ANZIAM J. 57 (E) pp.E1–E38, 2016

Identification of the diffusion coefficient in a nonlinear parabolic PDE and numerical methods

Sameh Abidi[1] Jamil Satouri[2]

(Received 26 September 2014; revised 27 April 2016)

Abstract

This paper presents an iterative method to identify the diffusion in a semi-linear parabolic problem. This method can be generalized to other kind of problems, elliptic, parabolic and hyperbolic in two-dimensional and three-dimensional case. The diffusion is obtained by solving an optimal control problem. By imposing specific conditions to the data, we build a sequence of linear problems which converge to the exact solution. We discretize our problem by a finite element method in the first case and a spectral method in the second case, using the sensibility method for approximating the gradient of the functional. Some numerical experiments prove the efficiency of this method.
Keywords: Parabolic equation, diffusion, optimization, finite-elements, spectral method

http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/8835 gives this article, Published May 16, 2016. ISSN 1445-8810.

Contents

1 Introduction

In the phenomena of exchange at the sediment-water interface of the seabed or lagoons, the concentration $u(x,t)$ where x is the spatial coordinate system and t is the time, is governed by the one-dimensional diffusion equation

$$\frac{\partial u}{\partial t} - \nu\frac{\partial u}{\partial x} - \frac{\partial}{\partial x}\left(D\frac{\partial u}{\partial x}\right) = f(x,t,u),$$

where $D(x,t)$ is the diffusion coefficient that accounts both molecular diffusion and dispersion kinetics [4] and f is the production-destruction term.

Defining $Q_T = \Omega\times]0,T[$, for open interval $\Omega =]0,1[$ and T is a positive real number, χ_T is the concentration at the time T, and ν_0 is the concentration at time $t=0$. Neglecting the convective term, the problem in this paper is to find the coefficient of diffusion D minimizing the functional

$$J(D) = \frac{1}{2}\int_0^1 [u(x,T;D) - \chi_T(x)]^2 \, dx\,,$$

where for a given D, $u(x,t;D)$ is the solution of the problem

$$\begin{aligned} &\text{Find } u(x,t) : Q_T \to \mathbb{R}\,, \quad \text{such that} \\ &\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial u}{\partial x}\right) = f(x,t,u), \quad (x,t) \in Q_T\,, \\ &u(x,0) = \nu_0(x), \quad x \in \Omega\,, \\ &u(0,t) = u(1,t) = g(t), \quad t \in]0,T[\,, \end{aligned} \tag{$\mathcal{P}_D$}$$

where f, g and $\nu_0 > 0$ are given. Thus the problem

$$\min J(D) \quad \text{such that } D \in U_{ad}\,, \quad u(x,t;D) \text{ solves } (\mathcal{P}_D), \tag{Q_D}$$

is an optimization problem.

A similar problem in linear case was first studied by Bouchiba and Abidi [3]. However, here and elsewhere the formulation that is considered in this work deals with the nonlinear source term f. There are Several applications in this field, discussing parabolic and hyperbolic, linear and nonlinear systems describing numerical simulation results. These include control of distributed parameter systems [15, 16, 17], regularization for highly ill-posed distributed parameter estimation problems [14]. This performed numerical analysis based on an iterative Gauss–Newton method and solved by Levenberg–Marquardt-type method, which applied to the output least squares formulation.

In section 2, we study after linearizing, the existence and the uniqueness of a weak solution of the linearized problem of ($\mathcal{P}_n$) for a fixed D. In section 3, we prove the continuity of the map $\Phi : D \to u(x,t,D)$. In section 4, we establish the convergence of the sequence $(u_n)_{n\in\mathbb{N}}$ where u_n is solution of ($\mathcal{P}_n$) to u solution of ($\mathcal{P}_D$). In section 5, we describe the optimal control problem under constraints ($\mathcal{P}_D$) [10]. Indeed we prove the existence of a global optimal solution for the optimal control problem (Q) [11, 12]. In section 6, we give a numerical algorithm to solve ($\mathcal{P}_n$) by computing the gradient of the functional J [4], [3, linear version]. The numerical solution of the PDE system is generated by both the finite elements method [7] and the spectral method [8, 9]. Finally in Section 7, we present a comparative study between the two methods (finite elements method and spectral method) to see the best convergence.

2 The linearized problems

2.1 The Assumptions

Let O be an open set of $\mathbb{R}^d$, let $p \in [1,+\infty]$ and let k be a natural number. Define the Sobolev space

$$W^{k,p}(O) = \{u \in L^p(O) : \partial_\alpha u \in L^p(O) \text{ for all } |\alpha| \leqslant k\},$$

where α is a multi-index, and ∂_α is a partial derivative of u in the weak sense.

We suppose that the nonlinear source term f in the first equation of ($\mathcal{P}_D$), satisfies the following assumptions:

$$f \in \mathcal{C}^2\left(\bar{\Omega}\times]0,T[\times\mathbb{R}\right) \cap W^{2,\infty}\left(\bar{\Omega}\times]0,T[\times\mathbb{R}\right), \tag{H1}$$

$$\frac{\partial f}{\partial s}(x,t,s) \leqslant 0, \quad \text{for all } s \in \mathbb{R} \text{ and } (x,t) \in \Omega\times]0,T[, \tag{H2}$$

$$s \mapsto f(\cdot,\cdot,s) \text{ is concave on } \mathbb{R}, \tag{H3}$$

$$f(x,t,0) > 0\,. \tag{H4}$$

2.2 Linearized problems

Using the first equation of ($\mathcal{P}_D$), we denote the residual

$$R(\psi) = \frac{\partial \psi}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial \psi}{\partial x}\right) - f(\cdot,\cdot,\psi),$$

so that u is a solution of $R(u) = 0$.

By applying Newton's method linearized which has the advantage to converge rapidly towards our solution, we build a recurrent sequence of functions $(u_n)_n$ with $u_n : Q_T \to \mathbb{R}$, for $n \in \mathbb{N}$,

$$u_n = u_{n-1} - \frac{R(u_n)}{R'(u_n)}.$$

When u_0 is the first term, then

$$\frac{\partial u_n}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial u_n}{\partial x}\right) - \frac{\partial f}{\partial u}(\cdot,\cdot,u_{n-1})u_n = -\frac{\partial f}{\partial u}(\cdot,\cdot,u_{n-1})u_{n-1} + f(\cdot,\cdot,u_{n-1}).$$

We obtain a sequence of linear problems: initially

$$\begin{aligned}
&\frac{\partial u_0}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial u_0}{\partial x}\right) = k, \quad \text{in } Q_T\\
&u_0(x,0) = v_0, \quad x \in \Omega,\\
&u_0(0,t) = u_0(1,t) = g(t), \quad t \in]0,T[,
\end{aligned} \tag{$\mathcal{P}_0$}$$

where $k = \|f\|_\infty$. For $n \geqslant 1$,

$$\begin{aligned}
&\frac{\partial u_n}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial u_n}{\partial x}\right) - \frac{\partial f}{\partial u}(x,t,u_{n-1})u_n = F(x,t,u_{n-1}), \quad \text{in } Q_T,\\
&u_n(x,0) = v_0, \quad x \in \Omega,\\
&u_n(0,t) = u_n(1,t) = g(t), \quad t \in]0,T[,
\end{aligned} \tag{$\mathcal{P}_n$}$$

where

$$F(x,t,u_{n-1}) = f(x,t,u_{n-1}) - \frac{\partial f}{\partial u}(x,t,u_{n-1})u_{n-1}.$$

2.3 The variational Formulation of $(\mathcal{P}_n)$

Let O be an open set of $\mathbb{R}^d$ and let k be a natural number. Define the Sobolev space

$$H^k(O) = \left\{u \in L^2(O) : \partial_\alpha u \in L^2(O) \text{ for all } \alpha \in \mathbb{N}^d,\ |\alpha| \leqslant k\right\}$$

For a fixed D, the variational formulation of problem $(\mathcal{P}_n)$ is

$$\begin{aligned}
&\text{Find } u_n(\cdot,t) \in H^1(\Omega), \quad \text{with } u_n(\cdot,t) - \bar{g}(\cdot,t) \in H_0^1(\Omega), \quad \text{such that} \\
&\frac{d}{dt}(u_n(\cdot,t),v) + a_D(u_n(\cdot,t),v) = (F(\cdot,t,u_{n-1}),v), \text{ for all } v \in H_0^1(\Omega), \\
&u_n(\cdot,0) = v_0,
\end{aligned} \tag{$\mathcal{P}V_n$}$$

where $u_n(\cdot,t) : \Omega \to \mathbb{R}$ and $\bar{g}(\cdot,t)$ is the lifting of the boundary condition $g(t)$ in $H^1(\Omega)$ and coincides with $g(t)$ on $\{0,1\}$, the boundary condition is reflected by the fact that $u_n(\cdot,t) - \bar{g}(\cdot,t) \in H_0^1(\Omega)$ [9, Proposition 1.12].

The bilinear form, for all $v \in H_0^1(\Omega)$,

$$a_D(u_n,v) = \int_0^1 D\frac{\partial u_n(\cdot,t)}{\partial x}\frac{\partial v}{\partial x}\,dx - \int_0^1 \frac{\partial f}{\partial u}(\cdot,t,u_{n-1})u_n(\cdot,t)v\,dx\,.$$

So for $F(\cdot,\cdot,u_{n-1}) \in L^2(0,T,L^2(\Omega))$ and $\bar{g}(\cdot,t)$ continuous in $\bar{\Omega}$, $v_0 \in L^2(\Omega)$ and $D \in L^\infty(Q_T)$ such that $D \geqslant \alpha > 0$, problem $(\mathcal{P}V_n)$ admits a unique solution [1, 2],

$$u_n \in L^2\left(0,T,H^1(\Omega)\right) \cap C^0\left(0,T,L^2(\Omega)\right).$$

3 Continuity of the map Φ

The following Lemma is needed to prove the continuity of the map Φ.

