Numerical null controllability of the 1D heat equation: Carleman weights an duality

Enrique Fernandez-Cara, Arnaud Munch

▶ **To cite this version:**

Enrique Fernandez-Cara, Arnaud Munch. Numerical null controllability of the 1D heat equation: Carleman weights an duality. 2011. hal-00687887v2

HAL Id: hal-00687887

https://hal.science/hal-00687887v2

Preprint submitted on 3 Jul 2012 (v2), last revised 3 May 2013 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Numerical exact controllability of the 1D heat equation: Carleman weights and duality

Enrique Fernández-Cara* and Arnaud Münch†

Abstract

This paper deals with the numerical computation of distributed null controls for the 1D heat equation. The goal is to compute a control that drives (a numerical approximation of) the solution from a prescribed initial state at $t = 0$ exactly to zero at $t = T$. We extend the earlier contribution of Carthel, Glowinski and Lions [5], which is devoted to the computation of minimal L^2-norm controls. We start from some constrained extremal problems (involving unbounded weights in time of Carleman type) introduced by Fursikov and Imanuvilov [13]) and we apply appropriate duality techniques. Then, we introduce numerical approximations of the associated dual problems and we apply conjugate gradient algorithms. Finally, we present several experiments, we highlight the influence of the weights and we analyze this approach in terms of robustness and efficiency. We also compare the results to those in the previous paper [11], where primal methods were considered.

Keywords: Heat equation, null controllability, numerical solution, duality.

Mathematics Subject Classification (2010)- 35K35, 65M12, 93B40.

1 Introduction

We are concerned in this work with the null controllability problem for 1D heat equation. The state equation is the following:

$$\begin{cases} y_t - (a(x)y_x)_x + b(x,t)\,y = v1_\omega, & (x,t) \in (0,1)\times(0,T) \\ y(x,t) = 0, & (x,t) \in \{0,1\}\times(0,T) \\ y(x,0) = y_0(x), & x \in (0,1). \end{cases} \tag{1}$$

Here, $\omega \subset\subset (0,1)$ is a (small) non-empty open interval, 1_ω is the associated characteristic function, $T > 0$, $a \in L^\infty(0,1)$ with $a(x) \geq a_0 > 0$ a.e., $b \in L^\infty((0,1)\times(0,T))$ and $y_0 \in L^2(0,1)$. In (1), $v \in L^2(\omega\times(0,T))$ is the *control* and $y = y(x,t)$ is the associated *state.*

In the sequel, for any $\tau > 0$, we will denote by Q_τ, Σ_τ and q_τ the sets $(0,1)\times(0,\tau)$, $\{0,1\}\times(0,\tau)$ and $\omega\times(0,\tau)$, respectively. We will also use the following notation:

$$Ly := y_t - (a(x)y_x)_x + b(x,t)\,y, \qquad L^\star z := -z_t - (a(x)z_x)_x + b(x,t)\,z.$$

For any $y_0 \in L^2(0,1)$ and $v \in L^2(q_T)$, it is well-known that there exists exactly one solution y to (1), with the regularity

$$y \in C^0([0,T];L^2(0,1)) \cap L^2(0,T;H_0^1(0,1)).$$

*Dpto. EDAN, University of Sevilla, Aptdo. 1160, 41080 Sevilla, Spain. E-mail: cara@us.es. Partially supported by grant MTM2006-07932 (Spain).

†Laboratoire de Mathématiques, Université Blaise Pascal (Clermont-Ferrand 2), UMR CNRS 6620, Campus des Cézeaux, 63177 Aubière, France. E-mail: arnaud.munch@math.univ-bpclermont.fr.

Accordingly, for any final time $T > 0$, the associated null controllability problem at T is the following: for each $y_0 \in L^2(0,1)$, find $v \in L^2(q_T)$ such that the associated solution to (1) satisfies

$$y(x,T) = 0, \quad x \in (0,1). \tag{2}$$

The controllability of PDEs is an important area of research and has been the subject of many papers in recent years. For the most relevant references, in particular those concerning the existence and numerical approximation of null controls for linear and semilinear heat equations, see [11].

This paper is devoted to design and analyze numerical methods for the previous null controllability problem based on duality arguments. It may be viewed as a complement of [11], where the direct computation of null controls has been achieved; see below.

In the context of numerical controllability, so far, the approximation of controls of minimal L^2 norm has focused most of the attention. The first contribution was due to Carthel, Glowinski and Lions in [5], who replaced the original constrained minimization problem by an unconstrained extremal (dual) problem, *a priori* easier to solve. However, the resulting problem involves some dual spaces which are very difficult, if not impossible, to approximate numerically.

More precisely, the null control of minimal norm in $L^2(q_T)$ is given by $v = \varphi 1_\omega$, where φ solves the backward heat equation

$$\begin{cases} -\varphi_t - (a(x)\varphi_x)_x + b(x,t)\varphi = 0, & (x,t) \in (0,1)\times(0,T) \\ \varphi(x,t) = 0, & (x,t) \in \{0,1\}\times(0,T) \\ \varphi(x,T) = \varphi_T(x), & x \in (0,1) \end{cases} \tag{3}$$

and φ_T minimizes the strictly convex and coercive functional

$$\mathcal{I}(\varphi_T) := \frac{1}{2}\|\varphi\|^2_{L^2(q_T)} - (\varphi(\cdot\,,0), y_0)_{L^2(0,1)} \tag{4}$$

over the Hilbert space $\mathcal{H}$ defined by the *completion* of $L^2(0,1)$ with respect to the norm

$$\|\varphi_T\|_{\mathcal{H}} := \|\varphi\|_{L^2(q_T)}. \tag{5}$$

Notice that the mapping $\varphi_T \mapsto \|\varphi_T\|_{\mathcal{H}}$ defined by (5) is a semi-norm in $\mathcal{D}(0,1)$; in view of the unique continuation property, which holds for the solutions to (24), it is in fact a Hilbertian norm. Hence, we can certainly consider the completion of $\mathcal{D}(0,1)$ for this norm.

The coercivity of $\mathcal{I}$ in $\mathcal{H}$ is a consequence of the so-called *observability inequality*

$$\|\varphi(\cdot\,,0)\|^2_{L^2(0,1)} \le C \iint_{q_T} |\varphi|^2 \, dx\, dt \qquad \forall \varphi_T \in L^2(0,1), \tag{6}$$

that holds for some constant $C = C(\omega, T)$ and, in turn, this is a consequence of some appropriate *global Carleman* inequalities; see [13] and [9].

As discussed in length in [21] (see also [16, 19]), the minimization of $\mathcal{I}$ is numerically ill-posed, essentially because of the hugeness of $\mathcal{H}$. Notice that, in particular, $H^{-s}(0,1) \subset \mathcal{H}$ for any $s > 0$; see also [1], where the degree of ill-posedness is investigated in the boundary situation.

All this explains why in [5] the *approximate controllability* problem is considered and $\mathcal{I}$ is replaced by $\mathcal{I}_\epsilon$, where

$$\mathcal{I}_\epsilon(\varphi_T) := \mathcal{I}(\varphi_T) + \epsilon\|\varphi_T\|_{L^2(0,1)}$$

for any $\epsilon > 0$. Now, the minimizer $\varphi_{T,\epsilon}$ belongs to $L^2(0,1)$ and the corresponding control v_ϵ produces a state y_ϵ with $\|y_\epsilon(\cdot\,,T)\|_{L^2(0,1)} \le \epsilon$. But, as $\epsilon \to 0^+$, high oscillations are observed for the controls v_ϵ near the controllability time T, see [21].

In this paper, we consider the following well-posed extremal problem, introduced by Fursikov and Imanuvilov in [13]:

$$\left\{ \begin{array}{l} \text{Minimize} \quad J(y,v) := \dfrac{1}{2} \iint_{Q_T} \rho^2 |y|^2 \, dx \, dt + \dfrac{1}{2} \iint_{q_T} \rho_0^2 |v|^2 \, dx \, dt \\ \text{Subject to} \quad (y,v) \in \mathcal{C}(y_0, T). \end{array} \right. \tag{7}$$

Here, we denote by $\mathcal{C}(y_0, T)$ the linear manifold

$$\mathcal{C}(y_0, T) = \{ \, (y,v) : v \in L^2(q_T), \ y \text{ solves (1) and satisfies (2)} \, \}$$

and we assume (at least) that

$$\left\{ \begin{array}{l} \rho = \rho(x,t), \, \rho_0 = \rho_0(x,t) \text{ are continuous and } \geq \rho_* > 0 \text{ in } Q_T \text{ and} \\ \rho, \rho_0 \in L^\infty(Q_{T-\delta}) \quad \forall \delta > 0 \end{array} \right. \tag{8}$$

(hence, the weights ρ and ρ_0 can blow up as $t \to T^-$).

In order to find a solution to (7), we can apply methods of two kinds :

- Primal methods, that provide an optimal couple (y,v) satisfying the constraint $(y,v) \in \mathcal{C}(y_0,T)$ and usually rely on the characterization of optimality; they have been considered, analyzed and applied in [11].
- Dual methods, in the spirit of [5] (see also [15]), that rely on appropriate reformulations of (7) as unconstrained problems and use new (dual) variables. This is the objective of the present paper.

The paper is organized as follows.

In Section 2, we briefly overview the primal method based on optimality conditions for (7) and we then remind several technical results, namely Carleman type estimates, used in the sequel.

In Section 3, we apply the *Fenchel-Rockafellar* duality theory to (7). To this end, we first introduce some approximations that lead to well-posed dual problems (see proposition 3.2). We also prove that the solutions to the latter converge, in an appropriate sense, to the solution to the original problem (7) (see Propositions 3.1 and 3.3).

In Section 4, we apply gradient methods in this dual framework. More precisely, Section 4.1 is concerned with a conjugate gradient type algorithm, while Section 4.2 deals with the finite dimensional approximation of the control problems.

In Section 5, we present several numerical experiments that show that the behavior of the considered algorithms is satisfactory. Finally, some further comments, additional results and concluding remarks are given in Section 6.

2 Overview of the primal method, technical results and notations

In the sequel, it is assume that

$$a \in C^1([0,1]), \quad a(x) \geq a_0 > 0 \quad \forall x \in [0,1]. \tag{9}$$

We have the following:

THEOREM 2.1 *For any $y_0 \in L^2(0,1)$ and any $T > 0$, there exists exactly one solution $(\hat{y}, \hat{v})$ to (7).*

The proof is standard. It relies on the facts that $\mathcal{C}(y_0,T)$ is a non-empty closed convex set of $L^2(Q_T)\times L^2(q_T)$ and $(y,v)\mapsto J(y,v)$ is strictly convex, proper and lower semi-continuous on the space $L^2(Q_T)\times L^2(q_T)$.

In [11], (7) is solved using a direct approach. There, the following weights are considered:

$$\left\{\begin{array}{l}\rho(x,t)=\exp\left(\dfrac{\beta(x)}{T-t}\right),\quad \rho_0(x,t)=(T-t)^{3/2}\rho(x,t),\quad \beta(x)=K_1\left(e^{K_2}-e^{\beta_0(x)}\right)\\ \text{where the } K_i \text{ are sufficiently large positive constants (depending on } T,\ a_0 \text{ and } \|a\|_{C^1})\\ \text{and } \beta_0\in C^\infty([0,1]),\ \beta_0>0 \text{ in } (0,1),\ \beta_0(0)=\beta_0(1)=0,\ \text{Supp } \beta_0'\subset\omega.\end{array}\right. \tag{10}$$

These functions blow up exponentially at $t=T$ and provide a very suitable solution to the original null controllability problem; this property, which can be seen as a reinforcement of (2), ensures the well-posedness of the variational formulation associated to the primal method. Weights of this kind were determined and systematically used by Fursikov and Imanuvilov in [13].

The direct approach is based on the following result:

PROPOSITION 2.1 *Assume that a satisfies* (9) *and let* ρ *and* ρ_0 *be given by* (10). *Let* (y,v) *be the corresponding optimal state-control pair. Then there exists* $p\in P$ *such that*

$$y=\rho^{-2}L^\star p,\qquad v=-\rho_0^{-2}p|_{q_T}. \tag{11}$$

The function p *is the unique solution of*

$$\left\{\begin{array}{l}\displaystyle\iint_{Q_T}\rho^{-2}L^\star p\,L^\star q\,dx\,dt+\iint_{q_T}\rho_0^{-2}p\,q\,dx\,dt=\int_0^1 y_0(x)\,q(x,0)\,dx\\ \forall q\in P;\quad p\in P.\end{array}\right. \tag{12}$$

Let us explain why (12) possesses exactly one solution. In the sequel, unless otherwise specified, we take ρ and ρ_0 as in (10). Let us introduce the linear space $P_0=\{\,q\in C^2(\overline{Q}_T):q=0\ \text{ on }\ \Sigma_T\,\}$. Then the bilinear form

$$(p,q)_P:=\iint_{Q_T}\rho^{-2}L^\star p\,L^\star q\,dx\,dt+\iint_{q_T}\rho_0^{-2}p\,q\,dx\,dt$$

is a scalar product. Indeed, if we have $q\in P_0$, $L^\star q=0$ in Q_T and $q=0$ in q_T then, by the *unique continuation property,* we have $q\equiv 0$.

