Memoirs on Differential Equations and Mathematical Physics

Volume 91, 2024, 1–20

Soufiane Abid, Khalid Atifi, El-Hassan Essoufi

AN INVERSE BACKWARD PROBLEM FOR A HEAT EQUATION WITH A MEMORY TERM, SOLVED BY A DEEP LEARNING METHOD

Abstract. Our goal of this paper is the identification of an initial condition in a heat equation that contains a memory term from final data. To this aim, we first establish the well-posedness of the direct problem. Then we prove the continuity and the G-derivability of the cost function. Finally, we validate the results numerically by using a deep neural network. Our algorithm is meshfree.

2020 Mathematics Subject Classification. 35R30, 35Q93, 35R09, 68T07.

Key words and phrases. Inverse problem, memory term, optimization, deep learning.

რეზიუმე. ნაშრომის მიზანია საწყისი პირობის იდენტიფიცირება სითბოგამტარებლობის განტოლებაში, რომელიც შეიცავს მეხსიერების წევრს საბოლოო მონაცემებიდან. ამ მიზნით, თავდაპირველად დადგენილია პირდაპირი ამოცანის კორექტულობა. შემდეგ დამტკიცებულია ხარჯების ფუნქციის უწყვეტობა და G-წარმოებადობა. ბოლოს, ღრმა ნერვული ქსელის გამოყენებით რიცხობრივად შემოწმებულია შედეგები. ეს არის ბადისგარეშე ალგორითმი.

1 Introduction

Several physical phenomena are described by partial differential equations where the model's dynamics is influenced by the past history of some variable(s). This means that some quantities are averaged by means of an integral with respect to a positive summable function, called the memory kernel. The presence of memory can, in some cases, make the description of the phenomena more accurate. On the other hand, equations with memory are generally much more difficult to handle than the corresponding equations without memory. However, the integral term may appear in the boundary conditions or in the partial differential equation itself. Problems of this type can involve a wide range of scientific and technical fields, for example, chemical diffusion, thermoelasticity, heat conduction processes, population dynamics, vibration problems, nuclear reactor dynamics, medical sciences, biochemistry and some biological processes. Generally, the thermal flow in memory materials is governed by parabolic integrodifferential equations containing time-dependent and, in the non-homogeneous case, also space-dependent memory kernels.

Due to the great importance of this type of equations, the interest in the study of inverse problems and the identification of the parameters associated with them began in the early 20th century with the works of H. Grabmueller [11], M. F. Zedan [24] and A. Lorenzi [17]. Until now, interest in this model is still presented. But almost all studies deal with identifying the memory kernel [4,5,12,13,15,21–23], the source term [6,16], velocity or another parameter in the equation [7,14] using various methods and approaches carried out in different theoretical and numerical aspects of inverse problems, such as the existence, uniqueness, stability and validation of results by numerical simulations, and so on.

Nevertheless, works dealing with the recovery of initial conditions in this case are still rare (to our knowledge, they have not been considered before).

Among the papers that considered equations similar to our one, we refer, for example, to [8], where the authors investigated two inverse problems related to the one-dimensional equation

$$u_t - a^2 u_{xx} = \int_0^t k(\tau) u(x, t-\tau)\, d\tau + h(x,t), \quad x \in (0,l), \ \ 0 < t \le T; \tag{1.1}$$

the first problem consists in identifying the kernel k of the integral term by using additional information about the solution of the direct problem

$$\int_0^l u(x,t)\, dx = f(t). \tag{1.2}$$

In the second problem, using the additional information

$$u(x_0, t) = f(t), \quad x_0 \in (0,l), \ \ t \in (0,T], \tag{1.3}$$

they proved the unique solvability of this inverse problem by using the principle of contracting mapping.

Another example is the problem discussed in [14] by K. Karuppiah et al. reconstructed the time-independent coefficient q in the integrodifferential equation

$$\begin{cases} y_t - \nabla(d(x)\nabla y) + \displaystyle\int_0^t K(t,\tau) y(x,\tau)\, \mathrm{d}\tau + q(x) y = f(x,t), & (x,t) \in (0,l) \times [0,T], \\ y(0,t) = y(l,t) = 0, & t \in [0,T], \\ y(x,0) = y_0(x), & x \in (0,l), \end{cases} \tag{1.4}$$

from the final time overspecified data

$$y(x,T) = m(x) \quad \text{for all } \ x \in (0,l). \tag{1.5}$$

They transformed the parameter reconstruction problem into a minimization problem through the optimal control. They ended up with a stability estimate of the coefficient with the upper bound in terms of the final measurement derived by minimizing the cost function.

Regarding the numerical part, the presence of the integral term makes the problem complicated and difficult to discretize, mentioning that the discretization of the problem is not easy, a work similar to ours. In [18], the authors used a backwards Euler's method in combination with the rule of integration of Euler's product in time and cubic B-spline in space, while in [9], they applied Euler's method backwards in the time direction and, in addition, used the trapezoidal product integration rule. Here, to validate our results, we propose a meshfree deep learning algorithm. The method is similar in spirit to the Galerkin method, but with several key changes using ideas from machine learning. The Galerkin method is a widely-used computational method which seeks a reduced-form solution to a PDE as a linear combination of basis functions. The deep learning algorithm, or "Deep Galerkin Method" (DGM), uses a deep neural network instead of a linear combination of basis functions. The deep neural network is trained to satisfy the differential operator, initial condition and boundary conditions using stochastic gradient descent at randomly sampled spatial points. By randomly sampling spatial points, the authors avoid the need to form a mesh and instead convert the PDE problem into a machine learning problem.

Here, we study the inverse problem of determining the initial state in a parabolic equation with a memory term from the theoretical analysis and numerical computation angles. More precisely, we consider the following problem:

$$\begin{cases} \partial_t u + A(u) = f & \text{in } Q, \\ u(x,t) = 0, & x \in \partial\Omega, \quad t \in]0;T[\,, \\ u(x,0) = u_0(x), & x \in \Omega. \end{cases} \tag{1.6}$$

A is the operator defined as

$$A(u) = -u_{xx}(x,t) - a(x) \int_0^t u(x,s)\, ds,$$

where $\Omega := (0,1)$, $Q := \Omega \times (0,T)$, $T > 0$ is a fixed moment of time, $u_0 \in L^2(\Omega)$ is the initial condition, $a \in L^\infty(\Omega)$ is a positive coefficient depending on the space and $f \in L^2(Q)$ is the source term.

In the case $a(x) = 0$, problem (1.6) is already treated in [2] and more generally in [3] even in the degenerate case.

Let us assume

$$A_{ad} = \left\{ h \in H^1(\Omega) : \quad \|h\|_{H^1(\Omega)} \leqslant r \right\}, \text{ where } r \text{ is a real strictly positive constant.}$$

Evidently, the set A_{ad} is a bounded, closed and convex subset of $L^2(\Omega)$.

We have $H^1(\Omega) \underset{compact}{\hookrightarrow} L^2(\Omega)$. Since the set A_{ad} is bounded in $H^1(\Omega)$, A_{ad} is a compact in $L^2(\Omega)$.

Let us define our inverse problem.

Inverse Source Problem (ISP). Let u be a solution to (1.6). Determine the initial state u_0 from the measured data at the final time $u(T, \cdot)$.

Remark 1.1. It should be mentioned that we do not need the supplement distributed measurements to obtain the numerical solution of the inverse problem.

We treat Problem (ISP) by interpreting its solution as a minimizer of the following problem:

$$\text{find } \; u_0^\star \in A_{ad} \; \text{ such that } \; E(u_0^\star) = \min_{u_0 \in A_{ad}} E(u_0), \tag{1.7}$$

where the cost function E, defined as

$$E(u_0) = \frac{1}{2T} \|u(T) - u^{obs}\|^2_{L^2(\Omega)}$$

and subject to u, is the weak solution of the parabolic problem (1.6) with the initial state u_0. Let $u^{obs} \in L^2(\Omega)$ be the observation data with a noise.

