Big(\beta u_h^{k-1} + (1-2\beta)u_h^k + \beta u_h^{k+1}\Big) = 0, \quad k \in \mathbb{N}^*, \\ \Big(\dfrac{u_h^0 + u_h^1}{2}, \dfrac{u_h^1 - u_h^0}{\triangle t}\Big) = (u_{0h}, u_{1h}) \in V_h^2, \end{cases} \tag{7.5.1}$$

where u_h^k corresponds to an approximation of the solution u_h of (7.1.8) at time $t_k = k\triangle t$.

The energy of solutions u_h of (7.5.1), defined by

$$E_h^{k+1/2} = \frac{1}{2}\left\|A_{0h}^{1/2}\Big(\frac{u_h^k + u_h^{k+1}}{2}\Big)\right\|_h^2 + \frac{1}{2}\left\|\frac{u_h^{k+1} - u_h^k}{\triangle t}\right\|_h^2 + \frac{(\triangle t)^2}{8}(4\beta - 1)\left\|A_{0h}^{1/2}\Big(\frac{u_h^{k+1} - u_h^k}{\triangle t}\Big)\right\|_h^2, \quad k \in \mathbb{N}, \quad (7.5.2)$$

is constant.

Then we get the following observability result (see [11]):

Theorem 7.5.1. *Let A_0 be a self-adjoint positive definite unbounded operator with compact resolvent and $B \in \mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$, with $\kappa < 1/2$.*

Assume that the maps $(\pi_h)_{h>0}$ satisfy property (7.1.10). Let $\beta \geq 1/4$, and consider the fully discrete approximation scheme (7.5.1). Set σ as in (7.1.12), and $\delta > 0$.

Admissibility: *Assume that system (7.1.1)-(7.1.3) is admissible.*

Then, for any $\eta > 0$ and $T > 0$, there exists a positive constant $K_{T,\eta} > 0$ such that, for any $h > 0$ and $\triangle t > 0$, any solution of (7.5.1) with initial data

$$(u_{0h}, u_{1h}) \in \Big(\mathcal{C}_h(\eta/h^\sigma) \cap \mathcal{C}_h(\delta^2/(\triangle t)^2)\Big)^2 \quad (7.5.3)$$

satisfies

$$\triangle t \sum_{k\triangle t \in [0,T]} \left\|B_h\Big(\frac{u_h^{k+1} - u_h^k}{\triangle t}\Big)\right\|_Y^2 \leq K_{T,\eta} E_h^{1/2}. \quad (7.5.4)$$

Observability: *Assume that system (7.1.1)-(7.1.3) is admissible and exactly observable.*

Then there exist $\epsilon > 0$, a time T^* and a positive constant $k_* > 0$ such that, for any $h > 0$ and $\triangle t > 0$, any solution of (7.5.1) with initial data

$$(u_{0h}, u_{1h}) \in \Big(\mathcal{C}_h(\epsilon/h^\sigma) \cap \mathcal{C}_h(\delta^2/(\triangle t)^2)\Big)^2 \quad (7.5.5)$$

satisfies

$$k_* E_h^{1/2} \leq \triangle t \sum_{k\triangle t \in [0,T^*]} \left\|B_h\Big(\frac{u_h^{k+1} - u_h^k}{\triangle t}\Big)\right\|_Y^2. \quad (7.5.6)$$

Obviously, inequalities (7.5.4)-(7.5.6) are time discrete counterparts of (7.1.14)-(7.1.16). Remark that, as in Theorem 7.1.3, a filtering condition is needed, but which now depends on both time and space discretization parameters.

Also remark that if $(\triangle t)^2 h^{-\sigma}$ is small enough, then $\mathcal{C}_h(\epsilon/h^\sigma) \cap \mathcal{C}_h(\delta^2/(\triangle t)^2) = \mathcal{C}_h(\epsilon/h^\sigma)$. Roughly speaking, this indicates that under the CFL type condition $(\triangle t)^2 h^{-\sigma} \leq \epsilon/\delta^2$, system (7.5.1) behaves, with respect to the admissibility and observability properties, similarly as the space semi-discrete equations (7.1.8).

Remark 7.5.2. We restrict our presentation to the Newmark method, but similar results hold for a large range of time discrete approximation schemes of (7.1.8). We refer to [11], and in particular to Section 3, for the precise assumptions on the time-discrete approximation schemes under which we can guarantee uniform observability properties to hold.

7.6 Controllability properties

This section aims at discussing applications of Theorem 7.1.3 to controllability properties for space semi-discretizations of wave type equations such as (7.1.1). The approach presented below is strongly inspired by previous works [16, 19, 40, 41, 10], and closely follows [10].

In the whole section, we assume that the hypotheses of Theorem 7.1.3 are satisfied.

7.6.1 The continuous setting

Consider the following control problem: Given $T > 0$, for any $(w_0, w_1) \in \mathcal{D}(A_0^{1/2}) \times X$, find a control $v \in L^2(0,T;Y)$ such that the solution w of

$$\ddot{w} + A_0 w = B^* v(t), \quad t \in [0,T], \qquad w(0) = w_0, \quad \dot{w}(0) = w_1, \tag{7.6.1}$$

satisfies

$$w(T) = 0, \quad \dot{w}(T) = 0. \tag{7.6.2}$$

The controllability issue in time T for (7.6.1) is equivalent to the observability property in time T for (7.1.1)-(7.1.3) (see for instance [23]). Indeed, these two properties are dual, and this duality can be made precise using the Hilbert Uniqueness Method (HUM in short), see [23].

More precisely, the control of minimal $L^2(0,T;Y)$ norm for (7.6.1), that we will denote by v_{HUM}, is characterized through the minimizer of the functional $\mathcal{J}$ defined on $\mathcal{D}(A_0^{1/2}) \times X$ by:

$$\mathcal{J}(u_{0T}, u_{1T}) = \frac{1}{2}\int_0^T \|B\dot{u}(t)\|_Y^2 \; dt + < A_0^{1/2} u(0), A_0^{1/2} w_0 >_X + < \dot{u}(0), w_1 >_X, \tag{7.6.3}$$

where u is the solution of

$$\ddot{u} + A_0 u = 0, \quad t \in [0,T], \qquad u(T) = u_{0T}, \quad \dot{u}(T) = u_{1T}. \tag{7.6.4}$$

Indeed, if (u_{0T}^*, u_{1T}^*) is the minimizer of $\mathcal{J}$, then $v_{HUM}(t) = B\dot{u}^*(t)$, where u^* is the solution of (7.6.4) with initial data (u_{0T}^*, u_{1T}^*).

Besides, the only admissible control for (7.6.1) which can be written as $B\dot{u}(t)$ for a solution u of (7.6.4) is the HUM control v_{HUM}. This characterization will be used in the sequel.

Note that the observability property (7.1.5) for (7.1.1)-(7.1.3) implies the strict convexity and the coercivity of $\mathcal{J}$ and therefore guarantees the existence of a unique minimizer for $\mathcal{J}$.

7.6.2 The semi-discrete setting

The natural idea which consists in computing the discrete HUM controls for discrete versions of (7.6.1) may fail in providing good approximations of the HUM control for (7.6.1). We refer for instance to

the survey article [41] for a detailed presentation of this fact in the context of the 1d wave equation. We thus use filtering techniques developed for instance in [16, 19, 40, 41, 10] to overcome the problems created by the high-frequency components.

Our presentation closely follows the one in [10]. The proofs of the result below will be only sketched, and can be done similarly as in [10].

Since we assumed that the hypotheses of Theorem 7.1.3 hold, there exists a time T^* such that (7.1.16) holds for any solution of (7.1.8) with initial data in the filtered space $\mathcal{C}_h(\epsilon/h^\sigma)^2$.

We now fix $T \geq T^*$.

Following the strategy of HUM, we introduce the adjoint problem

$$\ddot{u}_h + A_{0h}u_h = 0, \quad t \in [0,T], \quad (u_h, \dot{u}_h)(T) = (u_{0Th}, u_{1Th}). \tag{7.6.5}$$

Method I

For any $h > 0$, we consider the following control problem: For any $(w_{0h}, w_{1h}) \in V_h^2$, find $v_h \in L^2(0,T;Y)$ of minimal $L^2(0,T;Y)$ such that the solution w_h of

$$\ddot{w}_h + A_{0h}w_h = B_h^* v_h(t), \quad t \in [0,T], \qquad w_h(0) = w_{0h}, \quad \dot{w}_h(0) = w_{1h}, \tag{7.6.6}$$

satisfies

$$P_h w_h(T) = 0, \quad P_h \dot{w}_h(T) = 0, \tag{7.6.7}$$

where P_h is the orthogonal projection in V_h on $\mathcal{C}_h(\epsilon/h^\sigma)$.

To deal with this problem, we introduce the functional $\mathcal{J}_h$ defined for (u_{0Th}, u_{1Th}) in $\mathcal{C}_h(\epsilon/h^\sigma)^2$ by

$$\mathcal{J}_h(u_{0Th}, u_{1Th}) = \frac{1}{2}\int_0^T \|B_h \dot{u}_h(t)\|_Y^2 \ dt + < A_{0h}^{1/2} w_{0h}, A_{0h}^{1/2} \ u_h(0) >_h + < w_{1h}, \dot{u}_h(0) >_h, \tag{7.6.8}$$

where u_h is the solution of (7.6.5).

For each $h > 0$, the functional $\mathcal{J}_h$ is strictly convex and coercive (see (7.1.16)), and thus has a unique minimizer $(u_{0Th}^*, u_{1Th}^*) \in \mathcal{C}_h(\epsilon/h^\sigma)^2$.

Besides, we have:

Lemma 7.6.1. *For all $h > 0$, let $(u_{0Th}^*, u_{1Th}^*) \in \mathcal{C}_h(\epsilon/h^\sigma)^2$ be the unique minimizer of $\mathcal{J}_h$ (on $\mathcal{C}_h(\epsilon/h^\sigma)^2$), and denote by u_h^* the corresponding solution of* (7.6.5)*. Then the solution of* (7.6.6) *with $v_h = B_h \dot{u}_h^*$ satisfies* (7.6.7)*.*

Sketch of the proof. We present briefly the proof, which is standard (see for instance [23]).

On one hand, multiplying (7.6.6) by $\dot{u}_h$ solution of (7.6.5) with initial data (u_{0Th}, u_{1Th}), we get, for all $(u_{0Th}, u_{1Th}) \in V_h^2$,

$$\int_0^T < v_h(t), B_h \dot{u}_h(t) >_Y \ dt + < A_{0h}^{1/2} w_{0h}, A_{0h}^{1/2} u_h(0) >_h + < w_{1h}, \dot{u}_h(0) >_h$$
$$- < A_{0h}^{1/2} w_h(T), A_{0h}^{1/2} u_{0Th} >_h - < \dot{w}_h(T), u_{1Th} >_h = 0. \tag{7.6.9}$$

On the other hand, the Fréchet derivative of the functional $\mathcal{J}_h$ at (u^*_{0Th}, u^*_{1Th}) yields:

$$\int_0^T < B_h \dot{u}^*_h(t), B_h \dot{u}_h(t) >_Y \ dt + < A_{0h}^{1/2} w_{0h}, A_{0h}^{1/2} u_h(0) >_h + < w_{1h}, \dot{u}_h(0) >_h = 0,$$
$$\forall (u_{0Th}, u_{1Th}) \in \mathcal{C}_h (\epsilon/h^\sigma)^2. \quad (7.6.10)$$

Therefore, setting $v_h = B_h \dot{u}^*_h$, subtracting (7.6.9) to (7.6.10), we obtain

$$< A_{0h}^{1/2} w_h(T), A_{0h}^{1/2} u_{0Th} >_h + < \dot{w}_h(T), u_{1Th} >_h = 0, \quad \forall (u_{0Th}, u_{1Th}) \in \mathcal{C}_h (\epsilon/h^\sigma)^2,$$

or, equivalently, (7.6.7). □

As in [10], we then investigate the convergence of the discrete controls v_h obtained in Lemma 7.6.1.

Theorem 7.6.2. *Assume that the hypotheses of Theorem 7.1.3 are satisfied. Also assume that*

$$Y_X = \left\{ y \in Y, \text{ such that } B^* y \in X \right\} \quad (7.6.11)$$

is dense in Y.

Let $(w_0, w_1) \in \mathcal{D}(A_0^{1/2}) \times X$, and consider a sequence $(w_{0h}, w_{1h})_{h>0}$ such that (w_{0h}, w_{1h}) belongs to V_h^2 for any $h > 0$ and

$$(\pi_h w_{0h}, \pi_h w_{1h}) \to (w_0, w_1) \quad \text{in } \mathcal{D}(A_0^{1/2}) \times X. \quad (7.6.12)$$

Then the sequence $(v_h)_{h>0}$ of discrete controls given by Lemma 7.6.1 converges in $L^2(0,T;Y)$ to the HUM control v_{HUM} of (7.6.1) associated to the initial data (w_0, w_1).

Remark that, for $w \in \mathcal{D}(A_0)$, in view of (7.1.10), the sequence $(w_h)_h = (\pi^*_h w)$ converges to w in $\mathcal{D}(A_0^{1/2})$ in the sense that the sequence $(\pi_h w_h)$ converges to w in $\mathcal{D}(A_0^{1/2})$. For $(w_0, w_1) \in \mathcal{D}(A_0^{1/2}) \times X$, one can then find a sequence $(w_{0h}, w_{1h})_{h>0}$ satisfying (7.6.12) and $(w_{0h}, w_{1h}) \in V_h^2$ for any $h > 0$ by using the density of $\mathcal{D}(A_0)^2$ into $\mathcal{D}(A_0^{1/2}) \times X$.

The technical assumption (7.6.11) on B is usually satisfied, and thus does not limit the range of applications of Theorem 7.6.2. Also note that when B is bounded from X to Y, the space Y_X coincides with Y and (7.6.11) is then automatically satisfied.

The proof of Theorem 7.6.2 uses precisely the same ingredients as the one in [10], and is briefly sketched for the convenience of the reader.

Sketch of the proof. Step 1. The discrete controls v_h are bounded in $L^2(0,T;Y)$. This follows from the inequality

$$\mathcal{J}_h(u^*_{0Th}, u^*_{1Th}) \leq \mathcal{J}_h(0,0) = 0,$$

and the observability inequality (7.1.16). Hence the controls are bounded, and, up to an extraction, the sequence (v_h) weakly converges to some function v in $L^2(0,T;Y)$. Besides, the sequence (u^*_{0Th}, u^*_{1Th}) is also bounded in $\mathcal{D}(A_0^{1/2}) \times X$, and therefore weakly converges in $\mathcal{D}(A_0^{1/2}) \times X$ to some couples of functions $(\tilde{u}_{0T}, \tilde{u}_{1T})$.

Step 2. The weak limit v is an admissible control for (7.6.1) associated to the data (w_0, w_1). This can be deduced, as in [10], from the convergence properties of the approximation schemes (7.1.8) (or equivalently (7.6.5)), which can be found for instance in [33, Section 8].

Step 3. The weak limit v is the HUM control for (7.6.1) associated to the data (w_0, w_1). This is also based on a convergence result which can be found in [33, Section 8], and which guarantees that $v = B\dot{\tilde{u}}$, where $\tilde{u}$ is the solution of (7.6.4) with initial data $(\tilde{u}_{0T}, \tilde{u}_{1T})$. This also proves that $(\tilde{u}_{0T}, \tilde{u}_{1T})$ coincides with the minimizer (u^*_{0T}, u^*_{1T}) of the continuous functional $\mathcal{J}$ in (7.6.3). Assumption (7.6.11) is needed in this step to identify the limit of $(B\dot{u}^*_h)$ with $B\dot{\tilde{u}}$.

Step 4. Finally, the strong convergence of the controls is proved using the convergence of the $L^2(0,T;Y)$ norms. Compute first the Fréchet derivative of $\mathcal{J}$ at (u^*_{0T}, u^*_{1T}): for $(u_{0T}, u_{1T}) \in \mathcal{D}(A_0^{1/2}) \times X$, we obtain

$$\int_0^T < B\dot{u}^*(t), B\dot{u}(t) >_Y \ dt + < A_0^{1/2}u(0), A_0^{1/2}w_0 >_X + < \dot{u}(0), w_1 >_X = 0. \tag{7.6.13}$$

Now, applying (7.6.10) to (u^*_{0Th}, u^*_{1Th}) and (7.6.13) to (u^*_{0T}, u^*_{1T}), the assumptions on the convergence of (w_{0h}, w_{1h}) imply the convergence of the $L^2(0,T;Y)$ norms of v_h to the $L^2(0,T;Y)$ norm of v. □

Method II

As in [10], one can prefer a method which does not involve a filtering process in the discrete setting. We thus recall the works [16, 41, 10], which propose an alternate process based on a Tychonoff regularization of $\mathcal{J}_h$.