Lemma 1. *If function u is a solution of problem $(\mathcal{P}_D)$, then u is bounded in $L^2(0,T,H_0^1(\Omega))$.*

Proof: The variational formulation of problem ($\mathcal{P}_D$) is

$$\begin{aligned} &\text{Find } u(\cdot,t) \in H^1(\Omega); u(\cdot,t) - \bar{g}(\cdot,t) \in H_0^1(\Omega) \quad \text{such that} \\ &\frac{d}{dt}(u(\cdot,t),v) + \tilde{a}_D(u(\cdot,t),v) = (f(\cdot,t,u),v), \quad \text{for all } v \in H_0^1(\Omega), \quad (\mathcal{P}V_D) \\ &u(\cdot,0) = v_0\,, \end{aligned}$$

where $\tilde{a}_D(u(\cdot,t),v) = \int_0^1 D\frac{\partial u(\cdot,t)}{\partial x}\frac{\partial v}{\partial x}\,dx$. We denote $\bar{u}(\cdot,t) = u(\cdot,t) - \bar{g}(\cdot,t)$. Then $u(\cdot,t)$ is solution of problem ($\mathcal{P}V_D$) if and only if $\bar{u}(\cdot,t)$ is solution of the problem

$$\begin{aligned} &\text{Find } \bar{u}(\cdot,t) \in H_0^1(\Omega) \quad \text{such that, for all } v \in H_0^1(\Omega), \\ &\frac{d}{dt}(\bar{u}(\cdot,t),v) + \tilde{a}_D(\bar{u}(\cdot,t),v) = (f(\cdot,t,\bar{u}+\bar{g}(t)),v) - \tilde{a}_D(\bar{g}(\cdot,t),v), \\ &u(\cdot,0) = v_0\,. \end{aligned} \tag{1}$$

Using (1) in the particular case for $v = \bar{u}(\cdot,t) \in H_0^1(\Omega)$, such that $D \geqslant \alpha > 0$ and the Cauchy–Schwarz inequality combined with Poincare inequality we prove that u is bounded in $L^2\left(0,T,H_0^1(\Omega)\right)$. ♠

Let the map

$$\begin{aligned} \Phi : L^\infty(Q) &\to L^2\left(0,T,H_0^1(\Omega)\right) \\ D &\to u \end{aligned}$$

where u is the unique solution of ($\mathcal{P}_D$). Let

$$U = \{\vartheta \in L^\infty(Q_T) : \text{there exists } \alpha \backslash \vartheta \geqslant \alpha > 0\}.$$

For $D_1, D_2 \in U$, $D_1 \neq D_2$, we denote $u_i = \Phi(D_i)_{i=1,2}$ and $f_i(x,t) = f(x,t,u_i)_{i=1,2}$, for all $(x,t) \in Q_T$.

Proposition 2. *The map Φ is continuous*

$$\| u_1 - u_2 \|_{L^2\left(0,T,H_0^1(\Omega)\right)} \leqslant C \| D_1 - D_2 \|_{L^\infty(Q_T)}, \tag{2}$$

where C is a constant independent of D_1 and D_2.

Proof: We denote $\bar{u}^* = \bar{u}_1 - \bar{u}_2$. Using (2), after subtracting for all $v \in H^1_0(\Omega)$,

$$\left(\frac{\partial \bar{u}^*(\cdot,t)}{\partial t}, v\right) + \int_0^1 (D_1 - D_2)\,\frac{\partial \bar{u}_1(\cdot,t)}{\partial x}\,\frac{\partial v}{\partial x}\,dx + \int_0^1 D_2 \frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\,\frac{\partial v}{\partial x}\,dx$$
$$= \int_0^1 \left(\bar{f}_1(\cdot,t) - \bar{f}_2(\cdot,t)\right) v\,dx\,. \tag{3}$$

In particular, equation (3) is true for $v = \bar{u}^*(\cdot,t) \in H^1_0(\Omega)$ and after integration between 0 and t

$$\frac{1}{2} \parallel \bar{u}^*(\cdot,t) \parallel^2_{L^2(\Omega)} + \int_0^t \left(\int_0^1 D_2 \left|\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\right|^2 dx\right) d\sigma$$
$$= \int_0^t \left[\int_0^1 \left(\bar{f}_1(\cdot,t) - \bar{f}_2(\cdot,t)\right) \bar{u}^*(\cdot,t)\,dx\right] d\sigma$$
$$+ \int_0^t \left[\int_0^1 (D_2 - D_1)\frac{\partial \bar{u}_1(\cdot,t)}{\partial x}\,\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\,dx\right] d\sigma\,.$$

We denote

$$A = \frac{1}{2} \parallel \bar{u}^*(\cdot,t) \parallel^2_{L^2(\Omega)}, \quad B = \int_0^t \left(\int_0^1 D_2 \left|\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\right|^2 dx\right) d\sigma\,.$$

Then, as $D \geqslant \alpha > 0$,

$$\int_0^t \left(\int_0^1 \left|\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\right|^2 dx\right) d\sigma \leqslant \frac{1}{\alpha}\int_0^t \left[\int_0^1 \left(\bar{f}_1(\cdot,t) - \bar{f}_2(\cdot,t)\right) \bar{u}^*(\cdot,t)\,dx\right] d\sigma$$
$$+ \frac{1}{\alpha}\int_0^t \left[\int_0^1 (D_2 - D_1)\,\frac{\partial \bar{u}_1(\cdot,t)}{\partial x}\,\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\,dx\right] d\sigma\,. \tag{4}$$

And also,

$$\frac{1}{2}\parallel \bar{u}^*(\cdot,t)\parallel^2_{L^2(\Omega)} \leqslant \int_0^t \left[\int_0^1 \left(\bar{f}_1(\cdot,t)-\bar{f}_2(\cdot,t)\right)\bar{u}^*(\cdot,t)\,dx\right]d\sigma + \int_0^t \left[\int_0^1 (D_2-D_1)\,\frac{\partial \bar{u}_1(\cdot,t)}{\partial x}\,\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\,dx\right]d\sigma\,.$$

The inequality (4) is true for all $t\in]0,T[$. Recall the assumption (H1) and using Taylor's formula, we obtain the following inequality which is related to B defined above

$$\left\|\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\right\|^2_{L^2(0,T,L^2(\Omega))} \leqslant \frac{1}{\alpha}\|D_2-D_1\|_{L^\infty(Q_T)}\left\|\frac{\partial \bar{u}_1(\cdot,t)}{\partial x}\right\|_{L^2(0,T,L^2(\Omega))}\left\|\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\right\|_{L^2(0,T,L^2(\Omega))} + \frac{c}{\alpha}\|D_2-D_1\|_{L^\infty(Q_T)}\|\bar{u}^*\|_{L^2(0,T,L^2(\Omega))}$$

By Lemma 1,

$$\left\|\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\right\|^2_{L^2(0,T,L^2(\Omega))} \leqslant \frac{1}{\alpha}c_1\|D_2-D_1\|_{L^\infty(Q_T)}\left\|\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\right\|_{L^2(0,T,L^2(\Omega))} + \frac{c}{\alpha}\|D_2-D_1\|_{L^\infty(Q_T)}\|\bar{u}^*(\cdot,t)\|_{L^2(0,T,L^2(\Omega))}\,.$$

We obtain

$$\left\|\frac{\partial \bar{u}^*(\cdot,t)}{\partial x}\right\|^2_{L^2(0,T,L^2(\Omega))} \leqslant C_0\|D_2-D_1\|_{L^\infty(Q_T)}\|\bar{u}^*(\cdot,t)\|_{L^2\left(0,T,H^1_0(\Omega)\right)}\,.$$

Similarly to (4), the following inequality in terms of the defined variable A,

$$\frac{1}{2}\|\bar{u}^*(\cdot,t)\|^2_{L^2(\Omega)} \leqslant C_1\|D_2-D_1\|_{L^\infty(Q_T)}\|\bar{u}^*(\cdot,t)\|_{L^2\left(0,T,H^1_0(\Omega)\right)}\,.$$

After integration between 0 and T,

$$\|\bar{u}^*(\cdot,t)\|^2_{L^2(0,T,L^2(\Omega))} \leqslant 2TC_1 \|D_2 - D_1\|_{L^\infty(Q_T)} \|\bar{u}^*(\cdot,t)\|_{L^2(0,T,H^1_0(\Omega))} .$$

We set $C = 2\sup(C_0, 2TC_1)$, and conclude that

$$\|\bar{u}^*(\cdot,t)\|_{L^2(0,T,H^1_0(\Omega))} \leqslant C \|D_2 - D_1\|_{L^\infty(Q_T)} ,$$

and finally

$$\|u_1 - u_2\|_{L^2(0,T,H^1_0(\Omega))} \leqslant C \|D_1 - D_2\|_{L^\infty(Q_T)} .$$

♠

4 The convergence

Remark 3.

- The symbol $X \hookrightarrow Y$ denotes the continuous and dense embedding of X into Y.
- The symbol $X \rightharpoonup Y$ denotes the weak convergence of X to Y.
- The symbol $X \rightarrow Y$ denotes the strong convergence of X to Y.

Consider now the sequence $(u_n)_{n\in\mathbb{N}}$ where u_n is solution of $(\mathcal{P}_n)$.

Proposition 4. *The sequence* $\{u_n\}_{n=0}^\infty$ *satisfies*

$$0 < \cdots \leqslant u_{n+1} \leqslant u_n \leqslant \cdots \leqslant u_0 .$$

Proof:

1. We first prove that $u_{n+1} - u_n \leqslant 0$. Let $\omega_0(x,t) = u_1(x,t) - u_0(x,t)$, for all $(x,t) \in Q_T$, with u_0 is the solution of ($\mathcal{P}_0$) and u_1 is the solution of $\mathcal{P}_1$, ($\mathcal{P}_n$) for $n = 1$, then

$$\frac{\partial u_1}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial u_1}{\partial x}\right) - \frac{\partial f}{\partial u}(\cdot,\cdot,u_0)u_1 = F(\cdot,\cdot,u_0),$$
$$\frac{\partial u_0}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial u_0}{\partial x}\right) = k\,.$$

After subtracting ($\mathcal{P}_0$) and $\mathcal{P}_1$, ($\mathcal{P}_n$) for $n = 1$, we get

$$\frac{\partial \omega_0}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial \omega_0}{\partial x}\right) - \frac{\partial f}{\partial u}(\cdot,,,u_0)\omega_0 = f(\cdot,\cdot,u_0) - k \leqslant 0\,.$$

Using the assumption (H2), the difference field $\omega_0 \leqslant 0$ via the maximum principle [1]. Subtracting $\mathcal{P}_{n+1}$ and ($\mathcal{P}_n$), setting $\omega_n = u_{n+1} - u_n$, with u_n is the solution of ($\mathcal{P}_n$) and u_{n+1} is the solution of $\mathcal{P}_{n+1}$, ($\mathcal{P}_n$) for $n \mapsto n+1$,

$$\frac{\partial \omega_n}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial \omega_n}{\partial x}\right) - \frac{\partial f}{\partial u}(\cdot,\cdot,u_n)\omega_n = G_n\,,$$

where via (H3), $G_n = f(\cdot,\cdot,u_n) - f(\cdot,\cdot,u_{n-1}) - \frac{\partial f}{\partial u}(\cdot,\cdot,u_{n-1})\omega_{n-1} \leqslant 0$. Then, using the assumption (H2) the maximum principle gives $\omega_n \leqslant 0$, for all $n \geqslant 1$.

2. We prove now that $u_n > 0$ for all n. In problem ($\mathcal{P}_0$), $k = \|f\|_\infty > 0$ and $v_0 > 0$ by the maximum principle $u_0 > 0$. In problem $\mathcal{P}_{n+1}$, ($\mathcal{P}_n$) for $n \mapsto n+1$, under assumption (H3) and (H4) $F(\cdot,\cdot,u_n) > 0$. Using the assumption (H2) and $v_0 > 0$, then via the maximum principle $u_n > 0$ for all $n \geqslant 1$.

