



# THÈSE DE DOCTORAT

École doctorale n° 184 : Mathématiques et Informatique  
Spécialité : Mathématiques appliquées

*pour obtenir le grade de docteur délivré par*

Aix-Marseille Université

*présentée et soutenue publiquement par*

**Damien Allonsius**

le Mercredi 26 Septembre 2018

## Étude spectrale d'opérateurs de Sturm-Liouville et applications à la contrôlabilité de problèmes paraboliques discrets et continus

Devant le jury composé de

**Karine BEAUCHARD,**  
**Assia BENABDALLAH,**  
**Franck BOYER,**  
**Nicolae CINDEA,**  
**Enrique FERNÁNDEZ-CARA,**  
**Otared KAVIAN,**  
**Morgan MORANCEY,**  
**Yannick PRIVAT,**

ENS Rennes  
Aix-Marseille Université  
Université Paul Sabatier  
Université Clermont Auvergne  
Universidad de Sevilla  
Université de Versailles  
Aix-Marseille Université  
Université de Strasbourg

Examinatrice  
Examinatrice  
Directeur  
Examinateur  
Rapporteur  
Examinateur  
Directeur  
Rapporteur



---

# Remerciements

De nombreuses personnes ont contribué à cette thèse, chacune à sa façon, mais toutes avec leur cœur. Ainsi je souhaite consacrer cet espace singulier, unique terre vierge de toute science de ce manuscrit, pour les remercier.

En premier lieu, j'adresse mes plus sincères remerciements à mes maîtres de thèse Franck Boyer et Morgan Morancey, pour leur expertise en théorie du contrôle, bien sûr, mais aussi leur sympathie, leur humour et leur immense soutien, malgré la distance, dans les moments difficiles où le temps pressait. Merci à eux pour avoir semé tout au long de ces trois années leurs nombreuses idées, véritable terreau sans lequel ces recherches n'auraient pu éclore. Mais surtout, merci à ces tuteurs pour m'avoir éduqué et corrigé en faisant preuve d'une patience indéfectible mêlée à une grande exigence, et cela, toujours avec un regard bienveillant sur mon travail.

C'est un honneur pour moi de voir ma thèse rapportée par Yannick Privat et Enrique Fernández Cara. Je remercie très chaleureusement mes rapporteurs pour leur travail minutieux. Merci également à Assia Benabdallah, Karine Beauchard, Nicolae Cindea et Otared Kavian d'avoir accepté de faire partie de mon jury de thèse.

Ensuite, bien sûr, vient ma famille, ma lumière, ces contributeurs invisibles qui m'ont apporté tant de réconfort, de repos et de joie pendant les moments difficiles. Merci à Maman et Christian, Papa et Christine, Chris, Raf et les filles pour tout ce qu'ils ont fait au quotidien pour moi. Je ne sais comment vous exprimer ma gratitude autrement qu'en vous disant que d'une certaine façon, cette thèse, c'est aussi la vôtre. Merci à Papi et La Rose, à mes cousins, cousines, oncles et tantes pour tous les bons moments passés ensemble, malheureusement souvent trop courts et bien trop rares.

Que seraient ces trois années sans les moments d'évasion avec mes machines préférées? Je salue toute l'équipe : Antoine, Camille et Clara, Ludie et Constance, Geoffroy et Nat'nat' (ou Florence Juyon, c'est au choix) pour leur bonne humeur et leurs rires pendant nos week-ends sportifs ou nos aventures à l'étranger. J'ai bien sûr une pensée spéciale pour Geoffroy, Antoine, Camille et Lydie pour toutes les SLEFs, les sessions à Arkose et les trails à Pralo par  $-17$  l'hiver, et  $+40$  l'été. Je tiens à exprimer toute ma gratitude à l'intention d'Antoine, véritable montagne (au moins aussi haute que le petit Mont Blanc, et c'est pas peu dire!) de générosité, pour son hospitalité dans les Alpes comme à Châtelet.

Je remercie également l'ensemble des doctorants du Groupe de Travail des Thésards pour leur accueil cordial au LJLL.

Je salue mes amis cachanais, que j'ai un peu perdu de vue en cours de route : Jessica et Romain, Benji, Ludo, Claire, Loïc, Alain, Nicolas et bien sûr Lilian. Merci à Pierre pour tous nos bons souvenirs à Paris, sur la croisette ou en bateau à Marseille. À quand les prochains à Barcelone ?

Je n'oublie pas mes nouveaux amis marseillais sans qui la thèse aurait été par moments une véritable traversée du désert. Il y a d'abord mes co-bureaux exceptionnels : Thomas, que je remercie en premier parce qu'avec un directeur en commun, on est par définition demi frères de thèse. Merci pour tous tes bons conseils en tous genres. Ensuite, Rémi à qui je dois toujours un croissant et à qui je compte bien payer ma dette avec le pot qui suivra la soutenance! Merci pour ton humour. Arnaud, la gentillesse, le calme et la douceur incarnés, tu resteras toujours pour moi un exemple de modestie et de persévérance. Puis vient Lydia, qui irradie invariablement le bureau de sa bonne humeur, et dont les conversations m'ont apporté souvent bien plus que ce qu'elle ne croît, tant du point de vue humain que mathématique. Merci aussi à toi, très cher Lev-Arcady, pour toutes nos conversations extrêmement intéressantes en algorithmique, dans les calanques, au CMI ou à Arkose. Tu m'as appris énormément de choses en si peu de temps, il est dommage que tu ne sois pas resté davantage parmi nous à Marseille, mais je parie qu'on se retrouvera bientôt. Merci aussi à Marie-Eve, pour tes conseils et tes rires aux éclats. Un immense merci à Sylvie (appelée aussi, soulignons-le, la machine du CMI) pour m'avoir fait découvrir les sentiers du magnifique domaine de l'étoile, que tu connais comme ta poche depuis fort longtemps. Tes gâteaux au chocolat, excellents au goût mais peut-être un peu moins pour la récup', me manqueront toujours! Olivier, merci pour tes petits conseils de toutes sortes, qui

---

m'ont vraiment permis de progresser en info. Merci à toi Michou pour ta sympathie, tes conseils en maths et ton expertise en réglage de pignons de vélo. Je salue aussi Dyonisis mon compagnon du marathon de Nice, les inséparables Antoine Pi. et Juliana, que je remercie au passage pour leur bonne humeur et leurs sourires qui ont égayé nos déjeuners, Antoine Pa. pour nos discussions atypiques sur le Bitcoin, le GR20 et ses voyages, ainsi que Benoît, Pierre-Antoine et Paolo pour tous les bons moments passés ensemble. Plus généralement, merci à toute l'équipe de l'I2M pour leur gentillesse et pour les innombrables parties de Coinche au CMI ou à Porquerolles. Merci aussi à Valérie D. et Valérie J., à Tino et Éric, Evelyne et Marie-Christine pour leur aide dans mes (trop nombreuses!) démarches administratives durant ces trois ans. Merci à mon fidèle ami Thibault pour m'avoir coaché en CAP pendant quelques mois, dans les marches du CMI, sur le tapis de course ou la corniche. Je garde un très bon souvenir de l'UTL 2016. À quand le prochain ?

À Mazel, qui mérite évidemment une place bien spéciale. Merci pour ton amitié sans mesure, qui n'a d'égal que ta gentillesse, ton enthousiasme et ta jovialité. Merci pour tous nos souvenirs. Je n'oublierai pas les nombreux doobby et pectris, les soirées LoR et sushis, ta soirée de thèse et tes singeries... On a bien ri!

Merci à vous tous d'avoir abreuvé ces trois années de votre sympathie et de vos rires.

Pour terminer, je place ici mes remerciements pour une personne bien spéciale, arrivée comme une fleur au milieu de tout cela. Merci à toi Fanny, mon rayon de soleil, pour ta chaleur, ton humeur joyeuse immuable et la confiance que tu m'apportes. Merci aussi à toute la famille Chaix pour les bons moments à Lyon et à Calvi.

Je remercie à l'identique tous ceux que j'oublie pour avoir nourri, guidé et éclairé ce travail ou devrais je dire, ce jeune boulot. Merci à tous d'avoir mis du coeur à l'ouvrage.

# Table des matières

<b>Table des matières</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Contexte Mathématique . . . . .	2
1.2 Problématiques et résultats . . . . .	6
1.3 Idées générales des preuves . . . . .	15
1.4 Perspectives de recherche . . . . .	21
1.5 Références . . . . .	21
<b>2 Spectral analysis and applications in control theory</b>	<b>25</b>
2.1 Introduction . . . . .	27
2.2 Spectral properties of elliptic operators. The continuous case . . . . .	34
2.3 Discrete spectral properties . . . . .	38
2.4 Numerical illustrations . . . . .	49
2.5 Applications in control theory . . . . .	52
2.6 Remarks and further results . . . . .	61
2.7 Appendix . . . . .	62
2.8 References . . . . .	83
<b>3 Boundary null-controllability in multi dimension</b>	<b>87</b>
3.1 Introduction . . . . .	89
3.2 Controllability for linear ODEs . . . . .	92
3.3 Controllability of tensorized systems . . . . .	95
3.4 Applications . . . . .	105
3.5 Appendix . . . . .	118
3.6 References . . . . .	121
<b>4 Temps minimal de contrôle à zéro de l'équation de Grushin</b>	<b>123</b>
4.1 Plan du Chapitre . . . . .	124
4.2 Contrôle à zéro de l'équation de Grushin lorsque $\{0\} \times (0, 1) \subset \omega$ . . . . .	125
4.3 Quelques résultats sur l'équation de Grushin . . . . .	129
4.4 Étude du temps minimal sur une bande lorsque $\gamma = 1$ . . . . .	134
4.5 Annexes . . . . .	149
4.6 Références . . . . .	156



# Chapitre 1

## Introduction

*” Ma cohabitation passionnée avec les mathématiques m’a laissé un amour fou pour les bonnes définitions, sans lesquelles il n’y a que des à-peu-près. ”*

Stendhal

### Sommaire

---

<b>1.1 Contexte Mathématique</b> . . . . .	<b>2</b>
1.1.1 Contrôle à zéro . . . . .	2
1.1.2 La méthode des moments . . . . .	3
1.1.3 La méthode HUM . . . . .	5
<b>1.2 Problématiques et résultats</b> . . . . .	<b>6</b>
1.2.1 Plan du mémoire . . . . .	6
1.2.2 Contrôle au bord et interne de systèmes d’équations paraboliques en cascade discrétisés en une dimension d’espace . . . . .	7
1.2.3 Contrôle au bord de systèmes d’équations paraboliques sur une géométrie cylindrique . . . . .	10
1.2.4 Contrôle à zéro de l’équation de Grushin . . . . .	12
<b>1.3 Idées générales des preuves</b> . . . . .	<b>15</b>
1.3.1 Étude spectrale d’opérateurs de Sturm-Liouville . . . . .	15
1.3.2 Contrôle à zéro au bord de systèmes paraboliques discrétisés sur un domaine $\Omega$ cylindrique . . . . .	17
1.3.3 L’équation de Grushin . . . . .	18
<b>1.4 Perspectives de recherche</b> . . . . .	<b>21</b>
<b>1.5 Références</b> . . . . .	<b>21</b>

---

Dans cette thèse, nous étudions la contrôlabilité à zéro d'équations paraboliques linéaires et de leurs discrétisations. Les travaux exposés dans ce manuscrit se regroupent en deux thèmes distincts. Le premier (chapitre 2 et chapitre 3) concerne les propriétés de contrôlabilité à zéro de systèmes d'équations paraboliques en cascade, semi-discrétisées en espace, contrôlés en particulier par le bord. Dans une seconde partie (chapitre 4), nous étudions les propriétés de temps minimal de contrôlabilité à zéro d'une équation dégénérée (l'équation de Grushin) sur un domaine rectangulaire, contrôlée sur une bande verticale.

Bien que de natures différentes, ces deux problèmes sont abordés en utilisant principalement le même outil : *la méthode des moments*. Cette méthode requiert des informations précises sur le spectre des opérateurs elliptiques en jeu. L'étude de ces deux problèmes de contrôle se ramènera, en grande partie, à l'étude du spectre d'opérateurs de Sturm-Liouville, continus ou discrets, et des fonctions propres associées.

Cette introduction se décompose en plusieurs parties. Nous commençons par énoncer les résultats obtenus dans cette thèse puis nous décrivons succinctement les idées et les outils développés pour démontrer ces théorèmes. Nous terminons cette introduction en donnant les perspectives de recherche.

## 1.1 Contexte Mathématique

Dans cette section nous introduisons différentes définitions relatives à la contrôlabilité des équations paraboliques puis nous décrivons la méthode des moments. On se basera sur l'équation de la chaleur pour illustrer nos propos.

### 1.1.1 Contrôle à zéro

Soient  $\Omega$  un ouvert borné, connexe de  $\mathbb{R}^D$  et de régularité  $\mathcal{C}^2$  et  $\omega \subset \Omega$  un ouvert de mesure non nulle. Nous nous donnons un temps final  $T > 0$  et nous posons  $Q_T = (0, T) \times \Omega$ . On considère l'équation de la chaleur, contrôlée sur la zone de contrôle  $\omega$ ,

$$\begin{cases} \partial_t y - \Delta y = 1_\omega u, & \text{dans } Q_T, \\ y(0) = y_0 \in L^2(\Omega) & \text{dans } \Omega, \\ y = 0 & \text{sur } (0, T) \times \partial\Omega. \end{cases} \quad (1.1)$$

Ici  $y_0$  représente la donnée initiale et l'équation est contrôlée par la fonction  $u$  que nous appellerons *contrôle*. Un des objectifs de la théorie du contrôle est de prouver l'existence d'un contrôle  $u$  tel que la solution correspondante ait certaines propriétés que nous allons préciser ci-dessous. Pour une introduction plus exhaustive de la théorie du contrôle, nous renvoyons à [14], [30] et [31].

Avant toute chose, rappelons le Théorème classique suivant adapté de [14, thm 2.63].

**Théorème 1.1.1** *Pour toute donnée initiale  $y_0 \in L^2(\Omega)$  et tout contrôle  $u \in L^2(Q_T)$ , l'équation (1.1) admet une unique solution notée  $S(t, y_0, u)$ , au sens suivant :  $\forall \varphi \in L^2(\Omega)$ ,  $\forall t \in (0, T)$ ,*

$$\langle S(t, y_0, u), \varphi \rangle_{L^2(\Omega)} - \langle y_0, e^{t\Delta} \varphi \rangle_{L^2(\Omega)} = \int_0^t \langle 1_\omega u, e^{(t-s)\Delta} \varphi \rangle_{L^2(\Omega)} ds. \quad (1.2)$$

*De plus, cette solution vérifie  $S(\bullet, y_0, u) \in \mathcal{C}^0([0, T]; L^2(\Omega))$  ainsi que*

$$\sup_{t \in [0, T]} \|S(t, y_0, u)\|_{L^2(\Omega)} \leq C (\|y_0\|_{L^2(\Omega)} + \|u\|_{L^2(Q_T)}).$$

**Remarque 1.1.1** *Nous considérerons également le cas d'un contrôle au bord  $v$  évoluant dans l'espace  $L^2((0, T) \times \Gamma)$  où  $\Gamma$  est une partie du bord  $\Gamma \subset \partial\Omega$ . Pour une donnée initiale  $y_0 \in H^{-1}(\Omega)$ , la solution de l'équation (1.1) est la fonction  $S(t, y_0, u)$  vérifiant pour tout  $\varphi \in H_0^1(\Omega)$ ,*

$$\langle S(t, y_0, u), \varphi \rangle_{H^{-1}, H_0^1} - \langle y_0, e^{t\Delta} \varphi \rangle_{H^{-1}, H_0^1} = \int_0^t \left\langle v, \frac{\partial}{\partial n} e^{(t-s)\Delta} \varphi \right\rangle_{L^2(\partial\Omega)} ds. \quad (1.3)$$

Dans cette thèse, nous nous intéresserons particulièrement à la notion de contrôlabilité à zéro.

**Définition 1.1.1 (Contrôlabilité à zéro)** *On dit que l'équation (1.1) est contrôlable à zéro au temps  $T > 0$  si pour toute donnée initiale  $y_0 \in L^2(\Omega)$ , il existe un contrôle  $u \in L^2(Q_T)$  tel que*

$$S(T, y_0, u) = 0.$$



Le résultat suivant est connu depuis 1974 et a été démontré par H.O. Fattorini et D.L. Russell (voir [19]) par la méthode des moments que nous détaillons dans la section 1.1.2 ci-dessous.

**Théorème 1.1.2** *L'équation (1.1) est contrôlable à zéro en tout temps  $T > 0$ , lorsque  $D = 1$ .*

Une autre approche pour démontrer la contrôlabilité à zéro de l'équation (1.1) repose sur le lien entre la contrôlabilité à zéro et l'inégalité d'observabilité (1.4) ci-dessous. La proposition suivante est due à S. Dolecki et à D.L. Russell (voir [16]).

**Proposition 1.1.1** *Soit  $T > 0$ . S'il existe  $C > 0$  telle que pour toute donnée initiale  $y_0 \in L^2(\Omega)$ , la solution de (1.1) avec contrôle nul vérifie*

$$\|e^{T\Delta}y_0\|_{L^2(\Omega)}^2 \leq C^2 \int_0^T \|e^{t\Delta}y_0\|_{L^2(\omega)}^2 dt, \quad (1.4)$$

alors l'équation (1.1) est contrôlable à zéro au temps  $T$ .

**Remarque 1.1.2** *On pose  $C_{obs}^2 = \inf_{y_0 \in L^2(\Omega)} \frac{\|e^{T\Delta}y_0\|_{L^2(\Omega)}^2}{\int_0^T \int_{\omega} (e^{t\Delta}y_0)^2 dx dt}$ . D'après [14, Thm 2.44], étant donné  $y_0 \in L^2(\Omega)$ , la norme du contrôle  $L^2(Q_T)$  minimal amenant à zéro la condition initiale  $y_0$  est majorée par  $C_{obs}\|y_0\|_{L^2(\Omega)}$ . Nous utilisons cette propriété dans le chapitre 4 afin de construire une suite de contrôles dont le coût est uniformément majoré en norme  $L^2(Q_T)$ .*

### 1.1.2 La méthode des moments

La démonstration de Fattorini et Russell du Théorème 1.1.2 consiste en l'application de la *méthode des moments*. Nous la détaillons ci-dessous afin de dégager une stratégie générale pour les systèmes de contrôle que nous étudions. Cette méthode repose sur l'existence et l'estimation de *familles biorthogonales*, dans  $L^2(0, T)$ , aux fonctions  $(t \mapsto e^{-\lambda_k t})_{k \geq 1}$  où  $(\lambda_k)_{k \geq 1}$  est une suite de réels croissante et positive. Pour cette preuve, la méthode requiert de choisir la suite  $\Lambda := (\lambda_k)_{k \geq 1}$  des valeurs propres de l'opérateur  $-\Delta$  avec conditions aux bords de Dirichlet, sur un intervalle borné de dimension 1. Plus précisément, une telle famille de fonctions (notée  $(q_j^\Lambda)_{j \geq 1}$ ) vérifie :

$$\int_0^T q_j^\Lambda(t) e^{-\lambda_k t} dt = \delta_{k,j}, \quad \text{pour tout } k, j \in \mathbb{N}^*. \quad (1.5)$$

Les origines de cette méthode remontent aux travaux de Müntz (voir [27] et aussi les travaux connexes de Schwartz dans [28]) où l'auteur prouve que la condition suivante est nécessaire pour qu'une telle famille biorthogonale existe

$$\sum_{k \geq 1} \frac{1}{\lambda_k} < +\infty. \quad (1.6)$$

Cette méthode sera utilisée à maintes reprises dans cette thèse. Ainsi, la condition (1.6) sera cruciale dans tout ce qui suit.

Pour fixer les idées, choisissons le domaine  $\Omega := (0, 1)$  dans l'équation (1.1). D'après l'égalité (1.2), la solution  $t \mapsto S(t, y_0, u)$  est nulle au temps  $T$  si et seulement si pour tout  $\varphi \in L^2(\Omega)$ ,

$$-(y_0, e^{T\Delta}\varphi)_{L^2(\Omega)} = \int_0^T (1_\omega u, e^{(T-t)\Delta}\varphi)_{L^2(\Omega)} dt. \quad (1.7)$$

Appelons  $(\phi_k)_{k \geq 1}$  les fonctions propres normalisées associées aux valeurs propres  $\Lambda := (\lambda_k)_{k \geq 1}$  de l'opérateur  $-\Delta$  avec conditions aux bords de Dirichlet. La solution de l'équation de la chaleur sans terme source et condition initiale  $\phi_k$ , notée  $t \mapsto e^{t\Delta}\phi_k$ , s'exprime alors simplement par

$$e^{t\Delta}\phi_k = e^{-\lambda_k t}\phi_k. \quad (1.8)$$

Les fonctions propres  $(\phi_k)_{k \geq 1}$  forment une base de Hilbert de  $L^2(\Omega)$ , donc en injectant l'expression (1.8) dans (1.7) avec  $\varphi = \phi_k$ , nous trouvons

$$S(T, y_0, u) = 0 \Leftrightarrow \forall k \in \mathbb{N}^*, -e^{-\lambda_k T} (y_0, \phi_k)_{L^2(\Omega)} = \int_0^T (1_\omega u, \phi_k e^{-\lambda_k(T-t)})_{L^2(\Omega)} dt. \quad (1.9)$$

L'idée consiste maintenant à exprimer  $u$  à l'aide d'une famille biorthogonale  $(q_j^\Lambda)_{j \geq 1}$ , vérifiant par définition les égalités (1.5). Nous supposons pour l'instant qu'une telle famille existe. Il est alors possible de rechercher le contrôle  $u$  sous deux formes différentes. La première dépend d'un profil  $f$  indépendant de l'indice  $j$ . Le contrôle  $u$  recherché est alors découpé en les variables  $x$  et  $t$  et il s'écrit

$$u_1(t, x) := f(x) \sum_{j=1}^{+\infty} \alpha_j q_j^\Lambda(t), \text{ avec } \alpha_j \in \mathbb{R}, \forall j \geq 1. \quad (1.10)$$

La seconde dépend d'un profil  $f_j$  dépendant de l'indice  $j$  et nous choisissons :  $f_j(x) = (1_\omega \phi_j)(x)$  :

$$u_2(t, x) := \sum_{j=1}^{+\infty} \alpha_j (1_\omega \phi_j)(x) q_j^\Lambda(t), \text{ avec } \alpha_j \in \mathbb{R}, \forall j \geq 1. \quad (1.11)$$

L'équivalence (1.9) permet alors d'exprimer les contrôles sous les formes

$$\begin{aligned} u_1(t, x) &= -f(x) \sum_{j=1}^{+\infty} \frac{e^{-\lambda_j T} (y_0, \phi_j)_{L^2(\Omega)}}{(f, \phi_j)_{L^2(\omega)}} q_j^\Lambda(t), \\ u_2(t, x) &= - \sum_{j=1}^{+\infty} \frac{e^{-\lambda_j T} (y_0, \phi_j)_{L^2(\Omega)}}{(\phi_j, \phi_j)_{L^2(\omega)}} (1_\omega \phi_j)(x) q_j^\Lambda(t). \end{aligned} \quad (1.12)$$

**Remarque 1.1.3** Cette distinction n'a pas lieu pour le contrôle au bord car dans ce cas le contrôle ne dépend pas de la variable  $x$ . L'expression du contrôle est obtenue à partir de la forme  $u_2$  en remplaçant le dénominateur  $(\phi_j, \phi_j)_{L^2(\omega)}$  par la dérivée normale de  $\phi_j$  au bord où l'on contrôle l'équation et enlevant le terme  $1_\omega \phi_j$ .

Outre l'existence de la famille biorthogonale  $(q_j^\Lambda)_{j \geq 1}$ , il faut s'assurer de la convergence des séries (1.12) en majorant la norme  $L^2(0, T)$  des fonctions  $q_j^\Lambda$  et en minorant les quantités  $(\phi_k, \phi_k)_{L^2(\omega)}$  ou  $(f, \phi_k)_{L^2(\omega)}$ .

L'existence et l'estimation de la famille  $(q_j^\Lambda)_{j \geq 1}$  se traitent grâce au Théorème 1.1.3 ci-dessous.

**Définition 1.1.2** Soit  $\rho > 0$  et  $\mathcal{N} : \mathbb{R}^+ \rightarrow \mathbb{N}$ . On définit  $\mathcal{L}(\rho, \mathcal{N})$ , l'ensemble des suites  $\Sigma = (\sigma_k)_{k \geq 1}$  telles que :

1.  $\forall k \geq 1, \sigma_{k+1} - \sigma_k \geq \rho$  (condition de gap),
2.  $\forall \delta > 0, \sum_{k=\mathcal{N}(\delta)}^{+\infty} \frac{1}{\sigma_k} \leq \delta$ .

Le théorème suivant est extrait de [19].

**Théorème 1.1.3 (Théorème de Fattorini-Russel)** Soit  $\rho > 0$  et  $\mathcal{N} : \mathbb{R}^+ \rightarrow \mathbb{N}$ .

$\forall \varepsilon > 0, \exists K_\varepsilon > 0, \forall \Sigma \in \mathcal{L}(\rho, \mathcal{N}), \exists (q_j^\Sigma)_{j \geq 1}$  (biorthogonale à la famille  $(t \mapsto e^{-\sigma_k t})_{k \geq 1}$ ),

$$\forall j \geq 1, \|q_j^\Sigma\|_{L^2(0, T)} \leq K_\varepsilon \exp(\varepsilon \sigma_j). \quad (1.13)$$

**Remarque 1.1.4** 1. Il est important de bien noter que la majoration (1.13) ne dépend pas de la famille  $\Sigma$  mais uniquement des paramètres  $\rho$  et  $\mathcal{N}$ .

2. Dans l'article [10] les auteurs obtiennent une borne plus précise que (1.13). En ajoutant des hypothèses sur l'ensemble  $\mathcal{L}(\rho, \mathcal{N})$ , ils précisent la dépendance en temps de la constante  $K_\varepsilon$  et démontrent que  $q_j^\Sigma$  est majoré par  $C e^{C \sqrt{\sigma_j} + \frac{C}{T}}$ . Cette information sur la dépendance en temps sera précieuse lorsque nous appliquerons la méthode de Lebeau-Robbiano au chapitre 3.

Ce théorème fournit l'existence et une borne sur la famille  $(q_j^\Lambda)_{j \geq 1}$  sitôt que l'on remarque que, en dimension 1 d'espace, les valeurs propres de l'opérateur  $-\Delta$  sur  $\Omega = (0, 1)$  sont  $\lambda_k = k^2 \pi^2$ . Il reste donc à estimer les dénominateurs des coefficients  $\alpha_j$  des contrôles  $u_1$  et  $u_2$ .

La première approche a été étudiée dans [18] et [5]. En choisissant un profil  $f = 1_{(a,b)}$  avec des réels  $a$  et  $b$  judicieusement choisis dans le domaine de contrôle  $\omega$ , les auteurs de [5] montrent qu'il existe  $C > 0$  tel que pour tout  $k \in \mathbb{N}^*$ ,

$$(f, \phi_k)_{L^2(\omega)} \geq \frac{C}{k^3}. \quad (1.14)$$

L'inégalité (1.14) suffit alors à montrer que la série définie par  $u_1$  converge pour tout temps  $T$  et ceci prouve le Théorème 1.1.2. Cette méthode souffre cependant de plusieurs inconvénients. D'une part, la construction des réels  $a$  et  $b$  est fortement basée sur la connaissance de l'expression des fonctions propres de l'opérateur  $-\Delta$  avec conditions aux bords de Dirichlet. A notre connaissance, il n'existe pas d'inégalité analogue à (1.14) pour des opérateurs de Sturm-Liouville plus généraux tels que (1.18). D'autre part, la première approche se transporte difficilement à des systèmes discrétisés en dehors du cas où l'opérateur  $\mathcal{A}^h$  est la discrétisation de l'opérateur  $-\Delta$  (voir l'annexe du chapitre 2, section 2.7.4).

La seconde méthode, celle qui consiste à chercher un contrôle sous la forme (1.11), s'applique beaucoup plus facilement car elle ne nécessite pas la construction d'un profil  $f$  mais simplement une minoration de  $(\phi_k, \phi_k)_{L^2(\omega)}$ . En choisissant un intervalle  $(a, b) \subset \omega$ , le calcul donne

$$(\phi_k, \phi_k)_{L^2(\omega)} \geq \int_a^b \frac{1 - \cos(2k\pi x)}{2} dx \sim \frac{|b-a|}{2} > 0,$$

ce qui suffit à conclure. De plus, dans le chapitre 2, nous élaborons une nouvelle méthode permettant d'évaluer  $(\phi_k, \phi_k)_{L^2(\omega)}$  dans le cadre continu et discret, pour des opérateurs de Sturm-Liouville assez généraux.

Retenons que dans tous les cas, pour montrer la contrôlabilité à zéro au temps  $T$  par la méthode des moments, il suffit de montrer

1. des propriétés permettant de construire et de majorer la famille biorthogonale  $(q_j^\Lambda)_{j \geq 1}$ , c'est-à-dire
  - (a) qu'il existe  $\rho > 0$  tel que  $\forall k \in \mathbb{N}^*$ ,  $\lambda_{k+1} - \lambda_k \geq \rho$  (propriété de gap),
  - (b) qu'il existe  $\mathcal{N} : \mathbb{R}_+ \rightarrow \mathbb{N}$  telle que  $\forall \delta > 0$ ,  $\forall k \in \mathbb{N}^*$ ,  $\sum_{k=\mathcal{N}(\delta)}^\infty \frac{1}{\lambda_k} \leq \delta$ .
2. soit une minoration de  $(\phi_k, \phi_k)_{L^2(\omega)}$  par un terme de la forme  $C_1 e^{-C_T \lambda_k}$ , avec  $C_T > 0$  strictement inférieur à  $T$ , soit qu'il existe un profil  $f$  tel que  $(f, \phi_k)_{L^2(\omega)}$  soit minoré de la même façon.

**Remarque 1.1.5** Dans le cas du contrôle au bord, la méthode des moments s'applique de la même façon en remplaçant le point 2. ci-dessus par

2.bis Il suffit de montrer une minoration de  $|\phi'_k(1)|$  par un terme de la forme  $C_1 e^{-C_T \lambda_k}$ , avec  $C_T > 0$  strictement inférieur à  $T$ . Pour le contrôle au bord  $\{x = 0\}$ , il faudra plutôt chercher une minoration de  $|\phi'_k(0)|$ .

### 1.1.3 La méthode HUM

La méthode des moments est un outil théorique permettant de prouver l'existence d'un contrôle à zéro de l'équation (1.1) et des estimations sur celui-ci. Comme nous le verrons dans les chapitres qui suivent, celle-ci s'adapte également aux systèmes semi-discrétisés en espace. Cependant, elle ne fournit pas en pratique un algorithme adapté au calcul numérique d'un contrôle à zéro uniforme sur une équation discrétisée (voir la définition 1.2.1 et la discussion dans les paragraphes qui suivent). La méthode HUM quant à elle comble ce manque. En guise d'introduction, nous la présentons succinctement ci-dessous dans le cas continu, nous renvoyons le lecteur intéressé à [11, 20] pour une étude plus approfondie de cette méthode.

Nous étudions une version pénalisée de la méthode HUM. Celle-ci consiste à minimiser la fonctionnelle (ou énergie) suivante, dépendant d'un petit paramètre  $\varepsilon > 0$ , sur l'espace de contrôle

$$F_\varepsilon(v) := \frac{1}{2} \|v\|_{L^2(0,T)}^2 + \frac{1}{\varepsilon} \|S(T, y_0, v)\|^2. \quad (1.15)$$

Notons  $v_\varepsilon$  le contrôle minimisant cette fonctionnelle. Quand le paramètre  $\varepsilon$  tend vers 0, le second terme du membre de droite tend vers  $+\infty$  si la solution  $S(T, y_0, v_\varepsilon)$  ne tend pas suffisamment vite

vers 0 lorsque  $\varepsilon \rightarrow 0$ . On *pénalise* donc l'écart à zéro de la solution au temps  $T$  à mesure que  $\varepsilon$  devient petit. Ainsi, intuitivement, nous nous attendons à ce que la limite quand  $\varepsilon$  tend vers 0 (si elle existe) du contrôle  $v_\varepsilon$  soit égale à un contrôle à zéro de l'équation (1.1). On peut montrer que le contrôle  $v_\varepsilon$  s'exprime par une formule, qui contrairement à la méthode des moments, pourra être calculée numériquement lorsque l'on considérera des systèmes discrétisés :

$$v_\varepsilon = -S^*(T, (\Lambda + \varepsilon I_d)^{-1} S(T, y_0, 0), 0),$$

où  $S^*(T, y_0, 0)$  est l'adjoint de  $y_0 \mapsto S(T, y_0, 0)$  et  $\Lambda$  est l'opérateur de Gram défini, dans le cas du contrôle interne, par

$$\Lambda \phi = S(T, 1_\omega S^*(T, \phi, 0), 0).$$

Dans [11], l'auteur montre que la contrôlabilité de certains systèmes est liée à une majoration uniforme de l'énergie optimale  $F_\varepsilon(v_\varepsilon)$  (voir [11, Théorème 1.7]). Ce théorème affirme, entre autre, qu'il existe un contrôle amenant la donnée initiale  $y_0$  à zéro si et seulement si

$$\lim_{\varepsilon \rightarrow 0} F_\varepsilon(v_\varepsilon) < +\infty. \tag{1.16}$$

Ce lien entre la contrôlabilité et la majoration de l'énergie optimale  $F_\varepsilon(v_\varepsilon)$  existe également dans le cadre discret (voir la définition 2.1, la proposition 2.2 et la section 3 de [11]). Nous donnons des simulations numériques basées sur la méthode HUM dans les annexes des chapitres 2 et 3 et nous examinerons le lien entre l'allure de la solution contrôlée et le comportement de l'énergie optimale.

## 1.2 Problématiques et résultats

### 1.2.1 Plan du mémoire

En dehors de cette introduction, ce mémoire est composé de trois chapitres.

- L'article [3] constitue la majorité du chapitre 2. Il est composé de deux parties principales. Dans un premier temps, nous montrons des estimées sur le spectre et les vecteurs propres d'un opérateur de Sturm-Liouville discrétisé par la méthode des différences finies sur différents maillages. Dans une seconde partie, nous montrons des résultats de contrôlabilité (notamment au bord) de systèmes en cascade d'équations paraboliques discrétisées, en appliquant la méthode des moments. Cette méthode possède une limitation : comme nous le verrons, en toute généralité, elle ne peut s'appliquer qu'à des domaines de dimension  $D$  égale à 1. Une Annexe vient compléter ce chapitre substantiellement. Le premier appendice fournit des précisions sur la Remarque 2.1.6. Il y est expliqué comment certains résultats du chapitre 2 posés sur des maillages uniformes peuvent être étendus à des maillages *réguliers*, c'est-à-dire des maillages dont les points sont les images des points d'une discrétisation uniforme du domaine par une fonction régulière. Une note en préparation (voir [2]) constitue le second appendice. On montre une borne générale sur la différence des valeurs propres d'un opérateur de Sturm-Liouville avec son équivalent discrétisé sur un maillage quasi-uniforme. Cette borne est utilisée dans le papier [3]. Dans le troisième appendice, nous étendons certains résultats de l'article [3] à des fonctions  $\gamma$  constantes par morceaux. Dans le quatrième appendice, nous construisons un contrôle à variables séparées pour contrôler à zéro l'équation de la chaleur sur un maillage uniforme. Dans la dernière, nous donnons quelques simulations numériques obtenues avec la méthode HUM pénalisée.
- Le chapitre 3 est constitué de l'article [1]. Comme on l'a évoqué ci-dessous, les méthodes employées dans le chapitre 2 sont, par nature, limitées à la dimension 1 d'espace. Il est donc nécessaire de développer de nouvelles approches pour traiter les cas multi-dimensionnels. Ainsi, en s'inspirant de [10], on établit dans le chapitre 3 des résultats dans ce cadre multi-D pour des systèmes semi-discrétisés. L'étude menée dans le chapitre 2 est un des ingrédients essentiels pour l'obtention de ces nouveaux résultats. Précisons que le théorème principale du chapitre 3 s'applique dans un cadre plus général que celui du chapitre 2 et permet également de retrouver le résultat principal de l'article [10] qui concerne un système d'équations paraboliques continues. Quelques simulations numériques dans un domaine  $\Omega$  de dimension 2 viennent illustrer nos propos à la fin de ce chapitre.
- Dans le chapitre 4, nous présentons nos travaux sur le temps minimal de contrôle à zéro sur une bande verticale de l'équation de Grushin posée sur un rectangle. Nous mettons en oeuvre la méthode des moments et montrons toutes les propriétés suffisantes à son application, hormis la propriété de gap uniforme, que nous ne sommes pas parvenus à démontrer. En effet, nous prouvons cette propriété sur l'ensemble du spectre, excepté une bande très localisée d'indices.

### 1.2.2 Contrôle au bord et interne de systèmes d'équations paraboliques en cascade discrétisés en une dimension d'espace

Dans cette section nous énonçons les principaux résultats du chapitre 2.

Considérons l'équation parabolique suivante où la variable d'espace varie dans l'ensemble  $\Omega := (0, 1)$ .

$$\begin{cases} \partial_t y + \mathcal{A}y = 1_\omega V_d, & \text{sur } Q_T, \\ y = 0, & \text{sur } \{0\} \times (0, T), \\ y = V_b, & \text{sur } \{1\} \times (0, T), \\ y(0) = y_0, & \text{sur } \Omega. \end{cases} \quad (1.17)$$

L'équation (1.17) est contrôlée sur un domaine  $\omega \subset \Omega$  par un contrôle distribué  $V_d \in L^2(Q_T)$  ou par un contrôle sur un bord, que l'on note  $V_b \in L^2(0, T)$ . Pour fixer les idées, nous choisissons un contrôle sur le bord droit  $\{x = 1\}$ , évidemment l'étude reste inchangée si l'on contrôle l'équation (1.17) sur l'autre bord. On pourra poser  $V_b = 0$  ou  $V_d = 0$  pour n'étudier que la contrôlabilité de (1.17) sur la zone  $\omega$  ou sur le bord. L'opérateur elliptique  $\mathcal{A}$  est un opérateur de *Sturm-Liouville* sur  $L^2(\Omega)$ , défini par

$$\mathcal{A} := -\partial_x(\gamma(x)\partial_x \bullet) + q(x) \bullet. \quad (1.18)$$

Par exemple, l'opérateur  $-\Delta$  considéré dans les sections précédentes est un opérateur de Sturm-Liouville. Dans le chapitre 2 on étudie la contrôlabilité de l'équation (1.17) semi-discrétisée en espace par la méthode des différences finies. On considérera des fonctions  $\gamma$  et  $q$  dont la régularité choisie dépendra du type de maillage considéré et l'on supposera toujours que

$$\gamma_{\min} := \inf_{\Omega} \gamma(x) > 0. \quad (1.19)$$

L'opérateur  $\mathcal{A}$  est discrétisé selon un schéma aux différences finies classique à 3 points et, évalué en  $U = (u_j)_{j \in \llbracket 1, N \rrbracket}$ , son expression sur un maillage uniforme est

$$(\mathcal{A}^h U)_j := -\frac{1}{h} \left( \gamma_{j+1/2} \frac{u_{j+1} - u_j}{h} - \gamma_{j-1/2} \frac{u_j - u_{j-1}}{h} \right) + q(jh)u_j.$$

L'équation (1.17) discrétisée s'écrit

$$\begin{cases} (y^h)'(t) + \mathcal{A}^h y^h(t) = 1_\omega^h V_d^h(t) + \mathcal{B}_r^h V_b^h(t), & \text{où } t \in (0, T], \\ y^h(0) = y^{0,h} \in \mathbb{R}^N, \end{cases} \quad (1.20)$$

avec  $1_\omega^h$  et  $\mathcal{B}_r^h$  les opérateurs discrets de contrôle distribué et au bord droit. On notera  $\|\cdot\|$  la norme  $L^2$  discrète sur le maillage et respectivement  $\partial_r$  la dérivée discrète au bord droit.

On parlera de contrôlabilité à zéro de l'équation (1.20) au sens suivant.

**Définition 1.2.1 (Contrôlabilité à zéro uniforme discrète)** *Le système (1.20) est contrôlable à zéro sur  $\omega$  (resp. au bord droit) si pour toute donnée initiale discrète  $y^{0,h} \in \mathbb{R}^N$ , il existe un contrôle  $V_d^h$  (resp.  $V_b^h$ ) tel que*

1.  $\int_0^T \|V_d^h(t)\|^2 dt \leq C \|y^{0,h}\|^2$  (resp.  $\int_0^T |V_b^h(t)|^2 dt \leq C \|y^{0,h}\|^2$ ),
2. la solution  $y^h$  de (1.20) avec  $V_b^h = 0$  (resp.  $V_d^h = 0$ ) vérifie  $y^h(T) = 0$ .

Lorsque  $\mathcal{A}^h = -\Delta^h$  et que le maillage est uniforme, l'équation (1.20) est contrôlable à zéro au bord (i.e. avec  $V_d^h = 0$ ). Ceci est prouvé dans [25]. Les Théorèmes du chapitre 2 étendent les résultats de [25] au contrôle interne mais aussi à des maillages et des opérateurs plus généraux ainsi qu'à des systèmes d'équations en cascade (voir (S<sub>2</sub><sup>h</sup>) plus bas pour la définition de ces systèmes).

Lorsque l'on souhaite amener la solution d'un système discrétisé à zéro au temps  $T$ , il est aisé de construire une famille biorthogonale ad hoc car l'ensemble des valeurs propres de l'opérateur discrétisé est fini (on peut penser par exemple au procédé d'orthonormalisation de Schmidt). Cependant, cette construction ne garantit aucune borne uniforme en le paramètre de discrétisation sur la famille biorthogonale finie ainsi construite et donc, aucune borne uniforme sur le coût du contrôle. Le Théorème 1.1.3 permet quant à lui de construire une suite de contrôles pour l'équation discrétisée (1.20) vérifiant de plus les conditions de majoration uniforme (i.e. les points 1. des définitions 1.2.1 et 1.2.2). En effet,

comme nous l'avons remarqué dans la Remarque 1.1.4, la borne fournie sur la famille biorthogonale  $(q_j^\Sigma)_{j \geq 1}$  ne dépend pas du choix de la suite  $\Sigma$ , mais uniquement des paramètres  $\rho$  et  $\mathcal{N}$  de la famille  $\mathcal{L}(\rho, \mathcal{N})$ . Les familles biorthogonales seront donc uniformément majorées par rapport au pas du maillage s'il existe  $\rho$  et  $\mathcal{N}$  indépendants de  $h$  tels que pour tout  $h > 0$ , l'ensemble des valeurs propres de l'opérateur  $\mathcal{A}^h$  convenablement complété en une suite infinie et noté  $\Lambda^h$ , vérifie  $\Lambda^h \in \mathcal{L}(\rho, \mathcal{N})$ .

Pour appliquer la méthode des moments, nous avons donc besoin d'informations sur le spectre et les vecteurs propres de l'opérateur  $\mathcal{A}^h$  que l'on note respectivement  $(\lambda_k^h)_{k \in \llbracket 1, N \rrbracket}$  et  $(\phi_k^h)_{k \in \llbracket 1, N \rrbracket}$ . Dans la première partie du travail présenté dans le chapitre 2, nous étudions donc les éléments du spectre de  $\mathcal{A}^h$  sur différents maillages et l'on obtient le Théorème ci-dessous dans le cas d'un maillage uniforme (voir Théorème 2.3.2 du chapitre 2 pour un énoncé plus précis).

**Théorème** *Sous certaines hypothèses de régularité de  $\gamma$  et  $q$ , il existe  $\alpha(\gamma, q) \in (0, 1]$  tel que pour tout maillage uniforme de paramètre  $h$  on a,*

$$\frac{1}{\sqrt{\lambda_k^h}} |\partial_r \phi_k^h| \geq C, \forall k \in \llbracket 1, \alpha N \rrbracket, \quad (1.21)$$

$$\|1_\omega \phi_k^h\|^2 \geq C, \forall k \in \llbracket 1, \alpha N \rrbracket, \quad (1.22)$$

et de plus, la propriété suivante dite propriété de gap est satisfaite

$$\lambda_{k+1}^h - \lambda_k^h \geq C \sqrt{\lambda_{k+1}^h}, \forall k \in \llbracket 1, \alpha N \rrbracket \quad (1.23)$$

Il est important de noter que les estimées obtenues ne sont pas montrées sur le spectre entier, mais bien sur une fraction  $\alpha$  de celui-ci.

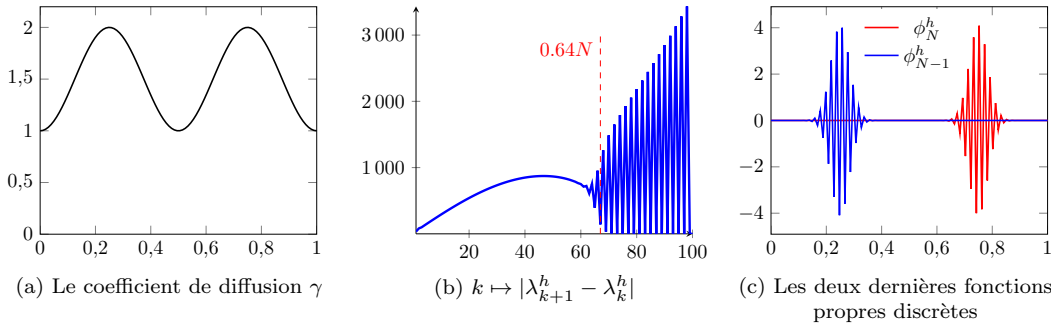


FIGURE 1.1 – Le gap et les deux dernières fonctions propres discrètes avec  $N = 100$ ,  $\gamma = 2 - \cos(2\pi x)^2$  et  $q = 0$ .

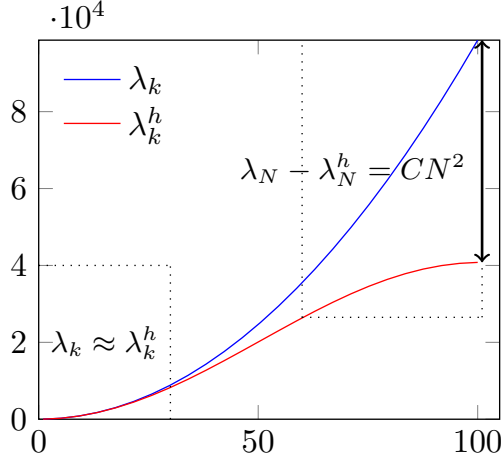
Dans la section 2.4 du chapitre 2, nous étudions numériquement l'optimalité de l'indice  $\alpha N$  en fonction de  $\gamma$  (voir la définition 2.29). Pour cela, choisissons par exemple un maillage uniforme avec un nombre de mailles variant de 50 à 400,  $\gamma = 2 - \cos(2\pi x)^2$  et  $q = 0$ . La définition (2.29) de la section 2.3.3 nous conduit à poser  $\alpha \approx 0.64$ . Le tableau 1.1 et la figure 1.1 montrent que ce choix semble raisonnablement optimal. En effet, les quantités discrètes du théorème précédent évaluées sur l'ensemble des indices  $\llbracket 1, N \rrbracket$  semblent converger vers 0 à mesure que le pas du maillage  $h$  tend vers 0 alors qu'elles semblent rester bornées lorsqu'on ne considère que les indices  $\llbracket 1, 0.64N \rrbracket$ . Dans le tableau suivant, nous donnons des estimations numériques du minimum du gap et des dérivées discrètes au bord droit et de l'intégrale discrète sur  $\omega = (0, 0.3)$  des fonctions  $\phi_i^h$  pour  $i \in \llbracket 1, 0.64N \rrbracket$  et  $i \in \llbracket 1, N \rrbracket$ .

Lorsque le coefficient de diffusion  $\gamma$  est constant, la propriété de gap est valable pour l'ensemble du spectre, autrement dit  $\alpha = 1$  (voir Théorème 2.3.3).

Nous montrons également le même type de minoration sur des maillages quasi-uniformes, c'est-à-dire des maillages non uniformes pour lesquels les mailles sont de tailles comparables (voir définition 2.1.2). Cette propriété de gap (1.23) s'obtient naturellement en comparant les valeurs propres  $\lambda_k^h$  de l'opérateur discrétisé  $\mathcal{A}^h$  aux valeurs propres  $\lambda_k$  de l'opérateur continu  $\mathcal{A}$ , pour lequel la propriété de gap est vraie pour l'ensemble des valeurs propres  $(\lambda_k)_{k \in \mathbb{N}^*}$ . La figure 1.2 montre que cette façon d'approcher les valeurs propres discrètes n'est valable que pour la partie basse du spectre, même lorsque la discrétisation est uniforme.

Nous obtenons ainsi la propriété de gap (1.23) seulement pour les indices allant de 1 à  $N^{2/5}$  (voir le Théorème 2.3.1). Nous démontrons ensuite dans la section 2.7.2 du chapitre 2 le résultat suivant à l'aide du principe du Min Max (voir Proposition 2.3.1 pour un énoncé plus précis).

$N$	$\min_{i \leq (\cdot)} \partial_r \phi_i^h$		$\min_{i \leq (\cdot)} 1_\omega^h \phi_i^h$		$\min_{i \leq (\cdot)} \lambda_{i+1}^h - \lambda_i^h$	
	$0.64N$	$N$	$0.64N$	$N$	$0.64N$	$N$
50	6.39	$1.74 \cdot 10^{-3}$	0.56	0.56	33.53	$4.51 \cdot 10^{-8}$
100	6.41	$7.18 \cdot 10^{-30}$	0.59	$1.75 \cdot 10^{-30}$	33.58	$2.91 \cdot 10^{-11}$
200	6.42	$3.80 \cdot 10^{-14}$	0.58	$8.90 \cdot 10^{-87}$	33.59	$2.91 \cdot 10^{-11}$
300	6.42	$9.47 \cdot 10^{-97}$	0.61	$4.41 \cdot 10^{-30}$	33.59	$1.16 \cdot 10^{-10}$
400	6.42	$5.30 \cdot 10^{-15}$	0.58	$7.38 \cdot 10^{-176}$	33.59	$6.98 \cdot 10^{-10}$

 TABLEAU 1.1 – Comportement du gap et des fonctions propres lorsque  $h \rightarrow 0$ 

 FIGURE 1.2 – Comparaison des valeurs propres de l'opérateur  $-\Delta$  et de sa version discrétisée sur maillage uniforme avec  $N = 100$ .

**Proposition** *Sous certaines hypothèses de régularité sur  $\gamma$  et  $q$ , et pour des maillages quasi-uniformes, on a l'estimation spectrale suivante*

$$|\lambda_k^h - \lambda_k| \leq Ch^2 \lambda_k^3, \forall k \in \llbracket 1, N \rrbracket.$$

Ce résultat, valable sur des maillages non uniformes et avec des coefficients variables, ne semblait pas connu dans la littérature. Il fait l'objet d'une note en préparation ([2]).

Dans la seconde partie du chapitre, nous utilisons ces théorèmes spectraux pour montrer des résultats de contrôlabilité sur l'équation discrétisée (1.20). Plus généralement, la méthode des moments nous permet de traiter des *systèmes d'équations paraboliques en cascade* discrétisés sur maillages quasi-uniformes avec des contrôles internes et sur le bord. Dans ces systèmes, la première équation est contrôlée par un contrôle au bord ou interne et les suivantes sont contrôlées par le biais de la solution de l'équation précédente. Il n'y a donc qu'un seul contrôle pour contrôler le système entier. Dans cette introduction, pour simplifier la présentation, nous considérerons ce type de système lorsqu'il n'est composé que de deux équations, il s'écrit alors :

$$\begin{cases} (y_1^h)'(t) + \mathcal{A}^h y_1^h(t) &= 1_\omega^h V_d^h(t) + \mathcal{B}_r^h V_b^h(t), \text{ pour } 0 < t \leq T, \\ (y_2^h)'(t) + \mathcal{A}^h y_2^h(t) + y_1^h(t) &= 0, \text{ pour } 0 < t \leq T, \\ \begin{pmatrix} y_1^h(0) \\ y_2^h(0) \end{pmatrix} &= Y^{0,h} \in (\mathbb{R}^N)^2, \end{cases} \quad (S_2^h)$$

Les méthodes précédentes développées par exemple dans [13], utilisant les techniques de Carleman, ne permettent pas, à notre connaissance, de traiter le cas des systèmes d'équations discrétisées et contrôlées au bord, dès qu'il y a moins de contrôles que d'équations.

Lorsque le maillage est uniforme et que le coefficient de diffusion  $\gamma$  est constant, la méthode des moments permet de montrer que le système  $(S_2^h)$  est contrôlable à zéro au sens de la définition 1.2.1 (voir Théorème 2.5.3).

Cependant, lorsque  $\gamma$  n'est pas constant (ou dans le cas plus général des maillages quasi-uniformes), l'estimation (1.23) du Théorème 1.23 n'est pas suffisante pour démontrer la contrôlabilité à zéro du système  $(S_2^h)$  en appliquant la méthode des moments. En effet, la propriété de gap n'est valable que pour une partie du spectre.

De plus, comme nous le verrons dans la section 1.2.3 la notion de contrôlabilité à zéro n'est pas adaptée pour contrôler des systèmes discrétisés en dimension d'espace  $D \geq 2$ .

Nous renonçons donc à démontrer la contrôlabilité à zéro de l'équation  $(S_2^h)$  pour des maillages non uniformes. La méthode des moments permet cependant d'annuler les  $\alpha N$  premiers modes de la solution  $Y^h = \begin{pmatrix} y_1^h \\ y_2^h \end{pmatrix}$  au temps  $T/2$ . On tire ensuite partie de la dissipation parabolique de la solution après le temps  $T/2$  pour montrer que la norme de la solution  $Y^h$  au temps  $T$  est majorée par la norme de la condition initiale  $Y^{0,h}$  multipliée par une fonction de  $h$ , que l'on notera  $\phi$ , de la forme

$$\phi(h) = \exp(-C/h^\delta), \quad (1.24)$$

où  $\delta$  est une constante positive.

On définit donc la notion d'uniforme  $\phi(h)$  contrôlabilité à zéro pour les systèmes discrétisés tels que  $(S_2^h)$ . Cette notion a été étudiée dans, par exemple, [11, 13, 23].

**Définition 1.2.2 (Uniforme  $\phi(h)$  contrôlabilité à zéro)** Soit  $\phi : (0, +\infty) \rightarrow (0, +\infty)$  une fonction telle que  $\lim_{h \rightarrow 0} \phi(h) = 0$ . Nous dirons que la famille de problèmes  $(S_2^h)$  (indexée par  $h$ ) est uniformément  $\phi(h)$  contrôlable à zéro par un contrôle distribué (resp. un contrôle au bord droit) s'il existe  $C > 0$ , tel que pour tout  $h > 0$  et pour toute donnée initiale  $y^{0,h} \in \mathbb{R}^N$ , il existe un contrôle  $V_d^h \in L^2(0, T, \mathbb{R}^N)$  (resp.  $V_b^h \in L^2(0, T)$ ) qui vérifie les points ci-dessous

1.

$$\int_0^T \|V_d^h(t)\|^2 dt \leq C \|Y^{0,h}\|^2, \left( \text{resp. } \int_0^T |V_b^h(t)|^2 dt \leq C \|Y^{0,h}\|^2 \right), \quad (1.25)$$

2. la solution  $Y^h$  de  $(S_2^h)$  avec  $V_b^h = 0$  (resp.  $V_d^h = 0$ ) vérifie

$$\|Y^h(T)\|^2 \leq C \phi(h) \|Y^{0,h}\|^2.$$

Pour ainsi dire, l'uniforme  $\phi(h)$  contrôlabilité à zéro consiste à contrôler la solution à zéro au temps  $T$  lorsque le pas du maillage tend vers zéro.

Nous démontrons dans le chapitre 2 que  $(S_2^h)$  est uniformément  $\phi(h)$  contrôlable à zéro, avec  $\phi$  de la forme (1.24), lorsque le maillage est uniforme (voir Théorème 2.5.4) ou quasi-uniforme (voir Théorème 2.5.5).

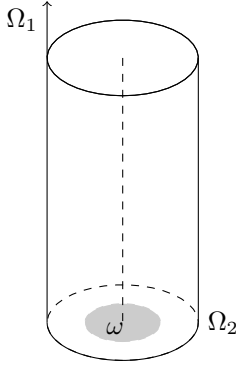
### 1.2.3 Contrôle au bord de systèmes d'équations paraboliques sur une géométrie cylindrique

A présent, nous souhaitons contrôler des systèmes discrétisés de dimension  $D \geq 2$ . Tout d'abord, il a été observé par O. Kavian qu'il n'est pas possible en général de contrôler à zéro l'équation de la chaleur en dimension  $D = 2$  sur un maillage cartésien uniforme lorsque  $\gamma = 1$  et  $q = 0$  (voir l'article de E. Zuazua à ce sujet [32, section 9.2]). En effet, l'auteur exhibe un mode propre de l'opérateur de la chaleur discrétisé localisé sur la diagonale du domaine. De plus, soulignons que, en toute généralité, le champ d'application de la méthode des moments sur des systèmes paraboliques de type (1.17) se limite à la dimension d'espace  $D = 1$ . Effectivement, comme nous l'expliquons dans la section 1.1.1, la condition (1.6) est nécessaire à la construction de famille biorthogonales. Or les valeurs propres  $(\lambda_k)_{k \geq 1}$  de l'opérateur  $\mathcal{A}$  sont asymptotiquement égales à celles de l'opérateur  $-\Delta$  avec conditions aux bords de Dirichlet sur le domaine  $\Omega$ . Cependant, en utilisant l'asymptotique de Weyl, nous savons que

$$\lambda_k \sim_{k \rightarrow +\infty} C_D k^{2/D}.$$

Ainsi, la condition (1.6) n'est plus respectée dès lors que  $D \geq 2$ .





Afin de contourner ces difficultés, nous choisissons donc d'étudier l'uniforme  $\phi(h)$  contrôlabilité à zéro de  $(S_2^h)$ , plus adaptée aux systèmes discrétisés que la contrôlabilité à zéro. Sur des géométries particulières, il est possible d'étendre les résultats du chapitre 2 à des dimensions d'espace supérieures, en s'inspirant de la technique utilisée dans [10]. Le point clé est de considérer un espace  $\Omega$  de forme cylindrique, c'est à dire  $\Omega = \Omega_1 \times \Omega_2$  avec  $\Omega_1$  de dimension 1, et un opérateur elliptique  $\mathcal{A}$  s'écrivant comme la somme de deux opérateurs elliptiques, l'un, noté  $\mathcal{A}_1$ , agissant uniquement sur des fonctions de la variable  $x_1$  et l'autre, noté  $\mathcal{A}_2$ , sur des fonctions de la variable  $x_2$ . Nous pouvons alors combiner la méthode des moments sur  $\Omega_1$  avec une approche de type Lebeau-Robbiano sur  $\Omega_2$  (voir la section 1.3.2 pour plus de détails sur cette méthode). Dans le chapitre 3, nous parvenons ainsi à contrôler des systèmes d'équations paraboliques sur une section du bord de  $\Omega$ , au sens de la définition 1.2.2. Nous décrivons ci-dessous les résultats obtenus.

Le théorème principal que nous montrons dans le chapitre 3 est valable dans un cadre général et s'applique aussi bien sur des systèmes discrétisés que des systèmes continus. Il est énoncé dans la section 3.3.2 (voir Théorème 3.3.1). En substance, il affirme que si une inégalité de type Lebeau-Robbiano (voir [24]) existe sur l'opérateur  $\mathcal{A}_2$  et s'il est possible de contrôler au bord  $x_1 = \{0\}$  le système parabolique défini sur  $\Omega_1$  avec l'opérateur elliptique  $\mathcal{A}_1$ , tout en donnant une estimation précise du coût du contrôle, alors nous pouvons en déduire un résultat de contrôle sur le système complet posé sur  $\Omega$ . Ce résultat découle de la méthode de Lebeau-Robbiano et s'appuie sur l'équivalence entre une inégalité d'observabilité relaxée et des estimations sur le contrôle et la valeur au temps finale de la solution. Les dépendances des constantes intervenant dans ces inégalités par rapport aux données du problème sont précisément mentionnées. Ainsi, lorsque nous appliquerons cette méthode à des systèmes semi-discrétisés en espace, nous choisirons des normes discrètes pour lesquelles ces constantes sont indépendantes du paramètre de discrétisation  $h$ . Ceci permet d'obtenir un résultat d'uniforme  $\phi(h)$  contrôlabilité à zéro sur le système discrétisé posé sur  $\Omega$ . Dans le cadre continu, nous retrouvons le théorème principal de [10]. Nous illustrons ci-dessous le type de résultats obtenus sur un exemple de système semi-discrétisé.

### Contrôle au bord d'un système en cascade semi-discrétisé sur un rectangle

Soit deux ensembles  $\Omega_1 = (0, 1) \subset \mathbb{R}$  et  $\Omega_2 \subset \mathbb{R}^{D-1}$ . Considérons le système en cascade suivant, contrôlé sur le bord de  $\Omega := \Omega_1 \times \Omega_2$ ,

$$\begin{cases} \partial_t \alpha - \partial_{x_1}^2 \alpha - \Delta_2 \alpha = 0, & \text{dans } (0, T) \times \Omega_1 \times \Omega_2, \\ \partial_t \beta - \partial_{x_1}^2 \beta - \Delta_2 \beta + \alpha = 0, & \text{dans } (0, T) \times \Omega_1 \times \Omega_2, \\ \alpha = \beta = 0, & \text{sur } (0, T) \times \Omega_1 \times \partial\Omega_2, \\ \alpha = 1_{\{0\} \times \omega_2} V_b, & \text{sur } (0, T) \times \{0, 1\} \times \Omega_2, \\ \beta = 0, & \text{sur } (0, T) \times \{0, 1\} \times \Omega_2. \end{cases} \quad (1.26)$$

La composante  $\alpha$  est contrôlée au bord par le contrôle  $V_b$  alors que la seconde composante  $\beta$  est contrôlée par l'action de  $\alpha$  via le terme de couplage dans la seconde équation du système. Le théorème principal du chapitre 3 (Théorème 3.3.1) permet de montrer l'uniforme  $\phi(h)$  contrôlabilité à zéro au bord de l'équivalent semi-discrét de (1.26). Nous avons déjà rencontré ce type de système dans la section 1.2.2 (voir  $(S_2^h)$ ).

Plus précisément, nous discrétisons les opérateurs  $\partial_{x_1}^2$  et  $\Delta_2$  par la méthode des différences finies sur un maillage uniforme du domaine  $\Omega$ . Pour plus de clarté, nous supposons que  $D = 2$ , c'est-à-dire que dans cette configuration  $\Omega$  est un rectangle, et nous traçons le maillage sur la figure 1.3. Lorsque le domaine  $\Omega$  est carré avec  $h_1 = h_2$ , nous obtenons le système suivant d'inconnues  $(\alpha_{i,j})_{1 \leq i, j \leq N}$  et  $(\beta_{i,j})_{1 \leq i, j \leq N}$ ,

$$\begin{cases} \partial_t \alpha_{i,j} + \frac{4\alpha_{i,j} - \alpha_{i-1,j} - \alpha_{i+1,j} - \alpha_{i,j-1} - \alpha_{i,j+1}}{h^2} = 0, & \forall 1 \leq i, j \leq N, \\ \partial_t \beta_{i,j} + \frac{4\beta_{i,j} - \beta_{i-1,j} - \beta_{i+1,j} - \beta_{i,j-1} - \beta_{i,j+1}}{h^2} + \alpha_{i,j} = 0, & \forall 1 \leq i, j \leq N, \\ \alpha_{i,0} = \beta_{i,0} = \alpha_{i,N+1} = \beta_{i,N+1} = 0, & \forall 1 \leq i \leq N, \\ \alpha_{0,j} = v_j 1_{\omega_2}(jh), & \forall 1 \leq j \leq N, \\ \alpha_{N+1,j} = \beta_{0,j} = \beta_{N+1,j} = 0, & \forall 1 \leq j \leq N. \end{cases} \quad (1.27)$$

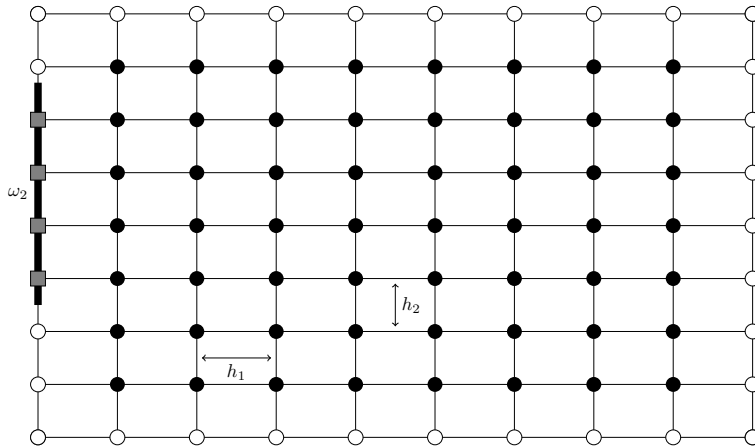


FIGURE 1.3 – Maillage cartésien uniforme sur un rectangle

En combinant les résultats du chapitre 2, introduits dans la section 1.2.2 précédente, avec les travaux de [12], nous parvenons à montrer que le système (1.27) est uniformément  $\phi(h)$  contrôlable à zéro avec, comme précédemment, une fonction  $\phi$  de la forme (1.24). Cette méthode, qui consiste à coupler la méthode des moments et la procédure de Lebeau-Robbiano, est mise en oeuvre dans le cas continu (c'est-à-dire pour un système d'équations couplées tel que (1.26)) dans l'article [10]. Notre résultat de contrôle sur le système discrétisé (1.27) s'inspire donc largement des techniques de [10]. Cependant, dans le chapitre 3, nous construisons un cadre plus général qui permet, à la fois, de montrer un résultat de contrôle sur le système discret (1.27) et de retrouver le résultat de contrôle à zéro de [10].

D'une part, ce résultat d'uniforme  $\phi(h)$  contrôlabilité à zéro permet de contourner l'obstacle posé par le contre-exemple de Kavian énoncé au début de cette section, qui affirme qu'en toute généralité, le système (1.27) n'est pas uniformément contrôlable à zéro. D'autre part, il montre que, combinée à la procédure de Lebeau-Robbiano, la méthode des moments peut en fait contribuer à montrer des résultats de contrôlabilité de systèmes posés sur un domaine  $\Omega$  de dimension strictement supérieure à 1. Ceci nuance donc les propos du premier paragraphe de cette section.

### 1.2.4 Contrôle à zéro de l'équation de Grushin

Le Théorème 1.1.2 affirme que l'équation de la chaleur est contrôlable à zéro en tout temps  $T > 0$ . Cependant il existe des systèmes paraboliques qui ne sont contrôlables à zéro qu'à partir d'un certain temps  $T^* > 0$  que l'on appelle *temps minimal* de contrôle à zéro (voir par exemple [4, 5, 15]). Dans cette thèse, nous étudions une équation parabolique scalaire et dégénérée dans laquelle ce phénomène apparaît et nous cherchons à préciser la valeur du temps minimal  $T^*$ .

Nous nous intéressons au problème de contrôle de l'équation de Grushin suivant :

$$\begin{cases} \partial_t f(t) + \mathcal{L}_\gamma f(t) = 1_\omega u(t), & \text{dans } Q \\ f(t) = 0, & \text{dans } (0, T) \times \partial\Omega \\ f(0) = f^0, & \text{dans } \Omega. \end{cases} \quad (1.28)$$

- L'opérateur  $\mathcal{L}_\gamma$  est un opérateur elliptique :  $\mathcal{L}_\gamma := -\partial_{xx} - |x|^{2\gamma}\partial_{yy}$ , et  $\gamma > 0$  est un paramètre réel.
- L'espace physique  $\Omega$  est un sous ensemble de  $\mathbb{R}^2$  défini par  $\Omega := (-1, 1) \times (0, 1)$  et  $\omega$  est un ouvert de  $\Omega$ .
- On note  $Q := (0, T) \times \Omega$ .
- le contrôle  $u$  appartient à l'espace  $L^2(Q)$ .
- Le temps final  $T > 0$  est une donnée du problème.
- La donnée initiale vérifie  $f^0 \in L^2(\Omega)$ .

L'équation (1.28) est-elle *contrôlable à zéro au temps*  $T$ ? Sous certaines conditions sur  $\gamma$  et sur la géométrie de l'ouvert de contrôle, cette équation possède un temps minimal de contrôle à zéro, c'est-à-dire qu'il peut exister un temps final  $T^* > 0$  tel que

- lorsque  $T < T^*$ , le système (1.28) n'est pas contrôlable,
- lorsque  $T > T^*$ , le système (1.28) est contrôlable.

Dans la configuration suivante :

$$\gamma = 1 \text{ et } \omega = (a, b) \times (0, 1) \text{ où } 0 < a < b \leq 1, \quad (1.29)$$

un temps minimal  $T^*$  non nul existe (voir [7]). On s'intéressera donc particulièrement à l'hypothèse (1.29) par la suite.

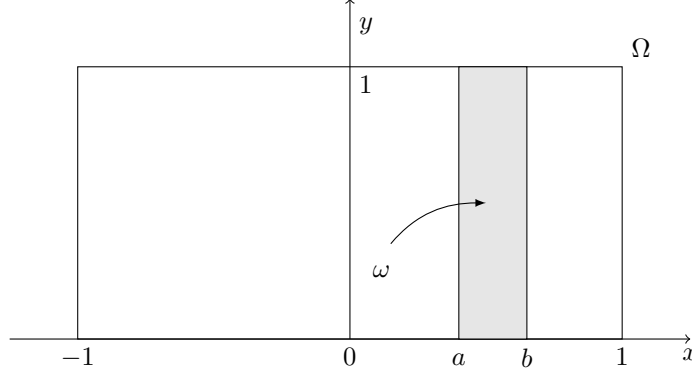


FIGURE 1.4 – Le domaine  $\Omega$  ainsi que la zone de contrôle  $\omega$  de l'équation de Grushin contrôlée sur une bande verticale.

Dans le chapitre 4, nous souhaitons démontrer par une méthode des moments la propriété suivante :

Sous les hypothèses (1.29), l'équation de Grushin admet pour temps minimal :

$$T^* = \frac{a^2}{2}. \quad (1.30)$$

**Remarque 1.2.1** *A ce stade nous n'avons pas réussi à conclure la preuve de cette propriété. En effet, l'inégalité de gap uniforme, nécessaire pour appliquer la méthode des moments, est établie, pour l'instant, sur l'ensemble du spectre, excepté au voisinage d'une valeur critique bien déterminée (voir le théorème de cette section). Des tests numériques semblent confirmer qu'il ne s'agit que d'un obstacle technique et que cette estimée spectrale est bien valable sur la totalité du spectre, ce qui permettrait de conclure la preuve de (1.30).*

*Dernièrement, ce problème a été résolu dans un article à paraître (voir [8]), par une approche totalement différente. Encore plus récemment, dans l'article soumis [17], les auteurs obtiennent des résultats de temps minimal de contrôle à zéro sur des ouverts  $\omega$  bien plus généraux, en se basant sur [22]. Malgré cela, il serait intéressant de compléter notre travail, car une fois terminé, celui-ci présentera l'avantage de pouvoir s'appliquer à des systèmes d'équations de Grushin en cascade.*

Pour tout entier non nul  $n$ , notons  $\Lambda_n := (\lambda_{j,n})_{j \geq 1}$  et  $(g_{j,n})_{j \geq 1}$  les valeurs propres et fonctions propres de l'opérateur elliptique  $\mathcal{A}_n := -\partial_{xx} + n^2 \pi^2 x^2$ , avec condition aux limites de Dirichlet, défini sur  $\Omega^1 := (-1, 1)$

$$\begin{cases} \mathcal{A}_n g_{j,n} = \lambda_{j,n} g_{j,n}, & \text{dans } \Omega^1, \\ g_{j,n} = 0, & \text{sur } \partial\Omega^1 = \{-1, 1\}, \\ \|g_{j,n}\|_{L^2(\Omega^1)} = 1. \end{cases} \quad (1.31)$$

La proposition suivante (voir [9]) indique comment démontrer la propriété (1.30)

**Proposition 1.2.1** *Si pour tout  $T > \frac{a^2}{2}$  il existe une constante  $K > 0$  telle que pour tout  $n \in \mathbb{N}^*$  et toute donnée initiale  $g_n^0 \in L^2(\Omega^1)$ , l'équation*

$$\begin{cases} \partial_t g_n + \mathcal{A}_n g_n = u_n(t, x) 1_{(a,b)}(x), & (t, x) \in (0, T) \times \Omega^1, \\ g_n(t, \pm 1) = 0, & t \in (0, T), \\ g_n(0, x) = g_n^0(x), & x \in \Omega^1. \end{cases} \quad (1.32)$$

est contrôlable à zéro avec un contrôle  $u_n$  vérifiant :

$$\|u_n\|_{L^2((0,T)\times\Omega^1)} \leq K \|g_n^0\|_{L^2(\Omega^1)}, \quad (1.33)$$

alors le temps minimal de l'équation de Grushin  $T^*$  est égal à  $\frac{a^2}{2}$ .

Ce résultat nous ramène au même problème de contrôle uniforme (cette fois-ci par rapport au paramètre  $n$ ) que celui de la section 1.2.2. Nous l'abordons donc de la même façon avec la méthode des moments. Ainsi, sous réserve d'existence de la famille biorthogonale  $(q_j^{\Lambda_n})_{j \geq 1}$  à l'ensemble des fonctions  $(t \mapsto e^{-\lambda_{j,n}(T-t)})_{j \geq 1}$  et de la convergence de la série, le contrôle  $u_n$  s'exprime sous la seconde forme de la façon qui suit :

$$u_n(t, x) := - \sum_{k \geq 1} \left( e^{-\lambda_{k,n}T} \frac{\int_{\Omega^1} g_n^0 g_{k,n}}{\int_a^b g_{k,n}^2} \right) \times (1_{(a,b)}(x) g_{k,n}(x)) \times q_k^{\Lambda_n}(t). \quad (1.34)$$

Comme précédemment, le problème de contrôle se réduit maintenant d'une part à prouver l'existence et à majorer convenablement les familles biorthogonales  $(q_j^{\Lambda_n})_{j \geq 1}$  et d'autre part à minorer le dénominateur  $\int_a^b g_{k,n}^2$ . Nous suivons la stratégie en deux points proposée tout à la fin de la section 1.1.2.

Rappelons que d'après le premier point de la Remarque 1.1.4 la majoration (1.13) de  $(q_j^{\Lambda_n})_{j \geq 1}$ , avec une constante  $K_\varepsilon$  indépendante de  $n$ , peut-être obtenue, notamment, grâce à une minoration uniforme du gap  $\lambda_{k+1,n} - \lambda_{k,n}$ . Dans le Chapitre 4, nous démontrons cette minoration à l'exception d'une fine bande d'indices.

**Théorème** *Pour tout réel  $\tau \in (0, 1)$  il existe  $\rho > 0$  tel que pour tout entier  $n \in \mathbb{N}^*$  et tout entier  $k$  tels que*

$$1 \leq k < \left\lfloor \tau \frac{\pi}{2} n \right\rfloor \text{ ou } k > \left\lceil \frac{\pi}{2} n \right\rceil,$$

*on a la minoration*

$$\lambda_{k+1,n} - \lambda_{k,n} \geq \rho. \quad (1.35)$$

Vraisemblablement, l'obstacle posé par la zone  $\left\lfloor \tau \frac{\pi}{2} n \right\rfloor \leq k \leq \left\lceil \frac{\pi}{2} n \right\rceil$ , avec  $\tau$  arbitrairement proche de 1, est purement technique comme l'illustrent les calculs numériques de la figure 1.5 ci-dessous. Ceci constitue un obstacle majeur et nous empêche de démontrer la propriété (1.30). Les autres propriétés requises pour faire fonctionner la méthode des moments sont cependant démontrées. En effet, l'existence de la fonction  $\mathcal{N}$  au sens de la définition 1.1.2 ne pose pas de difficultés majeures, et est prouvée dans la section 4.4.3. Enfin, dans la section 4.4.4, nous montrons une minoration de la forme suivante

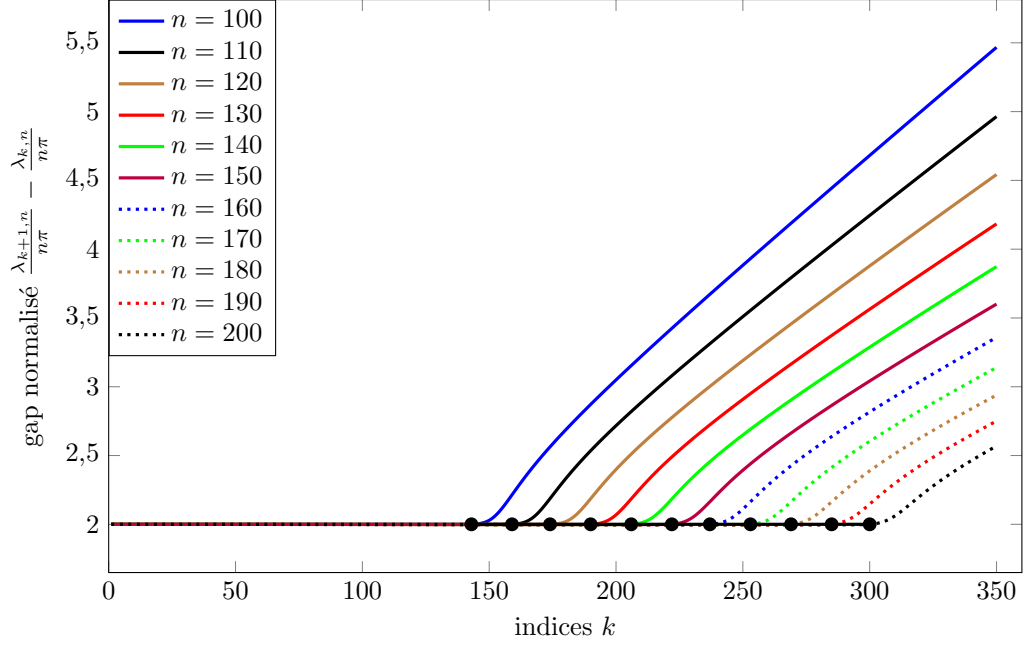
$$\int_a^1 g_{k,n}^2(x) dx \geq C \frac{e^{-n\pi a^2}}{n^2 \pi^2}.$$

Ceci suffirait à démontrer la Proposition 1.2.1 si la propriété de gap uniforme était vérifiée. En effet, en examinant la forme du contrôle  $u_n$  donnée par l'expression (1.34) et, en utilisant les estimations précédentes, on peut montrer que, pour tout  $\varepsilon > 0$ ,  $u_n$  est majoré par une série de terme général de la forme

$$e^{-\lambda_{k,n}T} \times n\pi e^{n\pi \frac{a^2}{2}} \times e^{\varepsilon \lambda_{k,n}}.$$

De plus, nous pouvons montrer que  $\lambda_{k,n} \geq n\pi$  et  $\lambda_{k,n} \geq k$  donc la convergence de cette série est assurée pour  $T > \frac{a^2}{2}$ , avec un coût du contrôle  $u_n$  indépendant de  $n$ .

En conclusion, il ne resterait plus qu'à démontrer la propriété de gap uniforme dans la bande  $\left\lfloor \tau \frac{\pi}{2} n \right\rfloor \leq k \leq \left\lceil \frac{\pi}{2} n \right\rceil$  pour obtenir la valeur du temps minimal  $T^*$ , y compris pour des systèmes en cascade d'équations de Grushin.


 FIGURE 1.5 – Gap normalisé  $\frac{\lambda_{k+1,n}}{n\pi} - \frac{\lambda_{k,n}}{n\pi}$  en fonction de l'indice  $k$  pour différents indices de potentiel.

### 1.3 Idées générales des preuves

#### 1.3.1 Étude spectrale d'opérateurs de Sturm-Liouville

Dans cette section, nous présentons la méthode du chapitre 2 pour obtenir des estimées sur le gap et sur les fonctions propres d'un opérateur continu  $\mathcal{A}$  ou discrétisé  $\mathcal{A}^h$ . Comme expliqué dans la section 1.2.2, celles-ci nous permettent de montrer ensuite, en particulier, des résultats de contrôlabilité au bord de systèmes d'équations en cascade comme  $(S_2^h)$ , par la méthode des moments.

Par souci de clarté, nous commençons par l'expliquer dans le cas continu, nous rappelons que  $\Omega = (0, 1)$ . L'idée consiste dans un premier temps à introduire un terme source  $f$  dans le problème aux valeurs propres et à écrire cette égalité comme une équation différentielle ordinaire vectorielle d'ordre 1.

Soit  $\lambda \in \mathbb{R}$  et  $u \in \mathcal{C}^2(\Omega)$  tels que

$$\mathcal{A}u(x) = \lambda u(x) + f(x), \forall x \in \Omega. \quad (1.36)$$

Posons  $U(x) = \begin{pmatrix} u(x) \\ \sqrt{\frac{\gamma(x)}{\lambda}} u'(x) \end{pmatrix}$ . L'équation (1.36) s'écrit

$$U'(x) = M(x)U(x) + Q(x)U(x) + f(x)F(x), \quad (1.37)$$

avec  $M(x) = \begin{pmatrix} 0 & \sqrt{\frac{\lambda}{\gamma(x)}} \\ -\sqrt{\frac{\lambda}{\gamma(x)}} & 0 \end{pmatrix}$ , la matrice  $Q(x)$  et le vecteur  $\sqrt{\lambda}F(x)$  sont uniformément bornés par rapport à  $\lambda$  et  $x$ . L'avantage de ce changement de variable est que pour tout  $x \in \Omega$ , la matrice  $M(x)$  est anti symétrique et commute avec  $M(y)$  pour tout  $y \in \Omega$ , rendant le semi groupe

$$S(y, x) := \exp\left(\int_x^y M(s)ds\right)$$

unitaire pour la norme sur  $M_2(\mathbb{R})$  :  $(\text{Tr}(M^t M))^{1/2}$ . En utilisant ensuite l'expression de la solution  $U(x)$  en fonction de  $U(y)$  et du semi groupe  $S(y, x)$  ainsi que le Lemme de Gronwall, nous montrons qu'il existe  $C > 0$  indépendant de  $x$  et de  $\lambda$  tels que

$$\forall x, y \in \Omega, \|U(y)\| \leq C \left( \|U(x)\| + \left| \int_x^y |f(s)| ds \right| \right) \quad (1.38)$$

Lorsque  $f = 0$  et  $u(x) = \phi_k(x)$ , nous avons  $\lambda = \lambda_k$ . L'estimation (1.38) devient

$$\forall x, y \in \Omega, \frac{\|U(x)\|^2}{\|U(y)\|^2} = \frac{\phi_k^2(x) + \frac{\gamma(x)}{\lambda_k} (\phi_k'(x))^2}{\phi_k^2(y) + \frac{\gamma(y)}{\lambda_k} (\phi_k'(y))^2} \geq \frac{1}{C},$$

et grâce au Lemme 2.2.1, nous obtenons les analogues continus des inégalités (1.21) et (1.22) pour l'ensemble des indices  $k \in \mathbb{N}^*$ . La propriété de gap s'obtient en posant  $f(x) = \phi_{k+1}'(1)\phi_k(x)(\lambda_{k+1} - \lambda_k)$ . La fonction  $u$  vaut alors  $u(x) = \phi_k'(1)\phi_{k+1}(x) - \phi_{k+1}'(1)\phi_k(x)$  et on remarque alors que  $U(1) = 0$ . L'inégalité (1.38) évaluée en  $x = 1$  fournit la propriété de gap annoncée.

Dans le chapitre 2, nous appliquons la même méthode sur l'équation discrétisée (1.20) et nous montrons une version discrète du Lemme 2.2.1 (voir le Lemme 2.3.1). La différence réside dans l'estimation du semi groupe discret. En effet, la version discrète de la matrice  $M(x)$ , notée  $(M_j^h)_{j=1}^N$ , n'est plus anti symétrique et vaut maintenant, sur un maillage uniforme,

$$M_j^h = \begin{pmatrix} -h \frac{\lambda}{\gamma_{j+1/2}} & \sqrt{\frac{\lambda}{\gamma_{j+1/2}}} \\ -\sqrt{\frac{\lambda}{\gamma_{j+1/2}}} & 0 \end{pmatrix}.$$

L'équation (1.38) s'écrit maintenant

$$U_{j+1} = (I^h + hM_j^h)U_j + hQ_j^h U_j + hf_j^h F_j^h.$$

Ainsi, le semi groupe  $S(y, x)$  est maintenant remplacé par la matrice

$$S_{i \leftarrow j}^h = (I + hM_{i-1}^h)(I + hM_{i-2}^h) \dots (I + hM_j^h), \text{ pour } i > j$$

non unitaire. Il s'agit donc d'évaluer la norme de ce semi groupe discret. Pour cela, l'idée principale est d'étudier la suite  $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$  définie par  $\begin{pmatrix} x_{i+1} \\ y_{i+1} \end{pmatrix} = (I^h + hM_i^h) \begin{pmatrix} x_i \\ y_i \end{pmatrix}$  et d'introduire la quantité  $H_i := x_i^2 + y_i^2 - h\sqrt{\frac{\lambda}{\gamma_{i+1/2}}}x_i y_i$ .

Supposons que la fonction  $\gamma$  est constante, la suite  $H_i$  est alors constante. Il existe  $C > 0$  ne dépendant que de  $\gamma$  et  $q$  telle que

$$\left(1 - \frac{h}{2}\sqrt{\frac{\lambda}{\gamma}}\right) (x_i^2 + y_i^2) \leq H_i = H_j \leq C (x_j^2 + y_j^2), \quad (1.39)$$

c'est-à-dire

$$\left(1 - \frac{h}{2}\sqrt{\frac{\lambda}{\gamma}}\right) \left\| S_{i \leftarrow j}^h \begin{pmatrix} x_j \\ y_j \end{pmatrix} \right\|^2 \leq C \left\| \begin{pmatrix} x_j \\ y_j \end{pmatrix} \right\|^2, \quad (1.40)$$

d'où  $\|S_{i \leftarrow j}^h\| \leq \frac{C}{\sqrt{1 - \frac{h}{2}\sqrt{\frac{\lambda}{\gamma}}}}$  et nous concluons comme dans le cas continu, en appliquant le Lemme

2.3.1. Nous obtenons alors le Théorème 2.3.3. Lorsque le maillage est uniforme mais que  $\gamma$  n'est pas une fonction constante, la suite  $H_i$  n'est plus constante mais vérifie une inégalité semblable à (1.40). Nous introduisons alors l'indice

$$k_{max}^h := \max \left\{ k \in \llbracket 1, N \rrbracket : \lambda_k^h < \frac{4}{h^2} \gamma_{\min} \right\},$$

afin de garantir que le membre de gauche de (1.40) soit minoré par une valeur strictement positive<sup>1</sup>. Ceci permet d'obtenir les inégalités (1.21) et (1.22) (voir Théorème 2.3.2). L'optimalité de cet indice est discutée dans la section 1.2.2.

Enfin, lorsque le maillage est quasi-uniforme, nous majorons grossièrement chaque matrice comme suit

$$\|I^h + hM_j^h\| \leq \exp\left(C h \sqrt{\lambda}\right),$$

ce qui suffit à obtenir les minoration du Théorème 2.3.1 sur les vecteurs propres discrets. Comme expliqué dans la section 1.2.2, rappelons que la propriété de gap est obtenue en comparant le spectre de l'opérateur discret  $\mathcal{A}^h$  à celui de l'opérateur continu  $\mathcal{A}$  pour  $k \in \llbracket 1, N^{2/5} \rrbracket$ .

1. en pratique, nous choisirons plutôt  $k_{max}^h$  dépendant d'un paramètre  $\varepsilon > 0$  (que l'on notera donc  $k_{max, \varepsilon}^h$ ) de sorte à ce que (1.40) soit minoré uniformément par le paramètre  $\varepsilon$ .

### 1.3.2 Contrôle à zéro au bord de systèmes paraboliques discrétisés sur un domaine $\Omega$ cylindrique

Dans cette section on présente les idées de la preuve du théorème principal du chapitre 3. Celles-ci sont largement tirées de l'article [10], nous les adaptions dans un cadre plus général et en déduisons des résultats de  $\phi(h)$  contrôlabilité sur des systèmes discrétisés. Le système général considéré est constitué de  $n$  équations couplées avec  $m$  contrôles. Le domaine est  $\Omega := \Omega_1 \times \Omega_2$  et il s'écrit

$$\begin{cases} \partial_t y + \mathcal{A}_1 y + \mathcal{A}_2 y + \mathbf{C}y = 0 \text{ sur } (0, T) \times \Omega, \\ y(0) = y_0 \text{ sur } \Omega, \\ y(t) = \mathcal{B}V_b(t) \text{ sur } \partial\Omega. \end{cases} \quad (1.41)$$

La matrice  $\mathbf{C} \in \mathbb{R}^{n \times n}$  est la matrice de couplage du système. L'opérateur  $\mathcal{B}$  est égal à  $\mathbf{B}1_{\{0\} \times \omega_2}$ , où  $\mathbf{B} \in \mathbb{R}^{n \times m}$  est la matrice de contrôle répartissant les  $m$  contrôles sur les  $n$  équations, et l'opérateur de Sturm-Liouville  $\mathcal{A}_1$  (resp.  $\mathcal{A}_2$ ) n'agit que sur des fonctions de la variable  $x_1$  (resp.  $x_2$ ) de l'ensemble  $\Omega_1$  (resp.  $\Omega_2$ ). Pour tout temps  $t$ , la solution  $y(t)$  évoluera dans un espace de dimension finie, noté  $E$ . Ainsi, le système (1.41) se décline en plusieurs équations possibles : il donne un système discrétisé sur un maillage de  $\Omega$  tel que (1.27) lorsque  $\mathcal{A}_1$  et  $\mathcal{A}_2$  sont des discrétisations d'opérateurs de Sturm-Liouville et on obtient le système (1.26) lorsque les opérateurs  $\mathcal{A}_1$  et  $\mathcal{A}_2$  sont des opérateurs différentiels et que  $y(t)$  appartient à un espace engendré par un nombre fini de produits de fonctions propres de  $\mathcal{A}_1$  et  $\mathcal{A}_2$ .

On fait les hypothèses suivantes.

1. La première hypothèse concerne le système suivant

$$\begin{cases} \partial_t y_1 + \mathcal{A}_1 y_1 + \mathbf{C}y_1 = 0 \text{ sur } (0, T) \times \Omega_1, \\ y_1(0) = y_{0,1} \text{ sur } \Omega_1, \\ y_1(t) = 1_{\{0\}} \mathcal{B}V_{1,b}(t) \text{ sur } \partial\Omega_1. \end{cases} \quad (1.42)$$

On suppose qu'il existe  $\mu_1 \in (0, +\infty]$  tel que pour tout  $y_{0,1}$ , il existe un contrôle  $V_{b,1} \in L^2(0, T)$  tel que

$$\begin{cases} \|V_{b,1}\| \leq e^{C/T} \|y_{0,1}\|_1 \\ \|y_1(T)\|_1 \leq e^{C/T} e^{-\mu_1 T} \|y_{0,1}\|_1, \end{cases} \quad (1.43)$$

où  $\|\cdot\|_0$  et  $\|\cdot\|_1$  sont des normes ad hoc sur les fonctions de  $\Omega_1$ .

2. La seconde hypothèse est une inégalité de Lebeau-Robbiano sur l'opérateur  $\mathcal{A}_2$  dont les valeurs propres et fonctions propres sont notées  $(\lambda_{k,2}, \phi_{k,2})$ . On suppose qu'il existe  $\mu_2 \in (0, +\infty]$  et  $M_{LR} > 0$  tels que pour tout  $\mu < \mu_2$ ,

$$\|\phi\|_2 \leq e^{M_{LR}\sqrt{\mu}} \|1_{\omega_2} \phi\|_2, \quad \forall \phi \in \text{span}(\phi_{k,2} : \lambda_{k,2} \leq \mu),$$

où  $\|\cdot\|_2$  est une norme sur les fonctions de  $\Omega_2$ .

Sur l'espace  $\Omega$  la norme des fonctions est notée  $\|\cdot\|$  et dépend des normes  $\|\cdot\|_1$  et  $\|\cdot\|_2$ . Celles-ci sont choisies de sorte à ce que les constantes  $M_{cont}$  et  $M_{adm}$  dans l'inégalité suivante

$$\sup_{t \in [0, T]} \|y(t)\| \leq M_{cont} \|y_0\| + M_{adm} \|V_b^h(t)\|_{L^2(0, T)}, \quad (1.44)$$

ne dépendent pas de certains paramètres du problème (par exemple du pas du maillage lorsque  $\mathcal{A}_1$  et  $\mathcal{A}_2$  sont des opérateurs discrets).