Theorem 7.6.3. *Assume that the hypotheses of Theorem 7.1.3 are satisfied. Also assume that $B \in \mathfrak{L}(X,Y)$, which, in particular, implies that $\sigma = 2\theta/3$.*
Let $(w_0, w_1) \in \mathcal{D}(A_0^{1/2}) \times X$, and consider a sequence $(w_{0h}, w_{1h})_{h>0}$ such that (w_{0h}, w_{1h}) belongs to V_h^2 for any $h > 0$ and (7.6.12) holds.

For any $h > 0$, consider the functionals $\mathcal{J}_h^*$, defined for $(u_{0Th}, u_{1Th}) \in V_h^2$ by

$$\mathcal{J}_h^*(u_{0Th}, u_{1Th}) = \frac{1}{2}\int_0^T \|B_h\dot{u}_h(t)\|_Y^2 \ dt + \frac{h^\sigma}{2}\Big(\left\|\tilde{A}_{0h}^{1/2}A_{0h}^{1/2}u_{0Th}\right\|_h^2 + \left\|\tilde{A}_{0h}^{1/2}u_{1Th}\right\|_h^2 \Big) + < A_{0h}^{1/2}w_{0h}, A_{0h}^{1/2}\ u_h(0) >_h + < w_{1h}, \dot{u}_h(0) >_h, \tag{7.6.14}$$

where

$$\tilde{A}_{0h} = A_{0h}(Id_{V_h} + h^\sigma A_{0h})^{-1}, \tag{7.6.15}$$

and u_h is the solution of (7.6.5) with initial data (u_{0Th}, u_{1Th}).

Then, for any $h > 0$, the functional $\mathcal{J}_h^*$ admits a unique minimizer (U_{0Th}, U_{1Th}) in V_h^2. Besides, setting $v_h(t) = B_h\dot{U}_h(t)$, where U_h is the solution of (7.6.5) with initial data (U_{0Th}, U_{1Th}), one gets the following convergence results:

$$v_h \longrightarrow v_{HUM} \quad \text{in } L^2(0,T;Y), \tag{7.6.16}$$

where v_{HUM} denotes the HUM control for (7.6.1).

Theorem 7.6.3 proposes a numerical process based on the minimization of the functional $\mathcal{J}_h^*$ defined for *any* element of V_h^2. Though, the functional $\mathcal{J}_h^*$ involves the regularizing term

$$h^\sigma \left\|\tilde{A}_{0h}^{1/2}u_{1Th}\right\|_h^2 + h^\sigma \left\|\tilde{A}_{0h}^{1/2}A_{0h}^{1/2}u_{0Th}\right\|_h^2.$$

This term is small for data in $\mathcal{C}_h(\epsilon/h^\sigma)$ and of unit order for frequencies higher than $1/h^\sigma$. Also note that this term can be computed easily since

$$h^\sigma \left\| \tilde{A}_{0h}^{1/2} \phi_h \right\|_h^2 = h^\sigma < \tilde{A}_{0h} \phi_h, \phi_h >_h = h^\sigma < A_{0h} \tilde{\phi}_h, \phi_h >_h,$$

where $\tilde{\phi}_h$ is the solution of

$$\Big(Id_{V_h} + h^\sigma A_{0h} \Big) \tilde{\phi}_h = \phi_h. \tag{7.6.17}$$

In other words, the operator $\tilde{A}_{0h}$ simply introduces an elliptic regularization of the data, and the regularizing terms can be computed explicitly by solving the elliptic equation (7.6.17).

Besides, from (7.6.15), $\tilde{A}_{0h}$ and A_{0h} commute, and $\tilde{A}_{0h}$ satisfies:

$$\begin{aligned} &\left\| h^{\sigma/2} \tilde{A}_{0h}^{1/2} \psi_h \right\|_h^2 \leq \|\psi_h\|_h^2, \quad \forall \psi_h \in V_h, \\ &\left\| h^{\sigma/2} \tilde{A}_{0h}^{1/2} \psi_h \right\|_h^2 \geq \frac{\delta}{1+\delta} \|\psi_h\|_h^2, \quad \forall \psi_h \in \mathcal{C}_h(\delta/h^\sigma)^\perp, \quad \forall \delta \geq 0. \end{aligned} \tag{7.6.18}$$

Let us check that the functionals $\mathcal{J}_h^*$ are uniformly coercive. For $(u_{0Th}, u_{1Th}) \in V_h^2$, using (7.1.16), we obtain

$$\begin{aligned} \int_0^T \|B_h \dot{u}_h(t)\|_Y^2 &\geq \frac{1}{2} \int_0^T \|B_h P_h \dot{u}_h(t)\|_Y^2 - \int_0^T \left\| B_h \Big(P_h - Id_{V_h} \Big) \dot{u}_h(t) \right\|_Y^2 \\ &\geq \frac{k_T}{2} \Big(\left\| A_{0h}^{1/2} P_h u_{0Th} \right\|_h^2 + \|P_h u_{1Th}\|_h^2 \Big) - \int_0^T \|B\|_{\mathfrak{L}(X,Y)}^2 \left\| \Big(P_h - Id_{V_h} \Big) \dot{u}_h(t) \right\|_h^2 \\ &\geq \frac{k_T}{2} \Big(\left\| A_{0h}^{1/2} P_h u_{0Th} \right\|_h^2 + \|P_h u_{1Th}\|_h^2 \Big) \\ &\quad - T \|B\|_{\mathfrak{L}(X,Y)}^2 \Big(\left\| A_{0h}^{1/2} (P_h - Id_{V_h}) u_{0Th} \right\|_h^2 + \|(P_h - Id_{V_h}) u_{1Th}\|_h^2 \Big) \\ &\geq \frac{k_T}{2} \Big(\left\| A_{0h}^{1/2} P_h u_{0Th} \right\|_h^2 + \|P_h u_{1Th}\|_h^2 \Big) \\ &\quad - h^\sigma T \|B\|_{\mathfrak{L}(X,Y)}^2 \Big(\frac{1+\epsilon}{\epsilon} \Big) \Big(\left\| A_{0h}^{1/2} \tilde{A}_{0h}^{1/2} u_{0Th} \right\|_h^2 + \left\| \tilde{A}_{0h}^{1/2} u_{1Th} \right\|_h^2 \Big). \end{aligned}$$

Besides, for $(u_{0Th}, u_{1Th}) \in V_h^2$, using (7.6.18), we also have

$$\begin{aligned} &\left\| A_{0h}^{1/2} \Big(Id_{V_h} - P_h \Big) u_{0Th} \right\|_h^2 + \left\| \Big(Id_{V_h} - P_h \Big) u_{1Th} \right\|_h^2 \\ &\qquad \leq h^\sigma \Big(\frac{1+\epsilon}{\epsilon} \Big) \Big(\left\| A_{0h}^{1/2} \tilde{A}_{0h}^{1/2} u_{0Th} \right\|_h^2 + \left\| \tilde{A}_{0h}^{1/2} u_{1Th} \right\|_h^2 \Big). \end{aligned}$$

Combining these two inequalities, we prove that the functionals $\mathcal{J}_h^*$ are uniformly coercive.

The proof of Theorem 7.6.3 can now be done similarly as the one of Theorem 7.6.2, and thus is left to the reader.

Remark 7.6.4. Similar results can be obtained for fully discrete approximation schemes derived from Newmark time discretizations of (7.1.8) (or more general time discrete approximation scheme, see Remark 7.5.2). The proof can then be done similarly as in the time continuous setting, using the observability inequality (7.5.6) and convergence properties for the fully discrete approximation schemes, which can be found for instance in [33].

7.7 Stabilization properties

This section is mainly based on the articles [14, 13], in which stabilization properties are derived for abstract linear damped systems.

Below, we assume that A_0 is self-adjoint, definite positive and with compact resolvent, and that $B \in \mathfrak{L}(X, Y)$.

7.7.1 The continuous setting

Consider the following damped wave type equations:

$$\ddot{u} + A_0 u + B^* B\dot{u} = 0, \quad t \geq 0, \qquad (u(0), \dot{u}(0)) = (u_0, u_1) \in \mathcal{D}(A_0^{1/2}) \times X. \tag{7.7.1}$$

The energy of solutions of (7.7.1), defined by (7.1.2), satisfies the dissipation law

$$\frac{dE}{dt}(t) = -\left\|B\dot{u}(t)\right\|_Y^2, \quad t \geq 0. \tag{7.7.2}$$

System (7.7.1) is said to be exponentially stable if there exists positive constants μ and ν such that any solution of (7.7.1) with initial data $(u_0, u_1) \in \mathcal{D}(A_0^{1/2}) \times X$ satisfies

$$E(t) \leq \mu E(0) \exp(-\nu t). \tag{7.7.3}$$

It is by now well-known (see [17]) that this property holds if and only if the observability inequality (7.1.5) holds for solutions of (7.1.1).

7.7.2 The space semi-discrete setting

We now assume that system (7.1.1)-(7.1.3) is observable in the sense of (7.1.5), or, equivalently (see [17]), that system (7.7.1) is exponentially stable.

Then, combining Theorem 7.1.3 and the results in [14], we get:

Theorem 7.7.1. *Let B be a bounded operator in $\mathfrak{L}(X, Y)$, and assume that system (7.7.1) is exponentially stable in the sense of (7.7.3). Also assume that the hypotheses of Theorem 7.1.3 are satisfied.*

Then the space semi-discrete systems

$$\begin{cases} \ddot{u}_h + A_{0h} u_h + B_h^* B_h \dot{u}_h + h^{2\theta/3} A_{0h} \dot{u}_h = 0, \quad t \geq 0, \\ (u_h(0), \dot{u}_h(0)) = (u_{0h}, u_{1h}) \in V_h^2, \end{cases} \tag{7.7.4}$$

are exponentially stable, uniformly with respect to the space discretization parameter $h > 0$: there exist two positive constants μ_0 and ν_0 independent of $h > 0$ such that for any $h > 0$, any solution u_h of (7.7.4) satisfies, for $t \geq 0$,

$$\left\|A_{0h}^{1/2} u_h(t)\right\|_h^2 + \|\dot{u}_h(t)\|_h^2 \leq \mu_0 \left(\left\|A_{0h}^{1/2} u_h(0)\right\|_h^2 + \|\dot{u}_h(0)\|_h^2 \right) \exp(-\nu_0 t). \tag{7.7.5}$$

Here, several other viscosity operators could have been chosen: We refer to [14] for the precise assumptions required on the viscosity operator introduced in (7.7.4) for which we can guarantee uniform stabilization results.

Note that systems (7.7.4) are similar to the numerical approximation schemes of the 1d and 2d wave equations studied in [35, 34, 27], which were dealt with using multiplier techniques. In [35, 34, 27], the viscosity term $h^2 A_{0h}$, instead of $h^{2\theta/3} A_{0h}$ in our setting, has been proved to be sufficient to guarantee the uniform exponential decay of the energy. However, the range of applications of [35, 34, 27] is limited to the case of uniform meshes and of wave equations with constant velocity.

Systems (7.7.4) are also similar to the ones in [32], where uniform stabilization results are derived for general damped wave equations (7.7.1) using a non-trivial spectral conditions. Especially, it is proved in [32] that systems (7.7.4) are uniformly exponentially stable with a weaker viscosity term: Namely, the viscosity term needed in [32] is $h^\theta A_{0h}$ instead of $h^{2\theta/3} A_{0h}$. However, in [32], a non-trivial spectral gap condition on the eigenvalues of A_0 is needed, which restricts the range of direct applications to the 1d case only.

Thus, in many situations, our results are not sharp. However, they apply for a wide range of applications: Especially, no condition is required on the dimension or on the uniformity of the meshes.

Remark 7.7.2. One can use the results in [14] to derive fully discrete approximation schemes of (7.7.1) for which one can guarantee uniform (in both time and space discretization parameters) stabilization properties.

7.8 Other models

In this section, we mention two other models of interest, for which our methods apply and yield new results.

7.8.1 A wave equation observed through $y(t) = Bu(t)$

Here, rather than studying an observation operator which involves the time derivative of solutions of (7.1.1) as in (7.1.3), we focus on the case of an observation of the form

$$y(t) = Bu(t). \tag{7.8.1}$$

The operator B is now assumed to belong to $\mathfrak{L}(\mathcal{D}(A_0), Y)$, where Y is an Hilbert space.

Now, the admissibility property for (7.1.1)-(7.8.1) consists in the existence, for every $T > 0$, of a constant K_T such that any solution of (7.1.1) with initial data $(u_0, u_1) \in \mathcal{D}(A_0) \times \mathcal{D}(A_0^{1/2})$ satisfies

$$\int_0^T \|Bu(t)\|_Y^2 \, dt \le K_T \Big(\left\| A_0^{1/2} u_0 \right\|_X^2 + \|u_1\|_X^2 \Big). \tag{7.8.2}$$

In particular, when B belongs to $\mathfrak{L}(\mathcal{D}(A_0^{1/2}), Y)$, system (7.1.1)-(7.8.1) is obviously admissible because of the conservation of the energy (7.1.2).

The observability property for (7.1.1)-(7.8.1) now reads as follows: There exist a time T and a positive constant $k_T > 0$ such that

$$k_T\Big(\left\|A_0^{1/2}u_0\right\|_X^2 + \|u_1\|_X^2 \Big) \leq \int_0^T \|Bu(t)\|_Y^2\, dt. \tag{7.8.3}$$

Similarly as before, assuming that system (7.1.1)-(7.8.1) is admissible and exactly observable, one can ask if the discrete systems (7.1.8) observed through

$$y_h(t) = B\pi_h u_h(t), \tag{7.8.4}$$

are uniformly admissible and exactly observable in a convenient filtered class.

Below, we provide a partial answer to that question. As before, we can only consider operators B which belong to $\mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$ for $\kappa < 1/2$. This makes the admissibility properties obvious since the observation operators $B_h = B\pi_h$ are then uniformly bounded as operators from V_h endowed with the norm $\left\|A_{0h}^{1/2}\cdot\right\|_h = \left\|A_0^{1/2}\pi_h\cdot\right\|_X$ (see (7.3.6)) to Y.

We therefore focus on the observability properties of (7.1.8)-(7.8.4), for which we obtain the following:

Theorem 7.8.1. *Let A_0 be a self-adjoint positive definite operator with compact resolvent and $B \in \mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$ with $\kappa < 1/2$. Assume that the maps (π_h) satisfy property (7.1.10). Set $\varsigma = 2\theta/3$.*

Assume that system (7.1.1)-(7.8.1) is exactly observable. Then there exist $\epsilon > 0$, a time T^* and a positive constant $k_* > 0$ such that, for any $h > 0$, any solution of (7.1.8) with initial data $(u_{0h}, u_{1h}) \in \mathcal{C}_h(\epsilon/h^\varsigma)^2$ satisfies

$$k_*\Big(\left\|A_{0h}^{1/2}u_{0h}\right\|_h^2 + \|u_{1h}\|_h^2 \Big) \leq \int_0^{T^*} \|B\pi_h u_h(t)\|_Y^2\ dt. \tag{7.8.5}$$

The proof of Theorem 7.8.1 is based on the following spectral characterization, which can be deduced from Theorems 7.2.4-7.2.5:

Theorem 7.8.2. *Let A_0 be a self-adjoint positive definite operator on X with compact resolvent and $B \in \mathfrak{L}(\mathcal{D}(A_0), Y)$. Assume that system (7.1.1)-(7.8.1) is admissible in the sense of (7.8.2).*

Then the following statements are equivalent:

1. *System (7.1.1)-(7.8.1) is exactly observable.*
2. *There exist positive constants m and M such that*

$$M^2 \left\|(A_0 - \omega^2 I)u\right\|_X^2 + m^2 \|Bu\|_Y^2 \geq \left\|A_0^{1/2}u\right\|_X^2,\ \forall u \in \mathcal{D}(A_0),\ \forall \omega \in \mathbb{R}. \tag{7.8.6}$$

3. *There exist positive constants α and β such that*

$$\left\|A_0^{1/2}u\right\|_X^4 \leq \|u\|_X^2 \Big(\|A_0 u\|_X^2 + \alpha^2 \|Bu\|_Y^2 - \beta^2 \left\|A_0^{1/2}u\right\|_X^2 \Big),\ \forall u \in \mathcal{D}(A_0). \tag{7.8.7}$$

Besides, assuming that the first eigenvalue of A_0 is bounded from below by a positive constant $\gamma > 0$, if one of the statements 2 or 3 holds, then the time T and the constants k_T in (7.8.3) can be chosen explicitly as functions of γ, the admissibility constants and either (m, M) or (α, β).

