Strong convergent approximations of null controls for the 1D heat equation

Enrique Fernández-Cara & Arnaud Münch

SeMA Journal
Boletin de la Sociedad Española de Matemática Aplicada

ISSN 1575-9822
Volume 61
Number 1

SeMA (2013) 61:49-78
DOI 10.1007/s40324-013-0001-6

Springer

SeMA (2013) 61:49–78
DOI 10.1007/s40324-013-0001-6

Strong convergent approximations of null controls for the 1D heat equation

Enrique Fernández-Cara · Arnaud Münch

Received: 20 October 2012 / Accepted: 27 October 2012 / Published online: 1 February 2013

Abstract This paper deals with the numerical computation of distributed null controls for the 1D heat equation, with Dirichlet boundary conditions. The goal is to compute approximations to controls that drive the solution from a prescribed initial state at $t = 0$ to zero at $t = T$. Using ideas from Fursikov and Imanuvilov (Controllability of Evolution Equations, Lecture Notes Series, vol. 34. Seoul National University, Korea, pp. 1–163, 1996), we consider the control that minimizes a functional involving weighted integrals of the state and the control, with weights that blow up at $t = T$. The optimality system is equivalent to a differential problem that is fourth-order in space and second-order in time. We first address the numerical solution of the corresponding variational formulation by introducing a space-time finite element that is C^1 in space and C^0 in time. We prove a strong convergence result for the approximate controls and then we present some numerical experiments. We also introduce a mixed variational formulation and we prove well-posedness through a suitable *inf-sup* condition. We introduce a (non-conformal) C^0 finite element approximation and we provide new numerical results. In both cases, thanks to an appropriate change of variable, we observe a polynomial dependance of the condition number with respect to the discretization parameter. Furthermore, with this second method, the initial and final conditions are satisfied exactly.

Keywords One-dimensional heat equation · Null controllability · Finite element methods · Mixed finite elements · Carleman inequalities

Mathematics Subject Classification (2010) 35K35 · 65M12 · 93B40

E. Fernández-Cara was partially supported by grants MTM2006-07932 and MTM2010-15592 (DGES, Spain).

E. Fernández-Cara (✉)
Dpto. EDAN, University of Sevilla, Aptdo. 1160, 41080 Sevilla, Spain
e-mail: cara@us.es

A. Münch
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

1 Introduction: The null controllability problem

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

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

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., $A \in L^\infty((0,1)\times(0,T))$ 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 also introduce the following notation:

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

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

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

The associated null controllability problem (at time T) is the following: for each $y_0 \in L^2(0,1)$, find $v \in L^2(q_T)$ such that the 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. Some relevant references are [8,22,28,29]. For heat equations, we refer to [1,12,17,20,21].

This paper is devoted to design and analyze efficient numerical methods for the previous null controllability problem.

The numerical approximation of null controls for (1) is a difficult issue. As shown below, this is mainly due to the regularization effect of the heat kernel, that renders the numerical problem severely ill-posed.

So far, the approximation of the control of minimal L^2 norm has focused most of the attention. The first contribution was due to Carthel et al. [5], who made use of duality arguments. However, the resulting problems involve 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 $v = \phi\, 1_\omega$, where ϕ solves the backward heat equation

$$\begin{cases} -\phi_t - (a(x)\phi_x)_x + A(x,t)\phi = 0, & (x,t) \in (0,1)\times(0,T) \\ \phi(x,t) = 0, & (x,t) \in \{0,1\}\times(0,T) \\ \phi(x,T) = \phi_T(x), & x \in (0,1) \end{cases} \tag{3}$$

and ϕ_T minimizes the strictly convex and coercive functional

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

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

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

$$\|\phi(\cdot,0)\|^2_{L^2(0,1)} \le C \iint_{q_T} |\phi|^2\,dx\,dt \qquad \forall \phi_T \in L^2(0,1), \tag{5}$$

that holds for some constant $C = C(\omega, T)$; in turn, this is a consequence of appropriate *global Carleman* inequalities, see [12]. But, as discussed in length in [27], 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 [2,18,24], 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

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

and $\epsilon > 0$. Now, the minimizer $\phi_{T,\epsilon}$ 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 [27].

In this paper, we consider the following extremal problem, introduced by Fursikov and Imanuvilov [12]:

$$\begin{cases} \text{Minimize } \; J(y,v) = \dfrac{1}{2}\displaystyle\iint_{Q_T} \rho^2|y|^2\,dx\,dt + \dfrac{1}{2}\iint_{q_T} \rho_0^2|v|^2\,dx\,dt \\ \text{Subject to } \; (y,v) \in \mathcal{C}(y_0,T). \end{cases} \tag{6}$$

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

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

and we assume (at least) that

$$\begin{cases} \rho = \rho(x,t),\ \rho_0 = \rho_0(x,t) \text{ are continuous and } \ge \rho_* > 0 \text{ in } Q_T \text{ and} \\ \rho,\rho_0 \in L^\infty(Q_{T-\delta}) \text{ for all small } \delta > 0 \end{cases} \tag{7}$$

(hence, they can blow up as $t \to T^-$).

This paper is organized as follows.

In Sect. 2, we recall some results from [12] and we present the details of the variational approach to the null controllability problem. The optimal pair (y,v) is written in terms of a new function p, the unique solution to (15). In Sect. 3, we analyze the numerical approximation of the variational formulation (22), that is obtained from (15) after a change of variables, see (18). The main advantage of (22) is that it involves no weight growing exponentially as $t \to T^-$. The approximation makes use of a finite element that is C^1 in space and C^0 in time. We prove a convergence result as the discretization parameters go to zero and then we present some numerical experiments. In order to avoid C^1 in space finite elements, we introduce in Sect. 4 the mixed variational formulation (57), which is equivalent to (22) and we prove well-posedness, see Theorem 4.1. Some numerical experiments, based on a non conformal C^0 finite element, are discussed in Sect. 4.2. Finally, some further comments, additional results and concluding remarks are given in Sect. 5.

2 A variational approach to the null controllability problem

In the sequel, unless otherwise specified, it is assumed that

$$a \in C^1([0,1]), \quad a(x) \geq a_0 > 0 \qquad \forall x \in [0,1]. \tag{8}$$

Under this assumption, for any $A \in L^\infty(Q_T)$, the linear system (1) is null-controllable.

Let ρ and ρ_0 be functions satisfying (7) and let us consider the extremal problem (6). Then we have the following:

Theorem 2.1 *For any $y_0 \in L^2(0,1)$ and $T > 0$, there exists exactly one solution to* (6).

The proof is simple. Indeed, from the null controllability of (1), $\mathcal{C}(y_0, T)$ is non-empty. Furthermore, it is a closed convex set of $L^2(Q_T) \times L^2(q_T)$; in fact, it is a closed linear manifold, whose supporting space is the set of all (z, w) such that $w \in L^2(q_T)$,

$$\begin{cases} z_t - (a(x)z_x)_x + A(x,t)\,z = w 1_\omega, & (x,t) \in (0,1) \times (0,T) \\ z(x,t) = 0, & (x,t) \in \{0,1\} \times (0,T) \\ z(x,0) = 0, & x \in (0,1) \end{cases} \tag{9}$$

and

$$z(x,T) = 0, \quad x \in (0,1).$$

On the other hand, $(y, v) \mapsto J(y, v)$ is strictly convex, proper and lower semi-continuous on the space $L^2(Q_T) \times L^2(q_T)$ and $J(y, v) \to +\infty$ as $\|(y, v)\|_{L^2(Q_T)\times L^2(q_T)} \to +\infty$. Hence, the extremal problem (6) certainly possesses a unique solution.

Since we are looking for controls such that the associated states satisfy (2), it is a good idea to choose weights ρ and ρ_0 blowing up to $+\infty$ as $t \to T^-$; this can be viewed as a reinforcement of the constraint (2).

When (8) holds, there exist "good" weight functions ρ and ρ_0 that blow up at $t = T$ and provide a very suitable solution to the original null controllability problem. They were determined and systematically used by Fursikov and Imanuvilov and are the following:

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

The roles of ρ and ρ_0 are clarified by the following arguments and results, which are mainly due to Fursikov and Imanuvilov. First, let us set

$$P_0 = \{q \in C^\infty(\overline{Q}_T) : q = 0 \ \text{ on } \ \Sigma_T\}.$$

In this linear space, 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 well known *unique continuation property,* we necessarily have $q \equiv 0$.

Let P be the completion of P_0 for this scalar product. Then P is a Hilbert space and the following results hold:

Lemma 2.1 *Assume that* (8) *is satisfied and* ρ *and* ρ_0 *are given by* (10). *Let us set*

$$\rho_1(x,t) = (T-t)^{1/2}\rho(x,t), \quad \rho_2(x,t) = (T-t)^{-1/2}\rho(x,t). \tag{11}$$

Then there exists $C > 0$ *only depending on* $\omega, T, a_0, \|a\|_{C^1}$ *and* $\|A\|_{L^\infty}$*, such that one has the following for all* $q \in P$*:*

$$\iint_{Q_T} \left[\rho_2^{-2}\left(|q_t|^2 + |q_{xx}|^2\right) + \rho_1^{-2}|q_x|^2 + \rho_0^{-2}|q|^2\right] dx\,dt$$
$$\leq C \left(\iint_{Q_T} \rho^{-2}|L^\star q|^2\,dx\,dt + \iint_{q_T} \rho_0^{-2}|q|^2\,dx\,dt \right). \tag{12}$$

The proof is given in [12].

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* $\omega, T, a_0, \|a\|_{C^1}$ *and* $\|A\|_{L^\infty}$*, such that*

$$\|q(\cdot\,,0)\|^2_{H_0^1(0,1)} \leq C \left(\iint_{Q_T} \rho^{-2}|L^\star q|^2\,dx\,dt + \iint_{q_T} \rho_0^{-2}|q|^2\,dx\,dt \right) \tag{13}$$

for all $q \in P$.

Proof Let $\delta > 0$ be given and let us consider the Banach space $C^0([0, T-\delta]; L^2(0,1))$. Let q be given in P. Then, in view of Lemma 2.1 and the fact that all the weights ρ_i are bounded from above in $Q_{T-\delta}$, we see that

$$q, q_t, q_x, q_{xx} \in L^2(Q_{T-\delta}),$$

with norms in this space bounded by a constant times $\|q\|_P$.

In particular, $t \mapsto q(\cdot\,,t)$ and $t \mapsto q_t(\cdot\,,t)$, respectively regarded as a $H^2(0,1)$-valued and a $L^2(0,T)$-valued function, are square-integrable. This implies that $t \mapsto q(\cdot\,,t)$, regarded as a $H_0^1(0,1)$-valued function, is continuous on $[0,T)$. □

Proposition 2.1 *Assume that* (8) *is satisfied and let* ρ *and* ρ_0 *be given by* (10). *Let* (y, v) *be the solution to* (6). *Then there exists* $p \in P$ *such that*

$$y = \rho^{-2}L^\star p, \quad v = -\rho_0^{-2}p|_{q_T}. \tag{14}$$

The function p *is the unique solution to*

$$\begin{cases} \displaystyle\iint_{Q_T} \rho^{-2}L^\star p\,L^\star q\,dx\,dt + \iint_{q_T} \rho_0^{-2}p\,q\,dx\,dt = \int_0^1 y_0(x)\,q(x,0)\,dx \\ \forall q \in P; \quad p \in P. \end{cases} \tag{15}$$

Proof In view of Lemma 2.2 and the *Lax–Milgram Lemma,* there exists exactly one solution p to (15). Let us introduce y and v according to (14). We will check that (y, v) solves (6); this will prove the result.

First, notice that $y \in L^2(Q_T)$ and $v \in L^2(q_T)$. Also, in view of (15), we have

$$\begin{cases} \displaystyle\iint_{Q_T} y\, L^\star q\, dx\, dt = \iint_{q_T} v\, q\, dx\, dt + \int_0^1 y_0(x)\, q(x,0)\, dx \\ \forall q \in P; \quad y \in L^2(Q_T). \end{cases} \tag{16}$$

But this means that y is the solution to (1) in the transposition sense. Since $y_0 \in L^2(0, 1)$ and $v \in L^2(q_T)$, y must coincide with the unique weak solution to (1). In particular, $y \in C^0([0, T]; L^2(0, 1))$ and, taking into account (14), we find that (2) holds. In other words, $(y, v) \in \mathcal{C}(y_0, T)$.

Finally, let $(z, w) \in \mathcal{C}(y_0, T)$ be such that $J(z, w) < +\infty$. Then, it is immediate that

$$\begin{aligned} J(z, w) &\geq J(y, v) + \iint_{Q_T} \rho^2 y\, (z - y)\, dx\, dt + \iint_{q_T} \rho_0^2 v\, (w - v)\, dx\, dt \\ &= J(y, v) - \iint_{Q_T} L^\star p\, (z - y)\, dx\, dt - \iint_{q_T} p\, (w - v)\, dx\, dt \\ &= J(y, v). \end{aligned}$$

Hence, (y, v) solves (6).