♠

Theorem 5. *The sequence $(u_n)_{n\in\mathbb{N}}$ where u_n is solution of $(\mathcal{P}_n)$ converges to u, where u is a solution of $(\mathcal{P}_D)$.*

Proof: By Lemma 1, u_n is bounded in $L^2\left(0, T, H_0^1(\Omega)\right)$ then, we can extract a subsequence still denoted u_n such that

$$u_n \rightharpoonup u \quad \text{in } L^2\left(0, T, H_0^1(\Omega)\right) \quad \text{weakly.} \tag{5}$$

Using (H4), we prove that $u > 0$. But it appears difficult because from (5), it is generally not guaranteed that

$$\frac{\partial}{\partial x}\left(D\frac{\partial u_n}{\partial x}\right) \to \frac{\partial}{\partial x}\left(D\frac{\partial u}{\partial x}\right).$$

Then, we increase the regularity of u using the singular perturbation method. For that we introduce a new bilinear form, continuous on $H_0^2(\Omega)$,

$$(\varphi, \psi) \mapsto b(\varphi, \psi),$$

such that

$$b(\psi, \psi) \geqslant \beta\|\psi\|^2_{H_0^2(\Omega)}, \quad \text{for all } \psi \in H_0^2(\Omega).$$

Using Riez's Theorem, we represent the bilinear form b by an operator B such that

$$b(\varphi, \psi) = \langle B\varphi, \psi\rangle_{H^{-2}(\Omega), H_0^2(\Omega)}.$$

Similarly for the bilinear form a_D: it is represented by the operator $A(D)$ such that

$$a_D(\varphi, \psi) = \langle A(D)\varphi, \psi\rangle_{H^{-1}(\Omega), H_0^1(\Omega)}.$$

For ε sufficiently small and positive we denote

$$a_{\varepsilon,D}(\varphi, \psi) = \varepsilon b(\varphi, \psi) + a_D(\varphi, \psi), \quad \text{for all } \varphi, \psi \in H_0^2(\Omega).$$

We verify that

$$a_{\varepsilon,D}(\psi, \psi) \geqslant \varepsilon\|\psi\|^2_{H_0^2(\Omega)} + c\|\psi\|^2_{H_0^1(\Omega)},$$

where c is a positive constant. The singular perturbation problem is

$$\frac{\partial u_{n,\varepsilon}}{\partial t} + (\varepsilon B + A(D))u_{n,\varepsilon} = F(\cdot,\cdot,u_{n-1}) \quad \text{in } Q_T,$$
$$u_{n,\varepsilon}(x,0) = v_0(x) \quad \text{in } \Omega, \tag{$\mathcal{P}_{n,\varepsilon}$}$$

where $u_{n,\varepsilon} \in L^2\left(0,T,H_0^2(\Omega)\right)$. The operator $\frac{\partial}{\partial t} + (\varepsilon B + A(D))$ is the singular perturbation operator of $\frac{\partial}{\partial t} + A(D)$. By applying the theorem of Lions [2] such that $u_{n,\varepsilon} \in H_0^2(\Omega)$ and $a_D(u_n, v)$ is replaced by $a_{\varepsilon,D}(u_n, v)$ and using the assumption of uniform coercivity,

$$a_{\varepsilon,D}(\psi,\psi) \geqslant \varepsilon\|\psi\|^2_{H_0^2(\Omega)} + c\|\psi\|^2_{H_0^1(\Omega)}, \tag{6}$$

problem ($\mathcal{P}_{n,\varepsilon}$) admits a unique solution $u_{n,\varepsilon}$.

Proposition 6. *As $\varepsilon \to 0$ the following strong convergences hold:*

- $u_{n,\varepsilon} \to u$ *in* $L^2\left(0,T,H_0^1(\Omega)\right)$*;*
- $\sqrt{\varepsilon}u_{n,\varepsilon} \to 0$ *in* $L^2\left(0,T,H_0^2(\Omega)\right)$*;*
- $\frac{\partial}{\partial t}u_{n,\varepsilon} \to \frac{\partial}{\partial t}u$ *in* $L^2\left(0,T,H_0^{-2}(\Omega)\right)$*.*

Since ε is fixed positive, we deduce from (6) that

$$u_n \text{ is bounded in } L^2\left(0,T,H_0^2(\Omega)\right). \tag{7}$$

From the first equation of ($\mathcal{P}_{n,\varepsilon}$), we prove that

$$\frac{\partial}{\partial t}u_{n,\varepsilon}\text{is bounded in } L^2\left(0,T,H_0^{-2}(\Omega)\right). \tag{8}$$

From (7), we extract a sequence, still denoted by u_n, such that u_n converge in a weak sense to u in $L^2\left(0,T,H_0^2(\Omega)\right)$. Since the injection of $H_0^2(\Omega)$ in $H_0^1(\Omega)$ is compact, there exists a subsequence, still denoted u_n, which converge strongly to u in $L^2\left(0,T,H_0^1(\Omega)\right)$.

From (8), we conclude that $\frac{\partial u_n}{\partial t}$ converge weakly to $\frac{\partial u}{\partial t}$ in $L^2\left(0, T, H_0^{-2}(\Omega)\right)$, then

$$D\frac{\partial u_n}{\partial x} \to D\frac{\partial u}{\partial x} \quad \text{in } Q_T,$$

in a weak sense. Indeed, for all $\psi \in D(Q_T)$, $X = (x, t) \in Q_T$,

$$\int_{Q_T} D\frac{\partial u_n}{\partial x}\psi \, dX = \int_{Q_T} D\frac{\partial u}{\partial x}\psi \, dX + \int_{Q_T} D\left(\frac{\partial u_n}{\partial x} - \frac{\partial u}{\partial x}\right)\psi \, dX .$$

We conclude that

$$\int_{Q_T} D\frac{\partial u_n}{\partial x}\psi \, dX \to \int_{Q_T} D\frac{\partial u}{\partial x}\psi \, dX ,$$

because

$$\left|\int_{Q_T} D\left(\frac{\partial u_n}{\partial x} - \frac{\partial u}{\partial x}\right)\psi \, dX\right| \leqslant C\left\|\frac{\partial u_n}{\partial x} - \frac{\partial u}{\partial x}\right\|_{L^2(Q_T)} \|\psi\|_{L^2(Q_T)},$$

and

$$\left\|\frac{\partial u_n}{\partial x} - \frac{\partial u}{\partial x}\right\|_{L^2(Q_T)} \to 0 .$$

Since $f(Q_T)$ is C^2, then

$$\frac{\partial f}{\partial u}(X, u_n) \rightharpoonup \frac{\partial f}{\partial u}(X, u),$$

and

$$f(X, u_{n-1}) - \frac{\partial f}{\partial u}(X, u_{n-1})u_{n-1} \rightharpoonup f(X, u) - \frac{\partial f}{\partial u}(X, u)u .$$

We conclude that

$$\frac{\partial}{\partial t}u_n - \frac{\partial}{\partial x}\left(D\frac{\partial}{\partial x}u_n\right) + \frac{\partial f}{\partial u}(X, u_n)u_n \rightharpoonup \frac{\partial}{\partial t}u - \frac{\partial}{\partial x}\left(D\frac{\partial}{\partial x}u\right) + \frac{\partial f}{\partial u}(X, u)u ,$$

weakly in Q_T. Therefore u_n satisfies the two equations of $(\mathcal{P}_{n,\varepsilon})$. ♠

5 Optimal control

In this section we are interested by the optimal control problem under constraints ($\mathcal{P}_D$) [10]

$$\begin{aligned} &\min \quad J(D) = \frac{1}{2}\int_0^1 [u(x,T;D) - \chi_T(x)]^2 \, dx\,, \quad \text{such that} \\ &\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial u}{\partial x}\right) = f(x,t,u), \quad (x,t) \in Q_T\,, \\ &u(x,0) = v_0(x), \quad x \in \Omega\,, \\ &u(0,t) = u(1,t) = g(t), \quad t \in]0,T[\,, \\ &u \in \mathbf{C}\,. \end{aligned} \tag{Q}$$

The set $\mathbf{C}$ is a closed convex subset of $\mathbf{C}_0(\Omega) = \{\omega \in \mathbf{C}(\bar{\Omega}) : \omega = 0 \text{ on } \{0,1\}\}$, the space of continuous functions on Ω vanishing on $\{0,1\}$.

Remark 7. Sometimes we adopt the notation $u(x,t;D)$ as a reminder that u implicitly depends on D.

The set of admissible solutions is defined as

$$U_{ad} = \{D \in L^\infty(Q_T) : u(x,t;D) \text{ satisfies the state equation in } (\mathcal{P}_D)\}.$$

The function $u(x,T;D)$, for all $(x,t) \in Q_T$, is the solution of problem ($\mathcal{P}_D$) at $t = T$, χ_T is given in $L^2(\Omega)$. We now prove the main result of this section.

Theorem 8. *If U_{ad} is non-empty, then there exists a global optimal solution for the optimal control problem (Q).*

Proof: Because U_{ad} is not empty, we may take a minimizing sequence $D_n \in U_{ad}$. We obtain that $\| D_n \|_\infty < \infty$ which implies that D_n is uniformly bounded in $L^\infty(Q_T)$. Then we may extract a weakly convergent subsequence, also denoted by D_n, which converge in a weak-star sense to $\bar{D} \in L^\infty(Q_T)$. And we denote $u_n(x,t) = \bar{u}(x,t;D_n)$, for all $(x,t) \in Q_T$, by (1) and using

Lemma 1 combined with (H1), u_n is uniformly bounded in $H_0^1(\Omega)$. Then we extract a weakly convergent subsequence, also denoted by u_n, such that $u_n \rightharpoonup \bar{u}^0 \in H_0^1(\Omega)$. In order to see that $(\bar{u}^0; \bar{D})$ is a solution of the $(\mathcal{P}_D)$ equations, the only problem is to pass to the limit in the nonlinear form $a_n = \int_0^1 f(\cdot, t, u_n) v \, dx$ of (1). Due to the compact embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$ [2] and the continuity of the a_n, and without forgetting that a_n is bounded for all x, t and u_n, it follows that

$$\int_0^1 f(\cdot, t, u_n) v \, dx \to \int_0^1 f(\cdot, t, u) v \, dx \,.$$

Consequently, taking into account the linearity and continuity of all terms involved, the limit $(\bar{u}^0; \bar{D})$ satisfies the state equations. Since $\mathbf{C}$ is convex and closed, it is weakly closed, so $u_n \rightharpoonup \bar{u}^0 \in H_0^1(\Omega)$ and the embedding $H_0^1(\Omega) \hookrightarrow C_0(\Omega)$ imply that $\bar{u}^0 \in \mathbf{C}$. Taking into consideration that $J(D)$ is weakly lower semi continuous, then there exists a global optimal solution for the optimal control problem (Q). ♠

6 The discrete problem

6.1 Finite element method

In this section, to simplify, we study the problem $(\mathcal{P}V_n)$ in the case of homogeneous boundary conditions $(g = 0)$.