The proof of Theorem 7.8.2 is left to the reader. We only briefly indicate the method one can use to show Theorem 7.8.2.

To prove that statement 2 in Theorem 7.8.2 is equivalent to the exact observability of (7.1.1)-(7.8.1), one can follow the proof of Theorem 7.2.3 in [31] and use the refined version of Theorem 7.2.4 given in Theorem 7.2.5.

The equivalence of statements 2 and 3 follows from the same arguments as in Theorem 7.2.11.

Once Theorem 7.8.2 is proved, one only needs to prove that for $h > 0$ small enough, there exist positive constants α_* and β_* such that

$$\left\|A_{0h}^{1/2}u_h\right\|_h^4 \leq \|u_h\|_h^2 \left(\|A_{0h}u\|_h^2 + \alpha_*^2 \|B_h u_h\|_Y^2 - \beta_*^2 \left\|A_{0h}^{1/2}u_h\right\|_h^2 \right), \tag{7.8.8}$$

for any $u_h \in \mathcal{C}_h(\epsilon/h^\varsigma)$. The proof of (7.8.8) can be done similarly as in Subsection 7.3.2 and is also left to the reader.

Remark 7.8.3. When observing the solutions of the wave equation with Dirichlet boundary conditions via their normal derivative on a part of the boundary which satisfies the *Geometric Control Condition*, the observation operator is not continuous on $\mathcal{D}(A_0^{1/2})$, and thus our results do not apply. This issue deserves further work.

7.8.2 Applications to Schrödinger type equations

In this section, we focus on the consequences of Theorem 7.1.3 to the study of Schrödinger type equations

$$i\dot{z}(t) = A_0 z(t), \quad t \in \mathbb{R}, \qquad z(0) = z_0 \in X, \tag{7.8.9}$$

observed through

$$y(t) = Bz(t). \tag{7.8.10}$$

The admissibility property for (7.8.9)-(7.8.10) reads as

$$\int_0^T \|Bz(t)\|_Y^2 \, dt \leq K_T \|z_0\|_X^2, \qquad \forall\, z_0 \in \mathcal{D}(A_0), \tag{7.8.11}$$

and the exact observability property as

$$k_T \|z_0\|_X^2 \leq \int_0^T \|Bz(t)\|_Y^2 \, dt, \qquad \forall\, z_0 \in \mathcal{D}(A_0). \tag{7.8.12}$$

The results in [26] imply that if the system (7.1.1)-(7.1.3) is admissible and exactly observable in some time $T^* > 0$, then system (7.8.9)-(7.8.10) is admissible and exactly observable in any time $T > 0$.

Below, we adapt this strategy to deduce admissibility and exact observability results for the space semi-discrete approximation schemes of (7.8.9)-(7.8.10).

When discretizing (7.8.9) using finite element methods described by (V_h, π_h) as in the introduction, we obtain (see [10])

$$i\dot{z}_h = A_{0h} z_h, \quad t \in \mathbb{R}, \qquad z_h(0) = z_{0h} \in V_h. \tag{7.8.13}$$

The natural observation operator is then

$$y_h(t) = B_h z_h(t) = B\pi_h z_h(t). \tag{7.8.14}$$

We then prove the following result:

Theorem 7.8.4. *Let A_0 be a positive definite unbounded operator with compact resolvent and $B \in \mathfrak{L}(\mathcal{D}(A_0^\kappa), Y)$, with $\kappa < 1/2$. Assume that the approximations $(\pi_h)_{h>0}$ satisfy property* (7.1.10)*. Set σ as in* (7.1.12)*.*

Admissibility: *Assume that system* (7.1.1)-(7.1.3) *is admissible.*

Then, for any $\eta > 0$ and $T > 0$, there exists a positive constant $K_{T,\eta} > 0$ such that, for any $h > 0$, any solution of (7.8.13) *with initial data*

$$z_{0h} \in \mathcal{C}_h(\eta/h^\sigma) \tag{7.8.15}$$

satisfies

$$\int_0^T \|B_h z_h(t)\|_Y^2 \ dt \leq K_{T,\eta} \|z_{0h}\|_h^2 . \tag{7.8.16}$$

Observability: *Assume that system* (7.1.1)-(7.1.3) *is admissible and exactly observable.*

Then there exist $\epsilon > 0$, a time T^* and a positive constant $k_* > 0$ such that, for any $h > 0$, any solution of (7.8.13) *with initial data*

$$z_{0h} \in \mathcal{C}_h(\epsilon/h^\sigma) \tag{7.8.17}$$

satisfies

$$k_* \|z_{0h}\|_h^2 \leq \int_0^{T^*} \|B_h z_h(t)\|_Y^2 \ dt. \tag{7.8.18}$$

This result has to be compared with the ones in [10]. Indeed, in [10], under the assumption that system (7.8.9)-(7.8.10) is admissible and exactly observable, it is proved that finite element approximation schemes (7.8.13)-(7.8.14) are admissible and exactly observable for initial data filtered at the scale

$$\tilde{\sigma} = \theta \min\Big\{2(1-2\kappa), \frac{2}{5}\Big\}.$$

Theorem 7.8.4 then states a stronger result than [10], but under the stronger assumption that (7.1.1)-(7.1.3) is admissible and exactly observable.

Proof. Consider the wave system (7.1.1)-(7.1.3). Note that we are in the setting of Theorem 7.1.3. Below, we only prove the exact observability property for (7.8.13)-(7.8.14). The proof of the admissibility properties (7.8.16) is similar and is left to the reader.

Assume then that system (7.1.1)-(7.1.3) is admissible and exactly observable. Then, from Theorem 7.1.3, the admissibility and exact observability properties hold in a filtered class $\mathcal{C}_h(\epsilon/h^\sigma)$, uniformly with respect to $h > 0$, for systems (7.1.8).

By Theorem 7.2.4, there exist positive constants $\tilde{\alpha}$ and $\tilde{\beta}$ such that for all $h > 0$, for all $\tilde{\omega} \in \mathbb{R}$, for any wave packet

$$u_h = \frac{1}{\sqrt{2}} \sum_{\substack{|\mu_j^h - \tilde{\omega}| \leq \tilde{\alpha}, \\ \lambda_j^h \leq \epsilon/h^\sigma}} a_j \begin{pmatrix} \dfrac{i}{\mu_j^h} \Phi_j^h \\ \Phi_j^h \end{pmatrix} = \begin{pmatrix} u_{0h} \\ u_{1h} \end{pmatrix},$$

where $\mu_j^h = \sqrt{\lambda_j^h}$ for $j > 0$, and $-\sqrt{\lambda_j^h}$ for $j < 0$, the following inequality holds

$$\|B_h u_{1h}\|_Y^2 \geq \tilde{\beta}^2 \Big(\|u_{1h}\|_h^2 + \left\| A_{0h}^{1/2} u_{0h} \right\|_h^2 \Big) = 2\tilde{\beta}^2 \|u_{1h}\|_h^2 . \tag{7.8.19}$$

Now, take a positive number ω, and consider z_h a wave packet

$$z_h = \sum_{\substack{|\lambda_j^h - \omega| \leq \alpha, \\ \lambda_j^h \leq \epsilon / h^\sigma}} a_j \Phi_j^h, \tag{7.8.20}$$

where α will be chosen later on. Remark that, if

$$|\lambda_j^h - \omega| \leq \alpha,$$

then

$$|\mu_j^h - \sqrt{\omega}\,| = \left| \sqrt{\lambda_j^h} - \sqrt{\omega} \right| \leq \frac{\alpha}{\mu_j^h + \sqrt{\omega}} \leq \frac{\alpha}{\sqrt{\lambda_1^h}} \leq \frac{\alpha}{\sqrt{\lambda_1}},$$

where the last estimates come from the positivity of ω and (7.3.24).

Therefore, if $\alpha \leq \tilde{\alpha}\sqrt{\lambda_1}$, applying (7.8.19) in $\tilde{\omega} = \sqrt{\omega}$ to

$$u_h = \begin{pmatrix} \displaystyle\sum_{\substack{|\lambda_j^h - \omega| \leq \alpha, \\ \lambda_j^h \leq \epsilon / h^\sigma}} a_j \frac{1}{\sqrt{\lambda_j^h}} \Phi_j^h \\ \\ z_h \end{pmatrix},$$

we get that for all $\omega \in (0, \infty)$, for any wave packet z_h as in (7.8.20), with $\alpha \leq \tilde{\alpha}\sqrt{\lambda_1}$,

$$\|B_h z_h\|_Y \geq \sqrt{2}\tilde{\beta} \|z_h\|_h .$$

Criterion (7.2.20) for (7.8.13)-(7.8.14) follows, uniformly with respect to $h > 0$, by taking

$$\alpha = \min\{\tilde{\alpha}\sqrt{\lambda_1}, \sqrt{\lambda_1}\}, \quad \text{and } \beta = \sqrt{2}\tilde{\beta}.$$

Indeed, this choice guarantees that, for $\omega \leq 0$, $J_\alpha(\omega)$ is empty.

Therefore Theorem 7.2.5 applies and yields (7.8.18). □

Under the assumptions of Theorem 7.8.4, it is very likely that systems (7.8.13)-(7.8.14) are uniformly exactly observable in any time $T > 0$, but our methods do not yield this result. Indeed, the proof of [26] in the continuous setting does not apply in our case. It uses a compactness argument to deal with the low-frequency components of the solutions, and this cannot be done in our setting.

7.9 Further comments

1. One of the interesting features of the approach presented here is that it works in any dimension and in a very general setting. To our knowledge, this is the first work (namely with the companion paper [10]) which proves in a systematic way observability properties for space semi-discrete systems from the ones of the continuous setting.

2. A widely open question consists in finding the sharp filtering scale. We think that the works [6, 7], which present a study of the observability properties of the 1d wave equation in highly heterogeneous media, might give some insights to address this issue. In [6, Paragraph 3.3.1], it is interesting to notice that, as in Theorem 7.1.3, the exponent 2/3 appears naturally as a critical value when comparing the spectrum of the wave operators corresponding to the oscillating media and the one of the homogenized wave operator. Though, in [6], it is proved that observability properties still hold when filtering the data at a higher scale.

3. In this article, we assumed that the continuous systems are exactly observable. However, there are several important models of vibrations where the energy is only weakly observable. That is the case for instance for networks of vibrating strings [8] or when the *Geometric Control Condition* is not fulfilled (see [2, 22]). It would be interesting to address the observability issues for the space semi-discretizations of such systems. To our knowledge, this issue is widely open.

Acknowledgements. The author acknowledges Jean-Pierre Puel, Enrique Zuazua and Marius Tucsnak for their fruitful comments.

Bibliography

[1] H. T. Banks, K. Ito, and C. Wang. Exponentially stable approximations of weakly damped wave equations. In *Estimation and control of distributed parameter systems (Vorau, 1990)*, volume 100 of *Internat. Ser. Numer. Math.*, pages 1–33. Birkhäuser, Basel, 1991.

[2] C. Bardos, G. Lebeau, and J. Rauch. Sharp sufficient conditions for the observation, control and stabilization of waves from the boundary. *SIAM J. Control and Optimization*, 30(5):1024–1065, 1992.

[3] N. Burq and P. Gérard. Condition nécessaire et suffisante pour la contrôlabilité exacte des ondes. *C. R. Acad. Sci. Paris Sér. I Math.*, 325(7):749–752, 1997.

[4] N. Burq and M. Zworski. Geometric control in the presence of a black box. *J. Amer. Math. Soc.*, 17(2):443–471 (electronic), 2004.

[5] C. Castro and S. Micu. Boundary controllability of a linear semi-discrete 1-d wave equation derived from a mixed finite element method. *Numer. Math.*, 102(3):413–462, 2006.

[6] C. Castro and E. Zuazua. Low frequency asymptotic analysis of a string with rapidly oscillating density. *SIAM J. Appl. Math.*, 60(4):1205–1233 (electronic), 2000.

[7] C. Castro and E. Zuazua. Concentration and lack of observability of waves in highly heterogeneous media. *Arch. Ration. Mech. Anal.*, 164(1):39–72, 2002.

[8] R. Dáger and E. Zuazua. *Wave propagation, observation and control in* 1*-d flexible multi-structures*, volume 50 of *Mathématiques & Applications (Berlin)*. Springer-Verlag, Berlin, 2006.

[9] S. Ervedoza. Observability of the mixed finite element method for the 1d wave equation on non-uniform meshes. *To appear in ESAIM: COCV*, 2008. *Cf Chapitre 2.*

[10] S. Ervedoza. Admissibility and observability for Schrödinger systems: Applications to finite element approximation schemes. *To be published*, 2008. *Cf Chapitre 6.*

[11] S. Ervedoza, C. Zheng, and E. Zuazua. On the observability of time-discrete conservative linear systems. *J. Funct. Anal.*, 254(12):3037–3078, June 2008. *Cf Chapitre 3.*

[12] S. Ervedoza and E. Zuazua. Perfectly matched layers in 1-d: Energy decay for continuous and semi-discrete waves. *Numer. Math.*, 109(4):597–634, 2008. *Cf Chapitre 1.*

[13] S. Ervedoza and E. Zuazua. Uniform exponential decay for viscous damped systems. *To appear in Proc. of Siena "Phase Space Analysis of PDEs 2007", Special issue in honor of Ferrucio Colombini*, 2008. *Cf Chapitre 4.*

[14] S. Ervedoza and E. Zuazua. Uniformly exponentially stable approximations for a class of damped systems. *To appear in J. Math. Pures Appl.*, 2008. *Cf Chapitre 5.*

[15] R. Glowinski. Ensuring well-posedness by analogy: Stokes problem and boundary control for the wave equation. *J. Comput. Phys.*, 103(2):189–221, 1992.

[16] R. Glowinski, C. H. Li, and J.-L. Lions. A numerical approach to the exact boundary controllability of the wave equation. I. Dirichlet controls: description of the numerical methods. *Japan J. Appl. Math.*, 7(1):1–76, 1990.

[17] A. Haraux. Une remarque sur la stabilisation de certains systèmes du deuxième ordre en temps. *Portugal. Math.*, 46(3):245–258, 1989.

[18] O. Y.. Imanuvilov. On Carleman estimates for hyperbolic equations. *Asymptot. Anal.*, 32(3-4):185–220, 2002.

[19] J.A. Infante and E. Zuazua. Boundary observability for the space semi discretizations of the 1-d wave equation. *Math. Model. Num. Ann.*, 33:407–438, 1999.

[20] A. E. Ingham. Some trigonometrical inequalities with applications to the theory of series. *Math. Z.*, 41(1):367–379, 1936.

[21] V. Komornik. *Exact controllability and stabilization.* RAM: Research in Applied Mathematics. Masson, Paris, 1994. The multiplier method.

[22] G. Lebeau. Équations des ondes amorties. *Séminaire sur les Équations aux Dérivées Partielles, 1993–1994,École Polytech.*, 1994.

[23] J.-L. Lions. *Contrôlabilité exacte, Stabilisation et Perturbations de Systèmes Distribués. Tome 1. Contrôlabilité exacte*, volume RMA 8. Masson, 1988.

[24] K. Liu. Locally distributed control and damping for the conservative systems. *SIAM J. Control Optim.*, 35(5):1574–1590, 1997.

[25] F. Macià. The effect of group velocity in the numerical analysis of control problems for the wave equation. In *Mathematical and numerical aspects of wave propagation—WAVES 2003*, pages 195–200. Springer, Berlin, 2003.

[26] L. Miller. Controllability cost of conservative systems: resolvent condition and transmutation. *J. Funct. Anal.*, 218(2):425–444, 2005.

[27] A. Münch and A. F. Pazoto. Uniform stabilization of a viscous numerical approximation for a locally damped wave equation. *ESAIM Control Optim. Calc. Var.*, 13(2):265–293 (electronic), 2007.

[28] M. Negreanu, A.-M. Matache, and C. Schwab. Wavelet filtering for exact controllability of the wave equation. *SIAM J. Sci. Comput.*, 28(5):1851–1885 (electronic), 2006.

[29] M. Negreanu and E. Zuazua. Convergence of a multigrid method for the controllability of a 1-d wave equation. *C. R. Math. Acad. Sci. Paris*, 338(5):413–418, 2004.