Problem (1.7) is ill-posed in the sense of Hadamard, some regularization technique is needed in order to guarantee numerical stability of the computational procedure even with noisy input data [20]. The problem thus consists in minimizing a functional of the form

$$J(u_0) = \frac{1}{2T}\left\|u(T) - u^{obs}\right\|^2_{L^2(\Omega)} + \frac{\varepsilon}{2}\left\|u_0 - u^b\right\|^2_{L^2(\Omega)},$$

here, ε is a small positive regularizing coefficient that provides extra convexity to the functional J; u^b is an a priori (background state) knowledge of the state u_0^{exact}. The background error is then defined as $err = \|u_0^{exact} - u^b\|f_2$. u_0^{exact} is called the true state, and is the state to estimate.

2 Well-posedness

Theorem 2.1. *Assume that $u_0 \in L^2(\Omega)$ and $f \in L^2(Q)$. There exists a unique weak solution which solves problem (1.6) such that*

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

and we have the estimate

$$\sup_{t\in[0,T]} \|u(x,t)\|^2_{L^2(\Omega)} + \int_0^T \|u_x(x,t)\|^2_{L^2(\Omega)}\,dt \le C_T\Big(\|f\|^2_{L^2(Q)} + \|u_0(x)\|^2_{L^2(\Omega)}\Big) \tag{2.1}$$

with the constant C_1 depending on Ω, and T.

Proof. The existence and uniqueness of the weak solution of (1.6) is already seen in Proposition 3.1 in [1] (with the particular case $\mu = 0$), here we show only estimate (2.1).

We multiply the first equation of (1.6) by u and integrate over Ω, thus we get

$$\int_0^1 u_t(x,t)u(x,t)\,dx - \int_0^1 u_{xx}(x,t)u(x,t)\,dx$$
$$= \int_0^1 \Big(a(x)u(x,t)\int_0^t u(x,s)\,ds\Big)\,dx + \int_0^1 f(x,t)u(x,t)\,dx. \tag{2.2}$$

By integration by parts, we obtain

$$\frac{1}{2}\frac{d}{dt}\|u(x,t)\|^2_{L^2(\Omega)} + \int_0^1 u_x^2(x,t)\,dx = \int_0^1 \Big(a(x)u(x,t)\int_0^t u(x,s)\,ds\Big)\,dx + \int_0^1 f(x,t)u(x,t)\,dx. \tag{2.3}$$

We have

$$\int_0^1 \Big(a(x)u(x,t)\int_0^t u(x,s)\,ds\Big)\,dx \le \Big(\int_0^1 (a(x)u(x,t))^2\,dx\Big)^{\frac{1}{2}}\Big(\int_0^1\Big(\int_0^t u(x,s)\,ds\Big)^2 dx\Big)^{\frac{1}{2}}$$
$$\le \frac{1}{2}\|a(x)\|^2_{L^\infty(\Omega)}\|u(x,t)\|^2_{L^2(\Omega)} + \frac{T}{2}\int_0^t \|u(x,s)\|^2_{L^2(\Omega)}\,ds, \tag{2.4}$$

and by the Cauchy–Schwarz inequality, for every $t \in [0,T]$ we obtain

$$\int_0^1 f(x,t)u(x,t)\,dx \le \frac{1}{2}\|f(x,t)\|^2_{L^2(\Omega)} + \frac{1}{2}\|u(x,t)\|^2_{L^2(\Omega)}.$$

By returning to equation (2.3), we obtain

$$\frac{1}{2}\frac{d}{dt}\|u(x,t)\|^2_{L^2(\Omega)} + \|u_x(x,t)\|^2_{L^2(\Omega)}$$
$$\leq \frac{1}{2}\|f(x,t)\|_{L^2(\Omega)} + \frac{1}{2}\left(1+\|a(x)\|^2_{L^\infty(\Omega)}\right)\|u(x,t)\|^2_{L^2(\Omega)} + \frac{T}{2}\int_0^t \|u(x,s)\|^2_{L^2(\Omega)}\, ds. \quad (2.5)$$

We integrate over $[0,t]$ for all $t \in [0,T]$,

$$\|u(x,t)\|^2_{L^2(\Omega)} + 2\int_0^t \|u_x(x,s)\|^2_{L^2(\Omega)}\, ds$$
$$\leq \|f(x,t)\|^2_{L^2(Q)} + \|u_0(x)\|^2_{L^2(\Omega)} + \left(1+\|a(x)\|^2_{L^\infty(\Omega)} + T^2\right)\int_0^t \|u(x,s)\|^2_{L^2(\Omega)}\, ds. \quad (2.6)$$

Since

$$2\int_0^t \|u_x(x,s)\|^2_{L^2(\Omega)}\, ds \geq 0,$$

we have

$$\|u(x,t)\|^2_{L^2(\Omega)} \leq \|f(x,t)\|^2_{L^2(Q)} + \|u_0(x)\|^2_{L^2(\Omega)} + \left(1+\|a(x)\|^2_{L^\infty(\Omega)} + T^2\right)\int_0^t \|u(x,s)\|^2_{L^2(\Omega)}\, ds.$$

Using Gronwall's inequality, we get

$$\|u(x,t)\|^2_{L^2(\Omega)} \leq \left(1+\|a(x)\|^2_{L^\infty(\Omega)} + T^2\right)e^T\Big(\|f(x,t)\|^2_{L^2(Q)} + \|u_0(x)\|^2_{L^2(\Omega)}\Big). \quad (2.7)$$

From (2.6) and (2.7), there exists a constant $M > 0$ such that

$$\int_0^T \|u_x(x,t)\|^2_{L^2(\Omega)}\, dt \leq M\Big(\|f(x,t)\|^2_{L^2(Q)} + \|u_0(x)\|^2_{L^2(\Omega)}\Big). \quad (2.8)$$

From (2.7) and (2.8), there exists a constant $C_T > 0$ such that

$$\sup_{t\in[0,T]}\|u(x,t)\|^2_{L^2(\Omega)} + \int_0^T \|u_x(x,t)\|^2_{L^2(\Omega)}\, dt \leq C_T\Big(\|f\|^2_{L^2(Q)} + \|u_0(x)\|^2_{L^2(\Omega)}\Big). \qquad \square$$

Lemma 2.1. *Let u be a weak solution of (1.6) with the initial condition u_0. The function*

$$\phi : u_0 \in L^2(\Omega) \longrightarrow u \in C([0,T]; L^2(\Omega)) \cap L^2(0,T; H^1_0(\Omega))$$

is Lipschitz continuous.

An automatic result of Lemma 2.1 is the following

Theorem 2.2. *Under the same assumptions of Theorem 2.1, the functional J is continuous on A_{ad}, and there exists a unique minimizer $u_0^\star \in A_{ad}$, i.e.,*

$$J(u_0^\star) = \min_{u_0\in A_{ad}} J(u_0).$$

Proof of Lemma 2.1. Let $\delta u_0 \in L^2(\Omega)$ be a small perturbation of u_0 such that $u_0 + \delta u_0 \in A_{ad}$.