This ends the proof. □

Remark 1 In this proposition, the regularity assumption on the diffusion coefficient a can be relaxed. Indeed, when a is piecewise C^1 and satisfies the ellipticity hypothesis $a \geq a_0 > 0$, it is also possible to construct weights ρ and ρ_0 such that the previous results hold; see [1].

Remark 2 In view of (14) and (15), it is clear that the function p furnished by proposition 2.1 solves, at least in the distributional sense, the following differential problem, that is second-order in time and fourth-order in space:

$$\begin{cases} L(\rho^{-2} L^\star p) + \rho_0^{-2} p\, 1_\omega = 0, & (x, t) \in (0, 1) \times (0, T) \\ p(x, t) = 0, \ \ (-\rho^{-2} L^\star p)(x, t) = 0 & (x, t) \in \{0, 1\} \times (0, T) \\ (\rho^{-2} L^\star p)(x, 0) = y_0(x), \ \ (\rho^{-2} L^\star p)(x, T) = 0, \ x \in (0, 1). & \end{cases} \tag{17}$$

Notice that, here, no information is obtained on $p(\cdot\, , T)$.

3 A first method: solving a variational equality

In the sequel, we take ρ and ρ_0 as in (10). In account of Proposition 2.1, a strategy to find the solution (y, v) to (6) is to first solve (15) and then use (14).

We know that the solution p to (15) belongs to $C^0([0, T]; H_0^1(0, 1))$. However, for any $s \geq 0$, there is no reason to have $p(\cdot\, , t)$ bounded in $H^{-s}(0, 1)$ as $t \to T^-$. This means that it can be difficult to approximate with robustness the variational equality (15). Hence, it will appear very efficient to perform a change of variable, so as to, somehow, we "normalize" the space P.

3.1 An equivalent variational reformulation

The idea is to rewrite the variational equality (15) in terms of a new variable z, given by

$$z(x,t) = (T-t)^{-\alpha}\rho_0^{-1}(x,t)p(x,t), \quad (x,t) \in Q_T \tag{18}$$

for some appropriate $\alpha \geq 0$. We define Z as the completion of P_0 for the scalar product

$$(z,\overline{z})_Z := \iint_{Q_T} \rho^{-2} L^\star((T-t)^\alpha \rho_0 z) L^\star((T-t)^\alpha \rho_0 \overline{z})\,dx\,dt + \iint_{q_T} (T-t)^{2\alpha} z\overline{z}\,dx\,dt \tag{19}$$

or, equivalently, we set

$$Z = \{(T-t)^{-\alpha}\rho_0\, p : p \in P\}.$$

We see that

$$\rho^{-1}L^\star((T-t)^\alpha\rho_0 z) = A_1 z + A_2 z_t + A_3 z_x + A_4 z_{xx}, \tag{20}$$

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

$$\begin{cases} A_1 = \left((\alpha + \frac{3}{2}) - (a\beta_x)_x\right)(T-t)^{\alpha+1/2} - (a\beta_x^2 + \beta)(T-t)^{\alpha-1/2} + A(T-t)^{\alpha+3/2} \\ A_2 = -(T-t)^{\alpha+3/2}, \quad A_3 = -2a\,\beta_x(T-t)^{\alpha+1/2} - a_x\,(T-t)^{\alpha+3/2}, \\ A_4 = -a\,(T-t)^{\alpha+3/2} \end{cases} \tag{21}$$

(notice that β is given in (10)). Consequently, the variational equality (15) can be rewritten as follows:

$$\begin{cases} \iint_{Q_T} (A_1 z + A_2 z_t + A_3 z_x + A_4 z_{xx})(A_1\overline{z} + A_2\overline{z}_t + A_3\overline{z}_x + A_4\overline{z}_{xx})\,dx\,dt \\ \quad + \iint_{q_T} (T-t)^{2\alpha} z\overline{z}\,dx\,dt = T^\alpha \int_0^1 y_0(x)\rho_0(x,0)\overline{z}(x,0)\,dx \\ \forall \overline{z} \in Z;\ z \in Z. \end{cases} \tag{22}$$

The well-posedness of this formulation is an obvious consequence of the well-posedness of (15).

Proposition 3.1 *The variational equality (22) possesses exactly one solution $z \in Z$. Moreover, the unique solution (y, v) to (6) is given by*

$$y = \rho^{-1}(A_1 z + A_2 z_t + A_3 z_x + A_4 z_{xx}), \quad v = -(T-t)^\alpha \rho_0^{-1} z 1_\omega, \tag{23}$$

where $z \in Z$ solves (22).

In order to have all the coeffcicients A_i in $L^\infty(Q_T)$, it suffices to take $\alpha \geq 1/2$. Indeed, notice that, thanks to the previous change of variable, the functions ρ and ρ_0 in $\rho^{-1}L^\star p = \rho^{-1}L^\star((T-t)^\alpha\rho_0 z)$ compensate each other, so that no exponential function appears anymore in (22).

Unless otherwise specified, the condition $\alpha \geq 1/2$ will be assumed in the sequel.

3.2 Numerical analysis of the variational equality

Let us introduce the bilinear form $m(\cdot\,,\cdot)$, with

$$\begin{aligned} m(z,\overline{z}) := & \iint_{Q_T} (A_1 z + A_2 z_t + A_3 z_x + A_4 z_{xx})(A_1\overline{z} + A_2\overline{z}_t + A_3\overline{z}_x + A_4\overline{z}_{xx})\,dx\,dt \\ & + \iint_{q_T} (T-t)^{2\alpha} z\overline{z}\,dx\,dt \end{aligned} \tag{24}$$

and the linear form ℓ, with

$$\langle \ell, \overline{z}\rangle := T^{\alpha} \int_0^1 y_0(x)\rho_0(x,0)\overline{z}(x,0)\,dx\,dt. \tag{25}$$

Then (22) reads as follows:

$$m(z,\overline{z}) = \langle \ell, \overline{z}\rangle \quad \forall \overline{z} \in Z; \quad z \in Z. \tag{26}$$

3.2.1 Finite dimensional approximation

For any finite dimensional space $Z_h \subset Z$, we can introduce the following problem:

$$m(z_h,\overline{z}_h) = \langle \ell, \overline{z}_h\rangle \quad \forall \overline{z}_h \in Z_h; \quad z_h \in Z_h. \tag{27}$$

Obviously, (27) is well-posed. Furthermore, we have the following classical result:

Lemma 3.1 *Let $z \in Z$ be the unique solution to (26) and let $z_h \in Z_h$ be the unique solution to (27). We have*

$$\|z - z_h\|_Z \le \inf_{\overline{z}_h \in Z_h} \|z - \overline{z}_h\|_Z. \tag{28}$$

Proof We write that

$$\|z_h - z\|_Z^2 = m(z_h - z, z_h - z) = m(z_h - z, z_h - \overline{z}_h) + m(z_h - z, \overline{z}_h - z).$$

The first term term vanishes for all $\overline{z}_h \in Z_h$. The second one is bounded by $\|z_h - z\|_Z \|\overline{z}_h - z\|_Z$. So, we get

$$\|z - z_h\|_Z \le \|z - \overline{z}_h\|_Z \quad \forall \overline{z}_h \in Z_h$$

and the result follows. □

As usual, this result can be used to prove the convergence of z_h towards z as $h \to 0$ when the spaces Z_h are chosen appropriately.

More precisely, assume that $\mathcal{H} \subset \mathbf{R}^d$ is a *net* (i.e. a generalized sequence) that converges to zero and Z_h is as above for each $h \in \mathcal{H}$. Let us introduce the interpolation operators $\Pi_h : P_0 \to Z_h$ and let us assume that the finite dimensional spaces Z_h are chosen such that

$$\|\Pi_h z - z\|_Z \to 0 \ \text{ as } \ h \in \mathcal{H}, \ h \to 0, \quad \forall z \in P_0. \tag{29}$$

We then have:

Proposition 3.2 *Let $z \in Z$ be the solution to (26) and let $z_h \in Z_h$ be the solution to (27) for each $h \in \mathcal{H}$. Then*

$$\|z - z_h\|_Z \to 0 \ \ as \ \ h \in \mathcal{H}, \ \ h \to 0. \tag{30}$$

Proof Let us choose $\epsilon > 0$. From the density of P_0 in Z, there exists $z_\epsilon \in P_0$ such that $\|z - z_\epsilon\|_Z \le \epsilon$. Therefore, from Lemma 3.1, we find that

$$\begin{aligned}\|z - z_h\|_Z &\le \|z - \Pi_h z_\epsilon\|_Z \\ &\le \|z - z_\epsilon\|_Z + \|z_\epsilon - \Pi_h z_\epsilon\|_Z \\ &\le \epsilon + \|z_\epsilon - \Pi_h z_\epsilon\|_Z.\end{aligned} \tag{31}$$

From (29), $\|z_\epsilon - \Pi_h z_\epsilon\|_Z$ goes to zero as $h \in \mathcal{H}$, $h \to 0$ and the result follows. □

3.2.2 The finite dimensional spaces Z_h

We now indicate which are the good spaces Z_h.

The spaces Z_h have to be chosen such that $\rho^{-1}L^\star((T-t)^\alpha \rho_0 z_h)$ belongs to $L^2(Q_T)$ for any $z_h \in Z_h$. This means that z_h must possess first order time derivatives and first and second order spatial derivatives in $L^2_{\mathrm{loc}}(Q_T)$. Thus, an approximation based on a standard triangulation of Q_T requires spaces of functions that must be, at least, C^0 in t and C^1 in x.

For large integers N_x and N_t, we set $\Delta x = 1/N_x$, $\Delta t = T/N_t$ and $h = (\Delta x, \Delta t)$. We introduce the associated uniform quadrangulations $\mathcal{Q}_h$, with $Q_T = \bigcup_{K \in \mathcal{Q}_h} K$ and we assume that $\{\mathcal{Q}_h\}_h$ is a regular family. Then, we introduce the spaces Z_h as follows:

$$Z_h = \left\{ z_h \in C^{1,0}_{x,t}(\overline{Q_T}) : z_h|_K \in \mathbb{P}(K) \ \ \forall K \in \mathcal{Q}_h, \ z_h = 0 \ \text{ on } \Sigma_T \right\}. \tag{32}$$

Here, $C^{1,0}_{x,t}(\overline{Q_T})$ is the space of the functions in $C^0(\overline{Q_T})$ that possess continuous partial derivatives with respect to x in $\overline{Q_T}$ and

$$\mathbb{P}(K) = (\mathbb{P}_{3,x} \otimes \mathbb{P}_{1,t})(K), \tag{33}$$

where $\mathbb{P}_{\ell,\xi}$ is the space of polynomial functions of order ℓ in the variable ξ.

Obviously, the Z_h are finite dimensional subspaces of Z.

According to the specific geometry of Q_T, we shall analyze the situation for a uniform quadrangulation $\mathcal{Q}_h$. Each element $K \in \mathcal{Q}_h$ is of the form

$$K_{kl} = (x_k, x_{k+1}) \times (t_l, t_{l+1}),$$

with

$$x_{k+1} = x_k + \Delta x, \ t_{l+1} = t_l + \Delta t, \quad \text{for } \ k = 1, \ldots, N_x, \ l = 1, \ldots, N_t.$$

It is then easy to see that a function $z_h \in \mathbb{P}(K_{kl})$ is uniquely determined by the real numbers $\{z_h(x_{k+m}, t_{l+n})\}$ and $\{(z_h)_x(x_{k+m}, t_{l+n})\}$, with $m, n = 0, 1$.

More precisely, let us introduce the functions

$$\begin{aligned} L_{0k}(x) &= \frac{(\Delta x + 2x - 2x_k)(\Delta x - x + x_k)^2}{(\Delta x)^3}, \quad L_{1k}(x) = \frac{(x - x_k)^2(-2x + 2x_k + 3\Delta x)}{(\Delta x)^3}, \\ L_{2k}(x) &= \frac{(x - x_k)(\Delta x - x + x_k)^2}{(\Delta x)^2}, \quad L_{3k}(x) = \frac{-(x - x_k)^2(\Delta x - x + x_k)}{(\Delta x)^2} \end{aligned} \tag{34}$$

and

$$\mathcal{L}_{0l}(t) = \frac{t_l - t + \Delta t}{\Delta t}, \quad \mathcal{L}_{1l}(t) = \frac{t - t_l}{\Delta t}. \tag{35}$$

Then, the following result is not difficult to prove:

Lemma 3.2 *Let us assume that $u \in P_0$ and let us define the function $\Pi_h u$ as follows: on each $K_{kl} = (x_k, x_k + \Delta x) \times (t_l, t_l + \Delta t)$, we set*

$$\Pi_h u(x,t) := \sum_{i,j=0}^{1} L_{ik}(x)\mathcal{L}_{jl}(t)u(x_{i+k}, t_{j+l}) + \sum_{i,j=0}^{1} L_{i+2,k}(x)\mathcal{L}_{jl}(t)u_x(x_{i+k}, t_{j+l}). \tag{36}$$

Then $\Pi_h u$ is the unique function in Z_h that satisfies

$$\Pi_h u(x_k, t_l) = u(x_k, t_l), \quad (\Pi_h u(x_k, t_l))_x = u_x(x_k, t_l), \quad \forall k, l. \tag{37}$$

The linear mapping $\Pi_h : P_0 \mapsto Z_h$ is by definition the interpolation operator associated to Z_h.