[30] A. Osses. A rotated multiplier applied to the controllability of waves, elasticity, and tangential Stokes control. *SIAM J. Control Optim.*, 40(3):777–800 (electronic), 2001.

[31] K. Ramdani, T. Takahashi, G. Tenenbaum, and M. Tucsnak. A spectral approach for the exact observability of infinite-dimensional systems with skew-adjoint generator. *J. Funct. Anal.*, 226(1):193–229, 2005.

[32] K. Ramdani, T. Takahashi, and M. Tucsnak. Uniformly exponentially stable approximations for a class of second order evolution equations—application to LQR problems. *ESAIM Control Optim. Calc. Var.*, 13(3):503–527, 2007.

[33] P.-A. Raviart and J.-M. Thomas. *Introduction à l'analyse numérique des équations aux dérivées partielles.* Collection Mathématiques Appliquées pour la Maitrise. Masson, Paris, 1983.

[34] L. R. Tcheugoué Tebou and E. Zuazua. Uniform boundary stabilization of the finite difference space discretization of the $1-d$ wave equation. *Adv. Comput. Math.*, 26(1-3):337–365, 2007.

[35] L.R. Tcheugoué Tébou and E. Zuazua. Uniform exponential long time decay for the space semi-discretization of a locally damped wave equation via an artificial numerical viscosity. *Numer. Math.*, 95(3):563–598, 2003.

[36] L. N. Trefethen. Group velocity in finite difference schemes. *SIAM Rev.*, 24(2):113–136, 1982.

[37] M. Tucsnak and G. Weiss. Observation and control for operator semigroups, 2008.

[38] R. M. Young. *An introduction to nonharmonic Fourier series*. Academic Press Inc., San Diego, CA, first edition, 2001.

[39] X. Zhang. Explicit observability estimate for the wave equation with potential and its application. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 456(1997):1101–1115, 2000.

[40] E. Zuazua. Boundary observability for the finite-difference space semi-discretizations of the 2-D wave equation in the square. *J. Math. Pures Appl. (9)*, 78(5):523–563, 1999.

[41] E. Zuazua. Propagation, observation, and control of waves approximated by finite difference methods. *SIAM Rev.*, 47(2):197–243 (electronic), 2005.

Part IV

Miscellaneous

Chapter 8

Control and stabilization property for a singular heat equation with an inverse square potential

Abstract: The goal of this article is to analyze control properties of parabolic equations with a singular potential $-\mu/|x|^2$, where μ is a real number. When $\mu \leq (N-2)^2/4$, it was proved in [19] that the equation can be controlled to zero with a distributed control which surrounds the singularity. In the present work, using Carleman estimates, we will prove that this assumption is not necessary, and that we can control the equation from any open subset as for the heat equation. Then we will study the case $\mu > (N-2)^2/4$, and prove that the situation changes completely: Indeed, we will consider a sequence of regularized potentials $\mu/(|x|^2+\varepsilon^2)$, and prove that we cannot stabilize the corresponding systems uniformly with respect to $\varepsilon > 0$, due to the presence of explosive modes which concentrate around the singularity.

8.1 Introduction

Let $N \geq 3$ and consider a smooth bounded domain $\Omega \subseteq \mathbb{R}^N$ such that $0 \in \Omega$, and let $\omega \subset \Omega$ be a non-empty open set.
We are interested in the control and stabilization properties of the following equation

$$\begin{cases} \partial_t u - \Delta_x u - \dfrac{\mu}{|x|^2} u = f, & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) = u_0(x), & x \in \Omega, \end{cases} \tag{8.1.1}$$

where $u_0 \in L^2(\Omega)$. Here, $f \in L^2((0,T); H^{-1}(\Omega))$ is the control that we assume to be null in $\Omega \backslash \bar{\omega}$, that is

$$\forall \theta \in \mathcal{D}(\Omega \backslash \bar{\omega}), \quad \theta f = 0 \quad \text{in } L^2((0,T); H^{-1}(\Omega)). \tag{8.1.2}$$

First of all, let us briefly mention that the Cauchy problem with such singular potential is not straightforward. Indeed, it has been proved that there is a critical value $\mu^*(N) = (N-2)^2/4$ of μ

which determines the well-posedness of (8.1.1). Actually, this problem is strongly related to the Hardy inequality:

$$\forall u \in H_0^1(\Omega), \qquad \mu^*(N) \int_\Omega \frac{u^2}{|x|^2}\ \mathrm{dx} \leq \int_\Omega |\nabla u|^2\ \mathrm{dx}, \tag{8.1.3}$$

where $\mu^*(N)$ is the optimal constant. Note that equality in (8.1.3) is not attained.

The first work [1] on the Cauchy problem was considering positive initial data. In [1], it was proved that if $\mu \leq \mu^*(N)$ and if the initial data u_0 is positive, then equation (8.1.1) has a global weak solution whereas if $\mu > \mu^*(N)$, then equation (8.1.1) has no solution if $u_0 > 0$ and $f \geq 0$, even locally in time (see also [4]).

Actually, the Cauchy problem properties for equation (8.1.1) can be deduced from generalizations of the Hardy inequality (8.1.3). Studying more precisely (8.1.3), it is proved in [20] that the Cauchy problem is well-posed in $L^2(\Omega)$ for any $\mu \leq \mu^*(N)$. A precise functional setting is given even in the special case $\mu = \mu^*(N)$ (see [20]).

The objective of the present paper is twofold. First, when $\mu \leq \mu^*(N)$, we will prove the null-controllability of (8.1.1) with a control $f \in L^2((0,T); L^2(\omega))$. Second, we will show that when $\mu > \mu^*(N)$, there is no way to stabilize system (8.1.1) with a control supported in ω in a reasonable sense when $0 \notin \bar{\omega}$.

The null-controllability problem reads as follows: Given any $u_0 \in L^2(\Omega)$, find a function $f \in L^2(\omega \times (0,T))$ such that the solution of (8.1.1) satisfies

$$u(x,T) = 0, \qquad x \in \Omega. \tag{8.1.4}$$

The controllability issue was already discussed under the assumption $\mu \leq \mu^*(N)$ in the recent work [19], in the special case where ω contains an annulus centered in the singularity. The authors of [19] need this assumption since their proof strongly uses a decomposition in spherical harmonics which allows to reduce the problem to the study of 1-d singular equations. J. Le Rousseau mentioned an argument in [19] to relax this strong geometric assumption into these two conditions: ω circles the singularity, and the exterior part of ω contains an annular set centered in the singularity. Even with this improvement, a non-trivial geometric assumption on ω is needed. Our purpose is to prove that we can actually remove this assumption and consider any non-empty open subset ω of Ω.

Theorem 8.1.1. *Let μ be a real number such that $\mu \leq \mu^*(N)$.*

Given any non-empty open set $\omega \subset \Omega$, for any $T > 0$ and $u_0 \in L^2(\Omega)$, there exists a control $f \in L^2((0,T) \times \omega)$ such that the solution of (8.1.1) *satisfies* (8.1.4). *Besides, there exists a constant C_T such that*

$$\|f\|_{L^2((0,T)\times\omega)} \leq C_T \|u_0\|_{L^2(\Omega)}. \tag{8.1.5}$$

Following the by now classical HUM method ([16]), the controllability property is equivalent to an observability inequality for the adjoint system

$$\begin{cases} \partial_t w + \Delta_x w + \dfrac{\mu}{|x|^2} w = 0, & (x,t) \in \Omega \times (0,T), \\ w(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ w(x,T) = w_T(x), & x \in \Omega. \end{cases} \tag{8.1.6}$$

More precisely, when $\mu \leq \mu^*(N)$, we need to prove that there exists a constant C such that for all $w_T \in L^2(\Omega)$, the solution of (8.1.6) satisfies

$$\int_\Omega |w(x,0)|^2 \, \mathrm{dx} \leq C \iint_{\omega\times(0,T)} |w(x,t)|^2 \, \mathrm{dx} \, \mathrm{dt}. \tag{8.1.7}$$

In order to prove (8.1.7), we will use a particular Carleman estimate, which is by now a classical technique in control theory, see for instance [2, 9, 10, 11, 12, 13, 14]... Indeed, the Carleman estimate we will derive later implies that for any solution w of (8.1.6),

$$\iint_{\Omega\times(\frac{T}{4},\frac{3T}{4})} |w(x,t)|^2 \, \mathrm{dx} \, \mathrm{dt} \leq C \iint_{\omega\times(0,T)} |w(x,t)|^2 \, \mathrm{dx} \, \mathrm{dt}, \tag{8.1.8}$$

which directly implies inequality (8.1.7) since $t \mapsto \|w(t,.)\|^2_{L^2(\Omega)}$ is increasing by the Hardy inequality (8.1.3).

The Carleman estimate derived here is inspired by the works [5, 17] on 1-d degenerate heat equations, the recent paper [19] which is inspired from the methods and results in [5, 17] to obtain radial estimates, and the article [13] on the controllability of the heat equation in any dimension. As in [5, 17, 19, 13], the major difficulty is to choose a special weight function appearing in the Carleman estimate. In [19], this has been done in the 1d case only, using spherical harmonics to recover results in the multi-d case, but with an extra geometric condition on the support of the control region. We thus adapt the results in [19] to derive directly Carleman estimates without using a spherical harmonics decomposition, in order to avoid the use of the geometric condition needed in [19].

Let us briefly present the existing results concerning the observability properties of a parabolic equation with a potential V:

$$\begin{cases} \partial_t z + \Delta_x z + V\, z = 0, & (x,t) \in \Omega \times (0,T), \\ z(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ z(T) = z_T \in L^2(\Omega). \end{cases} \tag{8.1.9}$$

It has been proved in [13] using Carleman estimates that, for potentials $V \in L^\infty(\Omega\times(0,T))$, such systems are observable in the sense of (8.1.7) for any open set $\omega \subset \Omega$. Later, in [14], this result has been extended to the case $V \in L^\infty((0,T); L^{2N/3}(\Omega))$. To our knowledge, the case $V \in L^\infty((0,T); L^{N/2+\epsilon}(\Omega))$ with $\epsilon > 0$ is still open. Note that our work presents a case in which the potential $V = \mu/|x|^2$ is not in $L^{N/2}(\Omega)$, and therefore none of these results applies. In this context, it is worth mentioning the work [15] which proves the strong unique continuation property for system (8.1.9) for a general potential $V \in L^{(N+1)/2}(\Omega \times (0,T))$.

The second part of this work is devoted to the case $\mu > \mu^*(N)$. In this case, the Cauchy problem is severely ill-posed as proved in [1] and [4]. Indeed, if u_0 is positive and $f = 0$ in (8.1.1), there is complete instantaneous blow-up, which makes impossible to define a reasonable solution. However, it does not answer to the following stabilization problem:

Given $u_0 \in L^2(\Omega)$, can we find a control $f \in L^2((0,T); H^{-1}(\Omega))$ localized in ω such that there exists a solution $u \in L^2((0,T); H^1_0(\Omega))$ of (8.1.1) ?

In other words, we ask whether it is possible or not to prevent from blow-up phenomena by acting only on a subset. Before going further, note that if $u \in L^2((0,T); H^1_0(\Omega))$ satisfies (8.1.1) with

$f \in L^2((0,T);H^{-1}(\Omega))$, then $\partial_t u \in L^2((0,T);H^{-1}(\Omega))$, and therefore $u \in C([0,T];L^2(\Omega))$, and the equality $u(0) = u_0$ in (8.1.1) makes sense.

Following the ideas of optimal control, for any $u_0 \in L^2(\Omega)$, we consider the functional

$$J_{u_0}(u,f) = \frac{1}{2} \iint\limits_{\Omega\times(0,T)} |u(t,x)|^2 \, \mathrm{dx} \, \mathrm{dt} + \frac{1}{2} \int\limits_0^T \|f(t)\|^2_{H^{-1}(\Omega)} \, \mathrm{dt}, \tag{8.1.10}$$

defined on the set

$$\mathcal{C}(u_0) = \Big\{ (u,f) \in L^2((0,T);H^1_0(\Omega)) \times L^2((0,T);H^{-1}(\Omega)) \text{ such that } u \text{ satisfies (8.1.1) with } f \text{ as in (8.1.2)} \Big\}. \tag{8.1.11}$$

We say that we can stabilize system (8.1.1) if we can find a constant C such that

$$\forall u_0 \in L^2(\Omega), \qquad \inf_{(u,f)\in\mathcal{C}(u_0)} J_{u_0}(u,f) \leq C \, \|u_0\|^2_{L^2(\Omega)} \, . \tag{8.1.12}$$

Of course, this property strongly depends on the set ω where the stabilization is effective. Especially, when $0 \in \omega$, (8.1.12) holds (see Section 8.4 B1).

When $0 \notin \bar{\omega}$, the situation is more intricate. Therefore we focus our study on this particular case, and give a severe obstruction, in this case, to the stabilization property (8.1.12).

More precisely, for $\varepsilon > 0$, we approximate (8.1.1) by the systems

$$\begin{cases} \partial_t u - \Delta_x u - \dfrac{\mu}{|x|^2+\varepsilon^2} u = f, & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) = u_0(x), & x \in \Omega. \end{cases} \tag{8.1.13}$$

For these approximate problems, the Cauchy problem is well-posed. Therefore we can consider the functionals

$$J^\varepsilon_{u_0}(f) = \frac{1}{2} \iint\limits_{\Omega\times(0,T)} |u(x,t)|^2 \, \mathrm{dx} \, \mathrm{dt} + \frac{1}{2} \int\limits_0^T \|f(t)\|^2_{H^{-1}(\Omega)} \, \mathrm{dt}, \tag{8.1.14}$$

where $f \in L^2((0,T);H^{-1}(\Omega))$ is localized in ω in the sense of (8.1.2) and u is the corresponding solution of (8.1.13). We prove the following:

Theorem 8.1.2. *Assume that $\mu > \mu^*(N)$, and that $0 \notin \bar{\omega}$.*

There is no constant C such that for all $\varepsilon > 0$, and for all $u_0 \in L^2(\Omega)$,

$$\inf_{\substack{f \in L^2((0,T);H^{-1}(\Omega)) \\ f \text{ as in (8.1.2)}}} J^\varepsilon_{u_0}(f) \leq C \, \|u_0\|^2_{L^2(\Omega)} \, . \tag{8.1.15}$$

In particular, this result implies that the stabilization of (8.1.1) is impossible to attain through regularization processes when $\mu > \mu^*(N)$ and $0 \notin \bar{\omega}$, and that we cannot prevent the system from blowing up.

Let us briefly mention the related work [12], which presents a study of the control properties of weakly blowing-up semi-linear heat equations, which deals with a similar question as the one asked

here. In particular, in [12], examples of systems are given for which blow up may occur in finite time, but this blow-up can be controlled in any time for any initial data.

The structure of the paper is the following. In Section 8.2, we give the proof of Theorem 8.1.1 for $\mu \leq \mu^*(N)$, or, to be more precise, of inequality (8.1.7) for the solutions of the adjoint equation (8.1.6). In Section 8.3, we prove that when $\mu > \mu^*(N)$ we cannot uniformly stabilize system (8.1.1), in the sense of Theorem 8.1.2. In Section 8.4, we add some comments.

Acknowledgments.
The author acknowledges the hospitality and support of IMDEA Matemáticas, where this work was completed. The author would like to thank E. Zuazua for having invited him in the IMDEA several months and for having suggested this work. The author also thanks J.-P. Puel for fruitful discussions and remarks.

8.2 Null controllability in the case $\mu \leq \mu^*(N)$

First of all, to simplify the presentation, we assume that $0 \notin \bar{\omega}$, that can always be done, taking if necessary a smaller set. We also assume that the unit ball $\bar{B}(0,1)$ is included in Ω and $\bar{B}(0,1) \cap \bar{\omega}$ is empty. This can always be done by a scaling argument.

8.2.1 Carleman estimate

As said in the introduction, the main tool we use to address the observability inequality (8.1.8) is a Carleman estimate. However, since it is based on tedious computations, we postpone the proofs of several technical lemmas in Subsection 8.2.3.