Let P be the completion of P_0 for this scalar product. The well-posedness of (12) relies on the following two lemmas proved in [13] (see also [9]) and [11], respectively:

LEMMA 2.1 *Assume that a satisfies* (9) *and let* ρ *and* ρ_0 *be given by* (10). *Let us also set*

$$\rho_1(x,t)=(T-t)^{1/2}\rho(x,t),\quad \rho_2(x,t)=(T-t)^{-1/2}\rho(x,t).$$

Then there exists $C>0$, *only depending on* ω, T, a_0 *and* $\|a\|_{C^1}$, *such that*

$$\left\{\begin{array}{l}\displaystyle\iint_{Q_T}\left[\rho_2^{-2}\left(|q_t|^2+|q_{xx}|^2\right)+\rho_1^{-2}|q_x|^2+\rho_0^{-2}|q|^2\right]dx\,dt\\ \qquad\qquad\displaystyle\le C\left(\iint_{Q_T}\rho^{-2}|L^\star q|^2\,dx\,dt+\iint_{q_T}\rho_0^{-2}|q|^2\,dx\,dt\right)\end{array}\right. \tag{13}$$

for all $q\in P$.

LEMMA 2.2 *Let the assumptions of Lemma 2.1 hold. Then, for any* $\delta>0$, *one has*

$$P\hookrightarrow C^0([0,T-\delta];H_0^1(0,1)),$$

where the embedding is continuous. In particular, there exists $C > 0$, only depending on ω, T, a_0 and $\|a\|_{C^1}$, such that

$$\|q(\cdot\,,0)\|^2_{H_0^1(0,1)} \le C\left(\iint_{Q_T} \rho^{-2}|L^\star q|^2\,dx\,dt + \iint_{q_T} \rho_0^{-2}|q|^2\,dx\,dt\right) \tag{14}$$

for all $q \in P$.

In view of Proposition 2.1, the task is reduced to the resolution of the variational equality (12). This is the weak formulation of the following elliptic boundary value, which is fourth-order in space and second-order in time:

$$\begin{cases} L(\rho^{-2}L^\star p) + \rho_0^{-2} p\,1_\omega = 0, & (x,t) \in (0,1)\times(0,T) \\ p(x,t) = 0, \quad (-\rho^{-2}L^\star p)(x,t) = 0 & (x,t) \in \{0,1\}\times(0,T) \\ (-\rho^{-2}L^\star p)(x,0) = y_0(x), \quad (-\rho^{-2}L^\star p)(x,T) = 0, & x \in (0,1). \end{cases}$$

Let us emphasize that the variational problem (12) is equivalent to the minimization over P of the functional

$$I(p) := \frac{1}{2}\iint_{Q_T} \rho^{-2}|L^\star p|^2\,dx\,dt + \frac{1}{2}\iint_{q_T} \rho_0^{-2}|p|^2\,dx\,dt - \int_0^1 y_0(x)\,p(x,0)\,dx. \tag{15}$$

Once p is determined, the optimal state-control pair for J can be computed using (11). The variational equality (12) has been solved numerically in [11] by intriducing an appropriate finite element method in the space-time domain Q_T. This requires the inversion of a square, definite positive, symmetric matrix. Moreover, it is proved there that the approximation converges strongly as the discretization parameters go to zero. The main drawback is that this primal approach requires the use of C^1 finite elements in space.

3 Approximation and Duality

In this section, we use the *Fenchel-Rockafellar* duality approach to convex optimization (see [7]) in order to formulate the null controllability problem as an *unconstrained* extremal problem with good properties. The arguments we are going to present generalize those in [5] and [14].

The main reason for using duality in the context of (7) is that it is difficult to construct minimizing sequences; in fact, it is already difficult to construct couples (y, v) in $\mathcal{C}(y_0, T)$.

However, it is not clear how to apply the Fenchel-Rockafellar techniques to (7) directly, mainly because ρ blows up as $t \to T^-$; recall that the problem considered in [5] corresponds to $\rho \equiv 0$ and $\rho_0 \equiv 1$.

Accordingly, we first work with well chosen approximations depending on appropriate parameters and then analyze what happens in the limit.

For each $R > 0$, we first consider the following problem:

$$\begin{cases} \text{Minimize} \quad J_R(y,v) := \dfrac{1}{2}\displaystyle\iint_{Q_T} T_R(\rho)^2|y|^2\,dx\,dt + \frac{1}{2}\iint_{q_T} \rho_0^2|v|^2\,dx\,dt \\ \text{Subject to} \quad (y,v) \in \mathcal{C}(y_0,T). \end{cases} \tag{16}$$

Here, we have used the notation $T_R(\rho) := \min(\rho, R)$. Notice that (16) is a new constrained extremal problem; again, it possesses exactly one solution (y_R, v_R).

PROPOSITION 3.1 *For any $R > 0$, let (y_R, v_R) be the unique solution to (16). One has*

$$y_R \to \hat{y} \quad \textit{strongly in } L^2(Q_T) \quad \textit{and} \quad v_R \to \hat{v} \quad \textit{strongly in } L^2(q_T) \quad \textit{as} \quad R \to +\infty, \tag{17}$$

where $(\hat{y}, \hat{v})$ is the unique solution to (7).

PROOF: First, notice that

$$J_R(\hat{y},\hat{v}) = \frac{1}{2}\left(\iint_{Q_T} T_R(\rho)^2|\hat{y}|^2 + \iint_{q_T}\rho_0^2|\hat{v}|^2\right) \le J(\hat{y},\hat{v})$$

for all $R>0$. Consequently, the solutions to the problems (16) satisfy

$$J_R(y_R,v_R) = \frac{1}{2}\left(\iint_{Q_T} T_R(\rho)^2|y_R|^2 + \iint_{q_T}\rho_0^2|v_R|^2\right) \le J(\hat{y},\hat{v}).$$

This shows that $T_R(\rho)y_R$ is uniformly bounded in $L^2(Q_T)$ and $\rho_0 v_R$ is uniformly bounded in $L^2(q_T)$. Therefore, at least for some subsequence one has

$$\rho_0 v_R \to w \ \text{ weakly in } L^2(q_T) \ \text{ and } \ T_R(\rho)y_R \to z \ \text{ weakly in } L^2(Q_T). \tag{18}$$

Let us set $\tilde{y} = \rho^{-1}z$ and $\tilde{v} = \rho_0^{-1}w$. Then it is clear from (18) that

$$v_R = \rho_0^{-1}(\rho_0 v_R) \to \tilde{v} \ \text{ weakly in } L^2(q_T) \ \text{ and } \ y_R = T_R(\rho)^{-1}(T_R(\rho)y_R) \to \tilde{y} \ \text{ weakly in } L^2(Q_T).$$

In fact, $\tilde{y}$ is the state associated to $\tilde{v}$ and y_R converges strongly to $\tilde{y}$. For every $(y',v') \in \mathcal{C}(y_0,T)$, one has

$$\begin{aligned} J(\tilde{y},\tilde{v}) &\le \frac{1}{2}\liminf_{R\to+\infty}\left(\iint_{Q_T} T_R(\rho)^2|y_R|^2 + \iint_{q_T}\rho_0^2|v_R|^2\right) \\ &\le \frac{1}{2}\lim_{R\to+\infty}\left(\iint_{Q_T} T_R(\rho)^2|y'|^2 + \iint_{q_T}\rho_0^2|v'|^2\right) = J(y',v'). \end{aligned} \tag{19}$$

Hence, $(\tilde{y},\tilde{v}) = (\hat{y},\hat{v})$. Finally, we also deduce from (18) that

$$\limsup_{R\to+\infty}\left(\iint_{Q_T} T_R(\rho)^2|y_R|^2 + \iint_{q_T}\rho_0^2|v_R|^2\right) \le J(\tilde{y},\tilde{v}),$$

whence we see that (17) holds. □

Once again, it is difficult to construct a minimizing sequence for (16). On the other hand, as shown below, the constraint $y(\cdot\,,T)=0$ is related to the existence of *multipliers* in a (very) large space, difficult to handle in practice.

For these reasons, it is also convenient to consider for any $R>0$ and $\varepsilon>0$ the following unconstrained *penalized* problem:

$$\left\{\begin{array}{l} \text{Minimize } \ J_{R,\varepsilon}(y,v) := \dfrac{1}{2}\displaystyle\iint_{Q_T} T_R(\rho)^2|y|^2\,dx\,dt + \frac{1}{2}\iint_{q_T}\rho_0^2|v|^2\,dx\,dt + \frac{1}{2\varepsilon}\|y(\cdot\,,T)\|_{L^2}^2 \\ \text{Subject to } \ (y,v)\in\mathcal{A}(y_0,T) \end{array}\right. \tag{20}$$

where

$$\mathcal{A}(y_0,T) = \{\,(y,v) : v\in L^2(q_T), \ \ y \text{ solves (1)}\,\}.$$

Let us now associate to (20) a dual problem. To this end, we denote by $\overline{y}$ the solution to (1) with $v=0$ and we introduce the operators $M\in\mathcal{L}(L^2(q_T);L^2(Q_T))$ and $B\in\mathcal{L}(L^2(q_T);L^2(\Omega))$, with

$$Mv = z_v, \quad Bv = z_v(\cdot\,,T)$$

for all $v\in L^2(q_T)$, where z_v is the solution to

$$Lz_v = v\,1_\omega \quad \text{in} \quad Q_T, \qquad z_v = 0 \quad \text{on} \quad \Sigma_T, \qquad z_v(\cdot\,,0) = 0. \tag{21}$$

The solution to (1) may be decomposed as follows:

$$y = Mv + \overline{y}. \tag{22}$$

Obviously, M and B are linear and bounded on $L^2(q_T)$. Their *adjoint* operators $M^\star$ and $B^\star$ are given as follows:

- For each $\mu \in L^2(Q_T)$, $M^\star \mu = \zeta|_{q_T}$, where ζ is the solution to the backwards system

$$L^\star \zeta = \mu \quad \text{in} \quad Q_T, \qquad \zeta = 0 \quad \text{on} \quad \Sigma_T, \qquad \zeta(\cdot\,, T) = 0. \tag{23}$$

- For each $\varphi_T \in L^2(0,1)$, $B^\star \varphi_T = \varphi|_{q_T}$, where φ is the solution to

$$L^\star \varphi = 0 \quad \text{in} \quad Q_T, \qquad \varphi = 0 \quad \text{on} \quad \Sigma_T, \qquad \varphi(\cdot\,, T) = \varphi_T. \tag{24}$$

The announced dual (and well-posed) problem to (20) is the following one:

$$\left\{ \begin{array}{l} \text{Minimize} \quad J^\star_{R,\varepsilon}(\mu, \varphi_T) := \dfrac{1}{2}\left(\displaystyle\iint_{Q_T} T_R(\rho)^{-2} |\mu|^2 \, dx\, dt + \iint_{q_T} \rho_0^{-2} |\varphi|^2 \, dx\, dt \right) \\ \qquad\qquad\qquad\qquad + \displaystyle\int_0^1 \varphi(x,0)\, y_0(x)\, dx + \frac{\varepsilon}{2} \|\varphi_T\|^2_{L^2(0,1)} \\ \text{Subject to} \quad (\mu, \varphi_T) \in L^2(Q_T) \times L^2(0,1) \end{array} \right. \tag{25}$$

where, for any $(\mu, \varphi_T) \in L^2(Q_T) \times L^2(0,1)$, we have set $\varphi = M^\star \mu + B^\star \varphi_T$; therefore, φ is the solution to

$$L^\star \varphi = \mu \quad \text{in} \quad Q_T, \qquad \varphi = 0 \quad \text{on} \quad \Sigma_T, \qquad \varphi(\cdot\,, T) = \varphi_T. \tag{26}$$

In the following result, we explain how (25) is related to (20):

PROPOSITION 3.2 *The unconstrained extremal problem* (25) *is the dual problem of* (20)*, in the sense of the Fenchel-Rockafellar theory. Furthermore,* (20) *and* (25) *possess unique solutions. If we denote by* $(y_{R,\varepsilon}, v_{R,\varepsilon})$ *the unique solution to* (20)*, we denote by* $(\mu_{R,\varepsilon}, \varphi_{T,R,\varepsilon})$ *the unique solution to* (25) *and we set* $\varphi_{R,\varepsilon} = M^\star \mu_{R,\varepsilon} + B^\star \varphi_{T,R,\varepsilon}$*, the following relations hold:*

$$v_{R,\varepsilon} = \rho_0^{-2} \varphi_{R,\varepsilon}|_{q_T}, \quad y_{R,\varepsilon} = -T_R(\rho)^{-2} \mu_{R,\varepsilon}, \quad y(\cdot\,, T) = -\varepsilon \varphi_{T,R,\varepsilon}. \tag{27}$$

PROOF: In view of the structures of the previous extremal problems, the proof of this result can be obtained from standard results in optimal control theory; see for instance [4, 12, 24]. However, for completeness, we will provide a proof that uses Fenchel-Rockafellar theory.