In the next section, we will use the following result:

Lemma 3.3 *For any $u \in P_0$ and any $(x, t) \in K_{kl}$, one has:*

$$u - \Pi_h u = \sum_{i,j=0}^{1} m_{ij} u_x(x_{i+k}, t_{j+l}) + \sum_{i,j=0}^{1} L_{ik}\mathcal{L}_{jl}\mathcal{R}[u; x_{i+k}, t_{j+l}], \tag{38}$$

where

$$m_{ij}(x,t) := (L_{ik}(x)(x - x_i) - L_{i+2,k}(x))\mathcal{L}_j(t)$$

and

$$\left\{ \begin{aligned} \mathcal{R}[u; x_{i+k}, t_{j+l}](x,t) := &\int_{t_{j+l}}^{t} u_t(x_{i+k}, s)\,ds + (x - x_{i+k}) \int_{t_{j+l}}^{t} (t-s) u_{xt}(x_{i+k}, s)\,ds \\ &+ \int_{x_{i+k}}^{x} (x-s) u_{xx}(s,t)\,ds. \end{aligned} \right.$$

Proof The equality (38) is a straightforward consequence of the following Taylor expansion with integral remainder:

$$\begin{aligned} u(x,t) = u(x_i, t_j) + (x - x_i)u_x(x_i, t_j) + \int_{t_j}^{t} u_t(x_i, s)\,ds \\ + (x - x_i)\int_{t_j}^{t} (t-s)u_{xt}(x_i, s)\,ds + \int_{x_i}^{x} (x-s)u_{xx}(s,t)\,ds \end{aligned} \tag{39}$$

and the fact that $\sum_{i,j=0}^{1} L_{ik}(x)\mathcal{L}_{jl}(t) = 1$. □

3.2.3 An estimate of $\|z-\Pi_h z\|_Z$ and some consequences

We now prove that (29) holds.

Thus, let us fix $z \in P_0$ and let us first see that

$$\iint_{q_T} (T-t)^{2\alpha}|z-\Pi_h z|^2\,dx\,dt \to 0 \ \text{ as } \ \Delta x, \Delta t \to 0^+. \tag{40}$$

For each $K_{kl} \in \mathcal{Q}_h$ (denoted by K in the sequel), we write:

$$\iint_K ((T-t)^\alpha)^2|z-\Pi_h z|^2\,dx\,dt \le T^{2\alpha}\iint_K |z-\Pi_h z|^2\,dx\,dt. \tag{41}$$

Using Lemma 3.3, we have:

$$\begin{aligned}\iint_K |z-\Pi_h z|^2\,dx\,dt &= \iint_K \Big(\sum_{i,j} m_{ij} z_x(x_i,t_j) + \sum_{i,j} L_i\mathcal{L}_j\mathcal{R}[z;x_{i+k},t_{j+l}]\Big)^2\,dx\,dt\\ &\le 8\|z_x\|^2_{L^\infty(K)}\sum_{i,j}\iint_K |m_{ij}|^2\,dx\,dt + 8\sum_{i,j}\iint_K |L_i\mathcal{L}_j\mathcal{R}[z;x_{i+k},t_{j+l}]|^2\,dx\,dt,\end{aligned} \tag{42}$$

where we have omitted the indices k and l.

Moreover,

$$\begin{aligned}|\mathcal{R}[z;x_{i+k},t_{j+l}]|^2 \le\ & 3\|z_t(x_i,\cdot)\|^2_{L^2(t_1,t_2)}|t-t_j| + (x-x_i)^2|t-t_j|^3\|z_{xt}(x_i,\cdot)\|^2_{L^2(t_1,t_2)}\\ &+|x-x_i|^3\|z_{xx}(\cdot,t)\|^2_{L^2(x_1,x_2)}.\end{aligned} \tag{43}$$

Consequently, we get:

$$\begin{aligned}&\sum_{i,j}\iint_K |L_i\mathcal{L}_j\mathcal{R}[z;x_{i+k},t_{j+l}]|^2\,dx\,dt\\ &\le 3\sup_{x\in(x_1,x_2)}\|z_t(x,\cdot)\|^2_{L^2(t_1,t_2)}\sum_{i,j}\iint_K |L_i(x)\mathcal{L}_j(t)|^2|t-t_j|\,dx\,dt\\ &+\sup_{x\in(x_1,x_2)}\|z_{tx}(x,\cdot)\|^2_{L^2(t_1,t_2)}\sum_{i,j}\iint_K |L_i(x)\mathcal{L}_j(t)|^2|t-t_j|^3(x-x_i)^2\,dx\,dt\\ &+\sup_{t\in(t_1,t_2)}\|z_{xx}(\cdot,t)\|^2_{L^2(x_1,x_2)}\sum_{i,j}\iint_K |L_i(x)\mathcal{L}_j(t)|^2|x-x_i|^3\,dx\,dt.\end{aligned} \tag{44}$$

After some tedious computations, one finds that

$$\sum_{i,j} \iint_K |m_{ij}|^2\,dx\,dt$$
$$= \frac{8}{945}(\Delta x)^3 \Delta t \sum_{i,j} \iint_K |L_i(x)\mathcal{L}_j(t)|^2 |t - t_j|\,dx\,dt = \frac{13}{105}\Delta x(\Delta t)^2, \tag{45}$$

$$\sum_{i,j} \iint_K |L_i(x)\mathcal{L}_j(t)|^2 |t - t_j|^3 (x - x_i)^2\,dx\,dt = \frac{19}{9450}(\Delta x)^3(\Delta t)^4, \tag{46}$$

$$\sum_{i,j} \iint_K |L_i(x)\mathcal{L}_j(t)|^2 |x - x_i|^3\,dx\,dt = \frac{11}{630}(\Delta x)^4 \Delta t. \tag{47}$$

This leads to the estimate

$$\begin{aligned}\iint_K |z - \Pi_h z|^2\,dx\,dt \le{}& \frac{64}{945}(\Delta x)^3 \Delta t \|z_x\|^2_{L^\infty(K)} \\ &+ \frac{312}{105}\Delta x(\Delta t)^2 \sup_{x \in (x_1,x_2)} \|(r^{-1}p)_t(x,\cdot)\|^2_{L^2(t_1,t_2)} \\ &+ \frac{152}{9450}(\Delta x)^3(\Delta t)^4 \sup_{x \in (x_1,x_2)} \|z_{tx}(x,\cdot)\|^2_{L^2(t_1,t_2)} \\ &+ \frac{88}{630}(\Delta x)^4 \Delta t \sup_{t \in (t_1,t_2)} \|z_{xx}(\cdot\,,t)\|^2_{L^2(x_1,x_2)}.\end{aligned} \tag{48}$$

We deduce that

$$\begin{aligned}\iint_{q_T} |z - \Pi_h z|^2\,dx\,dt \le{}& K_1|q_T|\|z_x\|^2_{L^\infty(q_T)}(\Delta x)^2 + K_2|\omega|\|z_t\|^2_{L^2(0,T;L^\infty(\omega))}(\Delta t)^2 \\ &+ K_3|\omega|\|z_{tx}\|^2_{L^2(0,T;L^\infty(\omega))}(\Delta x)^2(\Delta t)^4 \\ &+ K_4 T\|z_{xx}\|^2_{L^\infty(0,T;L^2(\omega))}(\Delta x)^4\end{aligned} \tag{49}$$

for some finite positive constants K_i. Hence, for any $z \in P_0$ one has

$$\iint_{q_T} (T - t)^{2\alpha}|z - \Pi_h(z)|^2\,dx\,dt \to 0 \quad \text{as } \Delta x, \Delta t \to 0. \tag{50}$$

On the other hand, recalling that $\alpha \ge 1/2$ (and consequently the coefficients A_i are bounded), after similar computations, we find:

$$\begin{aligned}&\iint_K |A_1(z - \Pi_h z) + A_2(z - \Pi_h z)_t + A_3(z - \Pi_h z)_x + A_4(z - \Pi_h z)_{xx}|^2\,dx\,dt \\ &\le 4\|A_1\|^2_{L^\infty(Q_T)} \iint_K |z - \Pi_h z|^2\,dx\,dt \\ &\quad + 4\|A_2\|^2_{L^\infty(Q_T)} \iint_K |(z - \Pi_h z)_t|^2\,dx\,dt\end{aligned}$$

$$+ 4\|A_3\|^2_{L^\infty(Q_T)} \iint_K |(z - \Pi_h z)_x|^2\, dx\, dt$$
$$+ 4\|A_4\|^2_{L^\infty(Q_T)} \iint_K |(z - \Pi_h z)_{xx}|^2\, dx\, dt \tag{51}$$

and, proceeding as above, we see that all these quantities go to 0 as $h = (\Delta x, \Delta t) \to (0, 0)$ (it suffices to differentiate (38) with respect to t and x).

This proves that 29 holds.

Consequently, we have the following result:

Proposition 3.3 *Let $z_h \in Z_h$ be the unique solution to (27) and let y_h, v_h be the functions defined by*

$$y_h = \rho^{-1}(A_1 z_h + A_2 z_{h,t} + A_3 z_{h,x} + A_4 z_{h,xx}), \quad v_h = -(T-t)^\alpha \rho_0^{-1} z_h\, 1_\omega.$$

Then

$$\|v - v_h\|_{L^2(q_T)} \to 0 \ \text{ and } \ \|y - y_h\|_{L^2(Q_T)} \to 0 \quad \text{as} \quad h \to 0.$$

In practice, in order to obain an approximation of the solution to (22), we must introduce numerical approximations of the weights and other data. Accordingly, we will consider in the next section a third problem:

$$m_h(\hat{z}_h, \overline{z}_h) = \langle \ell_h, \overline{z}_h \rangle \quad \forall \overline{z}_h \in Z_h; \quad \hat{z}_h \in Z_h, \tag{52}$$

where

$$m_h(z_h, \overline{z}_h) := \iint_{Q_T} ((\pi_h A_1) z_h + (\pi_h A_2) z_{h,t} + (\pi_h A_3) z_{h,x} + (\pi_h A_4) z_{h,xx})$$
$$\times ((\pi_h A_1)\overline{z}_h + (\pi_h A_2)\overline{z}_{h,t} + (\pi_h A_3)\overline{z}_{h,x} + (\pi_h A_4)\overline{z}_{h,xx})\, dx\, dt$$
$$+ \iint_{q_T} \pi_{\Delta t}((T-t)^{2\alpha}) z_h \overline{z}_h\, dx\, dt \tag{53}$$

and

$$\langle \ell_h, \overline{z}_h \rangle := T^\alpha \int_0^1 \pi_{\Delta x}(y_0 \rho_0(\cdot\,, 0)) \overline{z}_h(x, 0)\, dx. \tag{54}$$

Here, for any $f \in C^0(\overline{Q_T})$, $\pi_h f$ denotes the piecewise linear function that coincides with f at all vertices of $\mathcal{Q}_h$. Similar (self-explanatory) meanings can be assigned to $\pi_{\Delta x} f$ and $\pi_{\Delta t} f$.

The strong convergence of $\hat{z}_h$ towards z and some a priori estimates, explicit in h and $\|z\|_Z$, will be the goal of a future work.

3.3 Numerical experiments (I)

We present now some numerical experiments concerning the solution of (52), that can in fact be viewed as a linear system involving a sparse, definite positive and symmetric matrix of order $2N_x N_t$. We denote by $\mathcal{M}_h$ this matrix, so that $(z_h, \overline{z}_h)_{Z_h} = (\mathcal{M}_h\{z_h\}, \{\overline{z}_h\})$.