Consider $\delta u = u^\delta - u$, where u^δ and u are, respectively, the weak solutions of (1.6) with initial conditions $u_0^\delta = u_0 + \delta u_0$ and u_0. Consequently, δu is the solution of the variational problem

$$\begin{cases} \int_0^1 (\delta u)_t v \, dx + \int_0^1 (\delta u)_x \, v_x \, dx = \int_0^1 \left(av \int_0^t \delta u \, ds \right) dx & \forall \, v \in H_0^1(\Omega), \\ \delta u(0,t) = \delta u(1,t) = 0 & \forall \, t \in [0,T], \\ \delta u(x,0) = \delta u_0 & \forall \, x \in \Omega. \end{cases} \tag{2.9}$$

Hence δu is the weak solution of (1.6) with the initial condition δu_0 and source term $\delta f = 0$. Applying the estimate in Theorem 2.1, we find that there is a constant $C_T > 0$ such that

$$\sup_{t\in[0,T]} \|\delta u\|_{L^2(\Omega)}^2 + \int_0^T \|\delta u(t)\|_{L^2(\Omega)}^2 \, dt \le C_T \|\delta u_0\|_{L^2(\Omega)}^2.$$

Then

$$\|\delta u\|_{C(0,T;L^2(\Omega))}^2 \le C_T \|\delta u_0\|_{L^2(\Omega)}^2$$

and

$$\|\delta u\|_{L^2(0,T;H_0^1)}^2 \le C_T \|\delta u_0\|_{L^2(\Omega)}^2.$$

This ends the demonstration. $\square$

Proof of Theorem 2.2. Let u and u^δ be respectively the weak solutions of (1.6) with the initial conditions u_0 and u_0^δ.

We know that

$$\begin{aligned} & J(u_0^\delta) - J(u_0) \\ & \quad = \frac{1}{2T} \left(\|u^\delta(T) - u^{obs}\|_{L^2(\Omega)}^2 - \|u(T) - u^{obs}\|_{L^2(\Omega)}^2 \right) + \frac{\varepsilon}{2} \left(\|u_0^\delta - u^b\|_{L^2(\Omega)}^2 - \|u_0 - u^b\|_{L^2(\Omega)}^2 \right). \end{aligned}$$

Then

$$\begin{aligned} & |J(u_0^\delta) - J(u_0)| \\ & \quad \le \frac{1}{2T} \left| \|u^\delta(T) - u^{obs}\|_{L^2(\Omega)}^2 - \|u(T) - u^{obs}\|_{L^2(\Omega)}^2 \right| + \frac{\varepsilon}{2} \left| \|u_0^\delta - u^b\|_{L^2(\Omega)}^2 - \|u_0 - u^b\|_{L^2(\Omega)}^2 \right|. \end{aligned}$$

We have

$$\begin{aligned} & \left| \|u_0^\delta - u^b\|_{L^2(\Omega)}^2 - \|u_0 - u^b\|_{L^2(\Omega)}^2 \right| \\ & \qquad = \left| \|u_0^\delta - u^b\|_{L^2(\Omega)} - \|u_0 - u^b\|_{L^2(\Omega)} \right| \left| \|u_0^\delta - u^b\|_{L^2(\Omega)} + \|u_0 - u^b\|_{L^2(\Omega)} \right|. \end{aligned}$$

Let us recall that $\forall \, a, b \in E$, where E is a normed space,

$$\big| \|a\| - \|b\| \big| \le \|a - b\|.$$

This gives

$$\begin{aligned} \left| \|u_0^\delta - u^b\|_{L^2(\Omega)}^2 - \|u_0 - u^b\|_{L^2(\Omega)}^2 \right| & \le \left(\|u_0^\delta - u^b\|_{L^2(\Omega)} + \|u_0 - u^b\|_{L^2(\Omega)} \right) \|u_0^\delta - u_0\|_{L^2(\Omega)} \\ & \le \left(\|u_0 - u^b + u_0^\delta - u_0\|_{L^2(\Omega)} + \|u_0 - u^b\|_{L^2(\Omega)} \right) \|u_0^\delta - u_0\|_{L^2(\Omega)} \\ & \le \left(\|u_0^\delta - u^b\|_{L^2(\Omega)} + 2\|u_0 - u^b\|_{L^2(\Omega)} \right) \|u_0^\delta - u_0\|_{L^2(\Omega)}. \end{aligned}$$

Also, we have

$$\begin{aligned}&\Big|\|u^\delta(T)-u^{obs}\|^2_{L^2(\Omega)}-\|u(T)-u^{obs}\|^2_{L^2(\Omega)}\Big|\\&\quad=\Big|\|u^\delta(T)-u^{obs}\|_{L^2(\Omega)}-\|u(T)-u^{obs}\|_{L^2(\Omega)}\Big|\,\Big|\|u^\delta(T)-u^{obs}\|_{L^2(\Omega)}+\|u(T)-u^{obs}\|_{L^2(\Omega)}\Big|.\end{aligned}$$

Then

$$\begin{aligned}&\|u^\delta(T)-u^{obs}\|^2_{L^2(\Omega)}-\|u(T)-u^{obs}\|^2_{L^2(\Omega)}|\\&\quad\le\Big(\|u^\delta(T)-u^{obs}\|_{L^2(\Omega)}+\|u(T)-u^{obs}\|_{L^2(\Omega)}\Big)\|u^\delta(T)-u(T)\|_{L^2(\Omega)}\\&\quad\le\Big(\|u(T)-u^{obs}+u^\delta(T)-u(T)\|_{L^2(\Omega)}+\|u(T)-u^{obs}\|_{L^2(\Omega)}\Big)\|u^\delta(T)-u(T)\|_{L^2(\Omega)}\\&\quad\le\Big(\|u^\delta(T)-u(T)\|_{L^2(\Omega)}+2\|u(T)-u^{obs}\|_{L^2(\Omega)}\Big)\|u^\delta(T)-u(T)\|_{L^2(\Omega)}.\end{aligned}\tag{2.10}$$

Since the evolution function

$$\phi:u_0\in L^2(\Omega)\longrightarrow u\in C([0,T];L^2(\Omega))\cap L^2(0,T;H^1_0(\Omega))$$

is Lipschitz continuous (Lemma 2.1), we deduce that there is a constant $C>0$ such that

$$\|u^\delta(T)-u(T)\|_{L^2(\Omega)}\le C\|u^\delta_0-u_0\|_{L^2(\Omega)}.\tag{2.11}$$

Therefore,

$$\begin{aligned}|J(u^\delta_0)-J(u_0)|\le\frac{\varepsilon}{2}&\Big(\|u^\delta_0-u^b\|_{L^2(\Omega)}+2\|u_0-u^b\|_{L^2(\Omega)}\Big)\|u^\delta_0-u_0\|_{L^2(\Omega)}\\&+\frac{C}{2T}\Big(C\|u^\delta_0-u^b\|_{L^2(\Omega)}+2\|u_0-u^b\|_{L^2(\Omega)}\Big)\|u^\delta_0-u_0\|_{L^2(\Omega)}.\end{aligned}\tag{2.12}$$

When u^δ_0 tends to u_0, the right-hand side of the inequality tends to 0. Accordingly, J is continuous on the compact A_{ad} and there exists a unique minimizer $u^\star_0\in A_{ad}$ for J. □

Proposition 2.1. *Let u be a weak solution of (1.6) with the initial condition u_0. The function*

$$\phi:u_0\in L^2(\Omega)\longrightarrow u\in C([0,T];L^2(\Omega))\cap L^2(0,T;H^1_0(\Omega))$$

is G-differentiable, which implies that the functional J is G-derivable on A_{ad}.