From the decomposition (22), we can write, for any $(y, v) \in \mathcal{A}(y_0, T)$ that $J_{R,\varepsilon}(y, v) = F_{R,\varepsilon}(Mv, Bv) + G(v)$ where the functions $F_{R,\varepsilon}$ and G are defined by

$$F_{R,\varepsilon}(z, z_T) = \frac{1}{2} \iint_{Q_T} T_R(\rho)^2 |z + \overline{y}|^2 \, dx\, dt + \frac{1}{2\varepsilon} \int_0^1 |z_T(x) + \overline{y}(x,T)|^2 \, dx$$

and

$$G(v) = \frac{1}{2} \iint_{q_T} \rho_0^2 |v|^2 \, dx\, dt.$$

Let $V := L^2(Q_T) \times L^2(q_T)$. The functions $F_{R,\varepsilon} : V \mapsto \mathbb{R}$ and $G : L^2(q_T) \mapsto \mathbb{R}$ are both convex and continuous and we can apply the duality Theorem of W. Fenchel and T.R. Rockafellar; see Theorem 4.2 p. 60 in [7]. We deduce that

$$\inf_{(y,v) \in \mathcal{A}(y_0,T)} J_{R,\varepsilon}(y, v) = - \inf_{(\mu, \varphi_T) \in V} \Big\{ G^\star(M^\star \mu + B^\star \varphi_T) + F^\star_{R,\varepsilon}(-(\mu, \varphi_T)) \Big\},$$

where $F^\star_{R,\varepsilon}$ and $G^\star$ are the convex conjugate functions of $F_{R,\varepsilon}$ and G, respectively.

Notice that

$$\begin{aligned} F^\star_{R,\varepsilon}(\mu,\varphi_T) &= \sup_V\left\{\iint_{Q_T}\mu\, z\,dx\,dt+\int_0^1\varphi_T(x)\,z_T(x)\,dx-F(z,z_T)\right\}\\ &= \frac12\iint_{Q_T}T_R(\rho)^{-2}|\mu|^2\,dx\,dt-\iint_{Q_T}\mu\,\overline{y}\,dx\,dt+\frac{\varepsilon}{2}\|\varphi_T\|^2_{L^2}-\int_0^1\varphi_T(x)\,\overline{y}(x,T)\,dx \end{aligned}$$

for all $(\mu,\varphi_T)\in V$. On the other hand,

$$G^\star(w)=\frac12\iint_{q_T}\rho_0^{-2}|w|^2\,dx\,dt$$

for all $w\in L^2(Q_T)$. Therefore,

$$\begin{aligned} G^\star(M^\star\mu+B^\star\varphi_T)+F^\star_{R,\varepsilon}(-(\mu,\varphi_T)) &= \frac12\iint_{Q_T}T_R(\rho)^{-2}|\mu|^2\,dx\,dt+\frac12\iint_{q_T}\rho_0^{-2}|\varphi|^2\,dx\,dt\\ &\quad+\frac{\varepsilon}{2}\|\varphi_T\|^2_{L^2}+\iint_{Q_T}\mu\,\overline{y}\,dx\,dt+\int_0^1\varphi_T(x)\,\overline{y}(x,T)\,dx \end{aligned}$$

where we have used again the notation $\varphi=M^\star\mu+B^\star\varphi_T$.

Finally, multiplying the state equation of (26) by $\overline{y}$ and integrating by parts, we obtain that

$$\iint_{Q_T}\mu\,\overline{y}\,dx\,dt+\int_0^1\varphi_T(x)\,\overline{y}(x,T)\,dx=\int_0^1\varphi(x,0)\,y_0(x)\,dx,$$

whence

$$\begin{aligned} G^\star(M^\star\mu+B^\star\varphi_T)+F^\star_{R,\varepsilon}(-(\mu,\varphi_T)) &= \frac12\left(\iint_{Q_T}T_R(\rho)^{-2}|\mu|^2\,dx\,dt+\iint_{q_T}\rho_0^{-2}|\varphi|^2\,dx\,dt\right)\\ &\quad+\int_0^1\varphi(x,0)\,y_0(x)\,dx+\frac{\varepsilon}{2}\|\varphi_T\|^2_{L^2}. \end{aligned}$$

This proves that (25) is the dual of (20).

It is also easy to check that (20) and (25) are stable and possess unique solutions. Indeed, the hypotheses of Theorem 4.2 in [7] are satisfied for (20) (notice that this is not the case for (16), since the interior of the constraint set $\mathcal{C}(y_0,T)$ is empty).

Finally, let us deduce that the optimality conditions (27) hold.

Let us set $(y,v)=(y_{R,\varepsilon},v_{R,\varepsilon})$ and $(\mu,\varphi_T)=(\mu_{R,\varepsilon},\varphi_{T,R,\varepsilon})$. Then, since (25) and (20) are dual to each other, one has:

$$\begin{aligned} 0 &= \frac12\iint_{Q_T}T_R(\rho)^2|y|^2\,dx\,dt+\frac12\iint_{q_T}\rho_0^2|v|^2\,dx\,dt+\frac{1}{2\varepsilon}\|y(\cdot\,,T)\|^2_{L^2}\\ &\quad+\frac12\iint_{Q_T}T_R(\rho)^{-2}|\mu|^2\,dx\,dt+\frac12\iint_{q_T}\rho_0^{-2}|\varphi|^2\,dx\,dt+(\varphi(\cdot\,,0),y_0)+\frac{\varepsilon}{2}\|\varphi_T\|^2_{L^2}\\ &= \frac12\iint_{Q_T}T_R(\rho)^2|y+T_R(\rho)^{-2}\mu|^2\,dx\,dt+\frac12\iint_{q_T}\rho_0^2|v-\rho_0^{-2}\varphi|^2\,dx\,dt+\frac{1}{2\varepsilon}\|y(\cdot\,,T)+\varepsilon\varphi_T\|^2_{L^2}\\ &\quad-\iint_{Q_T}T_R(\rho)^2\mu\,y\,dx\,dt+\iint_{q_T}\rho_0^2\,v\,\varphi\,dx\,dt-(y(\cdot\,,T),\varphi_T)_{L^2}+(\varphi(\cdot\,,0),y_0)_{L^2}. \end{aligned}$$

But the terms in the last line cancel, since $\varphi=M^\star\mu+B^\star\varphi_T$. Consequently,

$$\iint_{Q_T}T_R(\rho)^2|y+T_R(\rho)^{-2}\mu|^2\,dx\,dt+\iint_{q_T}\rho_0^2|v-\rho_0^{-2}\varphi|^2\,dx\,dt+\frac{1}{\varepsilon}\|y(\cdot\,,T)+\varepsilon\varphi_T\|^2_{L^2}=0$$

and we get (27). □

We now justify the introduction of the parameter ε by analyzing the behavior of the solutions to the problems (20) as $\varepsilon \to 0^+$.

PROPOSITION 3.3 *With the notation of Proposition 3.2, for each fixed $R > 0$ one has*

$$y_{R,\varepsilon} \to y_R \ \text{ strongly in } L^2(Q_T) \ \text{ and } \ v_{R,\varepsilon} \to v_R \ \text{ strongly in } L^2(q_T) \ \text{ as } \ \varepsilon \to 0^+. \tag{28}$$

PROOF: First, notice that, for each $R > 0$ and $\varepsilon > 0$, one has

$$\iint_{Q_T} T_R(\rho)^2 |y_{R,\varepsilon}|^2 \, dx\, dt + \iint_{q_T} \rho_0^{-2} |\varphi_{R,\varepsilon}|^2 \, dx\, dt + \frac{1}{\varepsilon} \|y_{R,\varepsilon}(\cdot\,, T)\|_{L^2}^2 = (\varphi_{R,\varepsilon}(\cdot\,, 0), y_0)_{L^2}. \tag{29}$$

Indeed, taking into account the equations satisfied by $y_{R,\varepsilon}$ and $\varphi_{R,\varepsilon}$ and the identities (27), we find that the sum of the two integrals in the left hand side of (29) is equal to

$$\begin{aligned} \iint_{Q_T} (L^\star \varphi_{R,\varepsilon}\, y_{R,\varepsilon} - \varphi_{R,\varepsilon}\, L y_{R,\varepsilon}) \, dx\, dt &= - \, (\varphi_{R,\varepsilon}(\cdot\,, t), y_{R,\varepsilon}(\cdot\,, t)_{L^2}|_{t=0}^{t=T} \\ &= (\varphi_{R,\varepsilon}(\cdot\,, 0), y_0)_{L^2} - \frac{1}{\varepsilon} \|y_{R,\varepsilon}(\cdot\,, T)\|_{L^2}. \end{aligned}$$

We deduce that the left hand side of (29) is uniformly bounded. Indeed, we have

$$\begin{aligned} &\iint_{Q_T} T_R(\rho)^2 |y_{R,\varepsilon}|^2 \, dx\, dt + \iint_{q_T} \rho_0^{-2} |\varphi_{R,\varepsilon}|^2 \, dx\, dt + \frac{1}{\varepsilon} \|y_{R,\varepsilon}(\cdot\,, T)\|_{L^2}^2 \\ &\quad \le \|\varphi_{R,\varepsilon}(\cdot\,, 0)\|_{L^2} \, \|y_0\|_{L^2} \\ &\quad \le C \, \|y_0\|_{L^2} \left(\iint_{Q_T} \rho^{-2} T_R(\rho)^4 |y_{R,\varepsilon}|^2 \, dx\, dt + \iint_{q_T} \rho_0^{-2} |\varphi_{R,\varepsilon}|^2 \, dx\, dt \right)^{1/2} \\ &\quad \le C \, \|y_0\|_{L^2} \left(\iint_{Q_T} T_R(\rho)^2 |y_{R,\varepsilon}|^2 \, dx\, dt + \iint_{q_T} \rho_0^{-2} |\varphi_{R,\varepsilon}|^2 \, dx\, dt \right)^{1/2}. \end{aligned}$$

Therefore, $T_R(\rho)\, y_{R,\varepsilon}$ is uniformly bounded in $L^2(Q_T)$, $\rho_0\, v_{R,\varepsilon} = \rho_0^{-1} \varphi_{R,\varepsilon}|_{q_T}$ is uniformly bounded in $L^2(q_T)$, $\|y_{R,\varepsilon}(\cdot\,, T)\|_{L^2} \le C\varepsilon^{1/2}$ and, at least for some subsequence, one has

$$T_R(\rho)\, y_{R,\varepsilon} \to z_R = T_R(\rho)\, \tilde{y}_R \ \text{ weakly in } L^2(Q_T) \ \text{ and } \ \rho_0\, v_{R,\varepsilon} \to w_R = \rho_0\, \tilde{v}_R \ \text{ weakly in } L^2(q_T) \tag{30}$$

as $\varepsilon \to 0^+$.

Obviously, $\tilde{y}_R$ is the state associated to $\tilde{v}_R$ and $y_{R,\varepsilon}$ converges strongly to $\tilde{y}_R$ in $L^2(Q_T)$. Moreover, $\tilde{y}(\cdot\,, T) = 0$, that is, $(\tilde{y}_R, \tilde{v}_R) \in \mathcal{C}(y_0, T)$.