Once the variable z_h is known, the control v_h is given by $v_h = -\pi_h((T-t)^\alpha \rho_0^{-1}) z_h\, 1_\omega$. The corresponding controlled state may be first obtained from (23). Then, this approximation y_h satisfies the controllability requirement (2) (that is, $y_h(\cdot\,, T) = 0$), but not exactly the initial

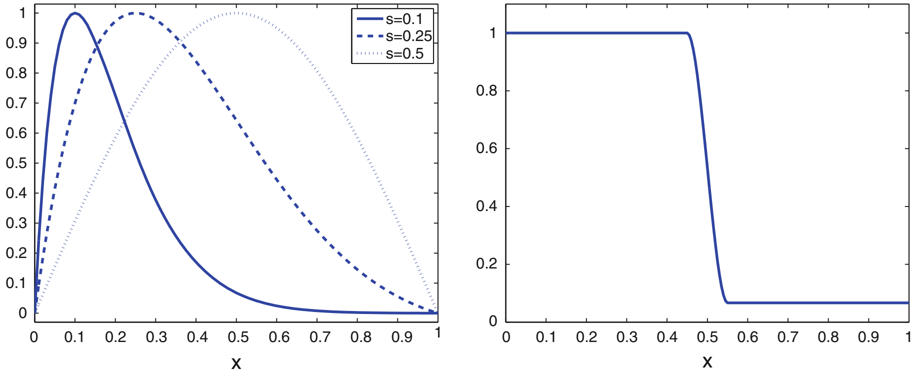

Fig. 1 *Left*: the function $\beta_{0,s}$ for $s = 0.1, s = 0.5$ and $s = 0.25$. *Right*: the non-constant C^1 diffusion coefficient a used in Sect. 4.2 (first case)

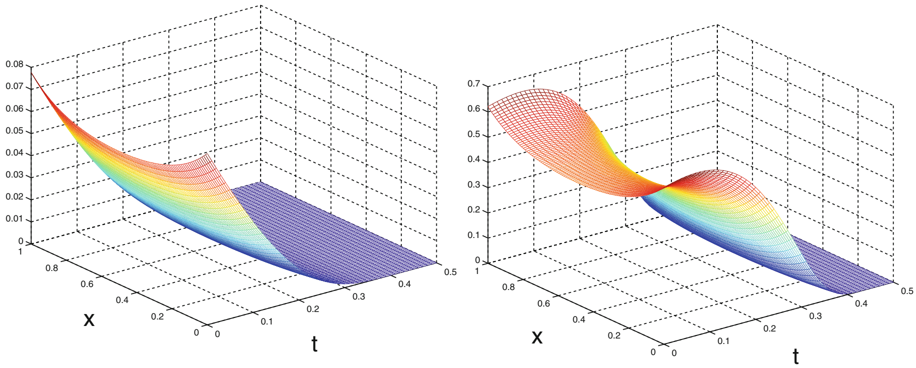

Fig. 2 The weights ρ^{-2} and ρ_0^{-2} defined by (10), with $\beta_0 = \beta_{0,1/2}$ defined by (55), $K_1 = 0.1$ and $K_2 = 2\|\beta_0\|_{L^\infty(0,1)}$

condition. Instead, in order to check the action of the control function v_h, the approximation y_h may be obtained by solving (1) using a finite element method in space and time in a standard way.

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

$$\beta_{0,s}(x) = \frac{x(1-x)e^{-(x-c_s)^2}}{s(1-s)e^{-(s-c_s)^2}}, \quad c_s = s - \frac{1-2s}{2s(1-s)}. \tag{55}$$

If s belongs to ω, we easily check that $\beta_{0,s}$ satisfies the conditions in (10). In the numerical experiments, we will take ρ and ρ_0 as in (10) with $\beta_0 = \beta_{0,s}$, s being the middle point of ω, $K_1 = 0.1$ and $K_2 = 2\|\beta_0\|_{L^\infty(0,1)} = 2$.

The function $\beta_{0,s}$ is plotted in Fig. 1-Left for $s = 0.1, 0.5$ and $s = 0.25$. The weights ρ^{-2} and ρ_0^{-2} corresponding to $s = 0.5$ are displayed in Fig. 2.

We use an exact integration method to compute the components of $\mathcal{M}_h$ and the (direct) Cholesky method with reordering to solve the linear system.

Le us consider a constant diffusion function $a \equiv a_0 = 0.1$ in $(0, 1)$. The initial state y_0 is the first eigenfunction of the Dirichlet–Laplacian, that is $y_0(x) \equiv \sin(\pi x)$ and $T = 1/2$. We take $A \equiv 1$ and $\alpha = 0.5$, so that all the coefficients appearing in the formulation belong

Table 1 Solution of (52), direct method

$\Delta x, \Delta t$	1/20	1/40	1/80	1/160	1/320
$\kappa(\mathcal{M}_h)$	1.28×10^9	1.00×10^{11}	7.04×10^{12}	4.71×10^{14}	3.07×10^{16}
$\|z_h\|_{L^2(Q_T)}$	1.804	2.083	2.309	2.462	2.559
$\|z_h(\cdot, T)\|_{L^2(0,1)}$	1.18×10^{-1}	4.08×10^{-2}	6.46×10^{-3}	4.98×10^{-3}	1.41×10^{-5}
$\|v_h\|_{L^2(q_T)}$	0.97	1.002	1.023	1.035	1.041
$\|y_h\|_{L^2(Q_T)}$	2.01×10^{-1}	1.998×10^{-1}	1.990×10^{-1}	1.986×10^{-1}	1.984×10^{-1}
$\|y_h(\cdot, T)\|_{L^2(0,1)}$	1.13×10^{-3}	3.00×10^{-4}	7.59×10^{-5}	1.89×10^{-5}	4.74×10^{-6}
$\|y - y_h\|_{L^2(Q_T)}$	6.47×10^{-3}	3.52×10^{-3}	1.59×10^{-3}	5.35×10^{-4}	–
$\|v - v_h\|_{L^2(q_T)}$	1.39×10^{-1}	7.42×10^{-2}	3.31×10^{-2}	1.11×10^{-2}	–

$\omega = (0.2, 0.8)$, $y_0(x) \equiv \sin(\pi x)$, $a(x) \equiv 10^{-1}$ – $\alpha = 1/2$

Table 2 Solution of (52), direct method

$\Delta x, \Delta t$	1/20	1/40	1/80	1/160	1/320
$\kappa(\mathcal{M}_h)$	1.36×10^9	1.05×10^{11}	6.87×10^{12}	4.67×10^{14}	3.07×10^{16}
$\|z_h\|_{L^2(Q_T)}$	9.58	16.18	24.22	33.46	44.11
$\|z_h(\cdot, T)\|_{L^2(0,1)}$	6.84×10^{-1}	1.90×10^{-1}	1.92×10^{-2}	8.05×10^{-3}	1.63×10^{-5}
$\|v_h\|_{L^2(q_T)}$	1.596	2.005	2.334	2.571	2.729
$\|y_h\|_{L^2(Q_T)}$	1.881×10^{-1}	1.837×10^{-1}	1.827×10^{-1}	1.827×10^{-1}	1.829×10^{-1}
$\|y_h(\cdot, T)\|_{L^2(0,1)}$	4.09×10^{-3}	1.65×10^{-3}	5.65×10^{-4}	1.68×10^{-4}	4.62×10^{-5}
$\|y - y_h\|_{L^2(Q_T)}$	7.92×10^{-2}	5.01×10^{-2}	2.70×10^{-2}	1.07×10^{-2}	–
$\|v - v_h\|_{L^2(q_T)}$	1.580	1.064	0.613	0.258	–

$\omega = (0.3, 0.6)$, $y_0(x) \equiv \sin(\pi x)$, $a(x) \equiv 10^{-1}$ – $\alpha = 1/2$

to $L^\infty(Q_T)$. Tables 1 and 2 collect relevant numerical values concerning the computed solutions to (52) for $\omega = (0.2, 0.8)$ and $\omega = (0.3, 0.6)$, respectively. For $\omega = (0.2, 0.8)$, we take $\beta_0 = \beta_{0,1/2}$. For $\omega = (0.3, 0.6)$, we take $\beta_0 = \beta_{0,0.45}$. Moreover, for simplicity, we always set $\Delta x = \Delta t$.

These tables clearly show that the solution z_h converges as $h \to 0$, as well as v_h and y_h. The influence of the size of ω on the norm of the control is also emphasized. The absolute errors, displayed in the last two rows, are computed assuming that $h = (1/320, 1/320)$ provides a reference solution.

For $\omega = (0.2, 0.8)$, we observe that $\|y - y_h\|_{L^2(Q_T)} = \mathcal{O}(h^{1.19})$ and $\|v - v_h\|_{L^2(q_T)} = \mathcal{O}(h^{1.25})$, while for $\omega = (0.3, 0.6)$ we observe a slightly slower convergence: $\|y - y_h\|_{L^2(Q_T)} = \mathcal{O}(h^{0.95})$ and $\|v - v_h\|_{L^2(q_T)} = \mathcal{O}(h^{0.85})$. We also check that the null controllability requirement (2) is very well satisfied: indeed, we observe that $\|y_h(\cdot, T) - y(\cdot, T)\|_{L^2(0,1)} = \mathcal{O}(h^{1.97})$ and $\|y_h(\cdot, T) - y(\cdot, T)\|_{L^2(0,1)} = \mathcal{O}(h^{1.65})$ for $\omega = (0.2, 0.8)$ and $\omega = (0.3, 0.6)$, respectively. Notice that, as a consequence of the change of variable (18), the L^2 norm of $z_h(\cdot, T)$ remains bounded.

In Tables 3 and 4 we consider the case $\alpha = 0$, once again with $\omega = (0.2, 0.8)$ and $\omega = (0.3, 0.6)$. In this case, the coefficient A_1 is weakly singular, like $(T - t)^{-1/2}$; see (21). This is simply handled by numerically replacing $(T - t)$ by $(T - t + 10^{-10})$. The approximation

Table 3 Solution of (52), direct method

$\Delta x, \Delta t$	$1/20$	$1/40$	$1/80$	$1/160$	$1/320$
$\kappa(\mathcal{M}_h)$	1.69×10^{12}	5.05×10^{12}	1.22×10^{13}	2.88×10^{13}	1.06×10^{14}
$\|z_h\|_{L^2(Q_T)}$	1.005	1.113	1.191	1.239	1.266
$\|z_h(\cdot,T)\|_{L^2(0,1)}$	5.37×10^{-10}	2.61×10^{-10}	6.27×10^{-11}	7.07×10^{-12}	2.87×10^{-13}
$\|v_h\|_{L^2(q_T)}$	0.971	1.003	1.023	1.035	1.041
$\|y_h\|_{L^2(Q_T)}$	2.011×10^{-1}	1.998×10^{-1}	1.990×10^{-1}	1.986×10^{-1}	1.984×10^{-1}
$\|y_h(\cdot,T)\|_{L^2(0,1)}$	6.17×10^{-4}	1.56×10^{-4}	3.83×10^{-5}	9.44×10^{-6}	2.35×10^{-6}
$\|y-y_h\|_{L^2(Q_T)}$	6.27×10^{-3}	3.43×10^{-3}	1.56×10^{-3}	5.28×10^{-4}	–
$\|v-v_h\|_{L^2(q_T)}$	1.36×10^{-1}	7.26×10^{-2}	3.25×10^{-2}	1.09×10^{-2}	–

$\omega = (0.2, 0.8)$, $y_0(x) \equiv \sin(\pi x)$, $a(x) \equiv 10^{-1} - \alpha = 0$

Table 4 Solution of (52), direct method

$\Delta x, \Delta t$	$1/20$	$1/40$	$1/80$	$1/160$	$1/320$
$\kappa(\mathcal{M}_h)$	3.70×10^{12}	1.12×10^{13}	3.33×10^{13}	1.01×10^{14}	3.03×10^{14}
$\|z_h\|_{L^2(Q_T)}$	4.664	7.725	10.98	14.30	17.63
$\|z_h(\cdot,T)\|_{L^2(0,1)}$	3.98×10^{-9}	1.36×10^{-9}	3.05×10^{-10}	1.19×10^{-11}	3.40×10^{-13}
$\|v_h\|_{L^2(q_T)}$	1.597	2.023	2.348	2.58	2.733
$\|y_h\|_{L^2(Q_T)}$	1.879×10^{-1}	1.834×10^{-1}	1.826×10^{-1}	1.827×10^{-1}	1.829×10^{-1}
$\|y_h(\cdot,T)\|_{L^2(0,1)}$	4.96×10^{-3}	1.82×10^{-3}	5.91×10^{-4}	1.71×10^{-4}	4.65×10^{-5}
$\|y-y_h\|_{L^2(Q_T)}$	7.52×10^{-2}	4.82×10^{-2}	2.62×10^{-2}	1.04×10^{-2}	–
$\|v-v_h\|_{L^2(q_T)}$	1.57	1.04	0.59	0.25	–

$\omega = (0.3, 0.6)$, $y_0(x) \equiv \sin(\pi x)$, $a(x) \equiv 10^{-1} - \alpha = 0$

Table 5 Solution of (52), direct method $\omega = (0.3, 0.6)$

$\Delta x, \Delta t$	$1/20$	$1/40$	$1/80$	$1/160$
$\kappa(\mathcal{M}_h)$	3.70×10^{12}	1.12×10^{13}	3.33×10^{13}	1.01×10^{14}
$\kappa(\mathcal{M}_{1,h})$	3.52×10^{15}	2.56×10^{27}	2.13×10^{50}	2.48×10^{95}

First line with change of variable, $\alpha = 0 - \kappa(\mathcal{M}_h) = O(h^{-1.58})$, *Second line* without change of variable $\kappa(\mathcal{M}_{1,h}) = O(e^{h^{-0.87}})$

z_h is different here, but we can check that the control function v_h and the corresponding solution y_h are independent of α, so that the rates of convergence are very similar.