The major problem when designing a Carleman estimate is the choice of a smooth weight function σ, which is in general assumed to be positive, and to blow up as t goes to zero and as t goes to T. Hence we are looking for a weight function σ that satisfies:

$$\begin{cases} \sigma(t,x) > 0, \quad (x,t) \in \Omega \times (0,T), \\ \lim\limits_{t\to 0^+} \sigma(t,x) = \lim\limits_{t\to T^-} \sigma(t,x) = +\infty, \quad x \in \Omega. \end{cases} \tag{8.2.1}$$

More precisely, we propose the weight

$$\sigma(t,x) = s\theta(t)\Big(e^{2\lambda \sup \psi} - \frac{1}{2}|x|^2 - e^{\lambda\psi(x)}\Big) \tag{8.2.2}$$

where s and λ are positive parameters aimed at being large,

$$\theta(t) = \Big(\frac{1}{t(T-t)}\Big)^3, \tag{8.2.3}$$

and ψ is a function satisfying

$$\begin{cases} \psi(x) = \ln(|x|), & x \in B(0,1), \\ \psi(x) = 0, & x \in \partial\Omega, \\ \psi(x) > 0, & x \in \Omega \backslash \bar{B}(0,1), \end{cases} \tag{8.2.4}$$

and there exists an open set ω_0 such that $\bar{\omega}_0 \subset \omega$ and $\delta > 0$ such that

$$|\nabla\psi(x)| \geq \delta, \qquad x \in \bar{\Omega}\backslash\omega_0. \tag{8.2.5}$$

The existence of such function ψ is not straightforward but can be easily deduced from the construction given in [13].

Indeed, there exists a smooth function which extends $\ln(|x|)$ outside the ball, which vanishes on the boundary, and with finitely many critical points, since this property is generically true. Then it is sufficient to consider such a function, and to move its critical points into ω_0 without modifying the function in $B(0,1)$. This can be done following the construction given in [13].

Note that the weight function σ defined by (8.2.2) indeed satisfies (8.2.1) and is smooth (at least in $C^4((0,T)\times\bar{\Omega})$) when λ is large enough.

To explain this choice for the weight function σ, we point out that in the ball $B(0,1)$, since ψ is negative, the weight function σ behaves like

$$s\theta(t)(C - \frac{1}{2}|x|^2)$$

when λ is large. This corresponds precisely to the weight given in [17] for dealing with singular 1-d heat-type equation and in [19] when dealing with the observability around the singularity. On the contrary, outside the unit ball, since ψ is positive, when λ is large enough, the weight is very close to the one used for the observability of the heat equation in [13].

To simplify notations, let us denote by ϕ the function

$$\phi(x) = e^{\lambda\psi(x)}, \tag{8.2.6}$$

by $\mathcal{O}$ the open set $\Omega\backslash(\bar{B}(0,1)\cup\bar{\omega}_0)$ and by $\tilde{\mathcal{O}}$ the open set $\Omega\backslash\bar{B}(0,1)$.

We are now in position to state the Carleman estimate.

Theorem 8.2.1. *There exist positive constants K and λ_0 such that for $\lambda \geq \lambda_0$, there exists $s_0(\lambda)$ such that for all $s \geq s_0$, any w solution of* (8.1.6) *satisfies*

$$\begin{aligned} s\lambda^2 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta\phi e^{-2\sigma}|\nabla w|^2 \,\mathrm{dx}\,\mathrm{dt} + s\iint\limits_{\Omega\times(0,T)} \theta e^{-2\sigma}\frac{|w|^2}{|x|}\,\mathrm{dx}\,\mathrm{dt} \\ + s^3\iint\limits_{\Omega\times(0,T)} \theta^3 e^{-2\sigma}|x|^2|w|^2\,\mathrm{dx}\,\mathrm{dt} + s^3\lambda^4\iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta^3\phi^3 e^{-2\sigma}|w|^2\,\mathrm{dx}\,\mathrm{dt} \\ \leq K\left(s\lambda^2\iint\limits_{\omega_0\times(0,T)} \theta\phi e^{-2\sigma}|\nabla w|^2\,\mathrm{dx}\,\mathrm{dt} + s^3\lambda^4\iint\limits_{\omega_0\times(0,T)} \theta^3\phi^3 e^{-2\sigma}|w|^2\,\mathrm{dx}\,\mathrm{dt}\right). \end{aligned} \tag{8.2.7}$$

Remark 8.2.2. Following the proof carefully, one can check that there exists a constant $s_1(\psi) > 0$ such that the choice

$$s_0(\lambda) = s_1 e^{3\lambda\sup\psi}$$

is convenient in Theorem 8.2.1.

Remark 8.2.3. We stated the Carleman estimate (8.2.7) in the restrictive setting that we need, but we can handle a source term. To be more precise, for any $w \in \mathcal{D}([0,T]\times\Omega)$, taking s and λ large enough, the following holds:

$$
\begin{aligned}
s\lambda^2 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta\phi e^{-2\sigma}|\nabla w|^2 \,\mathrm{dx}\,\mathrm{dt} + s\iint\limits_{\Omega\times(0,T)} \theta e^{-2\sigma}\frac{|w|^2}{|x|}\,\mathrm{dx}\,\mathrm{dt} + s(\mu^*(N)-\mu)\iint\limits_{\Omega\times(0,T)} \theta e^{-2\sigma}\frac{|w|^2}{|x|^2}\,\mathrm{dx}\,\mathrm{dt} \\
+ s^3\iint\limits_{\Omega\times(0,T)} \theta^3 e^{-2\sigma}|x|^2|w|^2\,\mathrm{dx}\,\mathrm{dt} + s^3\lambda^4\iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta^3\phi^3 e^{-2\sigma}|w|^2\,\mathrm{dx}\,\mathrm{dt} \\
\leq K\Bigg(\iint\limits_{\Omega\times(0,T)} e^{-2\sigma}\Big|\partial_t w + \Delta_x w + \frac{\mu}{|x|^2}w\Big|^2\,\mathrm{dx}\,\mathrm{dt} + s\lambda^2\iint\limits_{\omega_0\times(0,T)} \theta\phi e^{-2\sigma}|\nabla w|^2\,\mathrm{dx}\,\mathrm{dt} \\
+ s^3\lambda^4\iint\limits_{\omega_0\times(0,T)} \theta^3\phi^3 e^{-2\sigma}|w|^2\,\mathrm{dx}\,\mathrm{dt}\Bigg).
\end{aligned}
$$

Proof. We present the main ideas and steps of the proof of Theorem 8.2.1, using several technical Lemmas, that are proved later in Subsection 8.2.3.

Let us first remark that using the density the density of $H_0^1(\Omega)$ in $L^2(\Omega)$, if estimate (8.2.7) holds for any solution w of (8.1.6) with initial data $w_T \in H_0^1(\Omega)$, then (8.2.7) also holds for any solution w of (8.1.6) with initial data $w_T \in L^2(\Omega)$. We thus prove (8.2.7) only for solutions of (8.1.6) with initial data in $H_0^1(\Omega)$.

Now, let us assume that w is a solution of (8.1.6) for some initial data $w_T \in H_0^1(\Omega)$, and define

$$
z(t,x) = \exp(-\sigma(t,x))w(t,x), \tag{8.2.8}
$$

which obviously satisfies

$$
z(T) = z(0) = 0 \qquad \text{in } H_0^1(\Omega) \tag{8.2.9}
$$

due to the assumptions (8.2.1) on σ.

Then, plugging $w = z\exp(\sigma(t,x))$ in the equation (8.1.6), we obtain that z satisfies

$$
\partial_t z + \Delta_x z + \frac{\mu}{|x|^2}z + 2\nabla z\cdot\nabla\sigma + z\Delta_x\sigma + z\Big(\partial_t\sigma + |\nabla\sigma|^2\Big) = 0, \quad (x,t)\in\Omega\times(0,T), \tag{8.2.10}
$$

with the boundary condition

$$
z = 0, \quad (x,t)\in\partial\Omega\times(0,T). \tag{8.2.11}
$$

Let us define a smooth positive radial function $\alpha(x) = \alpha(|x|)$ such that

$$
\begin{aligned}
\alpha(x) = 0, \quad |x| \leq \frac{1}{2}, \qquad \alpha(x) = \frac{1}{N}, \quad |x| \geq \frac{3}{4}, \\
0 \leq \alpha(x) \leq \frac{1}{N}, \quad \frac{1}{2} \leq |x| \leq \frac{3}{4}.
\end{aligned} \tag{8.2.12}
$$

Setting

$$
Sz = \Delta_x z + \frac{\mu}{|x|^2}z + z\Big(\partial_t\sigma + |\nabla\sigma|^2\Big), \qquad Az = \partial_t z + 2\nabla z\cdot\nabla\sigma + z\Delta_x\sigma\Big(1+\alpha\Big), \tag{8.2.13}
$$

one easily deduces from (8.2.10) that

$$Sz + Az = -\alpha z \Delta_x \sigma, \qquad \|Sz\|^2 + \|Az\|^2 + 2 < Sz, Az >= \|\alpha z \Delta_x \sigma\|^2 ,$$

where $\|\cdot\|$ denotes the $L^2(\Omega \times (0,T))$ norm and $< \cdot, \cdot >$ the corresponding scalar product. Especially, the quantity

$$I =< Sz, Az > -\frac{1}{2} \|\alpha z \Delta_x \sigma\|^2 \tag{8.2.14}$$

is non positive.

Lemma 8.2.4. *The following equality holds:*

$$\begin{aligned} I = -2 \iint\limits_{\Omega\times(0,T)} & D^2\sigma(\nabla z, \nabla z) \text{ dx dt} + \iint\limits_{\partial\Omega\times(0,T)} |\partial_n z|^2 \ \partial_n \sigma \text{ds dt} \\ & - \iint\limits_{\Omega\times(0,T)} |\nabla z|^2 \Delta_x \sigma \ \alpha \text{ dx dt} + \frac{1}{2} \iint\limits_{\Omega\times(0,T)} |z|^2 \Delta_x^2 \sigma \Big(1+\alpha\Big) \text{ dx dt} \\ & + \iint\limits_{\Omega\times(0,T)} |z|^2 \nabla\alpha \cdot \nabla \Delta_x \sigma \text{ dx dt} + \frac{1}{2} \iint\limits_{\Omega\times(0,T)} |z|^2 \Delta_x \sigma \ \Delta_x \alpha \text{ dx dt} \\ & - \frac{1}{2} \iint\limits_{\Omega\times(0,T)} |z|^2 \Big(\partial_{tt}^2 \sigma + 2\partial_t \Big(|\nabla\sigma|^2\Big)\Big) \text{ dx dt} - 2 \iint\limits_{\Omega\times(0,T)} |z|^2 D^2\sigma\Big(\nabla\sigma, \nabla\sigma\Big) \text{ dx dt} \\ & + \iint\limits_{\Omega\times(0,T)} \alpha |z|^2 \Delta_x \sigma \Big(\partial_t \sigma + |\nabla\sigma|^2\Big) \text{ dx dt} - \frac{1}{2} \iint\limits_{\Omega\times(0,T)} \alpha^2 |z|^2 |\Delta_x \sigma|^2 \text{ dx dt} \\ & + \mu \iint\limits_{\Omega\times(0,T)} \frac{|z|^2}{|x|^2} \Delta_x \sigma \ \alpha \text{ dx dt} + 2\mu \iint\limits_{\Omega\times(0,T)} \frac{|z|^2}{|x|^3} \partial_r \sigma, \end{aligned} \tag{8.2.15}$$

where $\partial_n = \vec{n} \cdot \nabla$, $\vec{n}$ *being the normal outward vector on the boundary,* $\partial_r = \frac{x}{|x|} \cdot \nabla$ *and* ds *denotes the trace of the Lebesgue measure on* $\partial\Omega$.

For the proof, see Subsection 8.2.3.

Now, we will decompose the term I in (8.2.15) into several terms that we handle separately.

Let us define I_l as the sum of the integrals linear in σ which do not have any time derivative:

$$\begin{aligned} I_l = -2 \iint\limits_{\Omega\times(0,T)} & D^2\sigma(\nabla z, \nabla z) \text{ dx dt} + \iint\limits_{\partial\Omega\times(0,T)} |\partial_n z|^2 \ \partial_n \sigma \text{ds dt} \\ & - \iint\limits_{\Omega\times(0,T)} |\nabla z|^2 \Delta_x \sigma \ \alpha \text{ dx dt} + \frac{1}{2} \iint\limits_{\Omega\times(0,T)} |z|^2 \Delta_x^2 \sigma \Big(1+\alpha\Big) \text{ dx dt} \\ & + \iint\limits_{\Omega\times(0,T)} |z|^2 \nabla\alpha \cdot \nabla \Delta_x \sigma \text{ dx dt} + \frac{1}{2} \iint\limits_{\Omega\times(0,T)} |z|^2 \Delta_x \sigma \ \Delta_x \alpha \text{ dx dt} \\ & + \mu \iint\limits_{\Omega\times(0,T)} \frac{|z|^2}{|x|^2} \Delta_x \sigma \ \alpha \text{ dx dt} + 2\mu \iint\limits_{\Omega\times(0,T)} \frac{|z|^2}{|x|^3} \partial_r \sigma \text{ dx dt}. \end{aligned} \tag{8.2.16}$$

Then we have the following estimate:

Lemma 8.2.5. *There exist positive constants such that for λ large enough, we have:*

$$I_l \ge 2s \iint\limits_{\Omega\times(0,T)} \theta \frac{|z|^2}{|x|}\,\mathrm{dx\,dt} + sN \iint\limits_{\Omega\times(0,T)} \theta\alpha|\nabla z|^2\,\mathrm{dx\,dt} + C_1 s\lambda^2 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta\phi|\nabla z|^2\,\mathrm{dx\,dt} - C_2 s\lambda^2 \iint\limits_{\omega_0\times(0,T)} \theta\phi|\nabla z|^2\,\mathrm{dx\,dt} - C_3 s\lambda^4 \iint\limits_{\Omega\times(0,T)} \theta|z|^2\,\mathrm{dx\,dt} - C_4 s\lambda^4 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta\phi|z|^2\,\mathrm{dx\,dt}. \quad (8.2.17)$$

Again, the proof is given in Subsection 8.2.3. Note that the proof of Lemma 8.2.5 uses an improved form of the Hardy inequality (8.1.3), which can be found for instance in [18], namely:

Lemma 8.2.6. *There exists a positive constant $C_5 > 0$, such that*

$$\mu^*(N)\int_\Omega \frac{|z|^2}{|x|^2}\,\mathrm{dx} + \int_\Omega \frac{|z|^2}{|x|}\,\mathrm{dx} \le \int_\Omega |\nabla z|^2\,\mathrm{dx} + C_5\int_\Omega |z|^2\,\mathrm{dx}, \quad z \in H_0^1(\Omega). \quad (8.2.18)$$

Of course, this inequality also holds for $\mu < \mu^*(N)$.

We then consider the integrals involving non-linear terms in σ and without any time derivative, that is

$$I_{nl} = -2 \iint\limits_{\Omega\times(0,T)} |z|^2 D^2\sigma\Big(\nabla\sigma,\nabla\sigma\Big)\,\mathrm{dx\,dt} + \iint\limits_{\Omega\times(0,T)} \alpha|z|^2\Delta_x\sigma|\nabla\sigma|^2\,\mathrm{dx\,dt} - \frac{1}{2}\iint\limits_{\Omega\times(0,T)} \alpha^2|z|^2|\Delta_x\sigma|^2\,\mathrm{dx}. \quad (8.2.19)$$

Then, with σ as in (8.2.2), we obtain (see Subsection 8.2.3) that

Lemma 8.2.7. *There exist positive constants such that for λ large enough, for $s \ge s_0(\lambda)$,*

$$I_{nl} \ge C_6 s^3 \iint\limits_{\Omega\times(0,T)} \theta^3|x|^2|z|^2\,\mathrm{dx\,dt} + C_7 s^3\lambda^4 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta^3\phi^3|z|^2\,\mathrm{dx\,dt} - C_8 s^3\lambda^4 \iint\limits_{\omega_0\times(0,T)} \theta^3\phi^3|z|^2\,\mathrm{dx\,dt}. \quad (8.2.20)$$

We finally estimate the terms involving the time derivatives in σ:

$$I_t = -\frac{1}{2}\iint\limits_{\Omega\times(0,T)} |z|^2\Big(\partial_{tt}^2\sigma + 2\partial_t\Big(|\nabla\sigma|^2\Big)\Big)\,\mathrm{dx\,dt} + \iint\limits_{\Omega\times(0,T)} \alpha|z|^2\Delta_x\sigma\partial_t\sigma\,\mathrm{dx\,dt}. \quad (8.2.21)$$

We also add to I_t the integrals appearing in Lemma 8.2.5 that we want to get rid of and define

$$I_r = I_t - C_3 s\lambda^4 \iint\limits_{\Omega\times(0,T)} \theta|z|^2\,\mathrm{dx\,dt} - C_4 s\lambda^4 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta\phi|z|^2\,\mathrm{dx\,dt}. \quad (8.2.22)$$

Then we have to prove that I_r is negligible with respect to the positive terms in (8.2.17) and (8.2.20).