Now, arguing as in the proof of Proposition 3.1, it is not difficult to check that $(\tilde{y}_R, \tilde{v}_R)$ is the unique optimal pair of (16), i.e. $(\tilde{y}_R, \tilde{v}_R) = (y_R, v_R)$ and $v_{R,\varepsilon}$ also converges strongly. □

Remark 1 As a consequence of the way we have penalized the constraint (2), $J_{R,\varepsilon}$ is explicitly quadratic in $\|\varphi_T\|_{L^2(0,1)}$. This avoids the use of operator-splitting methods (see [14], Section 1.8.8). This does not affect the asymptotic limit in ε, since from (29), we get that the state $y_{R,\varepsilon}$ associated to $v_{R,\varepsilon} = \rho_0^{-2} \varphi_{R,\varepsilon}\, 1_\omega$ satisfies

$$\|y_{R,\varepsilon}(\cdot\,, T)\|_{L^2(0,1)} \le C_R \varepsilon^{1/2} \, \|y_0\|_{L^2(0,1)}$$

for some $C_R > 0$ that is uniformly bounded with respect to R. □

Remark 2 There are other ways to apply duality techniques to (7). For instance, we can use the fact that, if the first integral in (7) is finite, then (2) is necessarily satisfied. This leads to the extremal problem

$$\left\{ \begin{array}{l} \text{Minimize } \dfrac{1}{2}\displaystyle\iint_{q_T} \rho_0^{-2}|\zeta|^2\,dx\,dt + \frac{1}{2}\iint_{Q_T} \rho^{-2}|\mu|^2\,dx\,dt + \int_0^1 \zeta(x,0)\,y_0(x)\,dx \\ \text{Subject to } \mu \in L^2(Q_T) \end{array} \right. \tag{31}$$

where, for each $\mu \in L^2(Q_T)$, we have set $\zeta = M^\star\mu$; recall (23). However, this formulation is formal since the unique minimizer μ may not belong to $L^2(Q_T)$. □

Remark 3 Conversely, we can also get a dual problem to (20) where the unique (dual) variable is φ_T. Indeed, using again (22), we can decompose $J_{R,\varepsilon}$ as follows:

$$J_{R,\varepsilon}(y,v) = F_{1,R}(v) + F_{2,\varepsilon}(Bv),$$

where

$$F_1(v) = \frac{1}{2}\iint_{Q_T} T_R(\rho)^2|Mv+\overline{y}|^2\,dx\,dt + \frac{1}{2}\iint_{q_T}\rho_0^2|v|^2\,dx\,dt$$

and

$$F_{2,\varepsilon}(Bv) = \frac{1}{2\varepsilon}\|Bv+\overline{y}(\cdot\,,T)\|^2,$$

so that

$$\inf_{v\in L^2(q_T)}\{F_{1,R}(v)+F_{2,\varepsilon}(Bv)\} = -\inf_{\varphi_T\in L^2(0,1)}\{F_{1,R}^\star(B^\star\varphi_T)+F_{2,\varepsilon}^\star(-\varphi_T)\}.$$

By introducing the mappings $\mathcal{B}_R$ and $\mathcal{A}_R$, with $\mathcal{B}_R\varphi_T := B^\star\varphi_T - M^\star(T_R(\rho)^2\overline{y})$ and $\mathcal{A}_R := M^\star(T_R(\rho)^2M) + \rho_0^2\,1_\omega$, it is not difficult to check that

$$F_{1,R}^\star(B_R^\star\varphi_T) = \frac{1}{2}\iint_{Q_T}(\mathcal{A}_R^{-1}\mathcal{B}_R(\varphi_T))\,\mathcal{B}_R(\varphi_T)\,dx\,dt - \frac{1}{2}\iint_{Q_T}T_R(\rho)^2|\overline{y}|^2\,dx\,dt$$

and

$$F_{2,\varepsilon}^\star(-\varphi_T) = \frac{\varepsilon}{2}\|\varphi_T\|^2 + \int_0^1 \varphi_T(x)\,\overline{y}(x,T)\,dx$$

for all $\varphi_T \in L^2(0,1)$. Consequently, an extremal problem that can be put in duality with (20) is the following:

$$\left\{ \begin{array}{l} \text{Minimize } \dfrac{1}{2}\displaystyle\iint_{Q_T}(\mathcal{A}_R^{-1}\mathcal{B}_R(\varphi_T))\,\mathcal{B}_R(\varphi_T)\,dx\,dt + \int_0^1 \varphi_T(x)\,\overline{y}(x,T)\,dx + \frac{\varepsilon}{2}\|\varphi_T\|^2 \\ \text{Subject to } \varphi_T \in L^2(0,1). \end{array} \right. \tag{32}$$

We check that with the weights ρ and ρ_0 respectively replaced by 0 and 1, leading to $\mathcal{A}_R = 1_\omega$ and $\mathcal{B}_R = B^\star$, we recover exactly the formulation considered in [5].

Problem (32) involves minimization only with respect to the variable $\varphi_T \in L^2(0,1)$. However, it requires the inversion of a nonlocal operator $\mathcal{A}$ and is therefore *a priori* harder to solve. □

In view of these convergence results, it seems that an appropriate way to solve (7) is to first find the solution to (25) and then apply the relations (27) for small ε and large R. This is confirmed by the experiments in Section 5.

We also observe that the solutions to problems (16) and (20) are close for small ε. In fact, the experiments below will show that the parameter ε is in some sense useless, since the presence of the weighted integral of μ in the functional of (25) suffices by itself to stabilize this extremal problem.

The term in ε ensures that the second argument of the optimal pair $(\mu_{R,\varepsilon}, \varphi_{T,R,\varepsilon})$ belongs to $L^2(0,1)$ and therefore allows to apply duality techniques in a rigorous way, without using "abstract" or "nonstandard" spaces. Nevertheless, we can affirm that, in the limiting case, the analogous of the functional $\mathcal{I}$ in (4) is

$$J^\star(\mu, \varphi_T) = \frac{1}{2}\iint_{q_T} \rho_0^{-2}|\varphi|^2\,dx\,dt + \frac{1}{2}\iint_{Q_T} \rho^{-2}|\mu|^2\,dx\,dt + \int_0^1 \varphi(x,0)\,y_0(x)\,dx, \tag{33}$$

to be minimized over $\mathcal{V}$, defined as the completion of $\mathcal{D}(Q_T)\times\mathcal{D}(0,1)$ with respect to the norm

$$(\mu, \varphi_T) \to \left(\iint_{Q_T} \rho^{-2}|\mu|^2\,dx\,dt + \iint_{q_T} \rho_0^{-2}|\varphi|^2\,dx\,dt\right)^{1/2}.$$

Let us also emphasize that the link with the primal problem is now clear, since we have $J^\star(\mu,\varphi_T) = I(-\varphi)$ (see (15)).

4 Conjugate gradient and numerical approximation

In this section, we address the numerical solution to the minimization problem (25). Following [5], the method combines conjugate gradient algorithms with finite difference and finite element approximations.

The problem we want to solve reads as follows: for given $\varepsilon, R > 0$, $y_0 \in L^2(0,1)$ and $T > 0$, minimize over the Hilbert space $V = L^2(Q_T)\times L^2(0,1)$ the functional

$$J^\star_{R,\varepsilon}(\mu,\varphi_T) = \frac{1}{2}\iint_{q_T} \rho_0^{-2}|\varphi|^2\,dx\,dt + \frac{1}{2}\iint_{Q_T} T_R(\rho)^{-2}|\mu|^2\,dx\,dt + \int_0^1 \varphi(x,0)\,y_0(x)\,dx + \frac{\varepsilon}{2}\|\varphi_T\|^2_{L^2},$$

where $\varphi = M^\star\mu + B^\star\varphi_T$, that is, φ is the solution to (26). By definition, it will be said that φ is the *adjoint state* associated to μ and φ_T.

Notice that, in view of the optimality condition (27), $\mu_{R,\varepsilon}$ satisfies $\mu_{R,\varepsilon} + T_R(\rho)^2 y_{R,\varepsilon} = 0$ and therefore must vanish on Σ_T.

We apply a conjugate gradient method, more robust that gradient (steepest descent) method for small discretization parameters. In this respect, we get that the Fréchet derivative of $J^\star_{R,\varepsilon}$ at (μ,φ_T) in the direction $(\mu',\varphi'_T)\in V$ is given by

$$DJ^\star_{R,\varepsilon}(\mu,\varphi_T)\cdot(\mu',\varphi'_T) = \iint_{Q_T} (z + T_R(\rho)^{-2}\mu)\,\mu'\,dx\,dt + \int_0^1 (z(x,T) + \varepsilon\varphi_T(x))\,\varphi'_T(x)\,dx,$$

where z is the unique solution to the following system:

$$L\,z = \rho_0^{-2}\varphi\,1_\omega \quad \text{in} \quad Q_T, \qquad z = 0 \quad \text{on} \quad \Sigma_T, \qquad z(\cdot,0) = y_0. \tag{34}$$

4.1 The conjugate gradient algorithm

Let us introduce the following symmetric and continuous bilinear form on V:

$$\begin{aligned} a_{R,\varepsilon}((\mu,\varphi_T),(\mu',\varphi'_T)) = {} & \iint_{q_T} \rho_0^{-2}(M^\star\mu + B^\star\varphi_T)(M^\star\mu' + B^\star\varphi'_T)\,dx\,dt + \iint_{Q_T} T_R(\rho)^{-2}\mu\mu'\,dx\,dt \\ & + \varepsilon\int_0^1 \varphi_T(x)\,\varphi'_T(x)\,dx \qquad \forall (\mu,\varphi_T),(\mu',\varphi'_T)\in V, \end{aligned}$$

so that one has

$$J^\star_{R,\varepsilon}(\mu,\varphi_T) = \frac{1}{2}a_{R,\varepsilon}((\mu,\varphi_T),(\mu,\varphi_T)) + \int_0^1 (M^\star\mu + B^\star\varphi_T)(x,0)\,y_0(x)\,dx, \quad \forall(\mu,\varphi_T)\in V.$$

For any $\varepsilon > 0$ and $R \in \mathbb{R}^+$, the form $a_{R,\varepsilon}$ is coercive on V. It is therefore appropriate to apply conjugate gradient methods to (25); see [15]. The *Polak-Ribière* version reads as follows:

STEP 0: INITIALIZATION

Let κ be a small and strictly positive real number.

We choose $(\mu^0, \varphi_T^0) \in V$ and we compute the gradient $g^0 = (g_1^0, g_2^0)$ of $J^\star_{R,\varepsilon}$ at (μ^0, φ_T^0) :

$$g_1^0 = z^0 + T_R(\rho)^{-2}\mu^0, \quad g_2^0 = z^0(\cdot\,, T) + \varepsilon\varphi_T^0$$

where z^0 solves, together with φ^0, the cascade system

$$\left\{ \begin{array}{llllllll} L^\star\varphi^0 = \mu^0 & \text{in} & Q_T, & \varphi^0 = 0 & \text{on} & \Sigma_T, & \varphi^0(\cdot\,, T) = \varphi_T^0 \\ Lz^0 = \rho_0^{-2}\varphi^0\, 1_\omega & \text{in} & Q_T, & z^0 = 0 & \text{on} & \Sigma_T, & z^0(\cdot\,, 0) = y_0. \end{array} \right.$$

If $\|g^0\|_V / \|(\mu^0, \varphi_T^0)\|_V \leq \kappa$, then we take $(\mu, \varphi_T) = (\mu^0, \varphi_T^0)$ and we stop; otherwise, we set

$$w^0 = (w_1^0, w_2^0) = g^0.$$

Then, for $n \geq 0$, assuming that (μ^n, φ_T^n), g^n and w^n are given, with $g^n \neq 0$ and $w^n \neq 0$, we compute $(\mu^{n+1}, \varphi_T^{n+1})$, g^{n+1} and (if necessary) w^{n+1} performing the following steps.