A relevant feature of these tables is that they show that the condition number $\kappa(\mathcal{M}_h)$ of the matrix $\mathcal{M}_h$ depends polynomially on $h = (\Delta x, \Delta t)$. The condition number is defined here as follows:

$$\kappa(\mathcal{M}_h) = ||\mathcal{M}_h||_2||\mathcal{M}_h^{-1}||_2,$$

Table 6 $\omega = (0.3, 0.6)$ and $\alpha = 0$—numbers of iterates of GMRES needed to solve (22) (*second line*) and (15) (*third line*)

$\Delta x, \Delta t$	1/20	1/40	1/80	1/160
$2N_x N_t$	462	1722	6642	26082
$\#\text{GMRES}(\mathcal{M}_h)$	385	1297	4141	7201
$\#\text{GMRES}(\mathcal{M}_{1,h})$	411	1501	5890	11567

where the norm $||\mathcal{M}_h||_2$ stands for the largest singular value of $\mathcal{M}_h$. Thus, for $\alpha = 1/2$, we get $\kappa(\mathcal{M}_h) = \mathcal{O}(h^{-6.12})$ and $\kappa(\mathcal{M}_h) = \mathcal{O}(h^{-6.09})$, respectively for $\omega = (0.2, 0.8)$ and $\omega = (0.3, 0.6)$, while for $\alpha = 0$ we get $\kappa(\mathcal{M}_h) = \mathcal{O}(h^{-1.44})$ and $\kappa(\mathcal{M}_h) = \mathcal{O}(h^{-1.58})$ (notice that $\kappa(\mathcal{M}_h)$ is not too sensitive to ω).

This polynomial evolution contrasts with the exponential growth observed when we do not introduce the change (18) and we consider directly the formulation (15). Table 5 provides some numerical values in both situations and definitively highlights the influence of the change of variable on the condition number.

If we use an iterative method to solve the linear system, we can see that the relevance of the change of variable in terms of the number of iterates is also significative, although less important. For the same data considered in Table 5, Table 6 gives the number of iterates needed by GMRES (without restart and without preconditioner) obtained with and without change of variable. The tolerance is here $\sigma = 10^{-6}$.

The computed state and control for $h = (1/80, 1/80)$ and $\omega = (0.3, 0.6)$ are displayed in Figs. 3 and 4.

From these results, we see that the finite dimensional formulation (52) provides an efficient and robust method to approximate null controls for the heat equation (1). Let us however mention two drawbacks:

- First, for any fixed h, the controlled state computed by solving (1) numerically does not satisfy exactly the null controllability condition at time $t = T$; this is mainly explained by the fact that, in (17), this requirement appears as a Neumann condition.
- Secondly (and above all), the method requires a finite element approximation that must be C^1 in space (in higher dimension, this involves the use of specific and complex finite elements; see [6]).

We will see how to circumvent these two points in the next section.

4 A second method: solving a mixed variational formulation

Let us introduce the new variables $m = \rho^{-1}L^\star p$ and $r = \rho_0^{-1}p$ and let us rewrite (15) in the form

$$\begin{cases} \iint\limits_{Q_T} m\,\overline{m}\,dx\,dt + \iint\limits_{q_T} r\,\overline{r}\,dx\,dt = \int\limits_0^1 \rho_0(x,0)y_0(x)\,\overline{r}(x,0)\,dx \\ \forall(\overline{m},\overline{r}) \text{ with } \rho^{-1}L^\star(\rho_0\overline{r}) - \overline{m} = 0 \text{ and } \overline{r} \in \rho_0^{-1}P;\ \rho^{-1}L^\star(\rho_0 r) - m = 0 \text{ and } r \in \rho_0^{-1}P. \end{cases} \tag{56}$$

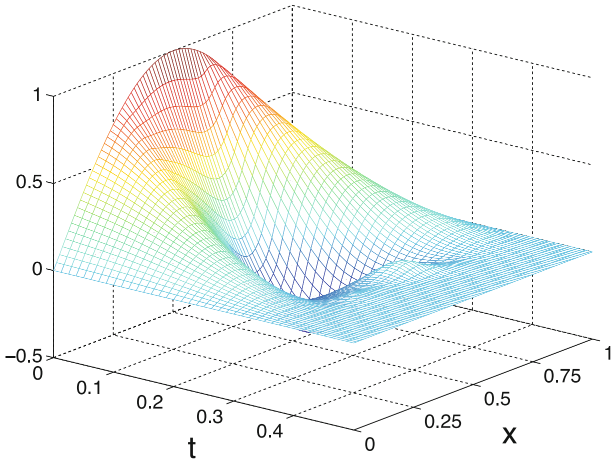

Fig. 3 $\omega = (0.3, 0.6)$. The computed state y_h

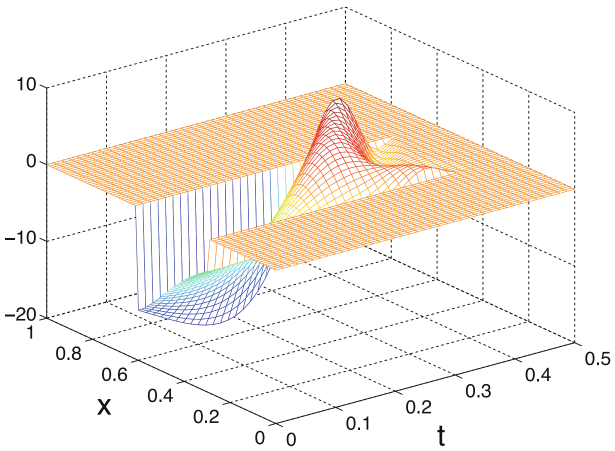

Fig. 4 $\omega = (0.3, 0.6)$. The computed control v_h

Let us introduce the spaces $M = L^2(Q_T)$, $R := \rho_0^{-1} P$ and $\tilde{M} := (T - t)^{1/2} M$, the bilinear forms

$$a(\,(m, r), (\overline{m}, \overline{r})\,) = \iint\limits_{Q_T} m\,\overline{m}\,dx\,dt + \iint\limits_{q_T} r\,\overline{r}\,dx\,dt \quad \forall (m, r), (\overline{m}, \overline{r}) \in M \times R$$

and

$$b(\,(\overline{m}, \overline{r}), \mu\,) = \iint\limits_{Q_T} (\rho^{-1} L^\star(\rho_0 \overline{r}) - \overline{m})\,\mu\,dx\,dt \quad \forall (\overline{m}, \overline{r}) \in M \times R, \quad \forall \mu \in \tilde{M}$$

and the linear form

$$\langle \ell, (\overline{m}, \overline{r}) \rangle = \int_0^1 \rho_0(x, 0) y_0(x)\,\overline{r}(x, 0)\,dx \quad \forall (\overline{m}, \overline{r}) \in M \times R.$$

Then, it is not difficult to check that $a(\cdot\,,\cdot)$, $b(\cdot\,,\cdot)$ and ℓ are well-defined and continuous and a mixed formulation of problem (56) is the following:

$$\begin{cases} a(\,(m,r),(\overline{m},\overline{r})\,)+b(\,(\overline{m},\overline{r}),\lambda\,)=\langle \ell,(\overline{m},\overline{r})\rangle \quad \forall(\overline{m},\overline{r})\in M\times R \\ b(\,(m,r),\mu\,)=0 \quad \forall \mu\in\tilde{M} \\ (m,r)\in M\times R, \quad \lambda\in\tilde{M}. \end{cases} \tag{57}$$

We can now state and prove an existence and uniqueness result:

Theorem 4.1 *There exists a unique solution* (m,r,λ) *to* (57)*. Moreover,* $y:=\rho^{-1}m$ *is, together with* $v:=-\rho_0^{-1}r|_{q_T}$*, the unique solution to* (6)*.*

Proof It is clear that, if (m,r,λ) solves (57), then $m=\rho^{-1}L^{\star}(\rho_0 r)$, $p=\rho_0 r$ is the unique solution to (15) and, consequently, the unique solution to (6) is given by $y=\rho^{-1}m$ and $v=-\rho_0^{-1}r|_{q_T}$.

Let us introduce the space

$$\begin{aligned} V &= \{(m,r)\in M\times R : b(\,(m,r),\mu\,)=0 \;\; \forall\mu\in\tilde{M}\} \\ &= \{(m,r)\in M\times R : m=\rho^{-1}L^{\star}(\rho_0 r)\}. \end{aligned}$$

In order to prove that (57) possesses exactly one solution, we will apply a general result concerning mixed variational problems. More precisely, we will check that

- $a(\cdot\,,\cdot)$ is coercive on V, that is:

$$a(\,(\overline{m},\overline{r}),(\overline{m},\overline{r})\,)\ge\kappa_1\|(\overline{m},\overline{r})\|^2_{M\times R} \quad \forall(\overline{m},\overline{r})\in V, \quad \kappa_1>0. \tag{58}$$

- $b(\cdot\,,\cdot)$ satisfies the usual "inf-sup" condition with respect to $M\times R$ and $\tilde{M}$, i.e.

$$\kappa_2 := \inf_{\mu\in\tilde{M}}\ \sup_{(\overline{m},\overline{r})\in M\times R}\frac{b(\,(\overline{m},\overline{r}),\mu\,)}{\|(\overline{m},\overline{r})\|_{M\times R}\|\mu\|_{\tilde{M}}}>0. \tag{59}$$

This will suffice to ensure existence and uniqueness; see for instance [4].

The proofs of (58) and (59) are straightforward. Indeed, we first notice that for any $(\overline{m},\overline{r})\in V$ one has $\overline{m}=\rho^{-1}L^{\star}(\rho_0\overline{r})$ and thus

$$\begin{aligned} a(\,(\overline{m},\overline{r}),(\overline{m},\overline{r})\,) &= \frac{1}{2}\iint_{Q_T}|\overline{m}|^2\,dx\,dt+\frac{1}{2}\iint_{Q_T}\rho^{-2}|L^{\star}(\rho_0\overline{r})|^2\,dx\,dt+\iint_{q_T}|\overline{r}|^2\,dx\,dt \\ &\ge \frac{1}{2}\,\|(\overline{m},\overline{r})\|^2_{M\times R} \end{aligned}$$

This proves (58).

On the other hand, for any $\mu\in\tilde{M}$ there exists $(\tilde{m},\tilde{r})\in M\times R$ such that

$$b(\,(\tilde{m},\tilde{r}),\mu\,)=\iint_{Q_T}(T-t)^{-1}|\mu|^2\,dx\,dt \quad\text{and}\quad \|(\tilde{m},\tilde{r})\|_{M\times R}\le C\|\mu\|_{\tilde{M}}.$$

For instance, it suffices to take $(\tilde{m},\tilde{r})=(-(T-t)^{-1/2}\mu,0)$. Since $\mu\in\tilde{M}$, we have $(T-t)^{-1/2}\mu\in L^2(Q_T)$ and, consequently, $(-(T-t)^{-1/2}\mu,0)\in M\times R$. Moreover,

$$\sup_{(\overline{m},\overline{r})\in M\times R}\frac{b(\,(\overline{m},\overline{r}),\mu\,)}{\|(\overline{m},\overline{r})\|_{M\times R}\|\mu\|_{\tilde{M}}}\ge\frac{b(\,(\tilde{m},\tilde{r}),\mu\,)}{\|(\tilde{m},\tilde{r})\|_{M\times R}\|\mu\|_{\tilde{M}}}\ge\frac{1}{C}$$

and we also have (59). □

As usual, solving (57) is equivalent to finding the *saddle-points* of a *Lagrangian.* In this case, the Lagrangian is given by

$$\begin{aligned}\mathcal{L}(\overline{m},\overline{r};\mu) &= \frac{1}{2}a((\overline{m},\overline{r}),(\overline{m},\overline{r})) + b((\overline{m},\overline{r}),\mu) - \langle \ell,(\overline{m},\overline{r})\rangle \\ &= \frac{1}{2}\left(\iint_{Q_T}|\overline{m}|^2\,dx\,dt + \iint_{q_T}|\overline{r}|^2\,dx\,dt\right) + \iint_{Q_T}(\rho^{-1}L^\star(\rho_0\overline{r}) - \overline{m})\,\mu\,dx\,dt \\ &\quad - \int_0^1 \rho_0(x,0)\,y_0(x)\,\overline{r}(x,0)\,dx \end{aligned} \tag{60}$$

for all $(\overline{m},\overline{r},\mu) \in M \times R \times \tilde{M}$. This can be viewed as the starting point of a large family of iterative methods for the solution of (57).[1]

Remark 3 The variable r coincides with the variable z (given by (18)) for $\alpha = 0$. Therefore, the term $\rho L^\star(\rho^{-1}\overline{r})$ in the bilinear form $b(\cdot\,,\cdot)$ possesses a singularity of the kind $(T-t)^{-1/2}$, see (20)–(21), that we can easily cancel by replacing the multiplier μ by $(T-t)^\gamma\mu$, with $\gamma \geq 1/2$.