Proof. Let δu_0 be a small amount such that $u_0+\delta u_0\in A_{ad}$, we define the function

$$F'(u_0):=\delta u_0\in A_{ad}\to\delta u,\tag{2.13}$$

where δu is the solution of the following variational problem:

$$\begin{cases}\displaystyle\int_0^1(\delta u)_t v\,dx+\int_0^1(\delta u)_x v_x\,dx=\int_0^1\Big(av\int_0^t\delta u\,ds\Big)dx & \forall\,v\in H^1_0(\Omega),\\ \delta u(0,t)=\delta u(1,t)=0 & \forall\,t\in[0,T],\\ \delta u(x,0)=\delta u_0 & \forall\,x\in\Omega.\end{cases}\tag{2.14}$$

We pose

$$\Phi(u_0)=F(u_0+\delta u_0)-F(u_0)-F'(u_0)\delta u_0.\tag{2.15}$$

We have to show that

$$\Phi(u_0)=O(\delta u_0).\tag{2.16}$$

It is easy to see that Φ verifies the following variational problem:

$$\begin{cases} \int\limits_0^1 \Phi_t v\,dx + \int\limits_0^1 \Phi_x v_x\,dx = \int\limits_0^1 \left(av \int\limits_0^t \Phi\,ds\right) dx & \forall\, v \in H_0^1(\Omega), \\ \Phi(0,t) = \Phi(1,t) = 0 & \forall\, t \in [0,T], \\ \Phi(x,0) = \delta u_0 - (\delta u_0)^2 & \forall\, x \in \Omega. \end{cases} \tag{2.17}$$

In the same way as used in the proof of continuity, we deduce that

$$\|\Phi\|^2_{C(0,T;L^2(\Omega))} \le C_T \|\delta u_0 - (\delta u_0)^2\|^2_{L^2(\Omega)}$$

and

$$\|\Phi\|^2_{L^2(0,T;H_0^1)} \le C_T \|\delta u_0 - (\delta u_0)^2\|^2_{L^2(\Omega)}.$$

This completes the proof of the proposition. □

3 Stability

In this section, we establish the stability of the solution of the inverse problem.

Lemma 3.1. *Let $u_0^\star$ be a minimizer of the functional J, then there exists a set of functions $(u^\star, w, u_0)$ such that*

$$\begin{cases} u_t^\star(x,t) - u_{xx}^\star(x,t) = a(x) \int\limits_0^t u^\star(x,s)\,ds + f(x,t) & \forall\, (x,t) \in Q, \\ u^\star(0,t) = u^\star(1,t) = 0 & \forall\, t \in (0,T), \\ u^\star(x,0) = u_0(x) & \forall\, x \in (0,1), \end{cases} \tag{3.1}$$

$$\begin{cases} w_t(x,t) - w_{xx}(x,t) = a(x) \int\limits_0^t w(x,s)\,ds & \forall\, (x,t) \in Q, \\ w(0,t) = w(1,t) = 0 & \forall\, t \in (0,T), \\ w(x,0) = \kappa(x) - u_0^\star(x) & \forall\, x \in (0,1), \end{cases} \tag{3.2}$$

and

$$\int\limits_0^T \int\limits_0^1 w(x,T)(u^\star(x,T) - \widetilde{u}(T))\,dx\,dt + \varepsilon \int\limits_0^1 u_0^\star(\kappa - u_0^\star\,dx \ge 0 \tag{3.3}$$

for any $\kappa \in A_{ad}$.

Proof. For any $\kappa \in A_{ad}$ and $0 \le \delta \le 1$, we pose

$$u_0^\delta = (1-\delta)u_0^\star + \delta\kappa \in A_{ad}.$$

Then there exists a solution u^δ of equation (1.6) with the initial condition u_0^δ satisfying

$$J_\delta = J(u^\delta) = \frac{1}{2} \int\limits_0^T \left\|u^\delta(x,T) - \widetilde{u}(T)\right\|^2_{L^2(\Omega)} dt + \frac{\varepsilon}{2} \|u_0^\delta\|^2_{L^2(\Omega)}.$$

Now, taking the Fréchet derivative of J_δ with optimal solution $u_0^\star$, we have

$$\left.\frac{dJ_\delta}{d\delta}\right|_{\delta=0} = \int\limits_0^T \int\limits_0^1 (u^\star(x,T) - \widetilde{u})\widehat{u}_\delta\,dx\,dt + \varepsilon \int\limits_0^1 u_0^\star(\kappa - u_0^\star)\,dx \ge 0, \tag{3.4}$$

where $\widehat{u}_\delta = \frac{du}{d\delta}|_{\delta=0}$ is the Fréchet derivative of u, which verifies the following equation:

$$\begin{cases} (\widehat{u}_\delta)_t(x,t) - (\widehat{u}_\delta)_{xx}(x,t) - a(x)\int_0^t \widehat{u}_\delta(x,s)\,ds = 0, & (x,t) \in \Omega \times [0,T], \\ \widehat{u}_\delta(0,t) = \widehat{u}_\delta(1,t) = 0 & \forall\, t \in [0;T], \\ \widehat{u_\delta}(x,0) = \kappa(x) - u_0^\star(x) & \forall\, x \in [0;1]. \end{cases} \tag{3.5}$$

Set $w = \widehat{u}_\delta$, then w satisfies

$$\begin{cases} w_t(x,t) - w_{xx}(x,t) = a(x)\int_0^t w(x,s)\,ds & \forall\,(x,t) \in Q, \\ w(0,t) = w(1,t) = 0 & \forall\, t \in (0,T)\,, \\ w(x,0) = \kappa(x) - u_0^\star(x) & \forall\, x \in (0,1)\,. \end{cases} \tag{3.6}$$

Combining (3.4) and (3.6), one can easily obtain

$$\int_0^T\int_0^1 w(x,T)\big(u^\star(x,T) - \widetilde{u}(T)\big)\,dx\,dt + \varepsilon\int_0^1 u_0^\star(\kappa - u_0^\star)\,dx \geq 0. \qquad \square$$

Theorem 3.1. *Suppose that $\widetilde{u}_1(T)$ and $\widetilde{u}_2(T)$ are two given functions in $L^2(\Omega)$. Let v_1 and v_2 be the minimizers of J corresponding to $\widetilde{u}_1(T)$ and $\widetilde{u}_2(T)$, respectively. If there exists a point $x_0 \in \Omega$ such that $v_1(x_0) = v_2(x_0)$, we have the following estimate:*

$$\|v_1(x) - v_2(x)\|^2_{L^2(\Omega)} \leq \Lambda \int_0^T \|\widetilde{u}_1(T) - \widetilde{u}_2(T)\|_{L^2(\Omega)}\,dt,$$

where the constant Λ depends only on Ω and ε.

Proof. In estimate (3.3) of Lemma 3.1, we take $\kappa = v_2$ and $u_0^\star = v_1$ and also $\kappa = u_0^1$ and $u_0^2 = u_0^\star$. We get

$$\int_0^T\int_0^1 w_1(x,T)\big(u_1^\star(x,T) - \widetilde{u}_1(T)\big)\,dx\,dt + \varepsilon\int_0^1 v_1(v_2 - v_1)\,dx \geq 0 \tag{3.7}$$

and

$$\int_0^T\int_0^1 w_2(x,T)\big(u_2^\star(x,T) - \widetilde{u}_2(T)\big)\,dx\,dt + \varepsilon\int_0^1 v_2(v_1 - v_2)\,dx \geq 0, \tag{3.8}$$

where $\{u_i^\star; w_i\}$ $(i = 1,2)$ are the solutions of systems (3.1) and (3.2) with the initial condition $u_0^\star = v_i$ $(i = 1,2)$, respectively. Setting

$$U = u_1^\star - u_2^\star, \ \ W = w_1 + w_2,$$

and taking $\kappa = v_2$ and $\kappa = v_1$, U and W satisfy

$$\begin{cases} U_t(x,t) - U_{xx}(x,t) - a(x)\int_0^t U(x,s)\,ds = 0 & \forall\,(x,t) \in Q, \\ U(0,t) = U(1,t) = 0 & \forall\, t \in (0,T)\,, \\ U(x,0) = v_1 - v_2 & \forall\, x \in (0,1)\,, \end{cases} \tag{3.9}$$

$$\begin{cases} W_t(x,t) - W_{xx}(x,t) - a(x)\int_0^t W(x,s)\,ds = 0 & \forall\,(x,t) \in Q, \\ W(0,t) = W(1,t) = 0 & \forall\, t \in (0,T), \\ W(x,0) = 0 & \forall\, x \in (0,1), \end{cases} \tag{3.10}$$