6.1.1 The variational Problem

We put $H = L^2(\Omega)$, $V = H_0^1(\Omega)$, $\Omega =]0,1[$. Then $V \subset H \subset V'$. We consider the variational problem $(\mathcal{P}V_n)$ which admits a unique solution

$$u_n \in L^2\left(0, T, H_0^1(\Omega)\right) \cap C^0([0,T], L^2(\Omega)).$$

To build an approximation $u_{n,h}$ of u_n, we choose a subspace of V composed of affine functions over an equal number of intervals of equal size. More precisely, let N be a natural number and $h = \frac{1}{N+1}$, $0 \leqslant i \leqslant N+1$. We denote $x_i = ih$, $0 \leqslant i \leqslant N+1$, that subdivides the interval $\bar{\Omega} = [0,1]$ to $N+1$ intervals $K_i = [x_i, x_{i+1}]$, $1 \leqslant i \leqslant N$, each of length h and

$$V_h = \{v \in C^0(\bar{\Omega}) : v(0) = v(1) = 0\,,\ v_{|K_i} \in P_1\,,\ 0 \leqslant i \leqslant N\},$$

a finite dimensional subspace of V [1].

Setting $F_{n-1}(t) = F(\cdot, t, u_{n-1})$, the approached problem is

$$\begin{aligned}
&\text{Find } u_{n,h}(\cdot,t) \in V_h\,,\ \text{such that for all } t \in [0,T],\\
&\frac{d}{dt}\langle u_{n,h}(\cdot,t), v_h\rangle + a_D(u_{n,h}(\cdot,t), v_h) = \langle F_{n-1}(t), v_h\rangle \quad \text{for all } v_h \in V_h\,,\\
&u_{n,h}(\cdot,0) = v_{0,h}\,.
\end{aligned} \tag{$\mathcal{P}V_{nh}$}$$

The problem ($\mathcal{P}V_{nh}$) admits an unique solution $u_{n,h}(\cdot,t) \in V_h$: the proof is obtained by direct application of Lions's Theorem [5]. We introduce the basis $(\varphi_i)_{1\leqslant i\leqslant N}$ of V_h, for $1 \leqslant j \leqslant N$. Then there exists $u^D_{n,0,j}$ and $u^D_{n,j}$ functions of $[0,T]$ in $\mathbb{R}$, such that, for all $(x,t) \in Q_T$,

$$v_{0,h}(x,t) = \sum_{j=1}^{N} u^D_{n,0,j}(t)\varphi_j(x), \quad u_{n,h}(x,t) = \sum_{j=1}^{N} u^D_{n,j}(t)\varphi_j(x).$$

Then problem ($\mathcal{P}V_{nh}$) is equivalent to the differential system

$$\begin{aligned}
&\sum_{j=1}^{N}(\varphi_j,\varphi_i)\frac{d}{dt}u^D_{n,j}(t) + \sum_{j=1}^{N} a_D(\varphi_j,\varphi_i)u^D_{n,j}(t) = F_{n-1,i}(t), \quad 1 \leqslant i \leqslant N\,,\\
&u^D_{n,i}(0) = u^D_{n,0,i}, \quad 1 \leqslant i \leqslant N\,,
\end{aligned} \tag{S_1}$$

where $F_{n-1,i}(t) = (F_{n-1}(t), \varphi_i)$. We denote

$$\begin{aligned}
R_D(t) &= (a_D(\varphi_j,\varphi_i))_{1\leqslant i,j\leqslant N} \text{ the stiffness matrix,}\\
M &= (\varphi_j,\varphi_i)_{1\leqslant i,j\leqslant N} \text{ the mass matrix,}\\
u^{*D}_n(t) &= (u^D_{n,1}(t), u^D_{n,2}(t), \ldots, u^D_{n,N}(t))^T, \quad \text{and}\\
\mathcal{F}_{n-1}(t) &= (F_{n-1,1}(t), F_{n-1,2}(t), \ldots, F_{n-1,N}(t))^T.
\end{aligned}$$

The system (S_1) is written as

$$M\frac{d}{dt}u_n^{*D}(t) + R_D(t)u_n^{*D}(t) = \mathcal{F}_{n-1}(t),$$
$$u_n^{*D}(0) = u_{n,0}^D\,. \tag{S_2}$$

A simple calculation gives the elements of M as

$$(\varphi_i,\varphi_i) = \frac{2h}{3}\,,\quad 1\leqslant i\leqslant N\,,$$
$$(\varphi_i,\varphi_{i+1}) = \frac{h}{6}\,,$$
$$(\varphi_{i+1},\varphi_i) = \frac{h}{6}\,,\quad 1\leqslant i\leqslant N\,,$$
$$(\varphi_i,\varphi_j) = 0\,,\quad |i-j|>1\,.$$

We put

$$D_{i+\frac{1}{2}} = \int_{x_i}^{x_{i+1}} D(t)\,dx\,.$$

Setting $K_{n-1}(t) = \int_0^1 \frac{\partial f}{\partial u}(\cdot,t,u_{n-1})u_n(\cdot,t)\nu\,dx$, then

$$K^-_{n-1,i+\frac{1}{2}} = \int_{x_i}^{x_{i+1}} K_{n-1}(t)(x_{i+1}-x)^2\,dx\,,$$
$$K^+_{n-1,i-\frac{1}{2}} = \int_{x_{i-1}}^{x_i} K_{n-1}(t)(x-x_{i-1})^2\,dx\,,$$
$$K_{n-1,i+\frac{1}{2}} = \int_{x_i}^{x_{i+1}} K_{n-1}(t)(x_{i+1}-x)(x-x_i)\,dx\,.$$

We obtain for the coefficients of R_D:

$$a_D(\varphi_i,\varphi_i) = \frac{1}{h^2}(D_{i+\frac{1}{2}} + D_{i-\frac{1}{2}} + K^-_{n-1,i+\frac{1}{2}} + K^+_{n-1,i-\frac{1}{2}}),\quad 1\leqslant i\leqslant N\,,$$
$$a_D(\varphi_{i+1},\varphi_i) = -\frac{1}{h^2}(D_{i+\frac{1}{2}} + K_{n-1,i+\frac{1}{2}}),$$
$$a_D(\varphi_i,\varphi_{i+1}) = -\frac{1}{h^2}(D_{i+\frac{1}{2}} + K_{n-1,i+\frac{1}{2}}),\quad 1\leqslant i\leqslant N-1\,,$$
$$a_D(\varphi_i,\varphi_j) = 0\,,\quad |i-j|>1\,. \tag{9}$$

Using the modified Euler method of integration,

$$\begin{aligned}\int_0^1 K_{n-1}(t)\varphi_i^2\,dx &= \frac{1}{h^2}\left(K^+_{n-1,i-\frac{1}{2}} + K^-_{n-1,i+\frac{1}{2}}\right)\\ &= \frac{2h}{3}\frac{\partial f}{\partial u}(x_{i-1}+h,t,u^D_{n-1,i}) + \frac{\partial f}{\partial u}(x_{i+1}+h,t,u^D_{n-1,i+2})\\ &\quad + \frac{h}{6}\frac{\partial f}{\partial u}\left(x_{i-1}+\frac{h}{2},t,\frac{u^D_{n-1,i-1}+u^D_{n-1,i}}{2}\right)\\ &\quad + \frac{\partial f}{\partial u}\left(x_{i+1}+\frac{h}{2},t,\frac{u^D_{n-1,i+1}+u^D_{n-1,i+2}}{2}\right),\end{aligned} \tag{10}$$

and

$$\int_0^1 K_{n-1}(t)\varphi_i\varphi_{i+1}\,dx = \frac{h}{24}\frac{\partial f}{\partial u}\left(x_i+\frac{h}{2},t,\frac{u^D_{n-1,i}+u^D_{n-1,i+1}}{2}\right). \tag{11}$$

The system (S_2) is solved using a fourth order Runge–Kutta method.

6.1.2 Calculating the gradient of J_n

Consider the problem

$$\min J_n(D),\quad D\in U_{ad},\quad u_n(\cdot,.;D)\text{ solves }(\mathcal{P}_n), \tag{Q_n}$$

where we define the functional

$$J_n(D) = \frac{1}{2}\int_0^1 [u_n(x,T;D) - \chi_{T,n}(x)]^2\,dx\,. \tag{12}$$

As $\chi_{T,n} = \sum_{i=1}^N \chi_{n,T}(x_i)\varphi_i$, we write

$$J_n(D) = \frac{1}{2}\int_0^1 \left[\sum_{i=1}^N \left(u^D_{n,i}(T) - \chi_{n,T}(x_i)\right)\varphi_i(x)\right]^2\,dx\,. \tag{13}$$

Expand the expression of J,

$$J_n(D) = \frac{h}{3}\sum_{i=1}^{N}\left(u^D_{n,i}(T) - \chi_{n,T}(x_i)\right)^2 + \frac{h}{6}\sum_{i=1}^{N-1}\left(u^D_{n,i+1}(T) - \chi_{n,T}(x_{i+1})\right)\left(u^D_{n,i}(T) - \chi_{n,T}(x_i)\right).$$

To calculate an approximation of the gradient J, we consider a mesh of the rectangle $\bar{\Omega} \times [0, T]$ by triangles T_{ij} with vertices

$$(ih, j\delta t),\ ((i+1)h, j\delta t),\ (ih, (j+1)\delta t),$$

and T'_{ij} with vertices

$$(ih, (j+1)\delta t),\ (ih, (j+1)\delta t),\ ((i+1)h, j\delta t),$$

where $\delta t = \frac{T}{M+1}$ is a time step. There $D_{i,j} \approx D(ih, j\delta t)$.