Lemma 8.2.8. *For any λ large enough, there exists $s_0(\lambda)$ such that for $s \geq s_0(\lambda)$,*

$$|I_r| \leq s \iint\limits_{\Omega\times(0,T)} \theta \frac{|z|^2}{|x|} \,\mathrm{dx}\,\mathrm{dt} + \frac{C_6}{2} s^3 \iint\limits_{\Omega\times(0,T)} \theta^3 |x|^2 |z|^2 \,\mathrm{dx}\,\mathrm{dt} + \frac{C_7}{2} s^3\lambda^4 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta^3\phi^3 |z|^2 \,\mathrm{dx}\,\mathrm{dt}, \tag{8.2.23}$$

where C_6 and C_7 are as in (8.2.20).

Using (8.2.14) and Lemmas 8.2.5, 8.2.7 and 8.2.8, whose proofs are postponed to Subsection 8.2.3, we obtain a Carleman estimate in the z variable. Undoing the change of variable (8.2.8) provides the Carleman estimate (8.2.7). □

8.2.2 From the Carleman estimate to the Observability inequality

In this Subsection, we explain why the Carleman estimate (8.2.7) implies the observability inequality (8.1.8).

We fix $\lambda > \lambda_0$ and $s > s_0(\lambda)$ such that (8.2.7) holds. These parameters now enter in the constant K:

$$\iint\limits_{\Omega\times(0,T)} \theta e^{-2\sigma} \frac{|w|^2}{|x|} \,\mathrm{dx}\,\mathrm{dt} \leq K \iint\limits_{\omega_0\times(0,T)} \theta\phi e^{-2\sigma} |\nabla w|^2 \,\mathrm{dx}\,\mathrm{dt} + K \iint\limits_{\omega_0\times(0,T)} \theta^3\phi^3 e^{-2\sigma} |w|^2 \,\mathrm{dx}\,\mathrm{dt}. \tag{8.2.24}$$

One easily checks the existence of a constant C such that

$$\begin{cases} \theta\ e^{-2\sigma} \dfrac{1}{|x|} & \geq & C, & (x,t) \in \Omega \times \left[\dfrac{T}{4}, \dfrac{3T}{4}\right], \\ \theta\ \phi\ e^{-2\sigma} & \leq & Ce^{-\sigma}, & (x,t) \in \omega_0 \times (0,T), \\ \theta^3\phi^3 e^{-2\sigma} & \leq & C, & (x,t) \in \omega_0 \times (0,T). \end{cases}$$

Thus, (8.2.24) implies

$$\iint_{\Omega\times(T/4,3T/4)} |w|^2 \,\mathrm{dx}\,\mathrm{dt} \leq K \iint\limits_{\omega_0\times(0,T)} e^{-\sigma} |\nabla w|^2 \,\mathrm{dx}\,\mathrm{dt} + K \iint\limits_{\omega_0\times(0,T)} |w|^2 \,\mathrm{dx}\,\mathrm{dt}. \tag{8.2.25}$$

Therefore to obtain inequality (8.1.8), it is sufficient to prove the following lemma:

Lemma 8.2.9 (Cacciopoli's inequality)**.** *Let $\bar{\sigma} : (0,T) \times \bar{\omega} \to \mathbb{R}_+^*$ be a smooth nonnegative function such that*

$$\bar{\sigma}(t,x) \to +\infty \text{ as } t \to 0^+ \text{ and as } t \to T^-.$$

There exists a constant C independent of $\mu \leq \mu^*(N)$ such that any solution w of (8.1.6) *satisfies*

$$\iint\limits_{\omega_0\times(0,T)} e^{-\bar{\sigma}} |\nabla w|^2 \,\mathrm{dx}\,\mathrm{dt} \leq C \iint\limits_{\omega\times(0,T)} |w|^2 \,\mathrm{dx}\,\mathrm{dt}. \tag{8.2.26}$$

The proof of this lemma is given for instance in [19, Lemma III.3]. This obviously implies (8.1.8) by taking $\bar{\sigma} = \sigma$ in Lemma 8.2.9, since σ satisfies (8.2.1). It follows that inequality (8.1.7) holds as well and, by the classical HUM duality ([16]), this proves Theorem 8.1.1.

8.2.3 Proofs of technical Lemmas

Here we present the proofs of the technical Lemmas stated in Subsection 8.2.1. This part can be skipped in a first lecture. In this subsection, all the constants are positive and independent of λ or s.

Proof of Lemma 8.2.4. To make the computations easier, we define

$$\begin{aligned} &S_1 z = \Delta_x z, \qquad S_2 z = \frac{\mu}{|x|^2} z, \qquad S_3 z = z\big(\partial_t \sigma + |\nabla \sigma|^2\big), \\ &A_1 z = \partial_t z, \qquad A_2 z = 2\nabla z \cdot \nabla \sigma, \qquad A_3 z = z\Delta_x \sigma\big(1+\alpha\big), \end{aligned} \tag{8.2.27}$$

and denotes by I_{ij} the scalar product $< S_i, A_j >$. We will compute each term using integration by parts and the boundary conditions (8.2.9) and (8.2.11).

Computation of I_{11}:

$$I_{11} = \iint\limits_{\Omega\times(0,T)} \Delta_x z \; \partial_t z \; \mathrm{dx}\; \mathrm{dt} = -\iint\limits_{\Omega\times(0,T)} \partial_t\Big(\frac{|\nabla z|^2}{2}\Big) \; \mathrm{dx}\; \mathrm{dt} = 0, \tag{8.2.28}$$

where the last identity is justified by (8.2.9).

Computation of I_{12}: Note that, since z vanishes on the boundary, its gradient ∇z on the boundary is normal to the boundary, and therefore $\nabla z = \partial_n z \; \vec{n}$, where $\vec{n}$ denotes the normal outward vector on the boundary.

$$\begin{aligned} I_{12} &= 2 \iint\limits_{\Omega\times(0,T)} \Delta_x z \; \nabla z \cdot \nabla \sigma \; \mathrm{dx}\; \mathrm{dt} \\ &= -2 \iint\limits_{\Omega\times(0,T)} \nabla z \cdot \nabla\Big(\nabla z \cdot \nabla \sigma\Big) \; \mathrm{dx}\; \mathrm{dt} + 2 \iint\limits_{\partial\Omega\times(0,T)} |\partial_n z|^2 \; \partial_n \sigma \mathrm{ds}\; \mathrm{dt}, \end{aligned}$$

Besides, one can check that

$$\nabla z \cdot \nabla\Big(\nabla z \cdot \nabla \sigma\Big) = \frac{1}{2}\nabla\Big(|\nabla z|^2\Big) \cdot \nabla \sigma + D^2\sigma(\nabla z, \nabla z).$$

It follows easily that

$$I_{12} = \iint\limits_{\Omega\times(0,T)} |\nabla z|^2 \Delta_x \sigma \; \mathrm{dx}\; \mathrm{dt} - 2 \iint\limits_{\Omega\times(0,T)} D^2\sigma(\nabla z, \nabla z) \; \mathrm{dx}\; \mathrm{dt} + \iint\limits_{\partial\Omega\times(0,T)} |\partial_n z|^2 \; \partial_n \sigma \mathrm{ds}\; \mathrm{dt}. \tag{8.2.29}$$

Computation of I_{13}:

$$I_{13} = \iint\limits_{\Omega\times(0,T)} \Delta_x z \; z\Delta_x \sigma\Big(1+\alpha\Big) \; \mathrm{dx}\; \mathrm{dt} = -\iint\limits_{\Omega\times(0,T)} \nabla z \cdot \nabla\Big(z\Delta_x \sigma\Big(1+\alpha\Big)\Big) \; \mathrm{dx}\; \mathrm{dt}.$$

Thus we obtain

$$\begin{aligned} I_{13} = -\iint\limits_{\Omega\times(0,T)} |\nabla z|^2 \Delta_x \sigma\Big(1+\alpha\Big) \; \mathrm{dx}\; \mathrm{dt} + \frac{1}{2}\iint\limits_{\Omega\times(0,T)} |z|^2 \Delta_x^2 \sigma\Big(1+\alpha\Big) \; \mathrm{dx}\; \mathrm{dt} \\ + \iint\limits_{\Omega\times(0,T)} |z|^2 \nabla\alpha \cdot \nabla \Delta_x \sigma \; \mathrm{dx}\; \mathrm{dt} + \frac{1}{2}\iint\limits_{\Omega\times(0,T)} |z|^2 \Delta_x \sigma \; \Delta_x \alpha \; \mathrm{dx}\; \mathrm{dt}. \end{aligned} \tag{8.2.30}$$

Computation of I_{21}: As in (8.2.28), using (8.2.9), one easily checks that

$$I_{21} = 0. \tag{8.2.31}$$

Computation of I_{22}:

$$\begin{aligned} I_{22} &= \mu \iint\limits_{\Omega\times(0,T)} \frac{1}{|x|^2}\nabla\Big(|z|^2\Big)\cdot\nabla\sigma \text{ dx dt} \\ &= -\mu \iint\limits_{\Omega\times(0,T)} \frac{|z|^2}{|x|^2}\Delta_x\sigma \text{ dx dt} + 2\mu \iint\limits_{\Omega\times(0,T)} \frac{|z|^2}{|x|^3}\partial_r\sigma \text{ dx dt}. \end{aligned} \tag{8.2.32}$$

Computation of I_{23}:

$$I_{23} = \mu \iint\limits_{\Omega\times(0,T)} \frac{|z|^2}{|x|^2}\Delta_x\sigma\Big(1+\alpha\Big) \text{ dx dt}. \tag{8.2.33}$$

Computation of I_{31}:

$$I_{31} = \frac{1}{2}\iint\limits_{\Omega\times(0,T)} \partial_t\Big(|z|^2\Big)\Big(\partial_t\sigma + |\nabla\sigma|^2\Big) \text{ dx dt} = -\frac{1}{2}\iint\limits_{\Omega\times(0,T)} |z|^2\partial_t\Big(\partial_t\sigma + |\nabla\sigma|^2\Big) \text{ dx dt}. \tag{8.2.34}$$

Computation of I_{32}:

$$I_{32} = \iint\limits_{\Omega\times(0,T)} \nabla\Big(|z|^2\Big)\cdot\nabla\sigma\Big(\partial_t\sigma + |\nabla\sigma|^2\Big) \text{ dx dt}.$$

It follows that

$$\begin{aligned} I_{32} = &-\iint\limits_{\Omega\times(0,T)} |z|^2\Delta_x\sigma\Big(\partial_t\sigma + |\nabla\sigma|^2\Big) \text{ dx dt} \\ &- \iint\limits_{\Omega\times(0,T)} |z|^2\nabla\sigma\cdot\nabla\Big(\partial_t\sigma\Big) - 2\iint\limits_{\Omega\times(0,T)} |z|^2 D^2\sigma\Big(\nabla\sigma,\nabla\sigma\Big) \text{ dx dt}. \end{aligned} \tag{8.2.35}$$

Computation of I_{33}:

$$I_{33} = \iint\limits_{\Omega\times(0,T)} |z|^2\Delta_x\sigma\Big(\partial_t\sigma + |\nabla\sigma|^2\Big)\Big(1+\alpha\Big) \text{ dx dt}. \tag{8.2.36}$$

Lemma 8.2.4 follows directly from these computations. □

Proof of Lemma 8.2.5. Since the integral I_l is linear in σ, we decompose σ as

$$\sigma = s\theta(t)e^{2\lambda\sup\psi} + \sigma_{x^2}(t,x) + \sigma_\phi(t,x),$$

with

$$\sigma_{x^2}(t,x) = -s\theta(t)\frac{|x|^2}{2}, \quad \sigma_\phi(t,x) = -s\theta(t)\phi(x).$$

Note that the term $s\theta \exp(2\lambda \sup \psi)$ in σ does not appear in the computations of I_l, since it is constant in the space variable, and each integral in (8.2.16) involves space derivatives.

We then define I_{l,x^2} and $I_{l,\phi}$ as the terms in I_l corresponding respectively to σ_{x^2} and σ_ϕ.

First, we compute I_{l,x^2}. In this case, all the computations are explicit:

$$\begin{aligned} I_{l,x^2} = 2s \iint_{\Omega\times(0,T)} \theta|\nabla z|^2 \,\mathrm{dx}\,\mathrm{dt} - s \iint_{\partial\Omega\times(0,T)} \theta|\partial_n z|^2 \vec{x}\cdot\vec{n}\mathrm{ds}\,\mathrm{dt} \\ + sN \iint_{\Omega\times(0,T)} \theta\alpha|\nabla z|^2 \,\mathrm{dx}\,\mathrm{dt} - s\frac{N}{2} \iint_{\Omega\times(0,T)} \theta|z|^2\Delta_x\alpha \,\mathrm{dx}\,\mathrm{dt} \\ - s\mu N \iint_{\Omega\times(0,T)} \theta\alpha\frac{|z|^2}{|x|^2} \,\mathrm{dx}\,\mathrm{dt} - 2s\mu \iint_{\Omega\times(0,T)} \theta\frac{|z|^2}{|x|^2} \,\mathrm{dx}\,\mathrm{dt}. \end{aligned}$$

Thus, from the Hardy improved inequality (8.2.18), since θ only depends on the time variable t and since α vanishes on $B(0,1/2)$ by (8.2.12), there exists a constant such that

$$\begin{aligned} I_{l,x^2} \geq 2s \iint_{\Omega\times(0,T)} \theta\frac{|z|^2}{|x|} \,\mathrm{dx}\,\mathrm{dt} + sN \iint_{\Omega\times(0,T)} \theta\alpha|\nabla z|^2 \,\mathrm{dx}\,\mathrm{dt} \\ - s \iint_{\partial\Omega\times(0,T)} \theta|\partial_n z|^2\vec{x}\cdot\vec{n}\mathrm{ds}\,\mathrm{dt} - Cs \iint_{\Omega\times(0,T)} \theta|z|^2 \,\mathrm{dx}\,\mathrm{dt}. \end{aligned} \tag{8.2.37}$$

Second, let us consider $I_{l,\phi}$. To simplify, we decompose this integral into the integrals $I_{l,\phi,1}$ in $B(0,1)$ and $I_{l,\phi,2}$ outside $B(0,1)$.

In the unit ball, $\phi(x) = |x|^\lambda$ and then, all the computations are explicit. Especially, ϕ is convex (at least for $\lambda > 1$, which can be assumed since λ is aimed at being large), and therefore $D^2\phi(\xi,\xi)$ is a positive quadratic form in ξ, and $\Delta_x\phi > 0$. Besides, all the terms

$$\Delta_x^2\phi, \ \nabla\Delta_x\phi, \ \Delta_x\phi, \ \frac{\Delta_x\phi}{|x|^2}, \ \frac{\partial_r\phi}{|x|^3}$$

are bounded by $C\lambda^4|x|^{\lambda-4}$ for λ large enough (namely $\lambda > 4$). Then

$$I_{l,\phi,1} \geq -Cs\lambda^4 \iint_{B(0,1)\times(0,T)} \theta(t)|x|^{\lambda-4}|z|^2 \,\mathrm{dx}\,\mathrm{dt}. \tag{8.2.38}$$

Outside the unit ball, the computations are more intricate. First, let us compute the first derivative of ϕ:

$$\begin{aligned} &\nabla\phi = \lambda\phi\nabla\psi, \qquad \partial^2_{i,j}\phi = \lambda\phi\partial^2_{i,j}\psi + \lambda^2\phi \ \partial_i\psi \ \partial_j\psi, \\ &\Delta_x\phi = \lambda\phi\Delta_x\psi + \lambda^2\phi|\nabla\psi|^2. \end{aligned} \tag{8.2.39}$$

Besides, due to the particular choice of ψ, and especially (8.2.5), one can get the following estimates :

$$\begin{aligned} 2D^2\phi(\xi,\xi)+\alpha\Delta_x\phi|\xi|^2 &\geq C\lambda^2\phi|\xi|^2, \qquad \xi\in\mathbb{R}^N,\ x\in\mathcal{O},\\ \Big|2D^2\phi(\xi,\xi)+\alpha\Delta_x\phi|\xi|^2\Big| &\leq C\lambda^2\phi|\xi|^2, \qquad \xi\in\mathbb{R}^N,\ x\in\omega_0,\\ |\Delta_x^2\phi|+|\Delta_x\phi|+|\nabla\phi|+|\partial_r\phi|+|\nabla\Delta_x\phi| &\leq C\phi\lambda^4, \qquad x\in\tilde{\mathcal{O}}, \end{aligned}$$

for λ large enough. Hence we deduce that

$$\begin{aligned} I_{l,\phi,2}\geq Cs\lambda^2\iint\limits_{\mathcal{O}\times(0,T)}\theta\phi|\nabla z|^2\,\mathrm{dx\,dt}-s\lambda\iint\limits_{\partial\Omega\times(0,T)}\theta\phi|\partial_n z|^2\nabla\psi\cdot\vec{n}\mathrm{ds\,dt}\\ -Cs\lambda^4\iint\limits_{\tilde{\mathcal{O}}\times(0,T)}\theta\phi|z|^2\,\mathrm{dx\,dt}-s\lambda^2\iint\limits_{\omega_0\times(0,T)}\theta\phi|\nabla z|^2\,\mathrm{dx\,dt}. \end{aligned} \tag{8.2.40}$$

Taking λ large enough, due to the properties (8.2.4) and (8.2.5), the sum of boundary terms in (8.2.37) and in (8.2.40) is positive. Indeed, from (8.2.4) and (8.2.5), $\nabla\psi\cdot\vec{n}=-|\nabla\psi|\leq-\delta$, and thus the choice $\lambda\geq\mathrm{diam}(\Omega)/\delta$, where $\mathrm{diam}(\Omega)$ is the diameter of Ω, is convenient.