Nous allons exposer ici les idées permettant d'aboutir à la conclusion suivante. Soit  $\mu^* = \min(\mu_1, \mu_2)$ , Pour toute donnée initiale  $y_0$ , il existe un contrôle  $V_b \in L^2((0, T) \times \Omega_2)$  tel que celui-ci et la solution  $y$  du système (1.41) vérifient

$$\begin{cases} \|V_b(t)\|_{L^2(0, T)} \leq e^{C/T} \|y_0\|, \\ \|y(T)\| \leq e^{C/T} e^{-\mu^* T/4} \|y_0\|. \end{cases} \quad (1.45)$$

Le point de départ de la preuve repose sur l'équivalence, pour tout sous espace  $F \subset E$  fixé, entre l'inégalité d'observabilité relaxée 1. et les estimations 2. ci-dessous.

1. Pour tout  $q_T \in F$ , la solution  $q$  de l'équation adjointe de (1.41) avec condition finale  $q_T$  vérifie :

$$\|q(0)\|^2 \leq M_{obs}^2 \|[\mathcal{B}^* q(t)]\|_{L^2(0, T)}^2 + M_{rel}^2 \|q_T\|^2.$$

2. Pour tout  $y_0$ , il existe un contrôle  $V_b$  tel que

$$\frac{1}{M_{obs}^2} \llbracket V_b \rrbracket_{L^2(0,T)}^2 + \frac{1}{M_{rel}^2} \|P_F(y(T))\|^2 \leq \|y_0\|^2,$$

où  $P_F$  est la projection orthogonale sur  $F$ .

Dans un premier temps, nous utilisons la structure tensorisée du système pour étendre l'hypothèse 1 à des fonctions définies sur l'espace  $\Omega$  tout entier (voir Proposition 3.3.2). Ceci s'obtient en traduisant d'abord l'hypothèse 1 en terme d'inégalité d'observabilité relaxée pour le système (1.42). On applique ensuite l'hypothèse 2 avec des fonctions de l'espace  $F_\mu$  engendré par les premières fonctions propres de l'opérateur  $\mathcal{A}_2$  de fréquences inférieures à  $\mu$  (avec  $\mu \leq \mu_2$ ). En utilisant de nouveau l'équivalence entre les points 1. et 2. ci-dessus, nous obtenons

$$\begin{cases} \llbracket V_b \rrbracket_{L^2(0,T)} \leq e^{C/T} e^{M_{LR}\sqrt{\mu}} \|y_0\| \\ \|P_{F_\mu} y(T)\| \leq e^{C/T} e^{-\mu_1 T} \|y_0\|. \end{cases} \quad (1.46)$$

Soit  $\tau > 0$ . La construction de Lebeau-Robbiano (voir [24]) consiste à contrôler dans un premier temps une certaine plage de fréquences  $\mu$  pendant un temps  $\tau/2$ , puis à laisser le système dissiper pendant une durée  $\tau/2$ . Ce faisant, avec un bon choix de  $\mu$  et de  $\tau$ , la norme de la solution décroît. Illustrons ceci sur l'intervalle de temps  $(0, \tau)$ . Sur  $(0, \frac{\tau}{2})$ , nous contrôlons l'équation comme précédemment, puis sur  $(\frac{\tau}{2}, \tau)$ , nous n'appliquons aucun contrôle. Celui-ci vérifie, d'après l'inégalité précédente,

$$\|V_b\|_{L^2(0,\tau)} = \|V_b\|_{L^2(0,\tau/2)} \leq e^{C/\tau} e^{M_{LR}\sqrt{\mu}} \|y_0\|.$$

Nous décomposons ensuite  $\|y(\tau)\|^2$  de la façon suivante :

$$\|y(\tau)\|^2 = \|P_{F_\mu} y(\tau)\|^2 + \|P_{F_\mu}^\perp y(\tau)\|^2.$$

Pour majorer le terme  $P_{F_\mu}^\perp y(\tau)$ , on utilise d'abord la dissipation parabolique de l'équation entre  $\tau/2$  et  $\tau$  (lorsque le contrôle est nul), puis les inégalités (1.44) et (1.46). Contrairement à la méthode classique, le terme  $P_{F_\mu} y(\tau)$  n'est pas annulé mais contrôlé par  $e^{-\mu_1 T}$  grâce à (1.46) et à (1.44). Nous choisissons ensuite  $\mu \leq \mu^*$ , il vient

$$\begin{cases} \llbracket V_b \rrbracket_{L^2(0,T)} \leq e^{C/\tau} e^{M_{LR}\sqrt{\mu}} \|y_0\|, \\ \|y(\tau)\| \leq e^{C/\tau} e^{-C\mu\tau + M_{LR}\sqrt{\mu}} \|y_0\|. \end{cases} \quad (1.47)$$

Nous répétons ce processus autant que possible, c'est-à-dire en veillant toujours à avoir  $\mu < \mu^*$  afin de garantir, en particulier, la validité de l'hypothèse 2. A partir de (1.47), un découpage judicieux de  $(0, T)$  et du spectre permet de conclure avec, de surcroît, des estimées précises des constantes.

Nous rappelons que (voir section 1.2.3), ce théorème s'applique à des systèmes d'équations discrétisées en cascade dont la dimension du domaine  $\Omega$  peut être quelconque, pourvu que ce système respecte la structure tensorisée de (1.41). Rappelons également que ce théorème permet aussi de retrouver le résultat principal de [10].

### 1.3.3 L'équation de Grushin

Cette section est divisée en deux parties. D'abord nous donnons un aperçu de la preuve qui permet d'obtenir la propriété de gap uniforme sur une partie du spectre (voir le théorème de la section 1.2.4). Ensuite, nous expliquons par quel moyen nous obtenons la minoration de la norme  $L^2$  des fonctions propres sur un sous intervalle de  $(-1, 1)$ .

Le théorème de la section 1.2.4 se montre en trois étapes. Soit  $\tau \in (0, 1)$  et définissons l'indice critique  $n_c(\tau) := \tau \frac{\pi}{2} n$ . D'abord nous établissons l'inégalité suivante, valable sur l'ensemble du spectre

$$\lambda_{k,n} \geq n\pi(2k-1), \quad \forall k, n \in \mathbb{N}^*.$$

Ensuite nous démontrons la propriété de gap pour la partie basse du spectre, c'est-à-dire, pour les indices  $k \leq n_c(\tau)$ . Enfin, nous considérons les indices de la partie haute du spectre :  $k \geq n_c(1)$ .



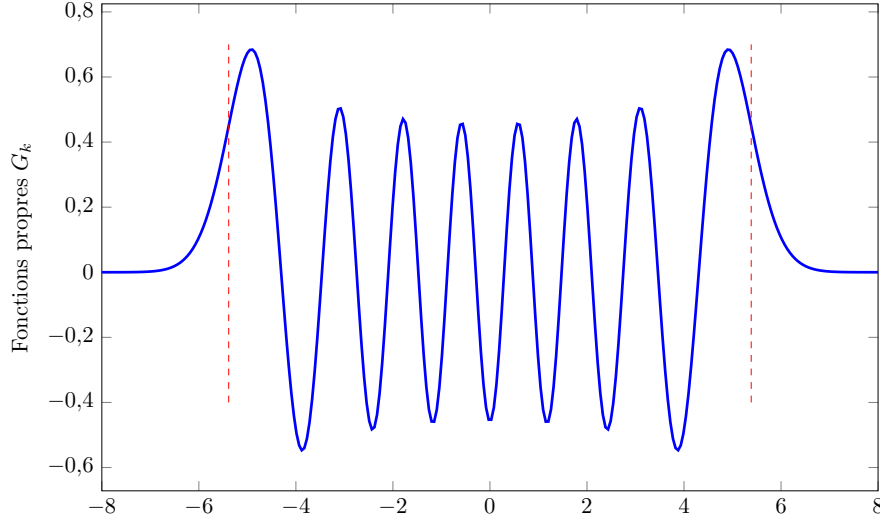


FIGURE 1.6 – Graphique représentant la fonction propre  $G_k$  de l'oscillateur harmonique. Les pointillés verticaux sont placés aux abscisses  $x = \pm\sqrt{\mu_k}$ , pour  $k = 15$ .

Cette distinction provient du fait que les valeurs propres  $\lambda_{k,n}$  ont un comportement assez différent avant et après l'indice critique. Effectivement, comme en témoigne la figure 1.5, le gap, dit normalisé,  $\frac{\lambda_{k+1,n}}{n\pi} - \frac{\lambda_{k,n}}{n\pi}$  semble être environ constant (égal à 2) avant l'indice critique et semble avoir un comportement linéaire après cet indice. En effet, remarquons que l'équation (1.31) peut-être réécrite, à l'aide du changement de variable  $v_{k,n}(x) := \frac{g_{k,n}(\frac{x}{\sqrt{n\pi}})}{(n\pi)^{1/4}}$ , de la façon suivante :

$$\begin{cases} -\partial_{xx}v_{k,n} + x^2v_{k,n} = \frac{\lambda_{k,n}}{n\pi}v_{k,n}, & \text{dans } (-\sqrt{n\pi}, \sqrt{n\pi}), \\ v_{k,n} = 0, & \text{sur } \{-\sqrt{n\pi}, \sqrt{n\pi}\}, \\ \|v_{k,n}\|_{L^2(-\sqrt{n\pi}, \sqrt{n\pi})} = 1. \end{cases} \quad (1.48)$$

Le gap normalisé n'est rien d'autre que le gap des valeurs propres du problème (1.48). Il paraît naturel de vouloir comparer ce problème au problème de l'oscillateur harmonique :

$$\begin{cases} -\partial_{xx}G_k + x^2G_k = \mu_k G_k, & \text{dans } \mathbb{R}, \\ \|G_k\|_{L^2(\mathbb{R})} = 1, \end{cases} \quad (1.49)$$

pour lequel le spectre est connu :  $\mu_k = 2k - 1, \forall k \in \mathbb{N}^*$ . Les fonctions propres  $G_k$  de l'oscillateur harmonique ont un comportement pseudo périodique dans l'intervalle  $(-\sqrt{\mu_k}, \sqrt{\mu_k})$  et décroissent exponentiellement en dehors de cet intervalle (voir figure 1.6 ci-dessous). Ceci se comprend intuitivement en regardant l'équation (1.49). Lorsque  $x^2$  est petit devant  $\mu_k$ , l'équation est proche de  $-\partial_{xx}G_k = \mu_k G_k$ , d'où le comportement périodique. Lorsque  $x^2$  est grand devant  $\mu_k$ , l'équation ressemble à  $-\partial_{xx}G_k + x^2G_k = 0$ , ce qui laisse penser que la solution aurait un comportement plutôt de type exponentiel (décroissant, pour assurer la condition de normalisation). Ainsi, lorsque  $(-\sqrt{\mu_k}, \sqrt{\mu_k}) \subset (-\sqrt{n\pi}, \sqrt{n\pi})$ , le problème (1.48) semble proche du problème (1.49) au sens où les fonctions  $G_k$  sont exponentiellement petites au bord de l'intervalle  $(-\sqrt{n\pi}, \sqrt{n\pi})$  (donc vérifie 'quasi-sim' le même problème aux valeurs propres que la fonction  $v_{k,n}$ ). En conclusion, cette discussion très informelle nous permet de deviner pourquoi lorsque  $\mu_k \leq n\pi$ , ou, de façon équivalente, lorsque  $k \leq n_c(1) + \frac{1}{2}$ , les valeurs propres vérifient  $\frac{\lambda_{k,n}}{n\pi} \approx \mu_k$  (d'où la valeur 2 pour le gap normalisé). Quand  $k > n_c(1)$  les valeurs propres normalisées  $\frac{\lambda_{k,n}}{n\pi}$  se rapprochent de leur comportement asymptotique, à savoir celui de l'opérateur  $-\Delta$ , d'où la linéarité du gap dans cette zone. La transition entre ces deux comportements est encore mal comprise, c'est en partie à cause de cela que le théorème de la section 1.2.4 est incomplet.

**Une minoration valable sur l'ensemble du spectre :**  $\lambda_{k,n} \geq n\pi(2k - 1)$

La comparaison suivante entre le spectre de (1.48) et de (1.49)

$$\forall k, n \in \mathbb{N}^*, \quad \frac{\lambda_{k,n}}{n\pi} \geq \mu_k, \quad (1.50)$$

s'établit en exprimant les valeurs propres des problèmes respectifs à l'aide du Théorème de Courant-Fisher (ou principe du max-min). L'expression de la valeur propre  $\frac{\lambda_{k,n}}{n\pi}$  fait alors apparaître un minimum sur un sous espace de dimension  $k$  de  $H_0^1(-\sqrt{n\pi}, \sqrt{n\pi})$ , alors que l'expression de la valeur propre  $\mu_k$  fait intervenir un minimum sur un autre ensemble  $V$ . Celui-ci contient le sous espace de  $H_0^1$  précédent, d'où le sens de l'inégalité (1.50). La démonstration est technique et est établie dans la section 4.4.2 du chapitre 4.

### Minoration sur la partie basse du spectre

Sur la partie basse du spectre, nous prouvons dans le chapitre 4 que,  $\tau \in (0, 1)$  étant fixé, pour  $n$  suffisamment grand et  $k \leq n_c(\tau)$ , on a

$$\frac{\lambda_{k,n}}{n\pi} \leq \mu_k + C_1 e^{-c_2 n}. \quad (1.51)$$

On montre alors avec l'estimation générale  $\lambda_{k,n} \geq n\pi(2k-1)$  que le gap est minoré par  $2 - C_1 e^{-c_2 n}$ .

Nous empruntons et généralisons une idée du papier [7] pour obtenir la majoration (1.51). Dans cet article, les auteurs tronquent la fonction  $G_1(x)$  pour se ramener aux conditions de Dirichlet homogènes et obtiennent une fonction qui vérifie le problème aux valeurs propres (1.48) avec pour valeur propre  $n\pi \times (2 \times 1 - 1)$  et un second membre supplémentaire, le commutateur provenant de la troncature, facilement contrôlable grâce à la décroissance exponentielle de  $n \mapsto G_1(\sqrt{n\pi})$ . Il est possible d'employer cette technique pour d'autres valeurs propres, pourvu que l'on puisse contrôler le second membre de l'équation de la même façon. Or, comme expliqué précédemment, nous démontrons dans le Lemme 4.4.4 que lorsque  $k \leq n_c(\tau)$ , les suites  $n \mapsto G_k(\sqrt{n\pi})$  sont exponentiellement décroissantes. Ceci s'obtient grâce à un développement asymptotique des fonctions  $G_k(x)$  dans une zone où  $x < \mu_k$  (voir [29]). La limitation technique nécessitant l'introduction du paramètre  $\tau$  provient de cette dernière expression.

On pourrait tenter de généraliser cette approche à l'ensemble des basses fréquences  $k \leq n_c(1)$ . Il faut alors étudier les fonctions  $G_k$  proche du point  $\sqrt{n\pi}$ . Comme le suggère la figure 1.6, la fonction  $G_k$  atteint un maximum aux alentours de ce point et les estimations disponibles dans la littérature (voir [6]) fournissent une estimation de ce maximum en  $n^{-1/3}$ . Cette borne n'est pas suffisante pour obtenir la propriété de gap voulue dans la preuve que nous proposons pour le Théorème 4.4.4.

### Minoration sur la partie haute du spectre

Nous expliquons comment obtenir la propriété de gap uniforme pour les valeurs propres vérifiant

$$\lambda_{k,n} \geq n^2 \pi^2 + n\pi. \quad (1.52)$$

Remarquons que cette condition est vérifiée lorsque  $k > \lceil \frac{\pi}{2} n \rceil$  en vertu de l'inégalité  $\lambda_{k,n} \geq n\pi(2k-1)$ . Nous parvenons à obtenir une propriété de gap plus forte, c'est-à-dire avec un minorant de la forme  $C\sqrt{\lambda_{k,n}}$ , si l'on considère une gamme d'indices plus restreinte :  $k \geq \frac{\pi}{2} n(1+C)$ , avec  $C > 0$  (voir le Théorème 4.5.1). Nous ne détaillerons pas ce dernier résultat dans cette section.

La méthode consiste d'abord à considérer la valeur propre dans l'équation (1.31) comme une variable du problème que l'on notera dorénavant  $t$ , et à transformer cette équation en un problème de Cauchy. Cette idée provient de l'article [21].

Pour ce faire, à  $t$  fixé, la solution du problème est maintenant notée  $h_n(t, x)$  et on change de plus les conditions de bord de l'équation (1.31) en les conditions initiales  $h_n(t, -1) = 0$  et  $h'_n(t, -1) = 1$ . La valeur de cette dernière condition sur  $h'_n$  n'a pas d'importance, car changer 1 en toute autre constante strictement positive n'a pour effet que de multiplier la fonction  $h_n$  par cette constante. On effectue ensuite une sorte de changement de variable polaire (dit *changement de variable de Prüfer modifié*) et on étudie l'équation différentielle vérifiée par la phase  $\phi_n$ . Plus précisément, la fonction  $h_n$  s'écrit

$$h_n(t, x) = r_n(t, x) \cos(\phi_n(t, x)) f_n(t, x),$$

où  $f_n(t, x)$  est une fonction ad hoc, qui n'est définie que lorsque  $t > n^2 \pi^2$ . La portée de cette méthode est donc, par nature, limitée à la partie haute du spectre.

Les fonctions  $r_n(t, \bullet)$  et  $f_n(t, \bullet)$  ne s'annulant jamais en  $x = 1$ , le paramètre  $t$  est une valeur propre  $\lambda_{k,n}$  dès que  $\phi_n(t, 1)$  est égal à  $\frac{\pi}{2}$  modulo  $\pi$ . En étudiant l'équation vérifiée par  $x \mapsto \phi(\lambda_{k,n}, x)$ , on montre que

$$\phi(\lambda_{k+1,n}, 1) - \phi(\lambda_{k,n}, 1) = -\pi.$$

La propriété de gap s'obtient alors en écrivant :  $-\pi = \int_{\lambda_{k,n}}^{\lambda_{k+1,n}} \partial_t \phi(t, 1) dt$  et en majorant uniformément par rapport à  $t$  la fonction  $t \mapsto |\partial_t \phi(t, 1)|$ , toujours sous l'hypothèse (1.52).

### Minoration de la norme $L^2$ des fonctions propres sur un intervalle de $(-1, 1)$

Dans le chapitre 4, nous prouvons qu'il suffit de montrer la propriété (1.30) avec l'ouvert  $(a, 1) \times (0, 1)$  pour en déduire la propriété (1.30) avec l'ouvert  $\omega = (a, b) \times (0, 1)$ . Il suffit donc de minorer convenablement  $\int_a^1 g_{k,n}^2$  pour obtenir le dernier ingrédient nécessaire à l'application de la méthode des moments suivant la stratégie énoncée dans la section 1.2.4.

Dans l'article [9], les auteurs parviennent, grâce à la méthode de transmutation, à montrer l'existence et à calculer le temps minimal de l'équation de Grushin lorsque l'ouvert de contrôle est constitué de deux bandes verticales : une bande  $(a, b) \times (0, 1)$ , avec  $0 \leq a < b \leq 1$ , et sa bande symétrique par rapport à l'axe (Oy). Nous notons cet ouvert symétrique  $\omega_{a,1}$ . Ils démontrent une inégalité proche de l'inégalité d'observabilité de l'équation des ondes. En injectant dans cette inégalité une solution judicieusement choisie à cette équation des ondes, nous montrons une minoration du type

$$\int_{\omega_{a,1}} g_{k,n}^2(x) dx \geq C e^{-n\pi(a^2+\delta)}, \quad (1.53)$$

avec  $\delta > 0$  arbitrairement petit. Or, le potentiel de l'opérateur différentiel d'ordre 2  $\mathcal{A}_n$  étant pair, il est facile de montrer que les fonctions propres  $g_{k,n}$  sont des fonctions paires ou impaires d'où  $\int_{\omega_{a,1}} g_{k,n}^2(x) dx = 2 \int_a^1 g_{k,n}^2(x)$ . La minoration voulue résulte alors de la minoration (1.53).

## 1.4 Perspectives de recherche

Les perspectives de cette thèse sont les suivantes.

- Tout d'abord, nous souhaiterions terminer la démonstration de la propriété de gap uniforme de l'équation de Grushin. Ceci permettrait de déduire ensuite des résultats de temps minimal de contrôle à zéro de systèmes d'équations de Grushin en cascade.
- Il serait également souhaitable de pouvoir visualiser numériquement ce temps minimal, grâce à la méthode HUM développée dans [11]. Nous pourrions ainsi établir d'autres conjectures sur l'existence et la valeur du temps minimal de contrôle à zéro de l'équation de Grushin lorsque celle-ci est contrôlée sur d'autres types d'ouverts  $\omega$  (par exemple certains ouverts  $\omega$  comme le 'domaine pincé' ou 'la cave' considérés dans [17], qui restent des problèmes ouverts) ou même lorsque celle-ci est posée sur d'autres domaines  $\Omega$  qu'un rectangle. Ces résultats numériques constitueraient une étude préliminaire pour démontrer des théorèmes de temps minimaux sur ces nouvelles géométries.
- La contrôlabilité au bord de l'équation des plaques discrétisée sur un domaine en une dimension est traitée dans l'article [26]. L'opérateur elliptique considéré est le bilaplacien sur un maillage uniforme. Les estimées spectrales du chapitre 2 semblent pouvoir s'adapter à ce type d'équation et pourraient permettre de généraliser leurs résultats à des opérateurs et des maillages plus généraux.
- Dans [5], les auteurs étudient le temps minimal de contrôle à zéro de systèmes d'équations paraboliques en cascade avec un terme de couplage dépendant de la variable d'espace  $x$ . Il serait intéressant d'étendre les résultats du chapitre 2 à des versions discrétisés de ces systèmes. Si l'on parvient à développer une méthode qui permet d'approcher numériquement le temps minimal de l'équation de Grushin dans les cas où celui-ci est déjà connu, il serait intéressant de l'appliquer à de tels systèmes. La difficulté principale proviendrait certainement du fait que le terme de couplage dans [5] est une fonction constante par morceaux avec un saut en un point d'abscisse irrationnelle.

## 1.5 Références

- [1] D. ALLONSIUS AND F. BOYER, *Boundary null-controllability of semi-discrete coupled parabolic systems in some multi-dimensional geometries*. working paper or preprint, July 2018, <https://hal.archives-ouvertes.fr/hal-01827044>. 6

- [2] D. ALLONSIUS AND F. BOYER, *Error estimate for the finite difference discretization of a sturm-liouville eigenvalue problem on quasi-uniform 1d meshes*, in preparation, (2018). 6, 9
- [3] D. ALLONSIUS, F. BOYER, AND M. MORANCEY, *Spectral analysis of discrete elliptic operators and applications in control theory*, Numerische Mathematik, (2018), doi:10.1007/s00211-018-0983-1, https://doi.org/10.1007/s00211-018-0983-1. 6
- [4] F. AMMAR KHODJA, A. BENABDALLAH, M. GONZÁLEZ-BURGOS, AND L. DE TERESA, *Minimal time for the null controllability of parabolic systems: the effect of the condensation index of complex sequences*, J. Funct. Anal., 267 (2014), pp. 2077–2151. 12
- [5] F. AMMAR KHODJA, A. BENABDALLAH, M. GONZÁLEZ-BURGOS, AND L. DE TERESA, *New phenomena for the null controllability of parabolic systems: minimal time and geometrical dependence*, J. Math. Anal. Appl., 444 (2016), pp. 1071–1113. 5, 12, 21
- [6] W. V. ASSCHE, *Entropy of hermite polynomials with application to the harmonic oscillator*. 20
- [7] K. BEAUCHARD, P. CANNARSA, AND R. GUGLIELMI, *Null controllability of Grushin-type operators in dimension two*, J. Eur. Math. Soc. (JEMS), 16 (2014), pp. 67–101, doi:10.4171/JEMS/428, http://dx.doi.org/10.4171/JEMS/428. 13, 20
- [8] K. BEAUCHARD, J. DARDÉ, AND S. ERVEDOZA, *Minimal time issues for the observability of Grushin-type equations*. working paper or preprint, Jan. 2018, https://hal.archives-ouvertes.fr/hal-01677037. 13
- [9] K. BEAUCHARD, L. MILLER, AND M. MORANCEY, *2D Grushin-type equations: minimal time and null controllable data*, J. Differential Equations, 259 (2015), pp. 5813–5845, doi:10.1016/j.jde.2015.07.007, http://dx.doi.org/10.1016/j.jde.2015.07.007. 13, 21
- [10] A. BENABDALLAH, F. BOYER, M. GONZÁLEZ-BURGOS, AND G. OLIVE, *Sharp estimates of the one-dimensional boundary control cost for parabolic systems and application to the  $n$ -dimensional boundary null controllability in cylindrical domains*, SIAM Journal on Control and Optimization, 52 (2014), pp. 2970–3001, doi:10.1137/130929680, http://dx.doi.org/10.1137/130929680. 4, 6, 11, 12, 17, 18
- [11] F. BOYER, *On the penalised HUM approach and its applications to the numerical approximation of null-controls for parabolic problems*, in CANUM 2012, 41e Congrès National d’Analyse Numérique, vol. 41 of ESAIM Proc., EDP Sci., Les Ulis, 2013, pp. 15–58. 5, 6, 10, 21
- [12] F. BOYER, F. HUBERT, AND J. L. ROUSSEAU, *Discrete carleman estimates for elliptic operators and uniform controllability of semi-discretized parabolic equations*, Journal de Mathématiques Pures et Appliquées, 93 (2010), pp. 240–276. 12
- [13] F. BOYER, F. HUBERT, AND J. L. ROUSSEAU, *Discrete carleman estimates for elliptic operators in arbitrary dimension and applications*, SIAM Journal on Control and Optimization, 48 (2010), pp. 5357–5397. 9, 10
- [14] J.-M. CORON, *Control and nonlinearity*, vol. 136 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2007. 2, 3
- [15] S. DOLECKI, *Observability for the one-dimensional heat equation*, Studia Math., 48 (1973), pp. 291–305. 12
- [16] S. DOLECKI AND D. L. RUSSELL, *A general theory of observation and control*, SIAM J. Control Optimization, 15 (1977), pp. 185–220, doi:10.1137/0315015, https://doi.org/10.1137/0315015. 3
- [17] M. DUPREZ AND A. KOENIG, *Control of the grushin equation: non-rectangular control region and minimal time*, 2018, arXiv:arXiv:1807.01241. 13, 21
- [18] H. O. FATTORINI AND D. L. RUSSELL, *Exact controllability theorems for linear parabolic equations in one space dimension*, Archive for Rational Mechanics and Analysis, 43 (1971), pp. 272–292, doi:10.1007/BF00250466, https://doi.org/10.1007/BF00250466. 5

- [19] H. O. FATTORINI AND D. L. RUSSELL, *Uniform bounds on biorthogonal functions for real exponentials with an application to the control theory of parabolic equations*, Quart. Appl. Math., 32 (1974/75), pp. 45–69. [3](#), [4](#)
- [20] R. GLOWINSKI, J.-L. LIONS, AND J. HE, *Exact and approximate controllability for distributed parameter systems*, vol. 117 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2008. A numerical approach. [5](#)
- [21] W. KIRSCH AND B. SIMON, *Universal lower bounds on eigenvalue splittings for one-dimensional schrödinger operators*, Comm. Math. Phys., 97 (1985), pp. 453–460, <https://projecteuclid.org/443/euclid.cmp/1103942135>. [20](#)
- [22] A. KOENIG, *Non-null-controllability of the grushin operator in 2d*, Comptes Rendus Mathématique, 355 (2017), pp. 1215 – 1235, [doi:10.1016/j.crma.2017.10.021](https://hal.archives-ouvertes.fr/hal-01654043), <https://hal.archives-ouvertes.fr/hal-01654043>. [13](#)
- [23] S. LABBÉ AND E. TRÉLAT, *Uniform controllability of semidiscrete approximations of parabolic control systems*, Systems & Control Letters, 55 (2006), pp. 597–609. [10](#)
- [24] G. LEBEAU AND L. ROBBIANO, *Contrôle exact de l'équation de la chaleur*, Séminaire Équations aux dérivées partielles (Polytechnique), (1994-1995), pp. 1–11, <http://eudml.org/doc/112118>. [11](#), [18](#)
- [25] A. LOPEZ AND E. ZUAZUA, *Some new results related to the null controllability of the 1-d heat equation*, in Séminaire sur les Équations aux Dérivées Partielles, 1997–1998, École Polytech., Palaiseau, 1998, pp. Exp. No. VIII, 22. [7](#)
- [26] S. MICU, I. ROVENȚA, AND L. E. TEMEREANȚĂ, *Approximation of the controls for the linear beam equation*, Math. Control Signals Systems, 28 (2016), pp. Art. 12, 53. [21](#)
- [27] C. H. MÜNTZ, *Über den Approximationssatz von Weierstraß*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1914, pp. 303–312, [doi:10.1007/978-3-642-50735-9\\_22](https://doi.org/10.1007/978-3-642-50735-9_22), [https://doi.org/10.1007/978-3-642-50735-9\\_22](https://doi.org/10.1007/978-3-642-50735-9_22). [3](#)
- [28] L. SCHWARTZ, *Étude des sommes d'exponentielles réelles*, NUMDAM, [place of publication not identified], 1943, [http://www.numdam.org/item?id=THESE\\_1943\\_\\_259\\_\\_1\\_0](http://www.numdam.org/item?id=THESE_1943__259__1_0). [3](#)
- [29] G. SZEGÓ, *Orthogonal polynomials*, American Mathematical Society, Providence, R.I., fourth ed., 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII. [20](#)
- [30] M. TUCSNAK AND G. WEISS, *Observation and Control for Operator Semigroups*, Birkhäuser Advanced Texts Basler Lehrbücher, Birkhäuser Basel, 2009, [https://books.google.fr/books?id=im\\_m0wmZJk0C](https://books.google.fr/books?id=im_m0wmZJk0C). [2](#)
- [31] J. ZABCZYK, *Mathematical control theory - an introduction*, Systems & Control: Foundations & Applications, Birkhäuser, 1992. [2](#)
- [32] E. ZUAZUA, *Propagation, observation, and control of waves approximated by finite difference methods*, SIAM Rev., 47 (2005), pp. 197–243. [10](#)



# Chapter 2

## Spectral analysis of discrete elliptic operators and applications in control theory

*” Celui qui trouve sans chercher est celui qui a longtemps cherché sans trouver. ”*

Gaston Bachelard

### Contents

---

<b>2.1</b>	<b>Introduction</b>	<b>27</b>
2.1.1	General context	27
2.1.2	Notations and discrete framework	28
2.1.3	Presentation of the results	29
<b>2.2</b>	<b>Spectral properties of elliptic operators. The continuous case</b>	<b>34</b>
<b>2.3</b>	<b>Discrete spectral properties</b>	<b>38</b>
2.3.1	General strategy	38
2.3.2	Estimates in the general case (S1)	41
2.3.3	Estimates for of a uniform grid, cases (S2) and (S3)	42
2.3.4	Proof of Lemma 2.3.1	48
<b>2.4</b>	<b>Numerical illustrations</b>	<b>49</b>
<b>2.5</b>	<b>Applications in control theory</b>	<b>52</b>
2.5.1	Null-controllability via the moments method in the continuous setting	52
2.5.2	Null controllability in the discrete setting	56
2.5.3	$\phi(h)$ -null controllability	58
<b>2.6</b>	<b>Remarks and further results</b>	<b>61</b>
<b>2.7</b>	<b>Appendix</b>	<b>62</b>
2.7.1	The case of non uniform but regular meshes	62
2.7.2	Error estimate for the finite difference discretization of a Sturm-Liouville eigenvalue problem on quasi-uniform 1D meshes	66
2.7.3	Estimates on discrete eigenfunctions with a piecewise constant diffusion coefficients	75
2.7.4	Null control with separated variables for the discretized heat equation	79
2.7.5	Simulations	82
<b>2.8</b>	<b>References</b>	<b>83</b>

---

### Résumé

Ce chapitre reprend en grande partie l'article [3] : *Spectral analysis of discrete elliptic operators and applications in control theory*, écrit en collaboration avec Franck Boyer et Morgan Morancey. Nous avons ajouté dans ce Chapitre l'Annexe 2.7 qui complète la Remarque 2.1.6, propose une preuve complète de la Proposition 2.3.1, précise la seconde remarque de la section 2.6 en analysant les propriétés spectrales des opérateurs elliptiques avec un coefficient de diffusion  $\gamma$  constant par morceaux et un potentiel nul et propose une façon de construire un contrôle à variables séparées pour l'équation de la chaleur semi-discrétisée en espace sur un maillage uniforme. Nous terminons en donnant des résultats numériques obtenus avec la méthode HUM pénalisée.

Ce papier est divisé en deux parties. Dans un premier temps, nous étudions des propriétés spectrales d'opérateurs elliptiques auto-adjoints d'ordre deux discrétisés par différences finies sur des maillages assez généraux en 1D. Nous analysons en particulier le comportement des fonctions propres et valeurs propres en établissant des estimations uniformes en le paramètre de discrétisation.

Dans une seconde partie, à l'aide de cette première étude, nous démontrons de nouveaux résultats sur la contrôlabilité à zéro d'EDPs paraboliques semi-discrétisées en espace par des contrôles uniformément bornés par rapport au paramètre de discrétisation. Nous obtenons en particulier des résultats sur le contrôle à zéro au bord de systèmes paraboliques avec moins de contrôles que d'équations. Nous établissons de plus des résultats de  $\phi(h)$  contrôlabilité à zéro lorsque le problème de contrôle est discrétisé sur certains maillages quasi-uniformes.

### Abstract

This Chapter includes the article [3] : *Spectral analysis of discrete elliptic operators and applications in control theory*, written in collaboration with Franck Boyer and Morgan Morancey. In this Chapter, we added Appendix 2.7 which completes Remark 2.1.6, proposes a complete proof of Proposition 2.3.1, precises the second Remark of section 2.6 analyzing some spectral properties of elliptic operators with a piecewise constant diffusion coefficient  $\gamma$  and a potentiel equal to zero and proposes a way to build a null control with separated variables to control the semi discretized in space heat equation on a uniform mesh. Finally, we give some numerical results obtained with the penalized HUM method.

In this paper we propose an analysis of discrete spectral properties for a finite difference discretization of quite general 1D second-order self-adjoint elliptic operators. We particularly investigate some (uniform with respect to the discretization parameter) qualitative behavior of eigenfunctions and eigenvalues.

With those estimates we manage to obtain new results for the construction of bounded families of controls for semi-discrete parabolic PDEs, in particular for boundary controls of a coupled parabolic system with fewer controls than equations. Moreover, we establish  $\phi(h)$  null controllability results when the control problem is discretized on some quasi-uniform meshes.



## 2.1 Introduction

### 2.1.1 General context

We consider the differential operator

$$\mathcal{A} := -\partial_x(\gamma(x)\partial_x \cdot) + q(x). \quad (2.1)$$

defined on the unit interval  $\Omega = (0, 1)$ . The real valued functions  $\gamma$  and  $q$  are given and satisfy, for some  $s \geq 0$ ,

$$q \in \mathcal{C}^s(\overline{\Omega}) \text{ and } \gamma \in \mathcal{C}^{s+1}(\overline{\Omega}), \text{ with } \gamma_{\min} := \inf_{\Omega} \gamma > 0. \quad (\text{H}_s)$$

We shall sometimes need to see  $\mathcal{A}$  as a self-adjoint unbounded operator in  $L^2(\Omega)$  whose domain is  $D(\mathcal{A}) = H_0^1(\Omega) \cap H^2(\Omega)$ . In this setting, it is well-known that there exists an orthonormal basis of  $L^2(\Omega)$  made of eigenfunctions of  $\mathcal{A}$  that we denote by  $(\phi_k)_{k \geq 1}$ , the associated (real) eigenvalues being denoted by  $(\lambda_k)_{k \geq 1}$ . Those eigenvalues are simple and we assume that they are sorted in increasing order.

For any subset  $O \subset \Omega$  and any  $u \in L^2(\Omega)$ , we define the semi-norm

$$\|u\|_{L^2(O)} := \left( \int_O |u(x)|^2 dx \right)^{\frac{1}{2}},$$

and for any smooth function  $u : \overline{\Omega} \rightarrow \mathbb{R}$ , we set

$$\partial_r u := u'(1), \text{ and } \partial_l u := -u'(0),$$

which are the normal derivatives of  $u$  at both ends of the interval  $\Omega$ .

In this section, we recall some qualitative properties of the eigenelements  $(\phi_k, \lambda_k)_k$  that we are interested in. Of course, there exist much more accurate descriptions of the asymptotic behavior of the eigenelements of such operators (see [22]) but we only deal here with the same level of accuracy as the one that we shall achieve in the discrete framework.

Throughout this paper  $C$  denotes a positive number that may vary from line to line. If necessary, we add information about dependency on parameters  $p_1, p_2, \dots, p_k$  in the following way :  $C := C(p_1, p_2, \dots, p_k)$ .

**Theorem 2.1.1** *Assume that  $(\text{H}_0)$  holds. There exists  $C_1(q, \gamma) > 0$ , such that*

1. For any  $k \geq 1$ ,

$$\lambda_{k+1} - \lambda_k \geq C_1 k.$$

2. For any  $k \geq 1$

$$|\partial_r \phi_k| \geq C_1 k, \text{ and } |\partial_l \phi_k| \geq C_1 k.$$

3. For any non-empty open subset  $\omega \subset \Omega$ , there exists  $C_2(q, \gamma, \omega) > 0$  such that for any  $k \geq 1$ ,

$$\|\phi_k\|_{L^2(\omega)} \geq C_2.$$

**Remark 2.1.1** *Considering  $\gamma = 1$  and  $q = 0$  one can easily check that the bounds given in Theorem 2.1.1 are sharp. Indeed, in this case, we know that  $\lambda_k = k^2 \pi^2$ ,  $\phi_k(x) = \sqrt{2} \sin(\sqrt{\lambda_k} x)$ . Thus,*

$$\lambda_{k+1} - \lambda_k = (2k+1)\pi^2 \text{ and } |\partial_b \phi_k| = \sqrt{2\lambda_k} = \sqrt{2}\pi k, \quad \forall k \geq 1, \quad \forall b \in \{l, r\},$$

and for  $\omega = (a, b) \subset (0, 1)$ , we have

$$\|\phi_k\|_{L^2(\omega)}^2 = \int_a^b 2 \sin^2(k\pi x) dx \xrightarrow[k \rightarrow \infty]{} b - a = |\omega|.$$

**Remark 2.1.2** *By using the min-max principle and the exact knowledge of the eigenvalues of the Laplace operator as recalled in the previous remark, we can show that the eigenvalues of  $\mathcal{A}$  satisfy*

$$\gamma_{\min} \pi^2 k^2 - \|q\|_{L^\infty} \leq \lambda_k \leq \|\gamma\|_{L^\infty} \pi^2 k^2 + \|q\|_{L^\infty}, \quad \forall k \geq 1. \quad (2.2)$$

The first point in Theorem 2.1.1 implies in particular the so-called *gap property* for the eigenvalues, which means that there exists some  $\rho > 0$  such that  $\lambda_{k+1} - \lambda_k \geq \rho$  for any  $k$ .

It appears that all those properties allow, by using the so-called *moments method* (see [15]), to prove some null-controllability results for the parabolic problem associated with  $\mathcal{A}$  (either for a distributed control or for a boundary control) as well as for some coupled parabolic systems associated with  $\mathcal{A}$  with fewer controls than components. This will be detailed in section 2.5.1.

The main goal of this paper is to study whether or not the three results given in Theorem 2.1.1 still hold for a discrete version  $\mathcal{A}^h$  of the differential operator uniformly with respect to the discretization parameter  $h > 0$ . As a consequence, we shall obtain new uniform controllability results for the associated semi-discrete parabolic equation and systems via the moments method adapted to the discrete setting.

## 2.1.2 Notations and discrete framework

For any two real numbers  $\alpha, \beta$  we define

$$\llbracket \alpha, \beta \rrbracket := [\min(\alpha, \beta), \max(\alpha, \beta)] \cap \mathbb{N},$$

to be the set of all the integers between  $\alpha$  and  $\beta$ .

Let us recall here the standard notation for the finite difference discretization of the differential operator  $\mathcal{A}$ . We consider an ordered set of  $N + 2$  distinct points of  $[0, 1]$  denoted by  $(x_i)_{i=0}^{N+1}$  and satisfying  $x_0 = 0$  and  $x_{N+1} = 1$ . Let  $(x_{i+1/2})_{i=0}^N$  be the dual mesh points defined by  $x_{i+1/2} := \frac{x_{i+1} + x_i}{2}$ . We set  $h_{i+1/2} := x_{i+1} - x_i$  for  $i \in \llbracket 0, N \rrbracket$  and  $h_i := x_{i+1/2} - x_{i-1/2} = \frac{h_{i+1/2} + h_{i-1/2}}{2}$ , for  $i \in \llbracket 1, N \rrbracket$ . As usual, the mesh size is defined by  $h := \max_{i \in \llbracket 0, N \rrbracket} h_{i+1/2}$ . We use the standard abuse of notation that consists in using the index  $h$  to denote any mesh-dependent quantity even though it obviously not only depends on  $h$  but also on the whole geometry of the mesh points. For such a given mesh we introduce

$$\Theta_h := \frac{h}{\min_{i \in \llbracket 0, N \rrbracket} h_{i+1/2}},$$

which is a measure of the (non-)uniformity of the mesh. A family of meshes such that  $(\Theta_h)_h$  is bounded, is called *quasi-uniform*.

For any open subset  $O \subset \Omega$ , we define the discrete  $L^2(O)$  semi-norm on  $\mathbb{R}^N$  by

$$\|U\|_{L^2(O_h)} := \left( \sum_{\substack{i \in \llbracket 1, N \rrbracket \\ x_i \in O}} h_i |u_i|^2 \right)^{\frac{1}{2}}, \quad \forall U = (u_i)_{i=1}^N \in \mathbb{R}^N,$$

and for a time-dependent function  $t \in (0, T) \mapsto U(t) \in \mathbb{R}^N$  we set

$$\|U\|_{L^2(O_h^T)} := \left( \int_0^T \|U(t)\|_{L^2(O_h)}^2 dt \right)^{\frac{1}{2}}.$$

It is easily seen that  $\|\cdot\|_{L^2(O_h)}$  is actually a Euclidean norm on  $\mathbb{R}^N$ , whose inner product is denoted by  $\langle \cdot, \cdot \rangle_{L^2(O_h)}$ .

We set  $q_i := q(x_i)$ ,  $\gamma_{i+1/2} := \gamma(x_{i+1/2})$  and we define the square matrix  $\mathcal{A}^h$  of size  $N$  by

$$(\mathcal{A}^h U)_j := -\frac{1}{h_j} \left( \gamma_{j+1/2} \frac{u_{j+1} - u_j}{h_{j+1/2}} - \gamma_{j-1/2} \frac{u_j - u_{j-1}}{h_{j-1/2}} \right) + q_j u_j, \quad \forall j \in \llbracket 1, N \rrbracket \quad (2.3)$$

for any discrete function  $U = (u_i)_{i=1}^N \in \mathbb{R}^N$ , and any  $j \in \llbracket 1, N \rrbracket$ , with the usual convention that  $u_0 = u_{N+1} = 0$  to take into account homogeneous Dirichlet boundary conditions.

By analogy with the continuous setting we define the normal derivative of a discrete function  $U \in \mathbb{R}^N$ , taking into account the Dirichlet boundary condition, as follows

$$\partial_r U := \frac{0 - u_N}{h_{N+1/2}}, \quad \text{and} \quad \partial_l U := -\frac{u_1 - 0}{h_{1/2}}.$$

It is well-known that the discrete operator  $\mathcal{A}^h$  is a second order accurate approximation of the differential operator  $\mathcal{A}$  (see [7]). Moreover,  $\mathcal{A}^h$  is self-adjoint in the space  $(\mathbb{R}^N, \langle \cdot, \cdot \rangle_{L^2(O_h)})$  so that

there exists a (finite) orthonormal family  $(\phi_k^h)_{k=1}^N \subset \mathbb{R}^N$  made of eigenvectors of  $\mathcal{A}^h$  (also referred to as *discrete eigenfunctions* in this paper), and associated with eigenvalues denoted by  $\Lambda^h := (\lambda_k^h)_{k=1}^N$ . In other words, we have the relations

$$\mathcal{A}^h \phi_k^h = \lambda_k^h \phi_k^h, \quad \text{and} \quad \langle \phi_k^h, \phi_j^h \rangle_{L^2(\Omega_h)} = \delta_{kj}, \quad \forall k, j \in \llbracket 1, N \rrbracket.$$

All those eigenvalues are simple and we assume that they are sorted in increasing order.

### 2.1.3 Presentation of the results

The results of this paper are split into two parts.

We first study spectral properties of the discrete operator  $\mathcal{A}^h$  similar to the ones given in Theorem 2.1.1 of Section 2.3 in the continuous setting. To prepare this analysis we shall give the corresponding proofs at the continuous level in Section 2.2. In Section 2.4, we propose some numerical simulations that illustrate our theoretical results. In a second part of the paper (Section 2.5), we give our main applications in discrete controllability theory for system of parabolic equations.

To begin with, notice that all computations are explicit in the case of discrete Laplace operator on a uniform mesh.

**Remark 2.1.3 (discrete Laplace operator on uniform meshes)** *Let  $h_* = 1/(N+1)$  and denote by  $\mathcal{A}^{h_*}$  the discrete operator obtained on the uniform mesh of size  $h_*$  with  $\gamma = 1$  and  $q = 0$  (which is nothing but the very usual discrete Laplace operator). Then, the eigenelements of  $\mathcal{A}^{h_*}$  are explicitly given by*

$$\lambda_k^{h_*} = \frac{4}{h_*^2} \sin^2 \left( \frac{k\pi h_*}{2} \right), \quad (\phi_k^{h_*})_i = \sqrt{2} \sin(k\pi i h_*), \quad \forall i, k \in \llbracket 1, N \rrbracket.$$

In particular, we have

$$4k^2 \leq \lambda_k^{h_*} \leq \pi^2 k^2, \quad \forall k \in \llbracket 1, N \rrbracket,$$

and

$$\lambda_k^{h_*} \leq \lambda_N^{h_*} = \frac{4}{h_*^2} \cos^2 \left( \frac{\pi}{2} h_* \right) \leq \frac{4}{h_*^2} - 4, \quad \forall k \in \llbracket 1, N \rrbracket.$$

Here we used the standard inequality  $\cos^2(t) \leq 1 - \frac{4}{\pi^2} t^2$  for any  $0 \leq t \leq 1$ .

By using the formulas recalled in the remark above, one can easily obtain the following result.

**Proposition 2.1.1** *Let  $h_*$  and  $\mathcal{A}^{h_*}$  be as defined in Remark 2.1.3. The following properties hold:*

— *Gap property:*

$$12 \leq \min_{k \in \llbracket 1, N-1 \rrbracket} (\lambda_{k+1}^{h_*} - \lambda_k^{h_*}) \leq 3\pi^2,$$

*moreover the minimum is achieved for  $k = 1$  and  $k = N - 1$ .*

— *Discrete normal derivative estimate: for  $b \in \{l, r\}$*

$$2\sqrt{2}k \leq |\partial_b \phi_k^{h_*}| \leq \sqrt{2}k\pi, \quad \forall k \in \llbracket 1, N/2 \rrbracket,$$

$$2\sqrt{2}(N+1-k) \leq |\partial_b \phi_k^{h_*}| \leq \sqrt{2}(N+1-k)\pi, \quad \forall k \in \llbracket N/2, N \rrbracket.$$

—  *$L^2$ -norm estimate: for any non empty open subset  $\omega \subset \Omega$ , there exists  $h_0(\omega) > 0$  and a  $C(q, \gamma, \omega)$  such that, for any  $h_* < h_0$ , we have*

$$\|\phi_k^{h_*}\|_{L^2(\omega_h)} \geq C, \quad \forall k \in \llbracket 1, N \rrbracket.$$

Some comments are in order. There are fundamental differences between Theorem 2.1.1 and Proposition 2.1.1. We first observe that the uniform gap property holds but not the stronger property 1 of Theorem 2.1.1 since the distance between the two largest discrete eigenvalues of  $\mathcal{A}^{h_*}$  does not tend to infinity. Similarly, the discrete normal derivative of the eigenfunction  $\phi_N^{h_*}$  for instance does not tend to infinity as  $N \rightarrow \infty$  (and simultaneously  $h_* \rightarrow 0$  of course). It thus already appears that the qualitative behavior of the eigenelements of  $\mathcal{A}^{h_*}$  for the upper part of the spectrum may be quite different from the one of the continuous operator, as we can see on Figure 2.1. However, for the lower part of the spectrum the properties are quite similar to the ones of  $\mathcal{A}$ .

Our aim is to generalize the three estimates of Proposition 2.1.1 to non uniform meshes and non constant coefficients operators for which we do not possess analytic formulas for the discrete eigenfunctions and eigenvalues.

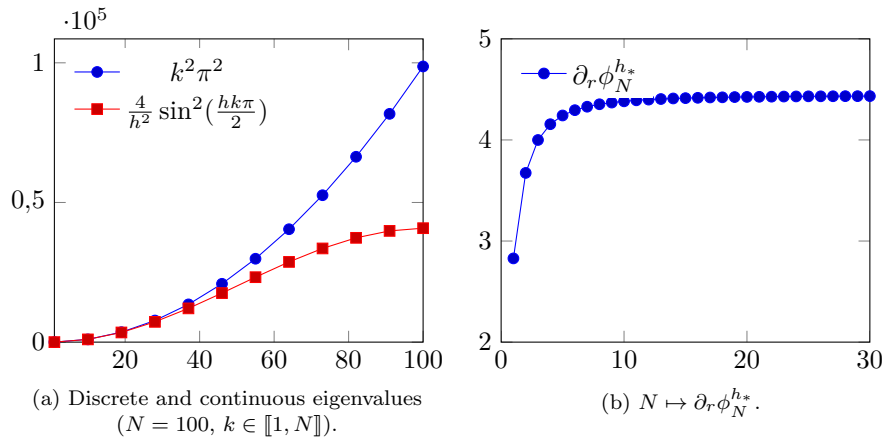


Figure 2.1 – Differences between discrete and continuous eigenlements.

**Remark 2.1.4 (Spectral properties for a given mesh and operator)** *It is worth noticing that, by straightforward linear algebra arguments, the following properties hold without any assumption on the mesh or on the coefficient of the operator*

$$\min_{1 \leq k \leq N-1} (\lambda_{k+1}^h - \lambda_k^h) > 0,$$

$$\min_{1 \leq k \leq N} |\partial_b \phi_k^h| > 0, \quad \forall b \in \{l, r\},$$

$$\min_{1 \leq k \leq N} \|\phi_k^h\|_{L^2(\omega_h)} > 0,$$

as soon as  $\omega$  contains at least two points of the mesh. However, those bounds from below in general tend to 0 as  $h \rightarrow 0$  (see the numerical experiments in Section 2.4). In that respect, the case of a constant coefficient operator on a uniform mesh is very particular (see Proposition 2.1.1).

Therefore, our main concern in this paper will be to obtain lower bounds for the quantities above that are uniform with respect to  $h$ . More precisely, we will try to identify the *maximal* subset of  $\llbracket 1, N \rrbracket$  for which such uniform bounds hold.

**Remark 2.1.5** *From the discussion above, it already appears that this problem cannot be tackled by classical numerical analysis arguments. Indeed, although it can be established that in the continuous setting  $\lambda_{k+1} - \lambda_k \geq Ck$  (see Theorem 2.1.1, first estimate), we cannot deduce directly from this estimate that a gap property holds in the discrete setting since Figure 2.1 reveals that the error of approximation  $|\lambda_N - \lambda_N^h|$  can be as large as  $CN^2$ , see also Proposition 2.3.1. However, for the low frequencies ( $k \in \{1, \dots, CN^\alpha\}$  with a suitable choice of  $\alpha < 1$  and  $C$  (see Theorem 2.3.1)), one can get that  $|\lambda_k - \lambda_k^h| \leq \epsilon k$ , with  $\epsilon$  sufficiently small so that estimate 1 of Theorem 2.1.1 allows us to deduce a uniform gap estimate in the discrete setting for this particular portion of the spectrum. We will apply this argument in the most general setting we consider. Nevertheless, with stronger assumptions, we develop a strategy to obtain better estimates.*

We will consider three slightly different sets of hypothesis for the regularity of the meshes and the coefficients of our operator, namely

$$\text{quasi-uniform meshes, } q \text{ and } \gamma \text{ satisfying (H}_2\text{)}, \quad (\text{S1})$$

$$\text{uniform meshes, } q \text{ and } \gamma \text{ satisfying (H}_0\text{)}, \quad (\text{S2})$$

$$\text{uniform meshes, } q \text{ and } \gamma \text{ satisfying (H}_0\text{)}, \text{ and } \gamma \text{ is constant.} \quad (\text{S3})$$

In the remark below we show that the results we shall obtain in Section 2.3 about operator  $\mathcal{A}^h$  on *uniform meshes* in the setting (S2) still hold, in fact, for a wide class of *regular but non uniform meshes*. Yet, to simplify the presentation of the next sections, we shall always deal with uniform meshes in the setting (S2).

**Remark 2.1.6 (Families of regular meshes)** *In this remark, we use the notation  $\mathcal{A}_{\gamma,q}^h$  for the discrete operator defined on any given mesh and for any couple of coefficients  $q$  and  $\gamma$ . We will also denote by  $\mathcal{A}_{\gamma,q}^{h,*}$  the discrete operator defined on the uniform mesh with the same number of cells and a mesh size  $h_* = \frac{1}{N+1}$ .*

We define the vector  $\alpha^h = (\alpha_i)_{1 \leq i \leq N}$  by

$$\alpha_i := \sqrt{\frac{h_i}{h_*}},$$

and we set

$$\bar{\gamma}_{i+1/2} := \frac{h_*^2}{h_{i+1/2} \sqrt{h_i h_{i+1}}} \gamma_{i+1/2}, \quad \forall i \in \llbracket 0, N \rrbracket,$$

and

$$\bar{q}_i := q_i - \frac{1}{\alpha_i} \mathcal{A}_{\bar{\gamma},0}^{h,*} \alpha^h, \quad \forall i \in \llbracket 1, N \rrbracket.$$

With those notations, one can check by direct algebraic manipulations that, if  $u^h \in \mathbb{R}^N$  and  $\lambda \in \mathbb{R}$  satisfy  $\mathcal{A}_{\gamma,q}^h u^h = \lambda u^h$ , then we have

$$\mathcal{A}_{\bar{\gamma},\bar{q}}^{h,*} (\alpha^h u^h) = \lambda (\alpha^h u^h).$$

In other words, the operators  $\mathcal{A}_{\gamma,q}^h$  and  $\mathcal{A}_{\bar{\gamma},\bar{q}}^{h,*}$  have exactly the same eigenvalues and the corresponding eigenfunctions are deduced one from each other by a term by term multiplication by the coefficients  $\alpha_i$ .

It follows that, any (uniform with respect to  $h$ ) spectral property proved on the modified operator  $\mathcal{A}_{\bar{\gamma},\bar{q}}^{h,*}$  on a uniform mesh, lead to the equivalent property for the original operator  $\mathcal{A}_{\gamma,q}^h$  on a non-uniform grid. The price to pay in this manipulation lies in the fact that the new coefficients  $\bar{\gamma}$  and  $\bar{q}$  actually depend on the geometry of the initial mesh. In particular, one needs to control their “regularity” for uniform estimates to hold. More precisely, we need to make sure that they are, in some sense, bounded in  $C^1$  and  $C^0$  respectively (as required in the assumption  $(H_0)$ ). Of course, this has to be carefully defined since those discrete coefficients are not, a priori, obtained by sampling some functions on the primal and dual meshes.

As an important example, it can be shown by using standard Taylor formulas that any mesh obtained as the image of a uniform mesh by a diffeomorphism  $\vartheta : [0, 1] \rightarrow [0, 1]$  of class  $C^3$  as follows

$$x_i = \vartheta(ih_*), \quad \forall i \in \llbracket 1, N \rrbracket,$$

fulfills all the required regularity properties for our results to hold.

Those particular meshes families were already considered for instance in [10] but also in the recent work [14].

Observe however that the same manipulation does not hold in the setting (S3) since it transforms a constant-diffusion problem into a variable and mesh-dependent diffusion problem.

## Discrete spectral properties

Let us first gather here the various results we will prove on qualitative properties of the eigenlements of  $\mathcal{A}^h$ . Without loss of generality we will assume that  $q \geq 0$  so that all the eigenvalues  $\lambda_k^h$  and  $\lambda_k$  are positive. Otherwise, one need to replace  $\lambda_k^h$  by  $\lambda_k^h + \|q\|_\infty$  in all the results below.

**Properties of discrete eigenfunctions** Let  $\omega$  be a fixed non empty open subdomain of  $\Omega$ .

— In the case (S1), we obtain (Theorem 2.3.1) bounds from below of the following kind

$$\|\phi_k^h\|_{L^2(\omega_h)}^2 \geq C_1 e^{-C_2 \sqrt{\lambda_k^h}}, \quad \forall k \in \llbracket 1, N \rrbracket,$$

for any non empty open subdomain  $\omega \subset \Omega$  and  $h$  small enough, and

$$|\partial_b \phi_k^h| \geq C_1 \sqrt{\lambda_k^h} e^{-C_2 \sqrt{\lambda_k^h}}, \quad \forall k \in \llbracket 1, N \rrbracket, \quad \forall b \in \{l, r\},$$

where  $C_1$  and  $C_2$  do not depend on the mesh size  $h$ .

- In the case (S2), we can refine those estimates on a constant lower fraction of the spectrum by proving (Theorem 2.3.2) that there exists a constant  $0 < \alpha \leq 1$  (depending only on  $\gamma$  and  $\omega$ ) such that

$$\begin{aligned} \|\phi_k^h\|_{L^2(\omega_h)}^2 &\geq C_1, \quad \forall k \in \llbracket 1, \alpha N \rrbracket \\ |\partial_b \phi_k^h| &\geq C_1 \sqrt{\lambda_k^h}, \quad \forall k \in \llbracket 1, \alpha N \rrbracket, \forall b \in \{l, r\}. \end{aligned}$$

Moreover, we can give a sharp estimate of the best constant  $\alpha$  for which those properties hold, as attested by numerical experiments in Section 2.4.

- Finally, in the case (S3), we can recover *optimal* (in a sense to be made precise later) estimates on the whole spectrum (Theorem 2.3.3)

$$\begin{aligned} \|\phi_k^h\|_{L^2(\omega_h)}^2 &\geq C_1 \left( 1 - \frac{h^2(\lambda_k^h - \|q\|_{L^\infty})}{4\gamma} \right), \quad \forall k \in \llbracket 1, N \rrbracket, \\ |\partial_b \phi_k^h| &\geq C_1 \sqrt{\lambda_k^h} \sqrt{1 - \frac{h^2(\lambda_k^h - \|q\|_{L^\infty})}{4\gamma}}, \quad \forall k \in \llbracket 1, N \rrbracket, \forall b \in \{l, r\}. \end{aligned}$$

Observe that in the settings (S2) and (S3), for a fixed value of  $k$  and  $h \rightarrow 0$ , since it is well-known that  $\lambda_k^h \rightarrow \lambda_k$ , we recover the same asymptotic lower bounds as for the continuous operator  $\mathcal{A}$ . The latter are provided by Theorem 2.1.1.

From those estimates, we will deduce in particular that for some  $C_2 > 0$ ,

$$\begin{aligned} \|\phi_k^h\|_{L^2(\omega_h)}^2 &\geq \frac{C_2}{\lambda_k^h}, \quad \forall k \in \llbracket 1, N \rrbracket, \\ |\partial_b \phi_k^h| &\geq C_2, \quad \forall k \in \llbracket 1, N \rrbracket, \forall b \in \{l, r\}. \end{aligned}$$

**Gap property for the eigenvalues** We study here the so-called uniform gap property, that is : is there a constant  $C > 0$ , independent of  $h > 0$ , such that

$$\lambda_{k+1}^h - \lambda_k^h \geq C, \quad \forall k \in \llbracket 1, N-1 \rrbracket \quad ? \quad (2.4)$$

In Proposition 2.1.1 we have shown that this property holds for the discrete Laplacian on a uniform mesh.

In the general cases where  $q$  and  $\gamma$  only satisfy (H<sub>0</sub>), numerical experiments (Section 2.4) reveal that the uniform gap property may fail, at least on the upper part of the spectrum. More precisely, one can exhibit two distinct eigenvalues that are exponentially close as  $h \rightarrow 0$  in the case where  $\gamma$  is not a constant function.

Thus, a reasonable question is: does a gap property hold if one only considers a lower fraction of the spectrum, or more precisely is there a positive increasing function  $F : \mathbb{N} \rightarrow \mathbb{R}$  such that for some  $C > 0$  we have for any  $h > 0$ ,

$$\lambda_{k+1}^h - \lambda_k^h \geq C, \quad \forall k \in \llbracket 1, F(N) \rrbracket \quad ? \quad (2.5)$$

In the case (S2), we prove existence of a sharp (according to numerical simulations) parameter  $\alpha \in (0, 1]$  such that  $F(N) = \alpha N$  satisfies the above requirements. In the case (S3), we prove a weaker property, namely that for any  $\beta > 0$  and any mesh satisfying  $\Theta_h \leq \beta$ , there exists  $\alpha := \alpha(\gamma, q, \beta) > 0$  such that  $F(N) = \alpha N^{2/5}$ .

### Applications to the numerical approximation of control problems

As a main application of the previous estimates we propose new uniform controllability results for 1D semi-discrete parabolic equations and systems either for a distributed control or for a boundary control. They are proved by using the moments method that heavily depends on the discrete spectral properties established in this paper.

We first consider the semi-discretized controlled scalar parabolic equation with a finite time horizon  $T > 0$

$$\begin{cases} (y^h)'(t) + \mathcal{A}^h y^h(t) = \mathcal{D}_\omega^h V_d^h(t) + \mathcal{B}_r^h V_b^h(t), & \text{for } 0 < t \leq T \\ y^h(0) = y^{0,h} \in \mathbb{R}^N. \end{cases} \quad (2.6)$$

Here, we consider two possible types of control

- $V_d^h : (0, T) \rightarrow \mathbb{R}^N$  is a distributed control function that acts on the system through the control operator  $\mathcal{D}_\omega^h$ , which is the diagonal  $N \times N$  matrix depending on the control domain  $\omega$  and defined by

$$(\mathcal{D}_\omega^h)_{i,i} := \begin{cases} 1 & \text{if } x_i \in \omega, \\ 0 & \text{if } x_i \notin \omega. \end{cases} \quad (2.7)$$

- $V_b^h : (0, T) \rightarrow \mathbb{R}$  is a boundary control function that acts on the system on the right boundary of the interval  $\Omega$  (that is at  $x = 1$ ) through the control operator  $\mathcal{B}_r^h$ , which is the column vector defined by

$$\mathcal{B}_r^h = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{\gamma_{N+1/2}}{h_N h_{N+1/2}} \end{pmatrix}.$$

The particular form of this vector comes from the fact that, the matrix  $\mathcal{A}^h$  was designed in (2.3) for homogeneous Dirichlet boundary condition and thus, we need to add a source term in the last component of the equation to account for a non-homogeneous Dirichlet boundary datum (which is precisely the control  $V_b^h$  we are looking for).

Of course, the same analysis can be carried out for a boundary control acting on the left boundary of  $\Omega$ , that is at  $x = 0$  but we will not give the details here.

System (2.6) is a discrete version of the following parabolic PDE control problem

$$\begin{cases} \partial_t y + \mathcal{A}y = 1_\omega V_d(t, x), & \text{for } 0 < t \leq T \\ y(t, 0) = 0, \quad y(t, 1) = V_b(t), & \text{for } 0 < t \leq T, \\ y(0, \cdot) = y^0 \in L^2(\Omega). \end{cases}$$

The typical questions we are interested in concern null controllability and may be roughly formulated as follows (for a given set of data  $\gamma, q, \omega, T, \dots$ )

- *Distributed control* : Is there a constant  $C > 0$  such that for any  $h > 0$  small enough and any  $y^{0,h} \in \mathbb{R}^N$  we can find a control  $V_d^h \in L^2(0, T, \mathbb{R}^N)$  such that the solution of (2.6) (with  $V_b^h = 0$ ) satisfies  $y^h(T) = 0$  and moreover

$$\|V_d^h\|_{L^2(\Omega_h^T)} \leq C \|y^{0,h}\|_{L^2(\Omega_h)}.$$

- *Boundary control* : Is there a constant  $C > 0$  such that for any,  $h > 0$  small enough, and any  $y^{0,h} \in \mathbb{R}^N$ , we can find a control  $V_b^h \in L^2(0, T)$  such that the solution of (2.6) (with  $V_d^h = 0$ ) satisfies  $y^h(T) = 0$  and moreover

$$\|V_b^h\|_{L^2(0, T)} \leq C \|y^{0,h}\|_{L^2(\Omega_h)}.$$

When  $\gamma$  is constant,  $q = 0$  and for a uniform mesh, the boundary control case was positively solved in [19]. The authors use explicit formulas for  $\lambda_k^h$  and  $\phi_k^h$  in that case (see Remark 2.1.3) to tackle the problem of null controllability with the moments method. In the present work we propose to use the same kind of strategy but replacing the exact knowledge of the eigenelements of  $\mathcal{A}^h$  by the estimates discussed above.

However, in the general case, it turns out that exact null-controllability at the discrete level formulated above is a too strong notion. Indeed, it may happen in some situations that the discrete problems are not uniformly controllable (we mention here the counter-example in 2D given in [24] for which (2.6) is not even approximately controllable for a given  $h$ ).

Finally, from a practical point of view, to obtain a bounded family of discrete controls that converges (at least weakly) to a null-control for the initial PDE problem, it is enough to weaken the requirement  $y^h(T) = 0, \forall h > 0$  replacing it by the convergence  $\|y^h(T)\|_{L^2(\Omega_h)} \rightarrow 0$  at a prescribed rate when  $h \rightarrow 0$ . Such a weaker requirement was already studied for instance in [10–12, 17] (yet with various naming conventions). A specific discussion on that topic was given in [9], from which we borrow the following definition.

**Definition 2.1.1** *Let  $\phi : (0, +\infty) \rightarrow (0, +\infty)$  be a function such that  $\lim_{h \rightarrow 0} \phi(h) = 0$ . We say that the family of Problems (2.6) is uniformly  $\phi(h)$ -null controllable with a distributed control (resp. a*

boundary control) if there exists a  $C > 0$ , such that, for any  $h$  small enough and any  $y^{0,h} \in \mathbb{R}^N$ , we can find a control  $V_d^h \in L^2(0, T, \mathbb{R}^N)$  (resp.  $V_b^h \in L^2(0, T)$ ) that satisfies

$$\|V_d^h\|_{L^2(\Omega_T)} \leq C \|y^{0,h}\|_{L^2(\Omega_h)}, \quad (\text{resp. } \|V_b^h\|_{L^2(0,T)} \leq C \|y^{0,h}\|_{L^2(\Omega_h)}),$$

and such that the solution  $y^h$  of (2.6) with  $V_b^h = 0$  (resp.  $V_d^h = 0$ ) satisfies

$$\|y^h(T)\|_{L^2(\Omega_h)}^2 \leq C\phi(h) \|y^{0,h}\|_{L^2(\Omega_h)}^2.$$

It is shown in [10–12] that, for a finite difference approximation on smooth grids with smooth coefficients, the uniform  $\phi(h)$ -null controllability property holds for a distributed control as soon as  $\phi$  does not tend to zero faster than some exponential  $h \mapsto e^{-C/h^\alpha}$  (see also [17] where the case of Galerkin approximation is discussed). However, in those references, the methods which are used (based on discrete Carleman estimates) do not allow to tackle the case of coupled parabolic systems with a boundary control.

With the techniques developed in the present paper (however restricted to the 1D case), we are able to obtain similar results for the scalar equation (2.6) both for distributed and boundary control problems, as well as for the coupled systems presented below, which is, up to our knowledge, the first result in that framework.

A typical semi-discrete parabolic system with two components we shall deal with is the following

$$\begin{cases} (Y^h)'(t) + \begin{pmatrix} \mathcal{A}^h & 0^h \\ \mathbf{I}^h & \mathcal{A}^h \end{pmatrix} Y^h(t) = \begin{pmatrix} \mathbf{I}^h \\ 0^h \end{pmatrix} (\mathcal{D}_\omega^h V_d^h(t) + \mathcal{B}_r^h V_b^h(t)), \text{ for } 0 < t \leq T, \\ Y^h(0) = Y^{0,h} \in (\mathbb{R}^N)^2. \end{cases} \quad (S_2^h)$$

In this system,  $\mathbf{I}^h$  (resp.  $0^h$ ) is the identity matrix (resp. the zero matrix) of size  $N \times N$ . The unknown  $Y^h(t) \in (\mathbb{R}^N)^2$  has now two components. The first one satisfies the same scalar equation as before and is controlled either by a distributed control  $V_d^h$  or by a boundary control  $V_b^h$ , whereas the second component satisfies a parabolic equation without explicit appearance of a control; actually this component of the solution is indirectly controlled by the first component itself by means of the coupling term which is here set to 1 and which appears in the matrix  $\begin{pmatrix} \mathcal{A}^h & 0^h \\ \mathbf{I}^h & \mathcal{A}^h \end{pmatrix}$ . The controllability questions we are interested in are now exactly the same: is it possible to find a bounded family of controls  $(V_d^h)_h$  (resp.  $(V_b^h)_h$ ) such that the two components of the corresponding solution of  $(S_2^h)$  are small at time  $T$ , that is

$$\|Y^h(T)\|_{L^2(\Omega_h)}^2 \leq C\phi(h) \|Y^{0,h}\|_{L^2(\Omega_h)}^2,$$

where the notation for the norms  $\|\cdot\|_{L^2(\Omega_h)}$  are adapted to the fact that  $Y^h$  has now two components. The corresponding precise theorems are stated in Section 2.5.3.

To conclude this introduction, we mention that we do not address here the question of the actual computation of the controls. Even though we shall use the moments method in the analysis, it has to be precised that this is not at all a constructive method that can be efficiently implemented. Instead, once a theoretical uniform  $\phi(h)$ -null controllability result is obtained, we can implement the penalized HUM method that provides, in a more natural and generic way, discrete controls that fulfill the properties of definition 2.1.1. This approach is discussed for instance in [9, 16].

## 2.2 Spectral properties of elliptic operators. The continuous case

In this section we present a proof of Theorem 2.1.1 even though this result is not new by itself (see for instance [22]). Our goal is just to introduce a quite simple proof that we will manage to adapt to the discrete setting in the following sections. In all this section, we assume that hypothesis  $(H_0)$  holds.

**Lemma 2.2.1** *Let  $\omega$  be a non-empty open subset of  $\Omega$ . There exists  $C_1(q, \gamma) > 0$  and  $C_2(q, \gamma, \omega) > 0$  such that we have, for any  $k \geq 1$ ,*

$$\frac{1}{\lambda_k} |\partial_b \phi_k|^2 \geq C_1 \mathcal{R}_k, \quad \forall b \in \{l, r\},$$

and

$$\|\phi_k\|_{L^2(\omega)}^2 \geq C_2 \mathcal{R}_k,$$



where we have defined

$$\mathcal{R}_k := \inf_{x,y \in \Omega} \frac{|\phi_k(x)|^2 + \frac{\gamma(x)}{\lambda_k} |\phi_k'(x)|^2}{|\phi_k(y)|^2 + \frac{\gamma(y)}{\lambda_k} |\phi_k'(y)|^2}. \quad (2.8)$$

This lemma says, in particular, that to obtain (possibly uniform in  $k$ ) lower bounds on  $|\partial_b \phi_k|/\sqrt{\lambda_k}$  and  $\|\phi_k\|_{L^2(\omega)}$ , it is enough to obtain lower bounds on  $\mathcal{R}_k$ . Its proof is postponed at the end of the section.

Our strategy is based on interpreting the equation satisfied by eigenfunctions as a particular first order ordinary differential system in such a way that the principal part (with respect to the large values of  $\lambda_k$ ) of the evolution matrix is skew-symmetric. It will let us obtain suitable estimates on the quantity  $\mathcal{R}_k$  defined in the previous lemma. In this direction, we begin with the following lemma whose proof is a straightforward computation and is left to the reader.

**Lemma 2.2.2** *Let  $f : \Omega \rightarrow \mathbb{R}$  be a continuous function and  $\lambda > 0$ . Let  $u \in \mathcal{C}^2(\Omega)$  be a real valued function satisfying the second-order differential equation (without any assumption on boundary conditions)*

$$\mathcal{A}u(x) = \lambda u(x) + f(x), \quad \forall x \in \Omega, \quad (2.9)$$

then the following equation holds

$$U'(x) = M(x)U(x) + Q(x)U(x) + F(x), \quad (2.10)$$

where we have defined the vectors

$$U(x) := \begin{pmatrix} u(x) \\ \sqrt{\frac{\gamma(x)}{\lambda}} u'(x) \end{pmatrix} \quad \text{and} \quad F(x) := \begin{pmatrix} 0 \\ -\frac{f(x)}{\sqrt{\gamma(x)\lambda}} \end{pmatrix}.$$

and the matrices

$$M(x) := \begin{pmatrix} 0 & \sqrt{\frac{\lambda}{\gamma(x)}} \\ -\sqrt{\frac{\lambda}{\gamma(x)}} & 0 \end{pmatrix} \quad \text{and} \quad Q(x) := \begin{pmatrix} 0 & 0 \\ \frac{q(x)}{\sqrt{\lambda\gamma(x)}} & \sqrt{\gamma(x)} \left(\frac{1}{\sqrt{\gamma}}\right)'(x) \end{pmatrix}.$$

The key-point of this formulation is that the large terms in  $\sqrt{\lambda}$  only appear in the skew-symmetric matrix  $M(x)$ , while the matrix  $Q(x)$  only contains bounded terms with respect to  $\lambda$ .

As a consequence of this particular structure, we can obtain the following estimates.

**Lemma 2.2.3** *With the same notations as in Lemma 2.2.2, and assuming that  $\lambda \geq 1$ , there exists  $C := C(\gamma, q)$ , independent of  $\lambda$ , such that for any  $x, y \in \Omega$ , we have*

$$\|U(y)\| \leq C \left( \|U(x)\| + \left| \int_x^y \|F(s)\| ds \right| \right). \quad (2.11)$$

### Proof

Let  $x, y \in \Omega$ . It is fundamental to notice that the matrices  $(M(s))_s$  pairwise commute, so that the resolvent operator associated with  $x \mapsto M(x)$  simply reads

$$S(y, x) := \exp \left( \int_x^y M(s) ds \right).$$

We can then use Duhamel's formula to deduce from equation (2.10) the following expression

$$U(y) = S(y, x)U(x) + \int_x^y S(y, s) (Q(s)U(s) + F(s)) ds. \quad (2.12)$$

We use now the fact that the matrix  $M(s)$  is skew-symmetric for any  $s$ , and so is  $\int_x^y M(s) ds$ . It follows that the resolvent  $S(y, s)$  satisfies  $\|S(y, s)\| = 1$  for any  $y, s$ . We get

$$\|U(y)\| \leq \|U(x)\| + \left| \int_x^y \|F(s)\| ds \right| + \left| \int_x^y \|Q(s)\| \|U(s)\| ds \right|.$$

Gronwall's lemma finally yields

$$\|U(y)\| \leq \left( \|U(x)\| + \left| \int_x^y \|F(s)\| ds \right| \right) \exp \left( \left| \int_x^y \|Q(s)\| ds \right| \right),$$

which gives the result since  $Q(s)$  is bounded uniformly in  $s$  and  $\lambda$ , by using assumption  $(H_0)$ .

■ We can now prove Theorem 2.1.1. Note that it is enough to prove the claims for  $k$  large enough and in particular, by (2.2), we can assume without any loss of generality that  $\lambda_k \geq 1$ .

- We begin with the proof of points 2. and 3. of the theorem, that is the properties which concern the eigenfunctions  $(\phi_k)$ . By definition,  $\phi_k$  is a solution of the equation

$$\mathcal{A}\phi_k = \lambda_k\phi_k,$$

which is exactly (2.9) with  $u = \phi_k$ ,  $\lambda = \lambda_k$ ,  $f = 0$ . From Lemma 2.2.3 we deduce that there exists  $C := C(\gamma, q)$ , independent of  $k$ , such that for any  $x, y \in \bar{\Omega}$ ,

$$|\phi_k(y)|^2 + \frac{\gamma(y)}{\lambda_k} |\phi'_k(y)|^2 \geq C \left( |\phi_k(x)|^2 + \frac{\gamma(x)}{\lambda_k} |\phi'_k(x)|^2 \right), \quad (2.13)$$

which exactly proves that the quantity  $\mathcal{R}_k$  defined in (2.8) is uniformly bounded from below. The claim thus immediately follows from Lemma 2.2.1.

- We shall now prove the first point in Theorem 2.1.1. For any index  $k \geq 1$  with  $\lambda_k \geq 1$ , we define

$$u(x) := \phi'_k(1)\phi_{k+1}(x) - \phi'_{k+1}(1)\phi_k(x),$$

in such a way that  $u(1) = u'(1) = 0$  and

$$\mathcal{A}u = \lambda_{k+1}u + f,$$

with

$$f(x) := \phi'_{k+1}(1)\phi_k(x)(\lambda_{k+1} - \lambda_k).$$

Using the notations introduced in Lemma 2.2.2, we observe that by construction we have  $U(1) = 0$  so that the estimate (2.11) specialized in  $x = 1$  leads to

$$\|U(y)\| \leq C \int_y^1 \|F(s)\| ds \leq C \int_0^1 \|F(s)\| ds, \quad \forall y \in \Omega.$$

Using the expression for  $F$  and  $f$ , we find that

$$\|U(y)\| \leq \frac{C}{\sqrt{\gamma_{min}}} \left( \frac{\lambda_{k+1} - \lambda_k}{\sqrt{\lambda_{k+1}}} |\phi'_{k+1}(1)| \right) \int_0^1 |\phi_k(s)| ds, \quad \forall y \in \Omega.$$

Thanks to the normalization condition  $\|\phi_k\|_{L^2(\Omega)} = 1$  and the expressions of  $U$  and  $u$ , we obtain for any  $y \in \Omega$ ,

$$|\phi'_k(1)\phi_{k+1}(y) - \phi'_{k+1}(1)\phi_k(y)|^2 \leq \frac{C}{\gamma_{min}} \left( \frac{\lambda_{k+1} - \lambda_k}{\sqrt{\lambda_{k+1}}} |\phi'_{k+1}(1)| \right)^2.$$

We integrate this inequality with respect to  $y \in (0, 1)$  and we use the  $L^2(\Omega)$  orthonormality of  $\phi_{k+1}$  and  $\phi_k$  to finally get

$$|\phi'_{k+1}(1)|^2 \leq (\phi'_{k+1}(1))^2 + (\phi'_k(1))^2 \leq \frac{C}{\gamma_{min}} \left( \frac{\lambda_{k+1} - \lambda_k}{\sqrt{\lambda_{k+1}}} |\phi'_{k+1}(1)| \right)^2,$$

and since  $\phi'_{k+1}(1) \neq 0$ , we conclude that

$$\lambda_{k+1} - \lambda_k \geq \bar{C} \sqrt{\lambda_{k+1}},$$

for some  $\bar{C}(\gamma, q) > 0$  independent of  $k$ . The claim follows by the lower bound in (2.2).

**Remark 2.2.1** Integrating (2.13) with respect to  $y$  over  $\Omega$  and using the min-max principle, we can easily obtain optimal  $L^\infty$  bounds on the eigenfunctions

$$\|\phi_k\|_{L^\infty(\Omega)} \leq C, \quad \|\phi'_k\|_{L^\infty(\Omega)} \leq C\sqrt{\lambda_k}, \quad \forall k \geq 1.$$

**Remark 2.2.2** Observe that the only point where the regularity of  $\gamma$  is used in this proof is in the uniform estimate of  $\|Q(s)\|$ . However, the only thing we need is that  $\int_\Omega \|Q(s)\| ds < +\infty$ , which proves that Theorem 2.1.1 still holds when  $\gamma$  is piecewise  $C^1$ .

It remains to give the proof of the lemma. Using Caccioppoli-like inequality or nodal sets of eigenfunctions of Sturm-Liouville operators, one can give a more direct proof of this lemma. The advantage of the proof given here is that it can be more easily extended to the discrete setting (see Section 2.3.4).

**Proof** [of Lemma 2.2.1]

From (2.8) we can write

$$\mathcal{R}_k \left( |\phi_k(y)|^2 + \frac{\gamma(y)}{\lambda_k} |\phi'_k(y)|^2 \right) \leq |\phi_k(x)|^2 + \frac{\gamma(x)}{\lambda_k} |\phi'_k(x)|^2, \quad \forall x, y \in \Omega. \quad (2.14)$$

We recall that  $\phi_k$  satisfies the Dirichlet boundary condition  $\phi_k(0) = \phi_k(1) = 0$ , so that by choosing  $x = 0$  and  $x = 1$  in the previous inequality, we obtain in particular

$$\frac{\gamma(0)}{\lambda_k} |\partial_l \phi_k|^2 \geq \mathcal{R}_k |\phi_k(y)|^2, \quad \forall y \in \Omega,$$

and

$$\frac{\gamma(1)}{\lambda_k} |\partial_r \phi_k|^2 \geq \mathcal{R}_k |\phi_k(y)|^2, \quad \forall y \in \Omega.$$

Since the left-hand side does not depend on  $y$ , and thanks to the normalization condition  $\|\phi_k\|_{L^2(\Omega)} = 1$ , we can integrate those inequalities with respect to  $y$  to obtain

$$\frac{\gamma(0)}{\lambda_k} |\partial_l \phi_k|^2 \geq \mathcal{R}_k, \quad \text{and} \quad \frac{\gamma(1)}{\lambda_k} |\partial_r \phi_k|^2 \geq \mathcal{R}_k.$$

Consider now any non empty open subset  $\omega \subset \Omega$ . Without any loss of generality we can assume that  $\omega$  is a non-empty interval. Split  $\omega$  in three consecutive disjoint intervals of identical measure :  $\omega_1, \omega_2$  and  $\omega_3$ .

— Suppose that there exists  $a_k \in \omega_1$  and  $b_k \in \omega_3$  such that

$$\phi_k(a_k)\phi'_k(a_k) \geq 0 \quad \text{and} \quad \phi_k(b_k)\phi'_k(b_k) \leq 0. \quad (2.15)$$

We multiply the equation  $\mathcal{A}\phi_k = \lambda_k\phi_k$  by  $\phi_k$  and integrate by parts on  $(a_k, b_k)$ . We get

$$-\underbrace{[\gamma(x)\phi'_k(x)\phi_k(x)]_{a_k}^{b_k}}_{\leq 0} + \int_{a_k}^{b_k} \left( \gamma(x) (\phi'_k(x))^2 + q(x)\phi_k^2(x) \right) dx = \lambda_k \int_{a_k}^{b_k} \phi_k^2(x) dx$$

thus,

$$\int_{a_k}^{b_k} \gamma(x) (\phi'_k(x))^2 dx \leq (\lambda_k + \|q\|_\infty) \int_{a_k}^{b_k} \phi_k^2(x) dx. \quad (2.16)$$

We come back to (2.14) that we integrate on the whole domain  $\Omega$  with respect to the variable  $y$  and on the interval  $(a_k, b_k)$  with respect to the variable  $x$ . With the normalization condition on  $\phi_k$  and (2.16) we obtain:

$$\mathcal{R}_k(b_k - a_k) \leq \int_{a_k}^{b_k} \phi_k^2(x) + \frac{\gamma(x)}{\lambda_k} (\phi'_k(x))^2 dx \leq \left( 2 + \frac{\|q\|_\infty}{\lambda_1} \right) \int_{a_k}^{b_k} \phi_k^2(x) dx.$$

Using that  $a_k \in \omega_1$  and  $b_k \in \omega_3$  we get  $b_k - a_k \geq |\omega_2| = \frac{|\omega|}{3}$  and we finally obtain the claimed lower bound

$$\mathcal{R}_k \frac{|\omega|}{3} \leq \mathcal{R}_k(b_k - a_k) \leq C \int_\omega \phi_k^2(x) dx.$$

— Suppose now that either

$$\forall x \in \omega_1, \phi_k(x)\phi'_k(x) < 0 \quad (2.17)$$

or

$$\forall x \in \omega_3, \phi_k(x)\phi'_k(x) > 0, \quad (2.18)$$

hold. We only consider the second case (2.18), since the case (2.17) is similar.

**Remark 2.2.3** Notice that from condition (2.18), we can deduce that either  $\phi_k$  is positive and increasing on  $\omega_3$  or  $\phi_k$  is negative and decreasing on  $\omega_3$ . Note also that this condition implies that  $x \rightarrow |\phi|(x)$  is strictly increasing on  $\omega_3$ .

We split  $\omega_3$  in two consecutive subintervals  $\omega_{3,1}$  and  $\omega_{3,2}$  of identical measure.

*First case :* Suppose that there exists  $x_0 \in \omega_{3,1}$  such that  $\frac{\gamma(x_0)}{\lambda_k} |\phi'_k(x_0)|^2 \leq \frac{\mathcal{R}_k}{2}$ . This assumption and inequality (2.14) give

$$|\phi_k(x_0)|^2 + \frac{\mathcal{R}_k}{2} \geq |\phi_k(x_0)|^2 + \frac{\gamma(x_0)}{\lambda_k} |\phi'_k(x_0)|^2 \geq \mathcal{R}_k |\phi_k(y)|^2, \forall y \in \Omega,$$

and integrating on  $\Omega$  with respect to the variable  $y$ , we get:

$$|\phi_k(x_0)|^2 \geq \frac{\mathcal{R}_k}{2}.$$

According to Remark 2.2.3,  $|\phi|^2$  is increasing on  $\omega_3$ . Thus, integrating on  $\omega_{3,2}$ , we obtain

$$\int_{\omega_{3,2}} |\phi_k(x)|^2 dx \geq C \mathcal{R}_k |\omega|.$$

*Second case :* Suppose now that  $\forall x_0 \in \omega_{3,1}, \frac{\gamma(x_0)}{\lambda_k} |\phi'_k(x_0)|^2 > \frac{\mathcal{R}_k}{2}$ . Let us integrate the square root of this inequality on  $\omega_{3,1}$

$$\int_{\omega_{3,1}} |\phi'_k(x_0)| dx_0 > C \sqrt{\lambda_k \mathcal{R}_k} |\omega_{3,1}|.$$

According to Remark 2.2.3,  $\phi_k$  is monotonic on  $\omega_3$  so  $\int_{\omega_{3,1}} |\phi'_k(x_0)| dx_0 = |\int_{\omega_{3,1}} \phi'_k(x_0) dx_0|$ . Denote by  $a_{3,1} < b_{3,1}$  the end points of  $\omega_{3,1}$ . We have

$$|\phi_k(b_{3,1}) - \phi_k(a_{3,1})| > C \sqrt{\lambda_k \mathcal{R}_k} |\omega_{3,1}|,$$

and thus, since  $|\phi_k|$  is increasing,

$$2|\phi_k(b_{3,1})| > C \sqrt{\lambda_k \mathcal{R}_k} |\omega_{3,1}|.$$

Hence,

$$\int_{\omega} |\phi_k|^2(x) dx \geq \int_{\omega_{3,2}} |\phi_k|^2(x) dx \geq |\phi_k(b_{3,1})|^2 |\omega_{3,2}| \geq C \lambda_k \mathcal{R}_k |\omega|^3,$$

which concludes the proof.  $\blacksquare$

## 2.3 Discrete spectral properties

### 2.3.1 General strategy

As previously stated, our goal here is to prove qualitative properties for the eigenlements of the discrete elliptic operator  $\mathcal{A}^h$ . We shall adapt Theorem 2.1.1 to the discrete setting by mimicking the proof given in Section 2.2. We start by the following preliminary result. Its proof is postponed to Section 2.3.4. We recall that we have assumed that  $q \geq 0$ .

**Lemma 2.3.1** *Assume that  $(\mathbf{H}_0)$  holds. Let  $\omega$  be a non-empty open subset of  $\Omega$  and  $\beta > 0$ . There exist  $h_1(q, \gamma, \beta) > 0$ ,  $h_2(q, \gamma, \omega, \beta) > 0$  and  $C_1(q, \gamma, \beta) > 0$ ,  $C_2(q, \gamma, \omega, \beta) > 0$  such that for any mesh of  $\Omega$  such that  $\Theta_h \leq \beta$ , we have*

$$\frac{1}{\lambda_k^h} |\partial_b \phi_k^h|^2 \geq C_1 \mathcal{R}_k^h, \quad \forall b \in \{l, r\}, \quad \forall k \in \llbracket 1, N \rrbracket,$$

if  $h \leq h_1$  and

$$\|\phi_k^h\|_{L^2(\omega_h)}^2 \geq C_2 \mathcal{R}_k^h, \quad \forall k \in \llbracket 1, N \rrbracket,$$

if  $h \leq h_2$ , with  $\mathcal{R}_k^h$  given by

$$\mathcal{R}_k^h := \min_{i,j \in \llbracket 1, N+1 \rrbracket} \frac{|\phi_{i,k}^h|^2 + \frac{\gamma_{i-1/2}}{\lambda_k^h} \left| \frac{\phi_{i,k}^h - \phi_{i-1,k}^h}{h_{i-1/2}} \right|^2}{|\phi_{j,k}^h|^2 + \frac{\gamma_{j-1/2}}{\lambda_k^h} \left| \frac{\phi_{j,k}^h - \phi_{j-1,k}^h}{h_{j-1/2}} \right|^2}. \quad (2.19)$$

This lemma shows that, to obtain (possibly uniform in  $k$  and  $h$ ) lower bounds on  $|\partial_b \phi_k^h|/\sqrt{\lambda_k^h}$  and  $\|\phi_k^h\|_{L^2(\omega_h)}$  it is enough to obtain lower bounds on  $\mathcal{R}_k^h$ .

To this end, we propose the following transformation of the discrete second order equation into a suitable first order system.

**Lemma 2.3.2** *We consider a given mesh of  $\Omega$ ,  $f^h$  a vector of  $\mathbb{R}^N$ ,  $\lambda > 0$  and we assume that  $u^h \in \mathbb{R}^N$  satisfies the discrete equation*

$$\mathcal{A}^h u^h = \lambda u^h + f^h. \quad (2.20)$$

Then, the following relation holds for any  $j \in \llbracket 1, N \rrbracket$

$$U_{j+1}^h = (\mathbf{I} + h_j M_j^h) U_j^h + h_j (Q_j^h U_j^h + F_j^h), \quad (2.21)$$

where we have defined the families of vectors

$$U_j^h := \begin{pmatrix} u_j^h \\ \sqrt{\frac{\gamma_{j-1/2}}{\lambda}} \frac{u_j^h - u_{j-1}^h}{h_{j-1/2}} \end{pmatrix} \quad \text{and} \quad F_j^h := \begin{pmatrix} -h_{j+1/2} \frac{f_j^h}{\gamma_{j+1/2}} \\ -\frac{f_j^h}{\sqrt{\gamma_{j+1/2} \lambda}} \end{pmatrix},$$

as well as the families of matrices

$$M_j^h := \begin{pmatrix} -h_{j-1/2} \frac{\lambda}{\gamma_{j+1/2}} & \frac{h_{j+1/2}}{h_j} \sqrt{\frac{\lambda}{\gamma_{j+1/2}}} \\ -\sqrt{\frac{\lambda}{\gamma_{j+1/2}}} & 0 \end{pmatrix}, \quad (2.22)$$

and

$$Q_j^h := \begin{pmatrix} h_{j+1/2} \frac{q_j}{\gamma_{j+1/2}} & h_{j+1/2} \sqrt{\frac{\lambda \gamma_{j-1/2}}{\gamma_{j+1/2}} \frac{\frac{1}{\sqrt{\gamma_{j+1/2}}} - \frac{1}{\sqrt{\gamma_{j-1/2}}}}{h_j}} \\ \frac{q_j}{\sqrt{\lambda \gamma_{j+1/2}}} & \sqrt{\frac{\lambda \gamma_{j-1/2}}{\gamma_{j+1/2}} \frac{\frac{1}{\sqrt{\gamma_{j+1/2}}} - \frac{1}{\sqrt{\gamma_{j-1/2}}}}{h_j}} \end{pmatrix}.$$

**Remark 2.3.1** *Note that the matrix  $(\mathbf{I} + h_j M_j^h)$  is a one step propagator which is the discrete analogue of  $\exp\left(\int_{x_j}^{x_{j+1}} M(s) ds\right)$ .*

We shall make use of the following estimates.