And we denote λ_i respectively (λ'_i), $i = 1, 2, 3$, the barycentric coordinates of a point from the triangle T_{ij}, respectively T'_{ij},

$$\begin{aligned} D_{ij|T_{ij}} &= \lambda_1 D_{ij} + \lambda_2 D_{i+1,j} + \lambda_3 D_{i,j+1}\,, \\ D_{ij|T'_{ij}} &= \lambda'_1 D_{i+1,j} + \lambda'_2 D_{i,j+1} + \lambda'_3 D_{i+1,j}\,. \end{aligned}$$

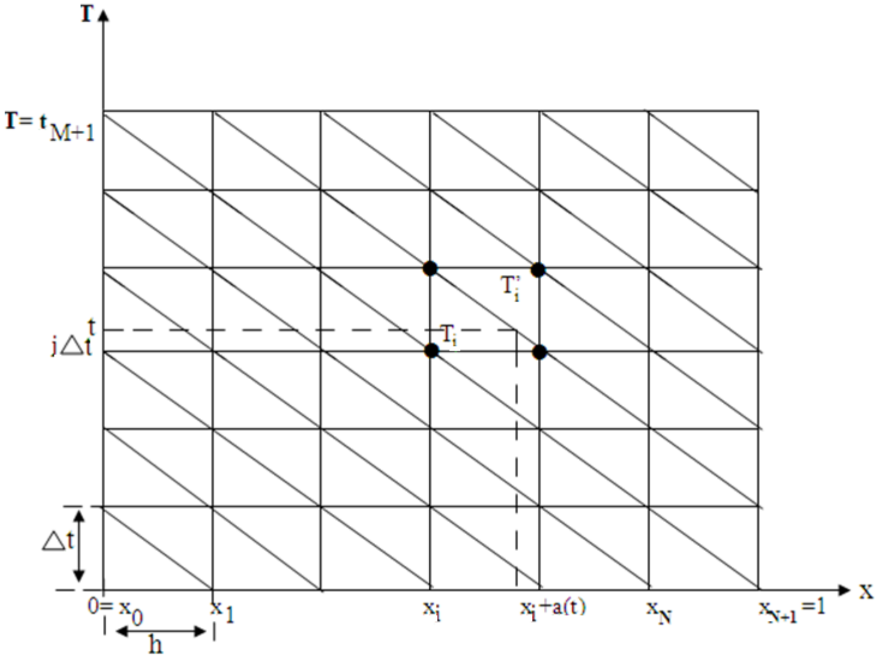
A simple calculation gives

$$\begin{aligned} \lambda'_1(x,t) &= \frac{x}{h} + \frac{t}{\Delta t} - (i+j+1) = -\lambda_1(x,t), \\ \lambda'_2(x,t) &= -\frac{x}{h} + i + 1 = -(\lambda_2(x,t) - 1), \\ \lambda'_3(x,t) &= -\frac{t}{\Delta t} + j + 1 = -(\lambda_3(x,t) - 1). \end{aligned}$$

The calculation of $D_{i+\frac{1}{2}}$ is more delicate when x varies between x_i and x_{i+1} and t between $j\delta t$ and $(j+1)\delta t$ the point $(x,t) \in T_{ij} \cup T'_{ij}$ so

$$\begin{aligned} &\text{if } x \in [x_i, x_i + a(t)] \text{ then } (x,t) \in T_i\,, \\ &\text{if } x \in [x_i + a(t), x_i] \text{ then } (x,t) \in T'_i\,. \end{aligned}$$

Figure 1: Mesh



where $a(t) = h(1 - s - j)$ (see Figure 1), with $s = t/\delta t$ so

$$D_{i+\frac{1}{2}} = \frac{h}{2}\left[\alpha_j(t)D_{ij} + \beta_j(t)D_{i+1,j} + \gamma_j(t)D_{i,j+1} + \delta_j(t)D_{i+1,j+1}\right], \tag{14}$$

with $\alpha_j(t) = (1+j-s)^2$, $\beta_j(t) = (1+j-s)(1-j+s)$, $\gamma_j(t) = (s-j)(2+j-s)$, and $\delta_j(t) = (s-j)^2$. Thus, knowledge of D_{ij}, $0 \leqslant i \leqslant N+1$, $0 \leqslant j \leqslant M+1$, approach $D(x,t)$ on Q_T and consequently reduce the problem of minimization in finite dimensions: D will equate to a vector $\mathbb{R}^k$ where $k = (N+2)(M+2)$. We denote

$$D = (D_{00}, \ldots, D_{(N+1)0}, D_{01}, \ldots, D_{(N+1)1}, \ldots, D_{0(M+1)}, \ldots, D_{(N+1)(M+1)})^T.$$

The problem (Q_n) becomes

$$\min J_n(D), \quad D \in \mathbb{R}^k, \quad D \geqslant \alpha \geqslant 0. \tag{Q_{nh}}$$

Then we obtain

$$\nabla J_n(D) = \left(\frac{\partial J_n}{\partial D_{00}}, \frac{\partial J_n}{\partial D_{01}}, \dots, \frac{\partial J_n}{\partial D_{0(N+1)}}, \dots, \frac{\partial J_n}{\partial D_{(N+1)(M+1)}}\right)^T .$$

As

$$\frac{\partial J_n}{\partial D_{ij}} = \sum_{k=1}^{N} \frac{\partial J_n}{\partial u^D_{n,k}} \frac{\partial u^D_{n,k}}{\partial D_{ij}}(t),$$

and

$$\frac{\partial u^{*D}_n}{\partial D_{ij}}(t) = \left(\frac{\partial u^D_{n,1}}{\partial D_{ij}}, \frac{\partial u^D_{n,2}}{\partial D_{ij}}, \dots, \frac{\partial u^D_{n,N}}{\partial D_{ij}}\right)^T .$$

The expression (13) gives

$$\frac{\partial J_n}{\partial u_{n,k}} = \frac{2}{3}h\left[u^D_{k,n}(T) - \chi_{T,n}(x_k)\right] + \frac{h}{6}\left[u^D_{k-1,n}(T) - \chi_{T,n}(x_{k-1})\right] + \frac{h}{6}u^D_{k+1,n}(T) - \chi_{T,n}(x_{k+1}).$$

Knowing that $\chi_T(x_0) = \chi_T(x_{N+1}) = 0$ and $u^D_0 = u^D_{N+1} = 0$, to calculate $\frac{\partial J_n}{\partial D_{ij}}$, we assume

$$v_{ij}(t) = \frac{\partial u^{*D}_n}{\partial D_{ij}}(t) = \left(\frac{\partial u^D_{n,1}}{\partial D_{ij}}, \frac{\partial u^D_{n,2}}{\partial D_{ij}}, \dots, \frac{\partial u^D_{n,N}}{\partial D_{ij}}\right)^T .$$

Then

$$\frac{\partial v_{ij}}{\partial t} = \frac{\partial}{\partial D_{ij}}\left(\frac{\partial u^{*D}_n}{\partial t}\right).$$

By using (S_2), it follows

$$\frac{\partial v_{ij}}{\partial t} = -M^{-1}\frac{\partial R_D}{\partial D_{ij}}u^{*D}_n(t) - M^{-1}R_D(t)\frac{\partial u^{*D}_n}{\partial D_{ij}}(t).$$

To calculate (v_{ij}), it suffices to solve the following system, using the fourth-order Runge–Kutta method,

$$\frac{\partial v_{ij}}{\partial t} = -M^{-1}\frac{\partial R_D}{\partial D_{ij}} u_n^{*D}(t) - M^{-1}R_D(t)v_{ij}\,, \quad v_{ij}(0) = 0\,.$$

The matrix coefficients $R_D(t)$ and $\frac{\partial R_D(t)}{\partial D_{ij}}$ are calculated using (10).

6.2 Spectral discretization of the problem $(\mathcal{P}V_n)$

In this section we recall some formulas in the spectral method in a reference field $\Lambda =]-1, 1[$, then we use these results to write the variational formulation in our field of study Ω.

Let N be an integer $\geqslant 2$. Denote the space $\mathbb{P}_N(\Lambda)$ of polynomials with degree $\leqslant N$ and the space $\mathbb{P}_N^0(\Lambda)$ of polynomials in $\mathbb{P}_N(\Lambda)$ vanishing on the boundary of Λ [8, 9]. We introduce the space $\mathbb{P}_N(-1, 1)$ of restrictions to Λ of polynomials with degree $\leqslant N$. Setting $\xi_0 = -1$ and $\xi_N = 1$, we introduce the $N-1$ nodes ξ_j, $1 \leqslant j \leqslant N-1$, and the $N+1$ weights ρ_j, $0 \leqslant j \leqslant N$, of the Gauss–Lobatto quadrature formula, recall that

$$\int_{-1}^{1} \phi(\zeta)d\zeta = \sum_{j=0}^{N} \phi(\xi_j)\rho_j\,. \tag{15}$$

We also recall [6, (13.20)] the following property, which is useful in what follows:

$$\text{for all } \varphi_N \in \mathbb{P}_N(-1, 1), \quad \|\varphi_N\|^2_{L^2(\Lambda)} \leqslant \sum_{j=0}^{N} \varphi_N^2(\zeta)\rho_j \leqslant 3\|\varphi_N\|^2_{L^2(\Lambda)}. \tag{16}$$

Relying on this formula, we introduce the discrete product [8, 9] defined on continuous functions u and v by

$$(u, v) = \sum_{j=0}^{N} u(\xi_j)v(\xi_j)\rho_j\,. \tag{17}$$

It follows from (16) that this discrete product is a scalar product on $\mathbb{P}_N(\Lambda)$. Let finally $\mathcal{I}_N$ denote the Lagrange interpolation operator at the nodes ξ_i, $0 \leqslant i \leqslant N$, with values in $\mathbb{P}_N(\Lambda)$.

We now assume that the function $F_{n-1}(t)$ is continuous on $\bar{\Omega}$. Thus the discrete problem is constructed from ($\mathcal{P}V_n$) using the Galerkin method combined with numerical integration. It reads

$$\begin{aligned}
&\text{Find } \ u_{n,N}(\cdot,t) \in \mathbb{P}_N(\Omega), \quad \text{with } u_{n,N}(\cdot,t) - \mathcal{I}_N \bar{g}(\cdot,t) \in \mathbb{P}_N^0(\Omega),\\
&\text{such that for all } v_N \in \mathbb{P}_N^0(\Omega)\\
&\frac{d}{dt}(u_{n,N}(\cdot,t), v_N)_N + a_{D,N}(u_{n,N}(\cdot,t), v_N) = (F_{n-1}(t), v_N)_N\,,\\
&u_{n,N}(\cdot,0) = v_{0,N}\,,
\end{aligned} \tag{18}$$

where the bilinear form

$$a_{D,N}(u_{n,N}, v_N) = \left(D_N \frac{\partial u_{n,N}(\cdot,t)}{\partial x}, \frac{\partial v_N}{\partial x} \right)_N . \tag{19}$$

It follows from ($\mathcal{P}V_n$), combined with Cauchy–Schwarz inequalities, that the form $a_{D,N}$ is continuous on $\mathbb{P}_N^0(\Omega) \times \mathbb{P}_N^0(\Omega)$, with norm bounded independently of N. Then to investigate the well-posedness of the problem we use Lions's theorem [2, 5]. To simplify the problem, we suppose that we know the values of $F_{n-1}(t)$ at the nodes of Ω. Let ℓ_i be Lagrange polynomials at the nodes $\tilde{\xi}_i$, $0 \leqslant i \leqslant N$ [6, Remark 1.5]. Then we calculate the coordinates $u_{n,i}(t)$ of solution $u_{n,N}(\cdot,t)$ in the base ℓ_i, for each t in $[0,T]$. So

$$u_{n,N}(x,t) = \sum_{i=0}^{N} u_{n,i}(t)\ell_i(x).$$

We note

$$v_N = \ell_r, \quad 1 \leqslant r \leqslant N-1\,,$$

and

$$D_N\,(x,t) = \sum_{m=0}^{N}\sum_{p=0}^{N} D_{mp}\ell_m\,(x)\,\ell_p\,(t)\,.$$

The variational problem (18)

$$\sum_{i=0}^{N}(\ell_i,\ell_r)\frac{\partial}{dt}u_{n,i}(t)+\sum_{i=0}^{N}a_{D,N}(\ell_i,\ell_r)u_{n,i}(t)=F_{n-1}(t)(\tilde{\xi}_r)\rho_r\,, \tag{20}$$

$$u_{n,i}(0)=v_{0,i}\,, \tag{21}$$

where $u_{n,i}(t)$ are the values of $u_{n,N}(t)$ at the nodes $\tilde{\xi}_i$ of the Gauss–Lobatto in Ω. We obtain $N-1$ equations with $N-1$ unknowns. So (19) is equivalent to the linear system

$$\begin{aligned} &M\frac{d}{dt}u_n^{*D}(t)+R_D(t)u_n^{*D}(t)=\mathcal{F}_{n-1}(t),\\ &u_n^{*D}(0)=\tilde{v}_0\,, \end{aligned} \tag{22}$$

where $u_n^{*D}(t)$ is the vector of coordinates $u_{n,i}(t)$, $1\leqslant i\leqslant N-1$. M is the mass matrix, $R_D(t)$ is the stiffness matrix. $\mathcal{F}_{n-1}(t)$ is the vector of coordinates

$$\begin{aligned} &\frac{1}{2}\left(f(\tilde{\xi}_r,t,u_{n-1}(\tilde{\xi}_r,t))-\frac{\partial f}{\partial s}(\tilde{\xi}_r,t,u_{n-1}(\tilde{\xi}_r,t))u_{n-1}(\tilde{\xi}_r)\right)\rho_r\\ &-a_{D,n}(\ell_0,\ell_r)u_n(\tilde{\xi}_0,t)+a_{D,n}(\ell_N,\ell_r)u_n(\tilde{\xi}_N,t),\quad 1\leqslant r\leqslant N-1\,. \end{aligned}$$