Hence, combining (8.2.37), (8.2.38) and (8.2.40) gives Lemma 8.2.5. □

Proof of Lemma 8.2.7. Again, we handle separately the integrals I_{nl1} in the unit ball and I_{nl2} outside the unit ball. This is needed since the terms $|x|^2$ and ϕ of σ (see (8.2.2)) do not have the same order inside and outside the unit ball.

Notice that, in the unit ball,

$$\begin{cases} \nabla\sigma=-s\theta x\Big(1+\lambda|x|^{\lambda-2}\Big),\\ \Delta_x\sigma=-s\theta\Big(N+\lambda(N+\lambda-2)|x|^{\lambda-2}\Big). \end{cases} \tag{8.2.41}$$

Hence we compute explicitly the terms appearing in the integrals for a radial vector ξ of $\mathbb{R}^N$, which is the case of $\nabla\sigma$ in the unit ball:

$$\alpha\Delta_x\sigma|\xi|^2-2D^2\sigma(\xi,\xi)=s\theta\Big((2-\alpha N)|\xi|^2+2\lambda|x|^{\lambda-2}|\xi|^2+\lambda|x|^{\lambda-4}|\xi|^2\big((2-\alpha)\lambda-4-\alpha(N+2)\big)\Big).$$

Thus we can take λ large enough such that

$$\begin{aligned} -2\iint\limits_{B(0,1)\times(0,T)}|z|^2D^2\sigma\Big(\nabla\sigma,\nabla\sigma\Big)\,\mathrm{dx\,dt}+\iint\limits_{B(0,1)\times(0,T)}\alpha|z|^2\Delta_x\sigma|\nabla\sigma|^2\,\mathrm{dx\,dt}\\ \geq Cs\iint\limits_{B(0,1)\times(0,T)}\theta|z|^2|\nabla\sigma|^2\,\mathrm{dx\,dt}\geq s^3\iint\limits_{B(0,1)\times(0,T)}\theta^3|x|^2|z|^2\,\mathrm{dx\,dt}. \end{aligned} \tag{8.2.42}$$

The last term in (8.2.19) can be absorbed, since from (8.2.41), we have

$$\big|\Delta_x\sigma\big|^2\leq Cs^2\theta^2\lambda^4.$$

Indeed, combined with the assumption (8.2.12) on the support of α, the last integral in (8.2.19) satisfies

$$\iint\limits_{B(0,1)\times(0,T)} \alpha^2|z|^2|\Delta_x\sigma|^2 \,\mathrm{dx}\,\mathrm{dt} \leq Cs^2\lambda^4 \iint\limits_{B(0,1)\times(0,T)} \theta^2|x|^2|z|^2 \,\mathrm{dx}\,\mathrm{dt}.$$

Then taking s large, for instance $s > C\lambda^4$, we can absorb the third term in (8.2.19), and we obtain that

$$I_{nl1} \geq Cs^3 \iint\limits_{B(0,1)\times(0,T)} \theta^3|x|^2|z|^2 \,\mathrm{dx}\,\mathrm{dt}. \tag{8.2.43}$$

Outside the unit ball, due to the particular choice of ψ, and especially (8.2.5), and since $\|\alpha\|_{L^\infty(\Omega)} < 2$, as in [13] we remark that, for s and λ large enough,

$$\begin{aligned} \alpha\Delta_x\sigma|\nabla\sigma|^2 - 2D^2\sigma(\nabla\sigma,\nabla\sigma) &\geq Cs^3\lambda^4\theta^3\phi^3, \qquad x \in \mathcal{O}, \\ \left|\alpha\Delta_x\sigma|\nabla\sigma|^2 - 2D^2\sigma(\nabla\sigma,\nabla\sigma)\right| &\leq Cs^3\lambda^4\theta^3\phi^3, \qquad x \in \omega_0, \end{aligned}$$

and

$$|\Delta_x\sigma|^2 \leq Cs^2\lambda^4\theta^2\phi^2, \qquad x \in \tilde{\mathcal{O}}.$$

Then, taking s large yields

$$I_{nl2} \geq Cs^3\lambda^4 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta^3\phi^3|z|^2 \,\mathrm{dx}\,\mathrm{dt} - Cs^3\lambda^4 \iint\limits_{\omega_0\times(0,T)} \theta^3\phi^3|z|^2 \,\mathrm{dx}\,\mathrm{dt}. \tag{8.2.44}$$

Hence the proof of Lemma 8.2.7 is completed. □

Proof of Lemma 8.2.8. First notice that

$$\left|\theta\theta'\right| \leq C\theta^3, \qquad \left|\theta'\right| \leq C\theta^3, \qquad \left|\theta''\right| \leq C\theta^{5/3}.$$

Then, since α vanishes in $B(0,1/2)$, bounding the integral in $B(0,1)$ and $\tilde{\mathcal{O}}$ using respectively (8.2.39) and (8.2.41),

$$\begin{aligned} \left| \iint\limits_{\Omega\times(0,T)} \alpha|z|^2\Delta_x\sigma\partial_t\sigma \,\mathrm{dx}\,\mathrm{dt}\right| &\leq Cs^2\lambda^2 e^{2\lambda\sup\psi} \iint\limits_{B(0,1)\times(0,T)} \theta^3|x|^2|z|^2 \,\mathrm{dx}\,\mathrm{dt} \\ &\quad + Cs^2\lambda^2 e^{2\lambda\sup\psi} \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta^3\phi|z|^2 \,\mathrm{dx}\,\mathrm{dt}. \end{aligned}$$

Similarly,

$$\begin{aligned} \left| \iint\limits_{\Omega\times(0,T)} |z|^2\partial_t\Big(|\nabla\sigma|^2\Big) \,\mathrm{dx}\,\mathrm{dt}\right| &\leq Cs^2\lambda^2 \iint\limits_{B(0,1)\times(0,T)} \theta^3|x|^2|z|^2 \,\mathrm{dx}\,\mathrm{dt} \\ &\quad + Cs^2\lambda^2 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta^3\phi^2|z|^2 \,\mathrm{dx}\,\mathrm{dt}. \end{aligned} \tag{8.2.45}$$

The remaining term

$$R = -\frac{1}{2}\iint\limits_{\Omega\times(0,T)} |z|^2\partial_{tt}^2\sigma \text{ dx dt} - C_3 s\lambda^4 \iint\limits_{\Omega\times(0,T)} \theta|z|^2 \text{ dx dt} - C_4 s\lambda^4 \iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta\phi|z|^2 \text{ dx dt}$$

satisfies for λ large enough

$$\Big|R\Big| \leq Cse^{2\lambda \sup \psi} \iint\limits_{\Omega\times(0,T)} \theta^{5/3}|z|^2 \text{ dx dt}. \tag{8.2.46}$$

Let us then estimate this last integral. Take β a positive number that we will choose later on. Then

$$\begin{aligned}\iint\limits_{\Omega\times(0,T)} \theta^{5/3}|z|^2 \text{ dx dt} &= \iint\limits_{\Omega\times(0,T)} \Big(\beta\theta|x|^{2/3}|z|^{2/3}\Big)\Big(\frac{1}{\beta}\theta^{2/3}|x|^{-2/3}|z|^{4/3}\Big) \text{ dx dt} \\ &\leq \frac{\beta^3}{3}\iint\limits_{\Omega\times(0,T)} \theta^3|x|^2|z|^2 \text{ dx dt} + \frac{2}{3\beta^{3/2}}\iint\limits_{\Omega\times(0,T)} \theta\frac{|z|^2}{|x|} \text{ dx dt},\end{aligned}$$

where we used the classical convexity inequality

$$ab \leq \frac{1}{3}a^3 + \frac{2}{3}b^{3/2}.$$

Then we get three constants such that

$$\begin{aligned}|I_r| \leq{}& c_1\Big(s^2\lambda^2 + s^2\lambda^2 e^{2\lambda\sup\psi} + se^{2\lambda\sup\psi}\beta^3\Big)\iint\limits_{\Omega\times(0,T)} \theta^3|x|^2|z|^2 \text{ dx dt} \\ &+ c_2\Big(s^2\lambda^2 e^{2\lambda\sup\psi} + s^2\lambda^2\Big)\iint\limits_{\tilde{\mathcal{O}}\times(0,T)} \theta^3\phi^3|z|^2 \text{ dx dt} + c_3 se^{2\lambda\sup\psi}\frac{1}{\beta^{3/2}}\iint\limits_{\Omega\times(0,T)} \theta\frac{|z|^2}{|x|} \text{ dx dt}.\end{aligned} \tag{8.2.47}$$

Thus, for a given $\lambda > 0$, choosing β such that

$$c_3 e^{2\lambda\sup\psi} = \beta^{3/2},$$

there exists $s_0(\lambda)$ such that for any $s \geq s_0(\lambda)$, inequality (8.2.23) holds. □

8.3 Non uniform stabilization in the case $\mu > \mu^*(N)$

The goal of this section is to prove Theorem 8.1.2. The proof is divided into two main steps.

First, we prove some basic estimates on the spectrum of the operator

$$L^\varepsilon = -\Delta_x - \frac{\mu}{|x|^2+\varepsilon^2} \tag{8.3.1}$$

on Ω with Dirichlet boundary conditions, especially on the first eigenvalue λ_0^ε and the corresponding eigenfunction ϕ_0^ε. This will be done in Subsection 8.3.1.

Second, we deduce Theorem 8.1.2 in Subsection 8.3.2 by giving a lower bound on the quantity $J^\varepsilon_{\phi_0^\varepsilon}$ that goes to infinity when $\varepsilon \to 0$.

8.3.1 Spectral estimates

Since for $\varepsilon > 0$, the function $1/(|x|^2 + \varepsilon^2)$ is smooth and bounded in Ω, the spectrum of L^ε is formed by a sequence of real eigenvalues $\lambda_0^\varepsilon \leq \lambda_1^\varepsilon \leq \cdots \leq \lambda_n^\varepsilon \leq \cdots$, with $\lambda_n^\varepsilon \to +\infty$. The corresponding eigenvectors ϕ_n^ε are a basis of $L^2(\Omega)$, orthonormal with respect to the L^2 scalar product. We choose ϕ_n^ε of unit L^2-norm.

In the sequel, we focus on the bottom of the spectrum -*the most explosive mode.*

Proposition 8.3.1. *Assume that $\mu > \mu^*(N)$. Then we have that*

$$\lim_{\varepsilon \to 0} \lambda_0^\varepsilon = -\infty. \tag{8.3.2}$$

and for all $\alpha > 0$,

$$\lim_{\varepsilon \to 0} \|\phi_0^\varepsilon\|_{H^1(\Omega \setminus \bar{B}(0,\alpha))} = 0. \tag{8.3.3}$$

Proof. We argue by contradiction, and assume that λ_0^ε is bounded from below for a subsequence by a real number C. Then, from the Rayleigh formula we get

$$\forall \varepsilon > 0, \forall u \in H_0^1(\Omega), \quad \mu \int_\Omega \frac{|u|^2}{|x|^2 + \varepsilon^2} \, \mathrm{dx} \leq \int_\Omega |\nabla u|^2 \, \mathrm{dx} - C \int_\Omega |u|^2 \, \mathrm{dx}.$$

Taking $u \in \mathcal{D}(\Omega)$, we pass to the limit $\varepsilon \to 0$ and get

$$\mu \int_\Omega \frac{|u|^2}{|x|^2} \, \mathrm{dx} \leq \int_\Omega |\nabla u|^2 \, \mathrm{dx} - C \int_\Omega |u|^2 \, \mathrm{dx}, \tag{8.3.4}$$

that must therefore hold for any $u \in H_0^1(\Omega)$ by a density argument.

Now, there exists $\alpha_0 > 0$ such that $B(0, \alpha_0) \subset \Omega$. We then choose $u \in H_0^1(B(0, \alpha_0))$ that we extend by 0 on $\mathbb{R}^N$, and define for $a \geq 1$

$$u_a(r) = a^N \, u(ar).$$

These functions are in $H_0^1(B(0, \alpha_0))$, and therefore in $H_0^1(\Omega)$, and we can apply (8.3.4) to them:

$$a^2 \Big(\mu \int_\Omega \frac{|u|^2}{|x|^2} \, \mathrm{dx} - \int_\Omega |\nabla u|^2 \, \mathrm{dx} \Big) \leq -C \int_\Omega |u|^2 \, \mathrm{dx}.$$

Passing to the limit $a \to \infty$, we obtain that

$$\forall u \in H_0^1(B(0, \alpha_0)), \qquad \mu \int_\Omega \frac{|u|^2}{|x|^2} \, \mathrm{dx} \leq \int_\Omega |\nabla u|^2 \, \mathrm{dx}.$$

Therefore we should have that $\mu \leq \mu^*(N)$, since this is the Hardy inequality (8.1.3) in the set $B(0, \alpha_0)$, and then we have a contradiction.

Now, consider the first eigenvector $\phi_0^\varepsilon \in H_0^1(\Omega)$ of L^ε:

$$-\Delta_x \phi_0^\varepsilon - \frac{\mu}{|x|^2 + \varepsilon^2} \phi_0^\varepsilon = \lambda_0^\varepsilon \phi_0^\varepsilon, \quad \text{in } \Omega. \tag{8.3.5}$$

Remark that since the potential is smooth in Ω, the function ϕ_0^ε is smooth by classical elliptic estimates.

Set $\alpha > 0$. Let η_α be a nonnegative smooth function that vanishes in $B(0,\alpha/2)$ and equals 1 in $\mathbb{R}^N \backslash B(0,\alpha)$ with $\|\eta_\alpha\|_\infty \leq 1$. Multiplying (8.3.5) by $\eta_\alpha \phi_0^\varepsilon$, we get:

$$\int_\Omega \eta_\alpha |\nabla \phi_0^\varepsilon|^2 \, \mathrm{dx} + |\lambda_0^\varepsilon| \int_\Omega \eta_\alpha |\phi_0^\varepsilon|^2 = \mu \int_\Omega \eta_\alpha \frac{|\phi_0^\varepsilon|^2}{|x|^2+\varepsilon^2} \, \mathrm{dx} + \frac{1}{2} \int_\Omega \Delta \eta_\alpha |\phi_0^\varepsilon|^2 \, \mathrm{dx}. \tag{8.3.6}$$

Therefore, since ϕ_0^ε is of unit L^2-norm, due to the particular choice of η_α, we get

$$|\lambda_0^\varepsilon| \int_{\Omega \backslash B(0,\alpha)} |\phi_0^\varepsilon|^2 \, \mathrm{dx} \leq \frac{4\mu}{\alpha^2} + \frac{1}{2} \|\Delta_x \eta_\alpha\|_{L^\infty(\Omega)}.$$

Since $|\lambda_0^\varepsilon| \to \infty$ when $\varepsilon \to 0$, we get that for any $\alpha > 0$,

$$\lim_{\varepsilon \to 0} \int_{\Omega \backslash B(0,\alpha)} |\phi_0^\varepsilon|^2 \, \mathrm{dx} = 0. \tag{8.3.7}$$

Besides, still using (8.3.6) and the particular form of η_α

$$\int_{\Omega \backslash B(0,\alpha)} |\nabla \phi_0^\varepsilon|^2 \, \mathrm{dx} \leq \Big(\frac{4\mu}{\alpha^2} + \frac{1}{2} \|\Delta_x \eta_\alpha\|_{L^\infty(\Omega)} \Big) \int_{\Omega \backslash B(0,\alpha/2)} |\phi_0^\varepsilon|^2 \, \mathrm{dx}.$$

Therefore the proof of (8.3.3) is completed by using (8.3.7) for $\alpha/2$ instead of α. □

8.3.2 Proof of Theorem 8.1.2

Fix $\varepsilon > 0$, and choose $u_0^\varepsilon = \phi_0^\varepsilon$, which is of unit L^2-norm. Our goal is to prove that

$$\inf_{\substack{f \in L^2((0,T);H^{-1}(\Omega)) \\ f \text{ as in (8.1.2)}}} J_{u_0^\varepsilon}^\varepsilon(f) \underset{\varepsilon \to 0}{\longrightarrow} \infty. \tag{8.3.8}$$

Let $f \in L^2((0,T);H^{-1}(\Omega))$ as in (8.1.2), and consider u the corresponding solution of (8.1.13) with initial data $u_0^\epsilon = \phi_0^\epsilon$.