4.1 A non-conformal mixed finite element approximation

For any $h = (\Delta x, \Delta t)$ as before, let us consider again the associated uniform quadrangulation $\mathcal{Q}_h$. We now introduce the following finite dimensional spaces:

$$M_h = \{z_h \in C^0(\overline{Q}_T) : z_h|_K \in (\mathbb{P}_{1,x}\otimes\mathbb{P}_{1,t})(K)\ \ \forall K \in \mathcal{Q}_h\}, \quad Q_h = M_h \ \text{ and} \tag{61}$$

$$\tilde{M}_h = \{\mu_h \in Q_h : \mu_h|_{t=T} \equiv 0\}. \tag{62}$$

We are now dealing with usual C^0 finite element spaces. We have $M_h \subset M$ but, contrarily, $Q_h \not\subset R = \rho_0^{-1}P$ (of course, this is the price we have to pay in order to use C^0 finite elements).

Let us introduce the bilinear form $b_h(\cdot\,,\cdot)$, with

$$\begin{cases} b_h((\overline{m}_h,\overline{r}_h),\mu_h) = \displaystyle\iint_{Q_T}\left(\rho^{-1}(-(\rho_0\overline{r}_h)_t + A\rho_0\overline{r}_h)\mu_h + a(\rho_0\overline{r}_h)_x(\rho^{-1}\mu_h)_x - \overline{m}_h\mu_h\right)dx\,dt \\ \forall(\overline{m}_h,\overline{r}_h) \in M_h \times Q_h,\ \ \forall\mu_h \in \tilde{M}_h. \end{cases}$$

Let us also set

$$\langle \ell_h,(\overline{m}_h,\overline{r}_h)\rangle = \int_0^1 \rho_0(x,0)\,y_0(x)\,\overline{r}_h(x,0)\,dx \quad \forall(\overline{m}_h,\overline{r}_h) \in M_h \times Q_h.$$

Then the mixed finite element approximation of (57) is the following:

$$\begin{cases} a((m_h,r_h),(\overline{m}_h,\overline{r}_h)) + b_h((\overline{m}_h,\overline{r}_h),\lambda_h) = \langle \ell,(\overline{m}_h,\overline{r}_h)\rangle & \forall(\overline{m}_h,\overline{r}_h) \in M_h \times Q_h \\ b_h((m_h,r_h),\mu_h) = 0 & \forall\mu_h \in \tilde{M}_h \\ (m_h,r_h) \in M_h \times Q_h, \quad \lambda_h \in \tilde{M}_h. \end{cases} \tag{63}$$

[1] In fact, one of these methods will be considered below, in Sect. 5 in the context of a problem similar to (57) in higher spatial dimension.

Table 7 Solution of (63), direct method

$\Delta x, \Delta t$	$1/20$	$1/40$	$1/80$	$1/160$	$1/320$
$\kappa(\mathcal{M}_h)$	1.47×10^5	8.30×10^5	6.48×10^6	5.02×10^7	–
$\Vert v_h\Vert_{L^2(q_T)}$	0.974	1.006	1.025	1.036	1.041
$\Vert y_h\Vert_{L^2(Q_T)}$	2.001×10^{-1}	1.996×10^{-1}	1.989×10^{-1}	1.986×10^{-1}	1.984×10^{-1}
$\Vert y - y_h\Vert_{L^2(Q_T)}$	6.32×10^{-3}	3.21×10^{-3}	1.41×10^{-3}	4.75×10^{-4}	–
$\Vert v - v_h\Vert_{L^2(q_T)}$	1.27×10^{-1}	6.56×10^{-2}	2.90×10^{-2}	9.72×10^{-3}	–

$\omega = (0.2, 0.8)$, $y_0(x) \equiv \sin(\pi x)$, $a(x) \equiv 10^{-1}$ – $\gamma = 1/2$

Table 8 Solution of (63), direct method

$\Delta x, \Delta t$	$1/20$	$1/40$	$1/80$	$1/160$	$1/320$
$\kappa(\mathcal{M}_h)$	1.39×10^5	8.78×10^5	6.62×10^6	4.76×10^7	–
$\Vert v_h\Vert_{L^2(q_T)}$	1.865	2.339	2.651	2.830	2.936
$\Vert y_h\Vert_{L^2(Q_T)}$	1.836×10^{-1}	1.814×10^{-1}	1.817×10^{-1}	1.822×10^{-1}	1.826×10^{-1}
$\Vert y - y_h\Vert_{L^2(Q_T)}$	7.13×10^{-2}	3.82×10^{-2}	1.78×10^{-2}	6.33×10^{-3}	–
$\Vert v - v_h\Vert_{L^2(q_T)}$	1.56	0.957	0.489	0.182	–

$\omega = (0.3, 0.6)$, $y_0(x) \equiv \sin(\pi x)$, $a(x) \equiv 10^{-1}$ – $\gamma = 1/2$

The good property of this approach is that there is no weight growing exponentially as $t \to T^-$. The worst behavior is found in the computation of $\rho^{-1}(\rho_0 \overline{r}_h)_t\, \mu_h$, which behaves at most like $(T-t)^{-1/2}$, but this singularity is weak, numerically acceptable and removable (see Remark 3).

Unfortunately, it is unknown at present whether or not a uniform "inf-sup" condition is satisfied by $b_h(\cdot\,,\cdot)$ in the product space $M_h \times Q_h$. As a consequence, we are not able to prove the convergence of the solution to (63) as $h \to 0$. However, in practice, the numerical experiments we have performed show that (63) possesses exactly one solution $(m_h, r_h, \lambda_h) \in M_h \times Q_h \times \tilde{M}_h$ for each h that is stable and converges.

4.2 Numerical experiments (II)

We present in this section some numerical experiments obtained by solving (63).

Three unknown functions, m_h, r_h and μ_h, are involved in the formulation (to be compared with (52), where only the variable z_h appears). The corresponding matrix, which is sparse and symmetric, is of order $3N_x N_t$. Once again, its components, as well as the terms in the right hand side, are computed with exact integration formulae. Moreover, the linear system can be solved again with a direct method. Once the triplet (m_h, r_h, μ_h) is computed, the numerical solution (y_h, v_h) is given directly by

$$y_h = \pi_h(\rho^{-1})m_h, \quad v_h = -\pi_h(\rho_0^{-1})r_h|_{q_T}. \tag{64}$$

First, we consider again the data of Sect. 3.3, that is, $y_0(x) \equiv \sin(\pi x)$, $a \equiv 0.1$ and $T = 1/2$. We also take $\gamma = 1/2$ (see Remark 3). Tables 7 and 8 give the norms of v_h and y_h for various h and $\omega = (0.2, 0.8)$ and $\omega = (0.3, 0.6)$, respectively (here, $\mathcal{M}_h$ denotes the matrix associated to (63)).

Table 9 Solution of (63), direct method

$\Delta x, \Delta t$	$1/20$	$1/40$	$1/80$	$1/160$	$1/320$
$\kappa(\mathcal{M}_h)$	4.18×10^5	5.02×10^5	6.04×10^5	1.17×10^6	–
$\\|v_h\\|_{L^2(q_T)}$	0.976	1.007	1.026	1.036	1.041
$\\|y_h\\|_{L^2(Q_T)}$	2.008×10^{-1}	1.996×10^{-1}	1.989×10^{-1}	1.986×10^{-1}	1.984×10^{-1}
$\\|y - y_h\\|_{L^2(Q_T)}$	6.24×10^{-3}	3.17×10^{-3}	1.41×10^{-3}	4.73×10^{-4}	–
$\\|v - v_h\\|_{L^2(q_T)}$	1.25×10^{-1}	6.48×10^{-2}	2.87×10^{-2}	9.67×10^{-3}	–

$\omega = (0.2, 0.8)$, $y_0(x) \equiv \sin(\pi x)$, $a(x) \equiv 10^{-1} - \gamma = 0$

Table 10 Solution of (63), GMRES

$\Delta x, \Delta t$	$1/20$	$1/40$	$1/80$	$1/160$	$1/320$
$\kappa(\mathcal{M}_h)$	1.43×10^6	3.29×10^6	8.75×10^6	4.67×10^7	–
$3N_x N_t$	693	2593	9963	39123	–
#GMRES(M_h)	457	1535	5678	10350	–
$\\|v_h\\|_{L^2(q_T)}$	1.849	2.335	2.650	2.833	2.936
$\\|y_h\\|_{L^2(Q_T)}$	1.84×10^{-1}	1.814×10^{-1}	1.817×10^{-1}	1.822×10^{-1}	1.826×10^{-1}
$\\|y - y_h\\|_{L^2(Q_T)}$	7.21×10^{-2}	3.83×10^{-2}	1.79×10^{-2}	6.35×10^{-3}	–
$\\|v - v_h\\|_{L^2(q_T)}$	1.587	0.962	0.491	0.183	–

$\omega = (0.3, 0.6)$, $y_0(x) \equiv \sin(\pi x)$, $a(x) \equiv 10^{-1} - \gamma = 0$

The numerical values agree with those obtained with the previous method, see Tables 1 and 2. If we compare closer, the case $\omega = (0.3, 0.6)$ suggests a faster convergence of the norms $\|v_h\|_{L^2(q_T)}$ and $\|y_h\|_{L^2(Q_T)}$. Assuming again that $h = (1/320, 1/320)$ provides a reference solution, we see that $\|v - v_h\|_{L^2(q_T)} = \mathcal{O}(h^{1.23})$ and $\|y - y_h\|_{L^2(Q_T)} = \mathcal{O}(h^{1.23})$ for $\omega = (0.2, 0.8)$ and $\|v - v_h\|_{L^2(q_T)} = \mathcal{O}(h^{1.15})$ and $\|y - y_h\|_{L^2(Q_T)} = \mathcal{O}(h^{1.02})$ for $\omega = (0.3, 0.6)$. Very similar values are observed for $\gamma = 0$, for which the coefficient B_1 is weakly singular at $t = T$.

The main difference observed with respect to the method described in Sect. 3 is the size of the condition numbers $\kappa(\mathcal{M}_h)$, which is significantly reduced. Once again, the $\kappa(\mathcal{M}_h)$ behave polynomially with respect to h (Table 9).

We also report in Table 10 the number of iterates leading to the convergence of GMRES versus h. These values can be compared to those in Table 6. Let us emphasize that, here, as a consequence of (64), the null controllability condition is exactly satisfied, that is, $y_h(x, T) \equiv 0$.

As we have seen, the measure of the support $|\omega|$ may affect the convergence of the approximation. Contrarily, , the regularity of the initial condition has no impact in practicedue to the regularizing effect of the heat operator. More determinant are the norm (and the sign) of the potential A, the size of the controllability time T and, of course, the size of the diffusion coefficient.

Let us consider a much more stiff situation. The diffusion coefficient will be now a non-constant C^1 function: we take $D_1 = (0, 0.45)$, $D_2 = (0.55, 1)$, $a_1 = 1$ and $a_2 = 1/15$ and we assume that a is the C^1 function that coincides with a polynomial of the third order in $(0, 1)\backslash(D_1 \cup D_2)$ and satisfies

$$a(x) \equiv a_i \text{ in } D_i.$$

Table 11 Solution of (63), GMRES

$\Delta x, \Delta t$	1/40	1/80	1/160	1/320	1/640
$\kappa(_h)$	1.05×10^8	8.28×10^9	2.32×10^{11}	7.06×10^{13}	–
$3N_x N_t$	861	3321	13041	51681	205761
$\#\text{GMRES}(M_h)$	631	2912	10211	33091	–
$\\|v_h\\|_{L^2(q_T)}$	14.44	19.70	23.48	25.41	26.01
$\\|y_h\\|_{L^2(Q_T)}$	3.58×10^{-1}	4.67×10^{-1}	5.59×10^{-1}	6.18×10^{-1}	6.43×10^{-1}
$\\|y - y_h\\|_{L^2(Q_T)}$	6.30×10^{-1}	3.21×10^{-1}	9.15×10^{-2}	4.76×10^{-2}	–
$\\|v - v_h\\|_{L^2(q_T)}/\\|v\\|_{L^2(q_T)}$	1.21	0.45	0.23	0.09	–

$\omega = (0.2, 0.4)$ – Numerical norms for a stiff case $\gamma = 0$

In particular, $\min(a_1, a_2) \leq a(x) \leq \max(a_1, a_2)$ in $(0, 1)$; the coefficient a is displayed in Fig. 1-right.