$\tilde{v}_0$ is constituted with the values $v_{0,i}$, $1\leqslant i\leqslant N-1$, at the nodes $\tilde{\xi}_i$ of the Gauss–Lobatto in Ω. The coefficients of M are written as [13, Appendix]

$$(\ell_i,\ell_r)=\begin{cases}0 & \text{if } 0\leqslant i<r\leqslant N\,,\\ \frac{2}{N(N+1)L_N^2(\tilde{\xi}_i)} & \text{if } 1\leqslant i=r\leqslant N-1\,.\end{cases}$$

And the coefficients of $R_D(t)$ are

$$\begin{aligned} a_{D,N}(\ell_i,\ell_r)=2\sum_{p=0}^{N}\bar{\ell}_p(t)\sum_{k=0}^{N}D_{kp}\ell_i'(\tilde{\xi}_k)\ell_r'(\tilde{\xi}_k)\rho_k\\ -\frac{1}{2}\frac{\partial f}{\partial s}(\tilde{\xi}_r,t,u_{n-1}(\tilde{\xi}_r,t))\delta_{ir}\rho_r\,. \end{aligned}$$

6.2.1 Calculating the gradient of J_n

Consider the problem (Q_n) as

$$u_{n,N}(t)(x) = \sum_{j=0}^{N} u^D_{n,j}\ell_j(x), \quad \chi_T(x) = \sum_{i=0}^{N} \chi_{T,i}\ell_i(x),$$

and setting $u^D_n(\tilde{\xi}_k)(T) = u^D_{n,k}(T)$, it follows

$$J_n(D) = \sum_{k=0}^{N} \left[u^D_{n,k}(T) - \chi_T(\tilde{\xi}_k)\right]^2 \rho_k\,. \tag{23}$$

To calculate an approximation of the gradient J_n (12), we consider a mesh of rectangle $\bar{\Omega} \times [0, T]$. Then

$$\frac{\partial J_n(D)}{\partial D_{mp}} = \sum_{k=0}^{N} \frac{\partial J_n(D)}{\partial u^D_{n,k}} \frac{\partial u^D_{n,k}}{\partial D_{mp}}\,.$$

We then observe that, via (23),

$$\frac{\partial J_n(D)}{\partial u_{n,k}} = 2(u^D_{n,k}(T) - \chi_{T,k})\rho_k\,.$$

And then find

$$\frac{\partial u^{*D}_n(t)}{\partial D_{mp}} = \left(\frac{\partial u^D_{n,1}(t)}{\partial D_{mp}}, \ldots, \frac{\partial u^D_{n,N-1}(t)}{\partial D_{mp}}\right),$$

we assume $v_{m,p}(t) = \frac{\partial u^{*D}_n(t)}{\partial D_{mp}}$, and using (22) gives

$$\frac{\partial v_{m,p}}{\partial t} = -M^{-1}\frac{\partial(R_D u^{*D}_n(t))}{\partial D_{mp}}, \quad v_{m,p}(0) = 0\,. \tag{24}$$

To solve (24),

$$R_D u_n^{*D} = 2 \sum_{m=0}^{N} \sum_{p=0}^{N} D_{mp} \left(\sum_{i=0}^{N} \sum_{j=0}^{N} u_{ij} \ell_i'(\tilde{\xi}_m) \bar{\ell}_j(t) \right) \ell_r'(\tilde{\xi}_m) \bar{\ell}_p(t)\, \rho_m \,,$$

and finally

$$\frac{\partial R_D u_n^{*D}(t)}{\partial D_{mp}} = 2 \left(\sum_{i=0}^{N} \sum_{j=0}^{N} u_{ij} \ell_i'(\tilde{\xi}_m) \bar{\ell}_j(t) \right) \ell_r'(\tilde{\xi}_m) \bar{\ell}_p(t)\, \rho_m \,.$$

7 Some numerical experiments

In this section we present some numerical experiments using MATLAB as a computational tool. The algorithm of resolution of the problem (Q), is divided into two parts.

- Firstly we solve the problem ($\mathcal{P}V_n$), in the matrix form (S_2) and (22) in both methods. The resolution of U_1 is done knowing $D_0 = D(\cdot, 0)$. The method used is a fourth order Runge–Kutta.
- Secondly, once U_1 is calculated, it is then introduced in the minimization problem (Q_D) to recover D_1. The best method used in our case is a Levenberg–Marquardt method based on the calculation of gradient of $J(D)$ detailed in subsections 6.1.2 and 6.2.1. We continue this process until we get the desired convergence.

The algorithm requires about 7% of total time for calculating the solution u and 93% to calculate D. Runge–Kutta is used three times in the program. In the Levenberg–Marquardt algorithm, the iteration number to achieve maximum convergence of D varies somewhat as function of N, and it does not exceed five iterations. We need also just a few iterations to assure the convergence of the global system. Table 1 reports the average CPU time (in

Table 1: The CPU time in seconds required for one iteration of Runge–Kutta.

N	FEM1	FEM2	SP.M
8	0.01	0.1	1
12	0.01	0.2	2
16	0.02	0.4	3
20	0.02	0.5	4
24	0.03	0.7	8
28	0.03	0.9	13
32	0.03	1.1	24
36	0.03	1.3	53

seconds) required for one iteration of Runge–Kutta using the three methods with N varying between 8 and 36.

We use the BICGSTABL method (GMRES method is also a suitable alternative). Several preconditioners can be found in literature. In our case a preconditioner performing an incomplete LU factorization is the best appropriate to be used for all three methods. The convergence is assured in four iterations at most. Without using any preconditioner the number of iterations increases as a function of N. For example by fixing a tolerance 10^{-6} the iteration number exceeds ten for N greater than 30. Without loss of generality, we consider the domain of study $\bar{Q}_T$. Numerically we prove the convergence of the same problem in a more general case taking $u(0,t) = g_1(t)$ and $u(1,t) = g_2(t)$ where g_1 and g_2 are given. We present a complete study of the problem, using finite element method (explained in subsection 6.1). We give in this method two different approaches. In the first approach (FEM1) we use the trapezoidal formula for approximating integrals. Regarding the second approach (FEM2) we employ the Gauss–Lobatto formula. Finally, we present a spectral method (SP.M), with a comparative study for each test case.

1. We consider the following analytical functions:

$$\begin{aligned} u(x,t) &= (x^2+2)(t+1), \quad D(x,t) = \frac{1}{24}(x+1)(t+1), \\ f(x,t) &= -s(x,t)u^2 + h(x,t), \text{ with } s(x,t) = (t^2+1)e^x, \\ &\quad \text{and } h(x,t) \text{ verifies the equation } (\mathcal{P}_D). \end{aligned} \tag{25}$$

   Figures 2 and 3 plot the two quantities $\log_{10}(\|u - u_N\|_{L^2(\Omega)})$ and $\log_{10}(\|D - D_N\|_{L^2(\Omega)})$ as functions of $\log_{10}(\delta t)$. We apply this for both the methods. In all these calculations we take $T = 0.1$ and $N = 30$.

2. We consider the following functions:

$$\begin{aligned} u(x,t) &= e^t(x+1), \quad D(x,t) = \frac{1}{2}e^{-t}(x^2+1), \\ f(x,t) &= -s(x,t)u^2 + h(x,t), \text{ with } s(x,t) = tx^2, \\ &\quad \text{and } h(x,t) \text{ verifies the equation } (\mathcal{P}_D). \end{aligned} \tag{26}$$

   Figures 4 and 5 plot the two quantities $\log_{10}(\|u - u_N\|_{L^2(\Omega)})$ and $\log_{10}(\|D - D_N\|_{L^2(\Omega)})$ as functions of $\log_{10}(\delta t)$. We apply this for both the methods. All these calculations are made by fixing $T = 0.1$ and $N = 35$. In both the cases (first test and second test) presented, we find that the spectral method is more precise than finite elements method. The slope of the curve estimations in approach 2 (finite element method) and the spectral method proves the good convergence in time of the solutions u and D. While in approach 1 (finite element method) the slopes of the curve estimations decreases very slowly with time, this result is due to the poor convergence space stops the evolution in time of errors.

3. For the same example as the first case, Table 2 presents the error $|u - u_N|$ and Figure 6 the approached solution u_N for times t varying between 0 and 1 in the nodes ξ_i, $i = 1, \ldots, 8$. The results are for $T = 1$, $\delta t = 10^{-3}$ and $N = 8$.

Figure 2: L^2 error between the exact solution u given in the first test and the experimental solution u_N using the three approaches.

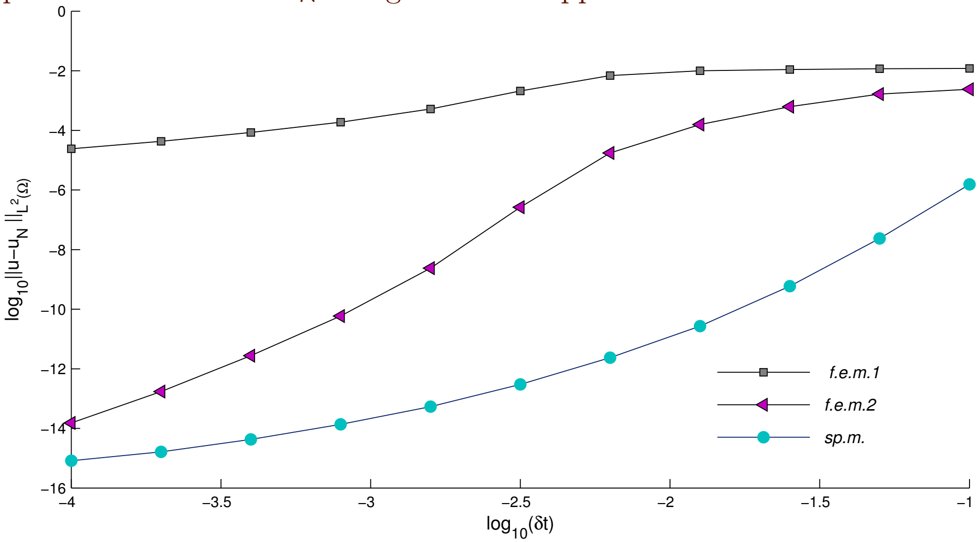


Figure 3: L^2 error between the exact solution D given in the first test and the experimental solution D_N.