STEP 1: STEEPEST DESCENT

We set

$$\eta^n = \frac{DJ^\star_{R,\varepsilon}(\mu^n, \varphi_T^n) \cdot w^n}{a_{R,\varepsilon}(w^n, w^n)}$$

and we take

$$(\mu^{n+1}, \varphi_T^{n+1}) = (\mu^n, \varphi_T^n) - \eta^n w^n.$$

Then, we compute the gradient $g^{n+1} = (g_1^{n+1}, g_2^{n+1})$ of $J^\star_{R,\varepsilon}$ at $(\mu^{n+1}, \varphi_T^{n+1})$:

$$g_1^{n+1} = z^{n+1} + T_R(\rho)^{-2}\mu^{n+1}, \quad g_2^{n+1} = z^{n+1}(\cdot\,, T) + \varepsilon\varphi_T^n$$

where z^{n+1} solves, together with φ^{n+1}, the cascade system

$$\left\{ \begin{array}{llllllll} L^\star\varphi^{n+1} = \mu^{n+1} & \text{in} & Q_T, & \varphi^{n+1} = 0 & \text{on} & \Sigma_T, & \varphi^{n+1}(\cdot\,, T) = \varphi_T^{n+1} \\ Lz^{n+1} = \rho_0^{-2}\varphi^{n+1}\, 1_\omega & \text{in} & Q_T, & z^{n+1} = 0 & \text{on} & \Sigma_T, & z^{n+1}(\cdot\,, 0) = y_0. \end{array} \right.$$

STEP 2: CONVERGENCE TEST AND CONSTRUCTION OF THE NEW DIRECTION

If $\|g^{n+1}\|_V / \|g^0\|_V \leq \kappa$, then we take $(\mu, \varphi_T) = (\mu^{n+1}, \varphi_T^{n+1})$ and we stop; otherwise, we compute

$$\gamma_n = \frac{(g^{n+1} - g^n, g^{n+1})_V}{\|g^n\|_V^2}, \tag{35}$$

we take

$$w^{n+1} = g^{n+1} + \gamma_n w^n$$

and we return to Step 1 with n replaced by $n + 1$.

Remark 4 In the present quadratic-linear situation, the gradients g^n are conjugate to each other, that is, $(g^m, g^n)_V = 0, \forall m, n \geq 0$, $m \neq n$. The parameter γ_n given by (35) can then be written in the form

$$\gamma_n = \frac{\|g^{n+1}\|_V^2}{\|g^n\|_V^2}. \tag{36}$$

For non necessarily quadratic-linear extremal problems, the choices (35) and (36) are not equivalent; they respectively lead to the *Polak-Ribiere* and the *Fletcher-Reeves* conjugate gradient algorithms. In our case, due to the numerical approximation, the orthogonality of the g^n is lost and strongly accentuated for small values of ε and large values of R. In that stiff case, we observed that the *Polak-Ribiere* version, mainly used in nonlinear situations, is much more robust. □

Remark 5 With $T_R(\rho)$ and ρ_0 respectively replaced by the constants 0 and 1, we obtain exactly the conjugate gradient algorithm considered in Section 1.8 in [15], designed for the computation of the control of minimal norm in $L^2(q_T)$. Notice that the present situation does not lead to a significative increase of the computational cost. □

4.2 Full discrete approximations

For "large" integers N_x and N_t, we set $\Delta x = 1/N_x$, $\Delta t = T/N_t$ and $h = (\Delta x, \Delta t)$. Let us denote by $\mathcal{P}_{\Delta x}$ the uniform partition of $[0,1]$ associated to Δx and let us denote by $\mathcal{Q}_h$ the uniform quadrangulation of Q_T associated to h so that in particular $\overline{Q}_T = \bigcup_{K\in\mathcal{Q}_h} K$.

The following (conformal) finite element approximation of $L^2(0,T;H^1(0,1))$ is introduced:

$$X_h = \{\, \varphi_h \in C^0([0,1]\times[0,T]) : \varphi_h|_K \in (\mathbb{P}_{1,x}\otimes\mathbb{P}_{1,t})(K) \ \ \forall K \in \mathcal{Q}_h \,\}.$$

Here, $\mathbb{P}_{m,\xi}$ denotes the space of polynomial functions of order m in the variable ξ. Accordingly, the functions in X_h reduce on each quadrangle $K \in \mathcal{Q}_h$ to a linear polynomial in both x and t. The space X_h is a conformal approximation of $L^2(Q_T)$. We also consider the space

$$X_{0h} = \{\, \varphi_h \in X_h : \varphi_h(0,t) = \varphi_h(1,t) = 0 \ \ \forall t \in (0,T) \,\}.$$

X_{0h} is a finite-dimensional subspace of $L^2(0,T;H^1_0(0,1))$ and the functions $\varphi_h \in X_{0h}$ are uniquely determined by their values at the nodes (x_j,t_j) of $\mathcal{Q}_h$ such that $0 < x_j < 1$.

Let us now introduce other finite dimensional spaces. First, we set

$$\Phi_{\Delta x} = \{\, z \in C^0([0,1]) : z|_k \in \mathbb{P}_{1,x}(k) \ \ \forall k \in \mathcal{P}_{\Delta x} \,\}.$$

Then, $\Phi_{\Delta x}$ is a finite dimensional subspace of $L^2(0,1)$ and the functions in $\Phi_{\Delta x}$ are uniquely determined by their values at the nodes of $\mathcal{P}_{\Delta x}$. For any function $v \in C^0([0,1])$, we will denote by $\pi_{\Delta x}(v)$ the associated interpolated function, given by

$$\pi_{\Delta x}(v) \in \Phi_{\Delta x}, \quad \pi_{\Delta x}(v) = v \ \text{ at all points in } \ \mathcal{P}_{\Delta x}.$$

Secondly, since the variable μ appears in the right hand side of the backward equation $L^\star\varphi = \mu$, it is natural to approximate $\mu \in L^2(Q_T)$ by a piecewise constant function. Thus, let M_h be the space defined by

$$M_h = \{\, \mu_h \in L^2(Q_T) : \mu_h|_K \in (\mathbb{P}_{0,x}\otimes\mathbb{P}_{0,t})(K) \ \ \forall K \in \mathcal{Q}_h \,\}.$$

M_h is a finite dimensional subspace of $L^2(Q_T)$ and the functions in M_h are uniquely determined by their (constant) values on the quadrangles $K \in \mathcal{Q}_h$. For any function $f \in L^2(Q_T)$, we will denote by $\pi_h(f)$ the associated interpolated function, defined by

$$\pi_h(f) \in M_h, \quad \iint_K \pi_h(f)\,dx\,dt = \iint_K f\,dx\,dt \quad \forall K \in \mathcal{Q}_h.$$

For any h, we therefore consider the following approximation of (25):

$$\left\{\begin{array}{l} \text{Minimize } \ J^\star_{R,\varepsilon,h}(\mu_h,\varphi_{\Delta x,T}) = \dfrac{1}{2}\left(\displaystyle\iint_{Q_T} \pi_h(T_R(\rho)^{-2})|\mu_h|^2\,dx\,dt + \iint_{q_T} \pi_h(\rho_0^{-2})|\varphi_h|^2\,dx\,dt\right) \\ \qquad\qquad + \displaystyle\int_0^1 \varphi_h(x,0)\,\pi_{\Delta x}(y_0(x))\,dx + \frac{\varepsilon}{2}\|\varphi_{\Delta x,T}\|^2_{L^2(0,1)} \\ \text{Subject to } \ (\mu_h,\varphi_{\Delta x,T}) \in M_h \times \Phi_{\Delta x}. \end{array}\right. \tag{37}$$

In (37), for every $\mu_h \in M_h$ and every $\varphi_{\Delta x,T} \in \Phi_{\Delta x}$, we have denoted by φ_h the associated *discrete adjoint state.* By definition, $\varphi_h \in X_{0h}$ is given as follows:

(i) Let us introduce the times $t_j = j\Delta t$. We have $T = t_{N_t}$ and we first set $\varphi_h|_{t=T} = \varphi_{\Delta x,T}$.

(ii) Secondly, $\varphi_h|_{t=t_{N_t-1}}$ is the solution to the linear problem

$$\left\{\begin{array}{l} \displaystyle\int_0^1 \frac{1}{\Delta t}(\varphi - \varphi_{\Delta x,T})z\,dx + \frac{1}{2}\int_0^1 (\pi_{\Delta x}(a(x))\varphi_x z_x + \pi_{\Delta x} b(x,t_{N_t-1})\varphi z)\,dx \\ \qquad + \displaystyle\frac{1}{2}\int_0^1 (\pi_{\Delta x}(a(x))\varphi_{\Delta x,T,x} z_x + \pi_{\Delta x} b(x,t_{N_t})\varphi_{\Delta x,T,x} z)\,dx \\ \qquad = \displaystyle\frac{1}{2}\int_0^1 (\mu_h(x,t_{N_t-1}) + \mu_h(x,t_{N_t}))z(x)\,dx \quad \forall z \in \Phi_{\Delta x}, \quad \varphi \in \Phi_{\Delta x}. \end{array}\right.$$

(iii) Then, for given $n = N_t - 1, \ldots, 2$, $\varphi^\star = \varphi_h|_{t=t_{n+1}}$ and $\overline{\varphi} = \varphi_h|_{t=t_n}$, $\varphi_h|_{t=t_{n-1}}$ is the solution to the linear problem

$$\left\{\begin{array}{l} \displaystyle\int_0^1 \frac{1}{2\Delta t}(3\varphi - 4\overline{\varphi} + \varphi^\star)z\,dx + \int_0^1 (\pi_{\Delta_x}(a(x))\varphi_x z_x + \pi_{\Delta_x}(b(x,t_{n-1}))\varphi z)\,dx \\ \qquad = \displaystyle\int_0^1 \mu_h(x,t_{n-1})z(x)\,dx \quad \forall z \in \Phi_{\Delta x}, \quad \varphi \in \Phi_{\Delta x}. \end{array}\right.$$

We are thus using the two-step *implicit Gear* algorithm to solve numerically the adjoint problem (25). As advocated in [5], where the influence of the time discretization is highlighted, this second order scheme ensures a better behavior of the underlying conjugate gradient algorithm than, for instance, implicit Euler or Crank-Nicolson scheme.

For the computation of the gradient of $J^\star_{R,\varepsilon,h}$, we also need to solve numerically systems of the form (34). This is done in a similar way.

For any R and ε, the functional $J^\star_{R,\varepsilon,h}$ enjoys the same properties than $J^\star_{R,\varepsilon}$ when V is replaced by $V_h := X_h \times \Phi_{\Delta x}$. In particular $J^\star_{R,\varepsilon,h}$ is coercive in V_h, uniformly with respect to h. Hence, (37) may be solved with the conjugate gradient algorithm stated in Section 4.1.

We do not present here any convergence result for the variables $\mu_{h,R,\varepsilon}, \varphi_{\Delta x,T,R,\varepsilon}$ (the minimizer of $J^\star_{R,\varepsilon,h}$) as $h \to 0$. Actually, only partial results have been obtained and concern the particular case of minimal L^2-minimal norm case, that is $\rho_0 = 1$ and $\rho = 0$.

In this context, we mention the work of [18], where the null controllability for the heat equation with constant diffusion is proved for finite difference schemes in one spatial dimension on uniform meshes. In higher dimensions, discrete eigenfunctions may be an obstruction to the null controllability; see [26], where a counter-example for finite differences due to O.Kavian is described. A result of null controllability for a constant portion of the lower part of the discrete spectrum is given in [2]. In [17], in the context of approximate controllability, a relaxed observability inequality is given for general semi-discrete (in space) schemes, with the parameter ε of the order of Δx. The work [3] extends the results in [17] to the full discrete situation and proves the convergence of full discrete (approximated) controls toward a semi discrete one, as the time step Δ_t tends to zero. Let us also mention [8], where the authors prove that any controllable parabolic equation, be it discrete or continuous in space, is null controllable after time discretization upon the application of an appropriate filtering of the high frequencies.

To our knowledge, in the framework of duality, a convergence result similar to those in [11] for a sequence of discrete controls towards a null control of the infinite dimensional system (1) is still missing.

5 Numerical experiments

We present in this section some numerical experiments for problem (37).

The main data will be the following : $\omega = (0.3, 0.6)$, $y_0(x) = \sin(\pi x)$, $a \equiv a_0 = 1/10$, $b \equiv 1$ and $T = 1/2$. Moreover, we take $\Delta t = \Delta x$. We first briefly discuss the behavior of the computed control with respect to ε, R. Then, we analyze the influence of the weights ρ and ρ_0 on the behavior of the conjugate gradient method as $h \to (0,0)$. Finally, we consider a change of variable and we discuss its influence on the behavior of the algorithm.

For any $s \in (0,1)$, we consider the following function $\beta_{0,s}$:

$$\beta_{0,s}(x) = \frac{x(1-x)e^{-(x-c_s)^2}}{s(1-s)e^{-(s-c_s)^2}}, \quad c_s = s - \frac{1-2s}{2s(1-s)}. \tag{38}$$

We easily check that if s belongs to ω, then $\beta_{0,s}$ satisfies the conditions in (10): $\beta_{0,s}(0) = \beta_{0,s}(1) = 0$, $\beta_{0,s} > 0$ in $(0,1)$ and $|\beta'_{0,s}| > 0$ except at $x = s$. In the sequel, we take ρ and ρ_0 as in (10) with $\beta_0 = \beta_{0,s}$ for some s the mid point of the interval $\omega \subset (0,1)$ and with $K_1 = 1/10$ and $K_2 = 2\|\beta_0\|_{L^\infty(0,1)} = 2$.