**Lemma 2.3.3** *For any mesh of size  $N$  and any  $\gamma, q$  satisfying  $(\mathbf{H}_0)$ , we have the inequalities*

$$\frac{\gamma_{\min}^{\min}}{(\Theta_h)^2} \lambda_k^{h*} - \|q\|_{L^\infty} \leq \lambda_k^h \leq (\Theta_h)^2 \|\gamma\|_{L^\infty} \lambda_k^{h*} + \|q\|_{L^\infty}, \quad \forall k \in \llbracket 1, N \rrbracket,$$

where  $\lambda_k^{h*}$  was defined in Remark 2.1.3. In particular, for any  $\beta > 0$ , there exists  $C_1, C_2, C_3 > 0$  depending only on  $\gamma, q$  and  $\beta$  such that for any mesh satisfying  $\Theta_h \leq \beta$  we have the inequalities

$$\begin{aligned} C_1 k^2 - C_2 &\leq \lambda_k^h \leq C_3 k^2 + C_2, \quad \forall k \in \llbracket 1, N \rrbracket, \\ h^2 \lambda_k^h &\leq C_2, \quad \forall k \in \llbracket 1, N \rrbracket. \end{aligned}$$

The proof relies on Remark 2.1.3 and the min-max principle. It is a straightforward computation that we leave to the reader. Those estimates can be seen as discrete versions of (2.2).

**Proof** [of lemma 2.3.2] The proof of (2.21) is a tedious but simple computation. The second component of  $U_{j+1}^h - U_j^h$  satisfies

$$\begin{aligned} & \sqrt{\frac{\gamma_{j+1/2}}{\lambda}} \frac{u_{j+1}^h - u_j^h}{h_{j+1/2}} - \sqrt{\frac{\gamma_{j-1/2}}{\lambda}} \frac{u_j^h - u_{j-1}^h}{h_{j-1/2}} \\ &= \frac{1}{\sqrt{\lambda \gamma_{j+1/2}}} \left( \gamma_{j+1/2} \frac{u_{j+1}^h - u_j^h}{h_{j+1/2}} - \gamma_{j-1/2} \frac{u_j^h - u_{j-1}^h}{h_{j-1/2}} \right) \\ & \quad + \frac{u_j^h - u_{j-1}^h}{h_{j-1/2}} \sqrt{\frac{\gamma_{j-1/2}}{\lambda}} \sqrt{\gamma_{j-1/2}} \left( \frac{1}{\sqrt{\gamma_{j+1/2}}} - \frac{1}{\sqrt{\gamma_{j-1/2}}} \right) \\ &= -h_j \left( \frac{\lambda - q_j}{\sqrt{\lambda \gamma_{j+1/2}}} \right) u_j^h - h_j \frac{f_j^h}{\sqrt{\lambda \gamma_{j+1/2}}} \\ & \quad + \frac{u_j^h - u_{j-1}^h}{h_{j-1/2}} \sqrt{\frac{\gamma_{j-1/2}}{\lambda}} h_j \sqrt{\gamma_{j-1/2}} \left( \frac{\frac{1}{\sqrt{\gamma_{j+1/2}}} - \frac{1}{\sqrt{\gamma_{j-1/2}}}}{h_j} \right), \end{aligned}$$

where we used the original equation (2.20) in the last equality, whereas for the first component, we simply write

$$u_{j+1}^h - u_j^h = \left( h_{j+1/2} \sqrt{\frac{\lambda}{\gamma_{j+1/2}}} \right) \sqrt{\frac{\gamma_{j+1/2}}{\lambda}} \frac{u_{j+1}^h - u_j^h}{h_{j+1/2}}.$$

It follows, in vectorial form, the following equality

$$\begin{aligned} U_{j+1}^h - U_j^h &= \begin{pmatrix} 0 & h_{j+1/2} \sqrt{\frac{\lambda}{\gamma_{j+1/2}}} \\ 0 & 0 \end{pmatrix} U_{j+1}^h + \begin{pmatrix} 0 & 0 \\ -h_j \sqrt{\frac{\lambda}{\gamma_{j+1/2}}} & 0 \end{pmatrix} U_j^h \\ & \quad + \begin{pmatrix} 0 & 0 \\ h_j \frac{q_j}{\sqrt{\lambda \gamma_{j+1/2}}} & h_j \sqrt{\gamma_{j-1/2}} \left( \frac{\frac{1}{\sqrt{\gamma_{j+1/2}}} - \frac{1}{\sqrt{\gamma_{j-1/2}}}}{h_j} \right) \end{pmatrix} U_j^h \\ & \quad + \begin{pmatrix} 0 & 0 \\ -h_j \frac{f_j^h}{\sqrt{\lambda \gamma_{j+1/2}}} & 0 \end{pmatrix}. \end{aligned}$$

This is not yet the required form since  $U_{j+1}^h$  appears in the right-hand side. Thus, we collect the terms in  $U_{j+1}^h$  in the left-hand side as follows

$$\begin{aligned} & \begin{pmatrix} 1 & -h_{j+1/2} \sqrt{\frac{\lambda}{\gamma_{j+1/2}}} \\ 0 & 1 \end{pmatrix} U_{j+1}^h = \begin{pmatrix} 1 & 0 \\ -h_j \sqrt{\frac{\lambda}{\gamma_{j+1/2}}} & 1 \end{pmatrix} U_j^h \\ & \quad + \begin{pmatrix} 0 & 0 \\ h_j \frac{q_j}{\sqrt{\lambda \gamma_{j+1/2}}} & h_j \sqrt{\gamma_{j-1/2}} \left( \frac{\frac{1}{\sqrt{\gamma_{j+1/2}}} - \frac{1}{\sqrt{\gamma_{j-1/2}}}}{h_j} \right) \end{pmatrix} U_j^h + \begin{pmatrix} 0 & 0 \\ -h_j \frac{f_j^h}{\sqrt{\lambda \gamma_{j+1/2}}} & 0 \end{pmatrix}, \end{aligned}$$

and we get the claim by multiplying the previous equality by

$$\begin{pmatrix} 1 & h_{j+1/2} \sqrt{\frac{\lambda}{\gamma_{j+1/2}}} \\ 0 & 1 \end{pmatrix},$$

which is nothing but the inverse of the  $2 \times 2$  matrix appearing in the left-hand side term.

For any  $j, k \in \llbracket 1, N \rrbracket$ , we denote by  $Q_{j,k}^h, M_{j,k}^h$  the matrices as defined in the previous lemma for the particular value  $\lambda = \lambda_k^h$ . Using the uniform bounds on the discrete eigenvalues obtained in Lemma 2.3.3, one can easily deduce that there exists a  $C(q, \gamma, \Theta_h) > 0$  but independent of  $h$  such that

$$\|Q_{j,k}^h\| \leq C, \text{ and } \|M_{j,k}^h\| \leq C \sqrt{\lambda_k^h}, \quad \forall h > 0, \forall j, k \in \llbracket 1, N \rrbracket, \quad (2.23)$$

and moreover, if  $\gamma$  is assumed to be constant then we have the additional estimate

$$\|Q_{j,k}^h\| \leq C \left( h_{j+1/2} + \frac{1}{\sqrt{\lambda_k^h}} \right), \quad \forall h > 0, \forall j, k \in \llbracket 1, N \rrbracket. \quad (2.24)$$

### 2.3.2 Estimates in the general case (S1)

We are now in position to prove our first results in the more general case we shall consider, that is the setting (S1).

**Theorem 2.3.1** *Let  $\gamma$  and  $q$  satisfying (H<sub>0</sub>),  $\omega$  a non empty open subset of  $\Omega$  and  $\beta > 0$ . There exists  $h_1(q, \gamma, \beta) > 0$ ,  $h_2(q, \gamma, \omega, \beta) > 0$  and  $C(q, \gamma, \beta) > 0$ ,  $C_1(q, \gamma, \beta) > 0$ ,  $C_2(q, \gamma, \omega, \beta) > 0$  such that the following lower bounds hold for any mesh of  $\Omega$  such that  $\Theta_h \leq \beta$*

$$\frac{1}{\sqrt{\lambda_k^h}} |\partial_b \phi_k^h| \geq C_1 e^{-C\sqrt{\lambda_k^h}}, \quad \forall k \in \llbracket 1, N \rrbracket, \forall b \in \{l, r\},$$

as soon as  $h \leq h_1$  and

$$\|\phi_k^h\|_{L^2(\omega_h)}^2 \geq C_2 e^{-C\sqrt{\lambda_k^h}}, \quad \forall k \in \llbracket 1, N \rrbracket,$$

as soon as  $h \leq h_2$ .

Moreover, if  $\gamma$  and  $q$  satisfy hypothesis (H<sub>2</sub>), there exists  $C_3(q, \gamma, \beta) > 0$  and  $\alpha(q, \gamma, \beta) > 0$  such that

$$\lambda_{k+1}^h - \lambda_k^h \geq C_3 k, \quad \forall k \in \llbracket 1, \alpha N^{2/5} \rrbracket.$$

**Proof** Let  $k \in \llbracket 1, N \rrbracket$ . We apply Lemma 2.3.2 with  $u^h := \phi_k^h = \left( \phi_{p,k}^h \right)_{p \in \llbracket 1, N \rrbracket}$ ,  $\lambda = \lambda_k^h$  and  $f^h = 0$ .

The vectors  $U_j^h$  are then denoted by  $U_{j,k}^h$  to keep track of the dependence on  $k$ . Moreover, with those notations, the quantity  $\mathcal{R}_k^h$  defined in (2.19) can be expressed as

$$\mathcal{R}_k^h = \inf_{i,j \in \llbracket 1, N+1 \rrbracket} \frac{\|U_{i,k}^h\|^2}{\|U_{j,k}^h\|^2}. \quad (2.25)$$

Therefore, to obtain the announced lower bound, we only have to obtain a bound of  $\|U_{j,k}^h\|$  by  $\|U_{i,k}^h\|$  for any two indices  $i, j$ .

From the estimate (2.23) and the standard inequality  $1 + t \leq e^t, \forall t \geq 0$ , we find the upper-bound

$$\|I + h_j M_{j,k}^h + h_j Q_{j,k}^h\| \leq \exp \left( h_j C \sqrt{\lambda_k^h} \right), \quad \forall j \in \llbracket 1, N \rrbracket.$$

Direct computations, similar to the ones in the proof Lemma 2.3.2, also show that

$$\| (I + h_j M_{j,k}^h + h_j Q_{j,k}^h)^{-1} \| \leq \exp \left( h_j C \sqrt{\lambda_k^h} \right), \quad \forall j \in \llbracket 1, N \rrbracket.$$

Then, using (2.21) with  $f^h = 0$ , we immediately obtain

$$\exp \left( -C \sqrt{\lambda_k^h} \right) \|U_{j,k}^h\| \leq \|U_{i,k}^h\|, \quad \forall i, j \in \llbracket 1, N+1 \rrbracket.$$

By (2.25) and Lemma 2.3.1, the claim follows.

We now prove the gap property for a part of the spectrum. To this end, we will use Proposition 2.3.1 below. Take  $k$  a positive integer such that  $k \leq \alpha N^{2/5}$ , where  $\alpha(q, \gamma, \beta) > 0$  will be chosen later on. Then,  $\exists \bar{C}(q, \gamma, \beta) > 0$ , such that

$$\begin{aligned} \lambda_{k+1}^h - \lambda_k^h &\geq \lambda_{k+1} - \lambda_k - |\lambda_{k+1}^h - \lambda_{k+1}| - |\lambda_k^h - \lambda_k| \\ &\geq \bar{C}k - Ch^2 \lambda_k^3 \\ &\geq \bar{C}k - Ch^2 k^6, \end{aligned}$$

and the last inequality comes from (2.2). Now since

$$h^2 k^6 = (h^2 k^5)k \leq (h^2 \alpha^5 N^2)k \leq \Theta_h^2 \alpha^5 k \leq \beta^2 \alpha^5 k,$$

we can write

$$\lambda_{k+1}^h - \lambda_k^h \geq \bar{C}k - C\beta^2 \alpha^5 k.$$

Now take  $\alpha$  such that  $\bar{C} \geq 2C\beta^2 \alpha^5$  to get

$$\lambda_{k+1}^h - \lambda_k^h \geq \frac{\bar{C}}{2}k.$$

This concludes the proof.  $\blacksquare$

We have used above the following somehow standard numerical analysis result, whose proof is postponed in Appendix in section 2.7.2. It can be obtained by similar techniques as in [8] and [23] and using the convergence analysis of finite difference methods on general grids given in [7].

**Proposition 2.3.1** *Assume that  $q$  and  $\gamma$  satisfy (H<sub>2</sub>). For any  $\beta > 0$ , there exists  $C(\gamma, q, \beta) > 0$ , such that for any mesh satisfying  $\Theta_h \leq \beta$ , we have*

$$|\lambda_k^h - \lambda_k| \leq Ch^2 \lambda_k^3, \forall k \in \llbracket 1, N \rrbracket.$$

### 2.3.3 Estimates for of a uniform grid, cases (S2) and (S3)

In the previous subsection, the exponentially small lower bound comes from the crude estimate of the norm of the propagation matrix  $I + h_j M_{j,k}^h$  by  $1 + Ch_j \sqrt{\lambda_k^h}$ . Besides, as noticed in Remark 2.3.1, the counter-part in the continuous case of this propagation matrix is  $\exp(\int_{x_j}^{x_{j+1}} M(s) ds)$  which is a unitary matrix (see the proof of Lemma 2.2.3).

Thus, our objective is now to improve those estimates in the case of a uniform mesh.

To this end, we introduce the discrete resolvent  $S_{i \leftarrow j, k}^h$  defined by

$$S_{i \leftarrow j, k}^h := \begin{cases} (I + hM_{i-1, k}^h)(I + hM_{i-2, k}^h) \cdots (I + hM_{j, k}^h) & \text{for } i > j \\ I & \text{for } i = j \\ (S_{j \leftarrow i, k}^h)^{-1} & \text{for } i < j \end{cases}$$

and for which we will be able to prove sharper estimates. Observe that each matrix  $I + hM_{i, k}^h$  is invertible since its determinant is equal to one for a uniform grid (see (2.22)).

**Lemma 2.3.4** *With the same notations as in Lemma 2.3.2, there exists  $C(q, \gamma) > 0$ , and  $h_0(q, \gamma) > 0$ , such that for any  $h < h_0$ , and any  $i, j \in \llbracket 1, N+1 \rrbracket$ ,*

$$\begin{aligned} \|U_{i, k}^h\| &\leq C \exp \left( C \sum_{p \in \llbracket i, j \rrbracket \setminus \{i\}} h \|S_{i \leftarrow p+1, k}^h\| \|Q_{p, k}^h\| \right) \\ &\quad \times \left[ \|S_{i \leftarrow j, k}^h\| \|U_{j, k}^h\| + \sum_{p \in \llbracket i, j \rrbracket \setminus \{i\}} h \|S_{i \leftarrow p+1, k}^h\| \|F_{p, k}^h\| \right]. \end{aligned} \quad (2.26)$$

**Proof** Let  $i, j \in \llbracket 1, N+1 \rrbracket$ . From (2.21), discrete Duhamel formula gives

$$\text{for } i > j: \quad U_{i, k}^h = S_{i \leftarrow j, k}^h U_{j, k}^h + h \sum_{p=j}^{i-1} S_{i \leftarrow p+1, k}^h (Q_{p, k}^h U_{p, k}^h + F_{p, k}^h), \quad (2.27)$$

$$\text{for } i < j: \quad U_{i, k}^h = S_{i \leftarrow j, k}^h U_{j, k}^h - h \sum_{p=i}^{j-1} S_{i \leftarrow p+1, k}^h (Q_{p, k}^h U_{p, k}^h + F_{p, k}^h). \quad (2.28)$$



So for any  $i, j \in \llbracket 1, N+1 \rrbracket$ ,

$$\|U_{i,k}^h\| \leq \|S_{i \leftarrow j,k}^h\| \|U_{j,k}^h\| + h \sum_{p \in \llbracket i,j \rrbracket} \|S_{i \leftarrow p+1,k}^h\| (\|Q_{p,k}^h\| \|U_{p,k}^h\| + \|F_{p,k}^h\|).$$

Using the bound (2.23) on  $\|Q_{i,k}^h\|$ ,

$$\|U_{i,k}^h\| (1 - Ch \|S_{i \leftarrow i+1,k}^h\|) \leq \|S_{i \leftarrow j,k}^h\| \|U_{j,k}^h\| + h \sum_{p \in \llbracket i,j \rrbracket \setminus \{i\}} \|S_{i \leftarrow p+1,k}^h\| (\|Q_{p,k}^h\| \|U_{p,k}^h\| + \|F_{p,k}^h\|).$$

Note also that  $\|S_{i \leftarrow i+1,k}^h\| = \|\mathbf{I} + hM_{i,k}^h\| \leq 1 + Ch\sqrt{\lambda_k^h}$ . Thus, by Lemma 2.3.3 and Remark 2.1.3 (remember that the mesh is assumed to be uniform), for  $h$  small enough and some  $C(q, \gamma) > 0$ , we have

$$\|U_{i,k}^h\| \leq C \|S_{i \leftarrow j,k}^h\| \|U_{j,k}^h\| + Ch \sum_{p \in \llbracket i,j \rrbracket \setminus \{i\}} \|S_{i \leftarrow p+1,k}^h\| (\|Q_{p,k}^h\| \|U_{p,k}^h\| + \|F_{p,k}^h\|).$$

We conclude with discrete Gronwall's Lemma.

■

We recall that in the continuous case we had the fundamental property that  $\|S(x_0, x)\| = 1$ . Here we will not obtain that the discrete resolvent  $S_{j \leftarrow i,k}^h$  is unitary but we are able to produce uniform bounds on this object at least for a well chosen portion of the discrete spectrum.

Suppose that  $q$  and  $\gamma$  satisfy assumption (H<sub>0</sub>). For any  $\varepsilon \geq 0$ , we define the following integer

$$k_{max,\varepsilon}^h := \max \left\{ k \in \llbracket 1, N \rrbracket; \lambda_k^h < \frac{4}{h^2} \gamma_{min} (1 - \varepsilon) \right\}. \quad (2.29)$$

**Proposition 2.3.2** *Let  $q$  and  $\gamma$  satisfying (H<sub>0</sub>). There exists  $h_0(q, \gamma) > 0$  and  $C(q, \gamma) > 0$  such that, for any  $\varepsilon > 0$  and any uniform mesh of size  $h < h_0$ , we have the estimate*

$$\|S_{i \leftarrow j,k}^h\| \leq C \exp\left(\frac{C}{\varepsilon}\right), \quad \forall i, j \in \llbracket 1, N+1 \rrbracket, \quad \forall k \in \llbracket 1, k_{max,\varepsilon}^h \rrbracket. \quad (2.30)$$

If we additionally assume that the diffusion coefficient  $\gamma$  is constant, then the following estimate holds for the whole spectrum of  $\mathcal{A}^h$ , for any  $h < h_0$

$$\|S_{i \leftarrow j,k}^h\| \leq \frac{C}{\sqrt{1 - \frac{h^2}{4\gamma}(\lambda_k^h - \|q\|_{L^\infty})}}, \quad \forall i, j \in \llbracket 1, N+1 \rrbracket, \quad \forall k \in \llbracket 1, N \rrbracket. \quad (2.31)$$

Note that, by Lemma 2.3.3 and Remark 2.1.3, there exists a  $\alpha(q, \gamma) > 0$  such that

$$k_{max,\varepsilon}^h \geq \alpha N \sqrt{1 - \varepsilon}, \quad \forall h > 0.$$

Thus, inequality (2.30) in the previous proposition gives actually a uniform estimate of  $S_{\bullet \leftarrow \bullet, k}^h$  for (at least) a constant portion of the eigenlements of the discrete operator  $\mathcal{A}^h$ .

**Proof** Since  $S_{i \leftarrow j,k}^h U_{j,k}^h = U_{i,k}^h$ , the study of the norm of  $S_{i \leftarrow j,k}^h$  amounts in obtaining a bound on  $U_{i,k}^h$  in function of  $U_{j,k}^h$ , for any  $i, j$ , where  $U_{\bullet,k}^h$  is any solution of the recurrence relation

$$U_{i+1,k}^h = (\mathbf{I} + hM_{i,k}^h) U_{i,k}^h, \quad \forall i \in \llbracket 1, N+1 \rrbracket. \quad (2.32)$$

To simplify the notation in this proof, we shall sometimes drop the indices  $h$  and  $k$  but we keep in mind that the condition  $k \leq k_{max,\varepsilon}^h$  implies that the eigenvalue  $\lambda = \lambda_k^h$  we are considering satisfies the bound

$$\lambda < \frac{4}{h^2} \gamma_{min} (1 - \varepsilon). \quad (2.33)$$

If we denote by  $x_i, y_i$  the two components of  $U_i$ , according to (2.22), relation (2.32) reads

$$\begin{cases} x_{i+1} = x_i \left(1 - \frac{h^2 \lambda}{\gamma_{i+1/2}}\right) + y_i h \sqrt{\frac{\lambda}{\gamma_{i+1/2}}}, \\ y_{i+1} = -x_i h \sqrt{\frac{\lambda}{\gamma_{i+1/2}}} + y_i. \end{cases} \quad (2.34)$$

The key-point is now to introduce the following quantity

$$H_i := x_i^2 + y_i^2 - h\sqrt{\frac{\lambda}{\gamma_{i+1/2}}}x_i y_i. \quad (2.35)$$

As we will see below, this quantity is conserved in the recurrence process ( $H_{i+1} = H_i$ ), when  $\gamma$  is constant. However, in the general case, this conservation does not hold anymore and we need sharper estimates.

After some tedious but straightforward computations, we end up with the following identity

$$H_{i+1} = H_i + \left( \frac{1}{\sqrt{\gamma_{i+1/2}}} - \frac{1}{\sqrt{\gamma_{i+3/2}}} \right) \times \left[ x_i^2 \frac{h^2 \lambda}{\sqrt{\gamma_{i+1/2}}} \left( \frac{h^2 \lambda}{\gamma_{i+1/2}} - 1 \right) + y_i^2 \frac{h^2 \lambda}{\sqrt{\gamma_{i+1/2}}} + x_i y_i h \sqrt{\lambda} \left( 1 - \frac{2h^2 \lambda}{\gamma_{i+1/2}} \right) \right]. \quad (2.36)$$

Using the regularity assumptions on  $\gamma$  and the mean-value theorem, we deduce that for some  $C(\gamma) > 0$

$$|H_{i+1}| \leq |H_i| + hC(\gamma) \left[ \frac{2h^2 \lambda}{\sqrt{\gamma_{\min}}} \left( \frac{h^2 \lambda}{\gamma_{\min}} + 1 \right) + \frac{h\sqrt{\lambda}}{2} \left( 1 + \frac{2h^2 \lambda}{\gamma_{\min}} \right) \right] (x_i^2 + y_i^2).$$

Moreover, we know from Lemma 2.3.3 that the quantity  $h^2 \lambda$  is uniformly bounded by some constant depending only on  $q$  and  $\gamma$ . It finally follows that, for some  $C(q, \gamma) > 0$ , we have

$$|H_{i+1}| \leq |H_i| + hC(x_i^2 + y_i^2), \quad \forall i \in \llbracket 1, N \rrbracket. \quad (2.37)$$

Besides, from definition (2.35) of  $H_i$ , we easily get that

$$(x_i^2 + y_i^2) \left( 1 - \frac{h}{2} \sqrt{\frac{\lambda}{\gamma_{i+1/2}}} \right) \leq H_i \leq (x_i^2 + y_i^2) \left( 1 + \frac{h}{2} \sqrt{\frac{\lambda}{\gamma_{i+1/2}}} \right). \quad (2.38)$$

Using again that  $h^2 \lambda$  is uniformly bounded, the right-hand side term can be replaced by some  $C(x_i^2 + y_i^2)$ . However, the factor in front of  $(x_i^2 + y_i^2)$  in the left-hand side may become negative for large eigenvalues. That is exactly the reason why the condition (2.33) on  $\lambda$  (which is nothing but the condition on  $k$  in the statement of the proposition) enters into the analysis. Indeed, if we assume (2.33), the above inequalities become

$$\frac{\varepsilon}{2} (x_i^2 + y_i^2) \leq H_i \leq C(q, \gamma) (x_i^2 + y_i^2). \quad (2.39)$$

From (2.37), we can thus infer that

$$|H_{i+1}| \leq |H_i| + \frac{C}{\varepsilon} h |H_i| \leq \exp(hC/\varepsilon) |H_i|, \quad \forall i \in \llbracket 1, N \rrbracket,$$

and finally, for any  $i \geq j$ , we get

$$|H_i| \leq \exp((i-j)hC/\varepsilon) |H_j| \leq \exp(C/\varepsilon) |H_j|.$$

Using again (2.39), we deduce that

$$(x_i^2 + y_i^2) \leq \frac{C}{\varepsilon} \exp\left(\frac{C}{\varepsilon}\right) (x_j^2 + y_j^2).$$

This exactly shows (2.30) in the case where  $i \geq j$ .

To prove the same equality for  $j \geq i$ , we come back to the equality (2.36) to get

$$|H_i| \leq |H_{i+1}| + Ch(x_i^2 + y_i^2). \quad (2.40)$$

Then, from (2.34), we find

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} 1 & -h\sqrt{\frac{\lambda}{\gamma_{i+1/2}}} \\ h\sqrt{\frac{\lambda}{\gamma_{i+1/2}}} & 1 - h^2 \frac{\lambda}{\gamma_{i+1/2}} \end{pmatrix} \begin{pmatrix} x_{i+1} \\ y_{i+1} \end{pmatrix}$$

and using again that  $h^2\lambda$  is uniformly bounded we obtain that, for some  $C(q, \gamma) > 0$ ,

$$x_i^2 + y_i^2 \leq C(x_{i+1}^2 + y_{i+1}^2), \quad \forall i \in \llbracket 1, N \rrbracket.$$

Hence, (2.40) becomes

$$|H_i| \leq |H_{i+1}| + Ch(x_{i+1}^2 + y_{i+1}^2) \leq |H_{i+1}| \left(1 + \frac{C}{\varepsilon}h\right),$$

and we can argue as before to prove our claim, when  $j \geq i$ .

Let us consider now the case where  $\gamma$  is constant.

— We first assume that the eigenvalue satisfies

$$\frac{h^2\lambda}{4\gamma} \leq 1 - \frac{h^2}{\gamma}. \quad (2.41)$$

Since  $\gamma$  is constant, we have  $H_i = H_j$  for any  $i, j$  (see (2.36)) and from (2.38) we have

$$\left(1 - \frac{h}{2}\sqrt{\frac{\lambda}{\gamma}}\right)(x_i^2 + y_i^2) \leq H_i = H_j \leq C(x_j^2 + y_j^2).$$

By (2.41), we observe that the first factor in the left-hand side is positive, so that we can write

$$x_i^2 + y_i^2 \leq \frac{C}{1 - \frac{h}{2}\sqrt{\frac{\lambda}{\gamma}}}(x_j^2 + y_j^2), \quad \forall i, j \in \llbracket 1, N+1 \rrbracket,$$

and the claim follows since, using again (2.41), we have

$$\frac{1}{1 - \frac{h}{2}\sqrt{\frac{\lambda}{\gamma}}} = \frac{1 + \frac{h}{2}\sqrt{\frac{\lambda}{\gamma}}}{1 - \frac{h^2\lambda}{4\gamma}} \leq \frac{C}{1 - \frac{h^2\lambda}{4\gamma}} \leq \frac{C\left(1 + \frac{\|q\|_\infty}{4}\right)}{1 - \frac{h^2(\lambda - \|q\|_\infty)}{4\gamma}}.$$

— Assume now that the eigenvalue does not satisfy (2.41). By Lemma 2.3.3 and Remark 2.1.3, this implies that

$$1 - \frac{h^2}{\gamma} < \frac{h^2\lambda}{4\gamma} \leq 1 - h^2 + \frac{h^2\|q\|_\infty}{4\gamma}. \quad (2.42)$$

We set  $\tilde{\lambda} = \lambda - \|q\|_\infty$ . We denote by  $\widetilde{M}_{\bullet, k}^h$  and  $\widetilde{S}_{\bullet \leftarrow \bullet, k}^h$  the same matrices as  $M_{\bullet, k}^h$  and  $S_{\bullet \leftarrow \bullet, k}^h$  with  $\lambda$  replaced by  $\tilde{\lambda}$ . An explicit computation shows that, for  $h$  small enough and using the left-hand side inequality of (2.42), we have

$$\|M_{i, k}^h - \widetilde{M}_{i, k}^h\| \leq C \frac{|\lambda - \tilde{\lambda}|}{\sqrt{\lambda} + \sqrt{\tilde{\lambda}}} \leq Ch, \quad \forall i \in \llbracket 1, N \rrbracket. \quad (2.43)$$

Since, by construction,  $\tilde{\lambda}$  satisfies the assumption (2.41), we have that

$$\|\widetilde{S}_{i \leftarrow j, k}^h\| \leq \frac{C}{\sqrt{1 - \frac{h^2\tilde{\lambda}}{4\gamma}}}, \quad \forall i, j \in \llbracket 1, N+1 \rrbracket. \quad (2.44)$$

With those notations, the recurrence relation (2.32) can be written as follows

$$U_{i+1, k}^h = \left(I + h\widetilde{M}_{i, k}^h\right)U_{i, k}^h + h(M_{i, k}^h - \widetilde{M}_{i, k}^h)U_{i, k}^h, \quad \forall i \in \llbracket 1, N+1 \rrbracket. \quad (2.45)$$

This formula has the same form as (2.21) with  $F_{i, k}^h = 0$  and  $Q_{i, k}^h = M_{i, k}^h - \widetilde{M}_{i, k}^h$  so that we can use the inequality (2.26) in this context, as well as (2.43) and (2.44), to finally get

$$\|U_{i, k}^h\| \leq \frac{C}{\sqrt{1 - \frac{h^2\tilde{\lambda}}{4\gamma}}} \exp\left(\frac{Ch}{\sqrt{1 - \frac{h^2\tilde{\lambda}}{4\gamma}}}\right) \|U_{j, k}^h\|, \quad \forall i, j \in \llbracket 1, N+1 \rrbracket.$$

Finally, since  $\tilde{\lambda}$  satisfies (2.41), we have  $\sqrt{1 - \frac{h^2\tilde{\lambda}}{4\gamma}} \geq \frac{h}{\sqrt{\gamma}}$  which proves that the exponential factor in the estimate above is actually uniformly bounded. It follows that

$$\|U_{i,k}^h\| \leq \frac{C}{\sqrt{1 - \frac{h^2\tilde{\lambda}}{4\gamma}}} \|U_{j,k}^h\|, \quad \forall i, j \in \llbracket 1, N+1 \rrbracket,$$

which exactly proves the claimed estimate.

■

We can now state and prove the following discrete spectral estimates in the setting (S2).

**Theorem 2.3.2 (The case of a uniform grid)** *Let  $\gamma$  and  $q$  satisfying (H<sub>0</sub>),  $\omega \subset \Omega$  a non empty open subset of  $\Omega$  and  $\varepsilon > 0$ . There exists  $h_1(q, \gamma) > 0$ ,  $h_2(q, \gamma, \omega) > 0$  and  $C_1(q, \gamma, \varepsilon) > 0$ ,  $C_2(q, \gamma, \omega, \varepsilon) > 0$  so that for any uniform mesh of size  $h$ , we have*

$$\frac{1}{\sqrt{\lambda_k^h}} |\partial_b \phi_k^h| \geq C_1, \quad \forall k \in \llbracket 1, k_{max}^h, \varepsilon \rrbracket, \quad \forall b \in \{l, r\}, \quad \forall h < h_1,$$

$$\|\phi_k^h\|_{L^2(\omega_h)}^2 \geq C_2, \quad \forall k \in \llbracket 1, k_{max}^h, \varepsilon \rrbracket, \quad \forall h < h_2,$$

and

$$\lambda_{k+1}^h - \lambda_k^h \geq C_1 \sqrt{\lambda_{k+1}^h}, \quad \forall k \in \llbracket 1, k_{max}^h, \varepsilon - 1 \rrbracket, \quad \forall h < h_1.$$

**Proof** We apply Lemma 2.3.2 with  $u^h = \phi_k^h$ ,  $\lambda = \lambda_k^h$  and  $f^h = 0$ . From Lemma 2.3.4 and inequality (2.30), we get for  $i, j \in \llbracket 1, N+1 \rrbracket$ ,

$$\|U_{i,k}^h\| \leq C \exp \left( \exp(C/\varepsilon) C \sum_{p \in \llbracket i, j \rrbracket \setminus \{i\}} h \|Q_{p,k}^h\| \right) \times \exp(C/\varepsilon) \|U_{j,k}^h\|.$$

According to (2.23), we have a uniform bound  $\|Q_{p,k}^h\| \leq C$ , so that we finally obtain

$$\|U_{i,k}^h\| \leq C(q, \gamma, \varepsilon) \|U_{j,k}^h\|, \quad \forall i, j \in \llbracket 1, N+1 \rrbracket.$$

By (2.25), we just have shown that

$$\mathcal{R}_k^h \geq \frac{1}{C(q, \gamma, \varepsilon)^2},$$

which gives the first two lower bounds by virtue of Lemma 2.3.1.

Let us now estimate the difference  $\lambda_{k+1}^h - \lambda_k^h$  for any  $k \in \llbracket 1, k_{max}^h, \varepsilon - 1 \rrbracket$ . We follow the same lines as in the proof in the continuous case in Section 2.2. We set  $u^h = (\partial_r \phi_k^h) \phi_{k+1}^h - (\partial_r \phi_{k+1}^h) \phi_k^h$  and  $f^h = (\lambda_{k+1}^h - \lambda_k^h) (\partial_r \phi_{k+1}^h) \phi_k^h$ , so that (2.20) is satisfied with  $\lambda = \lambda_{k+1}^h$ . We use the same notation as in Lemma 2.3.2, and we observe that

$$U_{N+1}^h = \begin{pmatrix} u_{N+1}^h \\ \sqrt{\frac{\gamma_{N+1/2}}{\lambda_{k+1}^h}} \partial_r u^h \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

since  $\phi_{k+1}^h$  and  $\phi_k^h$  satisfy the homogeneous boundary conditions and moreover, by construction,  $\partial_r u^h = 0$ .

We can then apply Lemma 2.3.4, with  $j = N+1$ , and the estimates (2.23) and (2.30) to get

$$\|U_i^h\| \leq C(q, \gamma, \varepsilon) \sum_{p \in \llbracket i, j \rrbracket \setminus \{i\}} h \|F_p^h\|, \quad \forall i \in \llbracket 1, N+1 \rrbracket.$$

By definition of  $F^h$  and  $f^h$ , we have

$$\|F_p^h\| \leq \frac{C(\gamma)}{\sqrt{\lambda_{k+1}^h}} (\lambda_{k+1}^h - \lambda_k^h) |\partial_r \phi_{k+1}^h| |\phi_{p,k}^h|, \quad \forall p \in \llbracket 1, N \rrbracket.$$

Since  $\phi_k^h$  is normalized in  $L^2(\Omega_h)$  we conclude that

$$\|U_i^h\| \leq \frac{C(q, \gamma, \varepsilon)}{\sqrt{\lambda_{k+1}^h}} (\lambda_{k+1}^h - \lambda_k^h) |\partial_r \phi_{k+1}^h|.$$

By definition of  $U_i^h$ , we have  $|u_i^h| \leq \|U_i^h\|$  for any  $i \in \llbracket 1, N \rrbracket$  and thus, the discrete  $L^2$  norm of  $u^h$  is estimated by

$$\|u^h\|_{L^2(\Omega_h)}^2 \leq \frac{C(q, \gamma, \varepsilon)}{\lambda_{k+1}^h} |\lambda_{k+1}^h - \lambda_k^h|^2 |\partial_r \phi_{k+1}^h|^2.$$

By definition of  $u^h$  and since  $\phi_{k+1}^h$  and  $\phi_k^h$  are orthonormal we have

$$\|u^h\|_{L^2(\Omega_h)}^2 = (\partial_r \phi_{k+1}^h)^2 + (\partial_r \phi_k^h)^2.$$

Combining the two previous inequalities (and using that  $\partial_r \phi_{k+1}^h \neq 0$ ) we finally get

$$1 \leq \frac{C(q, \gamma, \varepsilon)}{\lambda_{k+1}^h} |\lambda_{k+1}^h - \lambda_k^h|^2,$$

which gives the claim.  $\blacksquare$

Finally, in the case of a constant diffusion coefficient and uniform grids we obtain a stronger result with precise lower bounds for all the eigenfunctions of  $\mathcal{A}^h$ .

**Theorem 2.3.3 (Uniform grid - constant diffusion)** *Let  $q$  be a continuous function on  $\bar{\Omega}$  and assume that  $\gamma$  is a constant function. Let  $\omega \subset \Omega$  be a non-empty open subset of  $\Omega$ . There exist  $h_1(q, \gamma) > 0$ ,  $h_2(q, \gamma, \omega) > 0$  sufficiently small and  $C_1(q, \gamma) > 0$ ,  $C_2(q, \gamma, \omega) > 0$  such that for any uniform mesh of size  $h$ , we have for  $h < h_1$ ,*

$$\frac{1}{\sqrt{\lambda_k^h}} |\partial_b \phi_k^h| \geq C_1 \sqrt{1 - \frac{h^2(\lambda_k^h - \|q\|_{L^\infty})}{4\gamma}}, \quad \forall k \in \llbracket 1, N \rrbracket, \quad \forall b \in \{l, r\},$$

and

$$\|\phi_k^h\|_{L^2(\omega_h)}^2 \geq C_2 \left( 1 - \frac{h^2(\lambda_k^h - \|q\|_{L^\infty})}{4\gamma} \right), \quad \forall k \in \llbracket 1, N \rrbracket,$$

for  $h < h_2$ . Moreover,

$$\lambda_{k+1}^h - \lambda_k^h \geq C_1 \sqrt{\lambda_{k+1}^h} \sqrt{1 - \frac{h^2(\lambda_{k+1}^h - \|q\|_{L^\infty})}{4\gamma}}, \quad \forall k \in \llbracket 1, N-1 \rrbracket,$$

as soon as  $h < h_1$ .

**Proof** We shall use the notation  $\tilde{\lambda}_k^h = \lambda_k^h - \|q\|_{L^\infty}$ . We follow exactly the same lines as in the proof of the previous theorem, except that the estimate (2.30) is now improved into (2.31). It follows that, for any  $i, j \in \llbracket 1, N+1 \rrbracket$  we have

$$\|U_{i,k}^h\| \leq \frac{C}{\sqrt{1 - \frac{h^2 \tilde{\lambda}_k^h}{4\gamma}}} \exp \left( \frac{C}{\sqrt{1 - \frac{h^2 \tilde{\lambda}_k^h}{4\gamma}}} \sum_{p \in \llbracket i, j \rrbracket \setminus \{i\}} h \|Q_{p,k}^h\| \right) \|U_{j,k}^h\|.$$

According to (2.24), we have the bound  $\|Q_{p,k}^h\| \leq C \left( h + \frac{1}{\sqrt{\lambda_k^h}} \right)$ , for any  $p$  and thus

$$\|U_{i,k}^h\| \leq \frac{C}{\sqrt{1 - \frac{h^2 \tilde{\lambda}_k^h}{4\gamma}}} \exp \left( C \frac{h + \frac{1}{\sqrt{\lambda_k^h}}}{\sqrt{1 - \frac{h^2 \tilde{\lambda}_k^h}{4\gamma}}} \right) \|U_{j,k}^h\|.$$

It remains to prove that the exponential factor is actually uniformly bounded.

— Assume first that  $h^2\tilde{\lambda}_k^h \leq \gamma$ , so that we have

$$\sqrt{1 - \frac{h^2\tilde{\lambda}_k^h}{4\gamma}} \geq \sqrt{3/4},$$

and the exponential term above is clearly bounded.

— Assume now that  $\gamma \leq h^2\tilde{\lambda}_k^h$ . We have already seen that we also have  $h^2\tilde{\lambda}_k^h \leq 4\gamma(1 - h^2)$  so that

$$\sqrt{1 - \frac{h^2\tilde{\lambda}_k^h}{4\gamma}} \geq h,$$

and then

$$\frac{h + \frac{1}{\sqrt{\lambda_k^h}}}{\sqrt{1 - \frac{h^2\tilde{\lambda}_k^h}{4\gamma}}} \leq \left(1 + \frac{1}{h\sqrt{\lambda_k^h}}\right) \leq \left(1 + \frac{1}{h\sqrt{\tilde{\lambda}_k^h}}\right) \leq \left(1 + \frac{1}{\sqrt{\gamma}}\right),$$

and we reach the same conclusion.

To sum up, we have just proved that

$$\|U_{i,k}^h\| \leq \frac{C}{\sqrt{1 - \frac{h^2\tilde{\lambda}_k^h}{4\gamma}}} \|U_{j,k}^h\|,$$

which implies, with (2.25), that

$$\mathcal{R}_k^h \geq C \left(1 - \frac{h^2\tilde{\lambda}_k^h}{4\gamma}\right), \quad \forall k \in \llbracket 1, N \rrbracket.$$

With Lemma 2.3.1 the first two inequalities are proved.

As in the proof of Theorem 2.3.2, the gap property is proved by defining  $u^h = (\partial_r \phi_k^h) \phi_{k+1}^h - (\partial_r \phi_{k+1}^h) \phi_k^h$  and  $f^h = (\lambda_{k+1}^h - \lambda_k^h) (\partial_r \phi_{k+1}^h) \phi_k^h$ , and using the uniform estimate above on  $\exp\left(C \frac{h + \frac{1}{\sqrt{\lambda_k^h}}}{\sqrt{1 - \frac{h^2\tilde{\lambda}_k^h}{4\gamma}}}\right)$ .

■

A careful inspection of the arguments in the above proof shows that we have the estimate

$$\lambda_k^h \left(1 - \frac{h^2}{4\gamma} (\lambda_k^h - \|q\|_\infty)\right) \geq C(q, \gamma), \quad \forall k \in \llbracket 1, N \rrbracket.$$

This immediately gives the following corollary

**Corollary 2.3.1** *With the same assumptions and notations as in the previous theorem (the values of  $C_1$  and  $C_2$  being possibly different) we have*

$$|\partial_b \phi_k^h| \geq C_1, \quad \forall k \in \llbracket 1, N \rrbracket, \quad \forall b \in \{l, r\},$$

$$\lambda_{k+1}^h - \lambda_k^h \geq C_1, \quad \forall k \in \llbracket 1, N-1 \rrbracket,$$

as soon as  $h < h_1$ , and

$$\|\phi_k^h\|_{L^2(\omega_h)}^2 \geq \frac{C_2}{\lambda_k^h}, \quad \forall k \in \llbracket 1, N \rrbracket,$$

as soon as  $h < h_2$ .

### 2.3.4 Proof of Lemma 2.3.1

**Proof** We follow exactly the same arguments as the ones given in the proof of Lemma 2.2.1. For the sake of completeness, we provide here some technical precision.

— Let us prove that

$$\frac{1}{\lambda_k^h} |\partial_b \phi_k^h|^2 \geq C \mathcal{R}_k^h, \forall b \in \{l, r\}. \quad (2.46)$$

We recall, from (2.19), the following identity for  $i, j \in \llbracket 1, N+1 \rrbracket$

$$\mathcal{R}_k^h \left( |\phi_{i,k}^h|^2 + \frac{\gamma_{i-1/2}}{\lambda_k^h} \left| \frac{\phi_{i,k}^h - \phi_{i-1,k}^h}{h_{i-1/2}} \right|^2 \right) \leq |\phi_{j,k}^h|^2 + \frac{\gamma_{j-1/2}}{\lambda_k^h} \left| \frac{\phi_{j,k}^h - \phi_{j-1,k}^h}{h_{j-1/2}} \right|^2.$$

With  $j = N+1$ , using Dirichlet boundary condition in  $x = 1$  and summing the left hand side for  $i$  from 1 to  $N$ , we get (2.46) with  $b = r$ . For the case  $b = l$ , take  $j = 1$  and again, sum with respect to the variable  $i$  :

$$\mathcal{R}_k^h \leq |\phi_{1,k}^h|^2 + \frac{\gamma_{1/2}}{\lambda_k^h} \left| \frac{\phi_{1,k}^h - 0}{h_{1/2}} \right|^2 = \left( h_{1/2}^2 + \frac{\gamma_{1/2}}{\lambda_k^h} \right) \left| \frac{\phi_{1,k}^h - 0}{h_{1/2}} \right|^2$$

According to Lemma 2.3.3,  $h_{1/2}^2 \lambda_k^h \leq C(q, \gamma, \beta)$ , so  $\mathcal{R}_k^h \leq \frac{C(q, \gamma, \beta)}{\lambda_k^h} \left| \frac{\phi_{1,k}^h - 0}{h_{1/2}} \right|^2$ .

— Denote by  $\omega_j^h, j \in \{1, 2, 3\}$  the set of mesh points which belong to  $\omega_j$ . In the discrete setting, hypothesis (2.15) is replaced by: suppose that there exists  $x_{i_1}$  and  $x_{i_3}$  two points of  $\omega_1^h$  and  $\omega_3^h$  respectively, such that

$$\phi_{i_1,k}^h (\phi_{i_1,k}^h - \phi_{i_1-1,k}^h) \geq 0 \text{ and } \phi_{i_2,k}^h (\phi_{i_2+1,k}^h - \phi_{i_2,k}^h) \leq 0. \quad (2.47)$$

Thus, when multiplying the equation  $\mathcal{A}^h \phi_k^h = \lambda_k^h \phi_k^h$  by  $\phi_k^h$  and summing between  $i_1$  and  $i_3$ , the discrete integration by parts leads to boundary terms (which are the analogues of  $[\gamma(x) \phi_k'(x) \phi_k(x)]_{a_k}^{b_k}$ ) that are again non positive thanks to (2.47).

— Remark 2.2.3 still holds in the discrete setting. Indeed, denote by  $i_0$  the index of the first point of  $\omega_3^h$  and suppose that for all indices  $i \in \{1, \dots, N\}$  such that  $x_i \in \omega_3^h$ , we have  $\phi_{i,k}^h (\phi_{i+1,k}^h - \phi_{i,k}^h) > 0$ . If  $\phi_{i_0,k}^h > 0$  then  $(\phi_{i,k}^h)_{i, x_i \in \omega_3^h}$  is increasing and positive, else if  $\phi_{i_0,k}^h < 0$  then  $(\phi_{i,k}^h)_{i, x_i \in \omega_3^h}$  is decreasing and negative.

Moreover, given that for all  $i \in \{1, \dots, N\}$  such that  $x_i \in \omega_3^h$ ,

$$0 < \phi_{i,k}^h (\phi_{i+1,k}^h - \phi_{i,k}^h) = \frac{1}{2} ((\phi_{i+1,k}^h)^2 - (\phi_{i,k}^h)^2) - \frac{1}{2} [(\phi_{i+1,k}^h) - (\phi_{i,k}^h)]^2,$$

we conclude that  $(|\phi_{i,k}^h|^2)_{i, x_i \in \omega_3^h}$  is increasing on  $\omega_3^h$ .

■

## 2.4 Numerical illustrations

In this section we provide a few numerical simulations that aim to illustrate our theoretical results on the discrete spectral properties of the operator  $\mathcal{A}^h$  obtained in the previous section. We will consider the following two control/observation domains  $\omega^1 = (0, 0.3)$  and  $\omega^2 = (0.7, 1)$ .

Let us introduce the following notations for any  $K \in \llbracket 1, N \rrbracket$

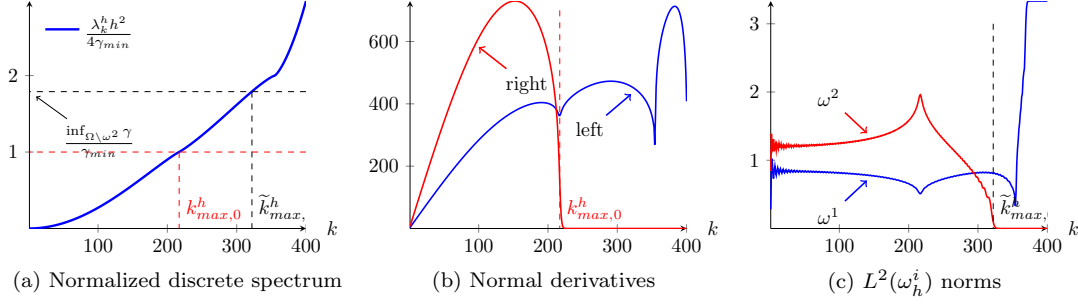
$$I_l^h(K) := \min_{k \leq K} |\partial_l \phi_k^h|, \quad I_r^h(K) := \min_{k \leq K} |\partial_r \phi_k^h|$$

$$I_1^h(K) := \min_{k \leq K} \frac{\|\phi_k^h\|_{L^2(\omega^1)}^2}{|\omega^1|}, \quad I_2^h(K) := \min_{k \leq K} \frac{\|\phi_k^h\|_{L^2(\omega^2)}^2}{|\omega^2|},$$

and

$$\Delta^h(K) := \min_{k \leq K-1} |\lambda_{k+1}^h - \lambda_k^h|.$$

We are mainly interested in the values of  $I_\bullet^h(N)$  (resp.  $I_\bullet^h(k_{max, \varepsilon}^h)$ ) in order to illustrate whether the corresponding quantities are bounded from below or not for the whole discrete spectrum (resp. for a constant portion of the spectrum characterized by  $k_{max, \varepsilon}^h$ , see (2.29)) when the mesh size tends to 0. In the tables below, each right-hand side sub-column (with a white background) corresponds to the whole spectrum estimate (i.e. for  $K = N$ ) whereas the left-hand side sub-column (with a gray background) corresponds to the partial spectrum estimate (i.e. for  $K = k_{max, \varepsilon}^h$ ). We have chosen  $\varepsilon = 0.05$ , except for some cases for which it is explicitly mentioned.


 Figure 2.2 – Case 1 -  $N = 400$ 

**Case 1** We consider  $\gamma(x) = 2 + \cos(\pi x)^3$ ,  $q(x) = 0$  and a family of uniform meshes. This case corresponds to the framework (S2) and to Theorem 2.3.2.

$N$	$k_{max,\varepsilon}^h$	$I_l^h(\cdot)$		$I_r^h(\cdot)$		$I_1^h(\cdot)$		$I_2^h(\cdot)$		$\Delta^h(\cdot)$	
		$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$
50	26	2.99	2.99	7.07	$1.01 \cdot 10^{-23}$	0.29	0.29	0.86	$8.71 \cdot 10^{-29}$	56.82	56.82
100	52	2.99	2.99	7.08	$2.46 \cdot 10^{-51}$	0.28	0.28	0.85	$1.26 \cdot 10^{-59}$	56.89	56.89
200	104	2.99	2.99	7.08	$4.16 \cdot 10^{-107}$	0.28	0.28	0.85	$1.97 \cdot 10^{-121}$	56.91	56.91
300	156	2.99	2.99	7.08	$4.22 \cdot 10^{-163}$	0.28	0.28	0.84	$2.70 \cdot 10^{-183}$	56.91	56.91
400	208	2.99	2.99	7.08	$3.47 \cdot 10^{-219}$	0.28	0.28	0.84	$3.50 \cdot 10^{-245}$	56.91	56.91

 Table 2.1 – Case 1 - behavior as  $h \rightarrow 0$ 

We observe in Table 2.1 that, in accordance with our theoretical results, all the partial spectrum quantities computed with  $K = k_{max,\varepsilon}^h$  (which is almost equal to  $N/2$  here) are bounded from below.

Interestingly enough, we observe that  $I_l^h(N)$  and  $I_1^h(N)$  are also uniformly bounded from below but it is not the case for  $I_r^h(N)$  and  $I_2^h(N)$ . This discrepancy seems to come from the fact that the diffusion coefficient  $\gamma$  is decreasing and maximal exactly on the left boundary of the domain. For those latter quantities, the only theoretical result we have is the one of Theorem 2.3.1. We can check numerically that the exponential bound given in this result seems to be sharp. Indeed we have observed that, for all the considered values of  $N$ , we have

$$\min_{k \in [k_{max,\varepsilon}^h, N]} \frac{\log(|\partial_r \phi_k^h|)}{\sqrt{\lambda_k^h}} \approx -0.36, \quad \min_{k \in [k_{max,\varepsilon}^h, N]} \frac{\log(\|\phi_k^h\|_{L^2(\omega_h^2)})}{\sqrt{\lambda_k^h}} \approx -0.20,$$

so that the quantities of interest actually behaves, in the upper part of the discrete spectrum, like  $e^{-C\sqrt{\lambda_k^h}}$  for some  $C$  independent of  $h$ .

Moreover, we observe in Figure 2.2b that the actual value of  $k_{max,\varepsilon}^h$  is sharp and that  $\partial_r \phi_k^h$  becomes exponentially small as soon as  $k \gtrsim k_{max,0}^h$ . However, in Figure 2.2c the exponential behavior of  $\|\phi_k^h\|_{L^2(\omega_h^2)}$  seems to appear for higher values of  $k$ . This can be explained (and actually it can be deduced by a careful observation of our proofs) by the fact that, due to the monotonicity of  $\gamma$ , the minimal value of  $\gamma$  that has to be taken into account in the evaluation of  $k_{max,\varepsilon}^h$  is not its infimum on  $\Omega$  but rather its infimum on  $\Omega \setminus \omega^2$ . The correct threshold in that case is thus the one defined by

$$\tilde{k}_{max,\varepsilon}^h := \max \left\{ k \in [1, N]; \lambda_k^h < \frac{4}{h^2} \left( \inf_{\Omega \setminus \omega^2} \gamma \right) (1 - \varepsilon) \right\}.$$

We represent the value of  $\tilde{k}_{max,0}^h$  in Figure 2.2a and 2.2c.

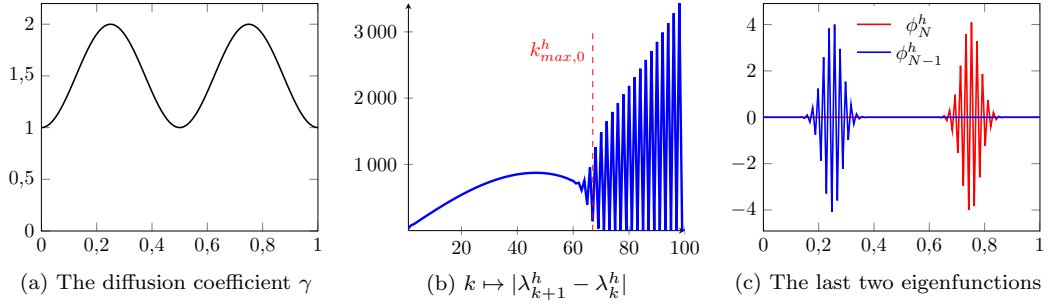
Last but not least, we observe on this particular example that the uniform gap condition seems to be satisfied. In particular, by the methods given in Section 2.5, we can then conjecture that the associated parabolic equations or systems are indeed uniformly null-controllable by either distributed or boundary control.

**Case 2** We consider now a case which is very much similar to the previous one by setting  $\gamma(x) = 2 - \cos(2\pi x)^2$ ,  $q(x) = 0$  and again a family of uniform meshes. The results are gathered in Table 2.2.



The main difference with the previous case lie in the fact that  $\gamma$  is no more monotonic and is actually symmetric with respect to the mid-point  $x = 1/2$  (see Fig. 2.3a). It appears, in this configuration, that the uniform gap property does not hold anymore and that neither  $I_1^h(N)$  nor  $I_2^h(N)$  are uniformly bounded from below. We illustrate further those behaviors by plotting the difference between two successive eigenvalues in Fig. 2.3b as well as the last two eigenfunctions in Fig. 2.3c. We observe that those two eigenfunctions are essentially supported in one half of the domain (which explains why their  $L^2$  norm on  $\omega^1$  or  $\omega^2$  can be exponentially small) and that they are almost symmetric to each other (which explains that the corresponding eigenvalues are very close).

$N$	$k_{max,\varepsilon}^h$	$I_l^h(\cdot)$		$I_r^h(\cdot)$		$I_1^h(\cdot)$		$I_2^h(\cdot)$		$\Delta^h(\cdot)$	
		$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$
50	32	6.39	$1.74 \cdot 10^{-3}$	6.39	$1.74 \cdot 10^{-3}$	0.56	0.56	0.6	0.6	33.53	$4.51 \cdot 10^{-8}$
100	64	6.41	$6.85 \cdot 10^{-15}$	6.41	$7.18 \cdot 10^{-30}$	0.59	$1.75 \cdot 10^{-30}$	0.59	$2.83 \cdot 10^{-42}$	33.58	$2.91 \cdot 10^{-11}$
200	126	6.42	$3.02 \cdot 10^{-63}$	6.42	$3.80 \cdot 10^{-14}$	0.58	$8.90 \cdot 10^{-87}$	0.58	$2.81 \cdot 10^{-30}$	33.59	$2.91 \cdot 10^{-11}$
300	187	6.42	$8.41 \cdot 10^{-15}$	6.42	$9.47 \cdot 10^{-97}$	0.61	$4.41 \cdot 10^{-30}$	0.58	$1.60 \cdot 10^{-131}$	33.59	$1.16 \cdot 10^{-10}$
400	250	6.42	$2.50 \cdot 10^{-130}$	6.42	$5.30 \cdot 10^{-15}$	0.58	$7.38 \cdot 10^{-176}$	0.58	$6.02 \cdot 10^{-30}$	33.59	$6.98 \cdot 10^{-10}$

 Table 2.2 – Case 2 - behavior as  $h \rightarrow 0$ 

 Figure 2.3 – Case 2 -  $N = 100$ 

**Case 3** In this case we consider a constant diffusion coefficient  $\gamma = 1$  but a variable potential  $q(x) = 50 \cos(\pi x^2)$ . We propose two subcases: in Subcase 3.1 we still use uniform meshes families whereas in Subcase 3.2 we use families of quasi-uniform meshes obtained by gluing a uniform mesh of  $(0, 1/2)$  with cells of size  $h$  and a uniform mesh of  $(1/2, 1)$  made of cells of size  $h/2$ ; the value of  $h$  is chosen so as we finally obtain the expected total number  $N$  of cells. For this test case, we have chosen  $\varepsilon = 0.1$ .

$N$	$k_{max,\varepsilon}^h$	$I_l^h(\cdot)$		$I_r^h(\cdot)$		$I_1^h(\cdot)$		$I_2^h(\cdot)$		$\Delta^h(\cdot)$	
		$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$
50	43	1.12	1.12	2.92	1.7	$9.97 \cdot 10^{-2}$	$9.97 \cdot 10^{-2}$	0.31	$8.77 \cdot 10^{-2}$	44.07	21.54
100	86	1.12	1.12	2.92	1.7	$9.73 \cdot 10^{-2}$	$9.73 \cdot 10^{-2}$	0.31	$8.66 \cdot 10^{-2}$	44.21	21.56
200	172	1.12	1.12	2.92	1.7	$9.60 \cdot 10^{-2}$	$9.60 \cdot 10^{-2}$	0.3	$8.61 \cdot 10^{-2}$	44.25	21.57
300	257	1.12	1.12	2.92	1.7	$9.55 \cdot 10^{-2}$	$9.55 \cdot 10^{-2}$	0.3	$8.59 \cdot 10^{-2}$	44.26	21.57
400	343	1.12	1.12	2.92	1.7	$9.53 \cdot 10^{-2}$	$9.53 \cdot 10^{-2}$	0.3	$8.58 \cdot 10^{-2}$	44.26	21.57

 Table 2.3 – Subcase 3.1 - behavior as  $h \rightarrow 0$ 

As predicted by Theorem 2.3.3 and Corollary 2.3.1, we observe in Table 2.3 that, in the subcase 3.1 all the quantities of interest are uniformly bounded from below. However, when considering a non uniform mesh we can see in Table 2.4 that the quantities  $I_l^h(N)$  and  $I_1^h(N)$  seem to be very small, but actually not exponentially small with respect to  $\sqrt{\lambda}$ . It is interesting to observe that the right derivative term  $I_r^h(N)$  is still uniformly bounded from below as well as  $I_2^h(N)$ . This is an illustration of the fact that the bounds from below are more likely to be uniform when it concerns the finest part of the mesh than the coarsest part.

$N$	$k_{max,\varepsilon}^h$	$I_l^h(\cdot)$		$I_r^h(\cdot)$		$I_1^h(\cdot)$		$I_2^h(\cdot)$		$\Delta^h(\cdot)$	
		$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$
50	26	1.12	$9.49 \cdot 10^{-17}$	2.91	2.91	0.11	$1.65 \cdot 10^{-18}$	0.25	$4.11 \cdot 10^{-2}$	43.9	43.9
100	50	1.12	$9.55 \cdot 10^{-16}$	2.92	2.92	0.1	$1.74 \cdot 10^{-30}$	0.13	$1.04 \cdot 10^{-2}$	44.17	44.17
200	96	1.12	$2.91 \cdot 10^{-15}$	2.92	2.92	$9.95 \cdot 10^{-2}$	$6.68 \cdot 10^{-30}$	0.28	$8.09 \cdot 10^{-4}$	44.24	44.24
300	144	1.12	$1.24 \cdot 10^{-14}$	2.92	2.92	$9.71 \cdot 10^{-2}$	$1.32 \cdot 10^{-29}$	0.25	$3.35 \cdot 10^{-4}$	44.25	44.25
400	190	1.12	$4.26 \cdot 10^{-15}$	2.92	2.92	$9.67 \cdot 10^{-2}$	$1.93 \cdot 10^{-29}$	0.24	$8.11 \cdot 10^{-4}$	44.26	44.26

 Table 2.4 – Subcase 3.2 - behavior as  $h \rightarrow 0$ 

**Case 4** We conclude this series of numerical illustrations by considering the case of a piecewise constant diffusion coefficient  $\gamma = 1]_{0,0.4[ + 2 \times 1]_{0.4,1[}$ , with  $q = 0$  and a family of uniform meshes. This case does not directly enter our analysis but our arguments can be adapted and we can prove in that case (see Section 2.6, Remark 2 and Appendix 2.7.3, for more details) that each quantity  $I_\bullet^h(N)$  is

- either exponentially small if it concerns a part of the domain where  $\gamma$  takes its lowest value (in the present example: the left normal derivative and the  $L^2$  norm on  $\omega^1$ )
- or uniformly bounded from below if it concerns a part of the domain where  $\gamma$  takes its highest value (here : the right normal derivative and the  $L^2$  norm on  $\omega^2$ ).

$N$	$k_{max,\varepsilon}^h$	$I_l^h(\cdot)$		$I_r^h(\cdot)$		$I_1^h(\cdot)$		$I_2^h(\cdot)$		$\Delta^h(\cdot)$	
		$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$	$k_{max,\varepsilon}^h$	$N$
50	35	4.41	$1.89 \cdot 10^{-14}$	3.79	3.79	0.55	$2.43 \cdot 10^{-11}$	0.12	0.12	44.89	44.89
100	68	5.37	$9.21 \cdot 10^{-30}$	3.79	3.79	0.68	$7.18 \cdot 10^{-20}$	$5.87 \cdot 10^{-2}$	$5.87 \cdot 10^{-2}$	44.62	44.62
200	131	5.37	$2.19 \cdot 10^{-60}$	3.79	3.79	0.67	$4.53 \cdot 10^{-36}$	$8.86 \cdot 10^{-2}$	$2.79 \cdot 10^{-2}$	44.47	44.47
300	194	5.37	$5.22 \cdot 10^{-91}$	3.79	3.79	0.67	$6.63 \cdot 10^{-52}$	$9.81 \cdot 10^{-2}$	$1.88 \cdot 10^{-2}$	44.42	44.42
400	257	5.37	$1.25 \cdot 10^{-121}$	3.79	3.79	0.67	$1.37 \cdot 10^{-67}$	0.1	$1.42 \cdot 10^{-2}$	44.4	44.4

 Table 2.5 – Case 4 - behavior as  $h \rightarrow 0$ 

## 2.5 Applications in control theory

The moments method has been successfully used to prove null-controllability of parabolic equations and systems, in particular with boundary controls, see for instance [5, 15]. In section 2.5.1 we present this method on the heat equation, then on cascade systems. Eventually, in Sections 2.5.2 and 2.5.3, we show how to adapt this strategy to the discrete setting.

### 2.5.1 Null-controllability via the moments method in the continuous setting

Let us fix  $y^0 \in H^{-1}(\Omega)$  and consider the following control problem

$$\begin{cases} \partial_t y(t, x) + \mathcal{A}y(t, x) = 1_\omega(x)V_d(t, x), & \text{in } (0, T) \times \Omega \\ y(t, 0) = 0, y(t, 1) = V_b(t) & \text{in } (0, T) \\ y(0, x) = y^0(x) & \text{in } \Omega, \end{cases} \quad (2.48)$$

with  $V_d \in L^2((0, T) \times \Omega)$ ,  $V_b \in L^2(0, T)$ . Of course, all the results remain unchanged if one controls (2.48) at the left boundary instead of the right boundary.

For a given pair of controls  $V_b \in L^2(0, T)$  and  $V_d \in L^2((0, T) \times \Omega)$ , we say that a function  $y \in C^0([0, T], H^{-1}(\Omega))$  is a solution of (2.48) if and only if for any  $k \geq 1$  and any  $t \in [0, T]$ , we have

$$\begin{aligned} \langle y(t), \phi_k \rangle_{H^{-1} \times H_0^1} &= \langle y^0, e^{-\lambda_k t} \phi_k \rangle_{H^{-1} \times H_0^1} \\ &= \int_0^t \int_\Omega 1_\omega(x)V_d(s, x)e^{-\lambda_k(t-s)}\phi_k(x)dxds - \int_0^t \gamma(1)V_b(s)e^{-\lambda_k(t-s)}\partial_x \phi_k(1)ds, \end{aligned}$$

where  $(\lambda_k, \phi_k)$  are the eigenlements of the self-adjoint operator  $\mathcal{A}$  as defined in the introduction.

It is proved in [13] that for any  $V_b$  and  $V_d$ , such a solution exists and is unique. Since we are interested in the controllability of such a system at time  $T$ , we will often make use of the above formula specialized at time  $t = T$

$$\begin{aligned} \langle y(T), \phi_k \rangle_{H^{-1} \times H_0^1} - \langle y^0, e^{-\lambda_k T} \phi_k \rangle_{H^{-1} \times H_0^1} \\ = \int_0^T \int_{\Omega} 1_{\omega}(x) V_d(t, x) z_k(t, x) dx dt - \int_0^T \gamma(1) V_b(t) \partial_x z_k(t, 1) dt, \end{aligned} \quad (2.49)$$

where we have defined  $z_k$  by  $z_k : (t, x) \mapsto e^{-\lambda_k(T-t)} \phi_k(x)$  which is nothing but the solution of the *adjoint problem* with final data  $\phi_k$

$$\begin{cases} -\partial_t z_k(t, x) + \mathcal{A} z_k(t, x) = 0 & \text{in } (0, T) \times \Omega, \\ z_k(t, 0) = z_k(t, 1) = 0 & \text{in } (0, T), \\ z_k(T, x) = \phi_k(x) & \text{in } \Omega. \end{cases} \quad (2.50)$$

— Consider first the *right-boundary null-control problem* : we set  $V_d = 0$ , and we look for a boundary control  $V_b \in L^2(0, T)$  such that the corresponding solution  $y$  of (2.48) satisfies  $y(T) = 0$ .

We write the decomposition of  $y^0$  in the basis  $(\phi_k)_{k \geq 1}$  as follows  $y^0 = \sum_{k \geq 1} y_k^0 \phi_k$  and from (2.49) we see that the problem amounts to finding  $V_b$  such that

$$y_k^0 e^{-\lambda_k T} = \gamma(1) \partial_r \phi_k \left( \int_0^T V_b(t) e^{-\lambda_k(T-t)} dt \right), \forall k \geq 1. \quad (2.51)$$

The set of equations (2.51) indexed by  $k$  is called a *moment problem*. The moments method consists in solving (2.51) using a *biorthogonal family* of the real exponentials  $(e^{-\lambda_k(T-t)})_{k \geq 1}$ .

Let  $\Lambda = (\lambda_k)_{k \geq 1}$  be a sequence of positive numbers. A biorthogonal family of  $(e^{-\lambda_k(T-t)})_{k \geq 1}$  is a set of functions in  $L^2(0, T)$  denoted by  $(q_l^\Lambda)_{l \geq 1}$  verifying

$$\int_0^T e^{-\lambda_k(T-t)} q_l^\Lambda(t) dt = \delta_{l,k}, \quad \forall l, k \geq 1. \quad (2.52)$$

We can now solve, at least formally, the moment problem (2.51) and give a possible expression of  $V_b$

$$V_b(t) := \sum_{k \geq 1} \frac{y_k^0 e^{-\lambda_k T}}{\gamma(1) \partial_r \phi_k} q_k^\Lambda(t). \quad (2.53)$$

— Now, we study the *distributed control problem* : we set  $V_b = 0$  and we look for a control  $V_d$  such that the corresponding solution  $y$  of (2.48) satisfies  $y(T) = 0$ . Using again (2.49), this amounts to finding  $V_d$  satisfying the following family of equalities

$$-y_k^0 e^{-\lambda_k T} = \int_0^T \int_{\omega} V_d(t, x) e^{-\lambda_k(T-t)} \phi_k(x) dx dt, \quad \forall k \geq 1. \quad (2.54)$$

Inspired by the boundary control case, we look for a suitable  $V_d$  in the following form

$$V_d(t, x) = \sum_{k \geq 1} \alpha_k q_k^\Lambda(t) \phi_k(x), \quad (2.55)$$

where  $(\alpha_k)_{k \geq 1}$  is a sequence of real numbers to be determined.

Injecting (2.55) in (2.54) and using (2.52) we finally get the following formula for the unknown coefficients

$$\alpha_k = -\frac{y_k^0 e^{-\lambda_k T}}{\int_{\omega} \phi_k^2(x) dx}.$$

Formally, the control problem is thus solved by defining

$$V_d(t, x) := \sum_{k \geq 1} \left( -\frac{y_k^0 e^{-\lambda_k T}}{\|\phi_k\|_{L^2(\omega)}^2} \right) q_k^\Lambda(t) \phi_k(x). \quad (2.56)$$

**Remark 2.5.1** *This strategy is not classic. In many papers (see [5], [6], [15]) when  $\mathcal{A}$  is the Laplace operator, the authors rather look for a distributed control of the form  $V_d(t, x) = u(t)f(x)$ , where  $f$  is a well chosen profile function, supported in  $\omega$ , and satisfying some lower bounds for its generalized Fourier coefficients  $|\int_{\Omega} f \phi_k dx|$ . However, it is not straightforward to find such a  $f$  when  $\mathcal{A}$  is a more general elliptic operator as in (2.1), since we do not have analytic formulas for the  $\phi_k$  (and, a fortiori, finding the equivalent of  $f$  in the discrete setting seems to be even more complicated).*

*Thus, the ansatz  $V_d(t, x) = \sum_{k \geq 1} \alpha_k q_k^{\Lambda}(t) \phi_k(x)$  we choose here is somehow more convenient since it does not require to find such a function  $f$ . Notice that the family  $(t, x) \mapsto q_k^{\Lambda}(t) \phi_k(x) / \|\phi_k\|_{L^2(\omega)}^2$  can be seen as a space-time biorthogonal family of  $(t, x) \mapsto e^{-(T-t)\lambda_k} \phi_k(x)$  in  $L^2((0, T) \times \omega)$ .*

*This form of distributed control has been used in [18] to prove exact controllability of the 1d-wave equation.*

To sum up, in order to justify the previous application of the moment method, we must check that

1. Such a biorthogonal family  $(q_k^{\Lambda})_{k \geq 1}$  exists.
2. The formal series (2.53) (resp. (2.56)) that defines  $V_b$  (resp.  $V_d$ ) converges. To this end, we need
  - (a) to estimate the  $L^2(0, T)$ -norms of  $q_k^{\Lambda}$ , for all  $k \geq 1$ ,
  - (b) to give lower bounds on  $\|\phi_k\|_{L^2(\omega)}$  and  $|\partial_r \phi_k|$ , that appear at the denominator in those formulas.

The lower bounds in the point 2b were stated in the assertions 2 and 3 of Theorem 2.1.1. We shall now tackle points 1 and 2a at the same time.

Problems of existence and bounds on the biorthogonal family have been studied in [15] and [4] and we recall below some useful results.

First, we need to extend the definition of a biorthogonal family given above, for the purpose of controlling *systems of coupled parabolic equations*.

**Definition 2.5.1** *Let  $\Sigma := (\sigma_k)_{k \geq 1}$  be a sequence of positive real numbers. Let  $T > 0$  and  $m \in \mathbb{N}$ . A biorthogonal family of  $((t - T)^i e^{-\sigma_k(T-t)})_{i \in \llbracket 0, m \rrbracket}^{k \geq 1}$  is a set of functions in  $L^2(0, T)$  denoted by  $(q_{j,l}^{\Sigma})_{j \in \llbracket 0, m \rrbracket}^{l \geq 1}$  satisfying*

$$\int_0^T (t - T)^i \exp(-\sigma_k(T - t)) q_{j,l}^{\Sigma}(t) dt = \delta_{k,l} \delta_{i,j}, \quad \forall k, l \geq 1, \forall i, j \in \llbracket 0, m \rrbracket.$$

Before stating an existence result and estimates for such biorthogonal families in Theorem 2.5.1, we need the following definition (see [15, Theorem 1.1])

**Definition 2.5.2 (class of sequences  $\mathcal{L}(\rho, \mathcal{N})$ )** *Let  $\rho > 0$  and let  $\mathcal{N} : \mathbb{R}^+ \rightarrow \mathbb{N}$ . We denote by  $\mathcal{L}(\rho, \mathcal{N})$  the class of all sequences of positive numbers  $\Sigma = (\sigma_k)_{k \geq 1}$  that satisfy the conditions:*

$$\sigma_{k+1} - \sigma_k \geq \rho, \quad \forall k \geq 1,$$

$$\sum_{k=\mathcal{N}(\xi)}^{\infty} \frac{1}{\sigma_k} \leq \xi, \quad \forall \xi > 0.$$

**Theorem 2.5.1** *Let  $T > 0$  and  $m \in \mathbb{N}$ . Let  $\rho > 0$  and  $\mathcal{N} : \mathbb{R}_+ \rightarrow \mathbb{N}$ .*

*For any  $\tau > 0$ , there exists  $K(\tau, T, \rho, \mathcal{N}, m) > 0$  such that for any sequence  $\Sigma := (\sigma_k)_{k \geq 1}$  in the class  $\mathcal{L}(\rho, \mathcal{N})$ , there exists a biorthogonal family  $(q_{j,l}^{\Sigma})_{j \in \llbracket 0, m \rrbracket}^{l \geq 1}$  in  $L^2(0, T)$  for  $((t - T)^i e^{-\sigma_k(T-t)})_{i \in \llbracket 0, m \rrbracket}^{k \geq 1}$  such that*

$$\|q_{j,l}^{\Sigma}\|_{L^2(0, T)} \leq K e^{\tau \sigma_l}, \quad \forall l \geq 1, \forall j \in \llbracket 0, m \rrbracket.$$

**Remark 2.5.2** *We emphasize that the upper bound  $K e^{\tau \sigma_l}$  does not depend on the choice of the particular sequence  $\Sigma$  in the class  $\mathcal{L}(\rho, \mathcal{N})$ . Thus, if there exist  $\rho$  and  $\mathcal{N}$  such that  $\forall h > 0, \Lambda^h \in \mathcal{L}(\rho, \mathcal{N})$ , then the sequences  $\|q_{j,l}^{\Lambda^h}\|_{L^2(0, T)}$  are uniformly bounded with respect to parameter  $h$ . This point is essential to get uniform bounds with respect to  $h$  on boundary and distributed controls  $V_b^h$  and  $V_d^h$ .*

For  $m = 0$ , this result is proved in [15, Theorem 1.1] and for  $m \geq 1$  it is proved in [4, Theorem 1.2]. In this last reference, the dependence of the bounds with respect to the sequence  $\Sigma$  is not precised but a careful inspection of the proof shows that the constants are actually uniform in the class  $\mathcal{L}(\rho, \mathcal{N})$  as in [15].

We can now completely justify the moments method applied to system (2.48) that we described formally above. For the sake of completeness, we state the following theorem in the more general case of a *cascade system of  $d \geq 1$  parabolic equations*. We introduce a control vector  $B \in \mathbb{R}^d$  and a coupling  $d \times d$  matrix  $C$  defined as follows

$$B = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & \text{---} & 0 \\ 1 & \diagdown & \\ 0 & \diagdown & \\ \vdots & \diagdown & \\ 0 & \text{---} & 0 & 1 & 0 \end{pmatrix}.$$

Let  $\omega$  be a non empty open subset of  $\Omega$  and  $T > 0$ . For any  $Y^0 \in (H^{-1}(\Omega))^d$ ,  $V_b \in L^2(0, T)$ ,  $V_d \in L^2((0, T) \times \Omega)$  the problem

$$\begin{cases} \partial_t Y + \mathcal{A}Y + CY = 1_\omega BV_d(t, x) \text{ in } (0, T) \times \Omega, \\ Y(t, 0) = 0, \quad Y(t, 1) = BV_b(t) \text{ in } (0, T), \\ Y(0, x) = Y^0(x) \text{ in } \Omega, \end{cases} \quad (2.57)$$

has a unique solution in  $C^0([0, T], (H^{-1}(\Omega))^d)$ . Observe that  $Y$  has  $d$  components (and  $\mathcal{A}$  acts component-by-component on  $Y$ ) but the controls  $V_b$  and  $V_d$  are scalar.

**Theorem 2.5.2** *Assume that  $(H_0)$  holds. For any  $Y^0 \in (H^{-1}(\Omega))^d$ , System (2.57) is null controllable at time  $T$  with either a distributed control  $V_d \in L^2((0, T) \times \Omega)$  (in this case, we set  $V_b = 0$ ) or a boundary control  $V_b \in L^2(0, T)$  (in this case, we set  $V_d = 0$ ).*

**Remark 2.5.3** *Notice that when  $d = 1$ , system (2.57) is nothing but system (2.48).*

*Notice also that, as usual in the study of controllability properties of systems with fewer controls than equations, not every equation is directly controlled. Indeed, the control directly intervenes only in the evolution equation for the first component of the solution. Thanks to the particular structure of the coupling matrix  $C$ , the second component is indirectly controlled by means of the first component itself and so on.*

**Proof** As an example, we start by considering the case  $d = 1$ . The first assertion of Theorem

2.1.1 as well as (2.2) ensure that Theorem 2.5.1 applies with  $m = 0$  and  $\Sigma = \Lambda$ . Choosing for instance  $\tau = T/2$ , and using the second and third assertions of Theorem 2.1.1, one can easily check that the series (2.53) (resp. the series (2.56)) converges in  $L^2(0, T)$  (resp. in  $L^2((0, T) \times \Omega)$ ), which justifies the formal approach and proves the claim.

The same kind of arguments apply to the case  $d > 1$ . First, note that the eigenvalues of the adjoint operator  $\mathcal{L} := \mathcal{A} + {}^t C$  in  $(L^2(\Omega))^d$  are the  $(\lambda_k)_{k \geq 1}$ , as in the case  $d = 1$ . They are geometrically simple and possess a  $d \times d$  Jordan bloc. More precisely if we define,  $\Phi_k^r := \phi_k e_r$  for  $r \in \llbracket 1, d \rrbracket$  where  $(e_1, \dots, e_d)$  is the canonical basis of  $\mathbb{R}^d$  and  $\Phi_k^0 = 0$ , then we have

$$\mathcal{L}\Phi_k^r = \mathcal{A}\Phi_k^r + {}^t C\Phi_k^r = \lambda_k \Phi_k^r + \Phi_k^{r-1}, \quad \forall r \in \llbracket 1, d \rrbracket,$$

so that  $\Phi_k^1$  is an eigenfunction, and the  $\Phi_k^r$ ,  $r \geq 2$  are generalized eigenfunctions.

With those notations we can explicitly compute the semi-group associated with  $\mathcal{L}$  for each initial data  $\Phi_k^r$ , as follows

$$e^{-s\mathcal{L}}\Phi_k^r = \sum_{l=0}^{r-1} \frac{(-s)^l}{l!} e^{-\lambda_k s} \Phi_k^{r-l}, \quad \forall s \geq 0.$$

— The *right-boundary null-control problem* consists in finding a control  $V_b \in L^2(0, T)$  such that  $\forall k \geq 1, \forall r \in \llbracket 1, d \rrbracket$ ,

$$\langle Y^0, e^{-T\mathcal{L}}\Phi_k^r \rangle_{H^{-1} \times H_0^1} = \sum_{l=0}^{r-1} (B, \partial_x \Phi_k^{r-l}(1)) \gamma(1) \int_0^T V_b(t) \frac{(t-T)^l}{l!} e^{-\lambda_k(T-t)} dt,$$

where  $(\cdot, \cdot)$  denotes the canonical inner product of  $\mathbb{R}^N$ . By definition of  $B$  and  $\Phi_k^{r-1}$ , we see that  $(B, \partial_x \Phi_k^{r-1}(1)) = \delta_{l,r-1} \partial_x \phi_k(1)$ . We are thus led to solve the following moments problem: find  $V_b \in L^2(0, T)$  such that for any  $k \geq 1$  and  $r \in \llbracket 1, d \rrbracket$

$$\langle Y^0, e^{-T\mathcal{L}} \Phi_k^r \rangle_{H^{-1} \times H_0^1} = \frac{\gamma(1) \partial_r \phi_k}{(r-1)!} \int_0^T V_b(t) (t-T)^{r-1} e^{-\lambda_k(T-t)} dt.$$

The gap estimate in Theorem 2.1.1 (first assertion) and the inequality (2.2) ensure that we can apply Theorem 2.5.1 with  $m = d-1$  and  $\tau = T/2$  (for instance) and obtain a biorthogonal family whose norms are bounded by  $e^{\lambda_k T/2}$ . Then, we observe that

$$\|e^{-T\mathcal{L}} \Phi_k^r\|_{H_0^1} \leq e^{-\lambda_k T} \left( \sum_{l=0}^{r-1} \frac{T^l}{l!} \right) \|\phi_k\|_{H_0^1} \leq C e^{-\lambda_k T} \sqrt{\lambda_k}.$$

Finally, with the uniform lower bound on  $\partial_r \phi_k$  given in Theorem 2.1.1 we can conclude that the following definition of  $V_b$

$$V_b(t) = \sum_{k \geq 1} \sum_{r=1}^d (r-1)! \frac{\langle -Y^0, e^{-T\mathcal{L}} \Phi_k^r \rangle_{H^{-1} \times H_0^1}}{\gamma(1) \partial_r \phi_k} q_{r-1,k}^\Lambda(t), \quad (2.58)$$

is actually a series that converges in  $L^2(0, T)$  and which is a solution of our boundary control problem.

— Considering now the *distributed control problem*, we look for a control  $V_d$  of the following form

$$V_d(t, x) = \sum_{k \geq 1} \sum_{r=1}^d \alpha_{k,r} q_{r-1,k}^\Lambda(t) \phi_k(x),$$

with the coefficients  $\alpha_{k,r}$  to be determined. By using formally this expression in the weak formulation of the problem, we obtain

$$\alpha_{k,r} = (r-1)! \frac{\langle -Y^0, e^{-T\mathcal{L}} \Phi_k^r \rangle_{H^{-1} \times H_0^1}}{\|\phi_k\|_{L^2(\omega)}^2}, \quad \forall k \geq 1, \forall r \in \llbracket 1, d \rrbracket.$$

Still using the bounds on the biorthogonal family  $(q_{r-1,k}^\Lambda)_{r \in \llbracket 0, d-1 \rrbracket}^{k \geq 1}$ , the bound on  $\|e^{-T\mathcal{L}} \Phi_k^r\|_{H_0^1}$  and the uniform lower bound on  $\|\phi_k\|_{L^2(\omega)}$  given in Theorem 2.1.1, we obtain that the series

$$V_d(t, x) = \sum_{k \geq 1} \sum_{r=1}^d (r-1)! \frac{\langle -Y^0, e^{-T\mathcal{L}} \Phi_k^r \rangle_{H^{-1} \times H_0^1}}{\|\phi_k\|_{L^2(\omega)}^2} q_{r-1,k}^\Lambda(t) \phi_k(x) \quad (2.59)$$

converges in  $L^2((0, T) \times \Omega)$  and is solution to our distributed control problem.

■

## 2.5.2 Null controllability in the discrete setting

We now want to apply this strategy to a discrete version of (2.57). For a given mesh, we consider the discrete control problem

$$\begin{cases} (Y^h)'(t) + \mathcal{A}^h Y^h(t) + C^h Y^h(t) = B^h (\mathcal{D}_\omega^h V_d^h(t) + \mathcal{B}_r^h V_b^h(t)), & \text{for } 0 < t \leq T, \\ Y^h(0) = Y^{0,h} \in (\mathbb{R}^N)^d, \end{cases} \quad (S_d^h)$$

where  $B^h$  is the  $(Nd) \times N$  matrix and  $C^h$  the  $(Nd) \times (Nd)$  matrix given by

$$B^h = \begin{pmatrix} \mathbf{I}^h \\ 0^h \\ \vdots \\ 0^h \end{pmatrix}, \quad C^h = \begin{pmatrix} 0^h & \text{---} & 0^h \\ \mathbf{I}^h & \diagdown & \vdots \\ 0^h & \diagdown & \vdots \\ \vdots & \diagdown & \vdots \\ 0^h & -0^h & \mathbf{I}^h & 0^h \end{pmatrix},$$

that approximate the control vector  $B$  and the coupling matrix  $C$ . Here we still use the notation  $\mathcal{A}^h$  for the component-by-component discrete elliptic operator.

The solutions of this system satisfy the discrete analogue of (2.49) that is:  $\forall k \in \llbracket 1, N \rrbracket, \forall r \in \llbracket 1, d \rrbracket$ ,

$$\begin{aligned} & \left\langle Y^h(T), \Phi_k^{r,h} \right\rangle_{L^2(\Omega_h)} - \left\langle Y^{0,h}, e^{-T\mathcal{L}^h} \Phi_k^{r,h} \right\rangle_{L^2(\Omega_h)} \\ &= \int_0^T \left\langle B^h \mathcal{D}_\omega^h V_d^h(t), e^{-(T-t)\mathcal{L}^h} \Phi_k^{r,h} \right\rangle_{L^2(\Omega_h)} dt \\ & \quad - \int_0^T V_b^h(t) \gamma_{N+1/2} \partial_r \phi_k^h \frac{(t-T)^{r-1}}{(r-1)!} \exp(-\lambda_k^h(T-t)) dt, \end{aligned} \quad (2.60)$$

where  $\mathcal{D}_\omega^h$  is defined in (2.7). Here,  $\Phi_k^{r,h}$  is a column vector of size  $dN$  and  $(\Phi_k^{r,h})_i = (\phi_k^h)_{i-(r-1)N}$  for  $i \in \llbracket 1 + (r-1)N, rN \rrbracket$  and 0 otherwise.

We have used here that the semi-group associated with the adjoint operator  $\mathcal{L}^h := \mathcal{A}^h + {}^t C^h$  satisfies

$$e^{-s\mathcal{L}^h} \Phi_k^{r,h} = \sum_{l=0}^{r-1} \frac{(-s)^l}{l!} e^{-\lambda_k^h s} \Phi_k^{r-l,h}.$$

In the discrete setting, we not only want to control the discrete system for any mesh but also to have uniform bounds on them, with respect to  $h$ , in order to be able to conclude that, at least in a weak sense, the discrete controls will eventually converge towards a control of the continuous problem.

Therefore, in this setting, for a given family of initial data  $(Y^{0,h})_h$ , the null-control problem consists in finding distributed controls  $(V_d^h)_{h>0}$  (resp. boundary controls  $(V_b^h)_{h>0}$ ) such that the corresponding solution  $Y^h$  with  $V_b^h = 0$  (resp.  $V_d^h = 0$ ) satisfies

$$Y^h(T) = 0, \forall h > 0, \quad (2.61)$$

and such that, for some  $C > 0$  depending only on the data  $(\gamma, q, \omega, \beta)$  and so on,  $(V_d^h)_{h>0}$  and  $(V_b^h)_{h>0}$  satisfy

$$\|V_d^h\|_{L^2(\Omega_h^T)} \leq C \|Y^{0,h}\|_{L^2(\Omega_h)}, \text{ and } \|V_b^h\|_{L^2(0,T)} \leq C \|Y^{0,h}\|_{L^2(\Omega_h)}, \quad \forall h > 0. \quad (2.62)$$

In the discrete case, we will refer to uniform null controllability as the combination of condition (2.61) and (2.62). We start by considering the setting (S3), which is the simplest one.

**Theorem 2.5.3** *Let  $d \in \mathbb{N}^*$  and  $T > 0$ . Consider a uniform mesh and suppose that  $\gamma$  is a constant function, while  $q$  is any continuous function. Then, the discrete cascade system  $(S_d^h)$  is uniformly null controllable at time  $T$  either with a distributed control, or with a boundary control.*

**Proof** Let us introduce the sequence

$$\tilde{\Lambda}^h := \begin{cases} \lambda_k^h & \text{for } k \in \llbracket 1, N \rrbracket, \\ \lambda_N^h + 4\gamma k^2 & \text{for } k \geq N + 1. \end{cases}$$

Notice that  $(\tilde{\Lambda}^h)_k \geq 4\gamma k^2 - \|q\|_\infty$  for any  $k \geq 1$  (see Lemma 2.3.3 and Remark 2.1.3). First, we prove that there exists  $\rho > 0$  and  $\mathcal{N}$ , an integer-valued function, such that  $\tilde{\Lambda}^h \in \mathcal{L}(\rho, \mathcal{N})$  for any  $h > 0$ . Let  $\mathcal{N} : \mathbb{R}_+ \rightarrow \mathbb{N}$  be a function satisfying

$$\sum_{k \geq \mathcal{N}(\eta)} \frac{1}{4\gamma k^2 - \|q\|_\infty} \leq \eta, \quad \forall \eta > 0. \quad (2.63)$$

According to Corollary 2.3.1 we know that there exists  $\kappa > 0$  such that, for any  $h > 0$ , we have

$$\lambda_{k+1}^h - \lambda_k^h \geq \kappa, \quad \forall k \in \llbracket 1, N-1 \rrbracket. \quad (2.64)$$

One can check that

$$\tilde{\Lambda}^h \in \mathcal{L}(\min(\kappa, 4\gamma), \mathcal{N}), \quad \forall h > 0. \quad (2.65)$$

Thus, we can apply Theorem 2.5.1 which states that given any  $\tau \in (0, T)$ , there exists a  $K > 0$ , such that for and any  $h > 0$ , there exists a biorthogonal family  $(q_{r,k}^{\tilde{\Lambda}^h})_{r \in \llbracket 0, d-1 \rrbracket}^{k \in \llbracket 1, N \rrbracket}$

satisfying

$$\|q_{r,k}^{\tilde{\Lambda}^h}\|_{L^2(0,T)} \leq K e^{\tau\lambda_k^h}, \quad \forall h > 0, \quad \forall k \in \llbracket 1, N \rrbracket, \quad \forall r \in \llbracket 0, d-1 \rrbracket. \quad (2.66)$$

It is fundamental to notice that the upper bound in (2.66) is valid for any  $h > 0$  even though the sequence of eigenvalues depends on  $h$ , thanks to (2.65), see also Remark 2.5.2.

From equation (2.60) and adapting the strategy of Section 2.5.1, we find the following expressions

$$V_b^h(t) = \sum_{k=1}^N \sum_{r=1}^d (r-1)! \frac{\langle Y^{0,h}, e^{-T\mathcal{L}^h} \Phi_k^{h,r} \rangle_{L^2(\Omega_h)}}{\gamma \partial_r \phi_k^h} q_{r-1,k}^{\tilde{\Lambda}^h}(t), \quad \text{and } V_d^h(t) = 0, \quad (2.67)$$

for the boundary control case and

$$V_d^h(t) = \sum_{k=1}^N \sum_{r=1}^d (r-1)! \frac{-\langle Y^{0,h}, e^{-T\mathcal{L}^h} \Phi_k^{h,r} \rangle_{L^2(\Omega_h)}}{\|\phi_k^h\|_{L^2(\omega_h)}^2} q_{r-1,k}^{\tilde{\Lambda}^h}(t) \phi_k^h, \quad \text{and } V_b^h(t) = 0. \quad (2.68)$$

for the distributed control case.

We deal now with finite sums and there is no series convergence problem as soon as  $\partial_r \phi_k^h \neq 0$  (this is always true) and  $\|\phi_k^h\|_{L^2(\omega_h)} \neq 0$  (which is true as soon as  $\omega$  contains at least two different points of the mesh, see Remark 2.1.4).