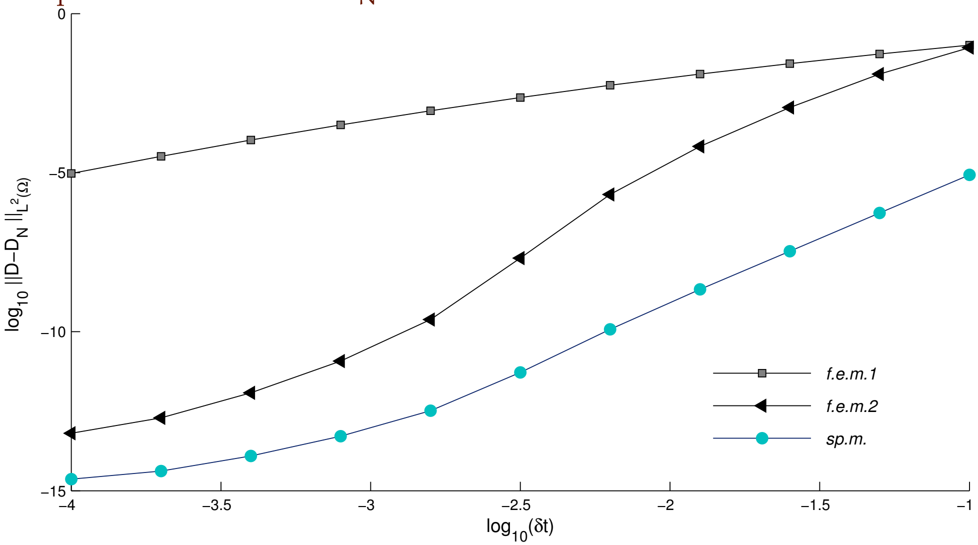


Figure 4: L^2 error between the exact solution u given in the second test and the experimental solution u_N.

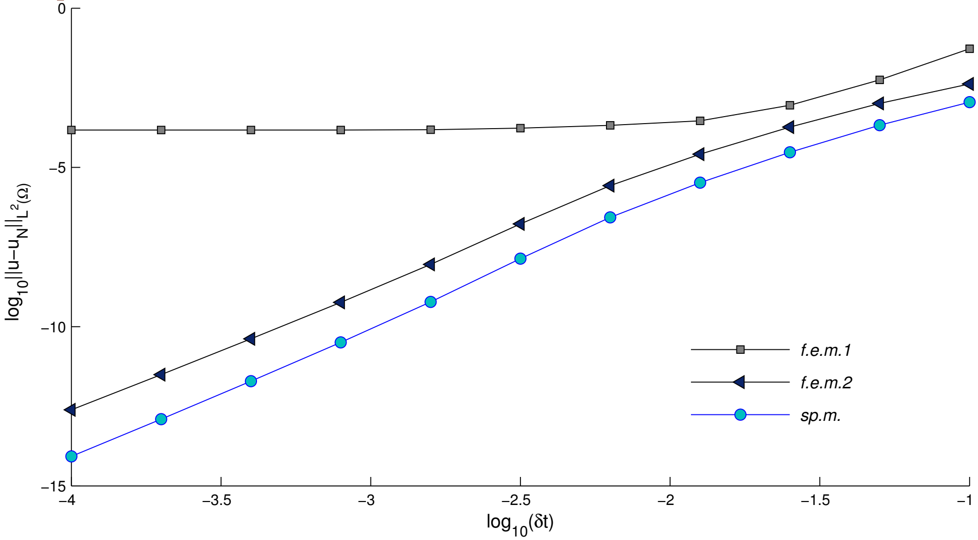


Figure 5: L^2 error between the exact solution D given in the second test and the experimental solution D_N.

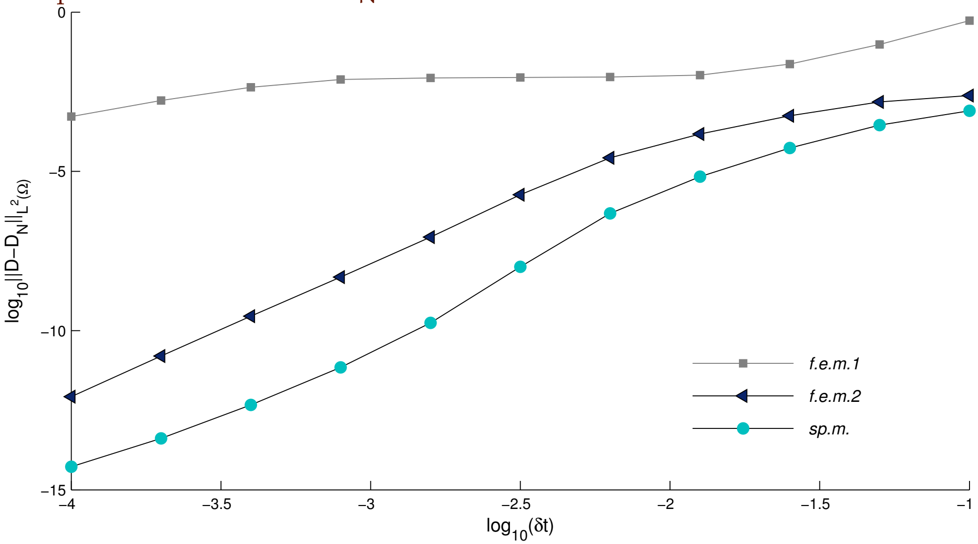


Table 2: The values $|u(\xi_i, .) - u_N(\xi_i, .)|$ in the third test calculated at the nodes ξ_i for different values of t, with $N = 8$ and $T = 1$.

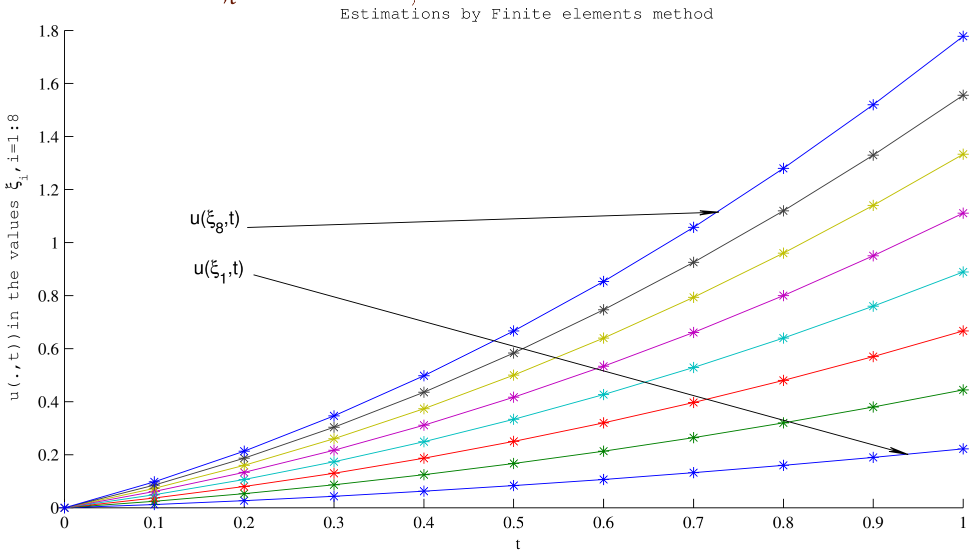
t	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	ξ_8
0	0	0	0	0	0	0	0	0
0.1	1E−13	1E−13	1E−12	7E−12	5E−12	1E−11	2E−11	3E−11
0.2	1E−13	1E−12	2E−12	7E−12	2E−11	5E−11	1E−10	1E−10
0.3	1E−12	2E−12	6E−12	1E−11	5E−11	1E−10	2E−10	3E−10
0.4	1E−12	4E−12	1E−11	3E−11	1E−10	2E−10	5E−10	6E−10
0.5	2E−12	7E−12	2E−11	5E−11	1E−10	4E−10	9E−10	1E−9
0.6	4E−12	1E−11	3E−11	1E−10	2E−10	6E−10	1E−9	1E−9
0.7	6E−12	1E−11	5E−11	1E−10	4E−10	1E−9	2E−9	2E−9
0.8	9E−12	2E−11	8E−11	2E−10	5E−10	1E−9	2E−9	3E−9
0.9	1E−11	4E−11	1E−10	3E−10	8E−10	1E−9	3E−9	4E−9
1	2E−11	6E−11	1E−10	4E−10	1E−9	2E−9	4E−9	5E−9

4. We study the convergence of solutions depending on the parameter n. Figures 7 and 8 show the evolution of curves as a function of N; N varying between 6 and 24. We take $T = 0.1$ and $\delta t = 10^{-2}$, and the exact solutions are

$$\begin{aligned} u(x,t) &= (t+1)(x^2+1)^{1/2} \quad D(x,t) = \frac{0.25}{1+t}(x^2+1)^{3/2}, \qquad (27)\\ f(x,t) &= -s(x,t)u^4 + h(x,t), \text{ with } s(x,t) = e^{t^2+1}(x^2+1), \\ &\text{and } h(x,t) \text{ verifies the equation } (\mathcal{P}_D). \end{aligned}$$

In this case u is a singular function. The maximum error of the solutions is reached at $N = 24$ in approach 1 and 2 of finite element methods, and at $N = 14$ in the spectral method. Knowing that, if we take u as a polynomial function, then the maximum is reached from $N = 4$ for the two methods.

Figure 6: The curves of $u_N(\xi_i, t)$ as a function of t, calculated in the third test at the nodes ξ_i with $N = 8$, $T = 1$ and $\delta t = 10^{-3}$.



5. In this test, we take

$$f(x,t) = -\left(\frac{u}{(x+1)(t+1)}\right)^2 + x + 10^{-5}, \quad u(x,0) = x+1, \quad (28)$$

$$g(t) = \begin{cases} (t+1) & \text{on } \{0\} \times \bar{\Omega}\,, \\ 2(t+1) & \text{on } \{1\} \times \bar{\Omega}\,, \end{cases} \quad (29)$$

$$D(x,0) = x+1\,.$$

The solution in this case is unknown. The calculations done with the second approach (finite element methods) and spectral method are presented in Figures 9 and 10. Figures 9 and 10 are almost identical, this proves the good convergence of the two algorithms.

Figure 7: L^2 error between the exact solution u given in the fourth test and the experimental solution u_N.