It is important to notice that the weights ρ and ρ_0 appear in the conjugate functional $J^\star_{R,\varepsilon}$ only through their inverse. Therefore, the behavior of ρ and ρ_0 as $t \to T^-$ does not lead *a priori* to any numerical pathology when we work with $J^\star_{R,\varepsilon}$.

A constant diffusion coefficient is here considered. We refer to [11], where other situations (in particular, piecewise constant diffusion coefficients) are considered and discussed.

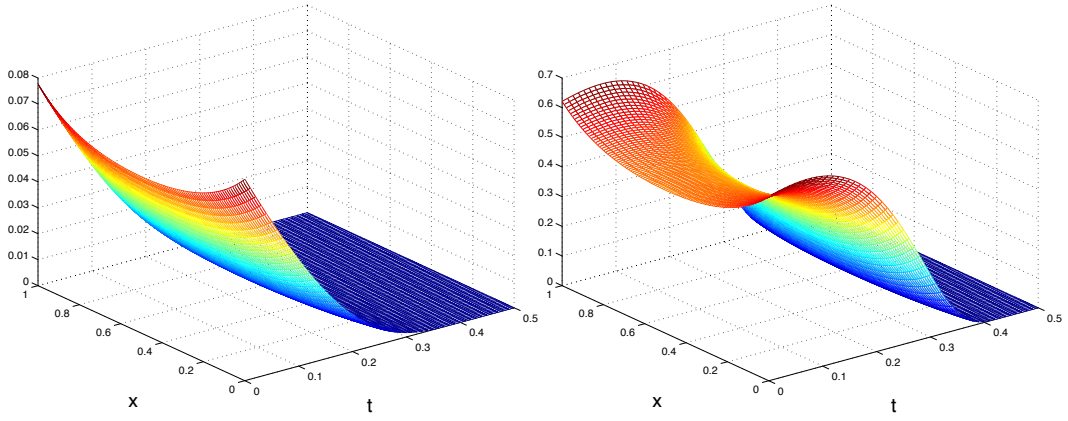

Figure 1: The weights ρ^{-2} and ρ_0^{-2} defined by (10), with $\beta_0 = \beta_{0,1/2}$ defined by (38), $K_1 = 1/10$ and $K_2 = 2\|\beta_0\|_{L^\infty(0,1)}$.

5.1 The behavior of $(v_{R,\varepsilon,h}, y_{R,\varepsilon,h})$ as $R \to +\infty$ and $\varepsilon \to 0$

For a fixed value of h sufficiently close to $(0,0)$, we first illustrate the convergence of the numerical functions $T_R(\rho)^{-2}\mu_{R,\varepsilon}$ and $\rho_0^{-2}\varphi_{R,\varepsilon}\,1_\omega$ as $R \to \infty$ and $\varepsilon \to 0$, in the sense stated in Propositions (3.1) and (3.3).

Once the unique minimizer $(\mu_{R,\varepsilon,h}, \varphi_{T,R,\varepsilon,h})$ of $J^\star_{R,\varepsilon,h}$ is obtained through the conjugate gradient algorithm described in Section 4.1, we compute the associated discrete adjoint solution $\varphi_{R,\varepsilon,h} = M^\star\mu_{R,\varepsilon,h} + B^\star\varphi_{T,R,\varepsilon,h}$ using $\mathbb{P}_1$-finite elements in space and the implicit Gear scheme in time, as discussed in Section 4.2. The control is then given by $v_{R,\varepsilon,h} = \rho_0^{-2}\varphi_{R,\varepsilon,h}\,1_\omega$. Finally, the controlled solution $y_{R,\varepsilon,h}$ is given by $y_{R,\varepsilon,h} = -T_R(\rho)^{-2}\mu_{R,\varepsilon,h}$ in Q_T, in accordance with the optimality relations (27).

For $h = (10^{-2}, 10^{-2})$, we show in Table 1 the behavior of the norms of $\mu_{R,\varepsilon,h}$ and $\varphi_{T,R,\varepsilon,h}$ with respect to ε and R. For each value of these parameters, we use the conjugate gradient algorithm with $\kappa = 10^{-4}$. This is small enough to guarantee a good approximation of the control but, obviously, does not allow

to fulfill exactly the optimality conditions (27). Notice however that the fact that κ (and h) is strictly positive allows to consider the limit cases $R = +\infty$ (for which $T_R(\rho)^2 = \rho^2$) and $\varepsilon = 0$ as well.

The algorithm is initialized with $\mu^0 \equiv 0$ and $\varphi_T^0 \equiv 0$.

	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$	$\varepsilon = 0$
$R = 10^4$	1.7754	2.0921	2.2570	2.3096
$R = 10^6$	2.1025	2.1583	2.3059	2.3330
$R = 10^8$	2.1619	2.1805	2.3127	2.3423
$R = +\infty$	2.1624	2.1807	2.3121	2.3410

	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$	$\varepsilon = 0$
$R = 10^4$	1.9251	1.8772	1.8613	1.8548
$R = 10^6$	1.8692	1.8651	1.8518	1.8493
$R = 10^8$	1.8636	1.8621	1.8510	1.8480
$R = +\infty$	1.8636	1.8621	1.8511	1.8482

Table 1: $L^2(q_T)$-norm of $\rho_0^{-2}\varphi_{R,\varepsilon,h}$ (**Top**) and $L^2(Q_T)$-norm of $-\rho_R^{-2}\mu_{R,\varepsilon,h}$ ($\times 10^{-1}$) (**Bottom**).

Table 1 reports $\|\rho_0^{-2}\varphi_{R,\varepsilon,h}\|_{L^2(q_T)}$ and $\|\rho_R^{-2}\mu_{R,\varepsilon,h}\|_{L^2(Q_T)}$ for $\varepsilon \in \{10^{-4}, 10^{-6}, 10^{-8}, 0\}$ and $R \in \{10^4, 10^6, 10^8, +\infty\}$. We check that these norms are uniformly bounded with respect to ε and R and both possess a limit as $\varepsilon \to 0$ and $R \to \infty$, in agreement with propositions 3.1 and 3.3.

For small values of ε (near $\varepsilon = 10^{-8}$), we observe that the parameter R has only a weak influence on the norm of $\rho_0^{-2}\psi_{R,\varepsilon,h}$; conversely, as soon as R is large enough (near $R = 10^6$), the norm of $\rho_R^{-2}\mu_{R,\varepsilon,h}$ is almost independent of ε. This is due to the choice of the weights ρ and ρ_0 and that any small ε and any large R reinforce in a suitable sense the null controllability property.

	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$	$\varepsilon = 0$
$R = 10^4$	652	1 427	4 447	7 532
$R = 10^6$	2 436	2 387	3 876	4 269
$R = 10^8$	2 928	2 595	3 112	5 662
$R = +\infty$	2 932	2 291	3 145	6 532

Table 2: The number of iterates to reach $\|g_h^n\|_V / \|g_h^0\|_V \leq \kappa = 10^{-4}$ vs. R and ε.

Table 2 provides the number of iterates needed to achieve $\|g_h^n\|_V / \|g_h^0\|_V \leq \kappa = 10^{-4}$, where g_h^n is the gradient of $J^\star_{R,\varepsilon,h}$. In agreement with the results and conclusions in [5] and [21], this number increases as $\varepsilon \to 0$ and/or $R \to +\infty$, which must be viewed as a numerical confirmation of the lack of uniform coercivity of $J^\star_{R,\varepsilon}$ in V. On the other hand, as soon as σ is small enough, depending on a_0, T and the size of ω, the conjugate algorithm fails to converge.

In agreement with the lack of uniform coercivity in V, the results in Tables 3 indicate that $\mu_{R,\varepsilon,h}$ is not uniformly bounded in $L^2(Q_T)$ with respect to R and $\varphi_{T,R,\varepsilon,h}$ is not uniformly bounded in $L^2(0,1)$ with respect to ε. Contrarily, we observe that the norm of $\mu_{R,\varepsilon,h}$ is bounded with respect to ε and the norm of $\varphi_{T,R,\varepsilon}$ is bounded with respect to R (Tables 3). This is due to the fact that, by definition of $J^\star_{R,\varepsilon}$, the weight ρ_R mainly acts on the variable $\mu_{R,\varepsilon,h}$ while ε^{-1} mainly acts on $\varphi_{T,R,\varepsilon}$.

In the limit as $\varepsilon \to 0$, the L^2-norm of $\varphi_{T,R,\varepsilon}$, which can be viewed as a multiplier associated to the constraint $y(\cdot\,, T) = 0$, does not belong anymore to $L^2(0,1)$. This is what we observe when we use the primal direct approach described in [11] and solve the formulation (12); as $h \to (0,0)$, we observe arbitrarily large values of the L^2-norm of $p_h(\cdot\,, T)$.

Table 4 depicts the L^2-norm of the computed state at time T. Note that this solution satisfies

	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$	$\varepsilon = 0$
$R = 10^4$	4.57×10^1	2.71×10^1	2.23×10^1	2.04×10^1
$R = 10^6$	3.36×10^2	1.94×10^2	2.95×10^2	3.22×10^2
$R = 10^8$	4.40×10^2	2.81×10^2	3.64×10^2	4.67×10^2
$R = +\infty$	4.41×10^2	2.82×10^2	3.63×10^2	4.60×10^2

	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$	$\varepsilon = 0$
$R = 10^4$	3.96×10^1	4.21×10^2	2.75×10^3	5.53×10^3
$R = 10^6$	6.17×10^0	3.16×10^2	1.99×10^3	3.17×10^3
$R = 10^8$	3.29×10^0	2.70×10^2	1.82×10^3	2.66×10^3
$R = +\infty$	3.27×10^0	2.69×10^2	1.81×10^3	2.61×10^3

Table 3: $L^2(Q_T)$-norm of $\mu_{R,\varepsilon,h}$ (**Top**) and $L^2(0,1)$-norm of $\varphi_{T,R,\varepsilon,h}$ vs. R and ε.

	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$	$\varepsilon = 0$
$R = 10^4$	3.96×10^{-3}	4.34×10^{-4}	5.76×10^{-5}	3.00×10^{-5}
$R = 10^6$	6.19×10^{-4}	3.07×10^{-4}	4.62×10^{-5}	3.09×10^{-5}
$R = 10^8$	3.29×10^{-4}	2.62×10^{-4}	4.28×10^{-5}	3.09×10^{-5}
$R = +\infty$	3.27×10^{-4}	2.62×10^{-4}	4.32×10^{-5}	3.10×10^{-5}

Table 4: $h = (10^{-2}, 10^{-2})$, $\omega = (0.3, 0.6)$, $y_0(x) \equiv \sin(\pi x)$. The $L^2(0,1)$-norm of $y_h(\cdot\,, T)$ vs. R and ε.

$y_h(\cdot\,, 0) = y_{0h}$ and *a priori* differs from the function $-T_R(\rho_h)^{-2}\mu_{R,\varepsilon,h}$. As expected, the weight $T_R(\rho)^2$ reinforces slightly the null controllability constraint (2) as R increases.

These Tables suggest that it is not actually necessary to take $R = +\infty$ and $\varepsilon = 0$ to achieve a good approximation of the controls. Due to the weights, the norms of the computed controls and controlled solutions change only slightly with respect to these parameters. The singular case $R = +\infty$ and $\varepsilon = 0$ ensures a better approximation of the null controllability requirement, but leads to a significative increase of iterates, as the coercivity of $J^\star_{R,\varepsilon}$ is lost.

The computed state and control are displayed in Figures 2 and 3.

5.2 Influence of the weights on the algorithm

We now discuss with more depth the influence of the weights ρ and ρ_0 on the behavior of the conjugate gradient algorithm. We take $R = +\infty$ and $\varepsilon = 0$. At the numerical level, this limit case still makes sense since, for any $h > 0$, the minimizer of $J^\star_{+\infty,0,h}$ obtained via a conjugate gradient method depends on the stopping parameter σ and does not actually satisfy the constraint $y_h(\cdot\,, T) = 0$ exactly. Note also that the numerical approximation we described in Section 4.2 remains consistent in that case, since the finite dimensional space $M_h \times \Phi_{\Delta x}$ is still a conformal approximation of the abstract space where $J^\star_{+\infty,0}$ is coercive, namely the completion of $\mathcal{D}(Q_T) \times \mathcal{D}(0,1)$ for the norm $\|(\mu, \varphi_T)\| := (\iint_{q_T} \rho_0^{-2}|\varphi|^2\, dx\, dt + \iint_{Q_T} \rho^{-2}|\mu|^2\, dxdt)^{1/2}$.