By the extremum principle, we know that (3.10) has only a zero solution and thus

$$w_1(x,t) = -w_2(x,t). \tag{3.11}$$

Moreover, w_1 satisfies the following equation:

$$\begin{cases} (w_1)_t(x,t) - (w_1)_{xx}(x,t) = a(x)\displaystyle\int_0^t w_1(x,s)\,ds & \forall\,(x,t)\in Q, \\ w_1(0,t) = w_1(1,t) = 0 & \forall\,t\in(0,T)\,, \\ w_1(x,0) = v_2 - v_1 & \forall\,x\in(0,1)\,, \end{cases} \tag{3.12}$$

Due to (3.9) and (3.12), we have

$$U(x,t) = -w_1(x,t). \tag{3.13}$$

From (3.7), (3.8), (3.11) and (3.13), we obtain

$$\begin{aligned} &\varepsilon\int_0^1 |v_1(x)-v_2(x)|^2\,dx \\ &\quad\le \int_0^T\!\!\int_0^1 w_1(x,T)\big(u_1^\star(x,T)-\widetilde{u}_1(T)\big)\,dx\,dt + \int_0^T\!\!\int_0^1 w_2(x,T)\big(u_2^\star(x,T)-\widetilde{u}_2(T)\big)\,dx\,dt \\ &\quad\le \int_0^T\!\!\int_0^1 U(x,T)w_1(x,t)\,dx\,dt + \int_0^T\!\!\int_0^1 (\widetilde{u}_2(T)-\widetilde{u}_1(T))w_1(x,T)\,dx\,dt \\ &\quad\le -\int_0^T\!\!\int_0^1 |w_1(x,t)|^2\,dx\,dt + \frac{1}{2}\int_0^T\!\!\int_0^1 |w_1(x,t)|^2\,dx\,dt + \frac{1}{2}\int_0^T\!\!\int_0^1 |\widetilde{u}_1(T)-\widetilde{u}_2(T)|^2\,dx\,dt \\ &\quad\le -\frac{1}{2}\int_0^T\!\!\int_0^1 |w_1(x,t)|^2\,dx\,dt + \frac{1}{2}\int_0^T\!\!\int_0^1 |\widetilde{u}_1(T)-\widetilde{u}_2(T)|^2\,dx\,dt. \end{aligned} \tag{3.14}$$

Then

$$\|v_1(x)-v_2(x)\|^2_{L^2(\Omega)} \le \Lambda\int_0^T \|\widetilde{u}_1(T)-\widetilde{u}_2(T)\|^2_{L^2(\Omega)}\,dt, \tag{3.15}$$

with $\Lambda = \frac{1}{2\varepsilon}$. □

Remark 3.1. From Theorem 3.1 we can easily deduce that if the final measurements of systems (1.6) and (3.1) are equal, then the data u_0 can be determined uniquely almost everywhere.

4 Numerical experiments

Consider the objective function

$$\begin{aligned} \widetilde{J}(y) &= \frac{1}{2}\|A(y)-f(x,t)\|^2_{L^2(\Omega)} + \frac{1}{2}\|y(x=0)\|^2_{L^2(0,T)} \\ &\quad + \frac{1}{2}\|y(x=1)\|^2_{L^2(0,T)} + \frac{1}{2}\|y(t=T)-u^{obs}\|^2_{L^2(\Omega)} + \frac{\varepsilon}{2}\|y(t=0)-u^b\|^2_{L^2(\Omega)}. \end{aligned} \tag{4.1}$$

The *DGM* algorithm approximates $u(t,x)$ with a deep neural network $y(t,x;\theta)$ where $\theta\in\mathbb{R}^k$ are the neural network's parameters. The goal is to find a set of parameters θ such that the function $y(t,x;\theta)$ minimizes the error $\widetilde{J}(y)$. If the error $\widetilde{J}(y)$ is small, then $y(t,x;\theta)$ will closely satisfy the PDE differential operator, boundary conditions and initial condition. Therefore, θ minimizing $\widetilde{J}(y(\cdot;\theta))$ produces a reduced-form model $y(t,x;\theta)$ which approximates the PDE solution $u(t,x)$.

In this sense, we recall the following

Theorem 4.1 ([19]). *Let the L^2 error $J(f)$ measure how well the neural network f satisfies the differential operator, boundary, initial and observability conditions.*

Define $\mathbb{C}^n$ as the class of neural networks with n hidden units and let f^n be a neural network with n hidden units which minimizes $J(f)$.

There exists $f^n \in \mathbb{C}^n$ such that $J(f^n) \to 0$ as $n \to \infty$, and $f^n \to u$ as $n \to \infty$.

Network architecture:

The first layer and the last of this neural network are fully connected. The rest is made up of *GRU* cells [10], which is a simplified version of the LSTM cell (Figure 1):

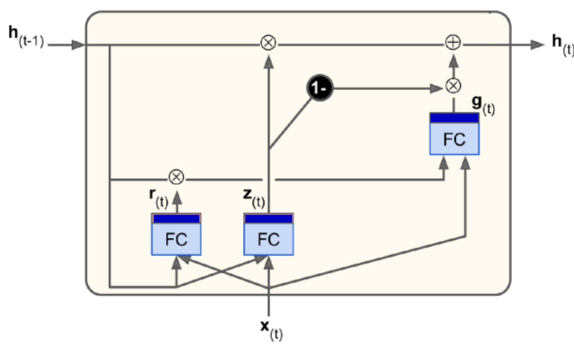

Figure 1: GRU cell

We have found the following network architecture:

$$\begin{aligned}
z_{(t)} &= \tan h\big(\theta_{xz}^T x_{(t)} + \theta_{hz}^T h_{(t-1)} + b_z\big),\\
r_{(t)} &= \tan h\big(\theta_{xr}^T x_{(t)} + \theta_{hr}^T h_{(t-1)} + b_r\big),\\
g_{(t)} &= \tan h\big(\theta_{xg}^T x_{(t)} + \theta_{hg}^T (r_{(t)} \otimes h_{(t-1)}) + b_g\big),\\
h_{(t)} &= z_{(t)} \otimes h_{(t-1)} + (1 - z_{(t)}) \otimes g_{(t)}.
\end{aligned}$$

The full algorithm is:

Algorithm 1. *DL-IP Algorithm:*

(1) *Define the boundary conditions,*

(2) *Define the architecture of neutral networks by setting the number of layers, number of neurons in each layer and activation functions,*

(3) *Generate random training set D_M,*

(4) *Initialize the parameter set θ_0 and the learning rate α_0,*

(5) *Repeat until convergence criterion is satisfied*

   1. *Randomly sample a mini-batch d_m of training examples from D_M,*
   2. *Compute the loss functional for the sampled mini-batch d_m:*

$$\widetilde{J}(\theta_n, d_m), \tag{4.2}$$

3. *Compute the gradient* $\nabla_{\theta_n} \widetilde{J}(\theta_n, d_m)$ *for the sampled mini-batch* d_m *using backpropagation,*
4. *Use the estimated gradient to take a descent step at* d_m *with learning rates to update* θ_{n+1}*:*

$$\theta_{n+1} = \theta_n - \alpha_n \nabla_{\theta_n} \widetilde{J}(\theta_n, d_m)$$

The parameters are updated by using the well-known ADAM algorithm with a decaying learning rate schedule.

(6) *Save the model to be used for any* $x \in \Omega$ *and* $t \in]0, T[$.