Set

$$a(t) = \int_\Omega u(t,x) \phi_0^\varepsilon(x) \, \mathrm{dx}, \qquad b(t) = < f(t), \phi_0^\varepsilon >_{H^{-1}(\Omega) \times H_0^1(\Omega)}.$$

Then $a(t)$ satisfies the equation

$$a'(t) + \lambda_0^\varepsilon a(t) = b(t), \qquad a(0) = 1.$$

Duhamel's formula gives

$$a(t) = \exp(-\lambda_0^\varepsilon t) + \int_0^t \exp(-\lambda_0^\varepsilon (t-s)) \, b(s) \mathrm{ds}.$$

Therefore

$$\begin{aligned} \iint_{\Omega \times (0,T)} |u(t,x)|^2 \, \mathrm{dx} \, \mathrm{dt} &\geq \int_0^T a(t)^2 \, \mathrm{dt} \\ &\geq \frac{1}{2} \int_0^T \exp(-2\lambda_0^\varepsilon t) \, \mathrm{dt} - \int_0^T \Big(\int_0^t \exp(-\lambda_0^\varepsilon (t-s)) b(s) \mathrm{ds} \Big)^2 \, \mathrm{dt}. \end{aligned} \tag{8.3.9}$$

Of course,

$$\frac{1}{2}\int_0^T \exp(-2\lambda_0^\varepsilon t)\ \mathrm{dt} = \frac{1}{4|\lambda_0^\varepsilon|}\Big(\exp(2|\lambda_0^\varepsilon|T)-1\Big).$$

The other term satisfies

$$\begin{aligned}\int_0^T \Big(\int_0^t \exp(-\lambda_0^\varepsilon(t-s))b(s)\mathrm{ds}\Big)^2\ \mathrm{dt} &\leq \int_0^T \Big(\int_0^t \exp(-2\lambda_0^\varepsilon(t-s))\mathrm{ds}\Big)\Big(\int_0^t |b(s)|^2\mathrm{ds}\Big)\ \mathrm{dt}\\ &\leq \int_0^T \frac{1}{2|\lambda_0^\varepsilon|}\exp(2|\lambda_0^\varepsilon|t)\Big(\int_0^t |b(s)|^2\mathrm{ds}\Big)\ \mathrm{dt}\\ &\leq \frac{1}{4|\lambda_0^\varepsilon|^2}\exp(2|\lambda_0^\varepsilon|T)\int_0^T |b(s)|^2\mathrm{ds}.\end{aligned}$$

Besides, from the definition of b and the assumption (8.1.2), we get that

$$|b(t)|^2 \leq \|f(t)\|^2_{H^{-1}(\Omega)}\,\|\phi_0^\varepsilon\|^2_{H^1(\omega)}.$$

Hence we deduce from (8.3.9) that

$$\frac{1}{4|\lambda_0^\varepsilon|}\Big(e^{2|\lambda_0^\varepsilon|T}-1\Big) \leq \iint\limits_{\Omega\times(0,T)} |u(t,x)|^2\ \mathrm{dx}\ \mathrm{dt} + \frac{\|\phi_0^\varepsilon\|^2_{H^1(\omega)}}{4|\lambda_0^\varepsilon|^2}e^{2|\lambda_0^\varepsilon|T}\int_0^T \|f(t)\|^2_{H^{-1}(\Omega)}\ \mathrm{dt}.$$

Therefore, either

$$\frac{1}{8|\lambda_0^\varepsilon|}\Big(e^{2|\lambda_0^\varepsilon|T}-1\Big) \leq \iint\limits_{\Omega\times(0,T)} |u(t,x)|^2\ \mathrm{dx}\ \mathrm{dt}$$

or

$$\frac{1}{8|\lambda_0^\varepsilon|}\Big(e^{2|\lambda_0^\varepsilon|T}-1\Big) \leq \frac{\|\phi_0^\varepsilon\|^2_{H^1(\omega)}}{4|\lambda_0^\varepsilon|^2}e^{2|\lambda_0^\varepsilon|T}\int_0^T \|f(t)\|^2_{H^{-1}(\Omega)}\ \mathrm{dt},$$

and in any case, for any f as in (8.1.2), we get

$$J^\varepsilon_{u_0^\varepsilon}(f) \geq \inf\left\{\frac{e^{2|\lambda_0^\varepsilon|T}-1}{16|\lambda_0^\varepsilon|},\ \frac{|\lambda_0^\varepsilon|}{4\,\|\phi_0^\varepsilon\|^2_{H^1(\omega)}}\Big(1-e^{-2|\lambda_0^\varepsilon|T}\Big)\right\}.$$

This bound blows up when $\varepsilon\to 0$ from the estimates (8.3.2). Indeed, since $0\notin\bar{\omega}$, we can choose $\alpha>0$ small enough such that $\omega\subset\Omega\backslash B(0,\alpha)$ and therefore

$$\|\phi_0^\varepsilon\|_{H^1(\omega)} \leq \|\phi_0^\varepsilon\|_{H^1(\Omega\backslash B(0,\alpha))} \underset{\varepsilon\to 0}{\longrightarrow} 0. \qquad \square$$

8.4 Comments

In this article we proposed a study of a parabolic equation with an inverse-square potential $-\mu/|x|^2$ from a control point of view, in the two cases $\mu\leq\mu^*(N)$, which corresponds to a subcritical case, and $\mu>\mu^*(N)$, the surcritical case.

A. When $\mu\leq\mu^*(N)$, we have addressed the null-controllability problem for a distributed control in an arbitrary open subset of Ω. To this end, we have derived a new Carleman inequality (8.2.7) inspired by the articles [19] and [13].

1. Our arguments can be adapted in much more general settings than presented here. For instance, one can handle several inverse-square singularities:

$$\begin{cases} \partial_t u - \Delta_x u - \sum_i \dfrac{\mu_i}{|x-x_i|^2} u = f, & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \end{cases} \tag{8.4.1}$$

where $\mu_i \leq \mu^*(N)$ for each i and f is localized in some open subset $\omega \subset \Omega$ in the sense of (8.1.2). In this case, the difficulty will again come from the choice of the weight. Let us assume that the points x_i satisfy the following properties

$$|x_i - x_j| \geq 3, \quad i \neq j, \qquad d(x_i, \partial\Omega) \geq 3.$$

Note that by a scaling argument, this can be assumed as soon as the set $\{x_i\}_i$ does not have any accumulation point in $\bar{\Omega}$, which is equivalent to say that they are in finite numbers since Ω is bounded. In this case, we propose a weight of the form

$$\sigma(t,x) = s\theta\Big(e^{2\lambda \sup \psi} - \frac{1}{2}\sum_i |x-x_i|^2 \gamma(x-x_i) - e^{\lambda\psi(x)}\Big),$$

where λ and s are positive parameters, θ is as in (8.2.3), ψ satisfies

$$\begin{cases} \psi(x) = \ln(|x-x_i|), & x \in B(x_i,1), \\ \psi(x) = 0, & x \in \partial\Omega, \\ \psi(x) > 0, & x \in \Omega \backslash \Big(\cup_i \bar{B}(x_i,1)\Big), \end{cases}$$

and (8.2.5), and $\gamma = \gamma(|x|)$ is a smooth cut-off function such that

$$\gamma(x) = 1, \quad |x| \leq 1, \qquad \gamma(x) = 0, \quad |x| \geq 3/2.$$

Using this weight and following the proof of Theorem 8.2.1, one can prove a Carleman estimate for the adjoint system of (8.4.1), which still directly implies (8.1.8). However it may occur that the system (8.4.1) is not dissipative (see [8] where a necessary and sufficient condition is given for a multipolar potential to be positive on $\mathbb{R}^n$), and therefore we need to explain why inequality (8.1.7) is still implied by (8.1.8). Following for instance [6, Lemma 2.1], one can prove that

$$F(t) = \int_\Omega |w(t,x)|^2 \, \mathrm{dx}$$

satisfies

$$F'(t) \geq -CF(t).$$

Thus a Gronwall inequality allows us to conclude (8.1.7) from (8.1.8).

2. Note also the dispersive properties (that is Strichartz estimates) of the operators $i\partial_t + P$ and $\partial_{tt}^2 + P$, with

$$P = -\Delta_x - \frac{\mu}{|x|^2},$$

were studied in the whole space $\mathbb{R}^N$, $N \geq 3$, in [3]. In [3], it is proved that Strichartz estimates hold for the Schrödinger and the wave equations provided $\mu < \mu^*(N)$. This result was generalized to the critical case $\mu = \mu^*(N)$ and to the multipolar case in [6]. To complete this picture, we mention [7], in which a positive potential V of order

$$\frac{\log(|x|)^2}{|x|^2}$$

was constructed in such a way that there exist quasi-modes for $P = -\Delta_x + V$ localized around the singularity. Note that in this case, the operator P is strongly elliptic since V is positive. To our knowledge, the controllability properties for the wave or Schrödinger equations with an inverse-square potential are widely open. Especially, it would be interesting to understand precisely the behavior of the rays of Geometric Optics around the singularities.

B. When $\mu > \mu^*(N)$, we have shown that we cannot uniformly stabilize regularized approximations of (8.1.1) with a control supported in ω when $0 \notin \bar{\omega}$.

1. To complete this result, we comment the case $0 \in \omega$, for which the stabilization property (8.1.10) holds. Given $u_0 \in L^2(\Omega)$, we claim that we can find $u \in L^2((0,T); H_0^1(\Omega))$ and $f \in L^2((0,T); H^{-1}(\Omega))$ as in (8.1.2) such that u is the solution of (8.1.1) and that $J_{u_0}(u,f) \leq C \|u_0\|^2_{L^2(\Omega)}$ (see (8.1.10)).

Indeed, denote by χ a smooth function that equals 1 in a neighborhood of 0 and vanishing outside ω. Then consider the solution u of

$$\begin{cases} \partial_t u - \Delta_x u - (1-\chi)\dfrac{\mu}{|x|^2} u = 0, & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) = u_0(x), & x \in \Omega. \end{cases}$$

which satisfies $u \in L^2((0,T); H_0^1(\Omega))$, and $\|u\|_{L^2(0,T;H_0^1(\Omega))} \leq C \|u_0\|_{L^2}$ for some constant C. Then taking $f = \mu\, \chi\, u/|x|^2 \in L^2((0,T); H^{-1}(\Omega))$ provides an admissible stabilizer with the required property (8.1.2).

The same argument can also be applied to derive the null-controllability property for (8.1.1) when $0 \in \omega$. Indeed, the results in [13] proves that there exists a control $v \in L^2((0,T) \times \omega)$ such that the solution of

$$\begin{cases} \partial_t u - \Delta_x u - (1-\chi)\dfrac{\mu}{|x|^2} u = v, & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) = u_0(x), & x \in \Omega. \end{cases}$$

satisfies $u(T) = 0$. Besides, the norms of v in $L^2((0,T) \times \omega)$ and u in $L^2((0,T); H_0^1(\Omega))$ are bounded by the norm of u_0 in $L^2(\Omega)$. Then, taking $f = v + \mu\, \chi\, u/|x|^2$ provides a control in $L^2((0,T); H^{-1}(\Omega))$ for (8.1.1) that drives the solution to 0 in time T.

2. Since we proved that we cannot uniformly stabilize (8.1.13) when $0 \notin \bar{\omega}$, there is no uniform observability properties such as (8.1.7) for the corresponding adjoint regularized systems.

Bibliography

[1] P. Baras and J. A. Goldstein. The heat equation with a singular potential. *Trans. Amer. Math. Soc.*, 284(1):121–139, 1984.

[2] A. Benabdallah, Y. Dermenjian, and J. Le Rousseau. Carleman estimates for the one-dimensional heat equation with a discontinuous coefficient and applications to controllability and an inverse problem. *J. Math. Anal. Appl.*, 336(2):865–887, 2007.

[3] N. Burq, F. Planchon, J. G. Stalker, and A. S. Tahvildar-Zadeh. Strichartz estimates for the wave and Schrödinger equations with the inverse-square potential. *J. Funct. Anal.*, 203(2):519–549, 2003.

[4] X. Cabré and Y. Martel. Existence versus explosion instantanée pour des équations de la chaleur linéaires avec potentiel singulier. *C. R. Acad. Sci. Paris Sér. I Math.*, 329(11):973–978, 1999.

[5] P. Cannarsa, P. Martinez, and J. Vancostenoble. Carleman estimates for a class of degenerate parabolic operators. *SIAM J. Control Optim.*, 47(1):1–19, 2008.

[6] T. Duyckaerts. Inégalités de résolvante pour l'opérateur de Schrödinger avec potentiel multipolaire critique. *Bull. Soc. Math. France*, 134(2):201–239, 2006.

[7] T. Duyckaerts. A singular critical potential for the Schrödinger operator. *Canad. Math. Bull.*, 50(1):35–47, 2007.

[8] V. Felli, E.M. Marchini, and S. Terracini. On Schrödinger operators with multipolar inverse-square potentials. *J. Funct. Anal.*, 250(2):265–316, 2007.

[9] E. Fernández-Cara, M. González-Burgos, S. Guerrero, and J.-P. Puel. Null controllability of the heat equation with boundary Fourier conditions: the linear case. *ESAIM Control Optim. Calc. Var.*, 12(3):442–465 (electronic), 2006.

[10] E. Fernández-Cara, S. Guerrero, O. Y. Imanuvilov, and J.-P. Puel. Some controllability results for the N-dimensional Navier-Stokes and Boussinesq systems with $N-1$ scalar controls. *SIAM J. Control Optim.*, 45(1):146–173 (electronic), 2006.

[11] E. Fernández-Cara and E. Zuazua. The cost of approximate controllability for heat equations: the linear case. *Adv. Differential Equations*, 5(4-6):465–514, 2000.

[12] E. Fernández-Cara and E. Zuazua. Null and approximate controllability for weakly blowing up semilinear heat equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 17(5):583–616, 2000.

[13] A. V. Fursikov and O. Y. Imanuvilov. *Controllability of evolution equations*, volume 34 of *Lecture Notes Series*. Seoul National University Research Institute of Mathematics Global Analysis Research Center, Seoul, 1996.

[14] O. Y. Imanuvilov and M. Yamamoto. Carleman inequalities for parabolic equations in Sobolev spaces of negative order and exact controllability for semilinear parabolic equations. *Publ. Res. Inst. Math. Sci.*, 39(2):227–274, 2003.

[15] F.-H. Lin. A uniqueness theorem for parabolic equations. *Comm. Pure Appl. Math.*, 43(1):127–136, 1990.

[16] J.-L. Lions. *Contrôlabilité exacte, Stabilisation et Perturbations de Systèmes Distribués. Tome 1. Contrôlabilité exacte*, volume RMA 8. Masson, 1988.

[17] P. Martinez and J. Vancostenoble. Carleman estimates for one-dimensional degenerate heat equations. *J. Evol. Equ.*, 6(2):325–362, 2006.

[18] V. G. Maz'ja. *Sobolev spaces*. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1985. Translated from the Russian by T. O. Shaposhnikova.

[19] J. Vancostenoble and E. Zuazua. Null controllability for the heat equation with singular inverse-square potentials. *J. Funct. Anal.*, 254(7):1864–1902, 2008.

[20] J. L. Vazquez and E. Zuazua. The Hardy inequality and the asymptotic behaviour of the heat equation with an inverse-square potential. *J. Funct. Anal.*, 173(1):103–153, 2000.