However, it remains to check that condition (2.62) is satisfied. According to the lower bounds in Corollary 2.3.1 and to (2.66), we find that for any  $h > 0$ , the distributed control in (2.68) satisfies

$$\|V_d^h\|_{L^2(\Omega_h^T)} \leq C \|Y^{0,h}\|_{L^2(\Omega_h)} \sum_{k=1}^N \lambda_k^h e^{-(T-\tau)\lambda_k^h} \leq C \|Y^{0,h}\|_{L^2(\Omega_h)} \sum_{k \geq 1} k^2 e^{-(T-\tau)4\gamma k^2},$$

and thus,

$$\|V_d^h\|_{L^2(\Omega_h^T)} \leq C \|Y^{0,h}\|_{L^2(\Omega_h)}.$$

The same computations hold for  $(V_b^h)_{h>0}$  in (2.67).

**Remark 2.5.4** Notice that, for the problem of control under consideration, we did not need so much precision on the lower bounds of Corollary 2.3.1. Indeed, controls  $(V_b^h)_{h>0}$  and  $(V_d^h)_{h>0}$  satisfy condition (2.62) as soon as the sums (2.67) and (2.68) are bounded uniformly in  $h$ . Thus, for this purpose, results of Theorem 2.3.1 would have been enough.

■

### 2.5.3 $\phi(h)$ -null controllability

As already mentioned in the last paragraph of Section 2.1.3 and observed in numerical simulations of Section 2.4, the gap property (2.4) as well as the uniform lower bounds for  $|\partial_b \phi_k^h|$  or  $\|\phi_k^h\|_{L^2(\omega_h)}$  may not be satisfied when the diffusion coefficient is not constant or when the mesh is not uniform. Nevertheless, by Remark 2.1.4, we may still follow the proof of Theorem 2.5.3 exactly the same way and produce semi-discrete controls for which the associated solution satisfies  $Y^h(T) = 0$ . Unfortunately, due to the lack of uniformity for the spectral properties mentioned above, we will not be able to produce a uniform bound for the control costs  $\|V_d^h\|_{L^2(\Omega_h^T)}$  or  $\|V_b^h\|_{L^2(0,T)}$ . It follows that such semi-discrete null-controls may be unstable when  $h \rightarrow 0$ .

Therefore, in these latter cases, we will no longer look for controls that lead to (2.61) and instead, we consider an adapted weaker definition of null-controllability, namely, we now investigate the  $\phi(h)$ -null controllability problem (see Definition 2.1.1).

**Remark 2.5.5** Note that the property  $\|Y^h(T)\|_{L^2(\Omega_h)}^2 \leq C \phi(h) \|Y^{0,h}\|_{L^2(\Omega_h)}^2$  ensures that  $Y^h(T) \rightarrow 0$  as  $h \rightarrow 0$  if the family of discrete initial data are bounded. Thus, the  $\phi(h)$ -null controllability problem really aims at approaching null-controls in the limit  $h \rightarrow 0$  and not approximate controls.

Let us first state a lemma on which the proofs of Theorem 2.5.4 and Theorem 2.5.5 both rely.

**Lemma 2.5.1** Let  $t_0 \in (0, T)$ ,  $K \geq 1$ , and  $\beta > 0$ . There exists a  $C(q, \gamma, t_0, \beta) > 0$  such that

— for any mesh satisfying  $\Theta_h \leq \beta$  and  $N > K$



— for any  $V_d^h \in L^2(0, T, \mathbb{R}^N)$  and  $V_b^h \in L^2(0, T)$  that vanish on  $(t_0, T)$  for which the associated solution  $Y^h$  of  $(S_d^h)$  satisfies

$$\left\langle \Phi_k^{r,h}, Y^h(t_0) \right\rangle_{L^2(\Omega_h)} = 0, \quad \forall k \in \llbracket 1, K \rrbracket, \quad \forall r \in \llbracket 1, d \rrbracket, \quad (2.69)$$

we have

$$\|Y^h(T)\|_{L^2(\Omega_h)} \leq C e^{-(T-t_0)\lambda_{K+1}^h} \left( \|Y^{0,h}\|_{L^2(\Omega_h)} + \frac{1}{h^{3/2}} \|V_b^h\|_{L^2(0,T)} + \|V_d^h\|_{L^2(\Omega_h^T)} \right).$$

**Proof** Since the family  $(\Phi_k^{r,h})_{r \in \llbracket 1, d \rrbracket}^{k \in \llbracket 1, N \rrbracket}$ , is an orthonormal basis of  $\mathbb{R}^{Nd}$ , we have

$$\|Y^h(t_0)\|_{L^2(\Omega_h)}^2 = \sum_{k=1}^N \sum_{r=1}^d \left\langle Y^h(t_0), \Phi_k^{r,h} \right\rangle_{L^2(\Omega_h)}^2.$$

Now, using (2.60) with  $T$  replaced by  $t_0$  we find that

$$\begin{aligned} \|Y^h(t_0)\|_{L^2(\Omega_h)}^2 &\leq 3 \sum_{k=1}^N \sum_{r=1}^d \left| \left\langle Y^{0,h}, \sum_{l=0}^{r-1} \Phi_k^{r-l,h} \frac{(-t_0)^l}{l!} \exp(-\lambda_k^h t_0) \right\rangle_{L^2(\Omega_h)} \right|^2 \\ &+ \left| \int_0^{t_0} \left\langle V_d^h(t), {}^t \mathcal{D}_\omega^h B^h \sum_{l=0}^{r-1} \Phi_k^{r-l,h} \frac{(t-t_0)^l}{l!} \exp(-\lambda_k^h(t-t_0)) \right\rangle_{L^2(\Omega_h)} dt \right|^2 \\ &+ \left( \frac{\Theta_h \|\gamma\|_\infty \|\phi_k^h\|_{L^\infty}}{h} \frac{1}{(r-1)!} \right)^2 \left| \int_0^{t_0} V_b^h(t) (t-t_0)^{r-1} \exp(-\lambda_k^h(t-t_0)) dt \right|^2 \end{aligned}$$

Thus,

$$\begin{aligned} \|Y^h(t_0)\|_{L^2(\Omega_h)}^2 &\leq C(t_0, \beta, \gamma) \sum_{k=1}^N \sum_{r=1}^d \left| \left\langle Y^{0,h}, \Phi_k^{r,h} \right\rangle_{L^2(\Omega_h)} \right|^2 \\ &+ \left| \int_0^{t_0} \left\langle V_d^h(t), {}^t \mathcal{D}_\omega \Phi_k^{1,h} \exp(-\lambda_k^h(t-t_0)) \right\rangle_{L^2(\Omega_h)} dt \right|^2 \\ &+ \frac{1}{h^3} \left| \int_0^{t_0} V_b^h(t) \exp(-\lambda_k^h(t-t_0)) dt \right|^2. \end{aligned}$$

In the last inequality we used that  $\|\phi_k^h\|_{L^\infty}^2 \leq \frac{\Theta_h}{h} \|\phi_k^h\|_{L^2(\Omega_h)}^2 \leq \frac{\beta}{h}$ .

$$\|Y^h(t_0)\|_{L^2(\Omega_h)}^2 \leq C \left[ \|Y^{0,h}\|_{L^2(\Omega_h)}^2 + \|V_d^h\|_{L^2(\Omega_h^T)}^2 \sum_{k=1}^N \frac{1}{\lambda_k^h} + \frac{1}{h^3} \|V_b^h\|_{L^2(0,T)}^2 \sum_{k=1}^N \frac{1}{\lambda_k^h} \right].$$

Finally, taking the square root and using the lower bounds for  $\lambda_k^h$  given in Lemma 2.3.3, we conclude that

$$\|Y^h(t_0)\|_{L^2(\Omega_h)} \leq C \left[ \|Y^{0,h}\|_{L^2(\Omega_h)} + \|V_d^h\|_{L^2(\Omega_h^T)} + \frac{1}{h^{3/2}} \|V_b^h\|_{L^2(0,T)} \right]. \quad (2.70)$$

Now, we take advantage of the assumption (2.69) that implies the exponential decay of coefficients  $\left\langle \Phi_k^{r,h}, Y^h(t) \right\rangle_{L^2(\Omega_h)}$  for  $t \geq t_0$ . Indeed, since both source terms  $V_d^h$  and  $V_b^h$  are null after time  $t_0$ , the solution at time  $T$  writes,

$$Y^h(T) = \sum_{r=1}^d \sum_{k=K+1}^N \left\langle \Phi_k^{r,h}, Y^h(t_0) \right\rangle_{L^2(\Omega_h)} \exp(-\lambda_k^h(T-t_0)) \Phi_k^{r,h}.$$

Thus,

$$\|Y^h(T)\|_{L^2(\Omega_h)} \leq \exp(-\lambda_{K+1}^h(T-t_0)) \|Y^h(t_0)\|_{L^2(\Omega_h)},$$

which gives the claim with estimate (2.70).  $\blacksquare$

Now we can state the main theorem of this section which applies in setting (S2).

**Theorem 2.5.4** *Let  $T > 0$  and suppose that the mesh is uniform and that  $q$  and  $\gamma$  satisfy hypothesis  $(H_0)$ . Let any function  $\phi : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$  such that*

$$\liminf_{h \rightarrow 0} [h^2 \log(\phi(h))] > -8\gamma_{\min} T. \quad (2.71)$$

*Then, system  $(S_d^h)$  is  $\phi(h)$ -null controllable in time  $T$ .*

**Proof** The proof follows the same lines as the ones of Theorem 2.5.3 except that we only set to 0 a given portion of the Fourier modes of the solution.

From (2.71) we can find  $t_0 \in (0, T)$  and  $\varepsilon \in (0, 1)$  (depending only on  $\phi$ ) such that

$$\liminf_{h \rightarrow 0} [h^2 \log(\phi(h))] > -8\gamma_{\min}(T - t_0)(1 - \varepsilon). \quad (2.72)$$

Let  $k_{\max, \varepsilon}^h$  as in (2.29). We define  $\tilde{\Lambda}_\varepsilon^h$  as follows

$$\tilde{\Lambda}_\varepsilon^h := \begin{cases} \lambda_k^h & \text{for } k \in \llbracket 1, k_{\max, \varepsilon}^h \rrbracket, \\ \lambda_{k_{\max, \varepsilon}^h}^h + 4\gamma_{\min} k^2 & \text{for } k \geq k_{\max, \varepsilon}^h + 1. \end{cases}$$

Thanks to the gap estimate in Theorem 2.3.2, there exists a  $\kappa_\varepsilon > 0$  such that

$$\tilde{\Lambda}_\varepsilon^h \in \mathcal{L}(\kappa_\varepsilon, \mathcal{N}), \quad \text{for any } h, \text{ where } \mathcal{N} \text{ is defined by (2.63).}$$

We apply Theorem 2.5.1 with  $T$  replaced by  $t_0$  and  $\tau = t_0/2$ , in such a way that the biorthogonal family we obtain in  $L^2(0, t_0)$  satisfies

$$\|q_{j,l}^{\tilde{\Lambda}_\varepsilon^h}\|_{L^2(0, t_0)} \leq K_{t_0, \varepsilon} e^{\lambda_l^h t_0/2}, \quad \forall l \in \llbracket 1, k_{\max, \varepsilon}^h \rrbracket, \forall j \in \llbracket 0, d-1 \rrbracket.$$

We define now the following controls on  $(0, t_0)$

$$V_b^h(t) = \sum_{k=1}^{k_{\max, \varepsilon}^h} \sum_{r=1}^d (r-1)! \frac{\langle Y^{0,h}, e^{-t_0 \mathcal{L}^h} \Phi_k^{h,r} \rangle_{L^2(\Omega_h)}}{\gamma(1) \partial_r \phi_k^h} q_{r-1, k}^{\tilde{\Lambda}_\varepsilon^h}(t), \quad \text{and } V_d^h(t) = 0 \quad (2.73)$$

in the boundary control case and

$$V_d^h(t) = \sum_{k=1}^{k_{\max, \varepsilon}^h} \sum_{r=1}^d (r-1)! \frac{-\langle Y^{0,h}, e^{-t_0 \mathcal{L}^h} \Phi_k^{h,r} \rangle_{L^2(\Omega_h)}}{\|\phi_k^h\|_{L^2(\omega_h)}^2} q_{r-1, k}^{\tilde{\Lambda}_\varepsilon^h}(t) \phi_k^h, \quad \text{and } V_b^h(t) = 0, \quad (2.74)$$

in the distributed control case. Using the above estimates on the biorthogonal family and the lower bounds on  $\partial_r \phi_k^h$  and  $\|\phi_k^h\|_{L^2(\omega_h)}$  given in Theorem 2.3.2, we obtain

$$\begin{cases} \|V_d^h\|_{L^2(\Omega_h^{t_0})} \leq C_{\varepsilon, t_0} \|Y^{0,h}\|_{L^2(\Omega_h)}, \\ \|V_b^h\|_{L^2(0, t_0)} \leq C_{\varepsilon, t_0} \|Y^{0,h}\|_{L^2(\Omega_h)}, \end{cases} \quad (2.75)$$

and moreover, by construction, we have

$$\left\langle \Phi_k^{r,h}, Y^h(t_0) \right\rangle_{L^2(\Omega_h)} = 0, \quad \forall k \in \llbracket 1, k_{\max, \varepsilon}^h \rrbracket, \forall r \in \llbracket 1, d \rrbracket.$$

Therefore, we can use Lemma 2.5.1 with  $K = k_{\max, \varepsilon}^h$  to conclude that

$$\begin{aligned} \|Y^h(T)\|_{L^2(\Omega_h)} &\leq C_{t_0, \varepsilon} \|Y^{0,h}\|_{L^2(\Omega_h)} \left(1 + \frac{1}{h^{3/2}}\right) e^{-(T-t_0)\lambda_{k_{\max, \varepsilon}^h}^h} \\ &\leq C_{t_0, \varepsilon} \|Y^{0,h}\|_{L^2(\Omega_h)} \frac{1}{h^{3/2}} e^{-(T-t_0)(1-\varepsilon)\frac{4\gamma_{\min}}{h^2}}. \end{aligned}$$

By (2.72), for  $h$  small enough, we conclude that

$$\|Y^h(T)\|_{L^2(\Omega_h)}^2 \leq C_{t_0, \varepsilon} \|Y^{0,h}\|_{L^2(\Omega_h)}^2 \phi(h).$$

With the bounds (2.75), the claim follows.  $\blacksquare$

We can now state an analogous theorem in the setting (S1).

**Theorem 2.5.5** *Let  $T > 0$ . Suppose that  $\gamma$  and  $q$  satisfy  $(H_2)$ . For any  $\beta > 0$ , there exists  $\alpha(q, \gamma, \beta) > 0$ , such that for any function  $\phi : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$  satisfying*

$$\liminf_{h \rightarrow 0} [h^{2/5} \log(\phi(h))] > -\alpha T, \quad (2.76)$$

and any mesh family such that  $\Theta_h \leq \beta$ , the system  $(S_d^h)$  is  $\phi(h)$ -null controllable at time  $T$ .

**Proof** This proof is similar to the proof of the previous theorem and the details are left to the reader. However, we precise below their main differences.

Let  $\alpha > 0$ , as in the last estimate of Theorem 2.3.1 and define

$$\tilde{\Lambda}^h := \begin{cases} \lambda_k^h & \text{for } k \in \llbracket 1, \alpha N^{2/5} - 1 \rrbracket, \\ \lambda_k^h + 4\gamma_{\min} k^2 & \text{for } k \geq \alpha N^{2/5}. \end{cases}$$

Thanks to the gap estimate of Theorem 2.3.1, there exists  $\kappa > 0$  such that

$$\tilde{\Lambda}^h \in \mathcal{L}(\kappa, \mathcal{N}), \quad \text{for any } h.$$

By constructing the same kind of controls as before, we get

$$\left\langle \Phi_k^{r,h}, Y^h(t_0) \right\rangle_{L^2(\Omega_h)} = 0, \quad \forall k \in \llbracket 1, \alpha N^{2/5} \rrbracket, \quad \forall r \in \llbracket 1, d \rrbracket,$$

which, using Lemma 2.5.1 with  $K = \alpha N^{2/5}$ , leads to

$$\|Y^h(T)\|_{L^2(\Omega_h)} \leq C_{t_0} \|Y^{0,h}\|_{L^2(\Omega_h)} \frac{1}{h^{3/2}} \exp\left(-\frac{(T-t_0)\alpha}{h^{2/5}}\right).$$

and the claim follows.

**Remark 2.5.6** *The notion of  $\phi(h)$ -null controllability has been introduced here to remedy the fact that the gap property does not hold for the entire spectrum in the general case. Note that if one is able to prove, on a particular choice of  $\gamma$  and  $q$ , that the gap property is valid for the whole spectrum, then the estimates of Section 2.3 allow to conclude that uniform null controllability holds.*

■

## 2.6 Remarks and further results

1. In [20] the authors consider the problem of null-controllability at the boundary for the semi-discretized in space linear beam equation with hinged boundary conditions and constant diffusion coefficient. They discretize the operator  $\partial_{xxxx}$  with finite differences in 1D on a uniform mesh. The equation writes:

$$\begin{cases} (Y^h)''(t) + (\mathcal{A}^h)^2 Y^h(t) = \mathcal{B}_r^h V_b^h(t), & t \in (0, T) \\ Y^h(0) = Y^{h,0}, & (Y^h)'(0) = Y^{h,1}, \end{cases}$$

where  $Y^{h,0}$  and  $Y^{h,1}$  are vectors in  $\mathbb{R}^N$ . The corresponding adjoint system with final datum  $Z_T^h \in \mathbb{R}^{2N}$  writes

$$\begin{cases} (Z^h)'(t) + L^h Z^h(t) = 0 \\ Z^h(T) = Z_T^h \end{cases}$$

where  $L^h = \begin{pmatrix} 0 & -I^h \\ (\mathcal{A}^h)^2 & 0 \end{pmatrix}$ . The authors show, via the moments method, that uniform null controllability holds for this equation for some initial data whose high frequencies have been filtered out. As for parabolic problems, their proof makes use of explicit computations on the eigenlements  $(\mu_k^h, \psi_k^h)_{1 \leq |k| \leq N}$  of the operator  $L^h$ .

Using the discrete spectral estimates obtained in Section 2.3 of the present paper, it is very likely that one can adapt the ideas of [20] to obtain similar results for more general second order elliptic operator  $\mathcal{A}^h$  and, more importantly, for non uniform grids.

2. We give here some additional information about the theoretical results mentioned in Section 2.4, test case 2.4, where the mesh is uniform,  $q = 0$  and  $\gamma$  is piecewise constant and takes its largest value on the right side of the domain :  $\mathcal{O} := ]0.4, 1[$ . First, note that the equation  $\mathcal{A}^h \phi_k^h = \lambda_k^h \phi_k^h$  defines a linear recurrence relation with constant coefficients for the indices  $i$  such that  $x_i \in \Omega \setminus \mathcal{O}$ , since  $\gamma$  is constant in this region. Using the homogeneous Dirichlet condition at  $x = 0$ , we can prove by explicit computations that  $I_r^h(N) \leq \exp(-\frac{C_1}{h})$  and  $I_1^h(N) \leq \exp(-\frac{C_1}{h})$ , where  $C_1$  depends only on  $\gamma$ . Next, we check that the computations done in Theorem 2.3.2 and Proposition 2.3.2 are still valid locally where  $\gamma$  is constant and in particular in  $\mathcal{O}$ . Doing this, we can show that  $I_r^h(N) \geq C_2$  and  $I_2^h(N) \geq C_2|\omega^2|$ , where  $C_2$  only depends on  $\gamma$ .
3. In Remark 2.5.4 we stressed that, for the distributed control problem on a uniform mesh, the exponentially small lower bound of Theorem 2.3.1 was enough to give a uniform bound on  $V_d^h$ . In [10, Th 6.1], the authors show that a partial discrete Lebeau-Robbiano inequality holds for a discretization of a second order elliptic operator of the form  $\mathcal{A} = -\partial_x(\gamma\partial_x\cdot)$ , on regular families of non-uniform meshes (as the ones considered in Remark 2.1.6). Thus, considering a uniform discretization, they obtain that there exist  $C > 0$  and  $\varepsilon > 0$ , both independent of  $h$ , such that for any eigenvalue  $\lambda_k^h$  satisfying  $\lambda_k^h \leq \varepsilon/h^2$ , we have  $\|\phi_k^h\|_{L^2(\omega_h)} \geq \frac{1}{C} \exp(-C\sqrt{\lambda_k^h})$ . This is the same lower bound as the one of Theorem 2.3.1, but only for a fraction of the spectrum, and only for regular grids and coefficients. Note also that the Lebeau-Robbiano inequalities do not give any information on the normal derivatives of the eigenfunctions.
4. The techniques used in this paper seem to be clearly restricted to the 1D situation. However, in a forthcoming paper [1], we manage to use those results in addition to discrete Lebeau-Robbiano techniques to obtain similar results for the boundary control of multi-dimensional coupled systems in Cartesian geometries.

Observe that the multi-dimensional situation has a more complex behavior since, contrary to the 1D case (see Remark 2.1.4 and the beginning of Section 2.5.3), it may happen that the semi-discrete problem is not even approximately controllable for a constant coefficient operator discretized on a uniform mesh. This is a consequence of the counter-example by O. Kavian reported in [24] that shows that it exists eigenfunctions of the 5-point 2D discrete Laplace operator that are localized along the diagonal of the domain. Therefore, the associated discrete normal derivative is zero on almost every point of the boundary of the domain.

## 2.7 Appendix

This appendix is an extension of article [3] and does not appear in the original version of paper [3]. Section 2.7.1 completes Remark 2.1.6. Appendix 2.7.2 proposes a complete proof of Proposition 2.3.1. In section 2.7.3, we precise the second Remark of section 2.6 analyzing some spectral properties of elliptic operators with a piecewise constant diffusion coefficient  $\gamma$  and a potential equal to zero. Finally, in section 2.7.4, we propose a way to build a null control with separated variables to control the semi discretized in space heat equation on a uniform mesh.

### 2.7.1 The case of non uniform but regular meshes

We give here additional information on Remark 2.1.6.

Let  $\vartheta : [0, 1] \rightarrow [0, 1]$  be a function of class  $\mathcal{C}^3$  such that  $\vartheta(0) = 0$  and  $\vartheta(1) = 1$ . Let

$$\vartheta'_0 := \min_{x \in [0, 1]} \vartheta'(x) \text{ and } \vartheta'_\infty := \max_{x \in [0, 1]} \vartheta'(x).$$

We suppose that  $\vartheta'_0 > 0$ . Let  $N \in \mathbb{N}^*$  and  $h_* = \frac{1}{N+1}$ . We define a mesh  $\{x_0, x_1, \dots, x_{N+1}\}$  on  $[0, 1]$  by

$$x_i := \vartheta(ih_*), \forall i \in \llbracket 0, N+1 \rrbracket.$$

As in section 2.1.2 we define the dual mesh and the variable steps as

$$\forall i \in \llbracket 0, N \rrbracket, x_{i+1/2} := \frac{x_{i+1} + x_i}{2} \text{ and } h_{i+1/2} := x_{i+1} - x_i.$$

For  $i \in \llbracket 1, N \rrbracket$ , we set also  $h_i := x_{i+1/2} - x_{i-1/2} = \frac{h_{i+1/2} + h_{i-1/2}}{2}$ .

Let  $k \in \llbracket 1, N \rrbracket$  and let us write, on this mesh, component  $i$  of equality  $\mathcal{A}^h \phi_k^h = \lambda_k^h \phi_k^h$

$$-\frac{1}{h_i} \left( \frac{\gamma_{i+1/2}}{h_{i+1/2}} (\phi_{k,i+1}^h - \phi_{k,i}^h) - \frac{\gamma_{i-1/2}}{h_{i-1/2}} (\phi_{k,i}^h - \phi_{k,i-1}^h) \right) + q_i \phi_{k,i}^h = \lambda_k^h \phi_{k,i}^h. \quad (2.77)$$

In the sequel, we drop subscript  $k$  and exponent  $h$  in (2.77). Let  $\hat{\gamma}$  be defined as

$$\frac{\hat{\gamma}_{i+1/2}}{h_*} := \frac{\gamma_{i+1/2}}{h_{i+1/2}}$$

then (2.77) becomes

$$-\frac{h_*}{h_i} \frac{1}{h_*} \left( \hat{\gamma}_{i+1/2} \frac{\phi_{i+1} - \phi_i}{h_*} - \hat{\gamma}_{i-1/2} \frac{\phi_i - \phi_{i-1}}{h_*} \right) + q_i \phi_i = \lambda \phi_i \quad (2.78)$$

Let  $\alpha_i := \sqrt{\frac{h_i}{h_*}}$ , we get from (2.78),

$$-\frac{1}{h_*} \left( \hat{\gamma}_{i+1/2} \frac{\phi_{i+1} - \phi_i}{h_*} - \hat{\gamma}_{i-1/2} \frac{\phi_i - \phi_{i-1}}{h_*} \right) + \alpha_i^2 q_i \phi_i = \alpha_i^2 \lambda \phi_i. \quad (2.79)$$

Now, we want to perform a change of variable to get rid of the term  $\alpha_i$  in the right hand side of (2.79) in order to get the same spectrum as the initial spectral problem. Let us analyze the situation in the continuous setting. Let  $f$  be a non vanishing  $\mathcal{C}^2$  function. Notice that

$$\begin{aligned} f \partial_x (\gamma \partial_x (f \phi)) &= f \partial_x (\gamma f \partial_x \phi) + f \partial_x (\gamma (\partial_x f) \phi) \\ &= f \partial_x (\gamma f \partial_x \phi) + f \gamma (\partial_x f) \partial_x \phi + f \partial_x (\gamma \partial_x f) \phi \\ &= \{f \partial_x (\gamma f \partial_x \phi) + (\partial_x f) \gamma f \partial_x \phi\} + f \partial_x (\gamma \partial_x f) \phi. \end{aligned}$$

Note that  $f \partial_x (\gamma f \partial_x \phi) + (\partial_x f) \gamma f \partial_x \phi = \partial_x ((\gamma f^2) \partial_x \phi)$ , therefore,

$$f \partial_x (\gamma \partial_x (f \phi)) = \partial_x ((\gamma f^2) \partial_x \phi) + f \partial_x (\gamma \partial_x f) \phi. \quad (2.80)$$

So suppose that  $\phi$  satisfies

$$-\partial_x (\gamma \partial_x \phi) + q \frac{\phi}{f^2} = \lambda \frac{\phi}{f^2}, \quad (2.81)$$

then

$$-f \partial_x (\gamma \partial_x \phi) + q \frac{\phi}{f} = \lambda \frac{\phi}{f}.$$

Defining

$$\psi := \frac{\phi}{f} \quad (2.82)$$

we get

$$-f \partial_x (\gamma \partial_x (f \psi)) + q \psi = \lambda \psi$$

and with equality (2.80),

$$-\partial_x (\bar{\gamma} \partial_x \psi) + \bar{q} \psi = \lambda \psi, \quad (2.83)$$

where,

$$\bar{\gamma} := \gamma f^2 \text{ and } \bar{q} := q - f \partial_x (\gamma \partial_x f). \quad (2.84)$$

From initial problem (2.81) we got (2.83) which is an eigenvalue problem for operator  $-\partial_x \bar{\gamma} \partial_x + \bar{q}$ .

We now apply this strategy in the discrete setting. We take back (2.79). Equality (2.81) suggests to choose  $f_i$  such that  $\frac{1}{f_i} = \alpha_i$ . Following (2.82) we define  $\psi_i = \frac{\phi_i}{f_i} = \phi_i \alpha_i$ . Then (2.79) becomes

$$\begin{aligned} &-\frac{1}{h_*} \left( \hat{\gamma}_{i+1/2} \frac{\frac{\psi_{i+1}}{\alpha_{i+1}} - \frac{\psi_i}{\alpha_i}}{h_*} - \hat{\gamma}_{i-1/2} \frac{\frac{\psi_i}{\alpha_i} - \frac{\psi_{i-1}}{\alpha_{i-1}}}{h_*} \right) + \alpha_i q_i \psi_i = \alpha_i \lambda \psi_i \\ &-\frac{1}{h_*} \left( \frac{\hat{\gamma}_{i+1/2}}{\alpha_{i+1}} \frac{\psi_{i+1} - \psi_i \frac{\alpha_{i+1}}{\alpha_i}}{h_*} - \frac{\hat{\gamma}_{i-1/2}}{\alpha_{i-1}} \frac{\psi_i \frac{\alpha_{i-1}}{\alpha_i} - \psi_{i-1}}{h_*} \right) + \alpha_i q_i \psi_i = \alpha_i \lambda \psi_i \\ &-\frac{1}{h_*} \left( \frac{\hat{\gamma}_{i+1/2}}{\alpha_i \alpha_{i+1}} \frac{\psi_{i+1} - \psi_i \frac{\alpha_{i+1}}{\alpha_i}}{h_*} - \frac{\hat{\gamma}_{i-1/2}}{\alpha_{i-1} \alpha_i} \frac{\psi_i \frac{\alpha_{i-1}}{\alpha_i} - \psi_{i-1}}{h_*} \right) + q_i \psi_i = \lambda \psi_i. \end{aligned}$$

Let  $\bar{\gamma}_{i+1/2} := \frac{\hat{\gamma}_{i+1/2}}{\alpha_i \alpha_{i+1}}$ , that is

$$\bar{\gamma}_{i+1/2} = \frac{h_*^2}{h_{i+1/2} \sqrt{h_i h_{i+1}}} \gamma_{i+1/2} \quad (2.85)$$

we have

$$-\frac{1}{h_*} \left( \bar{\gamma}_{i+1/2} \frac{\psi_{i+1} - \psi_i \frac{\alpha_{i+1}}{\alpha_i}}{h_*} - \bar{\gamma}_{i-1/2} \frac{\psi_i \frac{\alpha_{i-1}}{\alpha_i} - \psi_{i-1}}{h_*} \right) + q_i \psi_i = \lambda \psi_i$$

therefore,

$$\begin{aligned} & -\frac{1}{h_*} \left( \bar{\gamma}_{i+1/2} \frac{\psi_{i+1} - \psi_i}{h_*} - \bar{\gamma}_{i-1/2} \frac{\psi_i - \psi_{i-1}}{h_*} \right) \\ & + \frac{1}{\alpha_i} \frac{1}{h_*} \left( \bar{\gamma}_{i+1/2} \frac{\alpha_{i+1} - \alpha_i}{h_*} - \bar{\gamma}_{i-1/2} \frac{\alpha_i - \alpha_{i-1}}{h_*} \right) \psi_i + q_i \psi_i = \lambda \psi_i \end{aligned}$$

and,

$$-\frac{1}{h_*} \left( \bar{\gamma}_{i+1/2} \frac{\psi_{i+1} - \psi_i}{h_*} - \bar{\gamma}_{i-1/2} \frac{\psi_i - \psi_{i-1}}{h_*} \right) + \bar{q}_i \psi_i = \lambda \psi_i, \quad (2.86)$$

where

$$\bar{q}_i := \frac{1}{\alpha_i} \frac{1}{h_*} \left( \bar{\gamma}_{i+1/2} \frac{\alpha_{i+1} - \alpha_i}{h_*} - \bar{\gamma}_{i-1/2} \frac{\alpha_i - \alpha_{i-1}}{h_*} \right) + q_i. \quad (2.87)$$

Note that (2.85) and (2.87) are perfectly compatible with (2.84). Indeed,

$$-f \partial_x (\gamma \partial_x f) = -f \partial_x (\bar{\gamma} \times \frac{1}{f^2} \partial_x f) = f \partial_x (\bar{\gamma} \partial_x \frac{1}{f}),$$

and the first term of (2.87) is the discrete analogue of the previous expression. Thus, according to the notations of Remark 2.1.6, the initial eigenvalue problem is now recast as an eigenvalue equation on a uniform mesh :

$$\mathcal{A}_{\bar{\gamma}, \bar{q}}^{h_*} \psi = \lambda \psi \quad (2.88)$$

where  $\bar{\gamma}$  and  $\bar{q}$  satisfy respectively (2.85) and (2.87), and (2.87) can be compactly written as

$$\bar{q}_i = q_i - \frac{1}{\alpha_i} (\mathcal{A}_{\bar{\gamma}, 0}^{h_*} \alpha)_i,$$

like in Remark 2.1.6. Now, we want to apply the results of case (S2) on (2.88) to get estimates on the initial problem  $\mathcal{A}^h \phi = \lambda \phi$ . To do so, we need to check first that  $\bar{\gamma}$  and  $\bar{q}$  satisfy (H<sub>0</sub>), in the following sense

**Proposition 2.7.1** *There exist  $C > 0$  which depends on  $q$ ,  $\gamma$  and  $\vartheta$  and  $\bar{\gamma}_{\min} > 0$  which depends on  $\vartheta$  and  $\gamma$  such that for any uniform mesh of size  $h_*$ ,*

1. for any  $i \in \llbracket 1, N \rrbracket$ ,  $\bar{\gamma}_{i+1/2} \geq \bar{\gamma}_{\min}$  and  $\left| \frac{\bar{\gamma}_{i+1/2} - \bar{\gamma}_{i-1/2}}{h_*} \right| \leq C$ .
2. for any  $i \in \llbracket 1, N \rrbracket$ ,  $|\bar{q}_i| \leq C$ ,

### Proof

1. First,  $\gamma$  satisfies  $\inf_{[0,1]} \gamma > \gamma_{\min}$  for some positive constant  $\gamma_{\min}$ . Secondly, since  $\vartheta$  is  $\mathcal{C}^1$  then, according to the Mean Value Theorem, there exists  $c_{h_*} \in [0, 1]$  such that

$$h_{i+1/2} = \vartheta((i+1)h_*) - \vartheta(ih_*) = \vartheta'(c_{h_*})h_*,$$

so for any  $i \in \llbracket 0, N \rrbracket$ ,

$$\vartheta'_\infty \geq \frac{h_{i+1/2}}{h_*} \geq \vartheta'_0. \quad (2.89)$$

Moreover, since

$$\frac{h_i}{h_*} = \frac{h_{i+1/2} + h_{i-1/2}}{2h_*},$$

we have also for  $i \in \llbracket 1, N \rrbracket$ ,

$$\vartheta'_\infty \geq \frac{h_i}{h_*} \geq \vartheta'_0. \quad (2.90)$$

thus, (2.85) with (2.89) and (2.90) yield that for any  $i \in \llbracket 1, N \rrbracket$ ,  $\bar{\gamma}_{i+1/2} \geq \gamma_{min}$ . Note also that there exists  $\Gamma > 0$  such that for any mesh size  $h_*$  and any  $i \in \llbracket 1, N \rrbracket$ ,  $\bar{\gamma}_{i+1/2} \leq \Gamma$

Moreover,

$$\begin{aligned}
 \left| \frac{\bar{\gamma}_{i+1/2} - \bar{\gamma}_{i-1/2}}{h_*} \right| &= \left| \frac{\gamma_{i+1/2} h_*}{h_{i+1/2} \sqrt{h_i h_{i+1}}} - \frac{\gamma_{i-1/2} h_*}{h_{i-1/2} \sqrt{h_{i-1} h_i}} \right| \\
 &\leq h_* \left| \frac{\gamma_{i+1/2} - \gamma_{i-1/2}}{h_i} \sqrt{\frac{h_i}{h_{i+1}}} \frac{1}{h_{i+1/2}} \right| \\
 &\quad + h_* \left| \frac{\gamma_{i-1/2}}{h_i} \left( \sqrt{\frac{h_i}{h_{i+1}}} \frac{1}{h_{i+1/2}} - \sqrt{\frac{h_i}{h_{i-1}}} \frac{1}{h_{i-1/2}} \right) \right| \\
 &\leq \frac{1}{\vartheta'_0} \|\gamma'\|_\infty \sqrt{\frac{\vartheta'_\infty}{\vartheta'_0}} + h_* \left| \frac{\gamma_{i-1/2}}{\sqrt{h_i}} \left( \frac{1}{h_{i+1/2}^{3/2}} - \frac{1}{h_{i-1/2}^{3/2}} \right) \right| \\
 &\leq \frac{\sqrt{\vartheta'_\infty} \|\gamma'\|_\infty}{(\vartheta'_0)^{3/2}} + \frac{\sqrt{h_*} \|\gamma\|_\infty}{\sqrt{\vartheta'_0}} \frac{3}{2} |h_{i+1/2} - h_{i-1/2}| \sup_{x \in [h_{i-1/2}, h_{i+1/2}]} x^{-5/2} \\
 &\leq \frac{\sqrt{\vartheta'_\infty} \|\gamma'\|_\infty}{(\vartheta'_0)^{3/2}} + \frac{\|\gamma\|_\infty}{(\vartheta'_0)^3} \frac{3}{2} \left| \frac{h_{i+1/2} - h_{i-1/2}}{h_*^2} \right|
 \end{aligned}$$

Since the mesh is regular

$$\left| \frac{h_{i+1/2} - h_{i-1/2}}{h_*^2} \right| = \left| \frac{\vartheta((i+1)h_*) - 2\vartheta(ih_*) + \vartheta((i-1)h_*)}{h_*^2} \right| \leq \|\vartheta''\|_\infty,$$

thus,

$$\left| \frac{\bar{\gamma}_{i+1/2} - \bar{\gamma}_{i-1/2}}{h_*} \right| \leq \frac{\sqrt{\vartheta'_\infty} \|\gamma'\|_\infty}{(\vartheta'_0)^{3/2}} + \frac{\|\gamma\|_\infty}{(\vartheta'_0)^3} \frac{3}{2} \|\vartheta''\|_\infty$$

2. Applying (2.89) and (2.90) on (2.87),

$$\begin{aligned}
 |\bar{q}_i| &= \frac{1}{\sqrt{\vartheta'_0}} \frac{1}{h_*} \left| \bar{\gamma}_{i+1/2} \frac{\alpha_{i+1} - \alpha_i}{h_*} - \bar{\gamma}_{i-1/2} \frac{\alpha_i - \alpha_{i-1}}{h_*} \right| + \|q\|_\infty \\
 &\leq \frac{1}{\sqrt{\vartheta'_0}} \bar{\gamma}_{i+1/2} \frac{1}{h_*} \left| \frac{\alpha_{i+1} - \alpha_i}{h_*} - \frac{\alpha_i - \alpha_{i-1}}{h_*} \right| + \frac{1}{\sqrt{\vartheta'_0}} \left| \frac{\bar{\gamma}_{i+1/2} - \bar{\gamma}_{i-1/2}}{h_*} \right| \left| \frac{\alpha_i - \alpha_{i-1}}{h_*} \right| + \|q\|_\infty,
 \end{aligned}$$

therefore

$$|\bar{q}_i| \leq \frac{1}{\sqrt{\vartheta'_0}} \frac{\bar{\gamma}_{i+1/2}}{\sqrt{h_*}} \frac{1}{h_*} \left| \frac{\sqrt{h_{i+1}} - \sqrt{h_i}}{h_*} - \frac{\sqrt{h_i} - \sqrt{h_{i-1}}}{h_*} \right| + \frac{1}{\sqrt{\vartheta'_0}} \frac{C}{\sqrt{h_*}} \left| \frac{\sqrt{h_i} - \sqrt{h_{i-1}}}{h_*} \right| + \|q\|_\infty. \quad (2.91)$$

Moreover,

$$\begin{aligned}
 \left| \frac{\sqrt{h_i} - \sqrt{h_{i-1}}}{h_*} \right| &\leq \left| \frac{h_i - h_{i-1}}{h_*} \right| \sup_{x \in [h_{i-1}, h_i]} \frac{1}{2\sqrt{x}} \\
 &\leq \left| \frac{x_{i+1/2} - 2x_{i-1/2} + x_{i-3/2}}{h_*} \right| \sup_{x \in [h_{i-1}, h_i]} \frac{1}{2\sqrt{x}} \\
 &\leq h_* \left| \frac{\vartheta((i+1/2)h_*) - 2\vartheta((i-1/2)h_*) + \vartheta((i-3/2)h_*)}{h_*^2} \right| \frac{1}{2\sqrt{\vartheta'_0 h_*}} \\
 &\leq \|\vartheta''\|_\infty \frac{\sqrt{h_*}}{2\sqrt{\vartheta'_0}},
 \end{aligned}$$

thus,

$$\frac{1}{\sqrt{h_*}} \left| \frac{\sqrt{h_i} - \sqrt{h_{i-1}}}{h_*} \right| \leq \frac{\|\vartheta''\|_\infty}{2\sqrt{\vartheta'_0}}. \quad (2.92)$$

Similarly, there exists  $c_{h_*}^\pm \in (h_{i-1}, h_{i+1})$  such that

$$\frac{1}{h_*} \left| \frac{\sqrt{h_{i+1}} - \sqrt{h_i}}{h_*} - \frac{\sqrt{h_i} - \sqrt{h_{i-1}}}{h_*} \right| = \frac{1}{h_*} \left| \frac{h_{i+1} - h_i}{h_*} \frac{1}{2\sqrt{h_i}} + \frac{(h_{i+1} - h_i)^2}{2h_*} \frac{-1}{4(c_{h_*}^+)^{3/2}} \right. \\ \left. - \frac{h_i - h_{i-1}}{h_*} \frac{1}{2\sqrt{h_i}} + \frac{(h_i - h_{i-1})^2}{2h_*} \frac{-1}{4(c_{h_*}^-)^{3/2}} \right|$$

and,

$$\frac{1}{h_*} \left| \frac{\sqrt{h_{i+1}} - \sqrt{h_i}}{h_*} - \frac{\sqrt{h_i} - \sqrt{h_{i-1}}}{h_*} \right| \leq \frac{1}{2\sqrt{h_i}} \frac{1}{h_*} \left| \frac{h_{i+1} - h_i}{h_*} - \frac{h_i - h_{i-1}}{h_*} \right| \\ + \frac{1}{8(h_*)^{3/2}(\vartheta'_0)^{3/2}} \left( \left( \frac{h_{i+1} - h_i}{h_*} \right)^2 + \left( \frac{h_i - h_{i-1}}{h_*} \right)^2 \right) \\ \leq \frac{1}{2\sqrt{\vartheta'_0}} \frac{1}{2h_*^{5/2}} |x_{i+2} - 2x_{i+1} + 2x_{i-1} - x_{i-2}| + \frac{h_*^2 \|\vartheta''\|_\infty^2}{4(h_*)^{3/2}(\vartheta'_0)^{3/2}}.$$

There exists  $c_{h_*}^j$  for  $j \in \llbracket 1, 4 \rrbracket$  such that

$$|x_{i+2} - 2x_{i+1} + 2x_{i-1} - x_{i-2}| = \left| x_i + 2h_*\vartheta'(ih_*) + \frac{(2h_*)^2}{2}\vartheta''(ih_*) + \frac{(2h_*)^3}{6}\vartheta^{(3)}(c_{h_*}^1) \right. \\ \left. - 2x_i - 2h_*\vartheta'(ih_*) - 2\frac{(h_*)^2}{2}\vartheta''(ih_*) - 2\frac{(h_*)^3}{6}\vartheta^{(3)}(c_{h_*}^2) \right. \\ \left. + 2x_i - 2h_*\vartheta'(ih_*) + 2\frac{(h_*)^2}{2}\vartheta''(ih_*) - 2\frac{(h_*)^3}{6}\vartheta^{(3)}(c_{h_*}^3) \right. \\ \left. - x_i + 2h_*\vartheta'(ih_*) - \frac{(2h_*)^2}{2}\vartheta''(ih_*) + \frac{(2h_*)^3}{6}\vartheta^{(3)}(c_{h_*}^4) \right| \\ \leq 2\|\vartheta^{(3)}\|_\infty h_*^3.$$

Thus,

$$\frac{1}{h_*} \left| \frac{\sqrt{h_{i+1}} - \sqrt{h_i}}{h_*} - \frac{\sqrt{h_i} - \sqrt{h_{i-1}}}{h_*} \right| \leq \frac{\sqrt{h_*}\|\vartheta^{(3)}\|_\infty}{2\sqrt{\vartheta'_0}} + \frac{\sqrt{h_*}\|\vartheta''\|_\infty^2}{4(\vartheta'_0)^{3/2}}. \quad (2.93)$$

Finally, using (2.92) and (2.93) in (2.91), we get

$$|q_i| \leq \frac{\Gamma}{\sqrt{\vartheta'_0}} \left( \frac{\|\vartheta^{(3)}\|_\infty}{2\sqrt{\vartheta'_0}} + \frac{\|\vartheta''\|_\infty^2}{4(\vartheta'_0)^{3/2}} \right) + \frac{C\|\vartheta''\|_\infty}{2\vartheta'_0} + \|q\|_\infty,$$

and this concludes the proof.  $\blacksquare$

We can now apply Theorem 2.3.2 to get estimates on the gap of eigenvalues of spectral problem (2.88) which has the same eigenvalues of the initial problem (2.77). Theorem 2.3.2 gives also lower bounds on the discrete derivative and the discrete  $L^2$  norm on  $\omega_h$  of eigenvectors of (2.88). Thanks to (2.89) and (2.90), we recover the same estimates on eigenvectors of the initial problem (2.77). In conclusion, estimates of Theorem 2.3.2 also hold on regular meshes defined above.

## 2.7.2 Error estimate for the finite difference discretization of a Sturm-Liouville eigenvalue problem on quasi-uniform 1D meshes

This appendix summarizes the paper in preparation [2].

In this section we prove Proposition 2.3.1 which is recalled here.

**Proposition** *Assume that  $q$  and  $\gamma$  satisfy  $(H_2)$ . For any  $\beta > 0$ , there exists  $h_0 > 0$  and  $C(\gamma, q, \beta) > 0$ , such that for any mesh whose mesh size satisfies  $h < h_0$  and  $\Theta_h \leq \beta$ , we have*

$$|\lambda_k^h - \lambda_k| \leq Ch^2\lambda_k^3, \quad \forall k \in \llbracket 1, N \rrbracket. \quad (2.94)$$



Before to prove Proposition 2.3.1, we give the results of some numerical experiments to illustrate estimation (2.94). We consider the following non uniform mesh on  $[0, 1]$ ,

$$\left( h, \frac{h}{2}, h, \frac{h}{2}, \dots \right), \text{ where } h = \frac{1}{N+1},$$

$N$  being the number of mesh points in  $(0, 1)$ . One can easily check that  $\Theta_h = 2$ . We plot below the function

$$e : h \mapsto \max_{1 \leq k \leq k_0} \frac{|\lambda_k^h - \lambda_k|}{\lambda_k^3},$$

in log-log axis, for  $k_0 = 50$ . According to (2.94), function  $e$  is bounded above by a linear function whose slope is 2.

We start with operator  $\mathcal{A} = -\Delta$ , so  $\gamma = 1$  and  $q = 0$ . In this special case, we know the expression of the eigenvalues of  $\mathcal{A}$ :

$$\forall k \in \mathbb{N}^*, \lambda_k = k^2 \pi^2.$$

Parameter  $N$  varies between 100 and 1000 with a constant step of 50. In figure 2.4b, discrete eigenvalues  $\lambda_k^h$  are computed with the function `eighs` of Python 3. The slope of the mean square linear interpolation given by `polyfit` function is 2.0066.

We now change the operator  $\mathcal{A}$ . We choose  $q(x) = x^2$  and  $\gamma(x) = \cos(x) + 2$ . In this case, we cannot compute explicitly the spectrum of  $\mathcal{A}$ . Instead, we compute numerically  $\lambda_k^h$  with  $N = 5000$  to get a sufficiently good approximation of  $\lambda_k$ . Indeed, in the following table we compare the maximum over  $k \in \{1, \dots, k_0\}$  of the relative difference between  $\lambda_k^{h_1}$  and  $\lambda_k^{h_2}$  as  $h_1$  and  $h_2$  decrease.

$N_1$	$N_2$	$h_1$	$h_2$	$\max \frac{ \lambda_k^{h_1} - \lambda_k^{h_2} }{\lambda_k^{h_1}}$
1000	2000	$9.99 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$5.17 \cdot 10^{-4}$
2000	3000	$5 \cdot 10^{-4}$	$3.33 \cdot 10^{-4}$	$9.56 \cdot 10^{-5}$
3000	4000	$3.33 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	$3.35 \cdot 10^{-5}$
4000	5000	$2.5 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$1.55 \cdot 10^{-5}$

Table 2.6 – Values of  $\max_{k \in \{1, \dots, k_0\}} \frac{|\lambda_k^{h_1} - \lambda_k^{h_2}|}{\lambda_k^{h_1}}$ .

In figure 2.4a, we now plot function  $e$  replacing  $\lambda_k$  by  $\lambda_k^h$ , for  $N = 5000$ . Parameter  $N$  still varies from 100 to 1000 with a step of 50 and the slope of the mean square linear interpolation is now 2.0236.

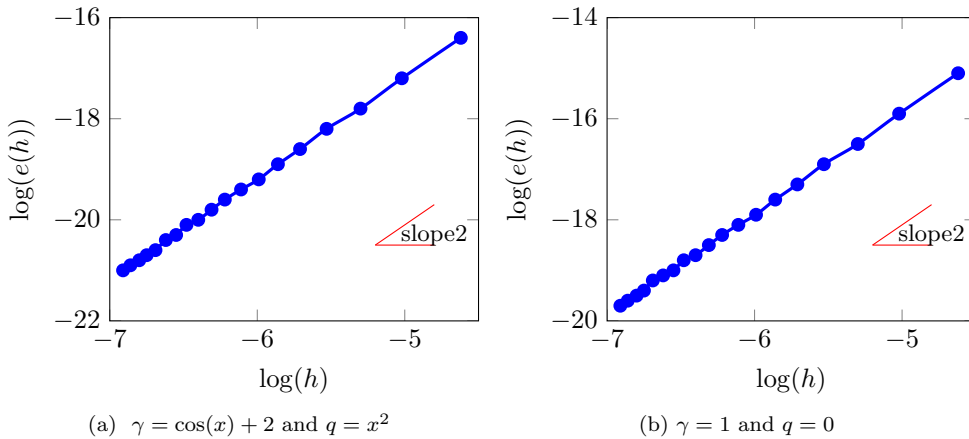


Figure 2.4 – Numerical illustration of Proposition (2.3.1)

**Outline of the proof of Proposition (2.3.1)**

The proof is split in two steps. First we show that,

$$\exists C_0 > 0, h^2 \lambda_k^h \leq C_0 \Rightarrow \lambda_k - \lambda_k^h \leq Ch^2 (\lambda_k^h)^2, \quad (2.95)$$

then, we prove that

$$\exists C_0 > 0, h \lambda_k \leq C_0 \Rightarrow \lambda_k^h - \lambda_k \leq Ch^2 \lambda_k^3, \quad (2.96)$$

where  $C_0$  depends only on  $q$  and  $\gamma$ . Inequalities (2.95) and (2.96) are valid on any non-uniform mesh.

We then conclude the proof in the last subsection by using the quasi-uniformity assumption on the meshes.

The main tool used in this proof is the Min Max principle. For convenience and without loss of generality, we can suppose that  $q \geq 1$ . Indeed, adding a constant  $q_0$  to  $q$  translates the continuous and the discrete spectrum of  $q_0$ , and this constant is irrelevant when comparing  $\lambda_k$  and  $\lambda_k^h$ .

**Proof of inequality (2.95)**

Let us define  $\langle \cdot, \cdot \rangle_0$  the inner product of  $L^2$  on  $\Omega$ . We recall that  $\langle \cdot, \cdot \rangle_{L^2(\Omega_h)}$  is the discrete  $L^2$  norm:

$$\langle u^h, v^h \rangle_{L^2(\Omega_h)} = \sum_{i=1}^N h_i u_i v_i.$$

Let  $k \in \llbracket 1, N \rrbracket$  and  $E_h^k$  the subspace of  $\mathbb{R}^N$ ,

$$E_h^k := \text{span}(\phi_1^h, \dots, \phi_k^h).$$

Let  $u^h \in E_h^k$  and define  $I^h : \mathbb{R}^N \rightarrow H_0^1(\Omega)$  the operator that builds from  $u^h$  the corresponding piecewise linear interpolation on  $\Omega$  :

$$\forall i \in \llbracket 1, N + 1 \rrbracket, \forall x \in [x_{i-1}, x_i], I^h u^h(x) = u_i^h \frac{x - x_{i-1}}{x_i - x_{i-1}} + u_{i-1}^h \frac{x_i - x}{x_i - x_{i-1}},$$

see figure 2.5 for an example.

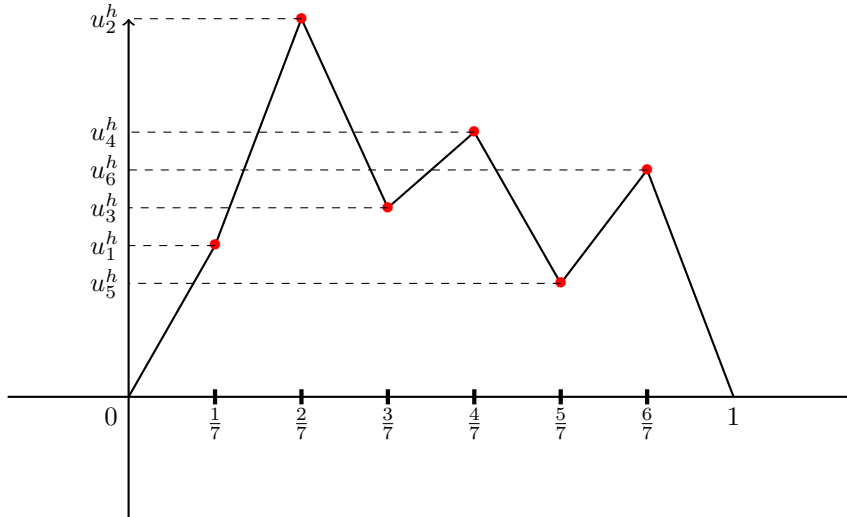


Figure 2.5 – Piecewise linear curve representing  $I^h(u^h)$  with  $N = 7$  and  $u^h = (2, 5, 2.5, 3.5, 1.5, 3)$ .

Let

$$\mathcal{R}(I^h u^h) := \frac{\langle \gamma \partial_x(I^h u^h), \partial_x(I^h u^h) \rangle_0 + \langle q I^h u^h, I^h u^h \rangle_0}{\langle I^h u^h, I^h u^h \rangle_0}$$

be the Rayleigh quotient of  $I^h u^h$ . We want to find an upper bound on  $\mathcal{R}(I^h u^h)$ .

— Let us consider first the denominator. We give a lower bound of the  $L^2$  norm of  $I^h u^h$ . We compute explicitly  $\int_{\Omega} (I^h u^h)^2$ :

$$\begin{aligned}
 \int_{x_{i-1}}^{x_i} (I^h u^h)^2 dx &= \int_{x_{i-1}}^{x_i} \left( u_{i-1}^h + \frac{x - x_{i-1}}{x_i - x_{i-1}} (u_i^h - u_{i-1}^h) \right)^2 dx \\
 &= \int_{x_{i-1}}^{x_i} (u_{i-1}^h)^2 + (x - x_{i-1})^2 \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2 + 2u_{i-1}^h (x - x_{i-1}) \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} dx \\
 &= h_{i-1/2} (u_{i-1}^h)^2 + \frac{h_{i-1/2}^3}{3} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2 + u_{i-1}^h h_{i-1/2}^2 \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \\
 &= h_{i-1/2} (u_{i-1}^h)^2 + \frac{h_{i-1/2}^3}{3} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2 + \frac{h_{i-1/2}}{2} (2u_{i-1}^h \times (u_i^h - u_{i-1}^h)) \\
 &= h_{i-1/2} (u_{i-1}^h)^2 + \frac{h_{i-1/2}^3}{3} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2 \\
 &\quad + \frac{h_{i-1/2}}{2} ((u_i^h)^2 - (u_{i-1}^h)^2 - (u_i^h - u_{i-1}^h)^2),
 \end{aligned}$$

we get

$$\int_{x_{i-1}}^{x_i} (I^h u^h)^2 dx = \frac{h_{i-1/2}}{2} (u_{i-1}^h)^2 + \frac{h_{i-1/2}}{2} (u_i^h)^2 - \frac{h_{i-1/2}^3}{6} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2.$$

Since,

$$\sum_{i=1}^{N+1} \frac{h_{i-1/2}}{2} (u_{i-1}^h)^2 + \frac{h_{i-1/2}}{2} (u_i^h)^2 = \sum_{i=1}^{N+1} \frac{h_{i+1/2} + h_{i-1/2}}{2} (u_i^h)^2 = \sum_{i=1}^N h_i (u_i^h)^2 = \langle u^h, u^h \rangle_{L^2(\Omega_h)},$$

we get by summing, for  $i \in \llbracket 1, N+1 \rrbracket$ ,

$$\int_{\Omega} (I^h u^h)^2 dx = \langle u^h, u^h \rangle_{L^2(\Omega_h)} - \sum_{i=1}^{N+1} \frac{h_{i-1/2}^3}{6} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2. \quad (2.97)$$

Remark that

$$\int_{\Omega} (I^h u^h)^2 dx \leq \langle u^h, u^h \rangle_{L^2(\Omega_h)}. \quad (2.98)$$

The last term of (2.97) looks like a discrete  $H^1$  norm weighted by a  $h^2$  factor. More precisely, there exists  $C_{\gamma} > 0$  such that

$$\sum_{i=1}^{N+1} \frac{h_{i-1/2}^3}{6} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2 \leq C_{\gamma} h^2 \sum_{i=1}^{N+1} \gamma_{i-1/2} h_{i-1/2} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2$$

and since  $q$  is positive,

$$\sum_{i=1}^{N+1} \frac{h_{i-1/2}^3}{6} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2 \leq C_{\gamma} h^2 \langle \mathcal{A}^h u^h, u^h \rangle_{L^2(\Omega_h)}.$$

Recall that  $u^h \in E_k^h$  so  $\langle \mathcal{A}^h u^h, u^h \rangle_{L^2(\Omega_h)} \leq \lambda_k^h \langle u^h, u^h \rangle_{L^2(\Omega_h)}$  hence,

$$\sum_{i=1}^{N+1} \frac{h_{i-1/2}^3}{6} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2 \leq C_{\gamma} h^2 \lambda_k^h \langle u^h, u^h \rangle_{L^2(\Omega_h)}.$$

Using this last expression in (2.97) we get,

$$\int_{\Omega} (I^h u^h)^2 dx \geq (1 - C_{\gamma} h^2 \lambda_k^h) \langle u^h, u^h \rangle_{L^2(\Omega_h)}. \quad (2.99)$$

— We now turn to the term  $\langle \gamma(\partial_x I^h u^h), \partial_x I^h u^h \rangle_0$ .

$$\begin{aligned} \int_{x_{i-1}}^{x_i} \gamma(\partial_x(I^h u^h))^2 dx &= h_{i-1/2} \gamma_{i-1/2} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2 \\ &\quad + \int_{x_{i-1}}^{x_i} (\gamma(x) - \gamma(x_{i-1/2})) \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2. \end{aligned}$$

By Taylor's formula, for any  $x \in (x_{i-1}, x_i)$ , there exists  $c \in (x_{i-1}, x_i)$ , such that

$$\gamma(x) = \gamma(x_{i-1/2}) + (x - x_{i-1/2})\gamma'(x_{i-1/2}) + \frac{(x - x_{i-1/2})^2}{2}\gamma''(c),$$

and since  $x_{i-1/2} = \frac{x_i + x_{i-1}}{2}$ , we have,  $\int_{x_{i-1}}^{x_i} (x - x_{i-1/2}) dx = 0$ , therefore,

$$\int_{x_{i-1}}^{x_i} \gamma(\partial_x(I^h u^h))^2 dx \leq (1 + h^2 C_\gamma) h_{i-1/2} \gamma_{i-1/2} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2,$$

and summing we find,

$$\int_{\Omega} \gamma(\partial_x(I^h u^h))^2 dx \leq (1 + h^2 C_\gamma) \sum_{i=1}^{N+1} h_{i-1/2} \gamma_{i-1/2} \left( \frac{u_i^h - u_{i-1}^h}{h_{i-1/2}} \right)^2. \quad (2.100)$$

— It remains to deal with the term  $\langle q I^h u^h, I^h u^h \rangle_0$ . We want an upper bound which depends on  $\langle q u^h, u^h \rangle_{L^2(\Omega_h)}$ . For any  $\varepsilon > 0$ ,

$$\begin{aligned} \int_{\Omega} q(I^h u^h)^2 &= \int_{\Omega} (I^h(\sqrt{q}u^h) + \sqrt{q}(I^h u^h) - I^h(\sqrt{q}u^h))^2 \\ &\leq (1 + \varepsilon) \int_{\Omega} (I^h(\sqrt{q}u^h))^2 + \left(1 + \frac{1}{\varepsilon}\right) \int_{\Omega} (\sqrt{q}(I^h u^h) - I^h(\sqrt{q}u^h))^2. \end{aligned}$$

According to (2.98) applied to  $\sqrt{q}u^h$ ,

$$\int_{\Omega} q(I^h u^h)^2 \leq (1 + \varepsilon) \langle \sqrt{q}u^h, \sqrt{q}u^h \rangle_{L^2(\Omega_h)} + \left(1 + \frac{1}{\varepsilon}\right) \int_{\Omega} (\sqrt{q}(I^h u^h) - I^h(\sqrt{q}u^h))^2. \quad (2.101)$$

Let us estimate the last term. Let  $x \in (x_{i-1}, x_i)$ ,

$$\begin{aligned} &\sqrt{q(x)}(I^h u^h)(x) - (I^h(\sqrt{q}u^h))(x) \\ &= \sqrt{q(x)} \left( \frac{x - x_{i-1}}{x_i - x_{i-1}} u_i^h + \frac{x_i - x}{x_i - x_{i-1}} u_{i-1}^h \right) \\ &\quad - \left( \frac{x - x_{i-1}}{x_i - x_{i-1}} \sqrt{q(x_i)} u_i^h + \frac{x_i - x}{x_i - x_{i-1}} \sqrt{q(x_{i-1})} u_{i-1}^h \right) \\ &= \frac{x - x_{i-1}}{x_i - x_{i-1}} u_i^h \left( \sqrt{q(x)} - \sqrt{q(x_i)} \right) + \frac{x_i - x}{x_i - x_{i-1}} u_{i-1}^h \left( \sqrt{q(x)} - \sqrt{q(x_{i-1})} \right). \end{aligned}$$

By Taylor's formula, there exist  $\xi^+, \xi^- \in (x_{i-1}, x_i)$  such that

$$\begin{aligned} \sqrt{q(x)} - \sqrt{q(x_i)} &= \sqrt{q(x)} - \sqrt{q(x_{i-1/2})} - \left( \sqrt{q(x_i)} - \sqrt{q(x_{i-1/2})} \right) \\ &= (x - x_{i-1/2})(\sqrt{q})'(x_{i-1/2}) + \frac{(x - x_{i-1/2})^2}{2}(\sqrt{q})''(\xi^+) \\ &\quad - \left( (x_i - x_{i-1/2})(\sqrt{q})'(x_{i-1/2}) + \frac{(x_i - x_{i-1/2})^2}{2}(\sqrt{q})''(\xi^-) \right) \\ &= (x - x_i)(\sqrt{q})'(x_{i-1/2}) + O(h^2). \end{aligned}$$

Similarly,

$$\sqrt{q(x)} - \sqrt{q(x_{i-1})} = (x - x_{i-1})(\sqrt{q})'(x_{i-1/2}) + O(h^2),$$

thus,

$$\begin{aligned} \sqrt{q(x)}(I^h u^h)(x) - (I^h(\sqrt{q}u^h))(x) &= (x - x_{i-1})(x - x_i) \left( \frac{u_i^h - u_{i-1}^h}{x_i - x_{i-1}} \right) (\sqrt{q})'(x_{i-1/2}) \\ &\quad + (u_i^h + u_{i-1}^h)O(h^2). \end{aligned}$$

Take the square and integrate on  $(x_{i-1}, x_i)$  to get

$$\begin{aligned} \int_{x_{i-1}}^{x_i} (\sqrt{q}I^h u^h - I^h(\sqrt{q}u^h))^2 &\leq 2 \int_{x_{i-1}}^{x_i} \left( (x - x_{i-1})(x - x_i) \left( \frac{u_i^h - u_{i-1}^h}{x_i - x_{i-1}} \right) (\sqrt{q})'(x_{i-1/2}) \right)^2 \\ &\quad + h_{i-1/2} ((u_i^h)^2 + (u_{i-1}^h)^2) O(h^4) \\ &\leq C_q h^4 h_{i-1/2} \left( \left( \frac{u_i^h - u_{i-1}^h}{x_i - x_{i-1}} \right)^2 + (u_i^h)^2 + (u_{i-1}^h)^2 \right). \end{aligned}$$

Summing, and still using that  $q$  is positive,

$$\begin{aligned} \int_{\Omega} (\sqrt{q}I^h u^h - I^h(\sqrt{q}u^h))^2 &\leq C_q h^4 \left( \sum_{i=1}^N h_{i-1/2} \left( \frac{u_i^h - u_{i-1}^h}{x_i - x_{i-1}} \right)^2 + \langle u^h, u^h \rangle_{L^2(\Omega_h)} \right) \\ &\leq C_{q,\gamma} h^4 (\langle \mathcal{A}^h u^h, u^h \rangle_{L^2(\Omega_h)} + \langle u^h, u^h \rangle_{L^2(\Omega_h)}) \\ &\leq C_{q,\gamma} h^4 (\lambda_k^h + 1) \langle u^h, u^h \rangle_{L^2(\Omega_h)}. \end{aligned}$$

We take back (2.101) with  $\varepsilon = h^2$ ,

$$\int_{\Omega} q(I^h u^h)^2 \leq (1 + h^2) \langle q u^h, u^h \rangle_{L^2(\Omega_h)} + \left( 1 + \frac{1}{h^2} \right) C_{q,\gamma} h^4 (\lambda_k^h + 1) \langle u^h, u^h \rangle_{L^2(\Omega_h)}.$$

Hypothesis  $q \geq 1$  yields that  $\lambda_k^h \geq 1$ . Therefore, for  $h$  small enough independent of  $k$ ,

$$\int_{\Omega} q(I^h u^h)^2 \leq (1 + h^2) \langle q u^h, u^h \rangle_{L^2(\Omega_h)} + C_{q,\gamma} h^2 \lambda_k^h \langle u^h, u^h \rangle_{L^2(\Omega_h)}. \quad (2.102)$$

Now we gather (2.99), (2.100) and (2.102) together and we get

$$\mathcal{R}(I^h u^h) \leq \frac{(1 + C_{\gamma} h^2) \langle \mathcal{A}^h u^h, u^h \rangle_{L^2(\Omega_h)} + C_{q,\gamma} h^2 \lambda_k^h \langle u^h, u^h \rangle_{L^2(\Omega_h)}}{(1 - C_{\gamma} h^2 \lambda_k^h) \langle u^h, u^h \rangle_{L^2(\Omega_h)}}.$$

Note that the following inequality holds for any  $x \in (0, \frac{1}{2})$ :

$$\frac{1}{1-x} \leq 1 + 2x.$$

Thus assuming that  $C_0$  is small enough so that  $C_{\gamma} h^2 \lambda_k^h \leq \frac{1}{2}$ , we find that

$$\mathcal{R}(I^h u^h) \leq \lambda_k^h (1 + C_{q,\gamma} h^2) (1 + C_{\gamma} h^2 \lambda_k^h).$$

For  $h$  small enough independent of  $k$ , we end up with

$$\mathcal{R}(I^h u^h) \leq \lambda_k^h (1 + C_{q,\gamma} h^2 \lambda_k^h). \quad (2.103)$$

Now recall that the Min Max principle gives

$$\lambda_k = \underset{V \subset H_0^1(\Omega), \dim(V)=k}{\text{Min}} \left( \underset{u \in V}{\text{Max}} \mathcal{R}(u) \right). \quad (2.104)$$

Since by hypothesis  $1 - C_{\gamma} h^2 \lambda_k^h > 0$ , according to (2.99), operator  $I^h$  is one-to-one. Therefore,  $I^h(E_k^h)$  is of dimension  $k$ . Moreover,  $I^h(E_k^h) \subset H_0^1(\Omega)$ , thus, according to (2.103) and (2.104),

$$\lambda_k \leq \underset{u \in I^h(E_k^h)}{\text{Max}} \mathcal{R}(u) \leq \lambda_k^h (1 + C_{q,\gamma} h^2 \lambda_k^h),$$

hence,

$$\lambda_k - \lambda_k^h \leq C_{q,\gamma} h^2 (\lambda_k^h)^2. \quad (2.105)$$

**Proof of inequality (2.96)**

Let  $E_k$  be the following subspace of dimension  $k$  of  $H_0^1(\Omega)$ :

$$E_k := \text{span}(\phi_1, \phi_2, \dots, \phi_k).$$

Denote by  $R^h : H_0^1(\Omega) \rightarrow \mathbb{R}^N$  the left inverse of  $I^h$  defined by

$$R^h(u) := (u(x_1), u(x_2), \dots, u(x_N)).$$

Let also  $P^h : H_0^1(\Omega) \cap C^2(\bar{\Omega}) \rightarrow \mathbb{R}^N$  be the operator defined by

$$\mathcal{A}^h(P^h u) = R^h(\mathcal{A}u)$$

To prove (2.96), we establish an upper bound for the discrete Rayleigh quotient

$$\mathcal{R}^h(P^h u) := \frac{\langle \mathcal{A}^h P^h u, P^h u \rangle_{L^2(\Omega_h)}}{\langle P^h u, P^h u \rangle_{L^2(\Omega_h)}}. \quad (2.106)$$

To begin with, we need some preliminary estimates.

Let  $u \in E_k$ . Since  $\gamma$  and  $q$  satisfy  $(H_2)$ ,  $u \in H^4$  and the following estimates hold

$$\begin{cases} \|u\|_{H^1} \leq C\sqrt{\lambda_k}\|u\|_{L^2} \\ \|u\|_{H^2} \leq C\|\mathcal{A}u\|_{L^2} \leq C\lambda_k\|u\|_{L^2} \\ \|u\|_{H^4} \leq C\|\mathcal{A}^2 u\|_{L^2} \leq C\lambda_k^2\|u\|_{L^2}. \end{cases} \quad (2.107)$$

Results of [7] can be used to prove that the following inequality holds, without making any restrictions on the non-uniformity of the mesh,

$$\|I^h(R^h u - P^h u)\|_{H^1} \leq Ch^2\|u\|_{H^4}, \quad \forall u \in H_0^1 \cap H^4. \quad (2.108)$$

**Remarque 2.7.1** We explain here how (2.108) can be derived. Denote by  $\overline{\mathcal{A}u}^h$  the vector whose component  $j$  is  $(\overline{\mathcal{A}u}^h)_j = \frac{1}{h_j} \int_{x_{j-1/2}}^{x_{j+1/2}} \mathcal{A}u(x) dx$ . Then, Theorem 3.1 of [7], with  $s = 2$ , gives

$$\|I^h(R^h u - s^h)\|_{H^1} \leq Ch^2\|u\|_{H^3}, \quad \forall u \in H_0^1 \cap H^4, \quad (2.109)$$

where  $s^h$  solves the equation  $\mathcal{A}^h s^h = \overline{\mathcal{A}u}^h$ . The proof of (2.109) consists in defining  $a^h(u^h, v^h) := \sum_{j=0}^{N-1} h_{j+1/2} \gamma_{j+1/2} (I^h u^h)'_{j+1/2} (I^h v^h)'_{j+1/2} + \langle qv^h, u^h \rangle_{L^2(\Omega_h)}$ , and, for any vector  $v^h$  of size  $N$ , by bounding  $a^h(R^h u - s^h, v^h)$  which is equal to  $a^h(R^h u, v^h) - \langle \overline{\mathcal{A}u}^h, v^h \rangle_{L^2(\Omega_h)}$ . The authors get

$$\left| a^h(R^h u, v^h) - \langle \overline{\mathcal{A}u}^h, v^h \rangle_{L^2(\Omega_h)} \right| \leq Ch^2\|u\|_{H^3} \|I^h v^h\|_{H^1}, \quad (2.110)$$

which gives (2.109) by applying Proposition 3.1 of [7].

Therefore, to get (2.108) it suffices to get an upper bound of  $a^h(R^h u - P^h u, v^h) = a^h(R^h u, v^h) - \langle R^h(\mathcal{A}u), v^h \rangle_{L^2(\Omega_h)}$ , for any vector  $v^h$ . According to (2.110), Remark 3.4 of [7] and (2.107), the following inequalities hold

$$\begin{aligned} \left| a^h(R^h u, v^h) - \langle R^h(\mathcal{A}u), v^h \rangle_{L^2(\Omega_h)} \right| &\leq \left| a^h(R^h u, v^h) - \langle \overline{\mathcal{A}u}^h, v^h \rangle_{L^2(\Omega_h)} \right| + \left| \langle \overline{\mathcal{A}u}^h - R^h(\mathcal{A}u), v^h \rangle_{L^2(\Omega_h)} \right| \\ &\leq C(h^2\|u\|_{H^3} + h^2\|\mathcal{A}u\|_{H^2}) \|I^h v^h\|_{H^1} \\ &\leq Ch^2 \left( \frac{\|u\|_{H^4}}{\sqrt{\lambda_k}} + \|u\|_{H^4} \right) \|I^h v^h\|_{H^1} \\ &\leq Ch^2\|u\|_{H^4} \|I^h v^h\|_{H^1}, \end{aligned}$$

and we get (2.108) by applying again Proposition 3.1 of [7].

Since  $H^1 \hookrightarrow L^\infty$ , we have

$$\max_{i \in \llbracket 1, N \rrbracket} |u(x_i) - (P^h u)_i| \leq Ch^2 \|u\|_{H^4}, \forall u \in H_0^1 \cap H^4.$$

Thus,

$$\langle R^h u - P^h u, R^h u - P^h u \rangle_{L^2(\Omega_h)}^{1/2} \leq Ch^2 \|u\|_{H^4}, \forall u \in H_0^1 \cap H^4. \quad (2.111)$$

Recall the trapezoid formula on  $[x_i, x_{i+1}]$  applied to  $fg$ ,

$$\int_{x_i}^{x_{i+1}} (fg)(x) dx = h_{i+1/2} \frac{(fg)(x_{i+1}) + (fg)(x_i)}{2} + \frac{1}{2} \int_{x_i}^{x_{i+1}} (x - x_{i+1})(x - x_i) (fg)''(x) dx,$$

thus, summing for  $i \in \{0, 1, \dots, N\}$  we get,

$$\left| \int_{\Omega} (fg)(x) dx - \sum_{i=0}^N h_{i+1/2} \frac{(fg)(x_{i+1}) + (fg)(x_i)}{2} \right| \leq Ch^2 \int_{\Omega} |(fg)''(x)| dx, \quad (2.112)$$

Notice that, since  $h_i = \frac{h_{i+1/2} + h_{i-1/2}}{2}$ ,

$$\begin{aligned} \langle R^h f, R^h g \rangle_{L^2(\Omega_h)} &= \sum_{i=1}^N h_i (fg)(x_i) \\ &= \frac{1}{2} \left[ \sum_{i=1}^N h_{i+1/2} (fg)(x_i) + \sum_{i=1}^N h_{i-1/2} (fg)(x_i) \right], \end{aligned}$$

thus, if  $f$  and  $g$  both vanish at  $x_0 = 0$  and  $x_{N+1} = 1$ ,

$$\begin{aligned} \langle R^h f, R^h g \rangle_{L^2(\Omega_h)} &= \frac{1}{2} \left[ \sum_{i=0}^N h_{i+1/2} (fg)(x_i) + \sum_{i=0}^N h_{i+1/2} (fg)(x_{i+1}) \right], \\ &= \sum_{i=0}^N h_{i+1/2} \frac{(fg)(x_i) + (fg)(x_{i+1})}{2}. \end{aligned}$$

so using (2.112),

$$\left| \int_{\Omega} fg - \langle R^h f, R^h g \rangle_{L^2(\Omega_h)} \right| \leq Ch^2 \int_{\Omega} |(fg)''(x)| dx \leq Ch^2 \sum_{s+t=2} \|f\|_{H^s} \|g\|_{H^t}. \quad (2.113)$$

Thus, in particular if  $f = g$ , then using (2.107) and the fact that  $\lambda_k \geq 1$ ,

$$\|f\|_{L^2}^2 - Ch^2 \|f\|_{H^2} \|f\|_{L^2} \leq \langle R^h f, R^h f \rangle_{L^2(\Omega_h)} \leq \|f\|_{L^2}^2 + Ch^2 \|f\|_{H^2} \|f\|_{L^2}. \quad (2.114)$$

We can now give an upper bound of (2.106).

— Let us consider first the numerator

$$\begin{aligned} \langle \mathcal{A}^h P^h u, P^h u \rangle_{L^2(\Omega_h)} &= \langle R^h \mathcal{A} u, P^h u \rangle_{L^2(\Omega_h)} \\ &= \langle R^h \mathcal{A} u, R^h u \rangle_{L^2(\Omega_h)} + \langle R^h \mathcal{A} u, P^h u - R^h u \rangle_{L^2(\Omega_h)}. \end{aligned}$$

Since  $u$  and  $\mathcal{A}u$  vanish at  $x = 0$  and  $x = 1$ , the first term can be estimated with (2.113),

$$\langle R^h \mathcal{A} u, R^h u \rangle_{L^2(\Omega_h)} \leq \langle \mathcal{A} u, u \rangle_0 + Ch^2 (\|\mathcal{A} u\|_{L^2} \|u\|_{H^2} + \|\mathcal{A} u\|_{H^1} \|u\|_{H^1} + \|\mathcal{A} u\|_{H^2} \|u\|_{L^2}).$$

Using (2.107), we get

$$\langle R^h \mathcal{A} u, R^h u \rangle_{L^2(\Omega_h)} \leq \lambda_k \|u\|_{L^2}^2 + Ch^2 \lambda_k^2 \|u\|_{L^2}^2. \quad (2.115)$$

The second one is bounded thanks to Cauchy-Schwarz inequality

$$\langle R^h \mathcal{A} u, P^h u - R^h u \rangle_{L^2(\Omega_h)} \leq \langle R^h \mathcal{A} u, R^h \mathcal{A} u \rangle_{L^2(\Omega_h)}^{1/2} \langle P^h u - R^h u, P^h u - R^h u \rangle_{L^2(\Omega_h)}^{1/2},$$

and (2.114) with (2.111)

$$\langle R^h \mathcal{A}u, P^h u - R^h u \rangle_{L^2(\Omega_h)} \leq (\|\mathcal{A}u\|_{L^2}^2 + Ch^2 \|\mathcal{A}u\|_{H^2} \|\mathcal{A}u\|_{L^2})^{1/2} Ch^2 \|u\|_{H^4},$$

and finally (2.107)

$$\langle R^h \mathcal{A}u, P^h u - R^h u \rangle_{L^2(\Omega_h)} \leq (1 + Ch^2 \lambda_k)^{1/2} Ch^2 \lambda_k^3 \|u\|_{L^2}^2. \quad (2.116)$$

Hence, with (2.115) and (2.116),

$$\langle \mathcal{A}^h P^h u, P^h u \rangle_{L^2(\Omega_h)} \leq (\lambda_k + Ch^2 \lambda_k^3 + Ch^3 \lambda_k^{\frac{7}{2}}) \|u\|_{L^2}^2. \quad (2.117)$$

— Let us now deal with the denominator. First, decompose it the following way

$$\begin{aligned} \langle P^h u, P^h u \rangle_{L^2(\Omega_h)} &= \langle R^h u + (P^h u - R^h u), R^h u + (P^h u - R^h u) \rangle_{L^2(\Omega_h)} \\ &= \langle R^h u, R^h u \rangle_{L^2(\Omega_h)} + 2\langle R^h u, P^h u - R^h u \rangle_{L^2(\Omega_h)} + \langle P^h u - R^h u, P^h u - R^h u \rangle_{L^2(\Omega_h)}. \end{aligned}$$

The first term is bounded from below by using (2.114) and (2.107)

$$\langle R^h u, R^h u \rangle_{L^2(\Omega_h)} \geq \|u\|_{L^2}^2 - Ch^2 \|u\|_{H^2} \|u\|_{L^2} \geq \|u\|_{L^2}^2 (1 - Ch^2 \lambda_k),$$

we estimate the second term the same way as we got inequality (2.116)

$$2\langle R^h u, P^h u - R^h u \rangle_{L^2(\Omega_h)} \geq - (1 + Ch^2 \lambda_k)^{1/2} Ch^2 \lambda_k^2 \|u\|_{L^2}^2 \geq - (1 + Ch \lambda_k^{1/2}) Ch^2 \lambda_k^2 \|u\|_{L^2}^2,$$

and the lower bound of the third one is directly obtained by applying (2.111)

$$\langle P^h u - R^h u, P^h u - R^h u \rangle_{L^2(\Omega_h)} \geq -Ch^4 \lambda_k^4 \|u\|_{L^2}^2.$$

Hence, there exists  $C$  which depends only on  $\gamma$  and  $q$  such that

$$\langle P^h u, P^h u \rangle_{L^2(\Omega_h)} \geq \|u\|_{L^2}^2 \left(1 - Ch^2 \lambda_k^2 - Ch^3 \lambda_k^{5/2} - Ch^4 \lambda_k^4\right). \quad (2.118)$$

Now suppose that  $h\lambda_k \leq C_0$ , for  $C_0 > 0$  small enough depending only on  $\gamma$  and  $q$ , such that

$$\langle P^h u, P^h u \rangle_{L^2(\Omega_h)} \geq \frac{1}{2} \|u\|_{L^2}^2. \quad (2.119)$$

We use inequalities (2.117) and (2.118) to estimate (2.114)

$$\mathcal{R}^h(P^h u) \leq \lambda_k \frac{1 + Ch^2 \lambda_k^2 + Ch^3 \lambda_k^3}{1 - Ch^2 \lambda_k^2 - Ch^3 \lambda_k^3 - Ch^4 \lambda_k^4}.$$

For  $C_0 > 0$  small enough,  $Ch^2 \lambda_k^2 + Ch^3 \lambda_k^3 + Ch^4 \lambda_k^4 \leq 1/2$  so we can use inequality  $\frac{1}{1-x} \leq 1+2x$  to get

$$\mathcal{R}^h(P^h u) \leq \lambda_k (1 + Ch^2 \lambda_k^2).$$

According to (2.119),  $P^h$  is one-to-one, so  $P^h E_k$  is a subspace of dimension  $k$  of  $\mathbb{R}^N$ . Thus, the Min Max principle (2.104) gives

$$\begin{aligned} \lambda_k^h &\leq \lambda_k (1 + Ch^2 \lambda_k^2) \\ \lambda_k^h - \lambda_k &\leq Ch^2 \lambda_k^3. \end{aligned}$$

## Conclusion

Suppose now that the mesh is quasi-uniform, that is  $\Theta_h \leq \beta$  for some given  $\beta$ . According to Lemma 2.3.3, there exists  $C > 0$  depending only on  $\gamma, q$  and  $\beta$  such that

$$\lambda_k^h \leq Ck^2,$$

and Remark 2.1.1 gives

$$\gamma_{\min} \pi^2 k^2 - \|q\|_{\infty} \leq \lambda_k,$$



thus we have

$$\lambda_k^h \leq C\lambda_k.$$

It follows that the condition  $h\lambda_k \leq C_0$  is more restrictive than hypothesis  $h^2\lambda_k^h \leq C_0$ . Therefore, there exists  $C_0 > 0$  (depending only on  $\gamma, q$  and  $\beta$ ) such that if  $h\lambda_k \leq C_0$  then

$$|\lambda_k^h - \lambda_k| \leq Ch^2\lambda_k^3.$$

On the other hand, if  $h\lambda_k > C_0$ , then

$$|\lambda_k^h - \lambda_k| \leq \lambda_k^h + \lambda_k \leq C\lambda_k \leq \frac{C}{C_0^2}\lambda_k C_0^2 \leq \frac{C}{C_0^2}h^2\lambda_k^3.$$

This concludes the proof of Proposition 2.3.1.

### 2.7.3 Estimates on discrete eigenfunctions with a piecewise constant diffusion coefficients

In this section we precise the second remark of Section 2.6. Let  $\gamma_{\min}$  and  $\gamma_{\max}$  be two positive real numbers such that  $\gamma_{\min} < \gamma_{\max}$  and  $x_0 \in (0, 1)$ . Define  $\gamma : [0, 1] \rightarrow \{\gamma_{\min}, \gamma_{\max}\}$  the following piecewise constant function

$$\gamma(x) := \begin{cases} \gamma_{\min} & \text{if } x \in [0, x_0], \\ \gamma_{\max} & \text{if } x \in (x_0, 1]. \end{cases}$$

Consider a uniform discretization of  $[0, 1]$  of parameter  $h$ . Let  $i_0 = \lceil \frac{x_0}{h} + \frac{1}{2} \rceil$ , so that  $h(i_0 - 1/2)$  is less than  $x_0$ . Define  $L_{\gamma_{\min}} := x_0$  and  $L_{\gamma_{\max}} := 1 - L_{\gamma_{\min}}$ . The following Theorem shows that the behaviour

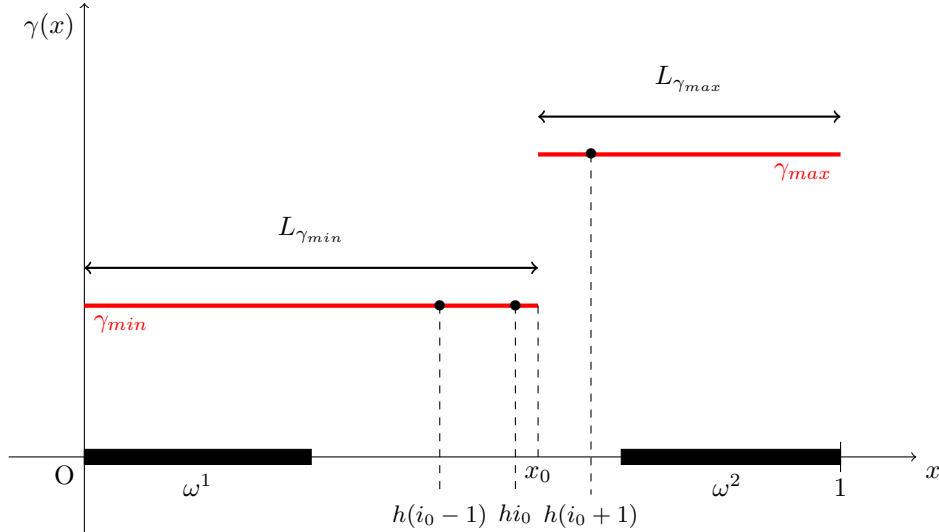


Figure 2.6 – Example of a piecewise constant diffusion function  $\gamma$  and some mesh points near  $x_0$ .

of the discrete eigenfunctions can be radically different in  $[0, x_0]$  and in  $[x_0, 1]$ .

**Theorem 2.7.1** *Let  $q = 0$  and assume that  $\gamma$  is a piecewise constant function as defined above. Let  $\omega^1$  be a non empty open subset of  $(0, x_0]$  and  $\omega^2$  be a non empty subinterval of  $[x_0, 1)$ . Let  $\lambda$  be an eigenvalue of  $\mathcal{A}^h$  satisfying*

$$\gamma_{\min} < \frac{\lambda h^2}{4} < \gamma_{\max}, \quad (2.120)$$

*and  $U = {}^t(u_1, u_2, \dots, u_N)$  the corresponding normalized discrete eigenfunction. There exists  $C > 0$  depending only on  $\gamma$ , such that*

$$|\partial_t U| \leq \exp(-C/h), \quad (2.121)$$

$$\|U\|_{L^2(\omega_h^1)} \leq \exp(-C/h), \quad (2.122)$$

whereas

$$|\partial_r U| \geq C \left(1 - \frac{\lambda h^2}{4\gamma_{\max}}\right), \quad (2.123)$$

$$\|U\|_{L^2(\omega_h^2)} \geq C \left(1 - \frac{\lambda h^2}{4\gamma_{\max}}\right). \quad (2.124)$$

**Remark 2.7.1** This result remains true if we interchange  $\gamma_{\min}$  and  $\gamma_{\max}$ ,  $l$  and  $r$  and  $\omega^1$  and  $\omega^2$ .

Estimations (2.121) and (2.122) explain the very small values obtained in columns  $I_l^h(N)$  and  $I_r^h(N)$  of table 2.5 of Section 2.4. We notice that  $I_r^h(N)$  seems bounded below by a quantity close to 3.79, so inequality (2.123) is probably not accurate. However, inequality (2.124), looks sharp. Indeed, notice that column  $I_2^h(N)$  decreases linearly in  $h$ :  $I_2^h(N) \approx 5.68 \times h$ , and so do  $C \left(1 - \frac{\lambda h^2}{4\gamma_{\max}}\right)$  when  $\lambda = \lambda_N^h$  since according to Remark 2.1.3,  $\left(1 - \frac{\lambda h^2}{4\gamma_{\max}}\right)$  is of order  $h$  for the highest frequencies.

### Proof

Recall that, since  $q = 0$ , the eigenvector  $U$  satisfies

$$-\left(\gamma_{i+1/2} \frac{u_{i+1} - u_i}{h^2} - \gamma_{i-1/2} \frac{u_i - u_{i-1}}{h^2}\right) = \lambda u_i. \quad (2.125)$$

### Estimate on the left side

For  $i \in \llbracket 1, i_0 - 1 \rrbracket$ , by choice of  $i_0$ ,  $h(i + 1/2) < x_0$ , thus  $\gamma_{i-1/2} = \gamma_{i+1/2} = \gamma_{\min}$  so equality (2.125) writes

$$u_{i+1} + \beta u_i + u_{i-1} = 0, \quad (2.126)$$

with

$$\beta = \frac{h^2 \lambda}{\gamma_{\min}} - 2. \quad (2.127)$$

Define  $U_i = \begin{pmatrix} u_{i-1} \\ u_i \end{pmatrix}$ , so that (2.126) writes,

$$\forall i \leq i_0 - 1, U_{i+1} = A_\beta U_i, \quad (2.128)$$

with

$$A_\beta = \begin{pmatrix} 0 & 1 \\ -1 & -\beta \end{pmatrix}.$$

We can now estimate  $u_i$  in function of  $u_1$  to get the upper bound (2.121).

Note that, according to (2.127) and with hypothesis (2.120),  $\beta > 2$  thus  $\beta^2 - 4 > 0$  which means that  $A_\beta$  has two distinct real eigenvalues :  $r_-$  and  $r_+$ . Since  $\beta > 0$ , equation (2.126) gives that

1.  $r_-$  and  $r_+$  are both negative,
2.  $r_- \times r_+ = 1$  thus  $r_- = \frac{1}{r_+}$ .
3. We set:  $|r_-| > 1$ , therefore  $|r_+| < 1$ .

We define two corresponding eigenvectors of  $A_\beta$ :

$$U_+ = \begin{pmatrix} 1 \\ r_+ \end{pmatrix} \text{ and } U_- = \begin{pmatrix} 1 \\ r_- \end{pmatrix}. \quad (2.129)$$

Let us decompose  $U_1$  in this basis:

$$U_1 = \alpha_+ U_+ + \alpha_- U_-.$$

Dirichlet boundary condition  $u_0 = 0$  imposes

$$\begin{cases} 0 = \frac{\alpha_+}{r_+} + \frac{\alpha_-}{r_-}, \\ u_1 = \alpha_+ + \alpha_-. \end{cases} \quad (2.130)$$

We inverse this system using that  $r_+ \times r_- = 1$  and we get

$$\begin{cases} \alpha_+ = \frac{u_1}{1 - r_-^2} \\ \alpha_- = \frac{u_1}{1 - r_+^2}. \end{cases} \quad (2.131)$$

Let us now decompose  $U_i$  in the same basis. According to (2.128),

$$\forall i \leq i_0, \quad U_i = \alpha_+ r_+^{i-1} U_+ + \alpha_- r_-^{i-1} U_-,$$

using (2.129) then first equation of (2.130), we get

$$\begin{aligned} u_i &= \alpha_+ r_+^{i-1} + \alpha_- r_-^{i-1}, \\ u_i &= \frac{\alpha_+}{r_+} (r_+^i - r_-^i), \end{aligned} \quad (2.132)$$

thus, according to (2.131)

$$\forall i \leq i_0, \quad u_i = \frac{u_1}{r_+ - r_-} (r_+^i - r_-^i), \quad (2.133)$$

and,

$$\forall i \leq i_0, \quad |u_i| = |u_1| |r_-|^{i-1} \left| \frac{1 - \left(\frac{r_+}{r_-}\right)^i}{1 - \frac{r_+}{r_-}} \right|.$$

Furthermore,  $\frac{r_+}{r_-} = r_+^2$ , so

$$\forall i \leq i_0, \quad |u_i| = |u_1| |r_-|^{i-1} \left( \frac{1 - (r_+^2)^i}{1 - r_+^2} \right). \quad (2.134)$$

Since  $|r_+| < 1$ ,

$$\begin{aligned} (r_+^2)^i &< r_+^2 \\ 1 - (r_+^2)^i &> 1 - r_+^2 > 0 \end{aligned}$$

and, with (2.134),

$$|u_i| > |u_1| |r_-|^{i-1}. \quad (2.135)$$

Notice that since  $h \sum_{i=1}^N |u_i|^2 = 1$ , we have  $|u_{i_0}| \leq \frac{1}{\sqrt{h}}$ . Letting  $i = i_0$  in (2.135), we end up with

$$\frac{1}{\sqrt{h}} > |u_{i_0}| > |u_1| \exp((i_0 - 1) \ln(r_-)),$$

which gives

$$r_-^2 \frac{\exp(-L_{\gamma_{\min}} \ln(r_-)/h)}{h^{3/2}} > \left| \frac{u_1}{h} \right| = |\partial_t u|, \quad (2.136)$$

and we obtain (2.121). To prove (2.122), remark that (2.134) gives

$$\begin{aligned} \sum_{ih \in \omega_1^h} h |u_i|^2 &< h \frac{|u_1|^2}{(1 - r_+^2)^2} \sum_{ih \in \omega_1^h} r_-^{2i-4} \\ &< h \frac{|u_1|^2}{(r_-^2 - 1)^2} \sum_{ih \in \omega_1^h} r_-^{2i} \\ &< h \frac{|u_1|^2 r_-^2}{(r_-^2 - 1)^2 (r_-^2 - 1)} \times r_-^{2i_1}, \end{aligned}$$

where  $i_1 = \max(\omega_1^h)/h$ . Using (2.136), we find

$$\begin{aligned} \sum_{ih \in \omega_1^h} h |u_i|^2 &< C_1 \exp\left(-\frac{2L_{\gamma_{\min}} \ln(r_-)}{h}\right) \times r_-^{2i_1} \\ &< C_1 \exp\left(-\frac{2L_{\gamma_{\min}} \ln(r_-)}{h}\right) \times \exp(2i_1 \ln(r_-)) \\ &< C_1 \exp\left(-\frac{2L_{\gamma_{\min}} \ln(r_-)}{h}\right) \times \exp\left(\frac{2 \max(\omega_1^h) \ln(r_-)}{h}\right) \end{aligned}$$

and since  $\max(\omega_1^h) < L_{\gamma_{\min}}$ , we get (2.122).