We use the same data as in the previous Section, except that we begin with a larger domain control $\omega = (0.2, 0.8)$. This allows to reach gradients closer to zero, i.e. to prescribe smaller values of σ.

In Tables 5 and 6, we collect some relevant results obtained respectively for $\kappa = 10^{-4}$ and $\kappa = 10^{-5}$. The conjugate gradient method is initialized with $\mu^0 = 0$ and $\varphi^0_T = 0$. The behavior of the method is shown for various $h = (\Delta x, \Delta t)$. In particular, the convergence of the control v_h as well the as the state y_h as $h \to (0,0)$ becomes clear. It is also easy to check that these numerical results are very similar to those obtained with the (primal) direct methods in [11].

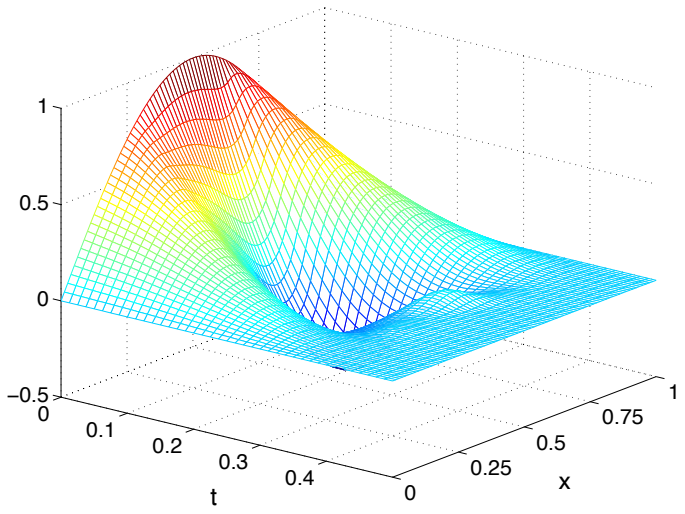

Figure 2: $\omega = (0.3, 0.6)$. The state y_h.

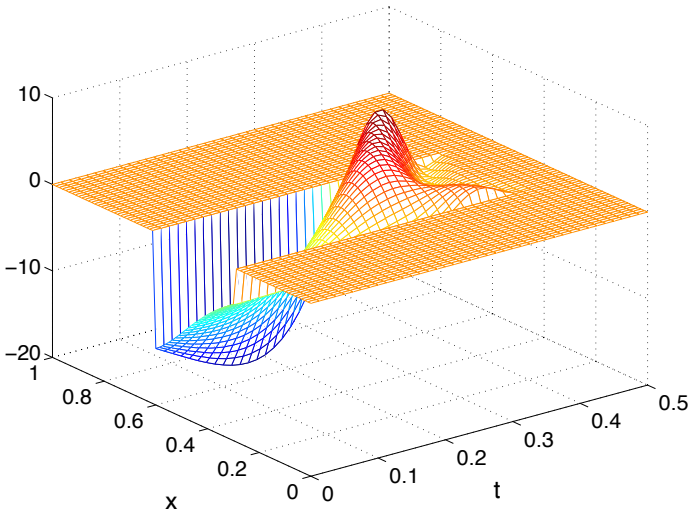

Figure 3: $\omega = (0.3, 0.6)$. The control v_h.

These control functions approximate in a satisfactory way the null controllability requirement: we obtain $\|y_h(\cdot,T)\|_{L^2(0,1)}$ of the order of 10^{-5} and 10^{-6} for $\kappa = 10^{-5}$ and $\kappa = 10^{-6}$, respectively. The number of iterates increases when σ is reduced, but we observe that this number is weakly dependent of the discretization parameter h. For $\kappa = 10^{-5}$ and $h = 1/160$, the evolution in $\log_{10}$ scale of the relative residue $r_h^n \equiv \|g_h^n\|_V/\|g_h^0\|_V$ is displayed in Figure 4. The evolution of the residues is nonlinear with respect to the iterates, as is usual for ill-posed parabolic problems, see for instance [21, 22]. Precisely, the slop reduces significantly after the first iterations and may even vanish for h too close to $(0,0)$.

The first iterates are devoted to compute the lower frequencies of the unknowns μ_h and $\varphi_{T,h}$: according to the regularizing effect of the operator $L^\star$, these low frequencies correspond for the backward solution ψ_h to the points $(x,t) \in \overline{Q_T}$ far enough from $t = T$. As we can see from Figure 4, this computation is achieved after a small number of iterates, almost independent of h. The remaining iterates are devoted to compute the high frequencies of the unknowns μ_h and $\varphi_{T,h}$, unavoidable and harder to capture. For μ_h, this corresponds to a neighborhood of $t = T$, say $(T-\delta, T]$ for some $\delta > 0$. This phenomenon, once again usual for ill-posed parabolic situations, is amplified by the behavior of the weights ρ^{-1} and ρ_0^{-1} near $t = T$. More precisely, since ρ^{-1} and ρ_0^{-1} are exponentially close to zero in $(0,1)\times(T-\delta,T)$, these high frequencies have a very weak effect on the values of $J^\star$. The high frequency components of the unknowns $\varphi_{T,h}$, which really do exist since the minimizer φ_T lives in abstract space much larger than $L^2(0,1)$, are damped out from $t = T$ to $t = T-\delta$ and, therefore, again does not affect the value of $J_h^\star$ significantly. This very low dependence explains the difficulty to capture such frequencies with a gradient method.

$\Delta x, \Delta t$	1/40	1/80	1/160	1/320
♯ CG iterates	559	383	471	504
$\|v_h\|_{L^2(q_T)}$	9.89×10^{-1}	1.006×10^{-1}	1.015×10^{-1}	1.021×10^{-1}
$\|y_h\|_{L^2(Q_T)}$	2.01×10^{-1}	2.004×10^{-1}	1.999×10^{-1}	1.996×10^{-1}
$\|\mu_h\|_{L^2(Q_T)}$	9.207	9.293	13.29	18.99
$\|\varphi_{T,h}\|_{L^2(0,1)}$	3.81×10^{1}	3.83×10^{1}	3.94×10^{1}	3.77×10^{1}
$\|y_h(\cdot,T)\|_{L^2(0,1)}$	2.24×10^{-5}	2.80×10^{-5}	3.01×10^{-5}	3.00×10^{-5}

Table 5: $\omega = (0.2, 0.8)$ - $\kappa = 10^{-4}$.

$\Delta x, \Delta t$	1/80	1/160	1/320
♯ CG iterates	3762	3620	3465
$\|v_h\|_{L^2(q_T)}$	1.016×10^{-1}	1.027×10^{-1}	1.032×10^{-1}
$\|y_h\|_{L^2(Q_T)}$	1.997×10^{-1}	1.992×10^{-1}	1.990×10^{-1}
$\|\mu_h\|_{L^2(Q_T)}$	$4.66\times10^{+1}$	$5.99\times10^{+1}$	$7.66\times10^{+1}$
$\|\varphi_{T,h}\|_{L^2(0,1)}$	1.05×10^{2}	1.74×10^{2}	1.53×10^{2}
$\|y_h(\cdot,T)\|_{L^2(0,1)}$	2.84×10^{-6}	3.14×10^{-6}	3.19×10^{-6}

Table 6: $\omega = (0.2, 0.8)$ - $\kappa = 10^{-5}$.

But the crucial point from the numerical viewpoint is that, since these high frequencies are damped out where the weight vanishes, they are not necessary to achieve a good approximation of the control v_h, the state controlled y_h and the associated cost. Consequently, a reasonable value of κ suffices. For instance, from Table 5 (for which $\kappa = 10^{-4}$) and Table 6 (where $\kappa = 10^{-5}$), we see that, for h close to $(0,0)$, $\|v_h\|_{L^2(q_T)}$ and $\|y_h\|_{L^2(Q_T)}$ are unchanged in practice.

Contrarily, the values of $\|\varphi_{T,h}\|_{L^2(0,1)}$ do change when κ is divided by 10, as it contains more high frequency modes. Table 7 displays relevant numerical values for $\kappa = 10^{-3}, 10^{-4}, 10^{-5}$ and $\kappa = 10^{-6}$ and emphasizes together with Table 6, that even if μ_h and φ_h do not converge in $L^2(Q_T)$, the weighted functions $\rho_h^{-2}\mu_h$ and $\rho_{0,h}^{-2}\varphi_h$ do.

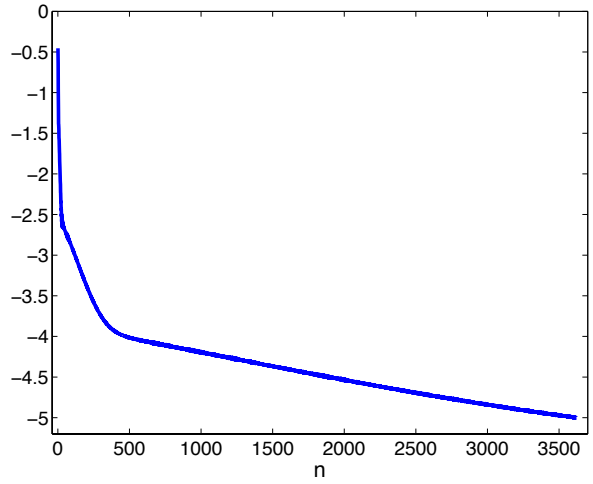

Figure 4: Evolution of $\log_{10}(r_h^n)$ with respect to the iterates with $\omega = (0.2, 0.8)$ and $y_0(x) \equiv \sin(\pi x)$ for $h = (1/160, 1/160)$.

σ	10^{-3}	10^{-4}	10^{-5}	10^{-6}
♯ CG iterates	16	471	3620	25631
$\|y_h(\cdot,T)\|_{L^2(0,1)}$	3.18×10^{-4}	3.01×10^{-5}	3.14×10^{-6}	2.81×10^{-7}
$\|\rho_0^{-2}\psi_h\|_{L^2(q_T)}$	1.0022	1.0159	1.0274	1.0309
$\|\rho^{-2}\mu_h\|_{L^2(Q_T)}$	2.0083×10^{-1}	1.9995×10^{-1}	1.9924×10^{-1}	1.9904×10^{-1}
$\|\psi_h\|_{L^2(Q_T)}$	2.64	5.89	1.85×10^{1}	2.48×10^{1}
$\|\varphi_h(\cdot,T)\|_{L^2(0,1)}$	1.51×10^{1}	3.94×10^{1}	1.74×10^{2}	2.46×10^{2}
$\|\mu_h\|_{L^2(Q_T)}$	7.55	1.32×10^{1}	5.99×10^{1}	1.62×10^{2}

Table 7: $\omega = (0.2, 0.8)$ and $h = (1/160, 1/160)$.

When we compute the null control of minimal L^2-norm , that is, when we solve a problem like (7) with $\rho \equiv 0$ and $\rho_0 \equiv 1$, the conjugate gradient method for the associated dual problem, which is much more sensitive to the numerical approximation, behaves very differently. This was discussed in length in [21]. In this case, the control depends much more strongly on the final adjoint state φ_T. Consequently, smaller values of the tolerance κ are needed, so as to capture high frequencies and this leads to a larger number of iterates. Moreover, the control of minimal L^2-norm, defined simply as $v = \varphi 1_{q_T}$ exhibits a highly oscillatory behavior in the time direction near $t = T$. It results that, for any fixed and small enough σ, the number of CG iterates is no more constant with respect to the discretization parameter, but blows up exponentially as $h \to (0, 0)$. In the present situation, the weight ρ_0^{-2} has the effect to destroy such oscillatory behavior so that $v_h = \rho_{0,h}^{-2} 1_{q_T}$ is smooth near T (see for instance Figure 3).

For $(\rho, \rho_0) = (0, 1)$, $h = (1/160, 1/160)$, Figure 5 depicts the evolution of the residue r_h^n and the corresponding final adjoint state, which minimizes $\mathcal{I}$, see (4). The evolution is similar to the one observed for weighted integrals (Figure 4), but the stopping test for $\kappa = 10^{-5}$ is achieved only after 9 671 iterates, instead of 3 620. The other numerical values (to be compared with those in Table 6, second column) are the following: $\|v_h\|_{L^2(q_T)} \approx 7.45 \times 10^{-1}$, $\|y_h\|_{L^2(Q_T)} \approx 1.52 \times 10^{-1}$, $\|y_h(\cdot\,, T)\|_{L^2(0,1)} \approx 2.61 \times 10^{-6}$ and $\|\varphi_T\|_{L^2(0,1)} \approx 5.05 \times 10^2$.