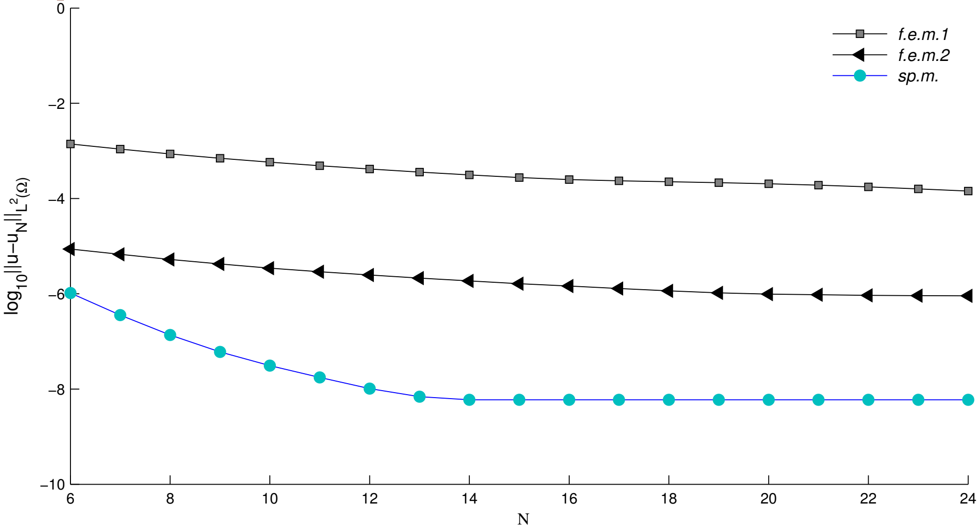


Figure 8: L^2 error between the exact solution D given in the fourth test and the experimental solution D_N.

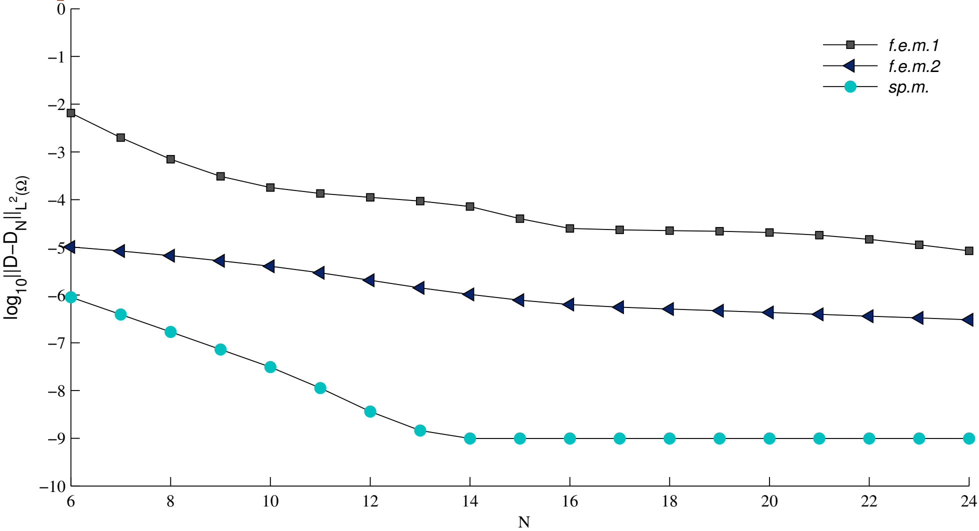
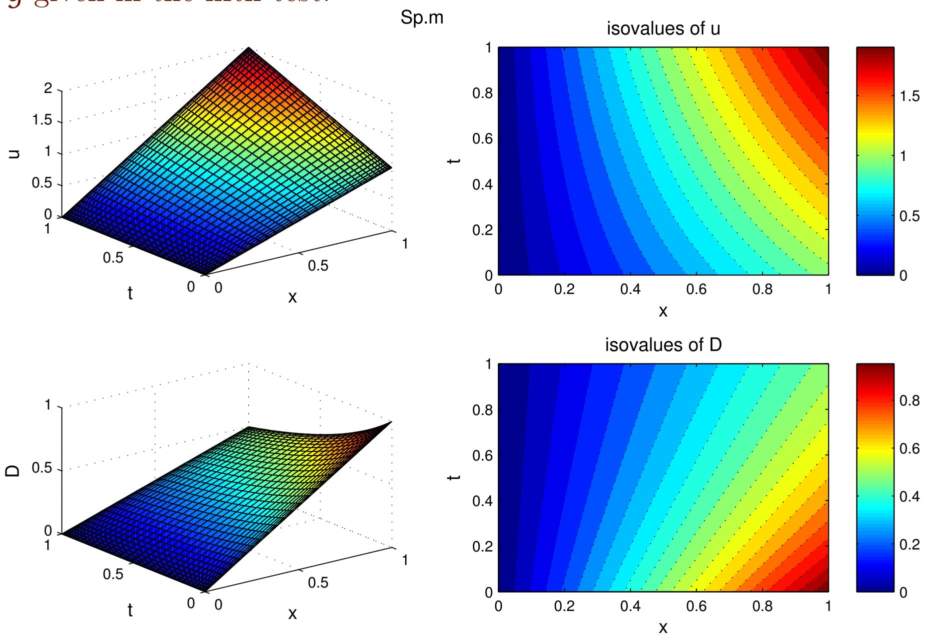


Figure 9: The discrete solution u_N and D_N using spectral method with data f and g given in the fifth test.



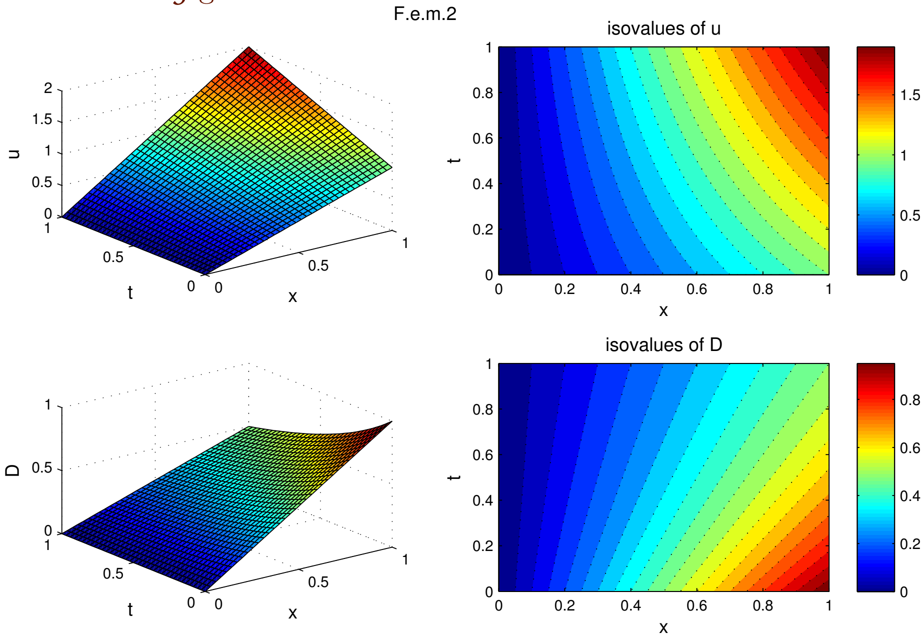
Conclusion The selection of the two algorithms introduced in the paper gives a good and stable numerical convergence of the two methods. The results of the numerical simulations that we performed confirm all theoretical results found.

References

[1] P. A. Raviart, D. J. M. Thomas, *Introduction à l'analyse numérique des équations aux dérivées partielles*, Masson (1983). E6, E11, E17

[2] J.L. Lions, E. Magenes, *Problèmes aux limites non homogènes et applications, Vol. 1*, Dunod (1968). E6, E13, E16, E24

Figure 10: The discrete solution u_N and D_N using finite elements method with data f and g given in the fifth test.



[3] M. Bouchiba, S. Abidi, Identification of the Diffusion in a lineair Parabolic problem, *International Journal of Applied Mathematics*, Vol. 23, No. 3, 2010, 491–501. E3, E4

[4] M. Kern, *Problèmes Inverses Aspects Numériques*, INRIA (2003). E3, E4

[5] J. E. Rakotoson, J. M. Rakotoson, *Analyse Fonctionnelle appliquée aux équations aux dérivées partielles*, Presse Universitaires de France, (1999). E17, E24

[6] C. Bernardi, Y. Maday, Spectral Methods, in *the Handbook of Numerical Analysis V*, P. G. Ciarlet & J. L. Lions eds., North-Holland, (1997), 209–485. E23, E24

[7] V. Girault, P.-A. Raviart, *Finite Element Methods for Navier–Stokes Equations, Theory and Algorithms*, Springer-Verlag (1986). E4

[8] C. Bernardi, Y. Maday, F. Rapetti, *Discrétisation Variationnelles de problèmes aux limites elliptiques*, Springer-Verlag (2004). E4, E23

[9] C. Bernardi, Y. Maday, *Approximations spectrales de problèmes aux limites elliptiques*, Springer-Verlag (1992). E4, E6, E23

[10] I. Ekeland, R. Teman, *Analyse Convexe et problèmes Variationnels*, Bordas (19674). E4, E15

[11] J. L. Lions, *Controle Optimal de systèmes gouvernés par des équations aux dérivées partielles*, Dunod, Paris, (1968). E4

[12] J. Henry, *Étude de la contrôlabilité de certaines équations paraboliques*, Thèse d'état, Université Paris VI, (1978). E4

[13] J. Satouri, *Méthodes déléments spectraux avec joints pour des géométries axisymétriques*, Thèse de l'Université Pierre et Marie Curie, Paris VI (2010). E25

[14] K. Van de Doel, U. M. Ascher, On level set regularization for highly ill-posed distributed parameter systems, *Journal of Computational Physics*, 2006, 216: 707–723.
http://people.cs.ubc.ca/~ascher/papers/da.pdf E3

[15] M. Böhm, M. A. Demetriou, S. Reich, and I. G. Rosen, Model Reference Adaptive Control of Distributed Parameter Systems, *SIAM J. Control Optim.*, 36(1), 33–81; doi:10.1002/rnc.1098 E3

[16] C. Jia, A note on the model reference adaptive control of linear parabolic systems with constant coefficients. *Journal of Systems Science and Complexity* 24, 1110–1117; doi:10.1007/s11424-011-0042-9 E3

[17] D. B. Pietria, M. Krstic, Output-feedback adaptive control of a wave PDE with boundary anti-damping. *Automatica* 50, 1407–1415; doi:10.1016/j.automatica.2014.02.040 E3

Author addresses

1. **Sameh Abidi**, University of Tunis El Manar, Faculty of Sciences of Tunis, Department of Mathematics, 1060 Tunis, Tunisia.
   `mailto:labidisameh@yahoo.fr.`
2. **Jamil Satouri**, IPEIT-University of Tunis, 2 street Jawaher Lel Nahra-1089, Monfleury, Tunisia.
   `mailto:jamil.satouri@yahoo.fr`