We take $\omega = (0.2, 0.4)$ (where the diffusion is higher), $\beta_0 = \beta_{0,0.3}$, $T = 1/2$ and we localize y_0 in D_2, where the diffusion is low: $y_0(x) \equiv e^{-100(x-3/4)^2}\, 1_{(0,1)}$. Finally, we take $A \equiv -1$ (of course, the effect of A is opposite to diffusion, which enhances the action of the control).

Table 11 collects some numerical values. The numerical convergence as $h \to 0$ is observed. Using now the solutions associated to $h = (1/640, 1/640)$ as a reference solution, we see that $\|v - v_h\| = \mathcal{O}(h^{0.94})$ and $\|y - y_h\| = \mathcal{O}(h^{1.1})$.

In Figs. 5 and 6, the computed controlled state and null control are displayed for $h = (1/80, 1/80)$. The action of the control is here much stronger than in previous examples; in particular, $\|v_h\|_{L^\infty(Q_T)} \approx 2.58 \times 10^2$. As a consequence, y_h takes relatively large values for $t < T$: $\|y_h\|_{L^\infty(Q_T)} \approx 3.24$, while $\|y_0\|_{L^\infty(0,1)}$ is only equal to one.

This situation can be amplified for (weakly explosive) semilinear heat equations; see [10] for more details. Let us also mention the work [25], that also rely on a variational reformulation of the controllability and allow to obtain both boundary and inner controls.

5 Further comments and concluding remarks

5.1 Numerical analysis and dual methods : related works

As mentioned in the introduction, we may find the solution to (6) by applying dual methods in the spirit of the pioneering contribution of Carthel, Glowinski and Lions in [5] (see also [16]); this relies on appropriate reformulations of (6) as unconstrained problems, that use new (dual) variables. This has been done in [11]. There, the influence of the weights is discussed and the numerical results are compared with the methods given in this paper. Very similar results are observed.

It should be noted that these two approaches are qualitatively different:

- The functions v_h computed from the solutions to 52 for various h are not *a priori* null controls for discrete systems (associated to the heat equation (1)), but simply approximations of the null control v furnished by the solution to (22). No heat equation has to be solved in order to compute the approximations v_h.

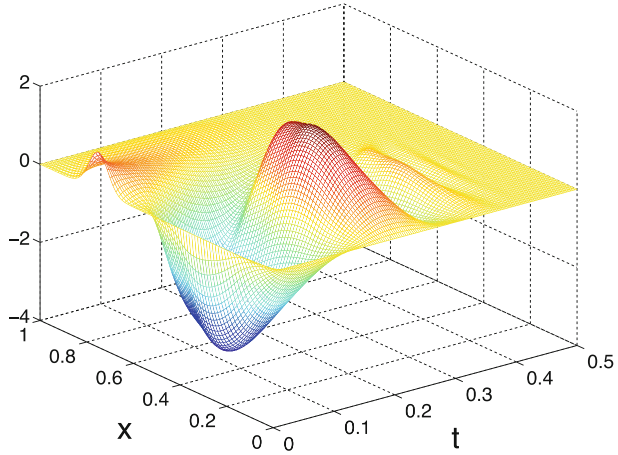

Fig. 5 $\omega = (0.2, 0.4) - y_0(x) = e^{-300(x-3/4)^2}$. The computed state y_h

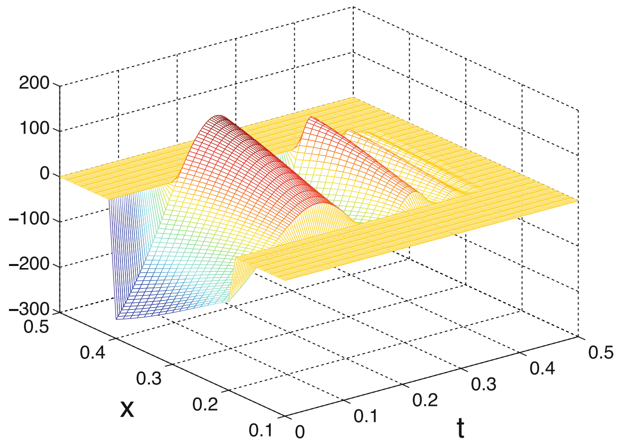

Fig. 6 $\omega = (0.2, 0.4) - y_0(x) = e^{-300(x-3/4)^2}$. The computed control v_h

- This is in contrast with the situation found by applying dual methods which, for any fixed h, provide a null control for a discrete finite dimensional system, a consistent approximation of the heat equation. Accordingly, the numerical analysis requires some uniform discrete Carleman estimates, that may not hold or may be difficult to prove for standard schemes.

Related to the dual framework, let us mention [23], 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 of a finite dimensional approximation to the heat equation; see [31], where a counter-example for finite differences due to O. Kavian is described.

In [19], 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 [19] to the fully discrete situation and proves the convergence towards a semi-discrete control, as the time step Δt tends to zero. Let us also mention [9], where the authors prove that any controllable parabolic equation, be it discrete or continuous in space, is null-controllable after time discretization through the application of an appropriate filtering of the high frequencies. As far as we know, a strong convergence result in the framework of dual method is still missing.

In terms of computational cost, the primal approach requires only the resolution of a linear system involving a definite positive, sparse and symmetric matrix, for which efficient LU solvers are well known. By contrast, the dual approach leads to the minimization of conjugate functions. Such minimization is numerically ill-posed, so that gradient-like methods may diverge as the discretization parameters Δx and Δt go to zero.

5.2 The lack of regularity of $p(\cdot, T)$

An important feature of the problem satisfied by p is that it gives no information on the regularity of $p(\cdot, T)$. There are several arguments that justify this lack of information. One of them is the following.

Assume that $\omega \neq (0, 1)$ and set $j(w) = J(y_w, w)$, where y_w is the (unique) solution to (1) with v replaced by w and let us set $Kw = z_w(\cdot, T)$, where z_w is the solution to (9). Then K can be viewed as a linear continuous mapping on $L^2(q_T)$ with values in a "very small" Hilbert space $R(K)$, that is dense in $L^2(0, 1)$. From the *Lagrange multipliers theorem,* we know that (y, v) solves (6) if and only if

- y solves (1) and
- There exists $\lambda \in R(K)'$ such that

$$(j'(v), w)_{L^2(q_T)} + \langle \lambda, z_w(\cdot, T)\rangle = 0 \quad \forall w \in L^2(q_T), \tag{65}$$

  where $\langle \cdot, \cdot \rangle$ stands for the duality pairing for $R(K)'$ and $R(K)$.

Obviously, $R(K)'$ can be viewed as a "large space" containing $L^2(0, 1)$. The multiplier λ belongs to $R(K)'$, but there is no reason to have $\lambda \in L^2(0, 1)$.

Let (y, v) be an optimal pair and let p be the solution to (17); we know that (y, v) and p satisfy (14). Let $\lambda \in R(K)'$ satisfy (65). From (14) and (17), we find that

$$\begin{aligned}(j'(v), w)_{L^2(q_T)} + \langle \lambda, z_w(\cdot, T)\rangle &= \iint_{q_T} \left(p + \rho_0^2 v\right) w \, dx\, dt + \langle \lambda - p(\cdot, T), z_w(\cdot, T)\rangle \\ &= \langle \lambda - p(\cdot, T), z_w(\cdot, T)\rangle\end{aligned}$$

for all $w \in L^2(q_T)$. Consequently, we should have $p(\cdot, T) = \lambda$, so that $p(\cdot, T)$ does not necessarily belong to $L^2(0, 1)$ (except in the particular case where the control acts on the whole space domain).

In what concerns the dual approach recalled in the introduction (see [11]) and, in particular, the minimal L^2-norm case (that is, $\rho \equiv 0$ and $\rho_0 \equiv 1$), we refer the reader to [2,18,24]. In particular, it is proved in [24] that the set of initial data y_0 for which the corresponding multiplier ϕ_T (the minimizer of $\mathcal{I}$, see (4)) does not belong to $\cup_{s\geq 0} H^{-s}$ is dense in $L^2(0, 1)$. We also refer to [11], where the influence of the exponential weights is discussed in the general case.

5.3 The role of the weights

The explicit introduction of y in the functional J in (6) allows to give expressions of the optimal control and state in terms of the solution p to (15). With $\rho = 0$, this would have not been possible.

The exponential behavior of these weights gives a meaning to the variational formulation (15), reinforces the controllability requirement (through ρ) and regularizes the behavior of the control near $t = T$ (through ρ_0), in contrast with the evolution of the control of minimal L^2-norm, that is highly oscillatory near T.

Carleman estimates ensure the well-posedness of the variational formulation for these specific weights, that blow up exponentially as $t \to T^-$. Notice that this can be ensured only for large K_1 and K_2. In practice, as K_1 increases, the computed control and state vanish at an earlier time (since the requirement $J(y, v) < +\infty$ becomes more severe).

On the other hand, it would be interesting to investigate which are the minimal values of these constants (since the control and state norms become smaller and the control process becomes less expensive as they decrease).

Numerically, using other weights in (15) leads to non-convergent sequences. This is not strange, since the choice $\rho_0 \equiv \rho \equiv 1$ is not supported by theoretical results: it leads to a bilinear form that is coercive in a space where it is unknown whether the linear form is continuous.

5.4 Numerical analysis and error estimates

The variational formulations (22) and (27) lead to strong convergence results for the approximate sequence $\{v_h\}$ as h goes to zero. To our knowledge, this is the first convergence result for the numerical approximation to the null controllability problem for the heat equation. After an appropriate regularity analysis of the solution to (15), one may further obtain estimates of $\|v - v_h\|_{L^2(q_T)}$ in terms of $|h|$ and a suitable norm of p.

The mixed approach provides numerical results for which the null controllability requirement is very well satisfied, even when the diffusion function is only piecewise constant. This is in contrast with the existing literature, mainly devoted to the approximate controllability issue. It will be interesting to analyze rigorously (63) from the viewpoints of stability and convergence. In particular, a key question is whether inequalities like (58) and (59) hold at the finite dimensional level, with constants κ_1 and κ_2 independent of h.

5.5 Some extensions

The methods used in this paper can be extended to cover null controllability problems for linear heat equations in higher spatial dimensions. More precisely, let $\Omega \subset \mathbf{R}^N$ be a regular, bounded, connected open set and let us consider the linear system

$$\begin{cases} y_t - \nabla \cdot (a(x)\nabla y) + A(x,t)\, y = v 1_{\mathcal{O}}, & (x,t) \in \Omega \times (0,T) \\ y(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T) \\ y(x,0) = y_0(x), & x \in \Omega \end{cases} \tag{66}$$

where $a \in C^1(\overline{\Omega})$ with $a(x) \geq a_0 > 0$, $A \in L^\infty(\Omega \times (0,T))$, $\mathcal{O} \subset \Omega$ is a (small) non-empty open set, $v \in L^2(\mathcal{O} \times (0,T))$ is the control and $y_0 \in L^2(\Omega)$ is the initial state. The null controllability problem for (66) is to find, for each $y_0 \in L^2(\Omega)$, a control v such that the

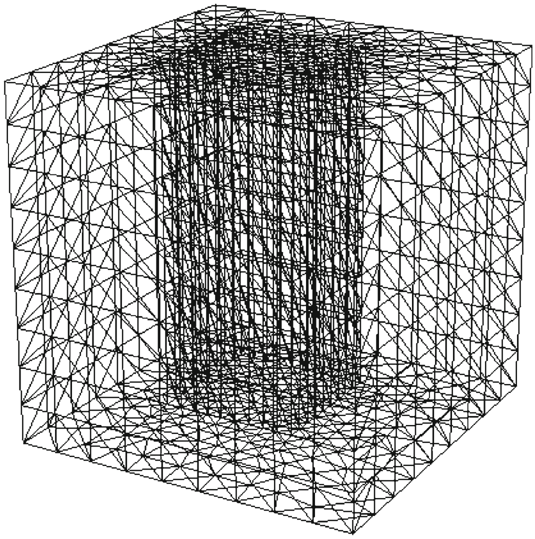

Fig. 7 The space-time domain and the mesh: 2 800 vertices. Total number of unknowns (the values of m_h, r_h and λ_h at the nodal points, see (63)): $6\,846 \times 3 = 20\,538$