The derivatives of $y(t, x; \theta)$ can be evaluated by using automatic differentiation since it is parametrizing as a neural network.

We implement the algorithm using TensorFlow, which is software libraries for deep learning. TensorFlow has reverse mode automatic differentiation which allows calculation of derivatives for a broad range of functions. For example, TensorFlow can be used to calculate the gradient of our neural network with respect to x or t, or θ. TensorFlow also allows the training of models on graphic processing units (GPUs).

4.1 The noise resistance of the proposed method

The data u^b and u^{obs} are assumed to be corrupted by measurement errors, which we will refer to as noise. In particular, we suppose that

$$u^b = u^{exact}(t = 0) + e \ \text{ and } \ u^{obs} = u^{exact}(t = T) + e^{obs}.$$

Let

$$err = \frac{\|e\|_2}{\|u^{exact}(t = 0)\|_2} \ \text{ and } \ err^{obs} = \frac{\|e^{obs}\|_2}{\|u^{exact}\|_2}.$$

To study the impact of err and err^{obs} on the construction of initial state, we consider $u^{exact} = \frac{x(x-1)}{T+t}$ and perfom two tests. First, we suppose $err^{obs} = 0$ and $err \in \{0\%; 3\%; 5\%\}$ and then we take $err = 0$ and $err^{obs} \in \{3\%; 5\%; 10\%\}$.

For the applications, we have found that the hyperparameter $L = 4$ (i.e., four hidden layers) is effective. The neural network parameters are initialized by using the keras.initializers.glorot_normal initialization .

In all tests below, the constructed state is drawing in red and the exact in blue. After 1000 epoch, we have found the following results.

4.1.1 Impact of err on construction of initial state

The tests (Figures 2 to 10) show that the proposed algorithm is uniformly stable to noise.

Case $err = 0\%$**.**

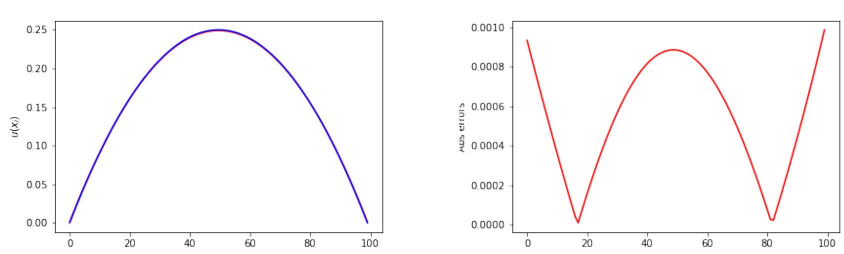

Figure 2: u_0 constructed and u_0^{exact} (left), absolute errors between exact and predicted initial states (right).

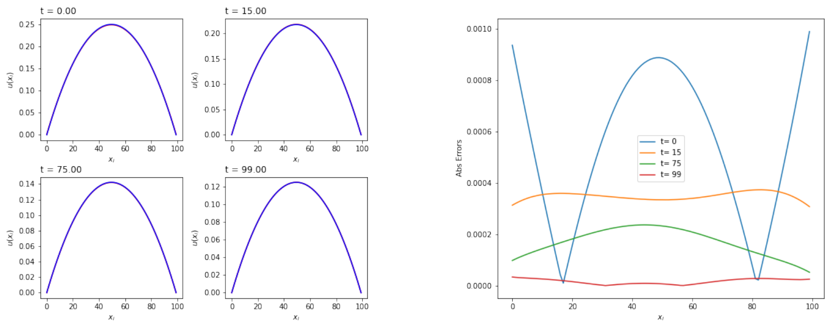

Figure 3: u constructed and u^{exact} (left), absolute error construction (right).

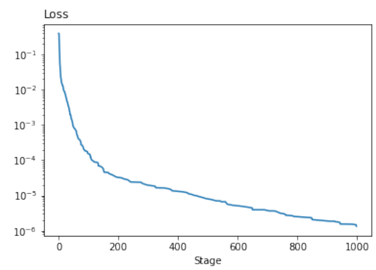

Figure 4: Value of the cost J. This figure shows that J converges to 0.

Case $err = 3\%$***.***

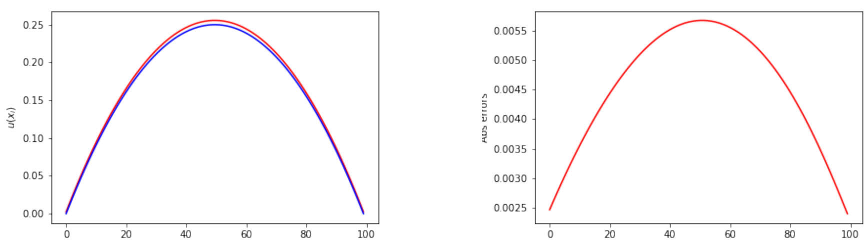

Figure 5: u_0 constructed and u_0^{exact} (left), absolute errors between exact and predicted initial states (right).

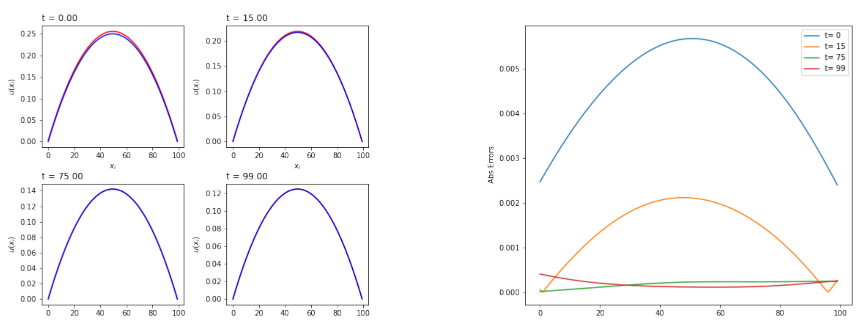

Figure 6: u constructed and u^{exact} (left), absolute error construction (right).

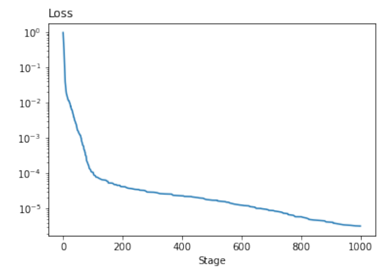

Figure 7: Values of the cost J.

Case $err = 5\%$***.***

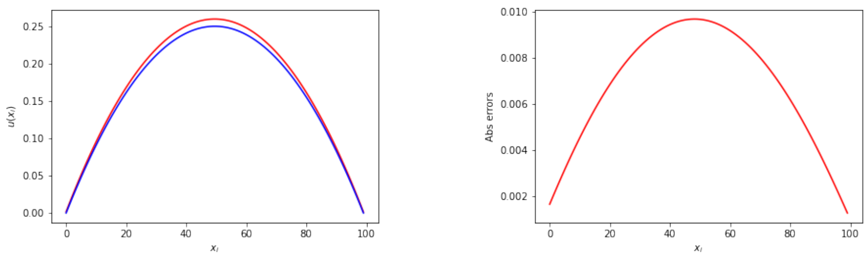

Figure 8: u_0 constructed and u_0^{exact} (left), absolute errors between exact and predicted initial states (right).

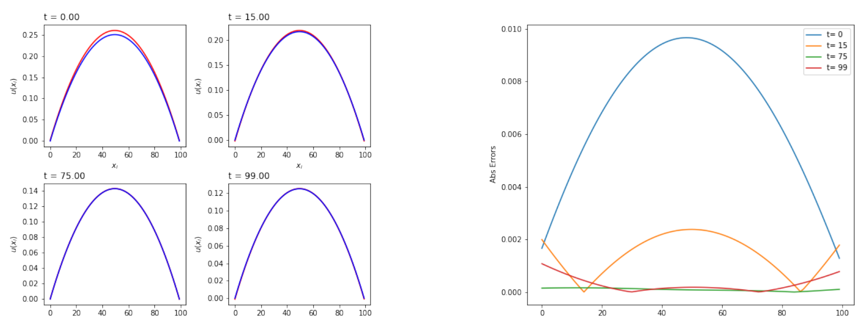

Figure 9: u constructed and u^{exact} (left), absolute error construction (right).