**Estimate on the right side**

First we show the following Lemmas.

**Lemma 2.7.1** *Let*

$$C_{h,\lambda,\gamma} := \sqrt{1 - \frac{\lambda h^2}{4\gamma_{\max}}}.$$

*There exists  $C > 0$  such that for any  $i \in \llbracket 1, N \rrbracket \setminus \{i_0\}$ ,*

$$|u_{i_0}| > CC_{h,\lambda,\gamma}|u_i|. \quad (2.137)$$

**Proof** Let  $i \in \llbracket 2, i_0 \rrbracket$ , according to (2.133),

$$\begin{aligned} \left| \frac{u_i}{u_{i-1}} \right| &= \left| \frac{r_+^i - r_-^i}{r_+^{i-1} - r_-^{i-1}} \right| \\ &= |r_-| \frac{1 - (r_+^2)^i}{1 - (r_+^2)^{i-1}} > 1, \end{aligned}$$

thus,

$$\forall i \in \llbracket 1, i_0 - 1 \rrbracket, |u_{i_0}| > |u_i|. \quad (2.138)$$

We now deal with indices  $i > i_0 + 1$  by propagating the discrete resolvent estimate from the right side of the domain to the left. To this end, let us recall the resolvent estimate (2.31) of Proposition 2.3.2, when  $\gamma$  is constant,

$$\|S_{i \leftarrow j, k}^h\| \leq \frac{C}{C_{h,\lambda,\gamma}}, \quad \forall i, j \in \llbracket i_0 + 1, N + 1 \rrbracket,$$

Define  $U_j$  as in Lemma 2.3.2

$$U_j := \left( \sqrt{\frac{\gamma_{\max}}{\lambda}} \frac{u_j - u_{j-1}}{h} \right).$$

By definition of the resolvent,  $\forall i \geq i_0 + 1$ ,

$$\begin{aligned} \|U_{i_0}\| &\geq CC_{h,\lambda,\gamma}\|U_i\| \\ |u_{i_0}| + \sqrt{\frac{\gamma_{\max}}{\lambda h^2}}(|u_{i_0}| + |u_{i_0-1}|) &\geq CC_{h,\lambda,\gamma}|u_i|, \end{aligned}$$

and according to (2.138),  $|u_{i_0}| > |u_{i_0-1}|$ , thus,

$$\forall i \in \llbracket i_0 + 1, N \rrbracket \quad |u_{i_0}| > CC_{h,\lambda,\gamma}|u_i|, \quad (2.139)$$

and (2.138) combined with (2.139) proves (2.137).  $\blacksquare$

**Lemma 2.7.2** *The vector  $U_{i_0+1}$  satisfies*

$$\|U_{i_0+1}\| \geq CC_{h,\lambda,\gamma}. \quad (2.140)$$

**Proof** Recall that  $\sum_{i=1}^N hu_i^2 = 1$ , so (2.137) gives

$$\begin{aligned} |u_{i_0}|^2 &\geq CC_{h,\lambda,\gamma}^2 \left( \sum_{i=1}^N h|u_i|^2 - h|u_{i_0}|^2 \right) \\ |u_{i_0}|^2 (1 + CC_{h,\lambda,\gamma}^2 h) &\geq CC_{h,\lambda,\gamma}^2 \end{aligned}$$

and for  $h$  small enough (notice that  $C_{h,\lambda,\gamma}$  is bounded above by  $\sqrt{1 - \frac{\gamma_{\min}}{\gamma_{\max}}}$ ),

$$|u_{i_0}| \geq CC_{h,\lambda,\gamma}. \quad (2.141)$$

Now,

$$\begin{aligned}
 \|U_{i_0+1}\|^2 &= u_{i_0+1}^2 + \frac{\gamma_{\max}}{\lambda h^2} (u_{i_0+1} - u_{i_0})^2 \\
 &= u_{i_0+1}^2 + \frac{1}{4} (u_{i_0+1}^2 + u_{i_0}^2 - 2u_{i_0}u_{i_0+1}) \\
 &\geq \frac{5}{4} u_{i_0+1}^2 + \frac{1}{4} u_{i_0}^2 - \frac{1}{4} \left( \frac{1}{\rho} u_{i_0}^2 + \rho u_{i_0+1}^2 \right) \\
 &\geq \frac{5-\rho}{4} u_{i_0+1}^2 + \frac{1}{4} \left( 1 - \frac{1}{\rho} \right) u_{i_0}^2,
 \end{aligned}$$

we choose, for example,  $\rho = 5$ , and we get

$$\|U_{i_0+1}\|^2 \geq \frac{1}{5} u_{i_0}^2.$$

Combining this inequality with (2.141), we get (2.140).

■

We can now conclude the proof of Theorem 2.7.1.

— The resolvent estimate gives

$$\|U_{N+1}\| \geq CC_{h,\lambda,\gamma} \|U_{i_0+1}\|,$$

thus, applying Lemma 2.7.2

$$\left| \frac{u_N}{h} \right| \geq CC_{h,\lambda,\gamma}^2,$$

and this gives (2.123)

— The resolvent estimates gives also

$$\forall i \geq i_0 + 1, \quad \|U_i\|^2 \geq C_{h,\lambda,\gamma}^2 \|U_{i_0+1}\|^2$$

and applying Lemma 2.7.2 again,

$$|u_i|^2 + |u_{i-1}|^2 \geq CC_{h,\lambda,\gamma}^4,$$

summing for  $ih \in \omega_h^2 \setminus \min(\omega_h^2)$ ,

$$\sum_{ih \in \omega_h^2} h |u_i|^2 \geq CC_{h,\lambda,\gamma}^4,$$

and this gives (2.124).

■

## 2.7.4 Null control with separated variables for the discretized heat equation

In this chapter, we show that the discretized heat equation on  $\Omega = (0, 1)$ , on a uniform mesh, can be null controlled on a domain  $\omega$  with the moments method by building a control of the second form (see (1.11) for the continuous analogue expression). In this appendix, we consider the case of a constant diffusion operator and a uniform mesh. We show how to construct a control of the first form (1.10) to get the same result. As in [19], we make use of the explicit formulas of the eigenfunctions. This is one of the reasons this approach is not useful in general.

### The distributed control problem

Let us recall the expression of the problem under study which is the semi-discretization in space of the heat equation with a distributed control acting on  $\omega$ .

$$\begin{cases} \partial_t y^h - \Delta^h y^h = 1_\omega^h V_d^h(t), & \text{for } t \in (0, T), \\ y_0^h(t) = y_{N+1}^h(t) = 0, & \text{for } t \in (0, T), \\ y^h(0) = y^{h,0}. \end{cases} \quad (2.142)$$

The eigenvalues of the discretized Laplace operator with Dirichlet boundary conditions  $-\Delta^h$  are denoted by  $\Lambda^h := (\lambda_k^h)_{k \geq 1}$  and the corresponding discrete eigenfunctions are defined by  $(\phi_k^h)_{k \geq 1}$ . We have chosen a uniform mesh, thus their expressions are known:

$$\lambda_k^h = \frac{4}{h^2} \sin^2 \left( \frac{kh\pi}{2} \right) \text{ and } (\phi_k^h)_j = \sqrt{2} \sin(h\pi k j), \forall k, j \in \llbracket 1, N \rrbracket. \quad (2.143)$$

By taking the discrete  $L^2$  inner product of equation (2.142) with the discrete function  $t \mapsto e^{-(T-t)\lambda_k^h} \phi_k^h$  and by integrating over  $(0, T)$ , we find that

$$y^h(T) = 0 \Leftrightarrow \forall k \in \llbracket 1, N \rrbracket, -e^{-\lambda_k^h T} \langle y^{h,0}, \phi_k^h \rangle_{L^2(\Omega_h)} = \int_0^T \langle V_d^h(t), e^{-\lambda_k^h(T-t)} \phi_k^h \rangle_{L^2(\omega_h)} \omega dt. \quad (2.144)$$

In order to build a biorthogonal family to functions  $(t \mapsto e^{-\lambda_k^h(T-t)})_{k \in \llbracket 1, N \rrbracket}$ , we extend the sequence  $(\lambda_k^h)_{k \in \llbracket 1, N \rrbracket}$  in such a way that the uniform gap condition and the existence of function  $\mathcal{N}$  hold (see Theorem 1.1.3). We propose, for instance, the same extension has the one given in [19]:

$$\lambda_k^h = \begin{cases} \frac{4}{h^2} \sin^2 \left( \frac{\pi h k}{2} \right), & \text{for } k \in \llbracket 1, N \rrbracket, \\ \pi^2 k^2, & \text{for } k > N. \end{cases}$$

Remark that for any  $k \in \llbracket 1, N-1 \rrbracket$ ,  $\lambda_{k+1}^h - \lambda_k^h \geq \lambda_N^h - \lambda_{N-1}^h \sim 3\pi^2$  thus, for example,  $\rho = 1$  is a suitable parameter. Moreover, since  $\lambda_k^h \geq 4k^2$  for any  $k \in \mathbb{N}^*$ , the value of  $\mathcal{N}(\delta)$  is chosen so as  $\sum_{k \geq \mathcal{N}(\delta)} \frac{1}{k^2} \leq 4\delta$ . Hence, for any  $h > 0$ ,  $\Lambda^h \in \mathcal{L}(1, \mathcal{N})$  and we can now use Theorem 1.1.3 to build a

biorthogonal family  $(q_j^{\Lambda^h})_{j \geq 1}$ , uniformly bounded with respect to the mesh parameter  $h$ . With the equivalence (2.144) we can write the null control under the form

$$V_d^h = - \sum_{k=1}^N \frac{e^{-\lambda_k^h T} \langle y^{0,h}, \phi_k^h \rangle_{L^2(\Omega_h)}}{\langle f^h, \phi_k^h \rangle_{L^2(\omega_h)}} f^h q_k^{\Lambda^h},$$

as long as we can build a discrete function  $f^h$  such that for any  $\varepsilon > 0$  there exists  $C_\varepsilon > 0$  independent of  $h$  such that  $\forall h > 0, \forall k \in \llbracket 1, N \rrbracket, \langle f^h, \phi_k^h \rangle_{L^2(\omega_h)} \geq C_\varepsilon e^{-\varepsilon \lambda_k^h}$ .

The construction of such discrete function  $f^h$  is based on the existence of a continuous profile  $f = 1_{(a,b)}$ , which gives the inequality (1.14). However, to our knowledge, the expression of such profile  $f$  such that  $(f, \phi_k)_{L^2(\omega)} \geq C e^{-\varepsilon \lambda_k}$ , is only known for the operator  $\mathcal{A} = -\Delta$ . So the following method does not apply to a discretized parabolic equation with any other Sturm Liouville operator  $\mathcal{A}$ . Because of that, the approach developed in chapter 2 with a control of the form (1.11) is more adapted to this kind of problem since it does not require to create a discrete function  $f^h$ .

### Construction of $f^h$

We show the following theorem which completes the section 1.1.2 of the introduction. As in the continuous case, we find a lower bound of the form  $\frac{C}{h^e}$ , with  $e \in \mathbb{N}$ , which is better than the bound  $C_\varepsilon e^{-\varepsilon \lambda_k^h}$  which would have been sufficient.

**Theorem 2.7.2** *There exist two constants (independent of  $h$ ) denoted by  $C > 0$  and  $e \in \mathbb{N}$ , such that for any  $h > 0$  sufficiently small, there exists a profile  $f^h$  supported in  $\omega$  and uniformly bounded with respect to  $h$  such that for any  $k \in \llbracket 1, N \rrbracket$ ,*

$$\langle f^h, \phi_k^h \rangle_{L^2(\omega_h)} \geq \frac{C}{h^e}.$$

**Proof** Let us choose  $j_-^h h := x_-^h$  (respectively  $j_+^h h := x_+^h$ ) the mesh points of the interval  $(a, b)$  which is the closest to  $a$  (respectively to  $b$ ) and let us define

$$(f^h)_j := (1 + r_x(j)) 1_{[x_-^h, x_+^h]}(jh),$$

where  $r_x(j)$  is defined by

$$r_x(j) := xh^4 \mathbf{1}_{\left\{\frac{j_-^h + j_+^h}{2}\right\}}(j) + x^2 h^5 \mathbf{1}_{\left\{\frac{j_-^h + j_+^h}{2} + 1\right\}}(j).$$

Without loss of generality, we can change  $j_-^h$  into  $j_-^h + 1$ , so that we can suppose that  $\frac{j_-^h + j_+^h}{2} \in \mathbb{N}^*$ . The real number  $x \in (0, 1)$  will be defined later on. Remark that  $r_h(j) = O(h^4)$ .

On the one hand

$$\begin{aligned} \left| \int_{\Omega} f(x) \sin(k\pi x) dx - \langle f^h, \phi_k^h \rangle_{L^2(\omega_h)} \right| &= \left| \sum_{j=j_-^h}^{j_+^h-1} \int_{j^h}^{(j+1)h} \sin(k\pi x) - \sin(k\pi h j) \right| + O(h^4) \\ &\leq k\pi h + O(h^4) \end{aligned}$$

We consider indices  $k$  verifying

$$k \leq \frac{m}{h^{1/4}} \Leftrightarrow h \leq \frac{m^4}{k^4},$$

where  $m > 0$  is a constant that we shall define later. Then, we have,

$$\left| \int_{\Omega} f(x) \sin(k\pi x) dx - \langle f^h, \phi_k^h \rangle_{L^2(\omega_h)} \right| \leq \frac{\pi m^4}{k^3} + O(h^4),$$

where  $m$  is chosen sufficiently small so that

$$\langle f^h, \phi_k^h \rangle_{L^2(\omega_h)} \geq \frac{C}{k^3}, \text{ with } C > \pi m^4, \text{ independent of } h.$$

On the other hand, to tackle the high frequencies (that is the indices  $k$  such that  $k > \frac{m}{h^{1/4}}$ ), it suffices to directly compute the sum  $\langle f^h, \phi_k^h \rangle_{L^2(\omega_h)}$ . We end up with

$$\begin{aligned} \langle f^h, \phi_k^h \rangle_{L^2(\omega_h)} &= h \frac{\sin\left(k\pi h \frac{j_-^h + j_+^h}{2}\right) \sin\left(k\pi h \frac{j_+^h - j_-^h + 1}{2}\right)}{\sin\left(\frac{k\pi h}{2}\right)} + xh^5 \sin\left(hk\pi \frac{j_-^h + j_+^h}{2}\right) \\ &\quad + x^2 h^6 \sin\left(hk\pi \left(\frac{j_-^h + j_+^h}{2} + 1\right)\right). \end{aligned}$$

Now we choose a real number  $x \in (\frac{1}{2}, 1)$  (depending on  $h$ ) such that this polynomial of variable  $x$  never vanishes for any  $k \in \llbracket 1, N \rrbracket$ . Hence, different cases can occur.

- In the case where  $\sin\left(k\pi h \frac{j_-^h + j_+^h}{2}\right) \sin\left(k\pi h \frac{j_+^h - j_-^h + 1}{2}\right) \neq 0$ ,

$$\begin{aligned} |\langle f^h, \phi_k^h \rangle_{L^2(\omega_h)}| &\geq h \left| \frac{\sin\left(k\pi h \frac{j_-^h + j_+^h}{2}\right) \sin\left(k\pi h \frac{j_+^h - j_-^h + 1}{2}\right)}{\sin\left(\frac{k\pi h}{2}\right)} \right| - O(h^4) \\ &\geq h \left| \sin^2\left(\frac{\pi h}{2}\right) \right| - O(h^4) \\ &\geq Ch^3. \end{aligned}$$

where  $C > 0$  does not depend on  $h$ .

- In the case where  $\sin\left(k\pi h \frac{j_-^h + j_+^h}{2}\right) \sin\left(k\pi h \frac{j_+^h - j_-^h + 1}{2}\right) = 0$ , we get a lower bound of order  $h^7$  thanks to function  $r_x$ .

Indeed, if  $\sin\left(k\pi h \frac{j_-^h + j_+^h}{2}\right) = 0$  then  $\sin\left(hk\pi \left(\frac{j_-^h + j_+^h}{2} + 1\right)\right) \neq 0$  and

$$|\langle f^h, \phi_k^h \rangle_{L^2(\omega_h)}| = x^2 h^6 \left| \sin\left(hk\pi \left(\frac{j_-^h + j_+^h}{2} + 1\right)\right) \right| = x^2 h^6 |\sin(hk\pi)| \geq Ch^7.$$

Finally, if  $\sin\left(k\pi h \frac{j_-^h + j_+^h}{2}\right) \neq 0$  then

$$|\langle f^h, \phi_k^h \rangle_{L^2(\omega_h)}| \geq Ch^5.$$

Since  $k > \frac{m}{h^{1/4}} \Leftrightarrow h > \frac{m^4}{k^4}$ , there exists  $e \in \mathbb{N}^*$  ( $e = 28$  for instance), and  $C > 0$  such that

$$\forall h > 0, \exists f^h, \forall k \in \llbracket 1, N \rrbracket, \langle f^h, \phi_k^h \rangle_{L^2(\omega_h)} \geq \frac{C}{k^e},$$

which concludes the proof. ■

## 2.7.5 Simulations

As mentioned in section 1.1.3 of the introduction, the HUM approach provides us with an algorithm to run simulations on discretized equations. We shall illustrate the HUM method by considering two controlled systems.

### Null control of an equation with an unstable mode

Consider the system

$$\begin{cases} \partial_t y - \Delta y - 1.5\pi^2 y = 0, & \text{for } (t, x) \in (0, T) \times (0, 1), \\ y(t, 0) = V_b(t), & \text{for } t \in (0, T), \\ y(t, 1) = 0, & \text{for } t \in (0, T), \\ y(0, x) = \sin^{10}(\pi x). \end{cases} \quad (2.145)$$

We fully discretize equation (2.145) with the finite differences method, on a uniform mesh. The number of mesh points  $N$  belongs to the set  $\{25, 50, 100, 200, 400\}$ . We take 300 time steps and  $T = 0.5$ . Note that the discrete  $L^2$  norm of the discretization of the initial condition  $\sin^{10}(\pi x)$  is roughly 0.42.

As we can see, the graph in figure 2.7 shows that, with the penalized HUM approach, we can compute a discrete control that makes the solution at time  $T$  become very small compared to the norm of its initial value. This is the result of the action of control  $V_b^h$  on the system because it can not be the consequence of a parabolic dissipation of energy. Indeed, notice that since the potential is constant, the first eigenvalue of the operator  $-\Delta^h - 1.5\pi^2$  is roughly equal to  $-0.5\pi^2$  for the range of values of  $N$  that we consider. Thus, if  $V_b^h = 0$  then the solution blows up when  $T$  becomes large. This can also be seen on the figure since the solution at time  $T$  is roughly equal to 1, that is more than twice larger than the norm of the initial condition.

We consider a discrete version of the energy (1.15) and we make the penalization parameter depend on  $h$  (see the discussions in [9, Section 1.7 and 2.2]). We define  $\varepsilon$  as a function of  $h$ , denoted by  $\phi$ , and we set  $\phi(h) = h^4$ . Following the notations of [9], the functional is now denoted by  $F_{h, \phi(h)}$ . As we can see on figure 2.7, the energy seems to remain bounded. Thus, a criterion given in [9], states that in this case, the fully discretized system is  $\phi(h)$  null controllable, which is consistent with the behaviour of the curve  $h \mapsto \|y^h(T)\|$  observed on figure 2.7.

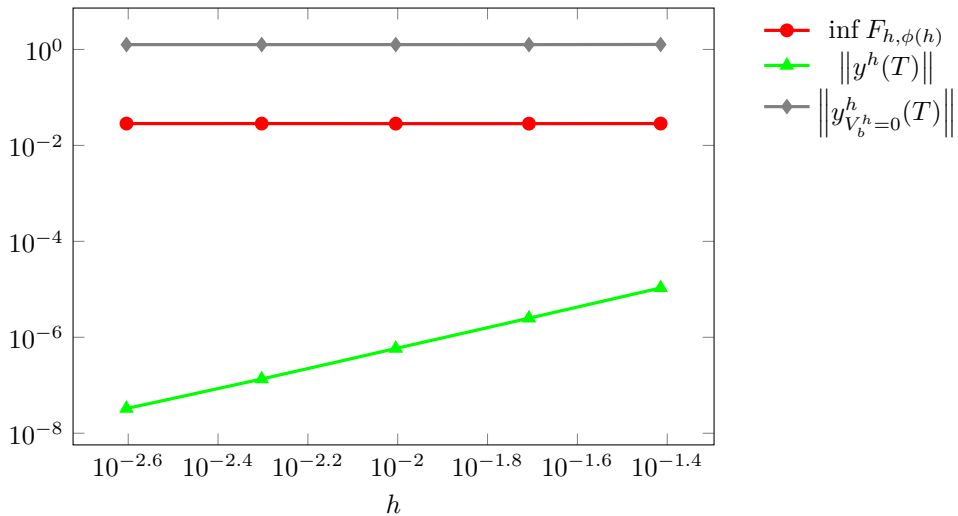


Figure 2.7 – System (2.145) controlled with a boundary control



### A system controllable only with both boundary and distributed controls

In this chapter, we proved that, under some hypothesis on the elliptic operator, a continuous cascade system of equations is null controllable at the boundary or with a distributed control. When the system is not in a cascade form, new phenomena may appear. In his PhD [21], Guillaume Olive builds a system that is neither null controllable at the boundary nor with a distributed control. However, it turns out that this system is null controllable when we activate both. This system is made of 3 equations and it reads

$$\begin{cases} \partial_t y - \Delta y = Cy + 1_\omega B_d V_d, & \text{on } (0, T) \times \Omega, \\ y(0) = y^0, & \text{on } \Omega, \\ y(t, 0) = B_l V_b, & \text{on } (0, T), \\ y(t, 1) = 0, & \text{on } (0, T), \end{cases} \quad (2.146)$$

with  $\Omega = (0, 1)$ ,  $C = \begin{pmatrix} 2 & 6 & 2 \\ 4 & 0 & -2 \\ 2 & -3/2 & 2 \end{pmatrix}$ ,  $B_l = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$  and  $B_d = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ . Those matrices are chosen

such that the conditions of the Theorem 1.1 of [4] hold for the equation (2.146) but become false if either  $B_l$  or  $B_d$  is set to zero. Let us see now how the discretized version of this system behaves in those cases. As before, we choose  $\phi(h) = h^4$  and we discretize the equation on a uniform mesh for  $N \in \{25, 50, 100, 200, 400\}$  mesh points. The number of time steps is set to 100. The initial condition is set to 1 for every component of the system and the final time  $T$  equals 1. As we can see on figures 2.8a and 2.8b the energy seems to increase as  $h \rightarrow 0$ . Thus, the functional does not seem to be uniformly bounded with respect to  $h$ . As a consequence, the norm of the solution does not seem to converge to zero, but stays quite small though. Hence, these graphics suggest that the fully discrete system is not null controllable. However in the last case 2.8c, the energy remains roughly constant and the solution seems to be strictly decreasing. According to this last figure the system controlled by  $V_d$  and  $V_b$  seems to be  $\phi(h)$  null controllable, as in the continuous case.

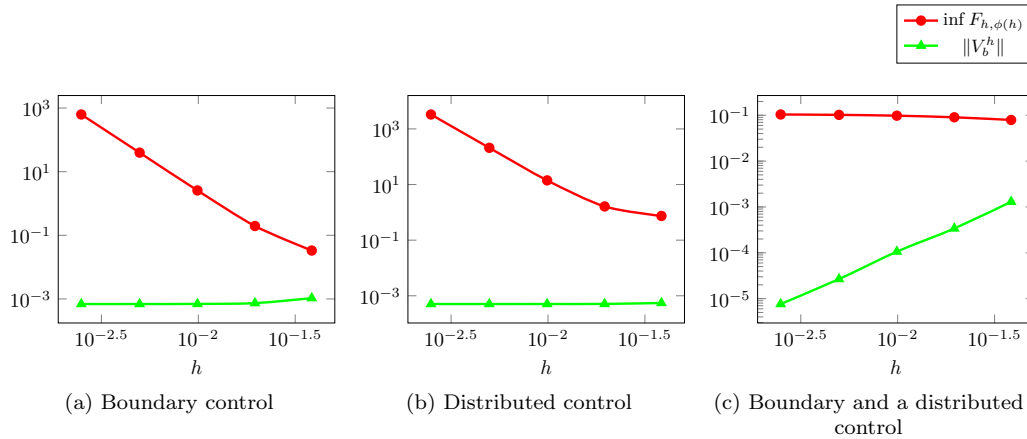


Figure 2.8 – Controllability of a discretized version of system (2.146)

## 2.8 References

- [1] D. ALLONSIUS AND F. BOYER, *Boundary null-controllability of semi-discrete coupled parabolic systems in some multi-dimensional geometries*. working paper or preprint, July 2018, <https://hal.archives-ouvertes.fr/hal-01827044>. 62
- [2] D. ALLONSIUS AND F. BOYER, *Error estimate for the finite difference discretization of a sturm-liouville eigenvalue problem on quasi-uniform 1d meshes*. In preparation, 2018. 66
- [3] D. ALLONSIUS, F. BOYER, AND M. MORANCEY, *Spectral analysis of discrete elliptic operators and applications in control theory*, Numerische Mathematik, (2018), doi:10.1007/s00211-018-0983-1, <https://doi.org/10.1007/s00211-018-0983-1>. 26, 62

- [4] F. AMMAR-KHODJA, A. BENABDALLAH, M. GONZÁLEZ-BURGOS, AND L. DE TERESA, *The Kalman condition for the boundary controllability of coupled parabolic systems. Bounds on biorthogonal families to complex matrix exponentials*, J. Math. Pures Appl. (9), 96 (2011), pp. 555–590. [54](#), [55](#), [83](#)
- [5] F. AMMAR KHODJA, A. BENABDALLAH, M. GONZÁLEZ-BURGOS, AND L. DE TERESA, *Minimal time for the null controllability of parabolic systems: the effect of the condensation index of complex sequences*, J. Funct. Anal., 267 (2014), pp. 2077–2151. [52](#), [54](#)
- [6] F. AMMAR KHODJA, A. BENABDALLAH, M. GONZÁLEZ-BURGOS, AND L. DE TERESA, *New phenomena for the null controllability of parabolic systems: minimal time and geometrical dependence*, J. Math. Anal. Appl., 444 (2016), pp. 1071–1113. [54](#)
- [7] S. BARBEIRO, J. FERREIRA, AND R. GRIGORIEFF, *Supraconvergence of a finite difference scheme for solutions in  $H^s(0, L)$* , IMA Journal of Numerical Analysis, 25 (2005), pp. 797–811. [28](#), [42](#), [72](#)
- [8] D. BOFFI, *Finite element approximation of eigenvalue problems*, Acta Numerica, 19 (2010), pp. 1–120. [42](#)
- [9] F. BOYER, *On the penalised HUM approach and its applications to the numerical approximation of null-controls for parabolic problems*, in CANUM 2012, 41e Congrès National d’Analyse Numérique, vol. 41 of ESAIM Proc., EDP Sci., Les Ulis, 2013, pp. 15–58. [33](#), [34](#), [82](#)
- [10] F. BOYER, F. HUBERT, AND J. L. ROUSSEAU, *Discrete carleman estimates for elliptic operators and uniform controllability of semi-discretized parabolic equations*, Journal de Mathématiques Pures et Appliquées, 93 (2010), pp. 240–276. [31](#), [33](#), [34](#), [62](#)
- [11] F. BOYER, F. HUBERT, AND J. L. ROUSSEAU, *Discrete carleman estimates for elliptic operators in arbitrary dimension and applications*, SIAM Journal on Control and Optimization, 48 (2010), pp. 5357–5397.
- [12] F. BOYER AND J. L. ROUSSEAU, *Carleman estimates for semi-discrete parabolic operators and application to the controllability of semi-linear semi-discrete parabolic equations*, Annales de l’Institut Henri Poincaré (C) Non Linear Analysis, 31 (2014), pp. 1035–1078. [33](#), [34](#)
- [13] J.-M. CORON, *Control and nonlinearity*, vol. 136 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2007. [53](#)
- [14] S. ERVEDOZA, A. MARICA, AND E. ZUAZUA, *Numerical meshes ensuring uniform observability of one-dimensional waves: construction and analysis*, IMA Journal of Numerical Analysis, 36 (2015), pp. 503–542. [31](#)
- [15] H. O. FATTORINI AND D. L. RUSSELL, *Uniform bounds on biorthogonal functions for real exponentials with an application to the control theory of parabolic equations*, Quart. Appl. Math., 32 (1974/75), pp. 45–69. [28](#), [52](#), [54](#), [55](#)
- [16] R. GŁOWINSKI, J.-L. LIONS, AND J. HE, *Exact and approximate controllability for distributed parameter systems*, vol. 117 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2008. A numerical approach. [34](#)
- [17] S. LABBÉ AND E. TRÉLAT, *Uniform controllability of semidiscrete approximations of parabolic control systems*, Systems & Control Letters, 55 (2006), pp. 597–609. [33](#), [34](#)
- [18] J. LAGNESE, *Control of wave processes with distributed controls supported on a subregion*, SIAM J. Control Optim., 21 (1983), pp. 68–85. [54](#)
- [19] A. LOPEZ AND E. ZUAZUA, *Some new results related to the null controllability of the 1-d heat equation*, in Séminaire sur les Équations aux Dérivées Partielles, 1997–1998, École Polytech., Palaiseau, 1998, pp. Exp. No. VIII, 22. [33](#), [79](#), [80](#)
- [20] S. MICU, I. ROVENȚA, AND L. E. TEMEREANȚĂ, *Approximation of the controls for the linear beam equation*, Math. Control Signals Systems, 28 (2016), pp. Art. 12, 53. [61](#)
- [21] G. OLIVE, *Contrôlabilité de systèmes paraboliques linéaires couplés*, PhD thesis, Aix-Marseille université, 2013, <http://www.theses.fr/2013AIXM4329>. Thèse de doctorat dirigée par Boyer, Franck et Benabdallah, Assia Mathématiques appliquées Aix-Marseille 2013. [83](#)

- [22] J. PÖSCHEL AND E. TRUBOWITZ, *Inverse spectral theory*, vol. 130 of Pure and Applied Mathematics, Academic Press, Inc., Boston, MA, 1987. [27](#), [34](#)
- [23] H. F. WEINBERGER, *Lower bounds for higher eigenvalues by finite difference methods*, Pacific J. Math., 8 (1958), pp. 339–368; erratum, 941. [42](#)
- [24] E. ZUAZUA, *Propagation, observation, and control of waves approximated by finite difference methods*, SIAM Rev., 47 (2005), pp. 197–243. [33](#), [62](#)



# Chapter 3

## Boundary null-controllability of semi-discrete coupled parabolic systems in some multi-dimensional geometries

” I: ‘Would your Lordship indicate or explain to me in what direction is the Third Dimension, unknown to me ?’ Stranger: ‘I came from it. It is up above and down below.’ I: ‘My Lord means seemingly that it is Northward and Southward.’ Stranger: ‘I mean nothing of the kind. I mean a direction in which you cannot look, because you have no eye in your side.’ ”

---

Flatland - Edwin Abbott

### Contents

---

<b>3.1 Introduction</b>	<b>89</b>
3.1.1 Motivating example	89
3.1.2 Passing to the discrete world	90
3.1.3 Main results and outline of the paper	91
<b>3.2 Controllability for linear ODEs</b>	<b>92</b>
3.2.1 Framework	92
3.2.2 Well-posedness	93
3.2.3 Relaxed observability inequalities	94
<b>3.3 Controllability of tensorized systems</b>	<b>95</b>
3.3.1 Notations	95
3.3.2 Main Theorem	96
3.3.3 Proof of the main result	97
<b>3.4 Applications</b>	<b>105</b>
3.4.1 Dirichlet boundary null control of a semi-discrete cascade system on a rectangle	105
3.4.2 Dirichlet boundary null control of a continuous n-dimensional system on a cylindrical domain of dimension d	115
<b>3.5 Appendix</b>	<b>118</b>
3.5.1 Proof of Lemma 3.2.1	118
3.5.2 Numerical Illustrations	119
<b>3.6 References</b>	<b>121</b>

---

### Résumé

Dans ce papier, nous étudions les propriétés de contrôlabilité de systèmes paraboliques couplés semi-discrétisés en espace avec moins de contrôles que d'équations, en dimension d'espace supérieure à 1. Nous nous intéressons particulièrement au cas du contrôle au bord qui est particulièrement plus difficile que le cas du contrôle distribué, bien que notre analyse soit plus générale.

L'hypothèse principale que nous faisons concerne la géométrie du domaine et l'équation d'évolution : on suppose que ceux-ci peuvent s'écrire sous une forme tensorisée. Dans ce cas, en suivant les idées de [4] et en utilisant une version adaptée de la procédure de Lebeau-Robbiano, nous prouvons des résultats de contrôlabilité pour ces systèmes semi-discrétisés (pourvu que la structure du terme de couplage satisfasse des conditions de Kalman nécessaires) avec des bornes uniformes sur les contrôles.

Pour cela, nous proposons un résultat abstrait sur des systèmes d'équations différentielles ordinaires et nous donnons des estimées de la solution et du coût du contrôle dont les dépendances par rapport aux paramètres du système sont soigneusement étudiées. En appliquant ces résultats à des ODE provenant de la discrétisation en espace d'un système parabolique, nous obtenons des estimées uniformes par rapport au paramètre de discrétisation.

### Abstract

The main goal of this paper is to investigate the controllability properties of semi-discrete in space coupled parabolic systems with less controls than equations, in dimension greater than 1. We are particularly interested in the boundary control case which is notably more intricate than the distributed control case, even though our analysis is more general.

The main assumption we make on the geometry and on the evolution equation itself is that it can be put into a tensorized form. In such a case, following [4] and using an adapted version of the Lebeau-Robbiano construction, we are able to prove controllability results for those semi-discrete systems (provided that the structure of the coupling terms satisfies some necessary Kalman condition) with uniform bounds on the controls.

To achieve this objective we actually propose an abstract result on ordinary differential equations with estimates on the control and the solution whose dependence upon the system parameters are carefully tracked. When applied to an ODE coming from the discretization in space of a parabolic system, we thus obtain uniform estimates with respect to the discretization parameters.

## 3.1 Introduction

### 3.1.1 Motivating example

In this paper, we are interested in null-controllability properties at any time  $T > 0$  of coupled linear parabolic equations at the continuous level as well as at the semi-discrete in space level. In this introduction, we will focus on the following prototype system (which is in the so-called cascade form)

$$\begin{cases} \partial_t \alpha - \Delta \alpha = 0, & \text{in } (0, T) \times \Omega, \\ \partial_t \beta - \Delta \beta + \alpha = 0, & \text{in } (0, T) \times \Omega, \\ \alpha = 1_\Gamma v, & \text{on } (0, T) \times \partial\Omega, \\ \beta = 0, & \text{on } (0, T) \times \partial\Omega, \end{cases} \quad (3.1)$$

where  $\Omega$  is a bounded domain of  $\mathbb{R}^d$  ( $d \geq 1$ ),  $\Gamma$  is a non empty part of the boundary  $\partial\Omega$ ,  $\alpha$  and  $\beta$  are the two components of the system, and  $v$  is the boundary control we are looking for. The main difficulty in the analysis of the controllability of such system comes from the fact that we only have one boundary control  $v$  to drive the two components  $(\alpha, \beta)$  to 0 at the final time. The coupling terms (here the term  $\alpha$  in the equation for  $\beta$ ) plays a key role in the problem and it can be seen as a kind of indirect control for the second component of the system. Note that this indirect controllability issue arises even if  $\Gamma = \partial\Omega$ .

Actually, it appears that the results for such systems may be quite different from the case of scalar equations or from the case of coupled systems with a distributed control. We refer for instance to the survey [2] for a review on that topic. In particular, it is explained in that reference that usual techniques based on Carleman estimates are useless on those problems. This is mainly because those techniques naturally gives the controllability of the system when there is as many controls as components of the system (for (3.1) it would consist in another boundary control for  $\beta$  on  $\Gamma$ ) and, in a second step, it is needed to prove that only one control is necessary. This is done, at the observability inequality level (see the discussion in Section 3.2.3) by removing one observation term thanks to the PDE itself. This last step cannot be done for boundary controls (or observations). This is why other approaches have to be developed.

Most of the results available up to now for such controllability problems are only proved in dimension  $d = 1$  by using the so-called moments method. This is a quite powerful method but, unfortunately, restricted to autonomous problem in space dimension 1. In the multi-dimensional case, one of the more advanced result available in the literature is proved in [4], in the case where the geometry and the diffusion operator can be tensorized and this is also the case we shall consider in the present work.

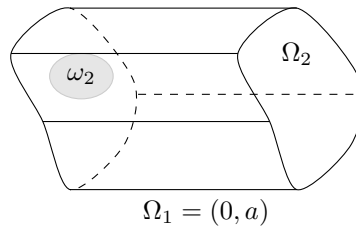


Figure 3.1 – Typical geometric situation

For problem (3.1), this structure assumption amounts to assume that the geometry is as follows:  $\Omega = \Omega_1 \times \Omega_2$  with  $\Omega_1 = (0, a)$  and  $\Omega_2 \subset \mathbb{R}^{d-1}$ , and  $\Gamma = \{0\} \times \omega_2$  (see Figure 3.1). Then, we can rewrite the system as follows

$$\begin{cases} \partial_t \alpha - \partial_{x_1}^2 \alpha - \Delta_2 \alpha = 0, & \text{in } (0, T) \times \Omega_1 \times \Omega_2, \\ \partial_t \beta - \partial_{x_1}^2 \beta - \Delta_2 \beta + \alpha = 0, & \text{in } (0, T) \times \Omega_1 \times \Omega_2, \\ \alpha = \beta = 0, & \text{on } (0, T) \times \Omega_1 \times \partial\Omega_2, \\ \alpha = 1_{\{0\} \times \omega_2} v, & \text{on } (0, T) \times \{0, a\} \times \Omega_2, \\ \beta = 0, & \text{on } (0, T) \times \{0, a\} \times \Omega_2, \end{cases} \quad (3.2)$$

where  $\Delta_2$  is the  $(d - 1)$ -dimensional Laplace operator in  $\Omega_2$ . The fact that the diffusion operator is split into two parts, each of them acting on different sets of variables, is crucial in the analysis. That

is the reason why we shall adopt a tensor product formalism that consists essentially in identifying  $L^2(\Omega)$  to  $L^2(\Omega_1) \otimes L^2(\Omega_2)$  and in writing the two equations above in the following equivalent form

$$\begin{cases} \partial_t \alpha + (-\partial_{x_1}^2) \otimes \mathcal{I} \alpha + \mathcal{I} \otimes (-\Delta_2) \alpha & = 0, & \text{in } (0, T), \\ \partial_t \beta + (-\partial_{x_1}^2) \otimes \mathcal{I} \beta + \mathcal{I} \otimes (-\Delta_2) \beta + \alpha & = 0, & \text{in } (0, T), \end{cases} \quad (3.3)$$

where the same symbol  $\mathcal{I}$  is used for the identity operator in  $L^2(\Omega_1)$  and  $L^2(\Omega_2)$ . All the necessary notations concerning tensor products will be recalled in Section 3.3.

By exploiting this tensor product structure, even though the tensor product formalism was not explicitly used, it was proved in [4], that the null-controllability of (3.3) holds at any time  $T > 0$ .

### 3.1.2 Passing to the discrete world

We are now interested in semi-discrete versions of the controllability result for (3.1) mentioned just before. To simplify the presentation in this introduction, we assume that  $d = 2$ , that  $\Omega_1 = \Omega_2 = (0, 1)$  and that the computation grid is made of  $N \times N$  uniformly distributed points  $(ih, jh)_{1 \leq i, j \leq N}$  with  $h = \frac{1}{N+1}$ . The semi-discrete system we consider is obtained by the finite difference method and reads

$$\left\{ \begin{array}{ll} \partial_t \alpha_{i,j} + \frac{4\alpha_{i,j} - \alpha_{i-1,j} - \alpha_{i+1,j} - \alpha_{i,j-1} - \alpha_{i,j+1}}{h^2} & = 0, & \forall 1 \leq i, j \leq N, \\ \partial_t \beta_{i,j} + \frac{4\beta_{i,j} - \beta_{i-1,j} - \beta_{i+1,j} - \beta_{i,j-1} - \beta_{i,j+1}}{h^2} + \alpha_{i,j} & = 0, & \forall 1 \leq i, j \leq N, \\ \alpha_{i,0} = \beta_{i,0} = \alpha_{i,N+1} = \beta_{i,N+1} & = 0, & \forall 1 \leq i \leq N, \\ \alpha_{0,j} = v_j \mathbf{1}_{\omega_2}(jh), & & \forall 1 \leq j \leq N, \\ \alpha_{N+1,j} = \beta_{0,j} = \beta_{N+1,j} & = 0, & \forall 1 \leq j \leq N. \end{array} \right. \quad (3.4)$$

The grid geometry is essentially the one described in Figure 3.2, where the control  $v = (v_j)_j$  is only appearing in the first equation of the system (the one for  $\alpha_{i,j}$ ) and only on the boundary points represented by the symbol  $\blacksquare$  corresponding to the subdomain  $\omega_2$ .

At each time  $t$ , both components  $\alpha_h = (\alpha_{i,j})_{i,j} \in \mathbb{R}^{N \times N}$  and  $\beta_h = (\beta_{i,j})_{i,j} \in \mathbb{R}^{N \times N}$  of the system are now considered as elements of the tensor product  $\mathbb{R}^N \otimes \mathbb{R}^N$  and we observe that the five-point discrete Laplace operator can be written as the tensor product  $A_h \otimes \mathbf{I} + \mathbf{I} \otimes A_h$  with the usual definition of the three-point discrete Laplace matrix

$$A_h = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & \diagdown & \diagup & & \\ 0 & \diagdown & \diagup & & \\ \vdots & \diagdown & \diagup & & \\ 0 & \dots & 0 & -1 & 2 \end{pmatrix}.$$

We finally end up with the following equivalent form of our semi-discrete system

$$\begin{cases} \partial_t \alpha_h + (A_h \otimes \mathbf{I} + \mathbf{I} \otimes A_h) \alpha_h & = B_h v_h, \\ \partial_t \beta_h + (A_h \otimes \mathbf{I} + \mathbf{I} \otimes A_h) \beta_h + \alpha_h & = 0, \end{cases} \quad (3.5)$$

where  $B_h$  is a matrix that accounts for the influence of the control  $v_h$  in the system through the boundary conditions in (3.4), the precise definition of which will be given in Section 3.4.1. This is the semi-discrete version of (3.3).

For this particular system, the main result of this paper is the following (see Section 3.4.1 for the precise statement and definition of the norms involved).

**Theorem** *There exist  $C > 0$  and  $h_0 > 0$  such that for any  $h < h_0$ , any time  $T > 0$  and any initial data  $\alpha_h^0, \beta_h^0 \in \mathbb{R}^N \otimes \mathbb{R}^N$ , (with  $h = \frac{1}{N+1}$ ) there exists a control  $v_h \in L^2(0, T, \mathbb{R}^N)$  such that*

$$\|v_h\|_{L^2(0, T, \mathbb{R}^N)} \leq C e^{\frac{C}{T}} (\|\alpha_h^0\| + \|\beta_h^0\|),$$

and the associated solution to (3.5) satisfies

$$\|\alpha_h(T)\| + \|\beta_h(T)\| \leq C e^{-C/h^2} e^{\frac{C}{T}} (\|\alpha_h^0\| + \|\beta_h^0\|).$$



It is well known, see [6, 16], that we cannot expect in general to achieve exactly  $\alpha_h(T) = \beta_h(T) = 0$  since the semi-discrete system may be not even approximately controllable. In this sense, achieving exponentially small targets with respect to  $h$  is an optimal result.

Our aim will be to provide similar results for more general semi-discrete systems. That is the reason why, in order to formulate them more conveniently for any number of coupled equations and to ease the reading of the proofs, we shall actually gather the two components  $(\alpha_h, \beta_h) \in (\mathbb{R}^N \otimes \mathbb{R}^N)^2$  into a single unknown  $y_h \in \mathbb{R}^N \otimes \mathbb{R}^N \otimes \mathbb{R}^n$ , with  $n = 2$  in the present case, in such a way that the considered system (3.5) will finally be written in the compact form

$$\partial_t y_h + (A_h \otimes I + I \otimes A_h) \otimes I y_h + I \otimes I \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} y_h = B_h \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} v_h.$$

In the sequel of this paper we shall not explicitly mention the subscript  $h$  for the notation of quantities related to the discretization process but taking into account the fact that the considered spaces, norms and operators are grid-dependent is a central point in the analysis.

### 3.1.3 Main results and outline of the paper

Considering the previous discussion we shall analyze in this paper the controllability of parabolic systems of  $n$  components and  $m$  controls of the following tensorized form

$$\partial_t y + \mathcal{A}_1 \otimes \mathcal{I} \otimes I y + \mathcal{I} \otimes \mathcal{A}_2 \otimes I y + \mathcal{I} \otimes \mathcal{I} \otimes C y = \mathcal{B}_1 \otimes \mathcal{B}_2 \otimes B v, \quad \text{in } (0, T), \quad (3.6)$$

where  $\mathcal{A}_i$  is a diffusion operator in  $\Omega_i$ ,  $\mathcal{B}_i$  is a (boundary or distributed) control operator in  $\Omega_i$ ,  $I$  is the  $n \times n$  identity matrix,  $C$  is a  $n \times n$  coupling matrix and  $B$  a  $n \times m$  control matrix.

Our main aim being to analyze semi-discrete versions of (3.6), we shall also consider linear ordinary differential equations of the similar form

$$\partial_t y + A_1 \otimes I \otimes I y + I \otimes A_2 \otimes I y + I \otimes I \otimes C y = B_1 \otimes B_2 \otimes B v, \quad \text{in } (0, T), \quad (3.7)$$

where the unknown  $y$  and the control  $v$  belong to a finite dimensional tensor space,  $A_i$  and  $B_i$  are linear operators on those space. All those objects depend, by nature, on discretization parameters, as in the example of Section 3.1.1. Therefore, it will be crucial to take care of all the constants in the estimates so as to obtain, at the end, controllability results for (3.7), that will not depend on those parameters.

**Remark 3.1.1** *We have used in (3.6) and (3.7) a convention that will be used all along the paper:*

- Operators acting in infinite dimensional function spaces are written with calligraphic letters:  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{I}$ , ...
- Operators (matrices) acting in finite dimensional spaces coming from discretization issues (whose dimension may be large and depends on the discretization parameters) are written with upright letters:  $A$ ,  $B$ ,  $I$ , ...
- Matrices acting in finite dimensional spaces coming the number of components or controls in the system (their dimension is fixed and independent of discretization parameters) are written with sans serif letters:  $B$ ,  $C$ ,  $I$ , ...

The outline of the paper is the following. Section 3.2 is dedicated to recalling the main results in the controllability theory for ODEs while paying a particular attention to the discrete functional setting that will be adapted to the analysis of finite difference approximations of parabolic PDEs. In section 3.3, we first review the material we need concerning tensorized operators that are central in the present work, then we state the precise assumptions we need and our main abstract result (Theorem 3.3.1). In short, we will assume for our tensorized system (3.6) (in the continuous setting) or (3.7) (in the discrete setting):

- that the associated sub-problem concerning only the first coordinate of the tensor product which is

$$\partial_t y + \mathcal{A}_1 \otimes I y + \mathcal{I} \otimes C y = \mathcal{B}_1 \otimes B v, \quad \text{in } (0, T), \quad (3.8)$$

or

$$\partial_t y + A_1 \otimes I y + I \otimes C y = B_1 \otimes B v, \quad \text{in } (0, T), \quad (3.9)$$

is null-controllable (or a relaxed version of this) at any time  $T$  with a precise control of the cost of the control.

If we come back to our motivating example, this subsystem reads

$$\begin{cases} \partial_t \alpha - \partial_{x_1}^2 \alpha = 0, & \text{in } (0, T) \times \Omega_1, \\ \partial_t \beta - \partial_{x_1}^2 \beta + \alpha = 0, & \text{in } (0, T) \times \Omega_1, \\ \alpha = 1_{\{0\}} v, & \text{on } (0, T) \times \partial\Omega_1, \\ \beta = 0, & \text{on } (0, T) \times \partial\Omega_1, \end{cases} \quad (3.10)$$

Note that this system does not depend on the control set  $\omega_2$ .

- that the diffusion operator  $\mathcal{A}_2$  (resp.  $A_2$ ) and the control operator  $\mathcal{B}_2$  (resp.  $B_2$ ) in the other direction satisfy a suitable spectral estimate similar to the Lebeau-Robbiano spectral inequality, except that we allow the inequality to hold only for a portion of the spectrum.

For system (3.2), this amounts to ask that the Lebeau-Robbiano spectral inequality holds, relative to the control set  $\omega_2$ , for the eigenfunctions of the operator  $-\Delta_2$  in  $\Omega_2$  with homogeneous boundary condition.

The complete proof of the theorem is given in Section 3.3.3. It consists, following the strategy developed in [4], to implement a construction of the control similar to the one originally proposed by Lebeau and Robbiano in [13]. We split the time interval into a suitable number of subintervals whose length is carefully chosen and, on each of those subintervals, we construct a partial control (obtained by combining the two assumptions above) that is able to damp out exponentially the part of the solution corresponding to the frequencies less than some threshold. Note that, contrary to the usual construction, we are not necessary able to drive this part of the solution exactly to zero at this stage. This threshold is then increased while the construction progresses towards the final time  $T$ ; this eventually gives the expected control.

The main novelties in the present work are that : we allow relaxed controllability and spectral inequalities in our assumptions and moreover we precisely take care of the dependence of all the quantities of interest (norms, constants, ...) with respect to parameters on which the problem may depend. Those two refinements of the proof in [4] are mandatory since we want to apply this abstract result to systems obtained by semi-discretization processes.

To conclude the paper, in section 3.4, we precisely explain how to use the abstract formalism developed here to achieve the uniform controllability results of semi-discrete coupled parabolic systems as announced in this introduction. As another example, we also show how to deduce the result of [4] (slightly generalized to variable coefficients operators) from the present abstract result and give some insights on other possible applications.

## 3.2 Controllability for linear ODEs

Before studying discrete versions of System (3.6), which is the main aim of the paper, we start by introducing the main notations and results that we shall use in the sequel concerning the controllability of linear ODEs. Most of this material is already well-known, however we propose a specific point of view adapted to our needs.

### 3.2.1 Framework

Let  $(E, \langle \bullet, \bullet \rangle_0)$  and  $(U, [\bullet, \bullet]_0)$  be two (finite dimensional) Euclidean spaces (each of them being identified with its own dual space). The corresponding norms are denoted by  $\|\bullet\|_0$  and  $\|\bullet\|_0$ . The presence of a subscript 0 in the notation is related to the fact that, in Section 3.2.2, those norms will be embedded in a scale of Sobolev-like norms.

We consider for the moment a general linear autonomous controlled system of the form

$$\begin{cases} y' + Ly = Bv, & \text{on } (0, T), \\ y(0) = y_0, \end{cases} \quad (3.11)$$

where  $L : E \rightarrow E$  and  $B : U \rightarrow E$  are two linear operators,  $y : [0, T] \rightarrow E$  is the state and  $v : [0, T] \rightarrow U$  is the control we are looking for.

In the sequel of this paper, different such systems will be considered, coming in particular from the discretization of multi-D parabolic control problems such as (3.7) or their reduced version (3.9). In

particular, the spaces  $E$ ,  $U$  and the operators  $L$  and  $B$  will depend on some discretization parameter  $h$ . We will be interested in properties of those systems that are uniform with respect to  $h$ , that is the reason why we will pay, in this section, a particular attention to the various constants appearing in the estimates. In section 3.4 we will propose a suitable framework ensuring that all those constants will be uniform with respect to  $h$ .

### 3.2.2 Well-posedness

It is clear that (3.11) is well-posed for any choice of  $y_0$  and  $v$  and that

$$\sup_{t \in [0, T]} \|y(t)\|_0 \leq e^{T\|L\|} (\|y_0\|_0 + \sqrt{T}\|B\| \|[v]\|_{L^2(0, T; U)}), \quad (3.12)$$

where  $\|L\|$  (resp.  $\|B\|$ ) is the operator norm of  $L : E \rightarrow E$  (resp.  $B : U \rightarrow E$ ). However, in the framework we are interested in which comes from the semi-discretization in space of an evolution PDEs, those operator norms will not be bounded in general with respect to the discretization parameter. This is the consequence at the discrete level of the fact that differential operators are naturally unbounded operators in Sobolev spaces. For example if  $L$  the discrete Laplace operator on a uniform mesh of size  $h$  and  $B$  the boundary control operator as defined in section 3.4.1, then  $\|B\|$  and  $\|L\|$  both behave like  $C/h^2$ . Thus, inequality (3.12) will not give usable estimates.

Therefore, we need to introduce adapted estimates and some kind of discrete Sobolev norms to take into account the particular geometry of the (discrete) control operators under study. To this end, we introduce  $D : E \rightarrow E$  a self-adjoint definite positive operator on  $E$  (one can think of the discrete Laplace operator for instance) and we define a scale of inner products in  $E$  defined, for any  $s \in \mathbb{R}$ , by

$$\langle u, v \rangle_{s, D} = \langle D^s u, v \rangle_0,$$

and  $\|\bullet\|_{s, D}$  is the associated norm. Observe that  $\|\bullet\|_0 = \|\bullet\|_{0, D}$  and that  $(E, \langle \bullet, \bullet \rangle_{-s, D})$  is naturally isometric to the dual of  $(E, \langle \bullet, \bullet \rangle_{s, D})$  since

$$\|u\|_{-s, D} = \sup_{\psi \in E} \frac{\langle u, \psi \rangle_0}{\|\psi\|_{s, D}}. \quad (3.13)$$

We shall now define, for given  $s \in \mathbb{R}$ , the two constants  $M_{s, \text{adm}} > 0$  and  $M_{s, \text{cont}} > 0$  that satisfy

$$\sup_{t \in [0, T]} \left\| e^{-tL^*} \psi \right\|_{s, D} \leq M_{s, \text{cont}} \|\psi\|_{s, D}, \quad \forall \psi \in E, \quad (3.14)$$

$$\left( \int_0^T \left[ \left\| B^* e^{-\tau L^*} \psi \right\|_0^2 d\tau \right] \right)^{\frac{1}{2}} \leq M_{s, \text{adm}} \|\psi\|_{s, D}, \quad \forall \psi \in E, \quad (3.15)$$

where the adjoint operators  $L^*$  and  $B^*$  are relative to the ambient inner product on  $E$  and  $U$ . Observe that (3.14) and (3.15) automatically holds since we consider finite dimension spaces, and the only interesting point is the uniformity (or not) of the constants with respect to the spaces and the operators involved. Depending on the targeted application (distributed control, Dirichlet boundary control or Neumann boundary control for instance) we will need to choose a convenient value of  $s$  and of the operator  $D$  to ensure that those constants are actually uniform with respect to the discretization parameter.

**Proposition 3.2.1** *For any  $v \in L^2(0, T; U)$  and any  $y_0 \in E$ , there exists a unique solution  $y$  to (3.11) and it satisfies*

$$\sup_{t \in [0, T]} \|y(t)\|_{-s, D} \leq M_{s, \text{cont}} \|y_0\|_{-s, D} + M_{s, \text{adm}} \|[v]\|_{L^2(0, T; U)}.$$

#### Proof

We write the Duhamel formula

$$y(t) = e^{-tL} y_0 + \int_0^t e^{-(t-s)L} B v(s) ds,$$

then we take the inner product with any  $\psi \in E$

$$\langle y(t), \psi \rangle_0 = \langle e^{-tL} y_0, \psi \rangle_0 + \int_0^t \langle e^{-(t-\tau)L} B v(\tau), \psi \rangle_0 d\tau.$$

It follows that

$$\langle y(t), \psi \rangle_0 = \langle y_0, e^{-tL^*} \psi \rangle_0 + \int_0^t [v(\tau), B^* e^{-(t-\tau)L^*} \psi]_0 d\tau,$$

and then

$$|\langle y(t), \psi \rangle_0| \leq \|y_0\|_{-s, D} \|e^{-tL^*} \psi\|_{s, D} + \int_0^t \llbracket v(\tau) \rrbracket_0 \llbracket B^* e^{-(t-\tau)L^*} \psi \rrbracket_0 d\tau.$$

By (3.14) and (3.15) we deduce that

$$|\langle y(t), \psi \rangle_0| \leq \left( M_{s, \text{cont}} \|y_0\|_{-s, D} + M_{s, \text{adm}} \llbracket v \rrbracket_{L^2(0, T; U)} \right) \|\psi\|_{s, D}.$$

Since this is valid for any  $\psi \in E$ , we deduce the expected estimate by the duality property (3.13). ■

**Remark 3.2.1** *During this work we will often use the following very standard duality formula that was given in the proof above*

$$\langle y(T), \psi \rangle_0 - \langle y_0, e^{-TL^*} \psi \rangle_0 = \int_0^T [v(t), B^* e^{-(T-t)L^*} \psi]_0 dt, \quad \forall \psi \in E. \quad (3.16)$$

### 3.2.3 Relaxed observability inequalities

It is well-known (see [10, 15] for instance) that System (3.11) is null controllable at time  $T$  for any initial condition  $y_0$  if and only if there exists  $C > 0$ , such that the following observability inequality for the adjoint problem is satisfied

$$\|e^{-TL^*} q_T\|_0^2 \leq C^2 \int_0^T \llbracket B^* e^{-(T-t)L^*} q_T \rrbracket_0^2 dt, \quad \forall q_T \in E. \quad (3.17)$$

The value of the constant  $C$  in this inequality is crucial since it appears in the measure of the control cost.

It happens that, when (3.11) comes from a discretization of a parabolic equation with the finite difference method then the null controllability of the semi-discrete system may not hold (in particular in a multi dimensional setting, see the example given by Kavian and reported in [16]). To tackle this problem, it was proposed (in [6–8, 11] for instance) to relax the controllability requirements by considering instead the  $\varphi(h)$ -null-controllability of (3.11). It consists in constructing uniformly bounded controls such that the solution  $y(T)$  does not identically vanish but is small enough with respect to the discretization parameter  $h$ . This approach is based on the penalized HUM construction, where the penalization parameter is a given function  $h \rightarrow \varphi(h)$  of the discretization parameter, given its name to this notion. Note that, the spaces  $E$ ,  $U$  and the operators  $L$  and  $B$  all depend on  $h$ , in particular the dimensions of  $E$  and  $U$  may increase when  $h$  tends to zero. This is one of the main difficulty that we need to take care of in the analysis.

The  $\varphi(h)$ -null-controllability property is equivalent to a relaxed version of inequality (3.17) and the following Lemma 3.2.1, whose proof is given in appendix 3.5.1, aims at establishing such an equivalence. This lemma is somehow related to [14, Lemma 3.4] or [3, Proposition 1] and is stated in a quite general framework : the constant  $s$  and the operator  $D$  can be chosen arbitrarily. Moreover using appropriate spaces  $F_0$  and  $F_T$  (which are defined below), one can show that Lemma 3.2.1 encompasses some already known situations (see Remark 3.2.3). Roughly speaking, this Lemma is about controlling the components in the final state space  $F_T$  of the solution of system (3.24) which starts from an initial condition  $y_0$  in the initial state space  $F_0$ . Even though we only state it in a finite dimensional setting, it is clear that infinite dimensional versions also hold, as in the references quoted above.

Let  $F_0$  and  $F_T$  be two subspaces of  $E$  and  $P_{F_0}$  (resp.  $P_{F_T}$ ) the orthogonal projection onto  $F_0$  (resp.  $F_T$ ) with respect to the inner product  $\langle \bullet, \bullet \rangle_{-s, D}$ .

We will denote the adjoint operators of the projectors  $P_{F_0}$  and  $P_{F_T}$  for the inner product  $\langle \bullet, \bullet \rangle_{0, D}$  by  $P_{F_0}^*$  and  $P_{F_T}^*$ . Observe that  $P_{F_0}^*$  and  $P_{F_T}^*$  are also the orthogonal projectors in  $(E, \langle \bullet, \bullet \rangle_{s, D})$  onto  $D^{-s}F_0$  and  $D^{-s}F_T$  respectively. In particular, we have

$$\begin{aligned} \|P_{F_0}^* y\|_{s, D} &\leq \|y\|_{s, D}, \quad \forall y \in E, \\ \|P_{F_T}^* y\|_{s, D} &\leq \|y\|_{s, D}, \quad \forall y \in E. \end{aligned}$$

**Remark 3.2.2** Observe that, if  $F_0$  (resp.  $F_T$ ) is stable by  $D$  then  $P_{F_0}$  (resp.  $P_{F_T}$ ) does not depend on  $s$ . In particular, those projections are orthogonal for the inner product  $\langle \bullet, \bullet \rangle_0$ .

Dealing with such subspaces will be crucial in the sequel when we will apply the Lebeau-Robbiano strategy since it requires to be able to control some precise components of the solution at each step (depending on eigenspaces of  $A_2$ ), see Section 3.3.3.

**Lemma 3.2.1 (Relaxed observability inequalities and controllability)** *We use the above notations and assume that  $s, D, F_0$  and  $F_T$  are given. Let  $M_{\text{obs}} > 0$  and  $M_{\text{rel}} \geq 0$  be two given numbers. The following two propositions are equivalent.*

1. For any  $y_0 \in F_0$  there exists a  $v \in L^2(0, T; U)$  satisfying

$$\frac{1}{M_{\text{obs}}^2} \|v\|_{L^2(0, T; U)}^2 + \frac{1}{M_{\text{rel}}^2} \|P_{F_T}(y(T))\|_{-s, D}^2 \leq \|y_0\|_{-s, D}^2, \quad (3.18)$$

where  $y$  is the corresponding solution of (3.11).

2. For any  $q_T \in D^{-s} F_T$ , the following relaxed observability inequality holds:

$$\|P_{F_0}^*(e^{-TL^*} q_T)\|_{s, D}^2 \leq M_{\text{obs}}^2 \int_0^T \|B^* e^{-(T-t)L^*} q_T\|_0^2 dt + M_{\text{rel}}^2 \|q_T\|_{s, D}^2. \quad (3.19)$$

Of course, for  $M_{\text{rel}} = 0$ , the inequality (3.18) should be understood as

$$\begin{cases} \frac{1}{M_{\text{obs}}^2} \|v\|_{L^2(0, T; U)}^2 \leq \|y_0\|_{-s, D}^2, \\ \|P_{F_T}(y(T))\|_{-s, D} = 0. \end{cases}$$

**Remark 3.2.3** Throughout this paper  $F_0$  will always be equal to the whole space  $E$ . However it is worth noticing that by specifying spaces  $F_0$  and  $F_T$ , one can recover usual inequalities related to different notions of controllability.

- When  $F_0 = F_T = E$ , inequality (3.19) is the usual relaxed inequality. If (3.19) holds with  $M_{\text{rel}} = 0$  then system (3.11) is null controllable. As explained above, we cannot always expect  $M_{\text{rel}}$  to be equal to zero when system (3.11) is discretized by finite differences method with a space domain of dimension greater than one. However, if this inequality holds with  $M_{\text{rel}}^2 = \varphi(h)$  and with  $M_{\text{obs}}$  independent of  $h$ , we recover the  $\varphi(h)$ -null-controllability notion briefly described above.
- When  $\dim(F_0) = 1$  and  $F_T = E$ , then proving (3.19) amounts to drive only one given initial condition to zero (when  $M_{\text{rel}} = 0$ ) or close to zero (when  $M_{\text{rel}}^2$  is small, like  $\varphi(h)$  for instance). This question is tackled for instance in [6].
- The partial null-controllability consists in driving to zero only some components of the solution of a system of parabolic PDEs. It amounts to prove inequality (3.19) for  $M_{\text{rel}} = 0$  and to choose an appropriate subspace  $F_T$ . This kind of controllability is studied for instance in [3] where related  $\varphi(h)$ -partial-null-controllability is also investigated.

## 3.3 Controllability of tensorized systems

### 3.3.1 Notations

Let  $(E_i, \langle \bullet, \bullet \rangle_{0, i})$ ,  $i = 1, 2$  two finite dimensional Euclidean spaces of dimensions  $N_1$  and  $N_2$ . The associated norms are denoted by  $\|\bullet\|_{0, i}$ . Let  $D_i$  be two positive definite self-adjoint operators in those spaces and for any  $s \in \mathbb{R}$ , we introduce the following scalar products

$$\langle u_i, v_i \rangle_{s, D_i} = \langle D_i^s u_i, v_i \rangle_{0, i}, \quad \forall u_i, v_i \in E_i,$$

and the associated norms  $\|\bullet\|_{s, D_i}$ .

In the case of a finite difference approximate system, the two spaces  $E_i$  have to be understood as the space of discrete in space (scalar) functions defined on a grid of  $\Omega_i$ . We will then consider the

tensor product space  $E_1 \otimes E_2$  as a natural discretization space for functions defined on the tensor product grid of  $\Omega$ . This space is equipped with the natural Euclidean structure defined by

$$\langle u_1 \otimes u_2, v_1 \otimes v_2 \rangle_0 = \langle u_1, v_1 \rangle_{0,1} \langle u_2, v_2 \rangle_{0,2}. \quad (3.20)$$

On  $E_1 \otimes E_2$  we consider the operator  $D = D_1 \otimes I + I \otimes D_2$ , where  $I$  stands for the identity operator in  $E_1$  and  $E_2$ . This is a positive definite self-adjoint operator, from which we can define the following natural inner products and associated norms

$$\langle u, v \rangle_{s,D} = \langle D^s u, v \rangle_0, \quad \forall u, v \in E_1 \otimes E_2$$

Note, in particular that we have

$$\langle u_1 \otimes u_2, v_1 \otimes v_2 \rangle_{1,D} = \langle u_1, v_1 \rangle_{1,D_1} \langle u_2, v_2 \rangle_{0,2} + \langle u_1, v_1 \rangle_{0,1} \langle u_2, v_2 \rangle_{1,D_2}, \quad \forall u_1, v_1 \in E_1, \forall u_2, v_2 \in E_2. \quad (3.21)$$

Since we will be interested in vector-valued discrete functions that aim at being approximations of the solution (at any time  $t$ ) of (3.6), we will naturally work with the space

$$E = E_1 \otimes E_2 \otimes \mathbb{R}^n,$$

which corresponds to the fact that all the  $n$  components of the system are approximated at each grid point.

We recall that if  $L_i$  is a linear operator in  $E_i$  and  $L$  a linear operator in  $\mathbb{R}^n$ , the tensor product  $L_1 \otimes L_2 \otimes L$  is a linear operator on  $E$  defined by

$$(L_1 \otimes L_2 \otimes L)(u_1 \otimes u_2 \otimes z) = L_1(u_1) \otimes L_2(u_2) \otimes L(z), \quad \forall u_i \in E_i, \forall z \in \mathbb{R}^n.$$

Let  $(\bullet, \bullet)$  be the Euclidean inner product in  $\mathbb{R}^n$  and  $|\bullet|$  the associated norm. The previous definitions are naturally extended to the space  $E$  as follows

$$\langle u \otimes y, v \otimes z \rangle_{s,D} = \langle u, v \rangle_{s,D} (y, z), \quad \forall u, v \in E_1 \otimes E_2, \quad \forall y, z \in \mathbb{R}^n,$$

and we do similar extensions for spaces  $E_1 \otimes \mathbb{R}^n$  and  $E_2 \otimes \mathbb{R}^n$ .

Recall that  $E$  can be identified to  $E_1 \otimes \mathbb{R}^n \otimes E_2$  and  $E_2 \otimes \mathbb{R}^n \otimes E_1$ . Therefore, if  $u := u_i \otimes y \in E_i \otimes \mathbb{R}^n$  and  $v_j \in E_j$  where  $i \neq j \in \{1, 2\}$ , one is allowed to consider  $u \otimes v_j$  as an element of  $E$ , although one should instead deal with the element  $u_i \otimes v_j \otimes y$ .

We now introduce two other finite dimensional Euclidean spaces  $(U_i, [\bullet, \bullet]_{0,i})$ ,  $i = 1, 2$ , with the associated norms  $[\bullet]_{0,i}$  that correspond to the control space for each subproblem.

We define  $U := U_1 \otimes U_2 \otimes \mathbb{R}^m$  and its inner product  $[\bullet, \bullet]_0$ , whose definition on  $U_1 \otimes U_2$  is analogous to (3.20) :

$$[u_1 \otimes u_2, v_1 \otimes v_2]_0 = [u_1, v_1]_{0,1} [u_2, v_2]_{0,2}, \quad (3.22)$$

and is extended to  $U$  as before.

### 3.3.2 Main Theorem

Let  $A_i$  be a symmetric definite positive operator in  $E_i$ . One can think of  $A_i$  as an approximation of the continuous operator  $\mathcal{A}_i$  but the statement of our result is generic and does not explicitly make use of such an assumption. The eigenvalues of  $A_i$  will be denoted by  $(\lambda_{i,k})_{k=1}^{N_i}$  and the corresponding eigenfunctions are  $(\phi_{i,k})_{k=1}^{N_i}$ . Those form an orthonormal family for  $(\bullet, \bullet)_{0,i}$ .

In our estimates, the following discrete Poincaré inequality for  $A_1$ , will be needed

$$\|u\|_0 = \|u\|_{0,A_1} \leq M_{P,1} \|u\|_{1,A_1}, \quad \forall u \in E_1, \quad (3.23)$$

for some  $M_{P,1}$ . Without loss of generality, we will assume that  $M_{P,1} \geq 1$ . Note that the best possible value for  $M_{P,1}$  is  $\lambda_{1,1}^{-1/2}$  but it is not sure that this value is greater than 1. Note also that the following generalized Poincaré estimate holds

$$\|u\|_0 \leq M_{P,1}^s \|u\|_{s,A_1}, \quad \forall u \in E_1, \forall s \geq 0.$$

For each  $i = 1, 2$ , we consider a control operator  $B_i : U_i \rightarrow E_i$ , a coupling matrix  $C$  and a control matrix  $B$ .

Our goal is to analyze the controllability properties (uniform with respect to any parameter on which the system may depend) of the following tensorized ODE

$$\begin{cases} \partial_t y + Ly = B_1 \otimes B_2 \otimes Bv, \\ y(0) = y_0 \in E, \end{cases} \quad (3.24)$$

where the control  $v$  belongs to  $L^2(0, T; U)$  and

$$L := A + I \otimes I \otimes C \text{ and } A := A_1 \otimes I \otimes I + I \otimes A_2 \otimes I.$$

Here, the symbol  $I$  stands for the identity operator on  $E_1$  or  $E_2$  and the symbol  $I$  is the identity operator of  $\mathbb{R}^n$ .

To this end, we will take benefit from the tensor product structure of the system and make two main assumptions:

1. the first one concerns the vector-valued sub-control-system in  $E_1$  defined by

$$\begin{cases} \partial_t y + A_1 \otimes I y + I \otimes C y = B_1 \otimes Bv, \\ y(0) = y_0 \in E_1 \otimes \mathbb{R}^n, \end{cases} \quad (3.25)$$

with  $y \in C^0([0, T], E_1 \otimes \mathbb{R}^n)$  and  $v \in L^2(0, T; U_1 \otimes \mathbb{R}^m)$ .

**Assumption 3.3.1** *For some  $M_{\text{obs},1} > 0$ ,  $\mu_{E_1} \in (0, +\infty]$ , we have that for any  $T > 0$ , for any  $y_0 \in E_1 \otimes \mathbb{R}^n$ , there exists a control  $v \in L^2(0, T; U_1 \otimes \mathbb{R}^m)$  for (3.25) such that*

$$\begin{cases} \|v\|_{L^2(0,T;U_1 \otimes \mathbb{R}^m)} \leq e^{M_{\text{obs},1}(1+\frac{1}{T})} \|y_0\|_{-s,A_1}, \\ \|y(T)\|_{-s,A_1} \leq e^{M_{\text{obs},1}(1+\frac{1}{T})} e^{-\mu_{E_1} T} \|y_0\|_{-s,A_1}. \end{cases}$$

2. the second one concerns a spectral property related with the pair of operators  $(A_2, B_2)$  similar to the well-known Lebeau-Robbiano inequality, excepted that it only holds for a certain portion of the spectrum of  $A_2$ .

**Assumption 3.3.2** *For some  $M_{\text{LR},2} > 0$ ,  $\mu_{E_2} \in (0, +\infty]$ , we have the following partial Lebeau-Robbiano spectral inequality*

$$\|\psi\|_0 \leq e^{M_{\text{LR},2}(1+\sqrt{\mu})} \|\mathbb{B}_2^* \psi\|_{0,2}, \quad \forall \psi \in \text{span}(\phi_{2,j}, \lambda_{2,j} \leq \mu), \quad (3.26)$$

for any  $0 < \mu < \mu_{E_2}$ .

The main theorem of this paper is the following.

**Theorem 3.3.1** *Assume that Assumptions 3.3.1 and 3.3.2 hold. Let  $\mu^* = \min(\mu_{E_1}, \mu_{E_2})$ .*

*There exist a  $M_{\text{obs}} > 0$  depending only on  $M_{\text{obs},1}$ ,  $M_{\text{LR},2}$ ,  $M_{s,\text{cont}}$ ,  $M_{s,\text{adm}}$  and  $\|C\|$  such that for any  $y_0 \in E$  and any  $T > 0$  there exists a control  $v \in L^2(0, T; U)$  satisfying*

$$\begin{cases} \|v\|_{L^2(0,T;U)} \leq e^{M_{\text{obs}}(1+\frac{1}{T}+T)} \|y_0\|_{-s,A}, \\ \|y(T)\|_{-s,A} \leq e^{M_{\text{obs}}(1+\frac{1}{T}+T)} e^{-\mu^* T/4} \|y_0\|_{-s,A}. \end{cases}$$

Note that in this theorem the constant  $M_{\text{obs},1}$  is built upon the Sobolev norms associated with  $D_1 = A_1$  and the constants  $M_{s,\text{cont}}$ ,  $M_{s,\text{adm}}$  with the Sobolev norms associated with  $D = A$ . Those norms are thus problem dependent.

In the case where  $A_i$  are discrete version of diffusion operators, it can be tempting to use instead usual discrete Sobolev norms, that are defined by using for  $D_1$  and  $D$  the discrete Laplace operators. However, the equivalence between those norms, for large values of  $s$ , is uniform with respect to the discretization parameter if and only if the diffusion coefficients  $\gamma_i$  are smooth enough. In the case of non smooth diffusion coefficients, using the norms given in the theorem is mandatory.

### 3.3.3 Proof of the main result

In this section, we will prove Theorem 3.3.1.

### Preliminary estimates

We start with some preliminary results.

**Lemma 3.3.1 (Norms comparison)** *For any  $s \geq 0$ , and any*

$$q = \sum_{j=1}^{N_2} q_j \otimes \phi_{2,j}, \quad q_j \in E_1 \otimes \mathbb{R}^n,$$

we have

$$\sum_{j=1}^{N_2} \|q_j\|_{s,A_1}^2 \leq \|q\|_{s,A}^2 \leq 2^s M_{P,1}^{2s} \sum_{j=1}^{N_2} (1 + \lambda_{2,j}^s) \|q_j\|_{s,A_1}^2.$$

### Proof

We write each  $q_j$  under the form

$$q_j = \sum_{k=1}^{N_1} \phi_{1,k} \otimes q_{k,j},$$

with  $q_{k,j} \in \mathbb{R}^n$  in such a way that

$$A^s q = \sum_{\substack{1 \leq k \leq N_1 \\ 1 \leq j \leq N_2}} (\lambda_{1,k} + \lambda_{2,j})^s \phi_{1,k} \otimes \phi_{2,j} \otimes q_{k,j}.$$

It follows that

$$\|q\|_{s,A}^2 = \sum_{\substack{1 \leq k \leq N_1 \\ 1 \leq j \leq N_2}} (\lambda_{1,k} + \lambda_{2,j})^s |q_{k,j}|^2,$$

and we immediately obtain

$$\|q\|_{s,A}^2 \geq \sum_{j=1}^{N_2} \left( \sum_{k=1}^{N_1} \lambda_{1,k}^s |q_{k,j}|^2 \right) = \sum_{j=1}^{N_2} \|q_j\|_{s,A_1}^2.$$

Moreover, we have

$$\begin{aligned} \|q\|_{s,A}^2 &\leq 2^s \sum_{\substack{1 \leq k \leq N_1 \\ 1 \leq j \leq N_2}} (\lambda_{1,k}^s + \lambda_{2,j}^s) |q_{k,j}|^2, \\ &= 2^s \sum_{j=1}^{N_2} \|q_j\|_{s,A_1}^2 + 2^s \sum_{j=1}^{N_2} \lambda_{2,j}^s \|q_j\|_{0,A_1}^2 \\ &\leq 2^s M_{P,1}^{2s} \sum_{j=1}^{N_2} (1 + \lambda_{2,j}^s) \|q_j\|_{s,A_1}^2, \end{aligned}$$

and the claim follows. ■

We define, for any  $\mu \geq 0$ , the following subspaces of  $E_2$

$$E_{\mu,2} := \text{span}\{\phi_{2,j} : \lambda_{2,j} \leq \mu\}, \text{ and } E_{\mu,2}^\perp := \text{span}\{\phi_{2,j} : \lambda_{2,j} > \mu\}$$

and the corresponding subspaces of  $E$

$$E_\mu := E_1 \otimes E_{\mu,2} \otimes \mathbb{R}^n, \text{ and } E_\mu^\perp := E_1 \otimes E_{\mu,2}^\perp \otimes \mathbb{R}^n.$$

Let  $P_\mu$  (resp.  $P_\mu^\perp$ ) be the orthogonal projection in  $E$  onto  $E_\mu$  (resp.  $E_\mu^\perp$ ) for the inner product  $\langle \bullet, \bullet \rangle_{-s,A}$ . By construction we have  $P_\mu + P_\mu^\perp = I$ .

Let us now state some properties of the uncontrolled system

$$\begin{cases} \partial_t y + Ly = 0, \\ y(0) = y_0 \in E. \end{cases} \quad (3.27)$$



**Proposition 3.3.1 (Dissipation estimates)**

1. For any  $\mu \geq 0$ , if  $y_0 \in E_\mu^\perp$ , then the unique solution to (3.27) satisfies  $y(t) \in E_\mu^\perp$  for any  $t$  and moreover, for any  $s \in \mathbb{R}$ , we have

$$\|y(t)\|_{-s,A} \leq e^{\|C\|t} e^{-t\mu} \|y_0\|_{-s,A}, \quad \forall t \geq 0.$$

2. For any  $\mu \geq 0$ , and any  $y_0 \in E$ , the unique solution to (3.27) satisfies, for any  $s \in \mathbb{R}$ ,

$$\begin{aligned} \|P_\mu y(t)\|_{-s,A} &\leq e^{\|C\|t} \|P_\mu y_0\|_{-s,A}, \\ \|P_\mu^\perp y(t)\|_{-s,A} &\leq e^{\|C\|t - \mu t} \|P_\mu^\perp y_0\|_{-s,A} \leq e^{\|C\|t - \mu t} \|y_0\|_{-s,A}. \end{aligned}$$

**Proof**

1. The first point is just a consequence of the fact that  $E_\mu$  and  $E_\mu^\perp$  are stable by  $L = A + I \otimes I \otimes C$ . A straightforward computation, using that  $A$  commutes with  $I \otimes I \otimes C$  shows that

$$\frac{1}{2} \partial_t \langle A^{-s} y, y \rangle_0 + \langle A A^{-\frac{s}{2}} y, A^{-\frac{s}{2}} y \rangle_0 = - \langle A^{-s} (I \otimes I \otimes C) y, y \rangle_0$$

Then we observe that,  $A^{-\frac{s}{2}} y(t) \in E_\mu^\perp$  for any  $t$ , and that by definition of  $E_\mu^\perp$  we have

$$\langle A z, z \rangle_0 \geq \mu \langle z, z \rangle_0, \quad \forall z \in E_\mu^\perp.$$

It follows that

$$\frac{1}{2} \partial_t \langle A^{-s} y, y \rangle_0 \leq (\|C\| - \mu) \langle A^{-s} y, y \rangle_0,$$

and the claim follows by the differential form of Gronwall's Lemma.

2. We simply observe that  $P_\mu y$  and  $P_\mu^\perp y$  solve the same equation as  $y$  with initial conditions  $P_\mu y_0 \in E_0^\perp$ ,  $P_\mu^\perp y_0 \in E_\mu^\perp$ . It is then enough to use the inequality of the first point to conclude.

■

**Partial controllability results**

The following proposition is a partial version of Theorem 3.3.1. More precisely, we establish the existence of a control  $v$  that reduces the norm of the projection on  $E_\mu$  of the final state  $y(T)$  as much as possible. However, the bound on the control is not uniform with respect to  $\mu$  yet.

**Proposition 3.3.2** *There exists a positive number  $M_{\text{part}}$  which depends only on  $s$ ,  $M_{P,1}$ ,  $M_{\text{obs},1}$ ,  $M_{\text{LR},2}$ ,  $M_{s,\text{cont}}$ ,  $M_{s,\text{adm}}$  such that for any  $\mu \in (0, \mu_{E_2})$  and any  $y_0 \in E$ , there exists a control  $v \in L^2(0, T; U)$  that satisfies*

$$\begin{cases} \|v\|_{L^2(0,T;U)} \leq e^{M_{\text{part}}(1+\frac{1}{T}+\sqrt{\mu})} \|y_0\|_{-s,A} \\ \|P_\mu(y(T))\|_{-s,A} \leq e^{M_{\text{part}}(1+\frac{1}{T})} e^{-\mu E_1 T} \|y_0\|_{-s,A}, \end{cases} \quad (3.28)$$

and

$$\|y(T)\|_{-s,A} \leq e^{M_{\text{part}}(1+\frac{1}{T}+\sqrt{\mu})} \|y_0\|_{-s,A}. \quad (3.29)$$

**Proof**

Let  $q_T \in E_\mu$  and  $q$  be the solution of the backward equation

$$\begin{cases} -\partial_t q + L^* q = 0 \\ q(T) = q_T. \end{cases}$$

We recall that  $A_1$  (resp.  $A_2$ ) is self-adjoint in  $E_1$  (resp.  $E_2$ ) so that

$$L^* = A_1 \otimes I \otimes I + I \otimes A_2 \otimes I + I \otimes I \otimes C^*.$$

Since  $E_\mu$  is stable by  $L^*$ , we can decompose  $q$  in the following way

$$q(t) = \sum_{\lambda_{2,j} \leq \mu} q_j(t) \otimes \phi_{2,j}, \text{ with } q_j(t) \in E_1 \otimes \mathbb{R}^n$$

where  $q_j$  satisfies

$$\begin{cases} -\partial_t q_j + (A_1 \otimes I)q_j + \lambda_{2,j}q_j + (I \otimes C^*)q_j = 0 \\ q_j(T) = q_{T,j} \in E_1 \otimes \mathbb{R}^n. \end{cases}$$

The new variable  $z_j(t) := q_j(t)e^{\lambda_{2,j}(T-t)}$  satisfies the following system

$$\begin{cases} -\partial_t z_j + (A_1 \otimes I)z_j + (I \otimes C^*)z_j(t) = 0 \\ z_j(T) = q_{T,j}. \end{cases}$$

Thanks to Assumption 3.3.1, we can apply Lemma 3.2.1 in the space  $E_1 \otimes \mathbb{R}^n$ , with  $F_0 = F_T = E_1 \otimes \mathbb{R}^n$ ,  $L = A_1 \otimes I + I \otimes C$ ,  $B = B_1 \otimes B$ ,  $D = A_1 \otimes I$ ,  $M_{\text{obs}} = e^{M_{\text{obs},1}(1+\frac{1}{T})}$  and  $M_{\text{rel}} = e^{M_{\text{obs},1}(1+\frac{1}{T})}e^{-\mu_{E_1}T}$ . Inequality (3.19) with  $q_T = q_{T,j}$  gives

$$\|z_j(0)\|_{s,A_1}^2 \leq e^{2M_{\text{obs},1}(1+\frac{1}{T})} \int_0^T \llbracket B_1^* \otimes B^* z_j(t) \rrbracket_{0,1}^2 dt + e^{2M_{\text{obs},1}(1+\frac{1}{T})} e^{-\mu_{E_1}2T} \|q_{T,j}\|_{s,A_1}^2,$$

and therefore, coming back to the variable  $q_j$ , we get

$$\|q_j(0)\|_{s,A_1}^2 e^{2\lambda_{2,j}T} \leq e^{2M_{\text{obs},1}(1+\frac{1}{T})} \int_0^T \llbracket B_1^* \otimes B^* q_j(t) \rrbracket_{0,1}^2 e^{2\lambda_{2,j}(T-t)} dt + e^{2M_{\text{obs},1}(1+\frac{1}{T})} e^{-\mu_{E_1}2T} \|q_{T,j}\|_{s,A_1}^2,$$

and thus,

$$\|q_j(0)\|_{s,A_1}^2 \leq e^{2M_{\text{obs},1}(1+\frac{1}{T})} \int_0^T \llbracket B_1^* \otimes B^* q_j(t) \rrbracket_{0,1}^2 dt + e^{2M_{\text{obs},1}(1+\frac{1}{T})} e^{-\mu_{E_1}2T} e^{-2\lambda_{2,j}T} \|q_{T,j}\|_{s,A_1}^2.$$

Using the second inequality in Lemma 3.3.1, we get

$$\|q(0)\|_{s,A}^2 \leq 2^s M_{P,1}^{2s} e^{2M_{\text{obs},1}(1+\frac{1}{T})} \sum_{\lambda_{2,j} \leq \mu} \left( (1 + \lambda_{2,j}^s) \int_0^T \llbracket B_1^* \otimes B^* q_j(t) \rrbracket_{0,1}^2 dt + (1 + \lambda_{2,j}^s) e^{-2\lambda_{2,j}T} e^{-\mu_{E_1}2T} \|q_{T,j}\|_{s,A_1}^2 \right).$$

Applying now the first inequality in Lemma 3.3.1 and the following inequality  $\lambda^s e^{-2\lambda T} \leq \frac{s^s}{(2T)^s}$ , we get

$$\|q(0)\|_{s,A}^2 \leq 2^s M_{P,1}^{2s} (1 + \mu^s) e^{2M_{\text{obs},1}(1+\frac{1}{T})} \sum_{\lambda_{2,j} \leq \mu} \int_0^T \llbracket B_1^* \otimes B^* q_j(t) \rrbracket_{0,1}^2 dt + 2^s M_{P,1}^{2s} e^{2M_{\text{obs},1}(1+\frac{1}{T})} \left( 1 + \frac{s^s}{(2T)^s} \right) e^{-\mu_{E_1}2T} \|q_T\|_{s,A}^2.$$

We shall now apply Assumption 3.3.2. To this end we choose any orthonormal basis  $(\Psi_k)_{k=1,\dots,K}$  of  $U_1 \otimes \mathbb{R}^n$  and we decompose each  $B_1^* \otimes B^* q_j(t)$  in this basis, the coefficients being denoted by  $a_{k,j}(t)$ . It follows

$$\sum_{\lambda_{2,j} \leq \mu} \llbracket B_1^* \otimes B^* q_j(t) \rrbracket_{0,1}^2 = \sum_{\lambda_{2,j} \leq \mu} \left\| \sum_{k=1}^K a_{k,j}(t) \Psi_k \right\|_{0,1}^2 = \sum_{k=1}^K \sum_{\lambda_{2,j} \leq \mu} (a_{k,j}(t))^2.$$

For any  $k \in \{1, \dots, K\}$ , given that  $0 < \mu < \mu_{E_2}$ , we can apply the discrete Lebeau-Robbiano inequality given by Assumption 3.3.2 to the vector  $\psi = \sum_{\lambda_{2,j} \leq \mu} a_{k,j}(t) \phi_{2,j}$ , to obtain

$$\begin{aligned} \sum_{\lambda_{2,j} \leq \mu} \llbracket \mathbf{B}_1^* \otimes \mathbf{B}^* q_j(t) \rrbracket_{0,1}^2 &\leq \sum_{k=1}^K e^{2M_{\text{LR},2}(1+\sqrt{\mu})} \left\| \sum_{\lambda_{2,j} \leq \mu} a_{k,j}(t) \mathbf{B}_2^* \phi_{2,j} \right\|_{0,2}^2 \\ &\leq e^{2M_{\text{LR},2}(1+\sqrt{\mu})} \sum_{k=1}^K \sum_{\lambda_{2,j} \leq \mu} \sum_{\lambda_{2,j'} \leq \mu} [a_{k,j}(t) \mathbf{B}_2^* \phi_{2,j}, a_{k,j'}(t) \mathbf{B}_2^* \phi_{2,j'}]_{0,2}. \end{aligned}$$

Apply now (3.22) and the fact that for any  $k$ ,  $\llbracket \Psi_k \rrbracket_{0,1} = 1$ :

$$\begin{aligned} &\sum_{\lambda_{2,j} \leq \mu} \llbracket \mathbf{B}_1^* \otimes \mathbf{B}^* q_j(t) \rrbracket_{0,1}^2 \\ &\leq e^{2M_{\text{LR},2}(1+\sqrt{\mu})} \sum_{k=1}^K \sum_{\lambda_{2,j} \leq \mu} \sum_{\lambda_{2,j'} \leq \mu} [a_{k,j}(t) \Psi_k \otimes \mathbf{B}_2^* \phi_{2,j}, a_{k,j'}(t) \Psi_k \otimes \mathbf{B}_2^* \phi_{2,j'}]_0 \\ &\leq e^{2M_{\text{LR},2}(1+\sqrt{\mu})} \sum_{\lambda_{2,j} \leq \mu} \sum_{\lambda_{2,j'} \leq \mu} [\mathbf{B}_1^* \otimes \mathbf{B}^* \otimes \mathbf{B}_2^* (q_j(t) \otimes \phi_{2,j}), \mathbf{B}_1^* \otimes \mathbf{B}^* \otimes \mathbf{B}_2^* (q_{j'}(t) \otimes \phi_{2,j'})]_0 \\ &\leq e^{2M_{\text{LR},2}(1+\sqrt{\mu})} \llbracket \mathbf{B}_1^* \otimes \mathbf{B}_2^* \otimes \mathbf{B}^* q(t) \rrbracket_0^2. \end{aligned}$$

Hence,

$$\begin{aligned} \|q(0)\|_{s,A}^2 &\leq 2^s M_{\text{P},1}^{2s} (1 + \mu^s) e^{2M_{\text{LR},2}(1+\sqrt{\mu}) + 2M_{\text{obs},1}(1+\frac{1}{T})} \int_0^T \llbracket \mathbf{B}_1^* \otimes \mathbf{B}_2^* \otimes \mathbf{B}^* q(t) \rrbracket_0^2 dt \\ &\quad + 2^s M_{\text{P},1}^{2s} e^{2M_{\text{obs},1}(1+\frac{1}{T})} \left(1 + \frac{s^s}{T^s}\right) e^{-\mu_{E_1} 2T} \|q_T\|_{s,A}^2. \end{aligned} \quad (3.30)$$

Note that, by construction,  $E_\mu$  is stable by A so that (3.30) is valid for any  $q_T \in F_T = D^{-s} E_\mu$ . Thus, we can apply Lemma 3.2.1 in the space  $E$  with  $D = A$ ,  $F_0 = E$  and  $F_T = E_\mu$ . It follows that there exists a control  $v \in L^2(0, T; U)$  such that

$$\begin{cases} \llbracket v \rrbracket_{L^2(0,T;U)} \leq 2^{\frac{s}{2}} M_{\text{P},1}^s \sqrt{1 + \mu^s} e^{M_{\text{LR},2}(1+\sqrt{\mu}) + M_{\text{obs},1}(1+\frac{1}{T})} \|y_0\|_{-s,A}, \\ \|\mathbf{P}_\mu(y(T))\|_{-s,A} \leq 2^{\frac{s}{2}} M_{\text{P},1}^s e^{M_{\text{obs},1}(1+\frac{1}{T})} \sqrt{1 + \frac{s^s}{T^s}} e^{-\mu_{E_1} T} \|y_0\|_{-s,A}. \end{cases}$$

Finally, inequality (3.29) comes from Proposition 3.2.1 and the estimate on the control cost just proved which give

$$\|y(T)\|_{-s,A} \leq \left( M_{s,\text{cont}} + M_{s,\text{adm}} 2^{\frac{s}{2}} M_{\text{P},1}^s \sqrt{1 + \mu^s} e^{M_{\text{LR},2}(1+\sqrt{\mu}) + M_{\text{obs},1}(1+\frac{1}{T})} \right) \|y_0\|_{-s,A}.$$

Those estimates can easily be put into the expected form for a suitable choice of  $M_{\text{part}}$ . ■

The following corollary contains the main idea of Lebeau and Robbiano's strategy: during the first half of a given time interval, we control the lowest frequencies then, during the second half of the time interval, we turn the control to zero to take advantage of the natural dissipation of the problem.