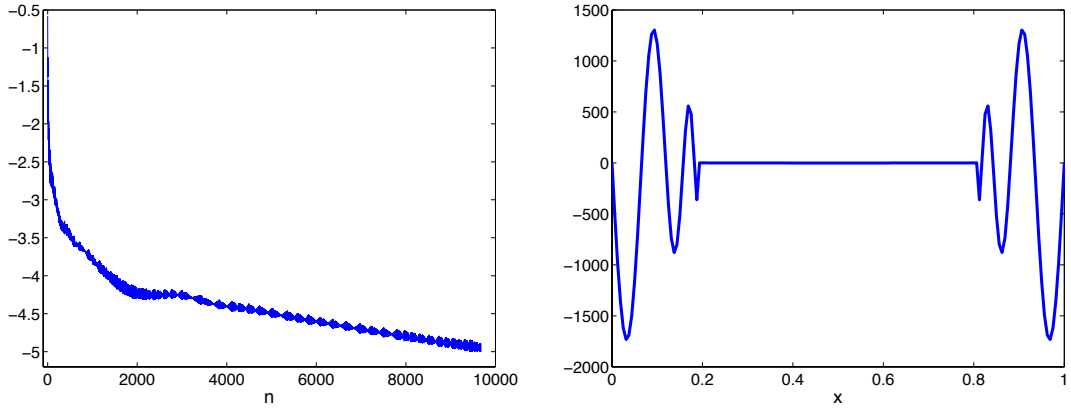

Figure 5: $\omega = (0.2, 0.8)$ - $(\rho, \rho_0) = (0, 1)$ - $\sigma = 10^{-5}$ - $h = (1/160, 1/160)$. Evolution of the residue r_h^n with respect to n (**Left**) and corresponding final adjoint state $\phi_{T,h}$ (**Right**).

The weights have a clear influence on the behavior of the iterative conjugate gradient algorithm. For $\varepsilon = 0$ and $R = \infty$, the minimization of $J^\star$ is numerically ill-posed, since the unique minimizer (μ, φ_T) lives in a singular and very large space, hard to approximate by a finite dimensional approach (for more severe data, many more iterates of the conjugate gradient algorithm are required to get a relative residue r_h^n of similar size).

6 Further comments and concluding remarks

Numerical analysis and error estimates - As mentioned above, by analogy with the methods introduced in [11] for the numerical solution of (7), it is reasonable to suspect that the strong convergence of $v_h = -\pi_h(\rho_0^{-2})\phi_h$ in $L^2(\omega \times (0, T))$ can be established. Observe that this issue is also open for the minimal L^2-norm situation (i.e. $\rho \equiv 0$ and $\rho_0 \equiv 1$).

Some extensions to other linear problems - The methods used in this paper can be extended to cover null controllability problems for linear heat equations in higher spatial dimensions. Since the related

computational work is reasonable, these dual methods can be adapted and extended to this setting.

The intrinsic ill-posedness of this problem is enhanced within the dual approach, at least when the variable φ is considered. There, as soon as the control support is sufficiently small, the conjugate gradient fails to converge. Indeed, for the dual problem to work reasonably, one has to be very careful, in particular in the time integration process.

Appropriate weight functions have to be used. In practice, what we have to be able to construct is a function that is positive in Ω, vanishes on $\partial\Omega$ and possesses nonzero gradient in $\overline{\Omega}\setminus\omega$. Such a function always exists (a result by Imanuvilov) and is relatively easy to construct for instance when Ω is convex.

Also, using finite element tools, we can without much difficulty get results in the case where ω is time-dependent, that is, q_T is replaced by a non-cylindrical set. It is not difficult to prove that null controllability holds as well for any time $T>0$ when the control is exerted on any open set

$$q_T = \{\,(x,t)\in Q_T : g(t)<x<h(t),\, t\in(0,T)\,\},$$

where g and h are smooths functions on $[0,T]$, with $0\le g\le h\le 1$ and $g(t)\not\equiv h(t)$. This opens the possibility to optimize numerically the domain q_T, as was done in a cylindrical situation in [20].

Change of variable - Preconditioning - We observed that the weights ρ and ρ_0 have a smoothing effect on the behavior of the descent algorithm. However, they do not prevent the problem (33) to be numerically ill-posed. From the numerical observations and also from the optimality conditions (27), it becomes clear that the relevant variables are not (μ,φ_T), others involving the weights ρ and ρ_0. More precisely, if we introduce

$$\eta = (T-t)^{-3/2}\rho_0^{-1}\varphi, \quad M=\rho^{-1}\mu, \tag{39}$$

then the functional (33) can be rewritten in the following equivalent form:

$$\overline{J}^\star(M,\eta_T) = \frac{1}{2}\left(\iint_{Q_T}|M|^2\,dx\,dt + \iint_{q_T}(T-t)^{-3}|\eta|^2\,dx\,dt\right) + T^{-3/2}\int_0^1 \rho_0(x,0)\,\eta(x,0)\,y_0(x)\,dx,$$

where η is now the solution to the backwards problem:

$$\rho^{-1}\,L^\star((T-t)^{-3/2}\rho_0\,\eta) = M \quad \text{in} \quad Q_T, \qquad \eta=0 \quad \text{on} \quad \Sigma_T, \qquad \eta(\cdot,T)=\eta_T.$$

Now, $\overline{J}^\star$ is to be minimized over a space that is expected to be much smaller than the one corresponding to $J^\star$. In particular, $M\in L^2(Q_T)$ (at least). If we denote by (M,η_T) the optimal pair for $\overline{J}^\star$, then the optimal control and controlled state are given by

$$v = (T-t)^{-3/2}\rho_0^{-1}\eta|_{q_T}, \quad y=-\rho^{-1}M.$$

We see that

$$\rho^{-1}L^\star((T-t)^{-3/2}\rho_0\eta) = L^\star\eta + A_1\eta_x + A_2\eta$$

where the $A_i=A_i(x,t)$ satisfy (see (10)) :

$$\begin{cases} A_1 = -2a(x)\beta_x(T-t)^{-1}, \\ A_2 = -(T-t)^{-1}\Big((a\beta_x)_x + (T-t)^{-1}(\beta(x)+a(x)\beta_x^2)\Big). \end{cases}$$

Note that no exponential function in time appears anymore. We observed in [11] that, in the context of the primal approach (see Section 2), a very similar change of variables leads to a significant reduction of the condition number of the matrix associated to (12). Accordingly, we may expect here that (39) acts as a pre-conditioner for the conjugate algorithm used to approximate the minimum of $\overline{J}^\star$. This important issue will be analyzed in a future work.

Additional extensions and future work

The methods can also be extended to cover many other controllable systems: non-scalar parabolic systems, Stokes and Stokes-like systems, etc.

It is also possible to extend the previous arguments and methods to the boundary null controllability case and to the exact controllability to trajectories (with distributed or boundary controls).

As first noticed in [13] and using in part the results by [23] and [25], the approach may also work for linear equations of the hyperbolic kind (typically, the classical wave equation). In this case, the null controllability problem is not always solvable; indeed, a geometric control condition involving ω and T must be satisfied. However, when this holds, we do not need unbounded weights in the analog of (7) and (12) in order to get a well-posed problem. As a consequence, contrarily to the situation found in this paper, no huge abstract functional space related to the null controllability constraint appears and everything is simpler. We refer to [6], where the primal method is addressed in the framework of the boundary controllability.

This work also opens the possibility to address the numerical solution of nonlinear control problems, the optimization of the control support ω, etc. In particular, we refer to [10] for the numerical approximation of null controls for a semi-linear heat equation.

References

[1] F. Ben Belgacem and S.M. Kaber, *On the Dirichlet boundary controllability of the 1-D heat equation: semi-analytical calculations and ill-posedness degre*, Inverse Problems 27 (2011), no. 5.

[2] F. Boyer, F. Hubert and J. Le Rousseau, *Discrete Carleman estimates for elliptic operators in arbitrary dimension and applications*, SIAM J. Control Optim. 48 (2010), no. 8, 5357–5397.

[3] F. Boyer, F. Hubert and J. Le Rousseau, *Uniform null-controllability properties for space/time-discretized parabolic equations*, Numerische Mathematik, Vol. 118, no 4, pp. 601-661 (2011).

[4] E. Casas, *Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations*, SIAM J. on Control and Optim. 35(4): 1297–1327, 1997.

[5] C. Carthel, R. Glowinski and J.-L. Lions, *On exact and approximate Boundary Controllabilities for the heat equation: A numerical approach*, J. Optimization, Theory and Applications 82(3), (1994) 429–484.

[6] N. Cîndea, E. Fernández-Cara and A. Münch, *Numerical controllability of the wave equation through a primal method and Carleman estimates*, Preprint (2012), http://hal.archives-ouvertes.fr/hal-00668951.

[7] I. Ekeland, R. Temam, *Convex analysis and variational problems*, Classics in Applied Mathematics 28, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1999.

[8] S. Ervedoza, J. Valein, *On the observability of abstract time-discrete linear parabolic equations*, Rev. Mat. Complut., 23 (2010), no. 1, 163–190.

[9] E. Fernández-Cara and S. Guerrero, *Global Carleman inequalities for parabolic systems and applications to controllability,* SIAM J. Control Optim. 45 (2006), no. 4, 1399–1446.

[10] E. Fernández-Cara and A. Münch, *Numerical null controllability of a semi-linear 1D heat equation via a least squares reformulation.*, C.R. Acad. Sci. Paris, Série. I, 349, (2011) 867–871.

[11] E. Fernández-Cara and A. Münch, *Numerical null controllability of the 1D heat equation: primal algorithms*, Preprint (2009), http://hal.archives-ouvertes.fr/hal-00687884.

[12] A.V. Fursikov, *Optimal control of distributed systems. Theory and applications,* Translations of Mathematical Monographs, 187. American Mathematical Society, Providence, 2000.

[13] A.V. Fursikov and O. Yu. Imanuvilov, *Controllability of Evolution Equations,* Lecture Notes Series, number 34. Seoul National University, Korea, (1996) 1–163.

[14] R. Glowinski and J.L. Lions, *Exact and approximate controllability for distributed parameter systems,* Acta Numerica (1996), 159–333.

[15] R. Glowinski, J.L. Lions and J. He, *Exact and approximate controllability for distributed parameter systems: a numerical approach* Encyclopedia of Mathematics and its Applications, 117. Cambridge University Press, Cambridge, 2008.

[16] S. Kindermann, *Convergence Rates of the Hilbert Uniqueness Method via Tikhonov regularization*, J. of Optimization Theory and Applications 103(3), (1999) 657-673.

[17] S. Labbé and E. Trélat, *Uniform controllability of semi-discrete approximations of parabolic control systems*, Systems and Control Letters 55 (2006) 597-609.

[18] A. López and E. Zuazua, *Some new results to the null controllability of the 1-d heat equation*, Séminaire sur les Equations aux dérivees partielles, 1997-1998, Exp. No. VIII, 22p., Ecole Polytech. Palaiseau (1998).

[19] S. Micu and E. Zuazua, *On the regularity of null-controls of the linear 1-d heat equation,* C.R.Acad. Sci. Paris, Ser. I (2011).

[20] A. Münch and F. Periago, *Optimal distribution of the internal null control for the 1D heat equation*, J. Differential Equations, 250, 95-111 (2011).

[21] A. Münch and E. Zuazua, *Numerical approximation of null controls for the heat equation : ill-posedness and remedies*, Inverse Problems, 26(8) 085018 (2010).

[22] A. Münch and P. Pedregal, *Numerical null controllability of the heat equation through a variational approach,* Preprint (2011).

[23] J.-P. Puel, *Global Carleman inequalities for the wave equations and applications to controllability and inverse problems,* Cours Udine Mod1-06-04-542968, Udine, 2011.

[24] F. Troltzsch, *Optimal control of partial differential equations,* Volume 112 of Graduate Studies in Mathematics, American Mathematical Society, Providence, 2010.

[25] X. Zhang, *Explicit observability inequalities for the wave equation with lower order terms by means of Carleman inequalities,* SIAM J. Control. Optim., 39 (2000) 812-834.

[26] E. Zuazua, *Control and numerical approximation of the wave and heat equations*, ICM2006, Madrid, Spain, Vol. III (2006) 1389–1417.