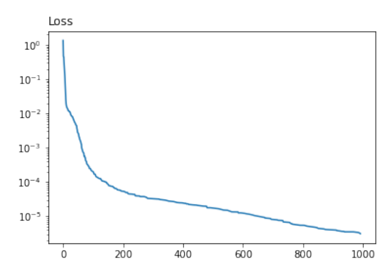

Figure 10: Values of the cost J.

4.1.2 Impact of err^{obs} on the construction of the initial state

The tests (Figures 11 to 16) show that the proposed algorithm is uniformly stable to observation noises. We observe that err^{obs} has an effect on the construction of final state and hasn't impact on the initial condition.

Case $err^{obs} = 3\%$.

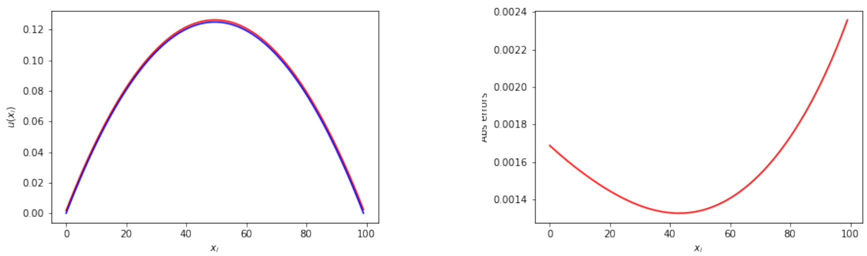

Figure 11: $u(t = T)$ constructed and $u^{exact}(t = T)$ (left), absolute errors between exact and predicted state at $t = T$ (right).

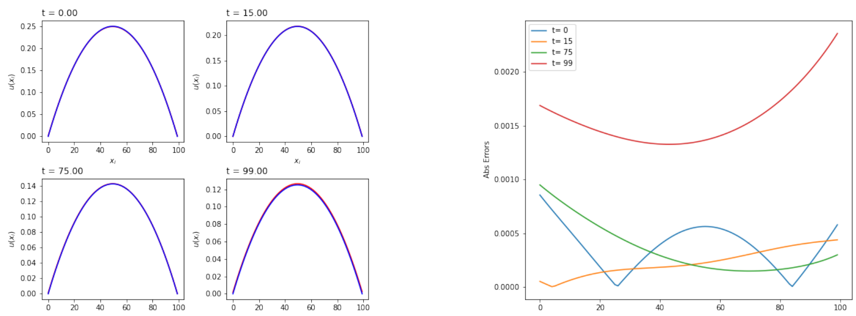

Figure 12: u constructed and u^{exact} (left), absolute error construction (right).

Case $err^{obs} = 5\%$.

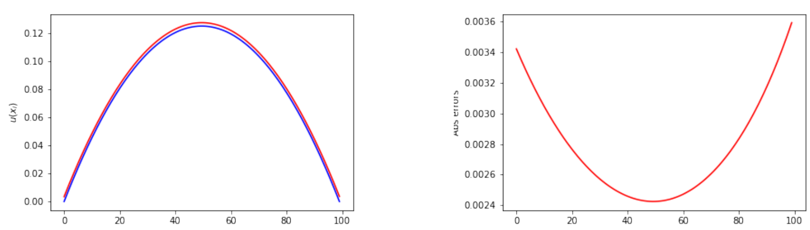

Figure 13: $u(t = T)$ constructed and $u^{exact}(t = T)$ (left), absolute errors between exact and predicted state at $t = T$ (right).

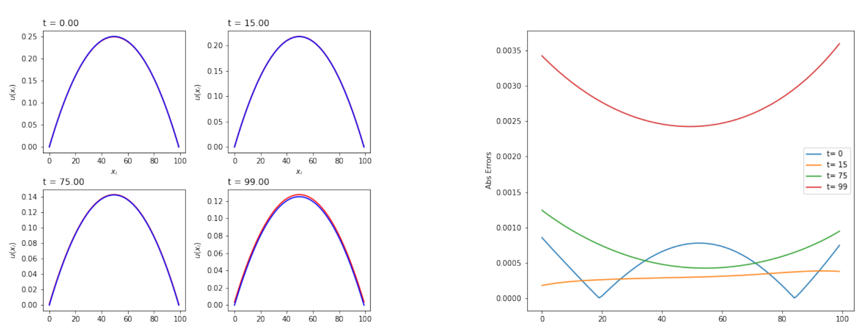

Figure 14: u constructed and u^{exact} (left), absolute error construction (right).

Case $err^{obs} = 10\%$.

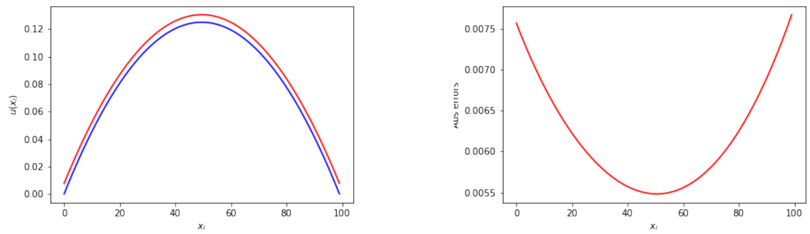

Figure 15: $u(t = T)$ constructed and $u^{exact}(t = T)$ (left), absolute errors between exact and predicted state at $t = T$ (right).

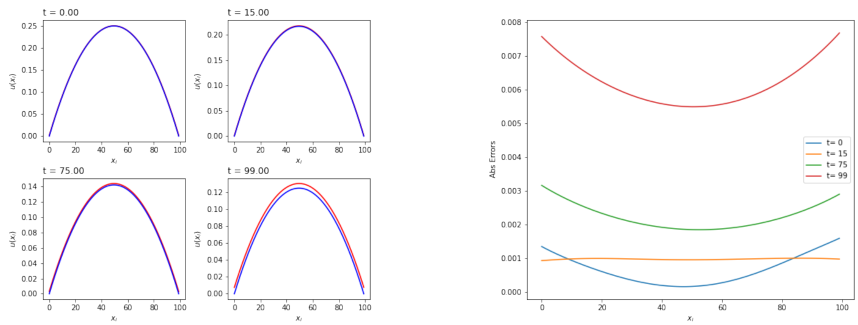

Figure 16: u constructed and u^{exact} (left), absolute error construction (right).

4.1.3 Other tests

Case $u^{exact} = (x - 1)(t + 1)(e^x - 1)$.

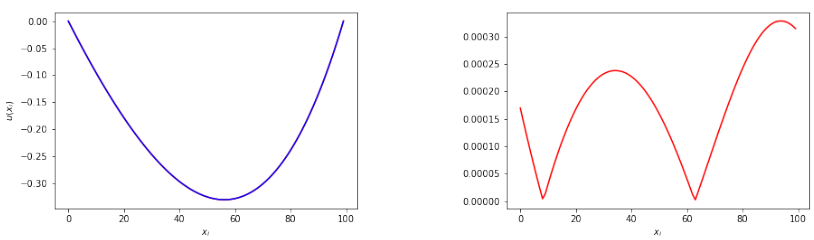

Figure 17: u_0 constructed and $u^{exact}(t = 0)$ (left), absolute errors between exact and predicted initial state (right).