**Corollary 3.3.1** *There exists a positive number  $M_{\text{LR}}$  which depends only on  $M_{\text{obs},1}$ ,  $M_{\text{LR},2}$ ,  $M_{s,\text{cont}}$ ,  $M_{s,\text{adm}}$  and  $\|C\|$  such that for any  $\tau \in (0, T)$ ,  $\mu \in (0, \mu_{E_2})$  and any  $y_0 \in E$ , there exists a control  $v \in L^2(0, \tau; U)$  that satisfies*

$$\begin{cases} \llbracket v \rrbracket_{L^2(0,\tau;U)} \leq e^{M_{\text{LR}}(1+\frac{1}{\tau}+\sqrt{\mu})} \|y_0\|_{-s,A} \\ \|y(\tau)\|_{-s,A} \leq e^{M_{\text{LR}}(1+\frac{1}{\tau}+\tau)} \left( e^{-\mu_{E_1} \tau/2} + e^{-\mu \tau/2 + M_{\text{LR}} \sqrt{\mu}} \right) \|y_0\|_{-s,A}. \end{cases}$$

**Proof**

We apply Proposition 3.3.2 on the time interval  $(0, \tau/2)$ . We get a control  $v$  and a solution  $y$  which satisfy (3.28) (with  $T$  replaced by  $\tau/2$ ). Then, on the interval  $(\tau/2, \tau)$ , we set the control to zero. Thus, we have constructed the following control

$$\bar{v}(t) := \begin{cases} v(t) & \text{for } t \in (0, \tau/2) \\ 0 & \text{for } t \in (\tau/2, \tau), \end{cases}$$

and the associated solution of the system is still denoted by  $y$ . Clearly, the  $L^2$  norm of  $\bar{v}$  on  $(0, \tau)$  is equal to that of  $v$  on  $(0, \tau/2)$ .

1. Since the control is 0 on  $(\tau/2, \tau)$ , the dissipation properties given in Proposition 3.3.1 give

$$\|P_\mu(y(\tau))\|_{-s, A} \leq e^{\|C\| \frac{\tau}{2}} \|P_\mu(y(\tau/2))\|_{-s, A}.$$

Then, we use (3.28) to get

$$\|P_\mu(y(\tau))\|_{-s, A} \leq e^{2M_{\text{part}}(1+\frac{1}{\tau})} e^{\|C\| \frac{\tau}{2}} e^{-\mu E_1 \frac{\tau}{2}} \|y_0\|_{-s, A}. \quad (3.31)$$

2. By the dissipation properties given in Proposition 3.3.1 we get

$$\|P_\mu^\perp(y(\tau))\|_{-s, A} \leq e^{(\|C\| - \mu) \frac{\tau}{2}} \|y(\tau/2)\|_{-s, A}.$$

Then inequality (3.29) of Proposition 3.3.2 (with still  $T$  replaced by  $\tau/2$ ) leads to

$$\|P_\mu^\perp(y(\tau))\|_{-s, A} \leq e^{2M_{\text{part}}(1+\frac{1}{\tau}+\sqrt{\mu})} e^{(\|C\| - \mu) \frac{\tau}{2}} \|y_0\|_{-s, A}. \quad (3.32)$$

We combine (3.31) and (3.32) to get the result for a suitable value of  $M_{\text{LR}}$  depending on  $M_{\text{part}}$ . ■ Observe that Corollary 3.3.1 is slightly different from the similar result in the classical Lebeau and Robbiano's strategy. Indeed, the modes corresponding to frequencies less than  $\mu$  of the final state  $y(\tau)$  are not cancelled; they still exist but are controlled by the small term  $e^{-\mu E_1 \frac{\tau}{2}}$ . When applying the complete strategy on the interval  $(0, T)$ , one has to make sure that the term  $e^{-\mu E_1 T/2}$  is smaller than  $e^{-\mu \tau/2}$ . This constraint is fulfilled in Theorem 3.3.1 by dealing with frequencies  $\mu$  smaller than  $\mu_{E_1}$ .

### Conclusion of the proof of the main theorem

We can now apply Lebeau-Robbiano's strategy, with a well chosen finite number of steps, and prove Theorem 3.3.1.

#### Proof

Without loss of generality we suppose that  $M_{\text{LR}} \geq \ln(2)$ . Let  $\alpha > 0$  such that

$$\frac{\alpha}{8M_{\text{LR}}} - 3\frac{\sqrt{\alpha}}{2} - 5 > 0. \quad (3.33)$$

Note that  $\alpha$  only depends on  $M_{\text{LR}}$ . Let  $\mu_j = \frac{\alpha}{T^2} (2^j)^2$ ,  $\tau_j = \frac{T}{2} \frac{1}{2^j}$  and  $T_j = \sum_{k=1}^j \tau_k$ .

Suppose that  $\mu^* \leq \mu_1$ . In this case we apply Corollary 3.3.1 with  $\mu = \mu^*$  on the whole time interval  $(0, T)$ . There exists a control  $v \in L^2(0, T, U)$  that satisfies

$$\begin{cases} \|v\|_{L^2(0, T; U)} \leq e^{M_{\text{LR}}(1+\frac{1}{T}+\sqrt{\mu^*})} \|y_0\|_{-s, A}, \\ \|y(T)\|_{-s, A} \leq e^{M_{\text{LR}}(1+\frac{1}{T}+T)} \left( e^{-\mu_{E_1} T/2} + e^{-\mu^* T/2 + M_{\text{LR}} \sqrt{\mu^*}} \right) \|y_0\|_{-s, A}. \end{cases}$$

Since  $e^{-\mu_{E_1} T/2} \leq e^{-\mu^* T/2}$  we have

$$\|y(T)\|_{-s, A} \leq e^{M_{\text{LR}}(1+\frac{1}{T}+T)} e^{-\mu^* T/2} \left( 1 + e^{M_{\text{LR}} \sqrt{\mu^*}} \right) \|y_0\|_{-s, A},$$

and moreover  $e^{M_{\text{LR}} \sqrt{\mu^*}} \leq e^{M_{\text{LR}} \sqrt{\mu_1}} \leq e^{\frac{2}{T} M_{\text{LR}} \sqrt{\alpha}}$ . Therefore,

$$\|y(T)\|_{-s, A} \leq e^{M_{\text{LR}}(1+2\sqrt{\alpha})(1+\frac{1}{T}+T)} e^{-\mu^* T/2} \|y_0\|_{-s, A}.$$

Thus, if  $\mu^* \leq \mu_1$ , the proof of Theorem 3.3.1 is complete since  $\alpha$  depends only on  $M_{\text{LR}}$ , which itself depends only on  $M_{\text{obs}, 1}$ ,  $M_{\text{LR}, 2}$ ,  $M_{s, \text{cont}}$ ,  $M_{s, \text{adm}}$  and  $\|C\|$ .

Suppose now that  $\mu_1 < \mu^*$  and define  $j^* \in \mathbb{N}^*$  such that  $\mu_{j^*} < \mu^* \leq \mu_{j^*+1}$ .

- During the time interval  $(0, T_1) = (0, \tau_1)$ , we apply a control  $v_1$  as given by Corollary 3.3.1 with  $\mu = \mu_1$  (which applies here since  $\mu_1 < \mu^* \leq \mu_{E_2}$ ) and we get

$$\begin{cases} \|v_1\|_{L^2(0, T_1; U)} \leq e^{M_{\text{LR}}(1 + \frac{1}{\tau_1} + \sqrt{\mu_1})} \|y_0\|_{-s, A}, \\ \|y(T_1)\|_{-s, A} \leq e^{M_{\text{LR}}(1 + \frac{1}{\tau_1} + \tau_1)} \left( e^{-\mu_{E_1} \tau_1 / 2} + e^{-\mu_1 \tau_1 / 2 + M_{\text{LR}} \sqrt{\mu_1}} \right) \|y_0\|_{-s, A}. \end{cases}$$

- For any index  $j \leq j^*$ , we continue this procedure by applying Corollary 3.3.1 on time interval  $(T_{j-1}, T_j) = (T_{j-1}, T_{j-1} + \tau_j)$  with  $\mu = \mu_j$ . We get a control  $v_j$  that satisfies

$$\begin{cases} \|v_j\|_{L^2(T_{j-1}, T_j; U)} \leq e^{M_{\text{LR}}(1 + \frac{1}{\tau_j} + \sqrt{\mu_j})} \|y(T_{j-1})\|_{-s, A}, \\ \|y(T_j)\|_{-s, A} \leq e^{M_{\text{LR}}(1 + \frac{1}{\tau_j} + \tau_j)} \times \\ \quad \left( e^{-\mu_{E_1} \tau_j / 2} + e^{-\mu_j \tau_j / 2 + M_{\text{LR}} \sqrt{\mu_j}} \right) \|y(T_{j-1})\|_{-s, A}. \end{cases} \quad (3.34)$$

Let us focus on the estimation of the term  $y(T_j)$ . First, we use that  $e^{-\mu_{E_1} \tau_j / 2} \leq e^{-\mu_j \tau_j / 2}$  and the relations

$$\begin{cases} \mu_j \tau_j = \frac{\alpha}{2T} 2^j = \frac{\alpha}{4} \frac{1}{\tau_j}, \\ \sqrt{\mu_j} = \frac{\sqrt{\alpha}}{2\tau_j}. \end{cases} \quad (3.35)$$

It follows

$$e^{-\mu_j \tau_j / 2 + M_{\text{LR}} \sqrt{\mu_j}} \leq e^{\frac{1}{\tau_j} \left( -\frac{\alpha}{8} + M_{\text{LR}} \frac{\sqrt{\alpha}}{2} \right)}.$$

Hence

$$\|y(T_j)\|_{-s, A} \leq 2e^{M_{\text{LR}}} e^{M_{\text{LR}}(\frac{1}{\tau_j} + \tau_j) + \frac{1}{\tau_j} \left( -\frac{\alpha}{8} + M_{\text{LR}} \frac{\sqrt{\alpha}}{2} \right)} \|y(T_{j-1})\|_{-s, A}, \quad (3.36)$$

since  $\ln(2) \leq M_{\text{LR}}$ , and  $1 \leq \frac{1}{\tau_j} + \tau_j$  leads to

$$2e^{M_{\text{LR}}} e^{M_{\text{LR}}(\frac{1}{\tau_j} + \tau_j)} \leq e^{2M_{\text{LR}}} e^{M_{\text{LR}}(\frac{1}{\tau_j} + \tau_j)} \leq e^{3M_{\text{LR}}(\frac{1}{\tau_j} + \tau_j)},$$

which, associated with (3.36), finally yields

$$\|y(T_j)\|_{-s, A} \leq e^{3M_{\text{LR}}(\frac{1}{\tau_j} + \tau_j) + \frac{1}{\tau_j} \left( -\frac{\alpha}{8} + M_{\text{LR}} \frac{\sqrt{\alpha}}{2} \right)} \|y(T_{j-1})\|_{-s, A}.$$

Now we set  $\beta := \frac{\alpha}{8M_{\text{LR}}} - \frac{\sqrt{\alpha}}{2} - 3$ , in such a way that the previous estimate reads

$$\|y(T_j)\|_{-s, A} \leq e^{3M_{\text{LR}} \tau_j} e^{-M_{\text{LR}} \frac{\beta}{\tau_j}} \|y(T_{j-1})\|_{-s, A}.$$

According to (3.33), we have  $\beta > 0$  since  $\frac{\alpha}{8M_{\text{LR}}} - (\frac{\sqrt{\alpha}}{2} + 3) > \frac{\alpha}{8M_{\text{LR}}} - (\frac{3\sqrt{\alpha}}{2} + 5) > 0$ .

Thus, we can estimate  $y(T_j)$  by using the previous upper-bound recursively,

$$\begin{aligned} \|y(T_j)\|_{-s, A} &\leq \left( e^{3M_{\text{LR}} \sum_{k=1}^j \tau_k} \right) \left( e^{-\beta M_{\text{LR}} \sum_{k=1}^j \frac{1}{\tau_k}} \right) \|y_0\|_{-s, A} \\ &\leq e^{3M_{\text{LR}} T_j} e^{-\frac{2}{T} \beta M_{\text{LR}} \sum_{k=1}^j 2^k} \|y_0\|_{-s, A} \\ &\leq e^{3M_{\text{LR}} T_j} e^{-\frac{2\beta M_{\text{LR}} 2^j}{T}} \|y_0\|_{-s, A}. \end{aligned}$$

Therefore,

$$\|y(T_j)\|_{-s, A} \leq e^{3M_{\text{LR}} T_j} e^{-\frac{\beta M_{\text{LR}}}{\tau_j}} \|y_0\|_{-s, A}, \quad (3.37)$$

and with (3.35),

$$\|y(T_j)\|_{-s, A} \leq e^{3M_{\text{LR}} T_j} e^{-2 \frac{\beta M_{\text{LR}}}{\sqrt{\alpha}} \sqrt{\mu_j}} \|y_0\|_{-s, A}.$$

When  $j = j^*$ , we end up with (recall that  $\mu^* < \mu_{j^*+1} = 4\mu_{j^*}$ )

$$\|y(T_{j^*})\|_{-s, A} \leq e^{3M_{\text{LR}} T_{j^*}} e^{-\frac{\beta M_{\text{LR}}}{\sqrt{\alpha}} \sqrt{\mu^*}} \|y_0\|_{-s, A}. \quad (3.38)$$

Let us now estimate the control  $v$ . Taking back (3.34) for  $j \geq 2$ , combined with (3.37) and (3.35),

$$\llbracket v_j \rrbracket_{L^2(T_{j-1}, T_j; U)} \leq e^{M_{\text{LR}} \left[ 1 + \frac{1}{\tau_j} \left( 1 + \frac{\sqrt{\alpha}}{2} - \frac{\beta}{2} \right) + 3T_{j-1} \right]} \|y_0\|_{-s, A}.$$

Note that, (3.33) gives that  $\bar{\beta} := \beta/2 - 1 - \frac{\sqrt{\alpha}}{2} = \frac{1}{2}(\beta - 2 - \sqrt{\alpha}) = \frac{1}{2} \left( \frac{\alpha}{8M_{\text{LR}}} - 3\frac{\sqrt{\alpha}}{2} - 5 \right) > 0$ .

This implies that

$$\begin{aligned} \llbracket v \rrbracket_{L^2(T_1, T_{j^*}; U)}^2 &= \sum_{j=2}^{j^*} \llbracket v_j \rrbracket_{L^2(T_{j-1}, T_j; U)}^2 \\ &\leq e^{2M_{\text{LR}}(1+3T)} \sum_{j \geq 0} e^{-\frac{M_{\text{LR}}\bar{\beta}}{T} 2^{j+2}} \|y_0\|_{-s, A}^2 \\ &\leq \left( 1 + \frac{T}{M_{\text{LR}}\bar{\beta}} \right) e^{2M_{\text{LR}}(1+3T)} \|y_0\|_{-s, A}^2, \end{aligned}$$

and the last inequality came from  $\frac{1}{1-e^{-x}} \leq 1 + \frac{1}{x}$ , for  $x > 0$ .

Using that

$$\llbracket v_1 \rrbracket_{L^2(0, T_1; U)}^2 \leq e^{2M_{\text{LR}}(1 + \frac{1}{\tau_1} + \tau_1 + \sqrt{\mu_1})} \|y_0\|_{-s, A}^2,$$

we have built a control  $v$  on  $(0, T_{j^*})$  that satisfies

$$\llbracket v \rrbracket_{L^2(0, T_{j^*}; U)} \leq e^{M_{\text{obs}}(1 + \frac{1}{T} + T)} \|y_0\|_{-s, A}, \quad (3.39)$$

with  $M_{\text{obs}}$  depending on  $M_{\text{LR}}$  but not on  $T$ .

Notice that  $T_{j^*} = \sum_{j=1}^{j^*} \tau_j = \frac{T}{2} \sum_{j=1}^{j^*} \frac{1}{2^j} \leq \frac{T}{2}$ . The last step of the proof consists in applying Corollary 3.3.1 on the interval  $(T_{j^*}, T)$  to recover the estimates of Theorem 3.3.1.

Corollary 3.3.1 with  $\mu = \mu^*$  gives now a control  $v \in L^2(T_{j^*}, T)$  such that

$$\begin{cases} \llbracket v \rrbracket_{L^2(T_{j^*}, T; U)} \leq e^{M_{\text{LR}}(1 + \frac{1}{T-T_{j^*}} + T - T_{j^*} + \sqrt{\mu^*})} \|y(T_{j^*})\|_{-s, A} \\ \|y(T)\|_{-s, A} \leq e^{M_{\text{LR}} \left( 1 + \frac{1}{T-T_{j^*}} + T - T_{j^*} \right)} \times \\ \quad \left( e^{-\mu_{E_1}(T-T_{j^*})/2} + e^{-\mu^*(T-T_{j^*})/2 + M_{\text{LR}}\sqrt{\mu^*}} \right) \|y(T_{j^*})\|_{-s, A}. \end{cases} \quad (3.40)$$

First, with (3.38), we have

$$\llbracket v \rrbracket_{L^2(T_{j^*}, T; U)} \leq e^{M_{\text{LR}}(1 + \frac{2}{T} + T - T_{j^*} + \sqrt{\mu^*})} e^{3M_{\text{LR}}T_{j^*}} e^{-\frac{\beta M_{\text{LR}}}{\sqrt{\alpha}} \sqrt{\mu^*}} \|y_0\|_{-s, A},$$

and by (3.33), we can check that

$$\beta > \sqrt{\alpha}. \quad (3.41)$$

Hence

$$\llbracket v \rrbracket_{L^2(T_{j^*}, T; U)} \leq e^{2M_{\text{LR}}(1 + \frac{1}{T} + T)} \|y_0\|_{-s, A}.$$

Finally this estimation and (3.39) give the first inequality of Theorem 3.3.1.

According to (3.40) and (3.38),

$$\begin{aligned} \|y(T)\|_{-s, A} &\leq e^{M_{\text{LR}} \left( 1 + \frac{1}{T-T_{j^*}} + T - T_{j^*} \right)} \left( e^{-\mu_{E_1}(T-T_{j^*})/2} + e^{-\mu^*(T-T_{j^*})/2 + M_{\text{LR}}\sqrt{\mu^*}} \right) \times \\ &\quad e^{3M_{\text{LR}}T_{j^*}} e^{-\frac{\beta M_{\text{LR}}}{\sqrt{\alpha}} \sqrt{\mu^*}} \|y_0\|_{-s, A} \\ &\leq e^{2M_{\text{LR}}(1 + \frac{1}{T} + T)} \left( e^{-\mu_{E_1}T/4} + e^{-\mu^*T/4 + M_{\text{LR}}\sqrt{\mu^*}} \right) e^{-\frac{\beta M_{\text{LR}}}{\sqrt{\alpha}} \sqrt{\mu^*}} \|y_0\|_{-s, A} \\ &\leq 2e^{2M_{\text{LR}}(1 + \frac{1}{T} + T)} e^{-\mu^*T/4} e^{M_{\text{LR}}\sqrt{\mu^*}} e^{-\frac{\beta M_{\text{LR}}}{\sqrt{\alpha}} \sqrt{\mu^*}} \|y_0\|_{-s, A}. \end{aligned}$$

Using (3.41), this reduces to

$$\|y(T)\|_{-s, A} \leq 2e^{2M_{\text{LR}}(1 + \frac{1}{T} + T)} e^{-\mu^*T/4} \|y_0\|_{-s, A},$$

and this concludes the proof. ■

## 3.4 Applications

### 3.4.1 Dirichlet boundary null control of a semi-discrete cascade system on a rectangle

In this section, we will apply the general framework introduced above to prove the  $\phi(h)$  null controllability properties for semi-discrete versions of the boundary control problem of the coupled system (3.6) in the case of a cascade form (that is for particular B and C given below). The penalization term  $\phi(h)$  will be exponentially small in  $h$ , just like in similar results obtained in the literature and quoted in the introduction section 3.1.1.

The results are valid for the finite difference discretization in space of the system in any dimension. However, for the simplicity of the presentation we will only state and prove our theorem in dimension  $d = 2$ . Observe that other usual techniques (based on Carleman estimates, or on moments methods as in [1, 7-9]) do not directly apply in this setting since we are considering boundary controls for multi-dimensional coupled systems with less controls than components in the system.

#### Presentation of the semi-discrete system

Let us rewrite the assumptions in the precise setting we consider here. Let  $a_1, a_2 \in \mathbb{R}_+^*$ ,  $\Omega_1 = (0, a_1)$ ,  $\Omega_2 = (0, a_2)$  and  $\Omega := \Omega_1 \times \Omega_2$ . Let  $\omega_2$  be a non empty open subset of  $\Omega_2$ . Without loss of generality, we assume that

$$\overline{\omega_2} \subset \Omega_2. \quad (3.42)$$

Let  $\gamma_1 \in \mathcal{C}^1(\overline{\Omega_1})$  and  $\gamma_2 \in \mathcal{C}^1(\overline{\Omega_2})$  be two functions which satisfy:

$$\gamma_{1,\min} := \inf_{\overline{\Omega_1}} \gamma_1 > 0 \text{ and } \gamma_{2,\min} := \inf_{\overline{\Omega_2}} \gamma_2 > 0.$$

Let  $\mathcal{A}_1 = -\partial_{x_1}(\gamma_1(x_1)\partial_{x_1}\bullet)$  (resp.  $\mathcal{A}_2 = -\partial_{x_2}(\gamma_2(x_2)\partial_{x_2}\bullet)$ ) be the self-adjoint unbounded operator in  $L^2(\Omega_1)$  (resp.  $L^2(\Omega_2)$ ) whose domain is  $D(\mathcal{A}_1) = H_0^1(\Omega_1) \cap H^2(\Omega_1)$  (resp.  $D(\mathcal{A}_2) = H_0^1(\Omega_2) \cap H^2(\Omega_2)$ ). Let  $n \geq 1$ , we define a self-adjoint unbounded operator in  $L^2(\Omega_1) \otimes L^2(\Omega_2) \otimes \mathbb{R}^n$  on the domain  $D(\mathcal{A}) = D(\mathcal{A}_1) \otimes D(\mathcal{A}_2) \otimes \mathbb{R}^n$  as follows

$$\mathcal{A} := \mathcal{A}_1 \otimes \mathbb{I} \otimes \mathbb{I} + \mathbb{I} \otimes \mathcal{A}_2 \otimes \mathbb{I}.$$

Note that  $\mathcal{A}$  can be seen as a self-adjoint operator in  $(L^2(\Omega))^n$  with domain  $(H_0^1(\Omega) \cap H^2(\Omega))^n$ . The control matrix  $B \in M_{n,1}(\mathbb{R})$  and the coupling matrix  $C \in M_n(\mathbb{R})$  are chosen in the so-called cascade form, given as follows

$$B = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & \text{---} & 0 \\ 1 & \diagdown & \vdots \\ 0 & \diagdown & \vdots \\ \vdots & \diagdown & \vdots \\ 0 & \text{---} & 0 & 1 & 0 \end{pmatrix}.$$

Let us now define the finite difference approximation that we consider for this system. For simplicity we present the results on uniform grids even though the same result will hold for regular enough families of meshes (see for instance the discussion in [1, Remark 1.6] or [8, Section 5]). Let  $N_1$  and  $N_2$  be two integers greater or equal to 2. For  $i \in \{1, 2\}$ , we set  $h_i := \frac{a_i}{N_i+1}$  and consider  $(x_{1,j})_{j=0}^{N_1+1}$  and  $(x_{2,j})_{j=0}^{N_2+1}$  a discretization of  $\Omega_1$  and  $\Omega_2$  respectively :  $x_{1,j} := jh_1$  for  $j \in \{0, \dots, N_1 + 1\}$ ,  $x_{2,j} := jh_2$  for  $j \in \{0, \dots, N_2 + 1\}$ . We define also, for  $j \in \{0, \dots, N_1\}$ ,  $x_{1,j+1/2} = x_{1,j} + \frac{h_1}{2}$  and for  $j \in \{0, \dots, N_2\}$ ,  $x_{2,j+1/2} = x_{2,j} + \frac{h_2}{2}$ .

The geometry of the grid and of the control domain is summarized in Figure 3.2: the interior grid points are shown in black circles  $\bullet$ , the homogeneous Dirichlet boundary points are shown in white circles  $\circ$  and the boundary points in the control region are shown in gray squares  $\blacksquare$ .

Associated with those grids, we can introduce the discrete functional spaces  $E_1 = \mathbb{R}^{N_1}$  (resp.  $E_2 = \mathbb{R}^{N_2}$ ) equipped with the Euclidean inner product inherited from the  $L^2$  inner product in  $\Omega_1$  (resp. in  $\Omega_2$ ) and defined by

$$\langle y, z \rangle_{0,i} := \sum_{j=1}^{N_i} h_i y_j z_j, \quad \forall y, z \in E_i, \forall i \in \{1, 2\}.$$

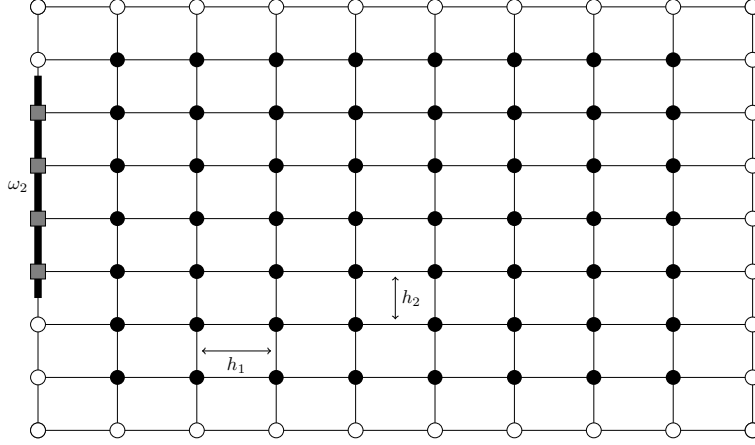


Figure 3.2 – Grid geometry

For any function  $f : \Omega_i \rightarrow \mathbb{R}$ , we will use the same letter  $f$  to denote the sampling of  $f$  on the grid  $(f(x_{i,j}))_j \in E_i$  or the *multiplication by  $f$*  operator in  $E_i$  which means that, for any  $y \in E_i$ , the vector  $fy \in E_i$  is defined by

$$(fy)_j = f(x_{i,j})y_j, \quad \forall j \in \{1, \dots, N_i\}. \quad (3.43)$$

Let  $A_i$  be the self-adjoint operator in  $E_i$  (that can be seen as a  $N_i \times N_i$  matrix) corresponding to the discretization of the scalar 1D operator  $\mathcal{A}_i$  by the finite difference method, which is defined for any  $y \in E_i$  by

$$(A_i y)_j := -\frac{1}{h_i} \left( \gamma_{i,j+1/2} \frac{y_{j+1} - y_j}{h_i} - \gamma_{i,j-1/2} \frac{y_j - y_{j-1}}{h_i} \right), \quad \forall j \in \llbracket 1, N_i \rrbracket, \quad (3.44)$$

with the usual convention that  $y_0 = y_{N_i+1} = 0$ . Here, we have used the notation  $\gamma_{i,j+1/2} := \gamma_i(x_{i,j+1/2})$ , for  $i = 1, 2$  and  $j \in \{0, \dots, N_i - 1\}$ , for the sampling of the diffusion coefficient  $\gamma_i$  on the dual grid of the mesh of  $\Omega_i$ . At some point we will also need to consider the discrete Laplace operators  $\Delta_i$  defined by

$$(-\Delta_i y)_j := -\frac{y_{j+1} - 2y_j + y_{j-1}}{h_i^2}, \quad \forall j \in \llbracket 1, N_i \rrbracket, \quad \forall y \in E_i.$$

The discretization of the vector-valued 2D operator  $\mathcal{A}$  we will be interested in is thus given by

$$A = A_1 \otimes I \otimes I + I \otimes A_2 \otimes I,$$

and, if we take into account the coupling terms, the complete discretization of the operator appearing in (3.6) is given by

$$L := A + I \otimes I \otimes C.$$

We introduce now the discrete control spaces and operators associated with our boundary control problem. We set  $U_2 = E_2$ , and to avoid confusions, we will use the notation  $[\bullet, \bullet]_{0,2}$  instead of  $\langle \bullet, \bullet \rangle_{0,2}$  when dealing with objects in  $U_2$ .

Using the convention (3.43), we define the discrete control operator in  $\Omega_2$  by

$$B_2 = 1_{\omega_2}.$$

This means that for any  $y \in E_2$ ,

$$(B_2 y)_j = \begin{cases} y_j, & \text{if } x_{2,j} \in \omega_2, \\ 0, & \text{if } x_{2,j} \notin \omega_2. \end{cases}$$

Observe that this operator is self-adjoint in  $E_2$ , that is  $B_2^* = B_2$ . Since the domain  $\Omega_1$  is a 1D interval, the corresponding boundary control  $v$  is in fact a scalar control that lives in the space  $U_1 = \mathbb{R}$ . However, the discretization  $A_1$  of the operator  $\mathcal{A}_1$  is built upon the assumption that we are dealing with homogeneous Dirichlet boundary condition. Therefore, in order to include non-homogeneous



Dirichlet boundary condition on the left end-point of  $\Omega_1$  in the discretization, we need to add a source term in the discretization which is given by

$$\mathbf{B}_1 v := \begin{pmatrix} -\frac{\gamma_{1,1/2}}{h_1^2} v \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in E_1.$$

For the analysis, it will be slightly more convenient to work with this operator in the following form

$$\mathbf{B}_1 v = \gamma_{1,1/2}(\Delta_1 r)v, \quad \forall v \in \mathbb{R}, \quad (3.45)$$

where  $r$  is (the sampling of) the affine map defined by  $x \in \Omega_1 \mapsto 1 - x/a_1$ . This is the discrete counterpart of the boundary control operator as analyzed in [15, Chapter 2].

A simple computation shows that the adjoint of  $\mathbf{B}_1$  is given by

$$\mathbf{B}_1^* q = \gamma_{1,1/2} \partial_l q,$$

where  $\partial_l q$  is the discrete *normal* derivative of  $q \in E_1$  on the left boundary of the domain defined by

$$\partial_l q = \frac{0 - q_1}{h_1} = -\frac{q_1}{h_1}.$$

Note that this formula takes into account implicitly the homogeneous Dirichlet boundary condition for  $q$ .

We can now precisely write the semi-discrete control problem that we consider

$$\begin{cases} y'(t) + Ly(t) = \mathbf{B}_1 \otimes \mathbf{B}_2 \otimes \mathbf{B}v(t) \\ y(0) = y_0 \in E, \end{cases} \quad (3.46)$$

where  $y(t) \in E = E_1 \otimes E_2 \otimes \mathbb{R}^n$  and  $v(t) \in U = \mathbb{R} \otimes U_2 \otimes \mathbb{R}$ . Note that, in this particular case, the control space  $U$  can be in fact identified with  $E_2$ .

The main theorem of this section is the following. The crucial point is that all the constants appearing in the estimates (3.47) do not depend on the discretization parameter  $h$ . In short, we prove that we can drive the semi-discrete system (3.46) to a target which is exponentially small with respect to  $h$  with controls that are uniformly bounded. Up to a subsequence, this results imply the weak convergence of the semi-discrete controls towards a control of the continuous problem which leads the solution to zero, as soon as the discrete initial data converges towards the suitable initial data.

**Theorem 3.4.1** *There exist  $C > 0$ ,  $\tilde{C} > 0$  and  $h_0 > 0$ , depending only on  $\gamma_1$ ,  $\gamma_2$  and  $\omega_2$  such that, for any  $T > 0$ , any mesh such that  $h < h_0$ , and any  $y_0 \in E$  there exists a control  $v \in L^2(0, T; U)$  satisfying*

$$\begin{cases} \|v\|_{L^2(0, T; U)} \leq e^{\tilde{C}(1 + \frac{1}{h} + T)} \|y_0\|_{-1, A}, \\ \|y(T)\|_{-1, A} \leq e^{\tilde{C}(1 + \frac{1}{h} + T)} e^{-C/h^2} \|y_0\|_{-1, A}, \end{cases} \quad (3.47)$$

where  $y$  is the corresponding solution to the semi-discrete problem (3.46) with control  $v$  and  $h = \max(h_1, h_2)$  is the space discretization parameter.

Once such a theorem is proved, even with a non explicit/constructive proof, one can produce an optimization algorithm, based on the penalized HUM approach, that is able to compute a control  $v$  satisfying (3.47). We can even relax the requirements by replacing the exponential factor  $e^{-C/h^2}$  by any more convenient  $\phi(h)$ , such as  $\phi(h) = h^p$  for some large enough integer  $p$ . Those questions are discussed in details for instance in [6] where some numerical illustrations are given.

### Additional notations and properties

In order to simplify the presentation of the following proofs we need to introduce a few more notations. For any  $i \in \{1, 2\}$ , we define  $\gamma_i^\pm$  to be the translated sampling of the diffusion coefficient  $\gamma_i$  defined by

$$(\gamma_i^\pm)_j = \gamma_{i, j \pm 1/2}, \quad \forall j \in \{1, \dots, N_i\}.$$

We also introduce the forward and backward difference operators  $\nabla_i^\pm$  defined, for any  $y \in E_i$ , by

$$\begin{aligned} (\nabla_i^+ y)_j &= \frac{y_{j+1} - y_j}{h_i}, \quad \forall j \in \{1, \dots, N_i\}, \\ (\nabla_i^- y)_j &= \frac{y_j - y_{j-1}}{h_i}, \quad \forall j \in \{1, \dots, N_i\}. \end{aligned}$$

For any  $i \in \{1, 2\}$ , we define  $(e_{i,j})_{1 \leq j \leq N_i}$  to be the canonical basis of  $E_i$  (each element corresponds to a point in the 1D grid of  $\Omega_i$ ). At some point in the forthcoming analysis, we shall need to work with *compactly supported discrete functions* in order to justify some discrete integration by parts. To this end, we introduce the subspaces  $E_i^{00} \subset E_i^0 \subset E_i$  defined by

$$\begin{aligned} E_i^0 &= \text{span}(e_{i,j}, \quad j = 2, \dots, N_i - 1), \\ E_i^{00} &= \text{span}(e_{i,j}, \quad j = 3, \dots, N_i - 2). \end{aligned}$$

With those notations, we observe that the gradient operators defined above satisfy the duality property

$$\langle \nabla_i^+ y, z \rangle_{0,i} = - \langle y, \nabla_i^- z \rangle_{0,i}, \quad \forall y, z \in E_i^0, \quad (3.48)$$

that we obtain by a summation by parts.

**Lemma 3.4.1** *There exists a  $C > 0$  depending only on  $\inf \gamma_i$ ,  $\sup \gamma_i$ ,  $\|\gamma'_i\|_{L^\infty}$ ,  $a_i$  such that: for any  $i \in \{1, 2\}$ , and any  $y \in E_i$ , we have*

$$\|y\|_{0,i} \leq C \|\nabla_i^\pm y\|_{0,i}, \quad (3.49)$$

$$\|\nabla_i^+ y\|_{0,i}^2 + \|\nabla_i^- y\|_{0,i}^2 \leq C \langle A_i y, y \rangle_{0,i}, \quad (3.50)$$

$$\|\Delta_i y\|_{0,i} \leq C \|A_i y\|_{0,i}. \quad (3.51)$$

**Remark 3.4.1** *By combining the above properties, we can obtain the following estimate*

$$\|y\|_{0,i} \leq C \|A_i y\|_{0,i}, \quad \forall y \in E_i, \quad (3.52)$$

where  $C$  depends only on  $\gamma_i$  and  $a_i$ . This implies in particular that the Poincaré inequality (3.23) holds with a constant  $M_{P,1}$  uniform with respect to the discretization parameters.

### Proof

The first inequality is very classical. We recall the sketch of proof: for  $i \in \{1, 2\}$ , we first write

$$y_j = y_j - y_0 = \sum_{k=1}^j (y_k - y_{k-1}) = \sum_{k=1}^j h_i (\nabla_i^- y)_k,$$

and by the Cauchy-Schwarz inequality, we obtain

$$|y_j|^2 \leq a_i \|\nabla_i^- y\|_{0,i}^2, \quad \forall j \in \{0, \dots, N_i\}.$$

The claim follows by multiplying by  $h_i$  and summing over  $j$ . The proof with the operator  $\nabla_i^+$ , is done in the same way but starting from the equality

$$y_j = y_j - y_{N_i+1} = - \sum_{k=j}^{N_i} (y_{k+1} - y_k) = \sum_{k=j}^{N_i} h_i (\nabla_i^+ y)_k.$$

For the third estimate, a simple calculation shows that

$$\langle A_i y, y \rangle_{0,i} = \sum_{j=0}^{N_i} h_i \gamma_{i,j+1/2} \left| \frac{y_{j+1} - y_j}{h_i} \right|^2, \quad \forall y \in E_i,$$

and we deduce (3.50) with  $C$  depending only on  $\min_{\Omega_i}(\gamma_i)$ .

Finally, a straightforward algebraic computation gives

$$A_i = \frac{\gamma_i^+ + \gamma_i^-}{2}(-\Delta_i) - \frac{\gamma_i^+ - \gamma_i^-}{h_i} \frac{\nabla_i^+ + \nabla_i^-}{2},$$

which implies

$$\|-\Delta_i y\|_{0,i} \leq C \|A_i y\|_{0,i} + C \|\gamma_i'\|_{L^\infty} (\|\nabla_i^+ y\|_{0,i} + \|\nabla_i^- y\|_{0,i}),$$

where  $C$  only depends on  $\min_{\Omega_i}(\gamma_i)$ . By using (3.50), the Cauchy-Schwarz inequality and the Young inequality, we finally deduce that

$$\|-\Delta_i y\|_{0,i} \leq C \|A_i y\|_{0,i} + C \|y\|_{0,i}, \quad \forall y \in E_i,$$

where  $C$  only depends on the function  $\gamma_i$  and we conclude by using (3.49) and (3.50). ■

### Proof of the semi-discrete controllability result

The proof of Theorem 3.4.1 consists essentially in applying the analysis above to a suitable setting. To achieve the results, it is just needed to check that all the assumptions of Theorem 3.3.1 are satisfied with constants that do not depend on the discretization parameter.

- In a first step, we will check that the discrete diffusion operator  $L$  and the discrete control operator  $B = B_1 \otimes B_2 \otimes B$  are compatible in the sense that inequalities (3.14) and (3.15) are satisfied with  $s = 1$ , and  $D = A$ , uniformly with respect to the mesh size.
- In a second step, using previous results in the literature, we shall prove that the discrete Lebeau-Robbiano spectral inequality (Assumption 3.3.2) and the semi-discrete controllability of the 1D system (Assumption 3.3.1) hold.

Let us start by proving the following result.

**Lemma 3.4.2** *There exist two constants  $M_{cont}$  and  $M_{adm}$  independent of  $h_1$  and  $h_2$  that satisfy:*

$$\sup_{t \in [0, T]} \|e^{-tL^*} \psi\|_{1,A} \leq M_{cont} \|\psi\|_{1,A}, \quad \forall \psi \in E, \quad (3.53)$$

$$\left( \int_0^T \left[ \mathbb{B}^* e^{-\tau L^*} \psi \right]_0^2 d\tau \right)^{\frac{1}{2}} \leq M_{adm} \|\psi\|_{1,A}, \quad \forall \psi \in E, \quad (3.54)$$

### Proof

The proof of (3.53) is exactly the same as the one of the dissipation estimates of Proposition 3.3.1, for  $\mu = 0$  and  $s = -1$ , except that we deal with the adjoint matrix  $C^*$ .

The more intricate part is now to prove (3.54). This will be a combination of discrete trace estimates and of discrete elliptic regularity properties for the operator  $A$ . In that proof the fact that we consider  $s = 1$  in the definition of the norms is crucial, this is the discrete counter-part of the fact that, in the continuous setting, we need to take  $H_0^1$  initial data to ensure that the normal derivative of the solution of the backward heat equation belongs to  $L^2$ .

1. First, we prove a **1D trace inequality**. More precisely, we show that there exists  $C > 0$  independent of  $h_1$  such that:

$$\left[ \mathbb{B}_1^* \otimes \mathbb{B}^* q \right]_{0,1}^2 \leq C \|A_1 \otimes I q\|_{0,1}^2, \quad \forall q \in E_1 \otimes \mathbb{R}^n. \quad (3.55)$$

To this end, we first use (3.45) and the fact that  $\Delta_1$  is self-adjoint to get

$$\mathbb{B}_1^* \otimes \mathbb{B}^* = \gamma_{1,1/2}(r \otimes \mathbb{B}^*)(\Delta_1 \otimes I).$$

It follows that, for any  $q \in E_1 \otimes \mathbb{R}^n$ ,

$$\left[ \mathbb{B}_1^* \otimes \mathbb{B}^* q \right]_{0,1} \leq \gamma_{1,1/2} \|r\|_{0,1} \|\mathbb{B}^*\| \|\Delta_1 \otimes I q\|_{0,1} \leq C \|\Delta_1 \otimes I q\|_{0,1},$$

where  $C$  depends only on  $\gamma_1$  and  $\Omega_1$ .

From (3.51) (that we apply on each component of the vector-valued unknown  $q$ ), we deduce that

$$\|\Delta_1 \otimes I q\|_{0,1} \leq C \|A_1 \otimes I q\|_{0,1},$$

which proves (3.55).

2. In the second step, we prove that the inequality (3.55) can be extended to **2D discrete unknowns**. More precisely, we show that there exists a constant  $C > 0$  independent of  $h_1$  and  $h_2$  such that

$$\|B^*q\|_0^2 = \|B_1^* \otimes B_2^* \otimes B^*q\|_0^2 \leq C \|A_1 \otimes I \otimes Iq\|_0^2, \forall q \in E. \quad (3.56)$$

We define  $f_j = e_{2,j}/\sqrt{h_2}$ , in such a way that  $(f_{2,j})_{1 \leq j \leq N_2}$  is an orthonormal basis of  $E_2$ . Note that the particular structure of  $B_2$  implies that  $B_2 f_{2,j} = \alpha_j f_{2,j}$  with  $\alpha_j \in \{0, 1\}$  ( $\alpha_j = 1$  if and only if the mesh point  $x_{2,j}$  lies in  $\omega_2$ ).

We can decompose any  $q \in E$  into the unique form

$$q = \sum_{j=1}^{N_2} q_j \otimes f_{2,j},$$

with  $q_j \in E_1 \otimes \mathbb{R}^n$  for any  $j$ . We have, by orthogonality of  $(f_{2,j})_j$ ,

$$\|B_1^* \otimes B_2^* \otimes B^*q\|_0^2 = \left\| \sum_{j=1}^{N_2} \alpha_j \left[ (B_1^* \otimes B^*)q_j \right] \otimes f_{2,j} \right\|_0^2 = \sum_{j=1}^{N_2} \|B_1^* \otimes B^*q_j\|_{0,1}^2 \alpha_j^2.$$

Using (3.55) for each  $j$ , and the fact that  $\alpha_j^2 \leq 1$ , we get

$$\|B_1^* \otimes B_2^* \otimes B^*q\|_0^2 \leq C \sum_{j=1}^{N_2} \|A_1 \otimes Iq_j\|_{0,1}^2 = C \sum_{j=1}^{N_2} \left\| \left[ (A_1 \otimes I) \otimes I \right] q_j \otimes f_{2,j} \right\|_0^2,$$

and, still by orthogonality of  $(f_{2,j})_j$ , we obtain the claimed inequality (3.56), the constant  $C$  being the same as in (3.55).

3. In the estimate (3.56) we only have the operator  $A_1 \otimes I \otimes I$  that appears in the right-hand side and not the complete discrete 2D elliptic operator  $A$ . The third step consists in proving a **discrete elliptic regularity property** that will allow us to get that there exists a constant  $C > 0$  independent of  $h_1$  and  $h_2$  such that

$$\|B^*q\|_0^2 \leq C \|Aq\|_0^2, \forall q \in E. \quad (3.57)$$

The main idea is based on the following well-known computation: for any smooth and compactly supported function  $f$  defined on  $\mathbb{R}^2$ , we can write

$$\begin{aligned} & \iint |\partial_{x_1} (\gamma_1(x_1)\partial_{x_1}f) + \partial_{x_2} (\gamma_2(x_2)\partial_{x_2}f)|^2 \\ &= \iint |\partial_{x_1} (\gamma_1(x_1)\partial_{x_1}f)|^2 + |\partial_{x_2} (\gamma_2(x_2)\partial_{x_2}f)|^2 \\ & \quad + 2 \iint \partial_{x_1} (\gamma_1(x_1)\partial_{x_1}f) \times \partial_{x_2} (\gamma_2(x_2)\partial_{x_2}f). \end{aligned}$$

Moreover, by a double integration by parts, we find

$$\iint \partial_{x_1} (\gamma_1(x_1)\partial_{x_1}f) \times \partial_{x_2} (\gamma_2(x_2)\partial_{x_2}f) = \iint \gamma_1(x_1)\gamma_2(x_2) |\partial_{x_1}\partial_{x_2}f|^2 \geq 0,$$

so that

$$\iint |\partial_{x_1} (\gamma_1(x_1)\partial_{x_1}f)|^2 \leq \iint |\partial_{x_1} (\gamma_1(x_1)\partial_{x_1}f) + \partial_{x_2} (\gamma_2(x_2)\partial_{x_2}f)|^2.$$

We want to apply the same idea to prove roughly speaking that

$$\|A_1 \otimes Iq\|_0^2 \leq \|A_1 \otimes Iq + I \otimes A_2q\|_0^2,$$

which, according to (3.56), would prove (3.57).

However, because of boundary terms, it is easier to prove this inequality for compactly supported discrete functions, that is for  $q \in E_1^{00} \otimes E_2^{00}$ . We will thus proceed by extension and truncation of the discrete functions under study.

Let  $\xi_1 : (0, a_1) \rightarrow \mathbb{R}$  be a smooth function such that  $\xi_1 = 1$  on  $(0, \frac{a_1}{3})$ , and  $\xi_1 = 0$  on  $(\frac{2a_1}{3}, a_1)$ , and  $\xi_2 : (0, a_2) \rightarrow \mathbb{R}$  be a smooth compactly supported function such that  $\xi_2 = 1$  on  $\omega_2$ . Such a

function exists thanks to the assumption (3.42). We introduce now the truncation operator on  $E_1 \otimes E_2$ , defined by  $T = \xi_1 \otimes \xi_2$ . The choice of  $\xi_1$  and  $\xi_2$  implies that

$$\|B_1^* \otimes B_2^* \otimes B^* q\|_0^2 = \|(B_1^* \otimes B_2^* \otimes B^*)(T \otimes I)q\|_0^2,$$

and thus by (3.56) we have

$$\|B_1^* \otimes B_2^* \otimes B^* q\|_0^2 \leq C \|(A_1 \otimes I \otimes I)(T \otimes I)q\|_0^2.$$

The discrete function  $(T \otimes I)q$  vanish near all the boundaries of the domain, except the one corresponding to  $\{x_1 = 0\}$  because of the choice of  $\xi_1$ . We will now introduce a symmetrization procedure that will let us work with a compactly supported discrete function.

We start by introducing the extended space in the first variable defined by

$$\bar{E}_1 := \mathbb{R}^{\{-N_1, \dots, N_1\}},$$

which stands for discrete functions defined on a uniform discretization of  $(-a_1, a_1)$  with a mesh size  $h_1$ . If we denote by  $(\bar{e}_{1,j})_{-N_1 \leq j \leq N_1}$  the canonical basis of  $\bar{E}_1$ , we can define the odd symmetrization operator  $S : E_1 \rightarrow \bar{E}_1$  by

$$S e_{1,j} = \bar{e}_{1,j} - \bar{e}_{1,-j}, \quad \forall j \in \{1, \dots, N_1\}.$$

With this notation, for  $q \in E_1 \otimes E_2$ ,  $(S \otimes I)q$  is the odd symmetrization of  $q$  with respect to the  $x_1$  variable in the extended 2D domain.

We consider  $\bar{\gamma}_1$  to be the even extension of  $\gamma_1$  to  $(-a_1, a_1)$  and let  $\bar{A}_1$  be the discrete diffusion operator associated with  $\bar{\gamma}_1$  and defined on  $\bar{E}_1$  in the same way as in (3.44).

A simple computation shows that the symmetrization is compatible with the diffusion operator definitions, that is

$$\bar{A}_1 S q = S A_1 q, \quad \forall q \in E_1. \quad (3.58)$$

In particular, using the same notation for the norm in  $\bar{E}_1$  as for the one in  $E_1$ , we have

$$\|\bar{A}_1 S q\|_{0,1}^2 = 2 \|A_1 q\|_{0,1}^2, \quad \forall q \in E_1.$$

As a consequence, all the previous estimates lead to the following inequality

$$\|B_1^* \otimes B_2^* \otimes B^* q\|_{0,2}^2 \leq C \|\bar{A}_1 \otimes I \otimes I \bar{q}\|_0^2, \quad \forall q \in E,$$

where, we have set  $\bar{q} = (S \otimes I \otimes I)(T \otimes I)q$ . Observe now that, by construction of the symmetrization and truncation operators, the discrete function  $\bar{q}$  belongs to  $\bar{E}_1^{00} \otimes E_2^{00} \otimes \mathbb{R}^n$ .

Note that for such *compactly supported* discrete functions we have the algebraic identities

$$(\bar{A}_1 \otimes I) \bar{q} = ((-\bar{\nabla}_1^- \bar{\gamma}_1^+ \bar{\nabla}_1^+) \otimes I) \bar{q},$$

$$(I \otimes A_2) \bar{q} = (I \otimes (-\nabla_2^- \gamma_2^+ \nabla_2^+)) \bar{q}.$$

We can now make the following computation

$$\begin{aligned} \|(\bar{A}_1 \otimes I + I \otimes A_2) \bar{q}\|_0^2 &= \|\bar{A}_1 \otimes I \bar{q}\|_0^2 + \|I \otimes A_2 \bar{q}\|_0^2 \\ &\quad + 2 \langle \bar{A}_1 \otimes I \bar{q}, I \otimes A_2 \bar{q} \rangle_0, \end{aligned}$$

and the double product term can be evaluated as follows, the discrete integration by parts being justified by (3.48) and the fact that  $\nabla_i^\pm E_i^{00} \subset E_i^0$ ,

$$\begin{aligned} \langle \bar{A}_1 \otimes I \bar{q}, I \otimes A_2 \bar{q} \rangle_0 &= \langle (\bar{\nabla}_1^- \bar{\gamma}_1^+ \bar{\nabla}_1^+) \otimes I \bar{q}, I \otimes (\nabla_2^- \gamma_2^+ \nabla_2^+) \bar{q} \rangle_0 \\ &= \langle (\bar{\gamma}_1^+ \otimes \gamma_2^+) (\bar{\nabla}_1^+ \otimes \nabla_2^+) \bar{q}, (\bar{\nabla}_1^+ \otimes \nabla_2^+) \bar{q} \rangle_0 \geq 0. \end{aligned}$$

At the end, we obtain that

$$\|\bar{A}_1 \otimes I \bar{q}\|_0^2 \leq \|(\bar{A}_1 \otimes I + I \otimes A_2) \bar{q}\|_0^2.$$

To conclude the proof of the final claim, we just need to show that

$$\|(\bar{A}_1 \otimes I + I \otimes A_2) \bar{q}\|_0^2 \leq C \|A\bar{q}\|_0^2,$$

for a  $C > 0$  independent of the mesh.

Recall that  $\bar{q} = (S \otimes I)Tq$  and by (3.58) we have

$$(\bar{A}_1 \otimes I + I \otimes A_2) \bar{q} = (S \otimes I)ATq,$$

and in particular we have

$$\|(\bar{A}_1 \otimes I + I \otimes A_2) \bar{q}\|_0^2 = 2 \|ATq\|_0^2.$$

It thus remains to evaluate the norm of  $ATq$  by the one of  $Aq$ . A simple algebraic computation leads to

$$\begin{aligned} ATq &= (A_1 \otimes I + I \otimes A_2)(\xi_1 \otimes \xi_2)q \\ &= (\xi_1 \otimes \xi_2)(Aq) + ((A_1 \xi_1) \otimes \xi_2)q + (\xi_1 \otimes (A_2 \xi_2))q \\ &\quad + (\gamma_1^+(\nabla_1^+ \xi_1) \otimes I)(\nabla_1^+ \otimes I)q + (\gamma_1^-(\nabla_1^- \xi_1) \otimes I)(\nabla_1^- \otimes I)q \\ &\quad + (I \otimes \gamma_2^+(\nabla_2^+ \xi_2))(I \otimes \nabla_2^+)q + (I \otimes \gamma_2^-(\nabla_2^- \xi_2))(I \otimes \nabla_2^-)q. \end{aligned}$$

Since  $\xi_1$  and  $\xi_2$  are smooth and  $\gamma_1$  and  $\gamma_2$  are bounded, we conclude, with the mean-value theorem that

$$\begin{aligned} \|ATq\|_0 &\leq C \|Aq\|_0 + C \|q\|_0 + C \|(\nabla_1^+ \otimes I)q\|_0 \\ &\quad + C \|(\nabla_1^- \otimes I)q\|_0 + C \|(I \otimes \nabla_2^+)q\|_0 + C \|(I \otimes \nabla_2^-)q\|_0. \end{aligned}$$

The conclusion follows by (3.49), (3.50), (3.52) and the Cauchy-Schwarz and Young inequalities.

4. The last step of the proof consists in showing (3.54) by using (3.57) and a classical energy estimate. Let  $\psi \in E$  and let  $q(t) = e^{-tL^*} \psi$ . We set  $z(t) = I \otimes I \otimes e^{tC^*} q(t)$ . One can check that  $z$  satisfies

$$\begin{cases} \partial_t z + Az = 0, \\ z(0) = \psi. \end{cases} \quad (3.59)$$

We multiply (3.59) by  $Az$  and integrate on  $(0, T)$ ,

$$\langle z(T), Az(T) \rangle_0 - \langle \psi, A\psi \rangle_0 + 2 \int_0^T \|Az(t)\|_0^2 dt = 0,$$

thus, since the operator  $A$  is positive,

$$\int_0^T \|Az(t)\|_0^2 dt \leq \|\psi\|_{1,A}^2.$$

Since  $A$  commutes with  $I \otimes I \otimes e^{tC^*}$ , we find that

$$\int_0^T \|I \otimes I \otimes e^{tC^*} Aq(t)\|_0^2 dt \leq \|\psi\|_{1,A}^2,$$

Moreover,  $\|I \otimes I \otimes e^{tC^*} Aq(t)\|_0^2 \geq e^{-2T\|C\|} \|Aq(t)\|_0^2$ , for any  $t \in (0, T)$  and so

$$\int_0^T \|Aq(t)\|_0^2 dt \leq e^{2T\|C\|} \|\psi\|_{1,A}^2.$$

Applying (3.57) to  $q(t)$  for any  $t \in (0, T)$  and integrating in time this inequality, we finally get

$$\int_0^T \llbracket B^* q(t) \rrbracket_0^2 dt \leq C e^{CT} \|\psi\|_{1,A}^2.$$

and this concludes the proof of (3.54).

■

**Remark 3.4.2** Note that we just proved that  $M_{adm}$  depends on  $T$  like  $e^{CT}$  and this is consistent with estimates of  $v$  and  $y(T)$  given by Theorem 3.4.1.

Now we can prove Assumption 3.3.2.

**Proof** [of Assumption 3.3.2]

In [7] the authors proved a discrete Lebeau-Robbiano inequality on quite general meshes. We translate the statement of their Theorem 6.1 in our setting.

**Theorem 3.4.2** There exists  $C > 0$ ,  $\varepsilon > 0$  and  $h_0 > 0$ , depending on  $\gamma_2, \omega_2$  such that if  $h_2 \leq h_0$ , we have for all  $0 < \mu \leq \varepsilon/h_2^2$ ,

$$\sum_{k:\lambda_{2,k} \leq \mu} |\alpha_k|^2 = \int_{\Omega_2} \left| \sum_{k:\lambda_{2,k} \leq \mu} \alpha_k \phi_{2,k} \right|^2 \leq C e^{C\sqrt{\mu}} \int_{\omega_2} \left| \sum_{k:\lambda_{2,k} \leq \mu} \alpha_k \phi_{2,k} \right|^2, \quad \forall (\alpha_k)_{1 \leq k \leq N_1} \subset \mathbb{C}$$

This result exactly yields that Assumption 3.3.2 is fulfilled in this setting with

$$\mu_{E_2} = \frac{\varepsilon}{h_2^2}.$$

■

Finally, we prove that Assumption 3.3.1 holds.

**Proof** [of Assumption 3.3.1]

We base our proof on the strategy developed in [1], where the semi-discretized (on a uniform mesh) boundary null-control problem in space dimension 1 with operator  $A_1$  is tackled by applying the moments method. However, the explicit dependence in  $T$  of the control cost by some bound in  $e^{C/T}$  was not given in that work. This is crucial in the present analysis.

We can actually obtain this precised bound by using the expression of the control obtained with this method, in setting (S2) described in [1], and by using the refined bounds on biorthogonal families to exponential functions given by Theorem 1.5 of [4]. We will just describe here the new estimate that we need to adapt the results of [1] to our needs.

We first set  $\varepsilon = 1/2$  for instance (any value between 0 and 1 would be acceptable), and we follow [1], to define

$$k_{max,\varepsilon}^h := \max \left\{ k \in \llbracket 1, N_1 \rrbracket; \lambda_{1,k} < \frac{4}{h_1^2} \gamma_{1,\min}(1 - \varepsilon) \right\}$$

and :

$$(\tilde{\Lambda}_\varepsilon^{h_1})_k := \begin{cases} \lambda_{1,k} & \text{for } k \in \llbracket 1, k_{max,\varepsilon}^h \rrbracket, \\ \lambda_{1,k_{max,\varepsilon}^h} + 4\gamma_{1,\min} k^2 & \text{for } k \geq k_{max,\varepsilon}^h + 1. \end{cases}$$

the sequence  $\tilde{\Lambda}_\varepsilon^{h_1}$  satisfies the following items required to apply Theorem 1.5 of [4]

1.  $(\tilde{\Lambda}_\varepsilon^{h_1})_k \neq (\tilde{\Lambda}_\varepsilon^{h_1})_n$ , for all  $k, n \in \mathbb{N}$  with  $k \neq n$ ;
- 2,3  $(\tilde{\Lambda}_\varepsilon^{h_1})_k \in \mathbb{R}_+^*$ , for every  $k \geq 1$ ;
4.  $(\tilde{\Lambda}_\varepsilon^{h_1})_k$  is non decreasing;
5. First, applying Theorem 3.2 of [1] there exists  $C > 0$  such that

$$(\tilde{\Lambda}_\varepsilon^{h_1})_{k+1} - (\tilde{\Lambda}_\varepsilon^{h_1})_k \geq C \sqrt{(\tilde{\Lambda}_\varepsilon^{h_1})_k}, \quad \forall k \geq 1.$$

Second, note that according to Lemma 3.3 of [1], there exist  $c_1$  and  $c_2$  two positive constants such that :

$$c_1 k^2 \leq \lambda_{1,k} \leq c_2 k^2, \quad \forall k \in \llbracket 1, N_1 \rrbracket. \quad (3.60)$$

Therefore,

$$(\tilde{\Lambda}_\varepsilon^{h_1})_{k+1} - (\tilde{\Lambda}_\varepsilon^{h_1})_k \geq Ck, \quad \forall k \geq 1.$$

Take  $k, n \in \mathbb{N}^*$ ,  $k > n$ , we have :

$$|(\tilde{\Lambda}_\varepsilon^{h_1})_k - (\tilde{\Lambda}_\varepsilon^{h_1})_n| \geq C \sum_{i=n}^{k-1} i \geq C(k^2 - n^2).$$

6. Finally, let  $\mathcal{N}$  the counting function associated with the sequence  $(\tilde{\Lambda}_\varepsilon^{h_1})_{k \geq 1}$ , defined by

$$\mathcal{N}(r) = \#\{k : (\tilde{\Lambda}_\varepsilon^{h_1})_k \leq r\}.$$

There exist  $\alpha > 0$ ,  $p_{\min} > 0$  and  $p_{\max} > 0$  such that for every  $r > 0$ ,

$$-\alpha + p_{\min}\sqrt{r} \leq \mathcal{N}(r) \leq \alpha + p_{\max}\sqrt{r}.$$

Indeed, from item 4, it follows easily from the definition of  $\mathcal{N}$  that  $(\tilde{\Lambda}_\varepsilon^{h_1})_{\mathcal{N}(r)} \leq r$  and  $(\tilde{\Lambda}_\varepsilon^{h_1})_{\mathcal{N}(r)+1} \geq r$ , and according to (3.60),

$$c_1\mathcal{N}(r)^2 \leq r \leq c_2(\mathcal{N}(r) + 1)^2,$$

thus,

$$\sqrt{\frac{r}{c_2}} - 1 \leq \mathcal{N}(r) \leq \sqrt{\frac{r}{c_1}}.$$

We can now apply Theorem 1.5 of [4] to conclude that there exists  $T_0 > 0$ , such that for every  $0 < T < T_0$ , there exists a family of functions in  $L^2(0, T)$  denoted by  $(q_{j,k}^{\tilde{\Lambda}_\varepsilon^{h_1}})_{k \geq 1, j \in \llbracket 1, n-1 \rrbracket}$ , satisfying for all  $k, l \geq 1$ , and  $i, j \in \llbracket 0, n-1 \rrbracket$ ,

$$\int_0^T q_{j,k}^{\tilde{\Lambda}_\varepsilon^{h_1}}(t)(t-T)^i \exp(-(\tilde{\Lambda}_\varepsilon^{h_1})_k(T-t)) dt = \delta_{k,l} \delta_{i,j},$$

and for all  $k \geq 1$  and  $j \in \llbracket 0, n-1 \rrbracket$ ,

$$\|q_{j,k}^{\tilde{\Lambda}_\varepsilon^{h_1}}\|_{L^2(0,T)} \leq C \exp\left(C\sqrt{(\tilde{\Lambda}_\varepsilon^{h_1})_k} + \frac{C}{T}\right), \quad (3.61)$$

where  $C$  does not depend on  $h_2$ .

**Remark 3.4.3** In [4], Theorem 1.5, hypothesis of 6 on the counting function is slightly different from item 6 given above. Indeed, in this reference the authors require the following condition : for some  $p, \alpha > 0$ ,

$$|p\sqrt{r} - \mathcal{N}(r)| \leq \alpha, \forall r > 0. \quad (3.62)$$

Actually, looking carefully at the proof of Theorem 1.5, we realize that we can use item 6 above instead of hypothesis (3.62). Indeed, we can just replace  $p$  by  $p_{\max}$  in the definition of the number  $d$  at page 2988 and in the proofs of Theorem 4.3 at page 2987, Lemma A.3 at page 2994 (note that the property  $\lim_{t \rightarrow +\infty} \frac{\mathcal{N}(t)}{t} = 0$  is preserved), and Lemma A.4 at page 2996 and change  $p$  into  $p_{\min}$  at page 2993 in the proof of Lemma A.1.

Now, in the proof of [1, Theorem 5.4], we can use the improved estimate (3.61) into the bound of the control (with  $t_0 = \frac{T}{2}$ ), and find that

$$\|v\|_{L^2(0,T;U)} \leq C_\varepsilon \|y_0\|_{-1,A_1} \sum_{k=1}^{k_{\max,\varepsilon}^h} e^{-\lambda_{1,k} \frac{T}{2}} e^{C\sqrt{\lambda_{1,k}} + \frac{C}{T}},$$

and using (3.60),

$$\|v\|_{L^2(0,T;U)} \leq C_\varepsilon e^{\frac{C}{T}} \|y_0\|_{-1,A_1} \sum_{k=1}^{+\infty} e^{-c_2 k^2 \frac{T}{2} + C\sqrt{c_1} k},$$

thus there exists  $C_1 > 0$  depending only on  $\varepsilon$  (whose value has been arbitrarily set to 1/2) and  $\gamma_1$  such that

$$\|v\|_{L^2(0,T;U)} \leq C_1 e^{\frac{C_1}{T}} \|y_0\|_{-1,A_1}. \quad (3.63)$$

Indeed if  $k \geq k_0 := \lceil \frac{C\sqrt{c_1}+1}{c_2 T/2} \rceil$ , we have  $-c_2 k^2 \frac{T}{2} + C\sqrt{c_1} k \leq -k$ , so

$$\begin{aligned} \sum_{k=1}^{+\infty} e^{-c_2 k^2 \frac{T}{2} + C\sqrt{c_1} k} &\leq \sum_{k=1}^{k_0} e^{-c_2 k^2 \frac{T}{2} + C\sqrt{c_1} k} + \sum_{k=k_0+1}^{+\infty} e^{-k} \\ &\leq k_0 e^{C\sqrt{c_1} k_0} + \frac{1}{1-e}, \end{aligned}$$



which gives (3.63) since  $k_0 \leq \frac{C_{\gamma_1, \varepsilon}}{T}$ .

Still using the estimates in the proof of Theorem 5.4 of [1], we get

$$\|y(T)\|_{-1, A_1} \leq e^{C_2(1+\frac{1}{T})} e^{-\mu_{E_1} T} \|y_0\|_{-1, A_1},$$

with

$$\mu_{E_1} = \frac{C_1}{h_1^2}$$

and  $C_1 > 0$  and  $C_2 > 0$  depending only on  $\gamma_1$ . ■

Now we can apply Theorem 3.3.1 with  $\mu_{E_1}$  and  $\mu_{E_2}$  both of the same order  $\frac{C}{h^2}$ . The proof of Theorem 3.4.1 is complete.

### 3.4.2 Dirichlet boundary null control of a continuous n-dimensional system on a cylindrical domain of dimension d

Our aim is to show how the finite dimensional framework developed in Section 3.3 actually applies to the study of null-controllability problems of tensorized parabolic systems.

As an illustration of this statement, we shall give a short proof of the main result of [4] by using Theorem 3.3.1 of the present article. This is of course not surprising since our approach is directly inspired from the one developed in [4]. However, it seems to us interesting to show how our general finite dimensional framework actually encompasses already known results through a spectral projection technique. This example should convince the reader that, by using the same strategy, one can easily adapt the proof to other kinds of tensorized controlled systems like, for instance, Neumann or Robin boundary controls, or even mixed (distributed and boundary) controls, as soon as we have in hand a suitable controllability result on the associated 1D system.

Finally, since we have taken care of all the constants in the proofs, this strategy can be used to derive controllability properties that are uniform with respect to some parameters present in the problem. As an illustration, in the case of a 1D Robin boundary control problem, we can prove estimates that are uniform in the Robin parameter (see [5]). By the present technique, those result will automatically be translated to the corresponding multi-D result, generalizing the one proved in [4].

Let us recall the statement of [4, Theorem 1.3] with the notation of the present paper.

**Théorème 3.4.1** *Let  $\Omega_1 = (0, 1)$ , and  $\Omega_2$  be a smooth bounded connected domain of  $\mathbb{R}^{d-1}$  and set  $\Omega = \Omega_1 \times \Omega_2$ .*

*We suppose given a coupling matrix  $C \in M_n(\mathbb{R})$  and a control matrix  $B \in M_{n,m}(\mathbb{R})$ . Assume that the following system of n equations*

$$\begin{cases} \partial_t y^1 + \mathcal{A}_1 \otimes I y^1 + \mathcal{I} \otimes C y^1 = 0, & \text{in } (0, T) \times \Omega_1, \\ y^1 = 1_{\{0\}} \otimes (B v^1) & \text{in } (0, T) \times \partial\Omega_1, \\ y^1(0) = y_0^1 & \text{in } \Omega_1, \end{cases}$$

*is null controllable for any  $y_0^1 \in H^{-1}(\Omega_1) \otimes \mathbb{R}^n$  and any time  $T > 0$  with, in addition, the following bound:*

$$\|v^1\|_{L^2(0, T; \mathbb{R}^m)} \leq e^{M_{\text{obs}, 1}(1+\frac{1}{T})} \|y_0^1\|_{H^{-1}(\Omega_1) \otimes \mathbb{R}^n}. \quad (3.64)$$

*Then, for any nonempty open set  $\omega_2 \subset \Omega_2$ , the following system of n equations:*

$$\begin{cases} \partial_t y + \mathcal{L} y = 0, & \text{in } (0, T) \times \Omega, \\ y = (1_{\{0\}} \times \omega_2 \otimes B) v & \text{in } (0, T) \times \partial\Omega, \\ y(0) = y_0 & \text{in } \Omega, \end{cases} \quad (3.65)$$

*where  $\mathcal{L} := \mathcal{A}_1 \otimes \mathcal{I} \otimes I + \mathcal{I} \otimes \mathcal{A}_2 \otimes I + \mathcal{I} \otimes \mathcal{I} \otimes C$ , is null controllable for any  $y_0 \in H^{-1}(\Omega) \otimes \mathbb{R}^n$  and any time  $T > 0$  with, in addition, the following bound:*

$$\|v\|_{L^2((0, T) \times \partial\Omega)^m} \leq e^{M_{\text{obs}, 1}(1+\frac{1}{T})} \|y_0\|_{H^{-1}(\Omega) \otimes \mathbb{R}^n}. \quad (3.66)$$

**Proof**

Let  $y_0 \in H^{-1}(\Omega) \otimes \mathbb{R}^n$ , recall that, by definition, the solution  $y$  of (3.65) satisfies for any  $\psi \in H_0^1(\Omega) \otimes \mathbb{R}^n$ , the equality

$$\begin{aligned} \langle y(T), \psi \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} - \langle y_0, e^{-T\mathcal{L}^*} \psi \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ = - \left[ \left( \frac{\partial}{\partial x_1} \Big|_{x_1=0} \otimes 1_{\omega_2} \otimes \mathbf{B}^* \right) e^{-(T-\cdot)\mathcal{L}^*} \psi, v \right]_{L^2((0,T) \times \partial\Omega)^m}. \end{aligned}$$

Therefore,  $v$  is a null-control for this system, if and only if, it satisfies

$$\begin{aligned} - \langle y_0, e^{-T\mathcal{L}^*} \psi \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ = - \left[ \left( \frac{\partial}{\partial x_1} \Big|_{x_1=0} \otimes 1_{\omega_2} \otimes \mathbf{B}^* \right) e^{-(T-\cdot)\mathcal{L}^*} \psi, v \right]_{L^2((0,T) \times \partial\Omega)^m}, \end{aligned}$$

for any  $\psi \in H_0^1(\Omega) \otimes \mathbb{R}^n$ . Actually, it is enough to check the previous equality for any  $\psi$  belonging to a total family of  $H_0^1(\Omega) \otimes \mathbb{R}^n$ .

Let us consider the total family of  $H_0^1(\Omega)$  made of the eigenfunctions of  $\mathcal{A}_1$  and  $\mathcal{A}_2$ . For  $i = 1, 2$ , we denote by  $(\phi_{i,j}, \lambda_{i,j})_{j \geq 1}$  the eigenfunctions and eigenvalues of the operator  $\mathcal{A}_i$  with homogeneous Dirichlet boundary conditions. We choose them to form an orthonormal basis of  $L^2(\Omega_i)$ . We will thus consider the total family of  $H_0^1(\Omega)$  defined by  $(\phi_{1,i} \otimes \phi_{2,j})_{i,j}$  that will be tensorized with the canonical basis of  $\mathbb{R}^n$ ,  $(e_k)_{k \in \{1, \dots, n\}}$  to finally produce a total family of  $H_0^1(\Omega) \otimes \mathbb{R}^n$ .

We are thus led to find a control satisfying,  $\forall i, j \geq 1$  and  $\forall k \in \{1, \dots, n\}$ ,

$$\begin{aligned} e^{-T(\lambda_{1,i} + \lambda_{2,j})} \langle y_0, \phi_{1,i} \otimes \phi_{2,j} \otimes (e^{-T\mathcal{C}^*} e_k) \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ = \left[ e^{-(T-\cdot)(\lambda_{1,i} + \lambda_{2,j})} (\phi_{1,i})'(0) \otimes (1_{\omega_2} \phi_{2,j}) \otimes (\mathbf{B}^* e^{-(T-\cdot)\mathcal{C}^*} e_k), v \right]_{L^2((0,T) \times \partial\Omega)^m}. \quad (3.67) \end{aligned}$$

In order to apply Theorem 3.3.1, we need to consider a projection of (3.65) on a finite dimensional space. For  $i = 1, 2$  and  $J \geq 1$ , we define the spaces

$$F_{i,J} = \text{span}(\phi_{i,j}, j \leq J),$$

and the operator  $A_i : F_{i,J} \rightarrow F_{i,J}$  as the restriction to  $F_{i,J}$  of  $\mathcal{A}_i$ , that is the unique isomorphism of  $F_{i,J}$  satisfying  $A_i \phi_{i,j} = \lambda_{i,j} \phi_{i,j}$ , for any  $j \leq J$ .

Let  $\mathcal{P}_{i,J}$  be the orthogonal projection from the space  $H^{-1}(\Omega_i)$  equipped with the inner product  $\langle \bullet, \bullet \rangle_{-1, \mathcal{A}_i}$  onto the space  $F_{i,J}$ .

We define also  $\mathcal{P}_J := \mathcal{P}_{1,J} \otimes \mathcal{P}_{2,J} \otimes \mathbf{I}$  which is an orthogonal projection from  $H^{-1}(\Omega) \otimes \mathbb{R}^n$  onto  $F_{1,J} \otimes F_{2,J} \otimes \mathbb{R}^n$ , with respect to  $\langle \bullet, \bullet \rangle_{-1, \mathcal{A}}$ .

**Remarque 3.4.1** *Since the eigenfunctions  $(\phi_{i,j})_{j \geq 1}$  are orthogonal with respect to the  $L^2(\Omega_i)$ ,  $H^{-1}(\Omega_i)$  and  $H_0^1(\Omega_i)$  inner products, we deduce that the projection  $\mathcal{P}_J$  (resp.  $\mathcal{P}_{i,J}$ ) is orthogonal with respect to  $\langle \bullet, \bullet \rangle_{s, \mathcal{A}}$  (resp.  $\langle \bullet, \bullet \rangle_{s, \mathcal{A}_i}$ ), for any  $s \in \{-1, 0, 1\}$ .*

Hypothesis (3.64) allows to apply Assumption 3.3.1 with  $E_1 = F_{1,J}$ ,  $U_1 = \mathbb{R}$ ,  $\mathcal{B}_1 = \Delta_1 r$ , where we have used the affine function  $r(x) = 1 - x$  (note that  $\mathcal{B}_1^* = -1_{\{0\}} \partial_{x_1}$ ) and  $\mu_{E_1} = +\infty$ . Moreover, thanks to the Lebeau-Robbiano's spectral inequality in the domain  $\Omega_2$ , (see [12, 13]), Assumption 3.3.2 is fulfilled with  $E_2 = F_{2,J}$ ,  $U_2 = F_{2,J}$ ,  $\mathcal{B}_2 : U_2 \rightarrow E_2$ ,  $\mathcal{B}_2 = \mathcal{P}_{2,J} 1_{\omega_2}$ , and  $\mu_{E_2} = +\infty$ . Indeed, the usual spectral inequality gives that there exists  $C > 0$  such that for any  $\mu > 0$ ,  $J \in \mathbb{N}^*$

$$\|\psi\|_{L^2(\Omega_2)}^2 \leq C e^{C\sqrt{\mu}} \int_{\Omega_2} (1_{\omega_2} \psi)^2, \quad \forall \psi \in \text{Span}(\phi_{2,j}, \lambda_{2,j} \leq \mu) \cap E_2.$$

Moreover,

$$\int_{\Omega_2} (1_{\omega_2} \psi)^2 = \int_{\Omega_2} (1_{\omega_2} \psi) \psi,$$

and according to Remark 3.4.1 and given that  $\psi \in E_2$ ,

$$\int_{\Omega_2} (1_{\omega_2} \psi) \psi = \int_{\Omega_2} (\mathcal{P}_{2,J} 1_{\omega_2} \psi) \psi = \int_{\Omega_2} (\mathcal{B}_2 \psi) \psi = \int_{\Omega_2} (\mathcal{B}_2^* \psi) \psi,$$

since  $\mathcal{B}_2$  is self-adjoint. It follows from the Cauchy-Schwarz inequality that

$$\|\psi\|_{L^2(\Omega_2)} \leq C e^{C\sqrt{\mu}} \|\mathcal{B}_2^* \psi\|_{L^2(\Omega_2)}, \quad \forall \psi \in \text{Span}(\phi_{2,j}, \lambda_{2,j} \leq \mu) \cap E_2.$$

Thus, for any fixed  $J$ , Theorem 3.3.1 applies with  $\mu^*$  as large as we want. Thus, letting  $\mu^*$  go to infinity, we finally build a null control for the projected finite dimensional problem denoted by  $v_J \in L^2(0, T; U_2) \otimes \mathbb{R}^m$  that drives solution  $y$  of (3.24) from  $P_J y_0$  to zero and which satisfies the following bound, uniformly with respect to  $J$ :

$$\|v_J\|_{L^2((0,T) \times \partial\Omega)^m} \leq e^{M_{\text{obs}}(1+\frac{1}{T}+T)} \|P_J y_0\|_{-1,A} \leq e^{M_{\text{obs}}(1+\frac{1}{T}+T)} \|y_0\|_{-1,A}. \quad (3.68)$$

where  $\|\bullet\|_{-1,A}^2 = \langle \bullet, \bullet \rangle_{-1,A} = \langle \bullet, \mathbf{A}^{-1} \bullet \rangle_{H^{-1}(\Omega), H_0^1(\Omega)}$ .

From (3.68) we can infer that there exists a subsequence of  $(v_J)_J$  denoted by  $(v_{J'})_{J'}$  which weakly converges in  $L^2((0, T) \times \partial\Omega)^m$  to a limit denoted by  $v$ , satisfying also

$$\|v\|_{L^2((0,T) \times \partial\Omega)^m} \leq e^{M_{\text{obs}}(1+\frac{1}{T}+T)} \|y_0\|_{-1,A}.$$

We claim that this limit drives the solution  $y$  of (3.65) from  $y_0$  to zero. Indeed, let  $k \in \{1, \dots, N\}$ ,  $i, j \geq 1$ , and  $J \geq \max(i, j)$ .

First, we have by definition of a solution of (3.24)

$$\begin{aligned} & - \left\langle \mathcal{P}_J y_0, e^{-T(\lambda_{1,i} + \lambda_{2,j})} \phi_{1,i} \otimes \phi_{2,j} \otimes (e^{-TC^*} e_k) \right\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ &= \left[ e^{-(T-\cdot)(\lambda_{1,i} + \lambda_{2,j})} (\mathbf{B}_1^* \phi_{1,i}) \otimes (\mathbf{B}_2^* \phi_{2,j}) \otimes (\mathbf{B}^* e^{-(T-\cdot)C^*} e_k), v_J \right]_{L^2((0,T) \times \partial\Omega)^m} \end{aligned} \quad (3.69)$$

Second, since  $J \geq i$  and  $J \geq j$ , and by definition of  $\mathcal{P}_J$ , we have

$$\begin{aligned} & - \left\langle \mathcal{P}_J y_0, e^{-T(\lambda_{1,i} + \lambda_{2,j})} \phi_{1,i} \otimes \phi_{2,j} \otimes (e^{-TC^*} e_k) \right\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ &= - \left\langle \mathcal{P}_J y_0, (\lambda_{1,i} + \lambda_{2,j}) e^{-T(\lambda_{1,i} + \lambda_{2,j})} \phi_{1,i} \otimes \phi_{2,j} \otimes (e^{-TC^*} e_k) \right\rangle_{-1,A} \\ &= - \left\langle y_0, (\lambda_{1,i} + \lambda_{2,j}) e^{-T(\lambda_{1,i} + \lambda_{2,j})} \phi_{1,i} \otimes \phi_{2,j} \otimes (e^{-TC^*} e_k) \right\rangle_{-1,A}. \end{aligned}$$

Hence, we get

$$\begin{aligned} & - \left\langle \mathcal{P}_J y_0, e^{-T(\lambda_{1,i} + \lambda_{2,j})} \phi_{1,i} \otimes \phi_{2,j} \otimes (e^{-TC^*} e_k) \right\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ &= - \left\langle y_0, e^{-T(\lambda_{1,i} + \lambda_{2,j})} \phi_{1,i} \otimes \phi_{2,j} \otimes (e^{-TC^*} e_k) \right\rangle_{H^{-1}(\Omega), H_0^1(\Omega)}. \end{aligned} \quad (3.70)$$

Third, since  $F_{2,J}$  is stable by  $\mathbf{A}_2$ ,  $\mathcal{P}_{2,J}$  is an orthogonal projection with respect to the  $L^2(\Omega_2)$  norm, see Remark 3.2.2. So, since  $v_J \in L^2(0, T; U_2) \otimes \mathbb{R}^m$  and  $U_2 = F_{2,J}$ , according to remark 3.4.1, we have for any  $t$

$$[\mathcal{B}_2^* \phi_{2,j}, v_J(t)]_{L^2(\Omega_2)^m} = [\mathcal{P}_{2,J} 1_{\omega_2} \phi_{2,j}, v_J(t)]_{L^2(\Omega_2)^m} = [1_{\omega_2} \phi_{2,j}, v_J(t)]_{L^2(\Omega_2)^m}.$$

Hence,

$$\begin{aligned} & \left[ (\mathbf{B}_1^* \phi_{1,i}) \otimes (\mathbf{B}_2^* \phi_{2,j}) \otimes (\mathbf{B}^* e^{-(T-\cdot)C^*} e_k), v_J \right]_{L^2((0,T) \times \partial\Omega)^m} \\ &= \left[ -(\phi_{1,i})'(0) \otimes 1_{\omega_2} \phi_{2,j} \otimes (\mathbf{B}^* e^{-(T-\cdot)C^*} e_k), v_J \right]_{L^2((0,T) \times \partial\Omega)^m}. \end{aligned} \quad (3.71)$$

Gathering (3.69), (3.70) and (3.71), we get

$$\begin{aligned} & - \left\langle y_0, e^{-T(\lambda_{1,i} + \lambda_{2,j})} \phi_{1,i} \otimes \phi_{2,j} \otimes (e^{-TC^*} e_k) \right\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ &= \left[ -\phi'_{1,i}(0) \otimes 1_{\omega_2} \phi_{2,j} \otimes (\mathbf{B}^* e^{-(T-\cdot)C^*} e_k), v_J \right]_{L^2((0,T) \times \partial\Omega)^m}. \end{aligned} \quad (3.72)$$

Considering the subsequence  $v_{J'}$  in (3.72) and letting  $J'$  go to infinity, we get (3.67) and this concludes the proof. ■

## 3.5 Appendix

### 3.5.1 Proof of Lemma 3.2.1

1.  $\Rightarrow$  2. For any  $q_T \in D^{-s}F_T$ , we set  $q(t) = e^{-(T-t)L^*} q_T$  and we consider the initial data  $y_0 := P_{F_0} D^s P_{F_0}^* q(0)$ ; by assumption there exists a control  $v$  and an associated solution  $y$  satisfying (3.18).

By (3.16) we find

$$\begin{aligned} \langle P_{F_0} D^s P_{F_0}^* q(0), q(0) \rangle_0 &= \langle y(T), q_T \rangle_0 - \int_0^T [v(t), B^* q(t)]_0 dt \\ \|P_{F_0}^*(q(0))\|_{s,D}^2 &= \langle y(T), D^s q_T \rangle_{-s,D} - \int_0^T [v(t), B^* q(t)]_0 dt. \end{aligned}$$

Since  $D^s q_T \in F_T$  and  $P_{F_T}$  is orthogonal with respect to  $\langle \bullet, \bullet \rangle_{-s,D}$ , we deduce

$$\begin{aligned} \|P_{F_0}^*(q(0))\|_{s,D}^2 &\leq \langle P_{F_T} y(T), D^s q_T \rangle_{-s,D} \\ &\quad + \llbracket v \rrbracket_{L^2(0,T;U)} \left( \int_0^T \llbracket B^* q(t) \rrbracket_0^2 dt \right)^{1/2}. \end{aligned}$$

The Cauchy-Schwarz inequality gives

$$\begin{aligned} \|P_{F_0}^*(q(0))\|_{s,D}^2 &\leq \left( \frac{1}{M_{\text{rel}}^2} \|P_{F_T} y(T)\|_{-s,D}^2 + \frac{1}{M_{\text{obs}}^2} \llbracket v \rrbracket_{L^2(0,T;U)}^2 \right)^{\frac{1}{2}} \\ &\quad \times \left( M_{\text{rel}}^2 \|D^s q_T\|_{-s,D}^2 + M_{\text{obs}}^2 \int_0^T \llbracket B^* q(t) \rrbracket_0^2 dt \right)^{\frac{1}{2}}. \end{aligned}$$

Using the estimate (3.18) given in the assumptions we get

$$\|P_{F_0}^*(q(0))\|_{s,D}^2 \leq \|y_0\|_{-s,D} \left( M_{\text{rel}}^2 \|D^s q_T\|_{-s,D}^2 + M_{\text{obs}}^2 \int_0^T \llbracket B^* q(t) \rrbracket_0^2 dt \right)^{\frac{1}{2}}.$$

Recall that we have chosen  $y_0 = P_{F_0} D^s P_{F_0}^* q(0)$ , and that  $P_{F_0}$  is the orthogonal projection onto  $F_0$  with respect to the inner product  $\langle \cdot, \cdot \rangle_{-s,D}$ . Therefore,

$$\|y_0\|_{-s,D} \leq \|D^s P_{F_0}^* q(0)\|_{-s,D} = \|P_{F_0}^* q(0)\|_{s,D}$$

and thus

$$\|P_{F_0}^*(q(0))\|_{s,D} \leq \left( M_{\text{rel}}^2 \|q_T\|_{s,D}^2 + M_{\text{obs}}^2 \int_0^T \llbracket B^* q(t) \rrbracket_0^2 dt \right)^{\frac{1}{2}},$$

which gives the claim.

2.  $\Rightarrow$  1. Let  $y_0$  be an element of  $F_0$ . We apply the quite usual penalized HUM approach (as described for instance in [6]) to find such a control  $v$ . To this end, we introduce the functional defined on  $D^{-s}F_T$  by

$$J(q_T) := \frac{1}{2} \int_0^T \llbracket B^* q(t) \rrbracket_0^2 dt + \frac{M_{\text{rel}}^2}{2M_{\text{obs}}^2} \|q_T\|_{s,D}^2 + \langle y_0, q(0) \rangle_0.$$

It is clear that  $J$  is continuous, strictly convex and coercive; we denote by  $q_T^{\text{opt}} \in D^{-s}F_T$  its minimizer and by  $q^{\text{opt}}(t) = e^{-(T-t)L^*} q_T^{\text{opt}}$  the solution to the backward equation with final condition  $q_T^{\text{opt}}$ . We set  $v(t) := B^* q^{\text{opt}}(t)$  and introduce the associated solution  $y$  to (3.11). Let us show the relation

$$P_{F_T}(y(T)) = -\frac{M_{\text{rel}}^2}{M_{\text{obs}}^2} D^s q_T^{\text{opt}}. \quad (3.73)$$

Take any  $q_T \in D^{-s}F_T$  and let  $q(t) = e^{-(T-t)L^*} q_T$  be the solution of the adjoint equation with final condition  $q_T$ . The associated Euler-Lagrange equation for  $J$  at the optimal point  $q_T^{\text{opt}}$  writes

$$\int_0^T [B^* q^{\text{opt}}(t), B^* q(t)]_0 dt + \frac{M_{\text{rel}}^2}{M_{\text{obs}}^2} \langle q_T^{\text{opt}}, q_T \rangle_{s,D} + \langle y_0, q(0) \rangle_0 = 0. \quad (3.74)$$

Combining (3.74) with (3.16) gives that for any  $q_T \in D^{-s}F_T$ ,

$$\langle y(T), q_T \rangle_0 = -\frac{M_{\text{rel}}^2}{M_{\text{obs}}^2} \langle q_T^{\text{opt}}, q_T \rangle_{s,D},$$

and thus,

$$\langle y(T), D^s q_T \rangle_{-s,D} = -\frac{M_{\text{rel}}^2}{M_{\text{obs}}^2} \langle D^s q_T^{\text{opt}}, D^s q_T \rangle_{-s,D}$$

and using that  $D^s q_T^{\text{opt}} \in F_T$ , and that  $D^s q_T$  is an arbitrary element in  $F_T$ , we finally find the equality (3.73).

Now we apply (3.74) with  $q_T = q_T^{\text{opt}}$  to get

$$\begin{aligned} \llbracket v \rrbracket_{L^2(0,T;U)}^2 + \frac{M_{\text{rel}}^2}{M_{\text{obs}}^2} \|q_T^{\text{opt}}\|_{s,D}^2 &= \langle y_0, q^{\text{opt}}(0) \rangle_0 \\ &= \langle y_0, P_{F_0}^* q^{\text{opt}}(0) \rangle_0 \\ &\leq \|y_0\|_{-s,D} \|P_{F_0}^* q^{\text{opt}}(0)\|_{s,D}, \end{aligned}$$

where the second equality comes from the fact that  $y_0 \in F_0$ . Using now the hypothesis (3.19), we obtain

$$\llbracket v \rrbracket_{L^2(0,T;U)}^2 + \frac{M_{\text{rel}}^2}{M_{\text{obs}}^2} \|q_T^{\text{opt}}\|_{s,D}^2 \leq \|y_0\|_{-s,D} \left( M_{\text{obs}}^2 \llbracket v \rrbracket_{L^2(0,T;U)}^2 + M_{\text{rel}}^2 \|q_T^{\text{opt}}\|_{s,D}^2 \right)^{1/2}$$

and thus

$$M_{\text{obs}}^2 \llbracket v \rrbracket_{L^2(0,T;U)}^2 + M_{\text{rel}}^2 \|q_T^{\text{opt}}\|_{s,D}^2 \leq M_{\text{obs}}^4 \|y_0\|_{-s,D}^2. \quad (3.75)$$

Note that taking the norm  $\|\cdot\|_{s,D}$  in (3.73) gives

$$\|q_T^{\text{opt}}\|_{s,D}^2 = \frac{M_{\text{obs}}^4}{M_{\text{rel}}^4} \|P_{F_T} y(T)\|_{-s,D}^2$$

so that with (3.75) we end up with

$$M_{\text{obs}}^2 \llbracket v \rrbracket_{L^2(0,T;U)}^2 + \frac{M_{\text{obs}}^4}{M_{\text{rel}}^2} \|P_{F_T} y(T)\|_{-s,D}^2 \leq M_{\text{obs}}^4 \|y_0\|_{-s,D}^2$$

and the claim is proved.

### 3.5.2 Numerical Illustrations

In this section we give a numerical illustration of our theorem using the HUM approach (see section 1.1.3 of the introduction and section 2.7.5 of chapter 2). We consider a discretization in time of system (3.4) of coupled heat equations on a unit square with a uniform discretization of 50 mesh points in both directions and 100 times steps. The time horizon is  $T = 0.5$ , the coupling coefficient is equal to 1 and the initial condition of the controlled equation is  $\alpha_0 = 0$  and the initial condition of the second component is  $\beta_0 = \sin(\pi x) \sin(\pi y)$ . The control zone is located on three edges of the square. In this fully discretized framework, we make the parameter  $\varepsilon$  of the functional  $F_\varepsilon$  depend on the mesh parameter  $h$  and we set  $\varepsilon = h^4$ . We give two sets of simulations below for different times  $t \in \{0, T/3, 2T/3, T\}$ .