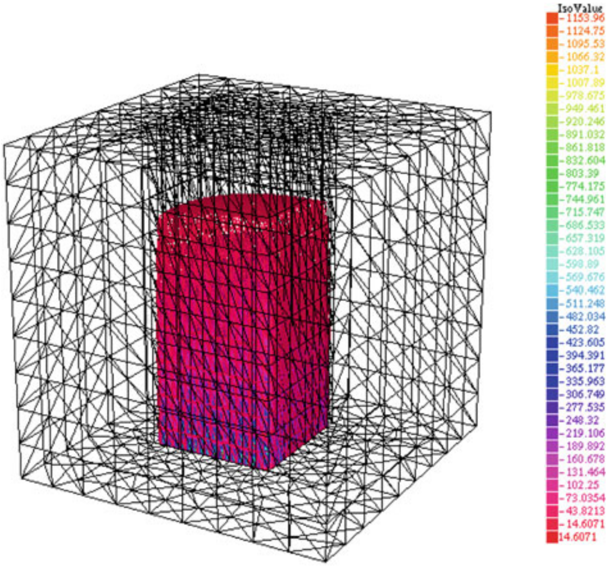

Fig. 8 The surface $u_h(x, t) = 0$ in the (x_1, x_2, t) space. The *region below* $u_h = 0$ is the real support of the computed control

associated solution satisfies

$$y(x, T) = 0, \quad x \in \Omega.$$

The situation is now more involved: first, notice that a result similar to Theorem 2.1 holds, but stronger regularity is needed in order to get Carleman estimates; secondly, observe that the analog of the space P_h in (32) is much more complex in this general case. Fortunately, this can be avoided using mixed formulations, as in Sect. 4.

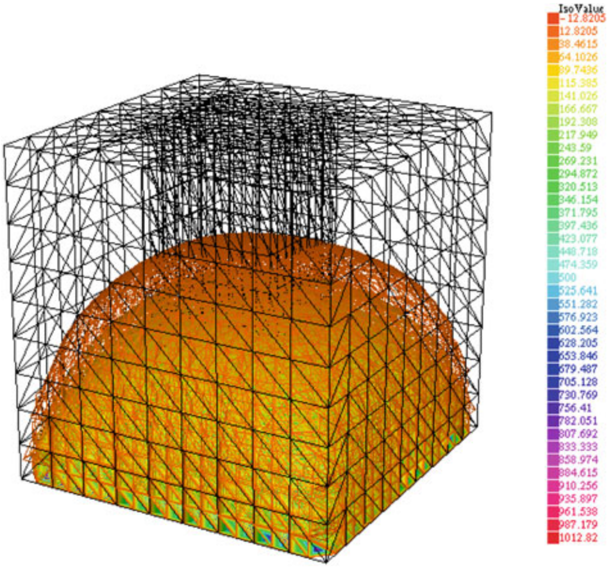

Fig. 9 The surface $y_h(x, t) = 0$ in the (x_1, x_2, t) space

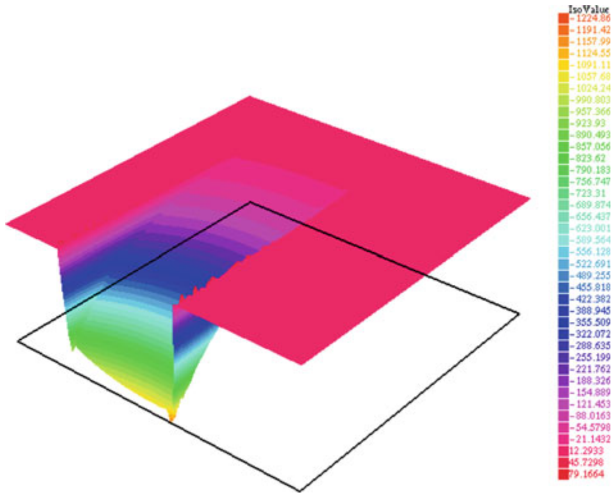

Fig. 10 A 3D view of the function $(x_2, t) \to v_h(x_1, x_2, t)$ at $x_1 = 0.36$. The *horizontal lines* are parallel to the x_2 and t axes and time grows from *left* to *right*

In order to illustrate the situation, let us present the results of an experiment. We solve numerically the null controllability problem for (66) with $N = 2, \Omega = (0, 1) \times (0, 1)$, $\mathcal{O} = (0.2, 0.6) \times (0.2, 0.6)$, $T = 1$, $a(x) \equiv 1$, $A(x, t) \equiv 1$ and $y_0(x) \equiv 1000$. The space-time domain and the mesh are displayed in Fig. 7. We have used a mixed formulation similar to (63), where M_h, Q_h and $\tilde{M}_h$ are standard finite element P_2-Lagrange spaces. The resulting system, in view of its size and structure, has been solved with the Arrow-Hurwicz method, that provides good results, better than a direct solver; see for instance [13, 14] (recall that solving (63) is equivalent to the computation of the saddle-points of a Lagrangian, see the

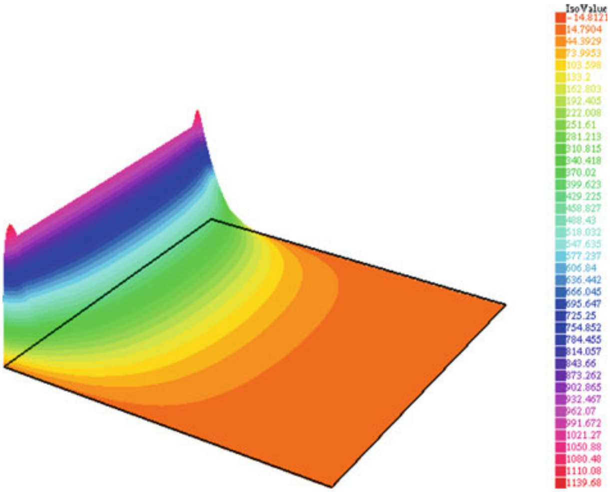

Fig. 11 A 3D view of the function $(x_2, t) \to y_h(x_1, x_2, t)$ at $x_1 = 0.4$. The *horizontal lines* are parallel to the x_2 and t axes and time grows from *left* to *right*

related argument in Sect. 4). The iterates have been stopped for a relative error of two consecutive iterates less than 10^{-5}. The computed control and state are shown in Figs. 8–11. The computations have been performed with the FreeFem++ package, see http://www.freefem.org/ff++. More information, a detailed analysis and other similar numerical experiments will appear in a forthcoming paper.

The previous methods can also be extended to cover many other controllable systems for which appropriate Carleman estimates are available: non-scalar parabolic systems, Stokes and Oseen, 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). Actually, as first noticed in [12] and using in part the results in [30], the approach may also work for linear equations of the hyperbolic kind, where the practical computation of exact controls remains a challenge, see [7]. This work may also be used to address the numerical solution of nonlinear control problems (see [10]), control support optimization (see [26]), etc.

References

1. Alessandrini, G., Escauriaza, L.: Null-controllability of one-dimensional parabolic equations. ESAIM Control Optim. Calc. Var. **14**(2), 284–293 (2008)
2. Belgacem, Ben: F., Kaber, S.M.: On the Dirichlet boundary controllability of the 1-D heat equation: semi-analytical calculations and ill-posedness degre. Inverse Problems **27**, 055012 (2011)
3. Boyer, F., Hubert, F., Le Rousseau, J.: Uniform null-controllability properties for space/time-discretized parabolic equations. Numer. Math. **118**(4), 601–661 (2011)
4. Brezzi, F., Fortin, M.: Mixed and hybrid finite element methods, Springer Series in Computational Mathematics, vol. 15. Springer, New York (1991)
5. Carthel, C., Glowinski, R., Lions, J.-L.: On exact and approximate boundary controllability for the heat equation: a numerical approach. J. Optim. Theory Appl. **82**(3), 429–484 (1994)
6. Ciarlet, P.G.: The finite element method for elliptic problems. Studies in Mathematics and its Applications, vol. 4. North-Holland, Amsterdam (1978)

7. Cîndea, N., Fernández-Cara, E., Münch, A.: Numerical null controllability of the wave equation with a primal approach: convergence results. ESAIM:COCV. doi:10.1051/cocv/2013046. http://hal.archives-ouvertes.fr/hal-00668951
8. Coron, J.-M.: Control and Nonlinearity. Mathematical Surveys and Monographs, vol. 136. AMS, Providence (2007)
9. Ervedoza, S., Valein, J.: On the observability of abstract time-discrete linear parabolic equations. Rev. Mat. Complut. **23**(1), 163–190 (2010)
10. Fernández-Cara, E., Münch, A.: Numerical null controllability of semi-linear 1D heat equations: fixed point, least squares and Newton methods. Math. Control Relat. Fields **2**(3), 217–246 (2012)
11. Fernández-Cara, E., Münch, A.: Numerical exact controllability of the 1D heat equation: Carleman weights and duality (preprint, 2011). http://hal.archives-ouvertes.fr/hal-00687887
12. Fursikov, A.V., Imanuvilov, O.Yu.: Controllability of Evolution Equations, Lecture Notes Series, vol. 34, pp. 1–163. Seoul National University, Korea (1996)
13. Glowinski, R.: Finite element methods for incompressible viscous flow. Handbook of numerical analysis, Vol. IX, 3–1176. North-Holland, Amsterdam (2003)
14. Glowinski, R., Le Tallec, P.: Augmented Lagrangian and operator-splitting methods in nonlinear mechanics. SIAM Studies in Applied Mathematics vol. 9, Philadelphia (1989)
15. Glowinski, R., Lions, J.-L.: Exact and approximate controllability for distributed parameter systems. Acta Numer. 159–333 (1996)
16. Glowinski, R., Lions, J.-L., He, J.: Exact and approximate controllability for distributed parameter systems: a numerical approach, Encyclopedia of Mathematics and its Applications, vol. 117. Cambridge University Press, Cambridge (2008)
17. Imanuvilov, O.Yu.: Controllability of parabolic equations (Russian). Mat. Sb. **186**(6), 109–132 (1995) (translation in. Sb. Math. **186**(6), 879–900 (1995))
18. Kindermann, S.: Convergence rates of the Hilbert uniqueness method via Tikhonov regularization. J. Optim. Theory Appl. **103**(3), 657–673 (1999)
19. Labbé, S., Trélat, E.: Uniform controllability of semi-discrete approximations of parabolic control systems. Syst. Control Lett. **55**, 597–609 (2006)
20. Lasiecka, I., Triggiani, R.: Control theory for partial differential equations: continuous and approximation theories. I. Abstract parabolic systems, Encyclopedia of Mathematics and its Applications, vol. 74. Cambridge University Press, Cambridge (2000)
21. Lebeau, G., Robbiano, L.: Contrôle exact de l'équation de la chaleur. Comm. Partial Diff. Equ. **20**(1–2), 335–356 (1995)
22. Lions, J.-L.: Contrôlabilité exacte, perturbations et stabilisation de systèmes distribués, Recherches en Mathématiques Appliquées, Tomes 1 et 2. Masson, Paris (1988)
23. López, A., Zuazua, E.: 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, 22 p. Ecole Polytech. Palaiseau (1998)
24. Micu, S., Zuazua, E.: On the regularity of null-controls of the linear 1-d heat equation. C. R. Acad. Sci. Paris, Ser. I 349, 673–677 (2011)
25. Münch, A., Pedregal, P.: Numerical null controllability of the heat equation through a variational approach (Preprint, 2011). http://hal.archives-ouvertes.fr/hal-00724035
26. Münch, A., Periago, F.: Optimal distribution of the internal null control for the 1D heat equation. J. Diff. Equ. **250**, 95–111 (2011)
27. Münch, A., Zuazua, E.: Numerical approximation of null controls for the heat equation: ill-posedness and remedies. Inverse Problems **26**(8), 085018 (2010)
28. Russell, D.L.: A unified boundary controllability theory for hyperbolic and parabolic partial differential equations. Studies in Appl. Math. **52**, 189–221 (1973)
29. Russell, D.L.: Controllability and stabilizability theory for linear partial differential equations. Recent progress and open questions. SIAM Rev. **20**, 639–739 (1978)
30. Zhang, X.: Explicit observability inequalities for the wave equation with lower order terms by means of Carleman inequalities. SIAM J. Control. Optim. **39**, 812–834 (2000)
31. Zuazua, E.: Control and numerical approximation of the wave and heat equations. International Congress of Mathematicians, Madrid, Spain **III**, 1389–1417 (2006)