In figures 3.3, we set the control to zero and we plot solutions  $\alpha$  and  $\beta$ . We see on figure 3.3d that the second component  $\beta$  of the system at time  $T$  is not equal to zero since the solution only decreases because of the dissipation of the heat equation. The first component  $\alpha$  remains equal to zero since its initial condition is zero and it is not controlled.

In the second set of figures 3.4, however, we see how the control affects the first component of the system  $\alpha$  so that it can control the second component of the system  $\beta$ . The solutions are eventually very close to zero.

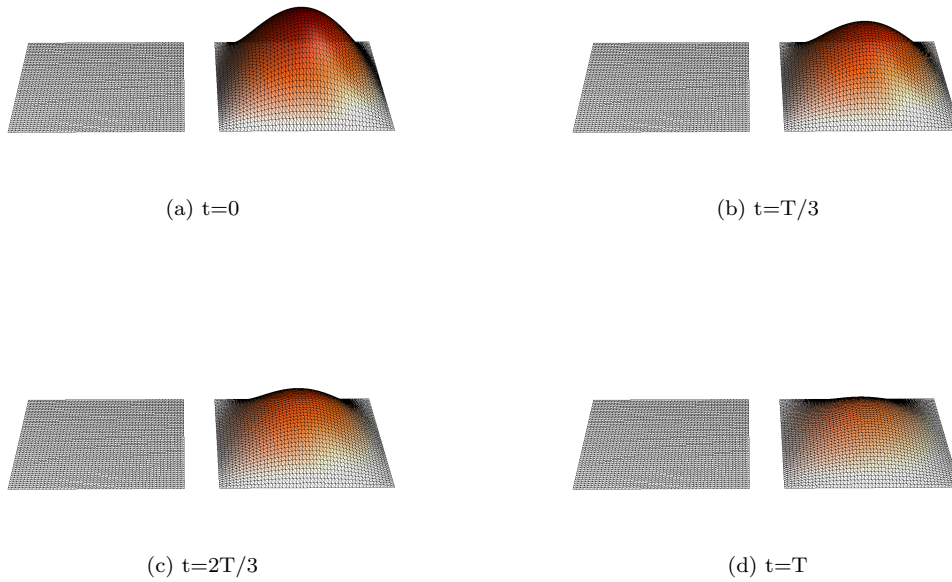


Figure 3.3 – Component  $\alpha$  (left) and  $\beta$  (right) of system (3.4) with no control

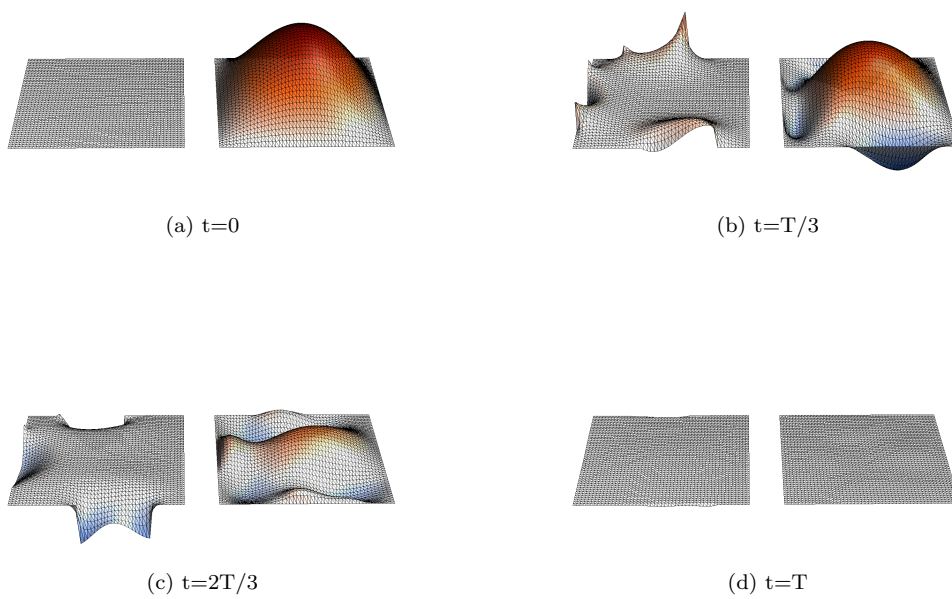


Figure 3.4 – Component  $\alpha$  (left) and  $\beta$  (right) of system (3.4) with a boundary control

### 3.6 References

- [1] D. ALLONSIUS, F. BOYER, AND M. MORANCEY, *Spectral analysis of discrete elliptic operators and applications in control theory*, Numerische Mathematik, (2018), doi:10.1007/s00211-018-0983-1, <https://doi.org/10.1007/s00211-018-0983-1>. 105, 113, 114, 115
- [2] F. AMMAR-KHODJA, A. BENABDALLAH, M. GONZÁLEZ-BURGOS, AND L. TERESA, *Recent results on the controllability of linear coupled parabolic problems: A survey*, Mathematical Control and Related Fields, 1 (2011), pp. 267–306, doi:10.3934/mcrf.2011.1.267. 89
- [3] F. AMMAR KHODJA, F. CHOULY, AND M. DUPREZ, *Partial null controllability of parabolic linear systems*, Math. Control Relat. Fields, 6 (2016), pp. 185–216, <https://doi.org/10.3934/mcrf.2016001>. 94, 95
- [4] A. BENABDALLAH, F. BOYER, M. GONZÁLEZ-BURGOS, AND G. OLIVE, *Sharp estimates of the one-dimensional boundary control cost for parabolic systems and application to the  $n$ -dimensional boundary null controllability in cylindrical domains*, SIAM Journal on Control and Optimization, 52 (2014), pp. 2970–3001, doi:10.1137/130929680, <http://dx.doi.org/10.1137/130929680>. 88, 89, 90, 92, 113, 114, 115
- [5] K. BHANDARI AND F. BOYER, *Boundary null-controllability of one-dimensional coupled parabolic system with robin conditions*, in preparation, (2018). 115
- [6] F. BOYER, *On the penalised HUM approach and its applications to the numerical approximation of null-controls for parabolic problems*, in CANUM 2012, 41e Congrès National d'Analyse Numérique, vol. 41 of ESAIM Proc., EDP Sci., Les Ulis, 2013, pp. 15–58. 91, 94, 95, 107, 118
- [7] F. BOYER, F. HUBERT, AND J. L. ROUSSEAU, *Discrete carleman estimates for elliptic operators and uniform controllability of semi-discretized parabolic equations*, Journal de Mathématiques Pures et Appliquées, 93 (2010), pp. 240–276. 105, 113
- [8] F. BOYER, F. HUBERT, AND J. L. ROUSSEAU, *Discrete carleman estimates for elliptic operators in arbitrary dimension and applications*, SIAM Journal on Control and Optimization, 48 (2010), pp. 5357–5397. 94, 105
- [9] F. BOYER AND J. L. ROUSSEAU, *Carleman estimates for semi-discrete parabolic operators and application to the controllability of semi-linear semi-discrete parabolic equations*, Annales de l'Institut Henri Poincaré (C) Non Linear Analysis, 31 (2014), pp. 1035–1078. 105
- [10] J.-M. CORON, *Control and nonlinearity*, vol. 136 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2007. 94
- [11] S. LABBÉ AND E. TRÉLAT, *Uniform controllability of semidiscrete approximations of parabolic control systems*, Systems Control Lett., 55 (2006), pp. 597–609. 94
- [12] J. LE ROUSSEAU AND G. LEBEAU, *On carleman estimates for elliptic and parabolic operators. applications to unique continuation and control of parabolic equations*, ESAIM: COCV, 18 (2011), pp. 712–747, doi:10.1051/cocv/2011168. 116
- [13] G. LEBEAU AND L. ROBBIANO, *Contrôle exact de l'équation de la chaleur*, Séminaire Équations aux dérivées partielles (Polytechnique), (1994-1995), pp. 1–11, <http://eudml.org/doc/112118>. 92, 116
- [14] L. MILLER, *A direct Lebeau-Robbiano strategy for the observability of heat-like semigroups*, Discrete Contin. Dyn. Syst. Ser. B, 14 (2010), pp. 1465–1485, <https://doi.org/10.3934/dcdsb.2010.14.1465>. 94
- [15] M. TUCSNAK AND G. WEISS, *Observation and control for operator semigroups*, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser Verlag, Basel, 2009, doi:10.1007/978-3-7643-8994-9, <http://dx.doi.org/10.1007/978-3-7643-8994-9>. 94, 107
- [16] E. ZUAZUA, *Control and numerical approximation of the wave and heat equations*, International Congress of Mathematicians, Madrid, Spain, III (2006), pp. 1389–1417. 91, 94





# Chapitre 4

## Étude du temps minimal de contrôle à zéro de l'équation de Grushin

*” Il n’y a pas des problèmes qu’on se pose, il y a des problèmes qui se posent. Il n’y a pas de problèmes résolus, il y a seulement des problèmes plus ou moins résolus. ”*

Henri Poincaré

### Sommaire

---

<b>4.1</b>	<b>Plan du Chapitre</b>	<b>124</b>
<b>4.2</b>	<b>Contrôle à zéro de l'équation de Grushin lorsque <math>\{0\} \times (0, 1) \subset \omega</math></b>	<b>125</b>
<b>4.3</b>	<b>Quelques résultats sur l'équation de Grushin</b>	<b>129</b>
<b>4.4</b>	<b>Étude du temps minimal sur une bande lorsque <math>\gamma = 1</math></b>	<b>134</b>
4.4.1	Plan de la section 4.4	134
4.4.2	Analyse spectrale des opérateurs $\mathcal{A}_n$ sur $(-1, 1)$ , étude de la propriété de gap	135
4.4.3	Existence de la fonction $\mathcal{N}$	146
4.4.4	Étude de la norme $L^2$ des vecteurs propres sur $(a, b)$	147
4.4.5	Conclusion	148
<b>4.5</b>	<b>Annexes</b>	<b>149</b>
4.5.1	Résultats concernant les fonctions propres de l'oscillateur harmonique (4.36)	149
4.5.2	Calculs de la preuve du Théorème 4.4.5	153
4.5.3	A propos de la Remarque 4.4.5	154
<b>4.6</b>	<b>Références</b>	<b>156</b>

---

Ce chapitre résume un travail effectué sur la conjecture formulée dans [2] (section 5, Conclusion and open problems). Celle-ci porte sur la valeur du temps minimal  $T^*$  de contrôlabilité à zéro de l'équation de Grushin, avec comme paramètre de dégénérescence  $\gamma = 1$ , posée sur un domaine rectangulaire et dont le domaine de contrôle est une bande verticale. Il est conjecturé que  $T^* = \frac{a^2}{2}$  où  $a$  est la distance du domaine de contrôle à l'axe  $Oy$ .

## 4.1 Plan du Chapitre

Nous nous intéressons au problème de contrôle de l'équation de Grushin suivant :

$$\begin{cases} \partial_t f(t) + \mathcal{L}_\gamma f(t) = 1_\omega u(t), & \text{dans } Q \\ f(t) = 0, & \text{dans } (0, T) \times \partial\Omega \\ f(0) = f^0, & \text{dans } \Omega. \end{cases} \quad (4.1)$$

- L'opérateur  $\mathcal{L}_\gamma$  est un opérateur elliptique :  $\mathcal{L}_\gamma := -\partial_{xx} - |x|^{2\gamma}\partial_{yy}$ , et  $\gamma > 0$  est un paramètre réel.
- L'espace physique  $\Omega$  est un sous ensemble de  $\mathbb{R}^2$  défini par  $\Omega := (-1, 1) \times (0, 1)$  et  $\omega$  est un ouvert de  $\Omega$ .
- On note  $Q := (0, T) \times \Omega$ .
- le contrôle  $u$  appartient à l'espace  $L^2(Q)$ .
- Le temps final  $T > 0$  est une donnée du problème.
- La donnée initiale vérifie  $f^0 \in L^2(\Omega)$ .

L'équation (4.1) est-elle *contrôlable à zéro au temps  $T$*  ?

Comme énoncé dans la Proposition 1.1.1, étant donné un paramètre  $\gamma$ , un ouvert de contrôle  $\omega$  et un temps  $T > 0$ , la contrôlabilité à zéro de (4.1) est équivalente à l'inégalité *d'observabilité* associée. Ici, étant donné  $g$  la solution de l'équation

$$\begin{cases} \partial_t g + \mathcal{L}_\gamma g = 0, & \text{dans } (0, T) \times \Omega \\ g = 0, & \text{dans } (0, T) \times \partial\Omega \\ g(0) = g^0, & \text{dans } \Omega, \end{cases} \quad (4.2)$$

l'inégalité d'observabilité s'écrit :

$$\exists C > 0, \forall g^0 \in L^2(\Omega), \int_\Omega |g(T)|^2 \leq C^2 \int_0^T \int_\omega |g(t)|^2. \quad (4.3)$$

Sous certaines conditions sur  $\gamma$  et sur la géométrie de l'ouvert de contrôle, cette équation, bien que de nature parabolique, a la particularité de posséder *un temps minimal de contrôle à zéro*. C'est-à-dire qu'il peut exister un temps final  $T > 0$  tel que

- lorsque  $T < T^*$ , le système (4.1) n'est pas contrôlable,
- lorsque  $T > T^*$ , le système (4.1) est contrôlable.

Dans la configuration suivante :

$$\gamma = 1 \text{ et } \omega = (a, b) \times (0, 1) \text{ où } 0 < a < b \leq 1, \quad (4.4)$$

un temps minimal  $T^*$  non nul existe. On s'intéressera donc particulièrement à l'hypothèse (4.4) par la suite (voir la figure 1.4 du chapitre (1)).

Détaillons l'articulation des sections de ce Chapitre. Dans la section 4.2 nous montrons pourquoi l'approche classique exposée dans [6, prop. 2.67] de contrôlabilité à zéro de l'équation de la chaleur ne s'applique pas à l'opérateur dégénéré  $\mathcal{L}_\gamma$ . Ensuite, nous rappelons quelques résultats existants sur ce type de problème dans la section 4.3. Enfin dans la dernière partie, nous proposons une approche basée sur la méthode des moments pour démontrer que  $T^* = \frac{a^2}{2}$ , sous l'hypothèse (4.4) (voir la section 4.4). Cette question a été résolue très récemment (voir [3]) par une méthode différente de celle que

nous proposons. Nous expliquerons leur méthode dans la section 4.3.

L'approche que nous proposons dans la section 4.4 pour calculer le temps minimal  $T^*$  repose sur l'étude du spectre d'une famille d'opérateurs de Sturm Liouville  $(\mathcal{A}_n)_{n \in \mathbb{N}^*}$ . Pour tout  $n \in \mathbb{N}^*$ , le spectre de l'opérateur  $\mathcal{A}_n$  est discret et est noté  $(\lambda_{k,n})_{k \in \mathbb{N}^*}$ . Nous démontrons dans la section 4.4 l'équivalence entre l'égalité  $T^* = \frac{\alpha^2}{2}$  et la propriété de gap uniforme sur les valeurs propres  $(\lambda_{k,n})_{k,n \in \mathbb{N}^*}$ , c'est-à-dire :

$$\exists \rho > 0, \forall k, n \in \mathbb{N}^*, \lambda_{k+1,n} - \lambda_{k,n} > \rho. \quad (4.5)$$

Nous prouvons dans le Théorème 4.4.1 que pour tout  $\tau \in (0, 1)$  et tout  $n \in \mathbb{N}^*$ , la propriété de gap uniforme (4.5) est valable pour les indices

$$1 \leq k < \left\lfloor \tau \frac{\pi}{2} n \right\rfloor \text{ et } k > \left\lceil \frac{\pi}{2} n \right\rceil.$$

Ainsi, notre approche n'est pas complète car les fréquences  $k$  comprises dans l'intervalle  $\llbracket \left\lfloor \tau \frac{\pi}{2} n \right\rfloor, \left\lceil \frac{\pi}{2} n \right\rceil \rrbracket$  ne sont pas couvertes par notre étude. Cependant, des résultats numériques de la section 4.4.2 viennent compléter le Théorème 4.4.1 et montrent que la limitation induite par le coefficient  $\tau$  est purement technique. Nous conjecturons donc que la propriété (4.5) est vraie.

**Remarque 4.1.1** *Pour les questions d'existence, d'unicité et de stabilité des solutions de l'équation (4.1), nous renvoyons à la section 2 de l'article [2].*

## 4.2 Contrôle à zéro de l'équation de Grushin lorsque $\{0\} \times (0, 1) \subset \omega$

Tout d'abord, notons que l'opérateur  $\mathcal{L}_\gamma$  est *dégénéré* c'est-à-dire qu'il ne vérifie pas l'hypothèse d'ellipticité uniforme

$$\exists \theta > 0, \forall \xi \in \mathbb{R}^2, \text{ p.s. } x \in \Omega, {}^t \xi M_{\mathcal{L}_\gamma}(x) \xi \geq \theta |\xi|^2, \quad (4.6)$$

où  $M_{\mathcal{L}_\gamma} := \text{diag}(1, |x|^{2\gamma})$  et  $|\cdot|$  désigne la norme euclidienne de  $\mathbb{R}^2$ . En effet, il suffit de remarquer que lorsque  $\xi := {}^t(0, 1)$ ,  ${}^t \xi M(x) \xi = |x|^{2\gamma} \rightarrow_{x \rightarrow 0} 0$ .

Le but de cette section est d'expliquer en quoi l'hypothèse de dégénérescence de l'opérateur  $\mathcal{L}_\gamma$  amène une difficulté supplémentaire quant au problème du contrôle à zéro de (4.1). Pour ce faire, nous commençons par démontrer que (4.1) est contrôlable à zéro en tout temps  $T > 0$ , par la méthode de Carleman, dans le cas particulier où l'opérateur  $\mathcal{L}_\gamma$  n'est pas dégénéré dans  $\Omega \setminus \omega$ . Nous choisissons pour cela un ouvert de contrôle  $\omega$  quelconque contenant le segment  $\{0\} \times (0, 1)$  comme illustré sur la figure ci-dessous. Ensuite, nous mettons en évidence comment l'hypothèse (4.6) intervient dans la preuve de la Proposition 2.67 de [6] afin d'expliquer pourquoi la méthode fait défaut dans le cas où  $\mathcal{L}_\gamma$  est dégénéré dans  $\Omega \setminus \omega$ , et ne vérifie donc pas (4.6).

**Remarque 4.2.1** *Le Théorème 4.2.1, dans lequel  $\omega$  contient le segment  $\{0\} \times (0, 1)$ , est également démontré dans l'Appendice de [2]. Les auteurs découpent l'ouvert  $\Omega$  en trois régions et traitent l'équation sur ces différentes zones. Ils rassemblent ensuite la solution sur le domaine complet en utilisant des fonctions plateaux.*

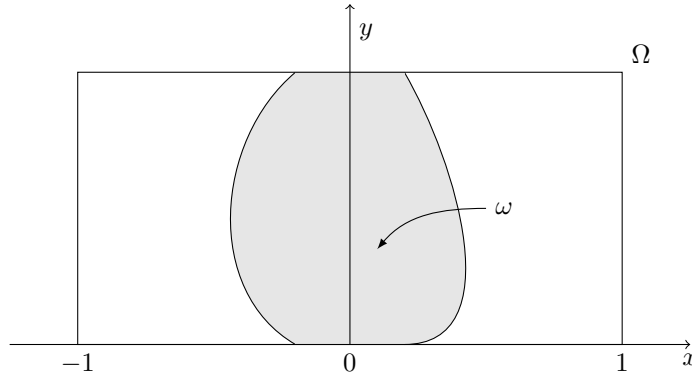


FIGURE 4.1 – Une zone de contrôle  $\omega$  contenant l'axe Oy pour le problème de contrôle de l'équation de Grushin.

**Théorème 4.2.1** Soient  $T > 0$ ,  $\gamma > 0$ ,  $a, b \in (-1, 1)$  tels que  $a < 0$  et  $b > 0$  et  $\omega$  un ouvert de contrôle tel que  $(a, b) \times (0, 1) \subset \omega$ . L'équation de Grushin (4.1) est contrôlable à zéro au temps  $T$ .

**Preuve**

La démonstration est classique et consiste à prouver l'inégalité d'observabilité (4.3).

La preuve repose sur la méthode de Carleman présentée dans [6] et s'appuie sur le Lemme suivant (prouvé dans [6, Lem. 2.68 p.80]):

**Lemme 4.2.1** Soit un ouvert non vide  $\omega_0 \subset \Omega$ . Il existe une fonction  $\Psi \in C^2(\bar{\Omega})$  telle que

$$\Psi(x, y) > 0, \forall (x, y) \in \Omega, \text{ et } \Psi(x, y) = 0 \text{ sur } \partial\Omega, \quad (4.7)$$

$$|\nabla\Psi(x, y)| > 0, \forall (x, y) \in \bar{\Omega} \setminus \omega_0, \quad (4.8)$$

Nous choisissons donc un ouvert non vide  $\omega_0 \subset \omega$  contenant la droite  $\{x = 0\}$  et nous posons

$$\begin{cases} \alpha(t, x, y) = \frac{e^{2\lambda\|\Psi\|_{C^0(\bar{\Omega})}} - e^{\lambda\Psi(x, y)}}{t(T-t)} > 0 \\ \phi(t, x, y) = \frac{e^{\lambda\Psi(x, y)}}{t(T-t)} \\ z(t, x, y) = e^{-s\alpha(t, x, y)} f(t, x, y), \forall t \in (0, T), \text{ et } z(T) = z(0) = 0, \end{cases}$$

avec  $s$  et  $\lambda$  deux paramètres que l'on fixera ultérieurement. Prenons  $w \in \{x, y\}$ ,

$$\partial_w \alpha = -\lambda\phi\partial_w \Psi,$$

Calculons les dérivées partielles de la fonction  $f$  :

$$\begin{cases} \partial_t f = e^{s\alpha} (sz\partial_t \alpha + \partial_t z) \\ -\partial_w f = -e^{s\alpha} \partial_w z + s\lambda\phi e^{s\alpha} z \partial_w \Psi \\ -\partial_{ww} f = e^{s\alpha} (-\partial_{ww} z + 2s\lambda\phi\partial_w z \partial_w \Psi + s\lambda\phi z \partial_{ww} \Psi \\ \quad + s\lambda^2 \phi z (\partial_w \Psi)^2 - s^2 \lambda^2 \phi^2 z (\partial_w \Psi)^2) \end{cases} \quad (4.9)$$

Nous définissons maintenant  $\tilde{\nabla}$  l'opérateur différentiel  $\tilde{\nabla} f = {}^t(\partial_x f, |x|^\gamma \partial_y f)$ . Les calculs (4.9) donnent

$$\begin{aligned} \mathcal{L}_\gamma f &= e^{s\alpha} \left( \mathcal{L}_\gamma z + 2s\lambda\phi\tilde{\nabla} z \tilde{\nabla} \Psi - s\lambda\phi z \mathcal{L}_\gamma \Psi + s\lambda^2 \phi z |\tilde{\nabla} \Psi|^2 - s^2 \lambda^2 \phi^2 z |\tilde{\nabla} \Psi|^2 \right) \\ &= e^{s\alpha} \left( \mathcal{L}_\gamma z - s^2 \lambda^2 \phi^2 z |\tilde{\nabla} \Psi|^2 + 2s\lambda\phi\tilde{\nabla} z \tilde{\nabla} \Psi - s\lambda^2 \phi z |\tilde{\nabla} \Psi|^2 - s\lambda\phi z \mathcal{L}_\gamma \Psi + 2s\lambda^2 \phi z |\tilde{\nabla} \Psi|^2 \right) \end{aligned}$$

En utilisant que  $f$  est solution de

$$\partial_t f + \mathcal{L}_\gamma f = 0,$$

on obtient que

$$P_1 + P_2 = P_3, \quad (4.10)$$

avec

$$P_1 = sz\partial_t \alpha + \mathcal{L}_\gamma z - s^2 \lambda^2 \phi^2 z |\tilde{\nabla} \Psi|^2$$

et

$$P_2 = \partial_t z + 2s\lambda\phi\tilde{\nabla} z \tilde{\nabla} \Psi - s\lambda^2 \phi z |\tilde{\nabla} \Psi|^2$$

$$P_3 = s\lambda\phi z \mathcal{L}_\gamma \Psi - 2s\lambda^2 \phi z |\tilde{\nabla} \Psi|^2$$

En élevant au carré (4.10), puis en intégrant sur  $Q = (0, T) \times \Omega$ , et enfin en utilisant l'inégalité  $2ab \leq (a+b)^2$ , on trouve

$$2 \iint_Q P_1 P_2 \leq \iint_Q P_3^2$$

Examinons un à un les 9 termes de  $2 \iint_Q P_1 P_2$ . On définit  $\gamma_1 = 0$  et  $\gamma_2 = \gamma$ , et on adopte la convention d'Einstein de sommation des indices dans les calculs qui suivent.

$$1. I_{1,1} := \iint_Q 2sz (\partial_t z) (\partial_t \alpha) = \iint_Q s (\partial_t z^2) (\partial_t \alpha) = - \iint_Q s (\partial_{tt} \alpha) z^2$$

2.

$$\begin{aligned} I_{1,2} &:= \iint_Q 4s^2 \lambda \phi z (\partial_t \alpha) \tilde{\nabla} z \tilde{\nabla} \Psi = \iint_Q 2s^2 \lambda \phi (\partial_t \alpha) \tilde{\nabla} z^2 \tilde{\nabla} \Psi \\ &= - \iint_Q 2s^2 \lambda |x|^{\gamma_i} \partial_{w_i} (\phi (\tilde{\nabla} \Psi)_i \partial_t \alpha) z^2 \end{aligned}$$

$$3. I_{1,3} := \iint_Q -2s^2 \lambda^2 \phi (\partial_t \alpha) |\tilde{\nabla} \Psi|^2 z^2$$

$$4. I_{2,1} := \iint_Q (\mathcal{L}_\gamma z) (\partial_t z) = - \iint_Q z (\partial_t \mathcal{L}_\gamma z) = - \iint_Q z (\mathcal{L}_\gamma \partial_t z) = - \iint_Q (\mathcal{L}_\gamma z) (\partial_t z).$$

Donc  $\iint_Q (\mathcal{L}_\gamma z) (\partial_t z) = 0$ .

5.

$$\begin{aligned} I_{2,2} &:= \iint_Q 4s\lambda (\mathcal{L}_\gamma z) \phi \tilde{\nabla} \Psi \tilde{\nabla} z = \int_0^T \int_{\partial\Omega} -2s\lambda \phi \left| \frac{\partial z}{\partial n} \right|^2 \frac{\partial \Psi}{\partial n} - \iint_Q 2s\lambda \widetilde{\text{div}}(\phi \tilde{\nabla} \Psi) |\tilde{\nabla} z|^2 \\ &\quad + \iint_Q 4s\lambda (\tilde{\nabla} z)_i (\tilde{\nabla} z)_j |x|^{\gamma_j} \partial_{w_j} (\phi |x|^{\gamma_i} \partial_{w_i} \Psi) + \iint_Q 4s\lambda (\partial_x z) 2\gamma |x|^{2\gamma-1} \partial_y \Psi \partial_y z \end{aligned}$$

(voir les calculs ci-dessous)

$$\begin{aligned} 6. I_{2,3} &:= - \iint_Q 4s\lambda^2 z (\mathcal{L}_\gamma z) |\tilde{\nabla} \Psi|^2 \phi = \iint_Q 2s\lambda^2 |\tilde{\nabla} z|^2 |\tilde{\nabla} \Psi|^2 \phi + \iint_Q 2s\lambda^2 z \tilde{\nabla} z \tilde{\nabla} (|\tilde{\nabla} \Psi|^2 \phi) \\ &= \iint_Q 2s\lambda^2 |\tilde{\nabla} z|^2 |\tilde{\nabla} \Psi|^2 \phi - \iint_Q s\lambda^2 \mathcal{L}_\gamma (|\tilde{\nabla} \Psi|^2 \phi) z^2 \end{aligned}$$

$$7. I_{3,1} := - \iint_Q 2s^2 \lambda^2 \phi^2 |\tilde{\nabla} \Psi|^2 z \partial_t z = \iint_Q 2s^2 \lambda^2 \phi (\partial_t \phi) |\tilde{\nabla} \Psi|^2 z^2$$

$$\begin{aligned} 8. I_{3,2} &:= - \iint_Q 4s^3 \lambda^3 \phi^3 |\tilde{\nabla} \Psi|^2 z \tilde{\nabla} z \tilde{\nabla} \Psi = - \iint_Q 2s^3 \lambda^3 \phi^3 |\tilde{\nabla} \Psi|^2 \tilde{\nabla} z^2 \tilde{\nabla} \Psi \\ &= \iint_Q 2s^3 \lambda^3 |x|^{\gamma_i} \partial_{w_i} (\phi^3 |\tilde{\nabla} \Psi|^2 (\tilde{\nabla} \Psi)_i) z^2 \end{aligned}$$

$$9. I_{3,3} := \iint_Q 2s^3 \lambda^4 \phi^3 z^2 |\tilde{\nabla} \Psi|^4$$

Le terme numéro 5,  $I_{2,2}$ , mérite un traitement à part :

$$\iint_Q 4s\lambda (\mathcal{L}_\gamma z) \phi \tilde{\nabla} \Psi \tilde{\nabla} z = 4s\lambda \left( \iint_Q \tilde{\nabla} z \tilde{\nabla} (\phi \tilde{\nabla} \Psi \tilde{\nabla} z) - \int_0^T \int_{\partial\omega_0} \frac{\partial z}{\partial n} \phi \tilde{\nabla} \Psi \tilde{\nabla} z \right).$$

où  $\frac{\partial}{\partial n} = n \cdot \tilde{\nabla}$ , où  $n$  est la normale sortante unitaire du domaine  $\Omega$  au point du bord considéré. La fonction intégrée du premier terme du second membre s'écrit

$$\begin{aligned} \tilde{\nabla} z \tilde{\nabla} (\phi \tilde{\nabla} \Psi \tilde{\nabla} z) &= (\partial_x z) \partial_x (\phi \partial_x \Psi \partial_x z + |x|^{2\gamma} \phi \partial_y \Psi \partial_y z) + |x|^{2\gamma} (\partial_y z) \partial_y (\phi \partial_x \Psi \partial_x z + |x|^{2\gamma} \phi \partial_y \Psi \partial_y z) \\ &= (\partial_x z) (\phi \partial_x \Psi \partial_{xx} z + |x|^{2\gamma} \phi \partial_y \Psi \partial_{xy} z) + |x|^{2\gamma} (\partial_y z) (\phi \partial_x \Psi \partial_{yx} z + |x|^{2\gamma} \phi \partial_y \Psi \partial_{yy} z) \\ &\quad + (\partial_x z) (\partial_x (\phi \partial_x \Psi) \partial_x z + |x|^{2\gamma} \partial_x (\phi \partial_y \Psi) \partial_y z) + |x|^{2\gamma} (\partial_y z) (\partial_y (\phi \partial_x \Psi) \partial_x z + |x|^{2\gamma} \partial_y (\phi \partial_y \Psi) \partial_y z) \\ &\quad + (\partial_x z) 2\gamma |x|^{2\gamma-1} \phi \partial_y \Psi \partial_y z \\ &= \frac{1}{2} (\phi \partial_x \Psi \partial_x (\partial_x z)^2 + |x|^{2\gamma} \phi \partial_y \Psi \partial_y (\partial_x z)^2) + \frac{1}{2} |x|^{2\gamma} (\phi \partial_x \Psi \partial_x (\partial_y z)^2 + |x|^{2\gamma} \phi \partial_y \Psi \partial_y (\partial_y z)^2) \\ &\quad + (\partial_x (\phi \partial_x \Psi) (\partial_x z)^2 + |x|^{2\gamma} \partial_x (\phi \partial_y \Psi) \partial_x z \partial_y z) + |x|^{2\gamma} (\partial_y (\phi \partial_x \Psi) (\partial_y z) (\partial_x z) + |x|^{2\gamma} \partial_y (\phi \partial_y \Psi) (\partial_y z)^2) \\ &\quad + (\partial_x z) 2\gamma |x|^{2\gamma-1} \partial_y \Psi \partial_y z. \end{aligned}$$

Ainsi, en notant  $\widetilde{\text{div}} f = \partial_x f + |x|^\gamma \partial_y f$ , on trouve :

$$\begin{aligned} \iint_Q \tilde{\nabla} z \tilde{\nabla} (\phi \tilde{\nabla} \Psi \tilde{\nabla} z) &= \frac{1}{2} \int_0^T \int_{\partial\Omega} |\tilde{\nabla} z|^2 \phi \frac{\partial \Psi}{\partial n} - \frac{1}{2} \iint_Q \widetilde{\text{div}}(\phi \tilde{\nabla} \Psi) |\tilde{\nabla} z|^2 + \int_Q (\tilde{\nabla} z)_i (\tilde{\nabla} z)_j \partial_{w_j} (\phi \partial_{w_i} \Psi) \\ &\quad + \iint_Q (\partial_x z) 2\gamma |x|^{2\gamma-1} \partial_y \Psi \partial_y z. \end{aligned}$$

et donc

$$\begin{aligned} \iint_Q 4s\lambda (\mathcal{L}_\gamma z) \phi \tilde{\nabla} \Psi \tilde{\nabla} z &= \int_0^T \int_{\partial\Omega} -4s\lambda \frac{\partial z}{\partial n} \phi \tilde{\nabla} \Psi \tilde{\nabla} z + 2s\lambda |\tilde{\nabla} z|^2 \phi \frac{\partial \Psi}{\partial n} \\ &\quad - \iint_Q 2s\lambda \widetilde{\text{div}}(\phi \tilde{\nabla} \Psi) |\tilde{\nabla} z|^2 + \int_Q 4s\lambda (\tilde{\nabla} z)_i (\tilde{\nabla} z)_j \partial_{w_j} (\phi \partial_{w_i} \Psi) \\ &\quad + \iint_Q 4s\lambda (\partial_x z) 2\gamma |x|^{2\gamma-1} \partial_y \Psi \partial_y z. \end{aligned} \tag{4.11}$$

Comme  $z$  et  $\Psi$  sont nuls au bord, on a

$$\begin{aligned} \int_0^T \int_{\partial\Omega} -4s\lambda \frac{\partial z}{\partial n} \phi \tilde{\nabla} \Psi \tilde{\nabla} z + 2s\lambda |\tilde{\nabla} z|^2 \phi \frac{\partial \Psi}{\partial n} &= \int_0^T \int_{\partial\Omega} -4s\lambda \frac{\partial z}{\partial n} \phi \frac{\partial \Psi}{\partial n} \frac{\partial z}{\partial n} + 2s\lambda \left| \frac{\partial z}{\partial n} \right|^2 \phi \frac{\partial \Psi}{\partial n} \\ &= \int_0^T \int_{\partial\Omega} -2s\lambda \phi \left| \frac{\partial z}{\partial n} \right|^2 \frac{\partial \Psi}{\partial n} \end{aligned}$$

Et de plus,

$$\int_0^T \int_{\partial\Omega} -2s\lambda \phi \left| \frac{\partial z}{\partial n} \right|^2 \frac{\partial \Psi}{\partial n} \geq 0$$

car  $\Psi$  est nulle au bord et strictement positive dans  $\Omega$ .

Nous allons maintenant estimer les intégrales  $\iint_Q P_1 P_2$  et  $\iint_Q P_3^2$  en vue d'obtenir (4.3). La remarque suivante est cruciale car elle explique pourquoi la méthode de Carleman ne fonctionne que dans le cas où  $x = 0$  est contenu dans  $\omega$ .

**Remarque 4.2.2** Comme  $\Psi \in C^1(\bar{\Omega})$ , la condition (4.8) implique qu'il existe une constante  $C > 0$  telle que

$$|\nabla \Psi(x, y)| > C, \forall (x, y) \in \bar{\Omega} \setminus \omega_0, \quad (4.12)$$

qui à son tour implique l'existence d'une autre constante  $C > 0$  telle que

$$|\tilde{\nabla} \Psi(x, y)| > C, p.p. (x, y) \in \bar{\Omega} \setminus \omega_0, \quad (4.13)$$

car  $\bar{\Omega} \setminus \omega_0$  est à distance non nulle de la droite  $x = 0$ . Comme nous le verrons par la suite, l'inégalité (4.13) est nécessaire pour pouvoir appliquer la technique de Carleman dans notre cas. De plus, nous pouvons montrer que l'inégalité (4.13) est fautive pour toute fonction  $\Psi \in C^2(\bar{\Omega})$  et vérifiant (4.7) lorsque  $\omega$  est à distance non nulle de la droite  $x = 0$ . En effet, puisque  $\Psi$  vérifie (4.7),  $\partial_x \Psi(0, 0) = 0$ , donc  $\tilde{\nabla} \Psi(0, 0) = 0$ . On conclut alors que (4.13) est faux par continuité des dérivées partielles de  $\Psi$ .

Le paramètre  $s$  est choisi plus grand que 1. Premièrement, les termes 1, 2, 3, 7 et le deuxième élément du terme 6, sont majorés par

$$Cs^2 \iint_Q \frac{|z|^2}{t^3(T-t)^3}. \quad (4.14)$$

Le terme  $\iint_Q P_3^2$  est également majoré par (4.14) pour un  $C > 0$  bien choisi.

Traitons maintenant 5 et le premier terme de 6 conjointement. Le premier élément de 5 est minoré par zéro grâce à l'hypothèse (4.7).

Découpons le domaine d'intégration du premier élément de 6 et des éléments restants de 5 en deux morceaux :  $(0, T) \times \Omega \setminus \omega_0$  et  $(0, T) \times \omega_0$ . Puisque le premier élément de 6 est en  $\lambda^2$  et que les éléments de 5 sont en  $\lambda$ , la somme des intégrales sur  $(0, T) \times \Omega \setminus \omega_0$  est minorée par zéro pour  $\lambda$  suffisamment grand, grâce à l'hypothèse (4.13). Le morceau restant (l'intégrale sur le domaine  $(0, T) \times \omega_0$ ) est majoré par

$$Cs \iint_{(0, T) \times \omega_0} \frac{|\tilde{\nabla} z|^2}{t^3(T-t)^3} \quad (4.15)$$

On procède au même découpage pour le terme 9 que l'on minore au préalable par

$$Cs^3 \lambda^4 \int_Q \frac{|z|^2}{t^3(T-t)^3}, \quad (4.16)$$

grâce à (4.13). On remarque que la fonction  $z \mapsto s^3 \lambda^4 \frac{|z|^2}{t^3(T-t)^3}$  intégrée sur le sous domaine  $(0, T) \times \Omega \setminus \omega_0$ , est minorée par

$$Cs^3 \lambda^4 \iint_{(0, T) \times \Omega \setminus \omega_0} \frac{|z|^2}{t^3(T-t)^3} \quad (4.17)$$

et l'opposé de la partie intégrée sur  $(0, T) \times \omega_0$  est majorée par

$$Cs^3 \iint_{(0, T) \times \omega_0} \frac{|z|^2}{t^3(T-t)^3} \quad (4.18)$$

Il reste enfin à traiter 8. La partie intégrée sur  $\Omega \setminus \omega_0$  est minorée par

$$-Cs^3\lambda^3 \iint_{\Omega \setminus \omega_0} \frac{|z|^2}{t^3(T-t)^3} \quad (4.19)$$

et la partie intégrée sur  $\omega_0$  est majorée par (4.18), pour un  $C > 0$  bien choisi.

Lorsque  $\lambda$  est suffisamment grand, la somme des termes (4.17) et (4.19) est minoré par un terme de type (4.17) avec un coefficient  $C$  adéquat.

En regroupant ces estimations, nous trouvons l'inégalité suivante

$$s^3 \iint_{(0,T) \times \Omega \setminus \omega_0} \frac{|z|^2}{t^3(T-t)^3} \leq Cs^2 \iint_Q \frac{|z|^2}{t^3(T-t)^3} + Cs^3 \iint_{(0,T) \times \omega_0} \frac{|z|^2 + |\tilde{\nabla}z|^2}{t^3(T-t)^3} \quad (4.20)$$

à partir de laquelle nous obtenons l'inégalité (4.3) par des arguments classiques figurant dans [6].

■

**Remarque 4.2.3** Notez que c'est le terme  $I_{3,3}$  qui permet d'obtenir (4.17) et (4.18) qui sont les composantes les plus importantes de (4.20). De plus, (4.17) et (4.18) reposent sur (4.16), obtenue grâce à l'hypothèse (4.13).

### 4.3 Quelques résultats sur l'équation de Grushin

- Dans l'article fondateur [2] les auteurs montrent que si la dégénérescence est 'trop importante' ( $\gamma > 1$ ) alors (4.1) n'est contrôlable à zéro pour aucun temps  $T > 0$ , alors que si le système est 'faiblement dégénéré' ( $0 < \gamma < 1$ ) alors le système est contrôlable à zéro pour tout temps  $T > 0$ . Dans le cas intermédiaire,  $\gamma = 1$ , un temps minimal de contrôle apparaît. Le résultat principal de [2] est le suivant

**Théorème 4.3.1** Soit  $\omega$  un ouvert tel que  $\bar{\omega} \subset (0, 1] \times [0, 1]$ .

1. Si  $\gamma \in (0, 1)$ , alors (4.1) est contrôlable à zéro pour tout temps  $T > 0$ .
2. Si  $\gamma > 1$ , alors, pour tout temps  $T > 0$ , (4.1) n'est pas contrôlable à zéro en temps  $T$ .
3. Si  $\gamma = 1$  et  $\omega = (a, b) \times (0, 1)$  avec  $0 < a < b \leq 1$  (c'est l'hypothèse (4.4)), alors il existe un temps minimal de contrôle noté  $T^*$  tel que si  $T < T^*$  alors (4.1) n'est pas contrôlable à zéro et si  $T > T^*$ , alors (4.1) est contrôlable à zéro. De plus,  $T^* \geq \frac{a^2}{2}$ .

**Remarque 4.3.1** L'hypothèse sur l'ouvert de contrôle  $\omega$  figurant dans le Théorème 4.3.1 est différente de celle de la section 4.2. On suppose dans le Théorème 4.3.1 que l'ouvert  $\omega$  ne contient pas le segment vertical  $\{0\} \times (0, 1)$ .

**Remarque 4.3.2** En reprenant les estimations de l'article [2], nous pouvons démontrer que l'équation de Grushin n'est pas contrôlable à zéro au temps  $T = \frac{a^2}{2}$ . En effet, la condition (25) de l'article [2], est vérifiée pour  $T = \frac{a^2}{2}$  grâce à (37) et (38). Cette remarque est formulée dans [4, Proposition 4.6] où les auteurs prouvent que l'équation (4.1) n'est pas contrôlable pour  $T \leq \frac{a^2}{2}$  sur  $\omega$ . Comme nous allons le voir par la suite, il a été prouvé récemment que  $T^* = \frac{a^2}{2}$ . Ainsi l'équation de Grushin est un cas rare d'équation parabolique où l'on a une information sur la contrôlabilité à zéro de l'équation au temps  $T = T^*$ .

Pour démontrer le Théorème 4.3.1 il suffit de démontrer l'inégalité (4.3). D'après la Proposition 1 de [2], la solution  $g$  de l'équation (4.2) est  $C([0, T]; L^2(\Omega))$ , ainsi  $y \mapsto g(t, x, y)$  appartient à  $L^2(0, 1)$ , p.p.  $(t, x) \in (0, T) \times (-1, 1)$ . On peut donc développer cette dernière en série de Fourier de la façon suivante :

$$g(t, x, y) = \sum_{n \in \mathbb{N}^*} g_n(t, x) \phi_n(y) \quad (4.21)$$

où  $\phi_n$  est la  $n^{\text{ième}}$  fonction propre normalisée du laplacien Dirichlet sur  $(0, 1)$  :

$$\phi_n(y) = \sqrt{2} \sin(n\pi y)$$

et

$$g_n(t, x) = \int_0^1 g(t, x, y) \phi_n(y) dy.$$

Supposons qu'il existe  $C > 0$  tel que pour tout  $n \in \mathbb{N}^*$ ,

$$\int_{-1}^1 |g_n(T, x)|^2 dx \leq C^2 \int_0^T \int_a^b |g_n(t, x)|^2 dx dt. \quad (4.22)$$

Alors on peut montrer facilement que l'inégalité d'observabilité (4.3) est vérifiée. En effet, d'après le Théorème de Parseval, en utilisant (4.21),

$$\int_{-1}^1 \int_0^1 |g(T, x, y)|^2 dy dx = \int_{-1}^1 \sum_{n \in \mathbb{N}^*} |g_n(T, x)|^2 dx = \sum_{n \in \mathbb{N}^*} \int_{-1}^1 |g_n(T, x)|^2 dx,$$

et donc, avec (4.22),

$$\int_{-1}^1 \int_0^1 |g(T, x, y)|^2 dy dx \leq C^2 \sum_{n \in \mathbb{N}^*} \int_0^T \int_a^b |g_n(t, x)|^2 dx dt = C^2 \int_0^T \int_a^b \sum_{n \in \mathbb{N}^*} |g_n(t, x)|^2 dx dt,$$

d'où (4.3) en utilisant encore l'égalité de Parseval.

Tout le problème se ramène donc à prouver les inégalité d'observabilité uniformes (4.22), indexés par  $n$ . Toujours dans [2], les auteurs montrent que  $g_n$  est l'unique solution faible de

$$\begin{cases} \partial_t g_n - \partial_{xx} g_n + (n\pi)^{2\gamma} |x|^{2\gamma} g_n = 0, & (t, x) \in (0, T) \times (-1, 1), \\ g_n(t, \pm 1) = 0, & t \in (0, T), \\ g_n(0, x) = g_n^0(x), & x \in (-1, 1), \end{cases} \quad (4.23)$$

où  $g_n^0 \in L^2(-1, 1)$  est donné par  $g_n^0 := \int_0^1 g_0(x, y) \phi_n(y) dy$ . Or (4.23) est le problème adjoint de

$$\begin{cases} \partial_t g_n - \partial_{xx} g_n + (n\pi)^{2\gamma} |x|^{2\gamma} g_n = u_n(t, x) 1_{(a,b)}(x), & (t, x) \in (0, T) \times (-1, 1), \\ g_n(t, \pm 1) = 0, & t \in (0, T), \\ g_n(0, x) = g_n^0(x), & x \in (-1, 1). \end{cases} \quad (4.24)$$

Montrer les inégalités d'observabilité uniformes (4.22) pour tout  $n \in \mathbb{N}^*$ , revient donc à contrôler les équations (4.24) pour tout  $n \in \mathbb{N}^*$ , et l'indépendance de  $C$  vis-à-vis de  $n$  impose que les contrôles  $u_n \in L^2((0, T) \times (-1, 1))$  recherchés doivent être majorés par une constante indépendante de  $n$  multipliée par  $\|g_n^0\|_{L^2(-1,1)}$ , d'après la remarque 1.1.2.

**Remarque 4.3.3** *En résumé, comme décrit dans la section 1.3.3 du chapitre 1, la stratégie pour contrôler (4.1) au temps  $T$  est la suivante. Il suffit de résoudre l'infinité de problèmes de contrôles (4.24) indexés par  $n \in \mathbb{N}^*$ , avec des contrôles  $u_n$  majorés comme suit :*

$$\exists K > 0, \forall n \in \mathbb{N}^*, \|u_n\|_{L^2((0,T) \times (-1,1))} \leq K \|g_n^0\|_{L^2(-1,1)},$$

*Nous emploierons cette méthode dans la section suivante lorsque nous étudierons le temps minimal de contrôle à zéro pour  $\gamma = 1$ .*

Pour prouver la première assertion du Théorème 4.3.1, les auteurs établissent une inégalité de Carleman avec des poids ad hoc. Ils parviennent grâce à cela à montrer également l'existence d'un temps à partir duquel l'équation (4.1) est contrôlable, dans le cas où  $\gamma = 1$ .

La seconde assertion du Théorème 4.3.1 est montrée dans [2] en niant l'inégalité d'observabilité uniforme (4.22) en choisissant une suite  $(g_n)_{n \geq 1}$  adéquate : pour tout  $n \geq 1$  et pour tout paramètre  $\gamma > 1$ , si l'on définit  $g_n$  comme le premier vecteur propre de l'opérateur elliptique  $-\partial_{xx} + (n\pi)^{2\gamma} |x|^{2\gamma}$  alors

$$\frac{\int_{-1}^1 |g_n(T, x)|^2 dx}{\int_0^T \int_a^b |g_n(t, x)|^2 dx dt} \rightarrow +\infty.$$

Ce rapport diverge aussi sous les hypothèses (4.4) avec  $T \leq \frac{a^2}{2}$ .



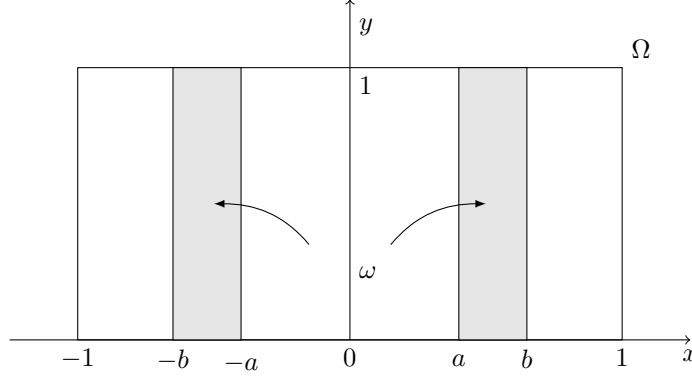


FIGURE 4.2 – Le domaine  $\Omega$  ainsi que la zone de contrôle  $\omega$  de l'équation de Grushin contrôlée sur deux bandes verticales symétrique par rapport à l'axe Oy.

- La question de la valeur exacte du temps minimal lorsque  $\gamma = 1$  est résolue dans [4] dans le cas où  $\omega$  est constitué de deux bandes verticales :  $\omega := ((-b, -a) \times (0, 1)) \cup ((a, b) \times (0, 1))$ , avec  $0 < a < b \leq 1$ .

Plus précisément, les auteurs montrent le Théorème suivant

**Théorème 4.3.2** *Pour tout  $0 < a < b \leq 1$ ,  $\omega = ((-b, -a) \times (0, 1)) \cup ((a, b) \times (0, 1))$  et  $\gamma = 1$ , le système (4.1) est contrôlable à zéro lorsque  $T > \frac{a^2}{2}$  et n'est pas contrôlable à zéro lorsque  $T \leq \frac{a^2}{2}$ .*

La démonstration repose sur la méthode de transmutation (voir [9]). Celle-ci permet de transformer l'équation considérée de type chaleur en une équation de type onde. Il est intéressant de noter que, pour cet ouvert  $\omega$ , il est possible de remplacer le terme en  $x^2$  de l'opérateur  $\mathcal{L}_1$  par une fonction de  $x$ , notée  $q^2$  et vérifiant des conditions plus générales incluant le cas  $q(x) = x$ , et de prouver l'existence d'un temps minimal de contrôle. De plus, les auteurs parviennent à fournir une expression de ce temps minimal. Celui-ci est relié à la distance de Agmon entre l'ouvert  $\omega$  et le segment  $\{0\} \times (0, 1)$  pour le potentiel  $q$  :

$$d_q(\omega) = \int_0^a q(s) ds.$$

Lorsque  $q(x) = x$ , on retrouve l'expression du temps minimal pour un contrôle sur 2 bandes :

$$d_{x^2}(\omega) = \frac{a^2}{2}.$$

- Dans son article (voir [11]), A. Koenig prouve que, lorsque  $\gamma = 1$ , s'il existe une bande horizontale qui a une intersection vide avec l'ouvert de contrôle  $\omega$  alors (4.1) n'est contrôlable pour aucun temps  $T > 0$ . Plus précisément, citons le Théorème 30 de l'Appendice C

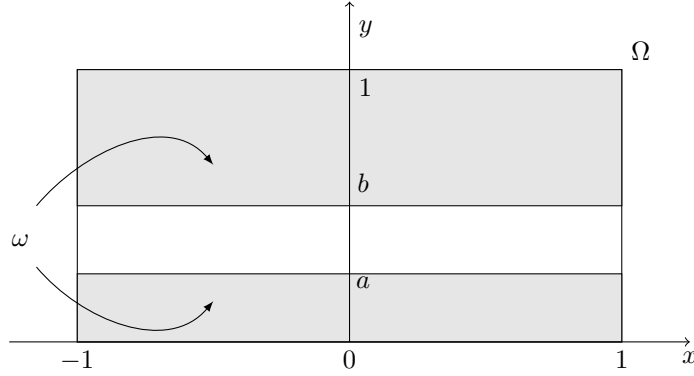
**Théorème 4.3.3** *Soit  $[a, b]$  un segment de  $(0, 1)$  non réduit à un point et  $T > 0$ . Soit la zone de contrôle  $\omega = (-1, 1) \times ((0, 1) \setminus [a, b])$ . L'équation de Grushin (4.1) sur  $\Omega$  pour  $\gamma = 1$  n'est pas contrôlable sur  $\omega$  au temps  $T$ .*

L'idée générale consiste à nier l'inégalité d'observabilité (4.3). Pour cela, l'auteur effectue un raisonnement par l'absurde. Il utilise, pour chaque  $n \geq 1$ , la première fonction propre de l'opérateur elliptique  $-\partial_{xx} + (n\pi)^2|x|^2$  pour montrer que l'inégalité d'observabilité (4.3) implique que, pour une certaine classe de polynômes complexes, l'inégalité suivante a lieu

$$\|f\|_{L^2(D_1)} \leq C \|f\|_{L^\infty(D_2)},$$

où  $D_1$  et  $D_2$  sont des ensembles du plan complexe judicieusement choisis. Ceci mène à une contradiction grâce au Théorème de Runge (voir [14, Théorème 13.9]).

- La conjecture formulée dans [2], sur la valeur du temps minimal ( $T^* = \frac{a^2}{2}$ ) pour un contrôle sur une unique bande, a été démontrée très récemment dans un article à paraître (voir [3]). Plus


 FIGURE 4.3 – La zone de contrôle  $\omega$  de l'équation de Grushin contrôlée sur  $\Omega$  privé d'une bande horizontale.

précisément, les auteurs considèrent un opérateur elliptique de la forme  $-\partial_{xx} - q(x)^2 \partial_{yy}$  sur un domaine  $\Omega^{L^\pm} = (-L_-, L_+) \times (0, \pi)$  où  $L_+$  et  $L_-$  sont deux constantes strictement positives et  $q$  satisfait les conditions suivantes

$$q(0) = 0, q \in C^3([-L_-, L_+]), \inf_{(-L_-, L_+)} \partial_x q > 0. \quad (4.25)$$

Les conditions (4.25) englobent le cas où  $q(x) = x$ . Ils montrent que l'équation de Grushin contrôlée sur le bord  $\{L_+\} \times (0, \pi)$  admet un temps minimal  $T^* = \frac{1}{q'(0)} \int_0^{L_+} q(s) ds$ . Ce temps minimal est égal à  $\frac{L_+^2}{2}$  lorsque  $q(x) = x$ .

La contrôlabilité pour  $T > T^*$  est établie en montrant l'inégalité d'observabilité suivante : il existe  $C > 0$  tel que pour toute condition initiale  $g^0 \in H_0^1(\Omega^{L^\pm})$ , la solution correspondante de (4.2) vérifie

$$\int_{\Omega} |g(T, x, y)|^2 \leq C^2 \int_0^T \int_0^\pi |\partial_x g(t, L_+, y)|^2 dy dt \quad (4.26)$$

Ceci est fait en deux temps. Ils montrent d'abord que pour tout temps  $T_0 > 0$  et  $\varepsilon > 0$ , il existe  $C > 0$  tel que pour tout  $n \in \mathbb{N}$  et  $g_n^0 \in H_0^1(-L_-, L_+)$ , la solution  $g_n$  de (4.23) (où la dégénérescence  $|x|^{2\gamma}$  est remplacée par  $q^2$ ) satisfait

$$\|g_n(T_0)\|_{L^2(-L_-, L_+)} \leq C \exp\left(n \left(\int_0^{L_+} q(s) ds + \varepsilon\right)\right) \|\partial_x g_n(\cdot, L_+)\|_{L^2(0, T_0)}. \quad (4.27)$$

Cette inégalité est obtenue à l'aide de deux estimées de type Carleman sur les intervalles  $(0, L_+)$  et  $(-L_-, 0)$ . L'inégalité (4.27) est ensuite démontrée à l'aide de fonctions de troncature sur l'intervalle  $(-L_-, L_+)$ . L'inégalité d'observabilité (4.27) n'est pas uniforme en  $n$  mais en utilisant une inégalité de décroissance parabolique exponentielle (voir [3, Lem. 3.7]), on obtient (4.26).

La non contrôlabilité à zéro pour  $T < T^*$  est prouvée en niant l'inégalité d'observabilité (4.26) à l'aide, encore une fois, des premières fonctions propres des opérateurs elliptiques  $-\partial_{xx} + n^2 q(x)^2$ .

On conclut enfin en montrant que ce temps minimal de contrôle à zéro au bord fournit le temps minimal espéré de contrôlabilité à zéro sur une bande, c'est-à-dire avec les hypothèses (4.4), à l'aide d'un argument de troncature. Pour cela, choisissons  $\varepsilon$ , un réel positif tel que  $a + \varepsilon < b$ , et définissons la fonction plateau  $\xi \in C^\infty(-1, 1)$  telle que

$$\xi(x) := \begin{cases} 0 & \text{sur } (-1, a) \\ 1 & \text{sur } (a + \varepsilon, 1) \\ \forall x \in [-1, 1], \xi(x) \in [0, 1]. \end{cases}$$

Soit  $T > \frac{L_+^2}{2}$ , et  $y_0 \in L^2(\Omega)$ . D'après le résultat précédent, il existe un contrôle  $u \in L^2(\{L_+\} \times (0, \pi))$  tel que la solution du système suivant (posé dans  $\Omega^{L^\pm} := (L_-, L_+) \times (0, \pi) := (-1, a +$

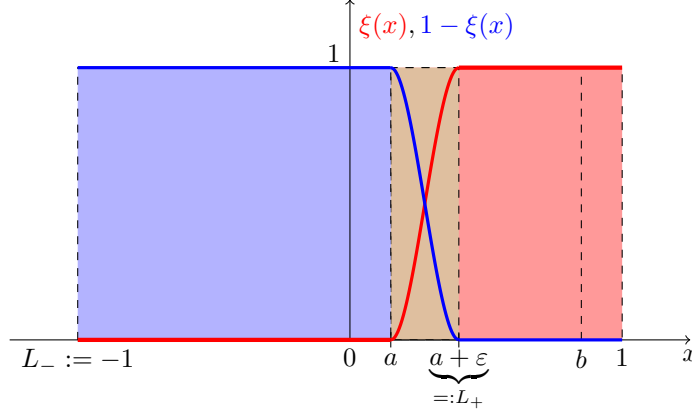


FIGURE 4.4 – Représentation de la fonction plateau  $\xi$  en rouge (resp.  $1 - \xi$  en bleu) valant 0 (resp. 1) sur  $[L_-, a]$  et 1 (resp. 0) sur  $[L_+, 1]$

$\varepsilon) \times (0, \pi)$ )

$$\begin{cases} \partial_t f_1 + \mathcal{L}_1 f_1 = 0 \text{ dans } \Omega^{L^\pm} \\ f_1(0) = y_0 \text{ dans } \Omega^{L^\pm} \\ f_1 = 0 \text{ sur } \partial\Omega^{L^\pm} \setminus \{L_+\} \times (0, \pi) \\ f_1 = u \text{ sur } \{L_+\} \times (0, \pi) \end{cases} \quad (4.28)$$

vérifie  $f_1(T) = 0$ . De plus, le système parabolique non dégénéré suivant (posé sur  $\Omega^{a,1} := (a, 1) \times (0, \pi)$  et contrôlé sur la bande  $\omega_{a,b} := (a, b) \times (0, \pi)$ )

$$\begin{cases} \partial_t f_2 + \mathcal{L}_1 f_2 = 1_{\omega_{a,b}} v \text{ dans } \Omega^{a,1} \\ f_2(0) = y_0 \text{ dans } \Omega^{a,1} \\ f_2 = 0 \text{ sur } \partial\Omega^{a,1} \end{cases} \quad (4.29)$$

est contrôlable à zéro en tout temps. Nous choisissons donc un contrôle  $v \in L^2(Q)$  tel que  $f_2(T) = 0$ . Posons alors  $f := \xi f_2 + (1 - \xi) f_1$ , où  $f_1$  et  $f_2$  sont étendus par 0 au domaine  $\Omega$  tout entier, et définissons le contrôle  $w := (\partial_t + \mathcal{L}_1) f$  supporté dans  $\omega_{a,b}$ . On peut vérifier que ce contrôle  $w$  est dans  $L^2(Q)$ . Ainsi,  $f$  est solution de

$$\begin{cases} \partial_t f + \mathcal{L}_1 f = w \text{ dans } \Omega \\ f(0) = y_0 \text{ dans } \Omega \\ f = 0 \text{ sur } \partial\Omega \end{cases}$$

et vérifie  $f(T) = 0$ . On a donc montré que (4.1) est contrôlable à zéro en  $T = \frac{(a+\varepsilon)^2}{2}$ , et ce quelque soit  $\varepsilon > 0$  tel que  $a < a + \varepsilon < b$ . Donc l'équation (4.1) est contrôlable à zéro pour tout  $T > \frac{a^2}{2}$ . Remarquer que dans [3], l'étude est menée sur  $(L_-, L_+) \times (0, \pi) = (-1, 1) \times (0, \pi)$  et non  $\Omega$ . Changer l'intervalle  $(0, \pi)$  selon Oy en  $(0, 1)$  ne change rien à l'étude et à la validité des résultats.

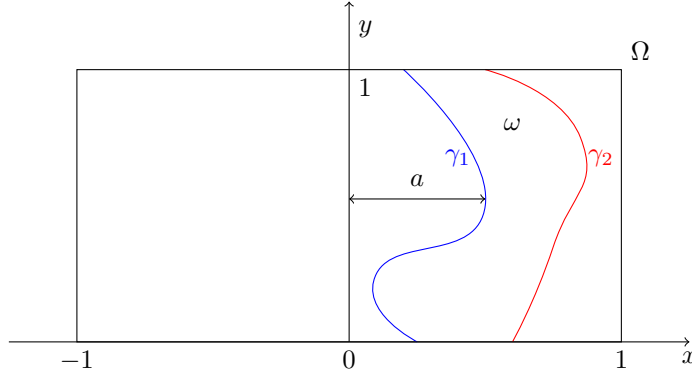
- Dans l'article soumis [8], les auteurs étudient le temps minimal de contrôle à zéro de l'équation de Grushin avec  $\gamma = 1$  sur des zones de contrôle  $\omega$  beaucoup plus générales. Plus précisément, ils considèrent un chemin continu  $\gamma$  reliant le bord horizontal inférieur  $(-1, 1) \times \{0\}$  du rectangle  $\Omega$  au bord horizontal supérieur  $(-1, 1) \times \{1\}$  et un voisinage de ce chemin

$$\omega_0 := \{z \in \Omega, d(z, \text{Im}(\gamma)) < \varepsilon\}.$$

En combinant les résultats de [11],[3] et la méthode de contrôle fictif, ils obtiennent l'existence et la valeur du temps minimal de contrôle à zéro de l'équation contrôlée sur un domaine  $\omega$  tel que

- l'ouvert  $\omega_0$  est un sous ensemble de  $\omega$ ,
- il existe  $y_0 \in (0, \pi)$  tel que  $\{(x, y_0), -a < x < a\}$  est disjoint de  $\bar{\omega}$  où  $a$  est donné par

$$a := \sup_{(x,y) \in \Omega \setminus \omega_0} \{|x| : \exists x_0 \in (-1, 1), |x| < |x_0|, \text{sgn}(x) = \text{sgn}(x_0), (x_0, y) \in \omega_0\}.$$


 FIGURE 4.5 – La zone de contrôle  $\omega$  de l'équation de Grushin définies par deux chemins.

Dans ce cas le temps minimal vaut  $\frac{a^2}{2}$ . Par exemple, lorsque la zone de contrôle est  $\omega := \{(x, y) : \gamma_1(y) < x < \gamma_2(y)\}$  avec  $\gamma_1$  et  $\gamma_2$  deux chemins joignant le bord horizontal inférieur au bord horizontal supérieur du domaine  $\Omega$  tels que  $\gamma_1(y) < \gamma_2(y)$ , alors, le temps minimal est égal à  $\frac{a^2}{2}$  où  $a = \max(\sup \gamma_2^-, \sup \gamma_1^+)$  avec  $\gamma^+ = \max(0, \gamma)$  et  $\gamma^- = \max(0, -\gamma)$  (voir la figure 4.5). Les auteurs considèrent aussi le cas où  $\Omega = \mathbb{R} \times (0, \pi)$  et obtiennent un résultat négatif, à savoir, que l'équation de Grushin n'est jamais contrôlable à zéro sur ce domaine lorsque la zone de contrôle  $\omega$  est définie par

$$\omega := \{(x, y) \in \mathbb{R} \times (0, \pi), |y - y_0| > f(x)\},$$

où  $y_0 \in (0, \pi)$  et  $f : \mathbb{R} \rightarrow \mathbb{R}_+^*$  est une fonction continue.

Dans la section qui suit, nous proposons une approche différente pour démontrer que le temps minimal de contrôle à zéro de l'équation de Grushin sous les hypothèses (4.4) vaut  $T^* = \frac{a^2}{2}$ . Celle-ci repose sur la méthode des moments.

## 4.4 Étude du temps minimal sur une bande lorsque $\gamma = 1$

Nous supposons dans toute la suite que  $\gamma$  et  $\omega$  vérifient (4.4).

Rappelons que, d'après le Théorème 4.3.1 de la section 4.3, démontrer que l'équation (4.1) est contrôlable pour tout temps  $T \in (\frac{a^2}{2}, +\infty)$ , implique que  $T^* = \frac{a^2}{2}$ . Pour montrer ceci, nous appliquons la stratégie décrite dans la Remarque 4.3.3 et la section 1.3.3 du chapitre 1. Nous utilisons la méthode des moments afin de construire des contrôles  $u_n$  telle que la suite  $\frac{\|u_n\|_{L^2(Q)}}{\|g_n^*\|_{L^2(-1,1)}}$  est uniformément majorée dans  $L^2(Q)$ . Rappelons que cette méthode exige de montrer

1. le point relatif à l'existence et la majoration de la famille biorthogonale, c'est-à-dire
  - (a) qu'il existe  $\rho > 0$  tel que  $\forall n, j \in \mathbb{N}^*$ ,  $\lambda_{j+1,n} - \lambda_{j,n} \geq \rho$  (propriété de gap),
  - (b) qu'il existe  $\mathcal{N} : \mathbb{R}_+ \rightarrow \mathbb{N}$  telle que  $\forall \delta > 0, \forall j, n \in \mathbb{N}^*$ ,  $\sum_{j=\mathcal{N}(\delta)}^{\infty} \frac{1}{\lambda_{j,n}} \leq \delta$ .
2. une minoration de la forme  $\int_a^b g_{j,n}^2 \geq C_1 e^{-\lambda_{j,n} a^2}$ .

### 4.4.1 Plan de la section 4.4

Dans la sous-section 4.4.2 nous montrons une partie du point 1a. Plus précisément, nous établissons le Théorème suivant

**Théorème 4.4.1** *Pour tout réel  $\tau \in (0, 1)$  il existe  $\rho > 0$  tel que pour tout entier  $n \in \mathbb{N}^*$  et tout entier  $k$  tels que*

$$1 \leq k < \left\lceil \tau \frac{\pi}{2} n \right\rceil \text{ ou } k > \left\lfloor \frac{\pi}{2} n \right\rfloor,$$

on a la minoration

$$\lambda_{k+1,n} - \lambda_{k,n} \geq \rho. \quad (4.30)$$

Dans la zone où l'indice  $k$  appartient à l'ensemble  $\llbracket \lceil \tau \frac{\pi}{2} n \rceil, \lfloor \frac{\pi}{2} n \rfloor \rrbracket$ , nous ne sommes pas parvenus à démontrer l'inégalité (4.30) de gap uniforme en le paramètre  $n$ . Pour montrer le Théorème 4.4.1, nous

procédons en plusieurs étapes. Tout d'abord nous établissons l'inégalité générale suivante à l'aide du principe du Max Min (ou formule de Courant-Fisher), valable pour tous entiers non nuls  $n$  et  $k$  :

$$\lambda_{k,n} \geq n\pi(2k - 1).$$

Ensuite, dans la sous-section intitulée 'Cas des basses fréquences', nous montrons la partie du Théorème 4.4.1 qui concerne les indices  $k$  compris entre 1 et  $\tau \frac{\pi}{2} n$ . L'inégalité (4.30) est ensuite démontrée dans la sous-section 'Cas des hautes fréquences' pour les indices  $k \geq \lceil \frac{\pi}{2} n \rceil + 1$ .

Nous montrons ensuite le point 1b dans la sous section 4.4.3. Puis, comme indiqué au point 2, nous estimons la norme  $L^2$  de la fonction propre sur un sous domaine de  $(0, 1)$  en établissant l'inégalité (4.61). Enfin nous rassemblons les différents résultats dans la partie 4.4.5 et concluons que  $T^* = \frac{\alpha^2}{2}$  en supposant que le point 1a est bien valable.

#### 4.4.2 Analyse spectrale des opérateurs $\mathcal{A}_n$ sur $(-1, 1)$ , étude de la propriété de gap

Nous cherchons dans cette section des estimations sur les ensembles  $\Lambda_{n \geq 1} = ((\lambda_{k,n})_{k \geq 1})_{n \geq 1}$  qui sont, rappelons-le, les spectres des opérateurs  $\mathcal{A}_n = -\partial_{xx} + (n\pi)^2|x|^2$ . Pour tout  $n \geq 1$ , les vecteurs propres correspondants sont notés  $(g_{k,n})_{k \geq 1}$  et vérifient (1.31). Comme suggéré dans [2], il peut être utile d'effectuer le changement de variable suivant

$$\forall k, n \geq 1, \forall x \in \Omega_n := (-\sqrt{n\pi}, \sqrt{n\pi}), v_{k,n}(x) := \frac{g_{k,n}(\frac{x}{\sqrt{n\pi}})}{(n\pi)^{1/4}}.$$

Ainsi définies, les fonction  $v_{k,n}$  vérifient alors

$$\begin{cases} -\partial_{xx} v_{k,n} + x^2 v_{k,n} = \frac{\lambda_{k,n}}{n\pi} v_{k,n}, \text{ dans } \Omega_n, \\ v_{k,n}(\pm\sqrt{n\pi}) = 0, \\ \|v_{k,n}\|_{L^2(\Omega_n)}^2 = 1. \end{cases} \quad (4.31)$$

Dans ce qui suit, pour tout  $n \geq 1$ , nous appellerons *valeurs propres normalisées* l'ensemble  $(\frac{\lambda_{k,n}}{n\pi})_{k \geq 1}$ . Avant de commencer l'étude spectrale des opérateurs  $\mathcal{A}_n$ , nous traçons les valeurs propres normalisées sur la figure 4.6.

**Remarque 4.4.1** Les figures 4.6, 4.7 et 4.8 sont obtenues à l'aide de la commande `spec` de Scilab. On calcule ainsi les 350 premières valeurs propres (notées  $(\lambda_{k,n}^N)_{k=1}^{350}$ ) des opérateurs  $\mathcal{A}_n$  discrétisés par la méthode des différences finies sur un maillage uniforme de  $N = 7200$  mailles, pour des valeurs du paramètre  $n$  allant de 100 à 200, par pas de 10. On s'assure que le maillage est suffisamment raffiné en calculant les écarts relatifs

$$Er_{N_1, N_2, n} := \frac{\max_{k=1}^{350} |\lambda_{k,n}^{N_2} - \lambda_{k,n}^{N_1}|}{\max(\lambda_{k_0, n}^{N_1}, \lambda_{k_0, n}^{N_2})}, \text{ où } k_0 = \operatorname{argmax}_{k=1}^{350} |\lambda_{k,n}^{N_2} - \lambda_{k,n}^{N_1}|,$$

pour différentes valeurs de  $n$  et différents nombres de mailles  $N_1$  et  $N_2$ . Ces valeurs sont reportées dans le tableau 4.1.

N1	N2	n=100	n=120	n=140	n=160	n=180	n=200
1000	2000	$6.8 \cdot 10^{-2}$	$6.5 \cdot 10^{-2}$	$6.3 \cdot 10^{-2}$	$6 \cdot 10^{-2}$	$5.8 \cdot 10^{-2}$	$5.5 \cdot 10^{-2}$
2000	3000	$1.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$1 \cdot 10^{-2}$
3000	4000	$4.7 \cdot 10^{-3}$	$4.5 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$	$4.1 \cdot 10^{-3}$	$3.8 \cdot 10^{-3}$	$3.6 \cdot 10^{-3}$
4000	5000	$2.2 \cdot 10^{-3}$	$2.1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$
5000	6000	$1.2 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$9.8 \cdot 10^{-4}$	$9.2 \cdot 10^{-4}$
6000	7000	$7.5 \cdot 10^{-4}$	$7.1 \cdot 10^{-4}$	$6.8 \cdot 10^{-4}$	$6.4 \cdot 10^{-4}$	$6 \cdot 10^{-4}$	$5.6 \cdot 10^{-4}$

TABLEAU 4.1 – Valeurs de  $Er_{N_1, N_2, n}$  pour  $N_1, N_2 \in \{1000, 2000, \dots, 7000\}$  et  $n \in \{100, 120, \dots, 200\}$ .

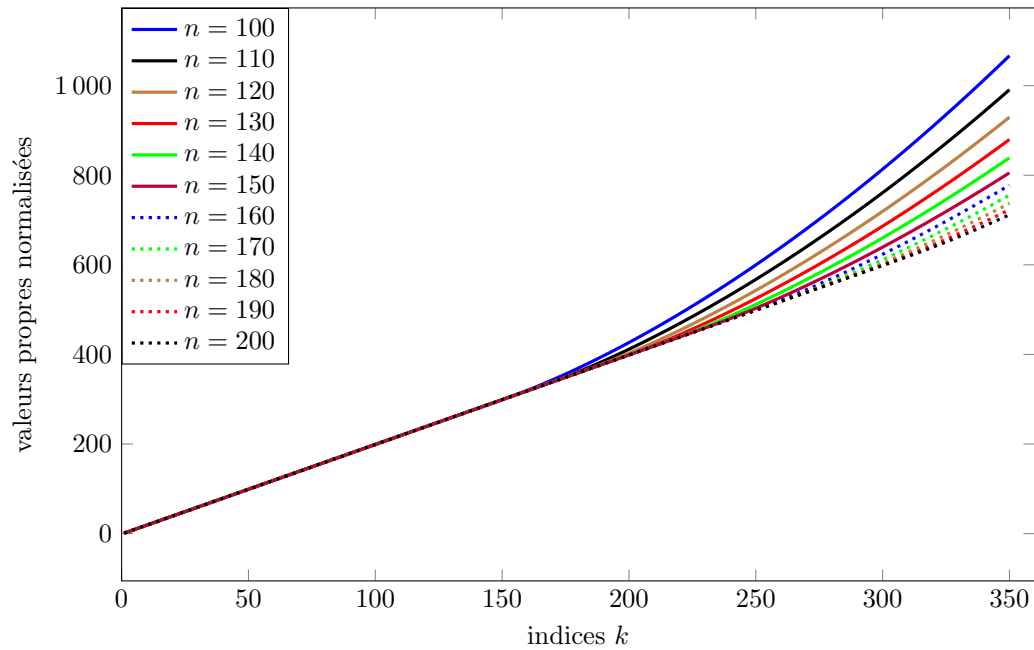


FIGURE 4.6 – Valeurs propres normalisées  $\frac{\lambda_{k,n}}{n\pi}$  en fonction de l'indice  $k$  pour différents indices  $n$  de potentiel

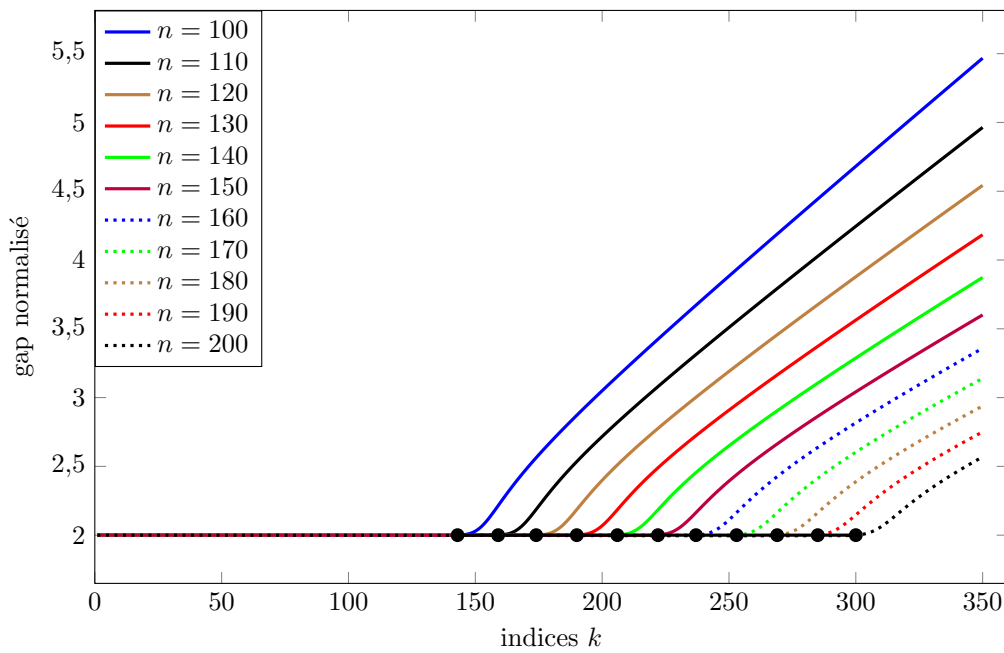
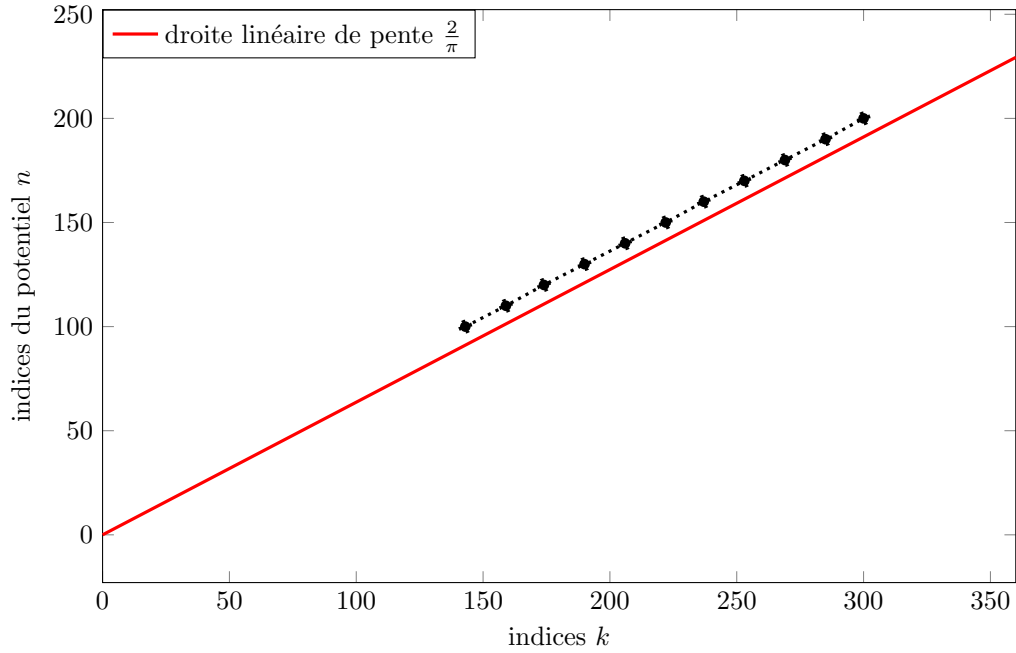


FIGURE 4.7 – Gap normalisé  $\frac{\lambda_{k+1,n}}{n\pi} - \frac{\lambda_{k,n}}{n\pi}$  en fonction de l'indice  $k$  pour différents indices de potentiel.


 FIGURE 4.8 – Indices du potentiel  $n$  en fonction des indices où les valeurs propres changent de comportement.

Nous constatons sur le graphique 4.6 l'existence de deux régimes. Pour les indices  $k$  'petits', les valeurs propres normalisées se comportent linéairement en  $k$  et lorsque  $k$  est 'grand', les valeurs propres semblent se comporter comme une fonction carrée en  $k$ . Ce comportement en hautes fréquences était attendu. En effet, l'équation (4.31) n'est autre qu'une équation aux valeurs propres avec un opérateur du type  $-\partial_{xx} \cdot + q(x) \cdot$  et la théorie de Sturm-Liouville (voir par exemple [12], [17]) affirme qu'asymptotiquement en  $k$ , les valeurs propres d'un tel opérateur sont celles de l'opérateur laplacien. Naturellement ceci s'observe également sur le graphique 4.7 représentant le gap des fonctions propres normalisées. Grâce au graphique 4.7, nous pouvons compléter nos observations. La partie linéaire du spectre normalisé semble avoir pour pente 2. Pour chaque valeur du paramètre  $n$ , les points noirs de la figure 4.7 correspondent aux indices  $k$  à partir desquels le gap varie significativement, ces indices sont notés  $k_n^*$ . A  $n$  fixé, le gap semble être constant et égal à 2 en dessous de l'indice  $k_n^*$  et tend à croître linéairement après  $k_n^*$ . Dans la figure 4.8, nous traçons les couples  $(k_n^*, n)$  afin de conjecturer une relation entre  $k_n^*$  et  $n$ .

D'après le graphique 4.8, il semblerait donc que les valeurs propres aient un comportement linéaire pour  $k < \frac{\pi}{2}n$  et que, pour  $k > \frac{\pi}{2}n$ , les valeurs propres se rapprochent d'une fonction carrée. Dans cette section, nous allons apporter des éléments de réponse à cette conjecture.

### Une inégalité générale

Nous établissons tout d'abord l'estimation suivante.

**Théorème 4.4.2** *Pour tous entiers non nuls  $k$  et  $n$ ,*

$$\lambda_{k,n} \geq n\pi(2k - 1) \quad (4.32)$$

Les courbes des valeurs propres normalisées sont donc toujours au-dessus de la droite  $k \mapsto 2k - 1$ . Cette inégalité s'établit à l'aide de la Formule de Courant-Fisher.

### Formule de Courant-Fisher (ou principe du max-min)

Extrait de [7, p.98]

- Soient  $V$  et  $\mathcal{H}$  deux espaces de Hilbert tels que  $V$  s'injecte dans  $\mathcal{H}$  de façon compacte :  $V \xrightarrow{K} \mathcal{H}$ . On suppose de plus que  $V$  est dense dans  $\mathcal{H}$ . On note  $|\cdot|$  et  $(\cdot, \cdot)$  la norme et le produit scalaire de  $\mathcal{H}$ . Ceux de  $V$  sont notés  $\|\cdot\|$  et  $((\cdot, \cdot))$ .
- Soit  $a(\cdot, \cdot)$  une forme sesquilinéaire sur  $V \times V$ , continue et coercive sur  $V \times V$  :  $\exists \alpha > 0, \forall v \in V, a(v, v) \geq \alpha \|v\|^2$ .

- Soit  $A$  un opérateur auto-adjoint de domaine  $D(A)$  dense dans  $\mathcal{H}$  défini par :

$$D(A) := \{u \in V, v \mapsto a(u, v) \text{ est continue sur } V \text{ pour la topologie de } H\} \quad (4.33)$$

et tel que  $a(u, v) = (Au, v)$  pour tout  $u \in D(A)$  et  $v \in V$  ( $Au$  est le représentant, dans  $\mathcal{H}$ , de  $a(u, \cdot)$ ). On rappelle qu'avec ces hypothèses,  $A$  est un isomorphisme de  $D(A)$  muni de la norme du graphe de  $A$  dans  $\mathcal{H}$  et que  $A^{-1} : \mathcal{H} \rightarrow \mathcal{H}$  est un opérateur compact.

**Théorème 4.4.3** (Extrait de [7, p.102], théorème 10) *Les valeurs propres de l'opérateur  $A$  vérifient :*

$$0 < \lambda_1 < \lambda_2 < \dots,$$

et la  $k^{\text{ième}}$  valeur propre vérifie l'égalité :

$$\lambda_k = \max_{\phi_1, \phi_2, \dots, \phi_{k-1} \in V} \left( \min_{u \in V \cap \text{span}(\phi_1, \dots, \phi_{k-1})^\perp, |u|=1} a(u, u) \right) \quad (4.34)$$

où  $\text{span}(\phi_1, \dots, \phi_{k-1})^\perp := \{u \in \mathcal{H}, (\phi_i, u) = 0, \forall i \in \llbracket 1, k-1 \rrbracket\}$ .

**Remarque 4.4.2** Dans [7], les auteurs énoncent le Théorème 4.4.3 en précisant que

$$\dim(\text{span}(\phi_1, \phi_2, \dots, \phi_{k-1})) = k - 1.$$

Remarquons cependant que cette condition n'est pas nécessaire. En effet, puisque la suite  $(\lambda_k)_{k \geq 1}$  est strictement croissante, le maximum dans (4.34) est nécessairement atteint pour des fonctions  $\phi_1, \phi_2, \dots, \phi_{k-1}$  générant un espace de dimension  $k - 1$ .

## Deux applications

Exprimons les valeurs propres normalisées à l'aide du Théorème 4.4.3. On pose

- $\mathcal{H} = L^2(\Omega_n)$  et  $V = H_0^1(\Omega_n)$ . L'espace  $V$  est dense dans  $\mathcal{H}$  et le théorème d'injection de Rellich assure que  $H_0^1(\Omega_n) \xrightarrow{K} L^2(\Omega_n)$ ,
- $a$  la forme bilinéaire, continue sur  $V \times V$  et coercive sur  $V$  définie par

$$a(u, v) = \int_{\Omega_n} \partial_x u(x) \partial_x v(x) + x^2 u(x) v(x) dx,$$

- $\mathcal{A}$  l'opérateur défini par  $\mathcal{A}u = -\partial_{xx}u + x^2u$  et de domaine  $D(\mathcal{A})$  défini par (4.33). Ce domaine est égal à  $H_0^1(\Omega_n) \cap H^2(\Omega_n)$  et est dense dans  $\mathcal{H}$ .

On a alors, d'après le Théorème 4.4.3 :

$$\frac{\lambda_{k,n}}{n\pi} = \max_{\phi_1, \phi_2, \dots, \phi_{k-1} \in H_0^1(\Omega_n)} \left( \min_{u \in H_0^1(\Omega_n) \cap \text{span}(\phi_1, \dots, \phi_{k-1})^\perp, \|u\|_{L^2(\Omega_n)}=1} a(u, u) \right) \quad (4.35)$$

On considère maintenant le problème aux valeurs propres d'inconnues  $(\mu_k)_{k \geq 1}$  et  $(G_k)_{k \geq 1}$

$$\begin{cases} -\partial_{xx}G_k(x) + x^2G_k(x) = \mu_k G_k(x), & x \in \mathbb{R}, \\ \|G_k\|_{L^2(\mathbb{R})} = 1 \end{cases} \quad (4.36)$$

**Remarque 4.4.3** Ce problème est complètement résolu dans, par exemple, [16, p.73] et [7, p.65].

Les solutions sont  $(G_k)_{k \geq 1} = \left( \frac{e^{-\frac{1}{2}x^2} H_k(x)}{\pi^{1/4} 2^{\frac{k-1}{2}} \sqrt{(k-1)!}} \right)_{k \geq 1}$  où  $H_k$  est le  $(k-1)^{\text{ième}}$  polynôme d'Hermite

et  $(\mu_k)_{k \geq 1} = (2k-1)_{k \geq 1}$ . Naturellement, puisque les problèmes aux valeurs propres (4.36) et (4.31) sont proches (ils diffèrent uniquement par leurs domaines :  $\Omega_n = (-\sqrt{n\pi}, \sqrt{n\pi})$  et  $\mathbb{R}$ ), leurs solutions le sont également. Nous établirons plus loin (voir Théorème 4.4.4) que les valeurs propres de (4.31) convergent, à  $k$  fixé, lorsque  $n \rightarrow +\infty$  vers celles de (4.36) :  $\forall k \geq 1, \frac{\lambda_{k,n}}{n\pi} \rightarrow_{n \rightarrow +\infty} \mu_k$ .

Appliquons le Théorème 4.4.3 à ce nouvel exemple.



- On pose cette fois  $\mathcal{H} = L^2(\mathbb{R})$  et  $V$  est défini de la façon suivante.

Définissons la forme bilinéaire  $a_{\mathbb{R}}$  sur  $\mathcal{D}(\mathbb{R}) \times \mathcal{D}(\mathbb{R})$  par

$$\forall u, v \in \mathcal{D}(\mathbb{R}), a_{\mathbb{R}}(u, v) = \int_{\mathbb{R}} \partial_x u(x) \partial_x v(x) + x^2 u(x) v(x) dx.$$

On pose alors  $V = \overline{\mathcal{D}(\mathbb{R})}^a$ , le complété de  $\mathcal{D}(\mathbb{R})$  pour la norme  $a_{\mathbb{R}}(u, u)^{1/2}$ . Ainsi défini,  $V$  est un espace de Hilbert pour le produit scalaire  $a_{\mathbb{R}}$  et de plus, d'après [7, p.64]  $V \xrightarrow{K} \mathcal{H}$  et  $V$  est dense dans  $\mathcal{H}$  (car  $\mathcal{D}(\mathbb{R}) \subset V$ ).

- On s'assure alors facilement que  $a_{\mathbb{R}}$  est une forme bilinéaire, continue sur  $V \times V$  et coercive.
- Soit  $\mathcal{A}_{\mathbb{R}}$  l'opérateur défini à partir de la forme bilinéaire  $a_{\mathbb{R}}$  et de domaine  $D(\mathcal{A}_{\mathbb{R}})$  donné par 4.33. Ce domaine est dense dans  $\mathcal{H}$  car en effet,  $\mathcal{D}(\mathbb{R}) \subset D(\mathcal{A}_{\mathbb{R}})$ .

On a alors :

$$\mu_k = \max_{\phi_1, \phi_2, \dots, \phi_{k-1} \in V} \left( \min_{u \in L^2(\mathbb{R}) \cap \text{span}(\phi_1, \dots, \phi_{k-1})^{\perp L^2(\mathbb{R})}, \|u\|_{L^2(\mathbb{R})} = 1} a_{\mathbb{R}}(u, u) \right)$$

**Preuve de l'inégalité  $\lambda_{k,n} \geq n\pi(2k-1)$**

Les entiers  $k$  et  $n$  sont dorénavant fixés. La preuve utilise le principe du max min mais n'est pas directe car les domaines des opérateurs  $\mathcal{A}$  et  $\mathcal{A}_{\mathbb{R}}$  sont différents. On s'inspire de la méthode Rayleigh-Ritz expliquée dans la théorème XIII.3 de [13, p.82].

On veut montrer que :

$$\begin{aligned} & \max_{\phi_1, \phi_2, \dots, \phi_{k-1} \in H_0^1(\Omega_n)} \left( \min_{u \in H_0^1(\Omega_n) \cap \text{span}(\phi_1, \dots, \phi_{k-1})^{\perp L^2(\Omega_n)}, \|u\|_{L^2(\Omega_n)} = 1} a(u, u) \right) \\ & \geq \max_{\phi_1, \phi_2, \dots, \phi_{k-1} \in V} \left( \min_{u \in L^2(\mathbb{R}) \cap \text{span}(\phi_1, \dots, \phi_{k-1})^{\perp L^2(\mathbb{R})}, \|u\|_{L^2(\mathbb{R})} = 1} a_{\mathbb{R}}(u, u) \right) \end{aligned}$$

On prouve tout d'abord que

**Lemme 4.4.1**  $H_0^1(\Omega_n)$  est un sous espace de  $V = \overline{\mathcal{D}(\mathbb{R})}^a$ .

Il est sous-entendu que l'on prolonge par zéro sur  $\mathbb{R} \setminus \Omega_n$  les fonctions de  $H_0^1(\Omega_n)$ .

**Preuve**

Soit  $v \in H_0^1(\Omega_n)$ , il existe  $(\phi_k)_{k \geq 1} \in \mathcal{D}(\Omega_n)^{\mathbb{N}}$  qui converge vers  $v$  :  $\|\phi_k - v\|_{H_0^1(\Omega_n)} \rightarrow 0$ . Montrons que  $(\phi_k)_{k \geq 1}$  est de Cauchy dans  $V$  pour la norme  $a_{\mathbb{R}}$ . Pour cela, il suffit de comparer  $a_{\mathbb{R}}$  et  $\|\cdot\|_{H_0^1(\Omega_n)}$  : pour  $u$  à support dans  $\Omega_n$ ,

$$\begin{aligned} a_{\mathbb{R}}(u, u) &= \int_{-\sqrt{n\pi}}^{\sqrt{n\pi}} |\partial_x u(x)|^2 + x^2 |u(x)|^2 dx \\ &\leq \|u\|_{H_0^1(\Omega_n)}^2 + n\pi \|u\|_{L^2(\Omega_n)}^2 \\ &\leq C_n \|u\|_{H_0^1(\Omega_n)}^2, \end{aligned}$$

par le Théorème de Poincaré. Donc  $(\phi_k)_{k \geq 1}$  est de Cauchy dans  $V$  et converge donc vers un élément  $w \in V$ . On conclut en montrant que  $v = w$  :

$$\begin{aligned} \|v - w\|_{H_0^1(\Omega_n)} &\leq \|v - \phi_k\|_{H_0^1(\Omega_n)} + \|w - \phi_k\|_{H_0^1(\Omega_n)} \\ &\leq \|v - \phi_k\|_{H_0^1(\Omega_n)} + a_{\mathbb{R}}(w - \phi_k, w - \phi_k)^{1/2} \xrightarrow{k \rightarrow +\infty} 0. \end{aligned}$$

■

**Corollaire 4.4.1** On pose  $E_{k,n} = \text{span}(v_{1,n}, v_{2,n}, \dots, v_{k,n})$  l'espace engendré par les  $k$  premières fonctions propres (normalisées et orthogonales) de l'opérateur  $\mathcal{A}$ . L'espace  $E_{k,n}$  forme un sous espace de  $H_0^1(\Omega_n)$  et donc de  $V$ .

Par ailleurs, on note

$$P_{k,n} : u \in V \mapsto \sum_{i=1}^k v_{i,n}(u, v_{i,n})_{L^2(\mathbb{R})} \in E_{k,n}$$

l'opérateur de projection de  $L^2(\mathbb{R})$  sur l'espace (fermé car de dimension finie)  $E_{k,n}$ , restreint à  $V$ . On a en particulier, pour tout  $u \in E_{k,n}$  et tout  $v \in L^2(\mathbb{R})$ ,  $(u, P_{k,n}v)_{L^2(\mathbb{R})} = (u, v)_{L^2(\mathbb{R})}$ .

Enfin, on sait que, dans l'égalité (4.35), le maximum est atteint pour les fonctions  $\phi_i = v_{i,n}$  pour  $i \in \llbracket 1, k-1 \rrbracket$ , et que le minimum est atteint pour  $u = v_{k,n} \in E_{k,n}$ .

On a alors (voir les explications détaillées ci-dessous dans la section justification) :

$$\frac{\lambda_{k,n}}{n\pi} = \min_{u \in E_{k,n} \cap \text{span}(v_{1,n}, \dots, v_{k-1,n})^{\perp L^2(\Omega_n)}, \|u\|_{L^2(\Omega_n)}=1} a(u, u) \quad (4.37)$$

$$= \max_{\phi_1, \phi_2, \dots, \phi_{k-1} \in V} \left( \min_{u \in E_{k,n} \cap \text{span}(P_{k,n}\phi_1, \dots, P_{k,n}\phi_{k-1})^{\perp L^2(\Omega_n)}, \|u\|_{L^2(\Omega_n)}=1} a(u, u) \right) \quad (4.38)$$

$$= \max_{\phi_1, \phi_2, \dots, \phi_{k-1} \in V} \left( \min_{u \in E_{k,n} \cap \text{span}(\phi_1, \dots, \phi_{k-1})^{\perp L^2(\mathbb{R})}, \|u\|_{L^2(\mathbb{R})}=1} a_{\mathbb{R}}(u, u) \right) \quad (4.39)$$

$$\geq \max_{\phi_1, \phi_2, \dots, \phi_{k-1} \in V} \left( \min_{u \in V \cap \text{span}(\phi_1, \dots, \phi_{k-1})^{\perp L^2(\mathbb{R})}, \|u\|_{L^2(\mathbb{R})}=1} a_{\mathbb{R}}(u, u) \right). \quad (4.40)$$

**Justifications des étapes (4.37)-(4.38)-(4.39)-(4.40)**

La preuve que l'on peut passer de (4.39) à (4.40) est immédiate avec le Corollaire 4.4.1.

Justifions (4.37)-(4.38). Puisque  $E_{k,n} \subset V$ , et que

$$\text{span}(v_{1,n}, v_{2,n}, \dots, v_{k-1,n}) = \text{span}(P_{k,n}v_{1,n}, P_{k,n}v_{2,n}, \dots, P_{k,n}v_{k-1,n}),$$

on a (4.38)  $\geq$  (4.37). Pour montrer l'inégalité inverse, on fixe  $\phi_1, \phi_2, \dots, \phi_{k-1}$  dans  $V$ . Comme  $P_{k,n}\phi_i \in E_{k,n}$ , et puisque dans l'égalité (4.35), le maximum est atteint pour les fonctions  $\phi_i = v_{i,n}$  pour  $i \in \llbracket 1, k-1 \rrbracket$ , on a

$$\begin{aligned} \min_{u \in E_{k,n} \cap \text{span}(P_{k,n}\phi_1, \dots, P_{k,n}\phi_{k-1})^{\perp L^2(\Omega_n)}, \|u\|_{L^2(\Omega_n)}=1} a(u, u) &\leq \\ \min_{u \in E_{k,n} \cap \text{span}(v_{1,n}, \dots, v_{k-1,n})^{\perp L^2(\Omega_n)}, \|u\|_{L^2(\Omega_n)}=1} a(u, u), & \end{aligned}$$

d'où (4.37)  $\geq$  (4.38) (qui est en fait le seul sens vraiment utile pour la preuve) en passant au max sur  $V$  dans le terme de gauche.

Il ne reste plus qu'à montrer la dernière égalité (4.38)-(4.39). Celle-ci provient de l'égalité

$$E_{k,n} \cap \text{span}(P_{k,n}\phi_1, \dots, P_{k,n}\phi_{k-1})^{\perp L^2(\Omega_n)} = E_{k,n} \cap \text{span}(\phi_1, \dots, \phi_{k-1})^{\perp L^2(\mathbb{R})},$$

et du fait que  $a$  et  $a_{\mathbb{R}}$  coïncident sur  $H_0^1(\Omega_n)$ . En effet, soit  $u \in E_{k,n}$ . On a :

$$u \in \text{span}(P_{k,n}\phi_1, \dots, P_{k,n}\phi_{k-1})^{\perp L^2(\Omega_n)},$$

si et seulement si, pour tout  $i \in \llbracket 1, k-1 \rrbracket$ ,

$$0 = (u, P_{k,n}\phi_i)_{L^2(\Omega_n)} = (u, P_{k,n}\phi_i)_{L^2(\mathbb{R})} = (u, \phi_i)_{L^2(\mathbb{R})}$$

si et seulement si  $u \in E_{k,n} \cap \text{span}(\phi_1, \dots, \phi_{k-1})^{\perp L^2(\mathbb{R})}$ .

### Cas des basses fréquences

On montre le théorème suivant qui décrit le comportement des valeurs propres en basses fréquences (c'est-à-dire dans une zone où  $k < \frac{\pi}{2}n$ ).

**Théorème 4.4.4** *Soit  $\tau \in (0, 1)$ . Il existe trois réels,  $n_0 \in \mathbb{N}^*$ ,  $c_1 > 0$  et  $c_2 > 0$ , tous ne dépendant que de  $\tau$ , tels que pour tout entier  $n \geq n_0$  et  $k \leq \tau \frac{\pi}{2}n$ ,*

$$\frac{\lambda_{k,n}}{n\pi} \leq \mu_k + c_1 e^{-c_2 n}. \quad (4.41)$$

On rappelle que la minoration  $\mu_k \leq \frac{\lambda_{k,n}}{n\pi}$  est établie au Théorème 4.4.2 sans hypothèse sur  $k$ . Ainsi, en utilisant le résultat du Théorème 4.4.4 précédent, on obtient immédiatement le Corollaire ci-dessous.

**Corollaire 4.4.2** *Soit  $\tau \in (0, 1)$ . Il existe trois réels,  $n_0 \in \mathbb{N}^*$ ,  $c_1 > 0$  et  $c_2 > 0$ , tous ne dépendant que de  $\tau$ , tels que pour tout entier  $n \geq n_0$  et  $k \leq \tau \frac{\pi}{2}n$ ,*

$$\mu_k \leq \frac{\lambda_{k,n}}{n\pi} \leq \mu_k + c_1 e^{-c_2 n}.$$

Ainsi, pour tout  $k < \lfloor \tau \frac{\pi}{2}n \rfloor$ , et tout  $n \geq n_0$ ,

$$\lambda_{k+1,n} - \lambda_{k,n} \geq 2 - c_1 e^{-c_2 n}.$$

Avant de prouver le Théorème 4.4.4, on commence avec un lemme décrivant la parité des fonctions propres  $g_{k,n}$ . Noter que ce lemme généralise le Lemme 2 de [2].

**Lemme 4.4.2** *Pour tout  $n \in \mathbb{N}^*$ , lorsque  $k$  est pair,  $g_{k,n}$  est impaire et lorsque  $k$  est impair,  $g_{k,n}$  est paire. Il en est de même pour les fonctions propres  $v_{k,n}$  et  $G_k$ .*

### Preuve

La preuve de ce lemme repose sur la parité du potentiel et sur le théorème d'oscillation de Sturm, suivant lequel, pour tout  $n \in \mathbb{N}^*$ , la fonction propre  $g_{k,n}$  possède  $k+1$  zéros dans  $[-1, 1]$  et, en chacun des zéros appartenant à l'ensemble  $(-1, 1)$ , s'il y en a, la fonction  $g_{k,n}$  s'annule en changeant de signe, d'après le théorème de Cauchy-Lipschitz.

On remarque que la fonction  $g_{k,n}(-x)$  est aussi une fonction propre de  $\mathcal{A}_n$  pour la valeur propre  $\lambda_{k,n}$  car le potentiel est une fonction paire. Donc, puisque les espaces propres de  $\mathcal{A}_n$  sont de dimension 1,  $g_{k,n}(-x)$  est proportionnelle à  $g_{k,n}$  et puisque les solutions sont normalisées,  $g_{k,n}(-x) = \pm g_{k,n}(x)$ . Autrement dit, les fonctions propres sont toutes, soit paires, soit impaires. Or une fonction paire ou impaire qui s'annule un nombre pair de fois sur  $[-1, 1]$  ne s'annule pas en zéro, donc est forcément paire. Donc, lorsque  $k$  est impair,  $g_{k,n}$  est paire. Inversement, si une fonction paire ou impaire s'annule un nombre impair de fois en changeant de signe sur  $[-1, 1]$ , alors elle s'annule en changeant de signe en 0, donc est forcément impaire. Donc, lorsque  $k$  est paire,  $g_{k,n}$  est impaire. ■

On peut maintenant prouver le résultat annoncé.

### Preuve

On fixe  $\tau \in (0, 1)$  et on prouve qu'il existe  $n_0, c_1$  et  $c_2$  ne dépendent que de  $\tau$  tels que pour  $n \geq n_0$ , on a

$$\forall k \leq \tau \frac{\pi}{2}n, \quad \frac{\lambda_{k,n}}{n\pi} \leq \mu_k + c_1 e^{-c_2 n},$$

par récurrence forte sur  $k$ .

Le cas  $k = 1$  a été traité dans [2], page 83. Les auteurs obtiennent :

$$\lambda_{1,n} \leq n\pi + O(n^{9/4} e^{-n\pi/2}),$$

ce qui donne bien le résultat voulu.

Supposons maintenant le résultat démontré jusqu'au rang  $k-1 \geq 1$ . On a alors  $\tau \frac{\pi}{2}n > k-1 \geq 1$ .

**Supposons  $k$  impair.** Ainsi  $G_k$  est paire. On pose maintenant

$$\tilde{G}_k = (G_k - G_k(\sqrt{n\pi})) - \sum_{p=1}^{k-1} v_{p,n} (G_k - G_k(\sqrt{n\pi}), v_{p,n})_{L^2(\Omega_n)}.$$

On a alors :  $\tilde{G}_k \in \text{span}(v_{1,n}, v_{2,n}, \dots, v_{k-1,n})^\perp$  et surtout  $\tilde{G}_k(\pm\sqrt{n\pi}) = 0$ , donc  $\tilde{G}_k \in H_0^1(\Omega_n)$ . Calculons l'image de  $\tilde{G}_k$  par l'opérateur de Sturm  $\mathcal{A}$

$$\begin{aligned} -\tilde{G}_k'' + x^2\tilde{G}_k &= -G_k'' + \sum_{p=1}^{k-1} v_{p,n}'' (G_k - G_k(\sqrt{n\pi}), v_{p,n})_{L^2(\Omega_n)} + x^2(G_k - G_k(\sqrt{n\pi})) \\ &\quad - \sum_{p=1}^{k-1} x^2 v_{p,n} (G_k - G_k(\sqrt{n\pi}), v_{p,n})_{L^2(\Omega_n)} \\ &= \mu_k G_k - x^2 G_k(\sqrt{n\pi}) - \sum_{p=1}^{k-1} \frac{\lambda_{p,n}}{n\pi} v_{p,n} (G_k - G_k(\sqrt{n\pi}), v_{p,n})_{L^2(\Omega_n)}. \end{aligned}$$

Alors en utilisant l'orthogonalité :  $\tilde{G}_k \in \text{span}(v_{1,n}, v_{2,n}, \dots, v_{k-1,n})^\perp$ ,

$$\begin{aligned} (\mathcal{A}\tilde{G}_k, \tilde{G}_k)_{L^2(\Omega_n)} &= \int_{\Omega_n} (\tilde{G}_k')^2 + x^2\tilde{G}_k^2 \\ &= \int_{\Omega_n} \mu_k G_k \tilde{G}_k - G_k(\sqrt{n\pi}) x^2 \tilde{G}_k \\ &= \int_{\Omega_n} \mu_k (G_k - G_k(\sqrt{n\pi})) \tilde{G}_k + \mu_k G_k(\sqrt{n\pi}) \tilde{G}_k - G_k(\sqrt{n\pi}) x^2 \tilde{G}_k \\ &= \int_{\Omega_n} \mu_k \tilde{G}_k^2 + G_k(\sqrt{n\pi}) (\mu_k - x^2) \tilde{G}_k \end{aligned}$$

d'où

$$\frac{(\mathcal{A}\tilde{G}_k, \tilde{G}_k)_{L^2(\Omega_n)}}{\|\tilde{G}_k\|_{L^2(\Omega_n)}^2} = \mu_k + G_k(\sqrt{n\pi}) \frac{\int_{\Omega_n} (\mu_k - x^2) \tilde{G}_k}{\int_{\Omega_n} \tilde{G}_k^2} \quad (4.42)$$

Avec l'inégalité de Cauchy-Schwarz, on obtient :

$$\frac{(\mathcal{A}\tilde{G}_k, \tilde{G}_k)_{L^2(\Omega_n)}}{\|\tilde{G}_k\|_{L^2(\Omega_n)}^2} \leq \mu_k + G_k(\sqrt{n\pi}) \frac{\|\mu_k - x^2\|_{L^2(\Omega_n)}}{\left(\int_{\Omega_n} \tilde{G}_k^2\right)^{1/2}}. \quad (4.43)$$

Il s'agit maintenant d'évaluer  $\|\tilde{G}_k\|_{L^2(\Omega_n)}$ .

Tout d'abord, notons que d'après l'égalité de Parseval :

$$\int_{\Omega_n} \tilde{G}_k^2 = \sum_{p \geq 1} (\tilde{G}_k, v_{p,n})^2,$$

puis, comme  $\tilde{G}_k \in \text{span}(v_{1,n}, v_{2,n}, \dots, v_{k-1,n})^\perp$ , on a :

$$\int_{\Omega_n} \tilde{G}_k^2 = \sum_{p \geq k} (\tilde{G}_k, v_{p,n})^2,$$

enfin, par orthogonalité de la famille  $(v_{k,n})_{k \geq 1}$ ,

$$\int_{\Omega_n} \tilde{G}_k^2 = \sum_{p \geq k} (G_k - G_k(\sqrt{n\pi}), v_{p,n})^2. \quad (4.44)$$

On va maintenant estimer la quantité  $(G_k - G_k(\sqrt{n\pi}), v_{p,n})^2$ . On a :

$$\begin{aligned} (G_k - G_k(\sqrt{n\pi}), v_{p,n}) \frac{\lambda_{p,n}}{n\pi} &= (G_k - G_k(\sqrt{n\pi}), \mathcal{A}v_{p,n}) \\ &= (\mathcal{A}G_k - x^2 G_k(\sqrt{n\pi}), v_{p,n}) \\ &= (\mu_k G_k - x^2 G_k(\sqrt{n\pi}), v_{p,n}) \\ &= \mu_k (G_k - G_k(\sqrt{n\pi}), v_{p,n}) + G_k(\sqrt{n\pi}) (\mu_k - x^2, v_{p,n}) \end{aligned}$$

donc,

$$(G_k - G_k(\sqrt{n\pi}), v_{p,n}) \left( \frac{\lambda_{p,n}}{n\pi} - \mu_k \right) = G_k(\sqrt{n\pi})(\mu_k - x^2, v_{p,n})$$

et enfin,

$$(G_k - G_k(\sqrt{n\pi}), v_{p,n})^2 \leq G_k(\sqrt{n\pi})^2 \frac{\|\mu_k - x^2\|^2}{\left(\frac{\lambda_{p,n}}{n\pi} - \mu_k\right)^2} \quad (4.45)$$

Considérons (4.45) pour  $p < k$ . Par hypothèse de récurrence,

$$\mu_p \leq \frac{\lambda_{p,n}}{n\pi} \leq \mu_p + c_1 e^{-c_2 n} < \mu_p + 1,$$

pour  $n$  suffisamment grand (ne dépendant que de  $\tau$ ). Donc  $\mu_k - \frac{\lambda_{p,n}}{n\pi} \geq \mu_k - (\mu_p + 1) \geq 2(k-p) - 1 > 0$ .  
Donc

$$\sum_{p < k} (G_k - G_k(\sqrt{n\pi}), v_{p,n})^2 \leq \frac{\pi^2}{6} G_k(\sqrt{n\pi})^2 \|\mu_k - x^2\|^2. \quad (4.46)$$

Mais on a par ailleurs,

$$\int_{\Omega_n} (G_k - G_k(\sqrt{n\pi}))^2 = \sum_{p < k} (G_k - G_k(\sqrt{n\pi}), v_{p,n})^2 + \sum_{p \geq k} (G_k - G_k(\sqrt{n\pi}), v_{p,n})^2$$

donc avec (4.46), (en prenant par exemple  $C = \frac{\pi^2}{6}$ )

$$\int_{\Omega_n} (G_k - G_k(\sqrt{n\pi}))^2 - C G_k(\sqrt{n\pi})^2 \|\mu_k - x^2\|^2 \leq \sum_{p \geq k} (G_k - G_k(\sqrt{n\pi}), v_{p,n})^2$$

ou encore avec (4.44),

$$\int_{\Omega_n} (G_k - G_k(\sqrt{n\pi}))^2 - C G_k(\sqrt{n\pi})^2 \|\mu_k - x^2\|^2 \leq \int_{\Omega_n} \tilde{G}_k^2.$$

On estime maintenant  $\int_{\Omega_n} (G_k - G_k(\sqrt{n\pi}))^2$  grâce au Lemme suivant dont la preuve est reportée en Annexe 4.5.1.

**Lemme 4.4.3** *Soit  $\delta > 0$ . Il existe un entier  $n_0 \in \mathbb{N}$  qui ne dépend que de  $\delta$ , tel que  $\forall n \geq n_0$ ,  $\forall k \leq \frac{\pi}{2}n$ ,*

$$\int_{\Omega_n} G_k^2 \geq 1 - \delta.$$

On prend par exemple  $\delta = \frac{1}{2}$  et, pour  $n$  suffisamment grand, (ne dépendant que de  $\delta = 1/2$  ici) tel que :

$$\int_{\Omega_n} (G_k - G_k(\sqrt{n\pi}))^2 \geq \frac{1}{2} \int_{\Omega_n} G_k^2 - 2\sqrt{n\pi} G_k^2(\sqrt{n\pi}) \geq \frac{1}{4} - 2\sqrt{n\pi} G_k^2(\sqrt{n\pi}).$$

On a donc (avec maintenant  $C = 2$ ):

$$\frac{1}{4} - C G_k^2(\sqrt{n\pi})(\sqrt{n\pi} + \|\mu_k - x^2\|) \leq \int_{\Omega_n} \tilde{G}_k^2. \quad (4.47)$$

Puisque  $\mu_k \leq n\pi$ , il existe  $C' > 0$  indépendant de  $n$  et de  $k$  (par exemple,  $C' = (2 + \frac{2}{5})\pi^{5/2}$ ) telle que :

$$\left( \int_{\Omega_n} (\mu_k - x^2)^2 \right)^{1/2} \leq C' n^{5/4}. \quad (4.48)$$

On énonce maintenant le Lemme suivant dont la preuve est aussi reportée en Annexe 4.5.1

**Lemme 4.4.4** *Soit  $\tau \in (0, 1)$ . Il existe un triplet  $(n_0, C_1, C_2) \in \mathbb{N}^* \times \mathbb{R}_+^* \times \mathbb{R}_+^*$ , ne dépendant que de  $\tau$ , tel que pour tout  $n \geq n_0$ , et tout  $x \geq \sqrt{n\pi}$ , on a*

$$\forall k \leq \tau n \frac{\pi}{2}, G_k(x) \leq C_1 e^{-C_2 n}. \quad (4.49)$$

Ainsi, d'après le Lemme 4.4.4 pour  $n$  suffisamment grand, ne dépendant que de  $\tau$ ,

$$\frac{1}{8} \leq \int_{\Omega_n} \tilde{G}_k^2.$$

Donc, en reprenant (4.43), en utilisant que  $\tilde{G}_k \in \text{span}(v_{1,n}, v_{2,n}, \dots, v_{k-1,n})^\perp$  et  $G_k(\pm\sqrt{n\pi}) = 0$  et d'après le principe du max min, on trouve :

$$\frac{\lambda_{k,n}}{n\pi} \leq \mu_k + C_u G_k(\sqrt{n\pi}) n^{5/4} \quad (4.50)$$

avec  $C_u = 8C'$ . On termine enfin ce cas en utilisant l'inégalité (4.49) du Lemme 4.4.4 qui fournit l'existence de  $C_1 > 0$  et  $C_2 > 0$  ne dépendant que de  $\tau$ , telles que  $G_k(\sqrt{n\pi}) \leq C_1 e^{-C_2 n}$ , pour tout  $k \leq \tau \frac{\pi}{2} n$ .

**Si  $k$  est pair** alors dans ce cas,  $G_k$  est impair et on pose plutôt :

$$\tilde{G}_k = \left( G_k - \frac{x}{\sqrt{n\pi}} G_k(\sqrt{n\pi}) \right) - \sum_{p=1}^{k-1} v_{p,n} \left( G_k - \frac{x}{\sqrt{n\pi}} G_k(\sqrt{n\pi}), v_{p,n} \right)_{L^2(\Omega_n)}.$$

L'inégalité (4.42) devient alors :

$$\frac{\int_{\Omega_n} (\tilde{G}_k')^2 + x^2 \tilde{G}_k^2}{\int_{\Omega_n} \tilde{G}_k^2} = \mu_k + G_k(\sqrt{n\pi}) \frac{\int_{\Omega_n} (\mu_k - \frac{x^3}{\sqrt{n\pi}}) \tilde{G}_k}{\int_{\Omega_n} \tilde{G}_k^2}$$

L'inégalité (4.47) est changée en :

$$\frac{1}{4} - C G_k^2(\sqrt{n\pi}) \left( \sqrt{n\pi} + \left\| \mu_k - \frac{x^3}{\sqrt{n\pi}} \right\|^2 \right) \leq \int_{\Omega_n} \tilde{G}_k^2,$$

ce qui conduit à poser plutôt  $C' = (2 + \frac{2}{7})\pi^{5/2}$  et on conclut de la même façon.

■

### Cas des hautes fréquences

Nous établissons le Théorème suivant :

**Théorème 4.4.5** *Il existe une constante universelle  $C > 0$  telle que*

$$\lambda_{k,n} \geq n^2 \pi^2 + n\pi \Rightarrow \lambda_{k+1,n} - \lambda_{k,n} \geq C. \quad (4.51)$$

**Remarque 4.4.4** *D'après le Théorème 4.4.2, la condition  $\lambda_{k,n} \geq n^2 \pi^2 + n\pi$  de (4.51) est vérifiée lorsque  $k \geq \lceil \frac{\pi}{2} n \rceil + 1$ .*

### Preuve

La preuve consiste à transformer le problème aux valeurs propres (1.31) en une méthode de tir dans laquelle l'inconnue  $\lambda_{k,n}$  est remplacée par une variable  $t$ , puis à appliquer le changement de Prüfer modifié à cette dernière équation.

On considère le système

$$\begin{cases} -\partial_{xx} h_n(t, x) + n^2 \pi^2 x^2 h_n(t, x) = t h_n(t, x) \\ h_n(t, -1) = 0 \\ \partial_x h_n(t, -1) = 1. \end{cases} \quad (4.52)$$

Notons que  $t$  est une valeur propre de l'opérateur  $\mathcal{A}_n$  lorsque  $h_n(t, 1) = 0$ . L'idée consiste à faire varier la source  $t$  jusqu'à obtenir la condition finale  $h_n(t, 1) = 0$ . La condition initiale  $\partial_x h_n(t, -1) = 1$  est arbitraire et ne change pas l'étude si on remplace 1 par n'importe quelle autre constante non nulle. En effet, changer la condition initiale  $\partial_x h_n(t, -1)$  dans le système (4.52) revient simplement à modifier la norme de la fonction propre  $h_n$ .

Le changement de variable de Prüfer modifié consiste à considérer 2 fonctions  $r_n$  et  $\phi_n$  définies par :

$$\begin{cases} h_n(t, x) = r_n(t, x) \cos(\phi_n(t, x))(t - n^2\pi^2x^2)^{-1/4}, \\ \partial_x h_n(t, x) = r_n(t, x) \sin(\phi_n(t, x))(t - n^2\pi^2x^2)^{1/4}. \end{cases} \quad (4.53)$$

Afin que les quantités apparaissant dans (4.53) soient bien définies, il convient d'imposer que la variable  $t$  vérifie  $t - n^2\pi^2x^2 > 0$ .

L'existence de  $\phi_n$  et de  $r_n$  est un résultat classique et on peut montrer comme dans [5] ou [10] que  $\phi_n$  est solution sur  $[-1, 1]$  de :

$$\partial_x \phi_n(t, x) = -\sqrt{t - n^2\pi^2x^2} + \frac{n^2\pi^2x}{2(t - n^2\pi^2x^2)} \sin(2\phi_n(t, x)). \quad (4.54)$$

On a le lemme suivant :

**Lemme 4.4.5** *Soit  $\lambda_{k,n}$  une valeur propre vérifiant la condition du Théorème 4.4.5. Il existe une fonction  $\phi_n$  répondant à (4.54) telle que*

$$\phi_n(\lambda_{k,n}, -1) = -\frac{\pi}{2} \text{ et } \phi_n(\lambda_{k,n}, 1) = -\pi(k + \frac{1}{2}).$$

### Preuve

Cherchons la condition initiale en  $x = -1$  vérifiée par  $\phi_n$ . Celle-ci est en fait contrainte par (4.53). On a nécessairement :  $\phi_n(\lambda_{k,n}, -1) = \frac{\pi}{2} + j\pi$ , avec  $j \in \mathbb{Z}$ , car  $h_n(\lambda_{k,n}, -1) = 0$ . La fonction  $\phi_n$  étant définie à  $2\pi$  près, on peut sans problème supposer que  $\phi_n(\lambda_{k,n}, -1) \in [-\pi, \pi]$ . Ainsi  $\phi_n(\lambda_{k,n}, -1) = -\frac{\pi}{2}$  ou  $\phi_n(\lambda_{k,n}, -1) = \frac{\pi}{2}$ . Quitte à changer le signe de la fonction  $r_n$ , on peut supposer que  $\phi_n(\lambda_{k,n}, -1) = -\frac{\pi}{2}$ .

En examinant (4.54), nous constatons que  $\partial_x \phi_n(\lambda_{k,n}, -1) < 0$ . D'après le Théorème d'oscillation de Sturm,  $h_n(\lambda_{k,n}, \cdot)$  s'annule  $k+1$  fois en changeant de signe sur  $[-1, 1]$ . Donc  $\phi_n(\lambda_{k,n}, \cdot)$  prend  $k+1$  fois la valeur  $\pi/2$  modulo  $\pi$  sur  $[-1, 1]$ . Au premier zéro de  $h_n(\lambda_{k,n}, \cdot)$ ,  $\phi_n(\lambda_{k,n}, \cdot)$  prend la valeur  $-\frac{\pi}{2}$ , au second zéro de  $h_n(\lambda_{k,n}, \cdot)$ , noté  $x_2$ ,  $\phi_n(\lambda_{k,n}, x_2)$  peut prendre les valeurs  $-\pi/2 + \pi$ ,  $-\pi/2$  ou  $-\pi/2 - \pi$ . Puisque d'après (4.54),  $\partial_x \phi_n(\lambda_{k,n}, -1) < 0$ , on peut directement exclure les possibilités  $\phi_n(\lambda_{k,n}, x_2) = -\pi/2 + \pi$  et  $\phi_n(\lambda_{k,n}, x_2) = -\pi/2$ . On a donc :  $\phi_n(\lambda_{k,n}, x_2) = -\pi/2 - \pi$ . En raisonnant de proche en proche, on montre que  $\phi_n(\lambda_{k,n}, 1) = -\frac{\pi}{2} - k\pi$ .

■

D'après le Lemme 4.4.5 :

$$\phi_n(\lambda_{k+1,n}, 1) - \phi_n(\lambda_{k,n}, 1) = -\pi = \int_{\lambda_{k,n}}^{\lambda_{k+1,n}} \partial_t \phi_n(t, 1) dt$$

d'où :

$$\pi \leq (\lambda_{k+1,n} - \lambda_{k,n}) \sup_{t \in [\lambda_{k,n}, \lambda_{k+1,n}]} |\partial_t \phi_n(t, 1)|, \quad (4.55)$$

il faut donc étudier cette borne supérieure. On reprend donc (4.54) et on dérive par rapport à  $t$  :

$$\partial_x \partial_t \phi_n = -\frac{1}{2\sqrt{t - n^2\pi^2x^2}} - \frac{n^2\pi^2x}{2(t - n^2\pi^2x^2)^2} \sin(2\phi_n) + \frac{n^2\pi^2x \partial_t \phi_n}{t - n^2\pi^2x^2} \cos(2\phi_n) \quad (4.56)$$

et ainsi :

$$\partial_x (\partial_t \phi_n) \leq \frac{n^2\pi^2|x|}{2(t - n^2\pi^2x^2)^2} + \frac{n^2\pi^2x \cos(2\phi_n)}{t - n^2\pi^2x^2} \partial_t \phi_n.$$

En appliquant le Lemme de Gronwall différentiel sur cette équation, on trouve :

$$\begin{aligned} & \partial_t \phi_n(t, 1) \exp\left(-\int_{-1}^1 \frac{n^2\pi^2s \cos(2\phi_n)}{t - n^2\pi^2s^2} ds\right) - \partial_t \phi_n(t, -1) \\ & \leq \int_{-1}^1 \frac{n^2\pi^2|y|}{2(t - n^2\pi^2y^2)^2} \exp\left(-\int_{-1}^y \frac{n^2\pi^2s \cos(2\phi_n)}{t - n^2\pi^2s^2} ds\right) dy \end{aligned}$$

et, on a  $\partial_t \phi_n(t, -1) = 0$ . Donc :

$$\partial_t \phi_n(t, 1) \leq \int_{-1}^1 \frac{n^2\pi^2|y|}{2(t - n^2\pi^2y^2)^2} \exp\left(\int_y^1 \frac{n^2\pi^2s \cos(2\phi_n)}{t - n^2\pi^2s^2} ds\right) dy$$

et

$$\partial_t \phi_n(t, 1) \leq \int_{-1}^1 \frac{n^2 \pi^2 |y|}{2(t - n^2 \pi^2 y^2)^2} \exp\left(\int_y^1 \frac{n^2 \pi^2 |s|}{t - n^2 \pi^2 s^2} ds\right) dy$$

La minoration s'effectue de la même façon. On reprend (4.56) et on a :

$$\partial_x(\partial_t \phi_n) \geq -\frac{1}{2\sqrt{t - n^2 \pi^2 y^2}} - \frac{n^2 \pi^2 |x|}{2(t - n^2 \pi^2 x^2)^2} + \frac{n^2 \pi^2 x \cos(2\phi_n)}{t - n^2 \pi^2 x^2} \partial_t \phi_n.$$

et de nouveau, avec le Lemme de Gronwall :

$$\begin{aligned} \partial_t \phi_n(t, 1) &\geq -\int_{-1}^1 \frac{n^2 \pi^2 |y|}{2(t - n^2 \pi^2 y^2)^2} \exp\left(\int_y^1 \frac{n^2 \pi^2 s \cos(2\phi_n)}{t - n^2 \pi^2 s^2} ds\right) dy \\ &\quad - \int_{-1}^1 \frac{1}{2\sqrt{t - n^2 \pi^2 y^2}} \exp\left(\int_y^1 \frac{n^2 \pi^2 s \cos(2\phi_n)}{t - n^2 \pi^2 s^2} ds\right) dy \end{aligned}$$

La majoration et la minoration ci-dessus montrent :

$$\begin{aligned} |\partial_t \phi_n(t, 1)| &\leq \int_{-1}^1 \frac{n^2 \pi^2 |y|}{2(t - n^2 \pi^2 y^2)^2} \exp\left(\int_y^1 \frac{n^2 \pi^2 s \cos(2\phi_n)}{t - n^2 \pi^2 s^2} ds\right) dy \\ &\quad + \int_{-1}^1 \frac{1}{2\sqrt{t - n^2 \pi^2 y^2}} \exp\left(\int_y^1 \frac{n^2 \pi^2 s \cos(2\phi_n)}{t - n^2 \pi^2 s^2} ds\right) dy \end{aligned}$$

Des calculs simples mais fastidieux menés dans l'Annexe 4.5.2 donnent :

$$|\partial_t \phi_n(t, 1)| \leq \frac{t}{6(t - n^2 \pi^2)^2} + \frac{1}{2(t - n^2 \pi^2)} + \frac{\sqrt{t}}{4n\pi\sqrt{t - n^2 \pi^2}} \ln\left(1 + \frac{2n\pi}{\sqrt{t - n\pi}}\right) + \frac{1}{2\sqrt{t - n^2 \pi^2}}.$$

En reprenant maintenant (4.55),

$$\begin{aligned} \pi &\leq (\lambda_{k+1,n} - \lambda_{k,n}) \times \\ &\left( \frac{\lambda_{k,n}}{6(\lambda_{k,n} - n^2 \pi^2)^2} + \frac{1}{2(\lambda_{k,n} - n^2 \pi^2)} + \frac{\sqrt{1 + \frac{n^2 \pi^2}{\lambda_{k,n} - n^2 \pi^2}}}{4n\pi} \ln\left(1 + \frac{2n\pi}{\sqrt{\lambda_{k,n} - n\pi}}\right) + \frac{1}{2\sqrt{\lambda_{k,n} - n^2 \pi^2}} \right). \end{aligned}$$

Maintenant, considérons l'hypothèse

$$\lambda_{k,n} \geq n^2 \pi^2 + n\pi. \quad (4.57)$$

On a :

$$\begin{aligned} \pi &\leq (\lambda_{k+1,n} - \lambda_{k,n}) \left( \frac{n^2 \pi^2 + n\pi}{6n^2 \pi^2} + \frac{1}{2n\pi} + \frac{\sqrt{1 + n\pi}}{4n\pi} \ln\left(1 + \frac{2n\pi(\sqrt{\lambda_{k,n}} + n\pi)}{\lambda_{k,n} - n^2 \pi^2}\right) + \frac{1}{2\sqrt{n\pi}} \right) \\ &\leq (\lambda_{k+1,n} - \lambda_{k,n}) \left( \frac{1}{6} + \frac{1}{6n\pi} + \frac{1}{2n\pi} + \frac{1}{2\sqrt{n\pi}} \ln\left(1 + 2(\sqrt{n^2 \pi^2 + n\pi} + n\pi)\right) + \frac{1}{2\sqrt{n\pi}} \right) \\ &\leq (\lambda_{k+1,n} - \lambda_{k,n}) \left( \frac{1}{6} + \frac{2}{3n\pi} + \frac{\ln(10n\pi)}{2\sqrt{n\pi}} + \frac{1}{2\sqrt{n\pi}} \right) \end{aligned}$$

Pour  $n$  suffisamment grand, on obtient :

$$\pi \leq \lambda_{k+1,n} - \lambda_{k,n}.$$

■

**Remarque 4.4.5** Moyennant des restrictions plus importantes sur les valeurs propres, on peut établir une inégalité du type  $\lambda_{k+1,n} - \lambda_{k,n} \geq C\sqrt{\lambda_{k,n}}$  (voir Annexe 4.5.3).

### 4.4.3 Existence de la fonction $\mathcal{N}$

Dans cette sous section, nous démontrons le point 1b de la page 134.

**Théorème 4.4.6** Il existe une fonction  $\mathcal{N} : \mathbb{R}_+^* \rightarrow \mathbb{N}^*$  telle que pour tout  $\varepsilon > 0$  et tout  $n \in \mathbb{N}^*$ ,

$$\sum_{k \geq \mathcal{N}(\varepsilon)} \frac{1}{\lambda_{k,n}} \leq \varepsilon$$



On étudie la série  $\sum \frac{1}{\lambda_{k,n}}$ . En considérant le potentiel  $n^2\pi^2x^2$  comme une perturbation du laplacien pour l'opérateur  $\mathcal{A}_n$ , les auteurs de [7, p.104, thm 11] obtiennent :

$$\lambda_{k,n} = \frac{k^2\pi^2}{4} + r_{k,n}$$

avec  $|r_{k,n}| \leq n^2\pi^2$  pour tout  $n, k \in \mathbb{N}^*$ . On va montrer qu'il existe  $c > 0$  tel que pour tout  $k, n \in \mathbb{N}^*$ ,  $\lambda_{k,n} \geq ck^2$ . On considère pour cela deux régimes.

1. Si  $k \geq 3n$  alors  $\lambda_{k,n} \geq \frac{k^2\pi^2}{4} - n^2\pi^2 \geq ck^2$
2. Si  $k < 3n$  alors d'après le Théorème 4.4.2  $\lambda_{k,n} \geq (2k-1)n\pi \geq kn \geq ck^2$ .

On fixe  $\varepsilon > 0$ . On choisit donc  $\mathcal{N}(\varepsilon) \in \mathbb{N}$  tel que

$$\frac{1}{c} \sum_{k \geq \mathcal{N}(\varepsilon)} \frac{1}{k^2} \leq \varepsilon,$$

on a ainsi,

$$\sum_{k \geq \mathcal{N}(\varepsilon)} \frac{1}{\lambda_{k,n}} \leq \frac{1}{c} \sum_{k \geq \mathcal{N}(\varepsilon)} \frac{1}{k^2} \leq \varepsilon.$$

#### 4.4.4 Étude de la norme $L^2$ des vecteurs propres sur $(a, b)$

Nous établissons une minoration de  $\int_a^b g_{j,n}^2$  à l'aide de la Proposition 4.5 de [4], que l'on rappelle ci-dessous.

**Proposition 4.4.1** *Soient  $L > 1$  et  $\delta \in (0, a)$  tels que  $a + 3\delta < 1$ . Il existe  $C > 0$  tel que pour tout  $\tilde{\delta} > 0$ ,  $q \in C^1(\Omega^1, \mathbb{R}^+) \setminus \{0\}$ ,  $(w_0, w_1) \in H_0^1(\Omega^1) \times L^2(\Omega^1)$ , on ait :*

$$\|w_0\|_{H_0^1}^2 + \|w_1\|_{L^2}^2 \leq C(1 + \|\tilde{q}\|_{L^\infty})O_W(a, q, \tilde{\delta}, \delta) \int_{-L}^L \int_{\omega_{a,1}} (w_s^2 + w^2)(s, x) dx ds, \quad (4.58)$$

où :

- $\tilde{q} : x \in \Omega^1 \mapsto q(x) + \tilde{\delta}^2 \|q\|_{L^\infty}$
- $O_W(a, q, \tilde{\delta}, \delta) = \max \left( e^{\int_0^{a+2\tilde{\delta}} [M(y)+2\sqrt{\tilde{q}(y)}] dy}, e^{\int_{-a-2\tilde{\delta}}^0 [M(y)+2\sqrt{\tilde{q}(y)}] dy} \right)$ , avec  $M : x \in \Omega^1 \mapsto \frac{|\tilde{q}'(x)|}{\tilde{q}(x)}$ ,
- $\omega_{a,1} = (-1, -a) \cup (a, 1)$ ,
- enfin,  $w$  est la solution de l'équation des ondes suivante

$$\begin{cases} w_{ss} - w_{xx} + q(x)w = 0, & (s, x) \in (-L, L) \times \Omega^1, \\ w(s, \pm 1) = 0, & s \in (-L, L), \\ (w, w_s)(0, x) = (w_0, w_1)(x), & x \in \Omega^1. \end{cases} \quad (4.59)$$

Prenons  $L = 2$  par exemple, et  $\delta > 0$  suffisamment petit ( $a$  est fixé). Soit  $\tilde{\delta} := \delta$  et  $q(x) = (n\pi)^2x^2$ . Choisissons  $w_0 = 0$  et  $w_1 = \sqrt{\lambda_{k,n}}g_{k,n}$  de sorte que

$$w(s, x) = \sin(s\sqrt{\lambda_{k,n}})g_{k,n}(x).$$

On obtient avec (4.58) :

$$\lambda_{k,n} \leq C(1 + n^2\pi^2)O_W(a, q, \tilde{\delta}, \delta)2L \int_{\omega_{a,1}} \lambda_{k,n}g_{k,n}^2(x)dx.$$

Or  $O_W(a, q, \tilde{\delta}, \delta) \leq \frac{(a+2\tilde{\delta})^2 + \tilde{\delta}^2}{\tilde{\delta}^2} \times e^{n\pi(a+2\tilde{\delta})(a+4\tilde{\delta})}$ , donc il existe  $C > 0$  dépendant de  $L, a$  et  $\delta$  et  $C_\delta > 0$  ne dépendant que de  $\delta$  et vérifiant  $C_\delta \rightarrow_{\delta \rightarrow 0} 0$ , tels que

$$\int_{\omega_{a,1}} g_{k,n}^2(x)dx \geq \frac{Ce^{-n\pi(a^2+C_\delta)}}{n^2\pi^2} \quad (4.60)$$

On sait également d'après le Lemme 4.4.2 que les fonctions propres  $g_{k,n}$  sont paires ou impaires. Ainsi,  $g_{k,n}^2$  est paire et donc

$$\int_a^1 g_{k,n}^2(x)dx \geq \frac{Ce^{-n\pi(a^2+C_\delta)}}{n^2\pi^2}. \quad (4.61)$$

#### 4.4.5 Conclusion

L'étude des valeurs propres dans le domaine des basses fréquences,  $k \leq \tau \frac{\pi}{2} n$  avec  $\tau \in (0, 1)$ , a révélé l'existence d'une borne uniforme sur le gap des fonctions propres de l'opérateur  $\mathcal{A}_n$ . Restreindre les indices  $k$  à des valeurs inférieures à  $\tau \frac{\pi}{2} n$  permet d'obtenir la majoration cruciale :  $G_k(\sqrt{n\pi}) \leq C_1 e^{-C_2 n}$ , fournie par le Lemme 4.4.4.

La méthode utilisée pour les hautes fréquences exige quant à elle de se restreindre aux fonctions propres vérifiant

$$\lambda_{k,n} \geq n^2 \pi^2 + n\pi, \quad (4.62)$$

afin que le changement de variable (4.53) puisse être correctement défini. La condition (4.62) est vérifiée lorsque  $k \geq \left\lceil \frac{\pi}{2} n \right\rceil + 1$ .

Ainsi, comme mentionné dans l'introduction, l'étude spectrale des opérateurs  $\mathcal{A}_n$  est inachevée : il reste à obtenir une inégalité uniforme sur le gap pour les valeurs propres dont l'indice  $k$  vérifie

$$\left\lceil \tau \frac{\pi}{2} n \right\rceil \leq k \leq \left\lceil \frac{\pi}{2} n \right\rceil, \quad (4.63)$$

avec une constante  $\tau \in (0, 1)$  arbitrairement proche de 1. Dans la suite, nous supposons qu'une telle minoration est établie pour l'ensemble des spectres  $(\Lambda_n)_{n \geq 1}$ . On suppose donc que

$$\exists C > 0, \forall n \geq 1, \forall k \geq 1, \lambda_{k+1,n} - \lambda_{k,n} > C. \quad (4.64)$$

**Remarque 4.4.6** *Naturellement, pour démontrer la propriété de gap sur les indices vérifiant (4.63) on aimerait vouloir étendre l'étude basses fréquences aux indices  $k \leq \left\lceil \frac{\pi}{2} n \right\rceil$ . Cependant, comme énoncé dans [1], page 7, le maximum de la fonction propre  $G_k$  est atteint en  $x_k \in [-\sqrt{\mu_k}, \sqrt{\mu_k}]$ , avec  $x_k \sim_{k \rightarrow +\infty} \sqrt{\mu_k}$  et de plus,  $G_k(x_k) = O(k^{-1/3})$ . Ainsi, lorsque  $k$  vérifie (4.63),  $G_k(x_k) = O(n^{-1/3})$ . Cet exposant  $n^{-1/3}$  ne permet alors pas de compenser dans (4.50) l'exposant  $n^{5/4}$  provenant de l'estimation (4.48).*

Nous allons montrer que sous l'hypothèse 4.64 l'équation de Grushin (4.1) est contrôlable pour tout temps  $T > \frac{a^2}{2}$  lorsque l'ouvert de contrôle est de la forme  $(a, 1) \times (0, 1)$ , puis nous montrerons, comme expliqué dans [4, prop. 4.1], que cela implique la contrôlabilité pour tout temps  $T > \frac{a^2}{2}$  avec l'ouvert de contrôle  $\omega$ .

Il suffit donc dans un premier temps de montrer que l'inégalité (4.61) est suffisante pour majorer uniformément  $u_n$ . Prenons  $T > \frac{a^2}{2}$  et  $\varepsilon > 0$  et  $\delta > 0$  suffisamment petits tels que

$$T > \varepsilon + \frac{C_\delta}{2} + \frac{a^2}{2}, \quad (4.65)$$

où  $C_\delta \rightarrow_{\delta \rightarrow 0} 0$  est défini en (4.60). Alors, nous obtenons grâce à (4.61),

$$\begin{aligned} \|u_n\|_{L^2(Q_1^T)} &\leq K_\varepsilon \|g_n^0\|_{L^2(\Omega^1)} \sum_{k \geq 1} \frac{e^{-\lambda_{k,n}(T-\varepsilon)}}{\left(\int_a^1 g_{k,n}^2\right)^{1/2}} \\ &\leq K_\varepsilon \|g_n^0\|_{L^2(\Omega^1)} \sum_{k \geq 1} n\pi e^{-\lambda_{k,n}(T-\varepsilon)} e^{n\pi\left(\frac{a^2}{2} + \frac{C_\delta}{2}\right)}. \end{aligned}$$

De plus, l'estimation du Théorème 4.4.2 a pour conséquence

$$\forall k, n \geq 1, \lambda_{k,n} \geq n\pi,$$

ainsi

$$\|u_n\|_{L^2(Q_1^T)}^2 \leq K_\varepsilon^2 \|g_n^0\|_{L^2(\Omega^1)}^2 n^2 \pi^2 \sum_{k \geq 1} e^{-2\lambda_{k,n}((T-\varepsilon) - \frac{a^2}{2} - \frac{C_\delta}{2})}.$$

Grâce à (4.65) nous avons  $(T - \varepsilon) - \frac{a^2}{2} - \frac{C_\delta}{2} > 0$ . En utilisant encore le Théorème 4.4.2, il vient

$$\forall k, n \geq 1, \lambda_{k,n} \geq k,$$

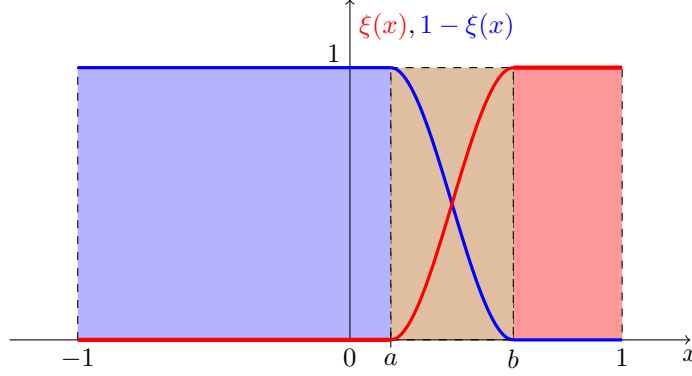


FIGURE 4.9 – Représentation de la fonction plateau  $\xi$  en rouge (resp.  $1 - \xi$  en bleu) valant 0 (resp. 1) sur  $[-1, a]$  et 1 (resp. 0) sur  $[b, 1]$

d'où

$$\|u_n\|_{L^2(Q_T^1)}^2 \leq K_\varepsilon^2 \|g_n^0\|_{L^2(\Omega^1)}^2 n^2 \pi^2 \sum_{k \geq 1} e^{-2k((T-\varepsilon) - \frac{a^2}{2} - \frac{c_\delta}{2})}.$$

On obtient ainsi la majoration uniforme des contrôles  $u_n$  et donc la contrôlabilité de (4.1) pour tout temps  $T > \frac{a^2}{2}$  sur l'ouvert  $(a, 1) \times (0, 1)$ .

Pour établir la contrôlabilité de (4.1) contrôlé sur l'ouvert  $\omega = (a, b) \times (0, 1)$ , on procède à un argument semblable à celui de la section 4.3. Posons cette fois-ci

$$\xi(x) = \begin{cases} 1 \text{ sur } (-1, a) \\ 0 \text{ sur } (b, 1) \\ \forall x \in [-1, 1], \xi(x) \in [0, 1] \end{cases}$$

Définissons les systèmes

$$\begin{cases} \partial_t f_1 + \mathcal{L}_1 f_1 = 1_\omega u_1, \text{ sur } \Omega^{-1,b} := (-1, b) \times (0, 1), \\ f_1(0) = y^0 \text{ sur } \Omega^{-1,b}, \\ f_1 = 0 \text{ sur } \partial\Omega^{-1,b}, \end{cases} \quad (4.66)$$

et

$$\begin{cases} \partial_t f_2 + \mathcal{L}_1 f_2 = 1_\omega u_2, \text{ sur } \Omega^{a,1} := (a, 1) \times (0, 1), \\ f_2(0) = y^0 \text{ sur } \Omega^{a,1}, \\ f_2 = 0 \text{ sur } \partial\Omega^{a,1}. \end{cases} \quad (4.67)$$

D'après les résultats précédents, le système (4.66) est contrôlable à zéro pour tout temps  $T > \frac{a^2}{2}$  et le système parabolique non dégénéré (4.67) est contrôlable à zéro pour tout temps  $T > 0$ . Soient  $T > \frac{a^2}{2}$  et deux contrôles  $u_1$  et  $u_2$  tels que la fonction  $f := \xi_1 f_1 + (1 - \xi_1) f_2$ , soit nulle à l'instant  $T$ . On prolonge par zéro à  $\Omega$  tout entier les fonctions  $f_1$  et  $f_2$ . Le contrôle  $\partial_t f + \mathcal{L}_1 f$  est supporté dans  $\omega$  et des arguments analogues à ceux évoqués sur la régularité de la solution de l'équation (4.29) montrent qu'il appartient à l'espace  $L^2(Q)$ . Il répond donc au problème de contrôle (4.1).

## 4.5 Annexes

### 4.5.1 Résultats concernant les fonctions propres de l'oscillateur harmonique (4.36)

Dans cette annexe, nous étudions quelques propriétés des fonctions propres  $(G_k)_{k \geq 1}$ . Nous établissons les Lemmes 4.4.4 et 4.4.3. Sur la figure 4.10, nous traçons quelques fonctions propres  $G_k$ .

Voici l'énoncé du **Lemme 4.4.4** :

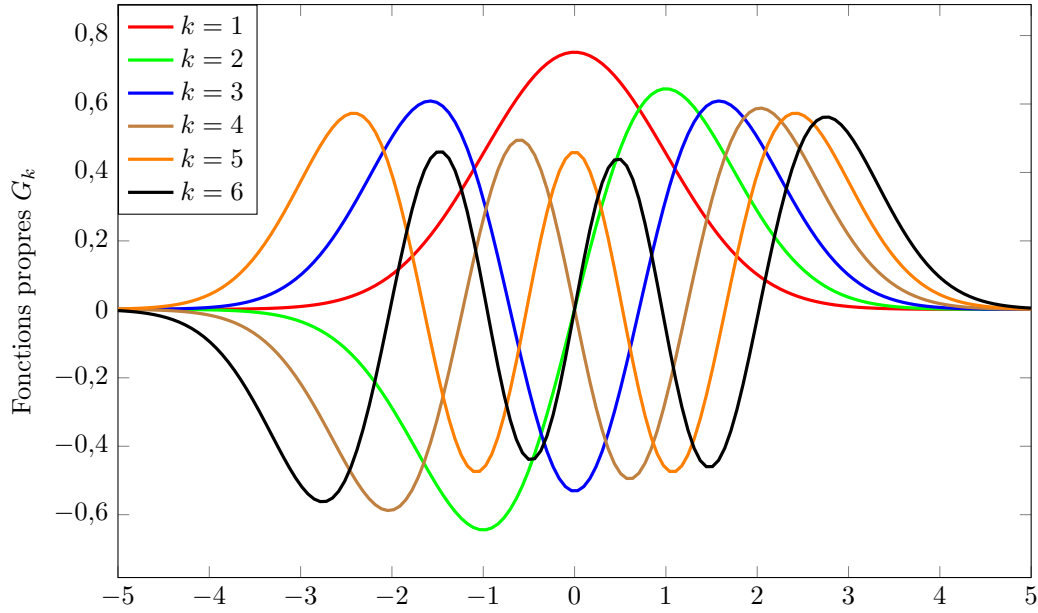


FIGURE 4.10 – Graphique représentant quelques fonctions propres de l'oscillateur harmonique.

**Lemme** Soit  $\tau \in (0, 1)$ . Il existe  $n_0 \in \mathbb{N}^*$  et  $C_1, C_2$  deux constantes strictement positives, ne dépendant que de  $\tau$ , telles que pour tout  $n \geq n_0$ , et tout  $x \geq \sqrt{n\pi}$ , on a

$$\forall k \leq \tau n \frac{\pi}{2}, G_k(x) \leq C_1 e^{-C_2 n}. \quad (4.68)$$

**Preuve**

On rappelle que les fonctions  $(G_k)_{k \geq 1}$  sont normalisées sur  $\mathbb{R}$  et vérifient :

$$G_k(x) = \frac{e^{-x^2/2} H_k(x)}{\pi^{1/4} 2^{\frac{k-1}{2}} \sqrt{(k-1)!}} \quad (4.69)$$

où  $H_k$  est le  $(k-1)$ <sup>ième</sup> polynôme de Hermite sous forme physique, c'est-à-dire défini par :

$$H_k(x) = (-1)^{k-1} e^{x^2} \frac{d^{k-1}}{dx^{k-1}} e^{-x^2}, \forall k \geq 1.$$

Ces polynômes sont orthogonaux pour le produit scalaire  $(f, g) = \int_{\mathbb{R}} f(x)g(x)e^{-x^2} dx$  et ne sont pas unitaires. On commence par montrer que

$$\forall x \geq \sqrt{n\pi}, \forall 1 \leq k \leq \frac{\pi}{2}n, G_{k+1}(x) \geq G_k(x) \quad (4.70)$$

Remarquons que cette inégalité peut s'observer sur la figure ci-dessus. En partant de la définition des polynômes de Hermite, on a pour  $k \geq 2$ :

$$\begin{aligned} e^{-x^2} H_k(x) &= (-1)^{k-1} \left( \frac{d}{dx} \right)^{k-1} e^{-x^2} \\ &= (-1)^{k-1} \left( \frac{d}{dx} \right)^{k-2} (-2xe^{-x^2}) \\ &= (-1)^{k-1} \left[ \binom{k-2}{0} (-2x) \left( \frac{d}{dx} \right)^{k-2} e^{-x^2} + \binom{k-2}{1} (-2) \times \left( \frac{d}{dx} \right)^{k-3} e^{-x^2} \right] \\ &= 2xe^{-x^2} H_{k-1}(x) - 2(k-2)e^{-x^2} H_{k-2}(x). \end{aligned}$$

Puis on utilise (4.69) pour obtenir :

$$\begin{aligned} G_k(x) &= 2x \frac{e^{-x^2/2} H_{k-1}(x)}{\pi^{1/4} 2^{\frac{k-1}{2}} \sqrt{(k-1)!}} - \frac{2(k-2)}{\pi^{1/4} 2^{\frac{k-1}{2}} \sqrt{(k-1)!}} e^{-x^2/2} H_{k-2}(x) \\ &= \frac{2x}{2^{1/2} \sqrt{k-1}} G_{k-1}(x) - \frac{2(k-2)}{2\sqrt{(k-1)(k-2)}} G_{k-2}(x) \\ &= \sqrt{\frac{2}{k-1}} x G_{k-1}(x) - \frac{\sqrt{k-2}}{\sqrt{k-1}} G_{k-2}(x). \end{aligned}$$

On fixe  $n$  et on procède maintenant par récurrence (finie) sur  $k$ . Puisque  $G_1(x) = \frac{e^{-x^2/2}}{\pi^{1/4}}$  et que  $G_2(x) = \frac{x e^{-x^2/2}}{\pi^{1/4} \sqrt{2}}$ , l'inégalité (4.70) est vraie pour le rang  $k = 1$ . Prenons  $k > 1$  et supposons (4.70) au rang  $k-1$  : Pour  $x \geq \sqrt{n\pi}$ ,  $G_k(x) \geq G_{k-1}(x)$ . On a d'après la formule de récurrence plus haut :

$$G_{k+1}(x) = \sqrt{\frac{2}{k}} x G_k(x) - \sqrt{\frac{k-1}{k}} G_{k-1}(x) \quad (4.71)$$

donc,

$$G_{k+1}(x) \geq \left( \sqrt{\frac{2}{k}} x - \sqrt{\frac{k-1}{k}} \right) G_k(x).$$

Puisque  $x \geq \sqrt{n\pi}$ , nous avons  $\sqrt{\frac{2}{k}} x - \sqrt{\frac{k-1}{k}} \geq \sqrt{\frac{2}{k}} \sqrt{n\pi} - 1 \geq \sqrt{\frac{2}{k}} \sqrt{2k} - 1 \geq 1$ .

On termine la preuve de (4.70) en montrant que  $G_k(x) \geq 0$  pour  $k$  et  $x$  vérifiant les contraintes imposées par (4.70). Pour tout  $k \geq 1$ , les zéros des fonctions  $H_k$  (et donc ceux des fonctions  $G_k$ ) sont compris dans l'intervalle  $(-\sqrt{\mu_k}, \sqrt{\mu_k})$  (voir [15, p. 130]). En effet, l'inégalité 6.32.3 de [15, p. 132] donne, en particulier (avec  $\nu = 1$ ), une borne sur  $x_1$ , le plus grand zéro de  $H_k$  (notons que dans cette inégalité,  $i_1$  est le plus petit zéro de la fonction d'Airy et que  $i_1 > 0$  d'après le Théorème 6.32 page 131, [15]). On en déduit immédiatement que

$$x_1 < \sqrt{\mu_k} < \sqrt{n\pi},$$

Ainsi, pour tout  $k \geq 1$ , lorsque  $x \geq \sqrt{n\pi}$ ,  $G_k(x) > 0$  car les zéros de  $G_k$  sont situés dans l'intervalle  $(-\infty, \sqrt{n\pi})$  et le polynôme  $H_k$  a un coefficient dominant positif.

Partant de l'inégalité (4.70), le Lemme 4.4.4 est démontré, dès lors que pour

$$k_0 = \lfloor \tau n \frac{\pi}{2} \rfloor,$$

on a

$$G_{k_0}(\sqrt{n\pi}) \leq C_1 e^{-C_2 n}. \quad (4.72)$$

Démontrons cette propriété. D'abord pour tout  $k$ ,  $G_k(x)$  est décroissante pour  $x \geq \sqrt{\mu_k}$ . En effet, lorsque  $x \geq \sqrt{\mu_k}$ ,  $H_k(x) > 0$  et donc  $G_k(x) > 0$ . De plus, pour un tel  $x$ ,  $G_k''(x) = (x^2 - \mu_k)G_k(x) \geq 0$ . La fonction  $G_k$  est donc convexe sur  $[\sqrt{\mu_k}, +\infty[$ . Montrons maintenant qu'une fonction convexe qui tend vers 0 en  $+\infty$  est décroissante. Supposons qu'il existe  $x_1 \geq \sqrt{\mu_k}$  tel que  $G_k'(x_1) > 0$ , et soit  $x \geq x_1$ . On a

$$\begin{aligned} G_k(x) - G_k(\sqrt{\mu_k}) &= \int_{\mu_k}^x G_k'(t) dt \\ &= \int_{\mu_k}^{x_1} G_k'(t) dt + \int_{x_1}^x G_k'(t) dt. \end{aligned}$$

Comme  $G_k$  est convexe, la fonction  $G_k'$  est croissante et :

$$G_k(x) - G_k(\sqrt{\mu_k}) \geq \int_{\mu_k}^{x_1} G_k'(t) dt + G_k'(x_1)(x - x_1).$$

Comme  $G_k(x) \rightarrow_{x \rightarrow +\infty} 0$ , on obtient une contradiction en faisant tendre  $x$  vers l'infinie dans l'inégalité ci-dessus. On vient donc de montrer que  $G_k$  est décroissante pour  $x \geq \sqrt{\mu_k}$ . Il suffit donc de démontrer que  $G_{k_0}(x_0) \leq C_1 e^{-C_2 n}$  avec  $\sqrt{n\pi} \geq x_0 \geq \sqrt{\mu_{k_0}}$ .

Nous concluons grâce au fait que  $\tau \in (0, 1)$ . Soit  $\alpha > 0$  tel que  $\cosh(\alpha) = \sqrt{\frac{1}{\tau}}$ . On pose  $x_0 := \sqrt{\mu_{k_0}} \cosh(\alpha)$ . On a ainsi pour tout  $n$  :

$$\sqrt{n\pi} \geq x_0 \geq \sqrt{\mu_{k_0}}$$

On peut estimer  $G_{k_0}(x_0)$  grâce à [15, p 201, formule 8.22.13] :

$$G_{k_0}(x_0) = \frac{\exp\left[\left(\frac{k_0}{2} - \frac{1}{4}\right)(2\alpha - \sinh(2\alpha))\right]}{\sqrt{\pi}2^{3/4}(k_0 - 1)^{1/4}\sqrt{\sinh(\alpha)}} \left(1 + O\left(\frac{1}{n}\right)\right),$$

où  $O\left(\frac{1}{n}\right)$  ne dépend que de  $n$  et de  $\tau$ . En utilisant que  $2\alpha - \sinh(2\alpha) < 0$ , on trouve, pour  $n \geq n_0$  avec  $n_0$  suffisamment grand,

$$G_{k_0}(x_0) \leq C_1 \exp(-C_2 n),$$

où  $C_1$  et  $C_2$  ne dépendent que de  $\tau$ .

■

Voici l'énoncé du lemme 4.4.3

**Lemme** Soit  $\delta > 0$ . Il existe un entier  $n_0 \in \mathbb{N}$  qui ne dépend que de  $\delta$ , tel que  $\forall n \geq n_0, \forall k \leq \frac{\pi}{2}n$ ,

$$\int_{\Omega_n} G_k^2 \geq 1 - \delta.$$

**Preuve**

Soit  $\delta > 0$ . Soit  $\varepsilon \in (0, \frac{\pi}{2})$  que l'on fixera plus tard, mais qui ne dépendra que de  $\delta$ . Prenons  $k \leq \frac{\pi}{2}n$ , on a alors  $\sqrt{n\pi} \geq \sqrt{\mu_k} \geq \sqrt{\mu_k} \cos(\varepsilon)$ . Par parité de  $G_k^2$ ,

$$\int_{\Omega_n} G_k^2(x) dx = 2 \int_0^{\sqrt{n\pi}} G_k^2(x) dx \geq 2 \int_0^{\sqrt{\mu_k} \cos(\varepsilon)} G_k^2(x) dx.$$

On fait alors le changement de variable :  $x = \sqrt{\mu_k} \cos(\phi)$  avec  $\phi \in [\varepsilon, \frac{\pi}{2}]$ ,

$$\int_{\Omega_n} G_k^2 \geq 2 \int_{\varepsilon}^{\frac{\pi}{2}} G_k^2(\sqrt{\mu_k} \cos(\phi)) \sqrt{\mu_k} \sin(\phi) d\phi$$

Pour tout  $\phi \in (\varepsilon, \frac{\pi}{2})$ , on a la formule suivante pour  $G_k(\sqrt{\mu_k} \cos(\phi))$  (voir [15], page 201, Théorème 8.22.9)

$$G_k(\sqrt{\mu_k} \cos(\phi)) = \frac{2^{1/4}}{\sqrt{\pi}(k-1)^{1/4}\sqrt{\sin(\phi)}} \left( \sin \left[ \left( \frac{k}{2} - \frac{1}{4} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{4} \right] + r_{\varepsilon,k} \right),$$

où  $r_{\varepsilon,k}$  ne dépend que de  $\varepsilon$  et  $k$  et, à  $\varepsilon$  fixé,  $r_{\varepsilon,k} = O\left(\frac{1}{k}\right)$ . Donc

$$\begin{aligned} \int_{\Omega_n} G_k^2 &\geq \frac{2\sqrt{2}\sqrt{\mu_k}}{\pi\sqrt{k-1}} \int_{\varepsilon}^{\frac{\pi}{2}} \sin^2 \left[ \left( \frac{k}{2} - \frac{1}{4} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{4} \right] + r_{\varepsilon,k} d\phi \\ &\geq \frac{2}{\pi} \int_{\varepsilon}^{\frac{\pi}{2}} 1 - \cos \left[ \left( \frac{k}{2} - \frac{1}{4} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{4} \right] + r_{\varepsilon,k} d\phi \\ &\geq \left(1 - \varepsilon \frac{2}{\pi}\right) - \frac{2}{\pi} \int_{\varepsilon}^{\frac{\pi}{2}} \cos \left[ \left( \frac{k}{2} - \frac{1}{4} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{4} \right] + r_{\varepsilon,k} d\phi \end{aligned}$$

Donnons quelques idées sur le reste de la preuve. Lorsque  $k$  est 'grand', le terme

$$\cos \left[ \left( \frac{k}{2} - \frac{1}{4} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{4} \right]$$

est très oscillant et donc son intégrale tend vers 0. Quand  $k$  est 'petit', c'est-à-dire majoré par  $K$ , ne dépendant pas de  $n$ , nous avons la convergence suivante  $\int_{\Omega_n} G_k^2 \rightarrow_{n \rightarrow +\infty} 1$ , uniforme pour tout

$k \leq K$ .

$$\begin{aligned}
 \int_{\Omega_n} G_k^2 &\geq 1 - \varepsilon \frac{2}{\pi} - \frac{2}{\pi} \int_{\varepsilon}^{\frac{\pi}{2}} \cos \left[ \left( k - \frac{1}{2} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{2} \right] \frac{(2k-1)(\cos(2\phi)-1)}{(2k-1)(\cos(2\phi)-1)} + r_{\varepsilon,k} d\phi \\
 &\geq 1 - \varepsilon \frac{2}{\pi} - \frac{2}{\pi} \left[ \frac{\sin \left[ \left( k - \frac{1}{2} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{2} \right]}{(2k-1)(\cos(2\phi)-1)} \right]_{\varepsilon}^{\frac{\pi}{2}} \\
 &\quad + \frac{2}{\pi(2k-1)} \int_{\varepsilon}^{\frac{\pi}{2}} \frac{\sin \left[ \left( k - \frac{1}{2} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{2} \right] 2 \sin(2\phi)}{(\cos(2\phi)-1)^2} d\phi + Cr_{\varepsilon,k} \\
 &\geq 1 - \varepsilon \frac{2}{\pi} + \frac{2}{\pi} \left[ \frac{\sin \left[ \left( k - \frac{1}{2} \right) (\sin(2\varepsilon) - 2\varepsilon) + \frac{3\pi}{2} \right]}{(2k-1)(\cos(2\varepsilon)-1)} \right] \\
 &\quad + \frac{2}{\pi(2k-1)} \int_{\varepsilon}^{\frac{\pi}{2}} \frac{\sin \left[ \left( k - \frac{1}{2} \right) (\sin(2\phi) - 2\phi) + \frac{3\pi}{2} \right] \cos(\phi)}{\sin^3(\phi)} d\phi + Cr_{\varepsilon,k} \\
 &\geq 1 - \varepsilon \frac{2}{\pi} - \frac{2}{\pi(2k-1)(1-\cos(2\varepsilon))} - \frac{1}{(2k-1)\sin^3(\varepsilon)} + Cr_{\varepsilon,k}
 \end{aligned}$$

1. On choisit  $\varepsilon > 0$  tel que  $\varepsilon \frac{2}{\pi} = \frac{\delta}{3}$ , c'est à dire  $\varepsilon = \frac{\pi\delta}{6}$ .
2. On prend  $K(\delta) \in \mathbb{N}$  ne dépendant que de  $\varepsilon$  (donc que de  $\delta$ ) suffisamment grand tel que

$$\forall k \geq K(\delta), \frac{2}{\pi(2k-1)(1-\cos(2\varepsilon))} + \frac{1}{(2k-1)\sin^3(\varepsilon)} + Cr_{\varepsilon,k} \leq \frac{\delta}{3}.$$

Notons que  $n$  doit être choisi suffisamment grand lui aussi car  $k \leq \frac{\pi}{2}n$ .

3. On a donc montré :

$$\exists n_1(\delta) \in \mathbb{N}, \forall n \geq n_1(\delta), \forall k \in \mathbb{N}^*, \frac{\pi}{2}n \geq k \geq K(\delta), \|G_k\|_{L^2(\Omega_n)}^2 \geq 1 - \delta.$$

4. On sait que  $\|G_k\|_{L^2(\mathbb{R})} = 1$ , donc :

$$\exists n_0(\delta) \in \mathbb{N}, n_0(\delta) \geq n_1(\delta), \forall n \geq n_0(\delta), \forall k < K(\delta), \|G_k\|_{L^2(\Omega_n)}^2 \geq 1 - \delta.$$

5. Conclusion :

$$\forall \delta > 0, \exists n_0(\delta) \in \mathbb{N}, \forall k \in \mathbb{N}^*, \frac{\pi}{2}n \geq k \geq 1, \|G_k\|_{L^2(\Omega_n)}^2 \geq 1 - \delta.$$

■

#### 4.5.2 Calculs de la preuve du Théorème 4.4.5

1. Soit  $f : [-1, 1] \rightarrow \mathbb{R}$  une fonction paire. Alors on a :

$$\begin{aligned}
 &\int_{-1}^1 f(y) \exp \left( \int_y^1 \frac{n^2 \pi^2 |s|}{t - n^2 \pi^2 s^2} \right) dy \\
 &\int_{-1}^0 f(y) \exp \left( \int_y^1 \frac{n^2 \pi^2 |s|}{t - n^2 \pi^2 s^2} \right) dy + \int_0^1 f(y) \exp \left( \int_y^1 \frac{n^2 \pi^2 |s|}{t - n^2 \pi^2 s^2} \right) dy = \\
 &\int_0^1 f(y) \exp \left( \int_{-y}^1 \frac{n^2 \pi^2 |s|}{t - n^2 \pi^2 s^2} \right) dy + \int_0^1 f(y) \exp \left( \int_y^1 \frac{n^2 \pi^2 s}{t - n^2 \pi^2 s^2} \right) dy = \\
 &\int_0^1 f(y) \exp \left( \int_{-y}^0 \frac{-n^2 \pi^2 s}{t - n^2 \pi^2 s^2} + \int_0^1 \frac{n^2 \pi^2 s}{t - n^2 \pi^2 s^2} \right) dy + \\
 &\int_0^1 f(y) \exp \left( \left[ \frac{-1}{2} \ln(t - n^2 \pi^2 s^2) \right]_y^1 \right) dy = \\
 &\int_0^1 f(y) \exp \left( \left[ \frac{1}{2} \ln(t - n^2 \pi^2 s^2) \right]_{-y}^0 + \left[ \frac{-1}{2} \ln(t - n^2 \pi^2 s^2) \right]_0^1 \right) dy + \\
 &\int_0^1 f(y) \exp \left( \ln \sqrt{\frac{t - n^2 \pi^2 y^2}{t - n^2 \pi^2}} \right) dy = \\
 &\int_0^1 f(y) \left( \frac{t}{\sqrt{(t - n^2 \pi^2 y^2)(t - n^2 \pi^2)}} + \sqrt{\frac{t - n^2 \pi^2 y^2}{t - n^2 \pi^2}} \right) dy.
 \end{aligned}$$

2. On veut évaluer

$$I_1 := \int_{-1}^1 \frac{n^2 \pi^2 |y|}{2(t - n^2 \pi^2 y^2)^2} \exp\left(\int_y^1 \frac{n^2 \pi^2 |s|}{t - n^2 \pi^2 s^2}\right) dy$$

On a donc d'après le premier point :

$$I_1 = \int_0^1 \frac{tn^2\pi^2 y}{2\sqrt{t - n^2\pi^2}(t - n^2\pi^2 y^2)^{5/2}} + \frac{n^2\pi^2 y}{2\sqrt{t - n^2\pi^2}(t - n^2\pi^2 y^2)^{3/2}} dy.$$

On peut calculer l'intégrale  $I_1$  explicitement :

$$\begin{aligned} I_1 &= \frac{t}{2\sqrt{t - n^2\pi^2}} \left[ \frac{1}{3(t - n^2\pi^2 y^2)^{3/2}} \right]_0^1 + \frac{1}{2\sqrt{t - n^2\pi^2}} \left[ \frac{1}{\sqrt{t - n^2\pi^2 y^2}} \right]_0^1 \\ &= \frac{1}{2\sqrt{t - n^2\pi^2}} \left[ \frac{t}{3(t - n^2\pi^2)^{3/2}} - \frac{1}{3\sqrt{t}} \right] + \frac{1}{2\sqrt{t - n^2\pi^2}} \left[ \frac{1}{\sqrt{t - n^2\pi^2}} - \frac{1}{\sqrt{t}} \right] \\ &= \frac{-2}{3\sqrt{t}\sqrt{t - n^2\pi^2}} + \frac{t}{6(t - n^2\pi^2)^2} + \frac{1}{2(t - n^2\pi^2)}. \end{aligned}$$

3. On veut maintenant évaluer

$$I_2 := \int_{-1}^1 \frac{1}{2\sqrt{t - n^2\pi^2 y^2}} \exp\left(\int_y^1 \frac{n^2\pi^2 s \cos(2\phi_n)}{t - n^2\pi^2 s^2}\right) dy$$

On utilise encore le premier point :

$$\begin{aligned} I_2 &= \int_0^1 \frac{t}{2(t - n^2\pi^2 y^2)\sqrt{t - n^2\pi^2}} + \frac{1}{2\sqrt{t - n^2\pi^2}} dy \\ &= \frac{1}{2\sqrt{t - n^2\pi^2}} \int_0^1 \frac{dy}{1 - \frac{n^2\pi^2 y^2}{t}} + \frac{1}{2\sqrt{t - n^2\pi^2}} \\ &= \frac{\sqrt{t}}{4n\pi\sqrt{t - n^2\pi^2}} \left[ \ln\left(\frac{1 + \frac{n\pi y}{\sqrt{t}}}{1 - \frac{n\pi y}{\sqrt{t}}}\right) \right]_0^1 + \frac{1}{2\sqrt{t - n^2\pi^2}} \\ &= \frac{\sqrt{t}}{4n\pi\sqrt{t - n^2\pi^2}} \ln\left(1 + \frac{2n\pi}{\sqrt{t} - n\pi}\right) + \frac{1}{2\sqrt{t - n^2\pi^2}} \end{aligned} \tag{4.73}$$

### 4.5.3 A propos de la Remarque 4.4.5

**Théorème 4.5.1** Soit  $\varepsilon > \frac{1}{\exp(\pi)-1} \approx 0.045$ . Il existe une constante  $c_\varepsilon > 0$  et un entier  $n_0 \in \mathbb{N}$ , tous deux ne dépendant que de  $\varepsilon$ , tels que  $\forall n \geq n_0, \forall k \geq \frac{\pi}{2}n(1 + \varepsilon)$ ,

$$\lambda_{k+1,n} - \lambda_{k,n} \geq c_\varepsilon \sqrt{\lambda_{k,n}}.$$

La preuve qui suit s'inspire de l'article [10].

#### Preuve

Avec le Lemme 4.4.5, nous avons,

$$\phi_n(\lambda_{k,n}, 1) - \phi_n(\lambda_{k,n}, -1) = -\pi k$$

et d'après (4.54), il vient

$$\int_{-1}^1 -\sqrt{\lambda_{k,n} - n^2\pi^2 x^2} + \frac{n^2\pi^2 x}{2(\lambda_{k,n} - n^2\pi^2 x^2)} \sin(2\phi_n(\lambda_{k,n}, x)) dx = -\pi k. \tag{4.74}$$

Posons

$$b_n(\lambda_{k,n}) = - \int_{-1}^1 \frac{n^2\pi^2 x}{2(\lambda_{k,n} - n^2\pi^2 x^2)} \sin(2\phi_n(\lambda_{k,n}, x)) dx.$$



Puisque  $k \geq \frac{\pi}{2}n(1 + \varepsilon)$ , on a d'après le Théorème 4.4.2 :  $\lambda_{k,n} \geq n^2\pi^2(1 + \varepsilon) - n\pi$ . Ainsi,

$$\begin{aligned} |b_n(\lambda_{k,n})| &\leq \int_0^1 \frac{n^2\pi^2 x}{(n^2\pi^2(1 + \varepsilon) - n\pi - n^2\pi^2 x^2)} dx \\ &= \frac{-1}{2} [\ln(n^2\pi^2(1 + \varepsilon) - n\pi - n^2\pi^2 x^2)]_0^1 \\ &= \frac{1}{2} \ln\left(\frac{1 + \varepsilon - \frac{1}{n\pi}}{\varepsilon - \frac{1}{n\pi}}\right) \end{aligned} \quad (4.75)$$

En reprenant (4.74), on trouve :

$$\sqrt{\lambda_{k,n}} \int_{-1}^1 \sqrt{1 - \frac{n^2\pi^2 x^2}{\lambda_{k,n}}} dx + b_n(\lambda_{k,n}) = \pi k. \quad (4.76)$$

et donc

$$\sqrt{\lambda_{k,n}} = \frac{\pi k - b_n(\lambda_{k,n})}{\int_{-1}^1 \sqrt{1 - \frac{n^2\pi^2 x^2}{\lambda_{k,n}}} dx}.$$

Cette dernière identité quantifie en fonction de  $\varepsilon$  la façon dont  $\lambda_{k,n}$  se rapproche du comportement asymptotique des valeurs propres du laplacien. En effet, en faisant tendre  $\varepsilon$  vers l'infini dans cette dernière égalité, on obtient bien que  $\lambda_{k,n} \sim_{k \rightarrow +\infty} \frac{k^2\pi^2}{4}$  car  $\sqrt{1 - \frac{n^2\pi^2 x^2}{\lambda_{k,n}}} \sim_{k \rightarrow +\infty} 1$ .

Faisons un développement en série entière du terme  $\sqrt{1 - \frac{n^2\pi^2 x^2}{\lambda_{k,n}}}$ . Posons pour  $l \geq 1$

$$\binom{1/2}{l} = \frac{1/2(1/2 - 1) \dots (1/2 - l + 1)}{l!},$$

et  $\binom{1/2}{0} = 1$ . On a alors pour tout  $x \in [-1, 1]$ :

$$\sqrt{1 - \frac{n^2\pi^2 x^2}{\lambda_{k,n}}} = \sum_{l \geq 0} \binom{1/2}{l} (-1)^l \left(\frac{n^2\pi^2 x^2}{\lambda_{k,n}}\right)^l$$

et comme  $\lambda_{k,n} > n^2\pi^2(1 + \varepsilon)$ , la série précédente converge uniformément (et même normalement) pour  $x \in [-1, 1]$ . Ainsi avec (4.76), et en intervertissant  $\sum$  et  $\int$ , on peut écrire que :

$$\begin{aligned} &\sqrt{\lambda_{k+1,n}} \sum_{l \geq 0} \binom{1/2}{l} \frac{2(-1)^l}{2l+1} \left(\frac{n^2\pi^2}{\lambda_{k+1,n}}\right)^l - \sqrt{\lambda_{k,n}} \sum_{l \geq 0} \binom{1/2}{l} \frac{2(-1)^l}{2l+1} \left(\frac{n^2\pi^2}{\lambda_{k,n}}\right)^l \\ &= \pi - b_n(\lambda_{k+1,n}) + b_n(\lambda_{k,n}), \end{aligned}$$

et,

$$\begin{aligned} &2(\sqrt{\lambda_{k+1,n}} - \sqrt{\lambda_{k,n}}) + \sum_{l \geq 1} \binom{1/2}{l} \frac{2(-1)^l}{2l+1} (n^2\pi^2)^l \left[ \left(\frac{1}{\sqrt{\lambda_{k+1,n}}}\right)^{2l-1} - \left(\frac{1}{\sqrt{\lambda_{k,n}}}\right)^{2l-1} \right] \\ &= \pi - b_n(\lambda_{k+1,n}) + b_n(\lambda_{k,n}). \end{aligned}$$

On remarque que  $\binom{1/2}{l} (-1)^l$  est toujours négatif pour  $l \geq 1$ , donc :

$$\begin{aligned} &2(\sqrt{\lambda_{k+1,n}} - \sqrt{\lambda_{k,n}}) + \sum_{l \geq 1} \left| \binom{1/2}{l} \right| \frac{2}{2l+1} (n^2\pi^2)^l \left[ (\sqrt{\lambda_{k,n}})^{-2l+1} - (\sqrt{\lambda_{k+1,n}})^{-2l+1} \right] \\ &= \pi - b_n(\lambda_{k+1,n}) + b_n(\lambda_{k,n}). \end{aligned} \quad (4.77)$$

On utilise maintenant le Théorème des accroissements finis sur l'intervalle  $[\sqrt{\lambda_{k,n}}, \sqrt{\lambda_{k+1,n}}] \subset \mathbb{R}_+^*$  avec la fonction  $x \mapsto (\sqrt{x})^{-2l+1}$  et la croissance de  $(\lambda_{k,n})_{k \geq 1}$  pour affirmer que

$$\frac{2l-1}{(\sqrt{\lambda_{k,n}})^{2l}} (\sqrt{\lambda_{k+1,n}} - \sqrt{\lambda_{k,n}}) \geq (\sqrt{\lambda_{k,n}})^{-2l+1} - (\sqrt{\lambda_{k+1,n}})^{-2l+1} \geq 0.$$

En combinant (4.77) avec (4.75),

$$2(\sqrt{\lambda_{k+1,n}} - \sqrt{\lambda_{k,n}}) \left( 1 + \sum_{l \geq 1} \frac{2l-1}{2l+1} \left| \binom{1/2}{l} \right| \left( \frac{n^2 \pi^2}{\lambda_{k,n}} \right)^l \right) \geq \pi - 2 \times \frac{1}{2} \ln \left( 1 + \frac{1}{\varepsilon} \right),$$

donc :

$$(\sqrt{\lambda_{k+1,n}} - \sqrt{\lambda_{k,n}}) \left( 1 + \sum_{l \geq 1} \left| \binom{1/2}{l} \right| \left( \frac{1}{1+\varepsilon} \right)^l \right) \geq \frac{\pi}{2} - \frac{1}{2} \ln \left( 1 + \frac{1}{\varepsilon} \right).$$

Ainsi lorsque  $\varepsilon > \frac{1}{\exp(\pi)-1}$ , la quantité  $c_\varepsilon := \pi - \ln \left( 1 + \frac{1}{\varepsilon} \right)$  est strictement positive, et donc

$$\sqrt{\lambda_{k+1,n}} - \sqrt{\lambda_{k,n}} \geq \frac{c_\varepsilon}{2},$$

d'où,

$$\lambda_{k+1,n} - \lambda_{k,n} \geq c_\varepsilon \sqrt{\lambda_{k,n}}.$$

■

## 4.6 Références

- [1] W. V. ASSCHE, *Entropy of hermite polynomials with application to the harmonic oscillator*. 148
- [2] K. BEAUCHARD, P. CANNARSA, AND R. GUGLIELMI, *Null controllability of Grushin-type operators in dimension two*, J. Eur. Math. Soc. (JEMS), 16 (2014), pp. 67–101, doi:10.4171/JEMS/428, <http://dx.doi.org/10.4171/JEMS/428>. 124, 125, 129, 130, 131, 135, 141
- [3] K. BEAUCHARD, J. DARDÉ, AND S. ERVEDOZA, *Minimal time issues for the observability of Grushin-type equations*. working paper or preprint, Jan. 2018, <https://hal.archives-ouvertes.fr/hal-01677037>. 124, 131, 132, 133
- [4] K. BEAUCHARD, L. MILLER, AND M. MORANCEY, *2D Grushin-type equations: minimal time and null controllable data*, J. Differential Equations, 259 (2015), pp. 5813–5845, doi:10.1016/j.jde.2015.07.007, <http://dx.doi.org/10.1016/j.jde.2015.07.007>. 129, 131, 147, 148
- [5] G. BIRKHOFF AND G. ROTA, *Ordinary differential equations*, Introductions to higher mathematics, Ginn, 1962, <https://books.google.fr/books?id=nfBvAAAAIAAJ>. 145
- [6] J.-M. CORON, *Control and nonlinearity*, vol. 136 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2007. 124, 125, 126, 129
- [7] R. DAUTRAY AND J.-L. LIONS, *Mathematical analysis and numerical methods for science and technology. Vol. 3*, Springer-Verlag, Berlin, 1990. Spectral theory and applications, With the collaboration of Michel Artola and Michel Cessenat, Translated from the French by John C. Amson. 137, 138, 139, 147
- [8] M. DUPREZ AND A. KOENIG, *Control of the grushin equation: non-rectangular control region and minimal time*, 2018, arXiv:arXiv:1807.01241. 133
- [9] S. ERVEDOZA AND E. ZUAZUA, *Sharp observability estimates for heat equations*, Archive for Rational Mechanics and Analysis, 202 (2011), pp. 975–1017, doi:10.1007/s00205-011-0445-8, <https://doi.org/10.1007/s00205-011-0445-8>. 131
- [10] B. HARRIS, *Asymptotics of eigenvalues for regular sturm-liouville problems*, Journal of Mathematical Analysis and Applications, 183 (1994), pp. 25 – 36, doi:http://dx.doi.org/10.1006/jmaa.1994.1128, <http://www.sciencedirect.com/science/article/pii/S0022247X84711280>. 145, 154
- [11] A. KOENIG, *Non-null-controllability of the grushin operator in 2d*, Comptes Rendus Mathématique, 355 (2017), pp. 1215 – 1235, doi:10.1016/j.crma.2017.10.021, <https://hal.archives-ouvertes.fr/hal-01654043>. 131, 133
- [12] J. PÖSCHEL AND E. TRUBOWITZ, *Inverse spectral theory*, vol. 130 of Pure and Applied Mathematics, Academic Press, Inc., Boston, MA, 1987. 137

- [13] M. REED AND B. SIMON, *Methods of modern mathematical physics. IV. Analysis of operators*, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1978. 139
- [14] W. RUDIN, *Analyse réelle et complexe: cours et exercices*, Sciences sup, Dunod, 1998, <https://books.google.fr/books?id=04nCPAAACAAJ>. 131
- [15] G. SZEGÓ, *Orthogonal polynomials*, American Mathematical Society, Providence, R.I., fourth ed., 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII. 151, 152
- [16] E. C. TITCHMARSH, *Eigenfunction expansions associated with second-order differential equations. Part I*, Second Edition, Clarendon Press, Oxford, 1962. 138
- [17] A. ZETTL, *Sturm-Liouville theory*, vol. 121 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2005. 137

## Résumé

Dans cette thèse, nous étudions la contrôlabilité à zéro de quelques systèmes paraboliques continus et semi-discretisés.

Nous considérons tout d'abord des systèmes en cascade d'équations paraboliques de la forme  $\partial_t - (\partial_x \gamma \partial_x + q)$ . La variable spatiale évolue dans un intervalle réel borné et ce système est semi-discretisé en espace par un schéma aux différences finies. En appliquant *la méthode des moments*, nous démontrons des résultats de contrôlabilité à zéro et de  $\phi(h)$  contrôlabilité à zéro, suivant les hypothèses formulées sur le maillage et les fonctions  $\gamma$  et  $q$ . Puis nous étendons ces résultats lorsque la variable d'espace évolue dans un domaine cylindrique, la zone de contrôle se situant dans une partie d'une section au bord du cylindre. Ce domaine cylindrique se décompose en un produit de deux espaces. Sur le premier, de dimension 1, nous appliquons les résultats décrits précédemment. Sur le second, nous appliquons *la méthode de Lebeau-Robbiano*. Cette approche permet à la fois de montrer que le problème discretisé est  $\phi(h)$  contrôlable à zéro et de retrouver un résultat de contrôlabilité à zéro sur le système continu.

Dans une autre partie, nous nous intéressons au temps minimal de contrôle à zéro de l'équation de Grushin posée sur un domaine rectangulaire dont le domaine de contrôle est une bande verticale. L'étude se ramène à une infinité dénombrable, indexée par le paramètre de Fourier  $n$ , de problèmes de contrôle à zéro d'équations paraboliques, traitée, ici encore, à l'aide de la méthode des moments.

Cette dernière requiert des estimations précises sur le spectre d'opérateurs de Sturm-Liouville. Nous établissons, d'une part, des minoration sur certaines quantités dépendant des fonctions propres de ces opérateurs et nous étudions d'autre part *la propriété de gap* de leurs valeurs propres. Pour appliquer la méthode des moments aux différents problèmes de contrôle présentés tout au long de ce mémoire, il est alors crucial que ces estimations soient uniformes tantôt en le paramètre de discrétisation tantôt en le paramètre  $n$ . La théorie spectrale de ces opérateurs constitue donc la clef de voûte de cette thèse. Les résultats présents dans ce mémoire sont illustrés et complétés par des calculs numériques, basés sur la méthode HUM.

## Abstract

In this thesis, we study the null controllability of some continuous and semi discretized parabolic systems.

We first consider cascade systems of parabolic equations of the form  $\partial_t - (\partial_x \gamma \partial_x + q)$ . The space variable belongs to a real and bounded interval and this system is semi-discretized in space by a finite differences scheme. Applying the so called *moments method*, we prove null controllability and  $\phi(h)$  null controllability results, depending on the hypotheses on the mesh and on functions  $\gamma$  and  $q$ . Then, we extend this results when the space variable belongs to a cylindrical domain which control zone is in a section at the border of the cylinder. This cylindrical domain is decomposed into a product of two spaces. On the first, of dimension 1, we apply the results described previously. On the second, we use *the Lebeau-Robbiano's procedure*. In this framework, we prove  $\phi(h)$  null controllability results on the discretized domain as well as null controllability results on the continuous problem.

In another section, we investigate the computation of minimal time of null controllability of Grushin's equation defined on a rectangular domain which control region is a vertical strip. This problem of control amounts to study a countably infinite family, indexed by the Fourier parameter  $n$ , of null control problems of parabolic equations, tackled, once again, with the moments method.

The latter requires precise estimates on the spectrum of Sturm-Liouville operators. We prove lower bounds on quantities depending on the eigenfunctions of these operators and we study *the gap property* of their eigenvalues. To tackle control problems addressed in this manuscript, it is crucial that our estimates are uniform with respect to the discretization parameter or the parameter  $n$ . Spectral theory of these operators is therefore the keystone of this thesis.

Our results are illustrated and complemented by numerical simulations, based on the HUM approach.