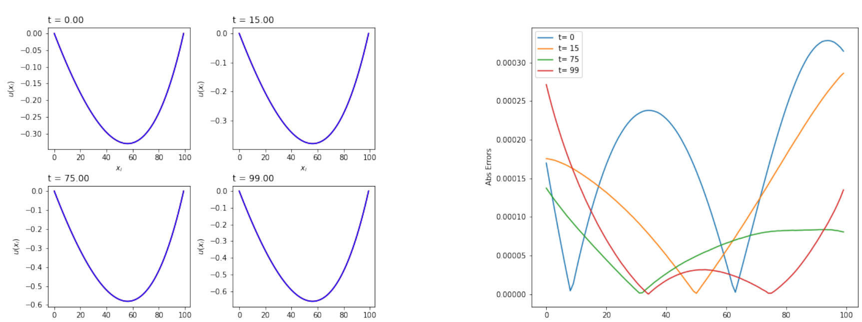

Figure 18: u constructed and u^{exact} (left), absolute error construction (right).

Case $u^{exact} = (x-1)(t+1)(\cos(x)-1)$.

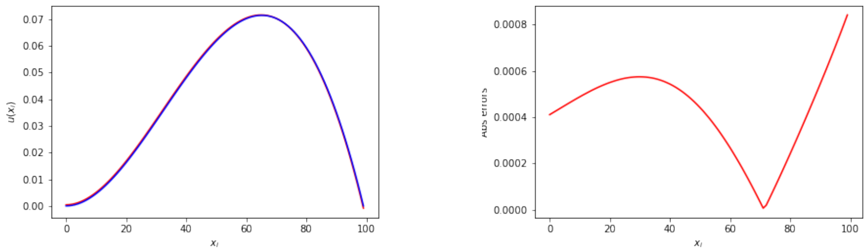

Figure 19: u_0 constructed and $u^{exact}(t=0)$ (left), absolute errors between exact and predicted initial state (right).

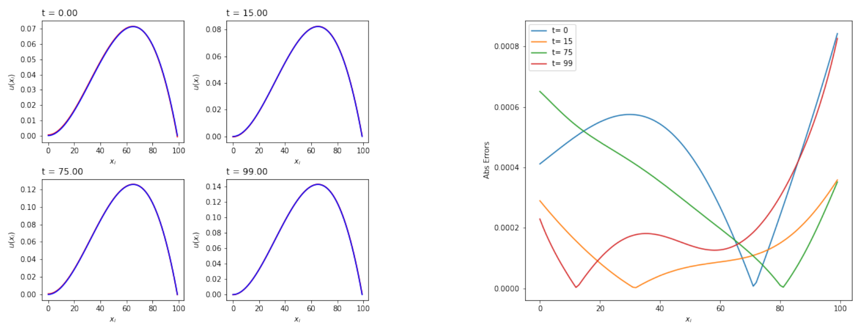

Figure 20: u constructed and u^{exact} (left), absolute error construction (right).

5 Conclusion

Our deep learning algorithm for solving PDEs is meshfree, which is a key since meshes become infeasible in higher dimensions. Instead of forming a mesh, the neural network is trained on batches of randomly sampled time and space points.

Moreover, deep learning algorithm doesn't have a rounding errors caused by the discretization which plays a very important roll for the construction of a solution.

The ease of implementing a deep learning algorithm and the independence of this algorithm to EDP equations make the method very efficient and easy to handle.

But there remains the problem of parametrization of a deep learning algorithm. The choice of a number of layers and a number of units in each layer turns out to be very important to have a good approximation of the solution.

References

[1] B. Allal, G. Fragnelli and J. Salhi, Null controllability for the singular heat equation with a memory term. *Preprint* arXiv:2005.04442; `https://arxiv.org/abs/2005.04442`.

[2] K. Atifi, I. Boutaayamou, H. Ould Sidi and J. Salhi, An inverse source problem for singular parabolic equations with interior degeneracy. *Abstr. Appl. Anal.* **2018**, Art. ID 2067304, 16 pp.

[3] Kh. Atifi, E.-H. Essoufi and B. Khouiti, Null controllability of degenerate/singular parabolic equations with degeneracy and singularity occurring in the interior of the spatial domain. *Eurasian Journal of Mathematical and Computer Applications* **8** (2020), no. 1, 4–29.

[4] S. A. Avdonin, S. A. Ivanov and J. M. Wang, Inverse problems for the heat equation with memory. *Preprint* arXiv:1612.02129; `https://arxiv.org/abs/1612.02129`.

[5] C. Cavaterra and D. Guidetti, Identification of a convolution kernel in a control problem for the heat equation with a boundary memory term. *Ann. Mat. Pura Appl. (4)* **193** (2014), no. 3, 779–816.

[6] C. Cavaterra, A. Lorenzi and M. Yamamoto, A stability result via Carleman estimates for an inverse source problem related to a hyperbolic integro-differential equation. *Comput. Appl. Math.* **25** (2006), no. 2-3, 229–250.

[7] A. M. Denisov and T. S. Shores, An inverse coefficient problem for an integro-differential equation. *Appl. Anal.* **81** (2002), no. 3, 725–752.

[8] D. K. Durdiev and Zh. Zh. Zhumaev, One-dimensional inverse problems of finding the kernel of integrodifferential heat equation in a bounded domain. *Ukrainian Math. J.* **73** (2022), no. 11, 1723–1740.

[9] A. Fahim, M. A. Fariborzi Araghi, Numerical solution of convection-diffusion equations with memory term based on sinc method. *Comput. Methods Differ. Equ.* **6** (2018), no. 3, 380–395.

[10] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow.* O'Reilly Media, Inc., 2022.

[11] H. Grabmueller, Linear Theorie der Waermeleitung in Medium mit Gedaechtnis; Existenz und Eindeutigkeit von Loesungen sum Inversen Problem. *Preprint* 226, 1975.

[12] J. Janno and L. von Wolfersdorf, An inverse problem for identification of a time- and space-dependent memory kernel of a special kind in heat conduction. *Inverse Problems* **15** (1999), no. 6, 1455–1467.

[13] Y. Jiang, J. Fan, S. Nagayasu and G. Nakamura, Local solvability of an inverse problem to the Navier-Stokes equation with memory term. *Inverse Problems* **36** (2020), no. 6, 065007, 14 pp.

[14] K. Karuppiah, J. K. Kim and K. Balachandran, Parameter identification of an integrodifferential equation. *Nonlinear Funct. Anal. Appl.* **20** (2015), no. 2, 169–185.

[15] A. Lorenzi, An inverse problem in the theory of materials with memory, II. *Semigroup theory and applications (Trieste, 1987)*, 261–290, Lecture Notes in Pure and Appl. Math., 116, *Dekker, New York*, 1989.

[16] A. Lorenzi and G. Mola, Identification of unknown terms in convolution integro-differential equations in a Banach space. *J. Inverse Ill-Posed Probl.* **18** (2010), no. 3, 321–355.

[17] A. Lorenzi and E. Sinestrari, An inverse problem in the theory of materials with memory. *Nonlinear Anal.* **12** (1988), no. 12, 1317–1335.

[18] S. Shahid and S. Arshed, Numerical solution of convection-diffusion integro-differential equations with a weakly singular kernel. *Journal of Basic and Applied Scientific Research* **3** (2013), no. 11, 106–120.

[19] J. Sirignano and K. Spiliopoulos DGM, A deep learning algorithm for solving partial differential equations. *ArXiv e-prints*, 2017.

[20] A. N. Tikhonov, On the solution of ill-posed problems and the method of regularization. (Russian) *Dokl. Akad. Nauk SSSR* **151** (1963), 501–504.

[21] Zh. D. Totieva, Determining the kernel of a viscoelasticity equation in a weakly horizontally inhomogeneous medium. (Russian) *Sibirsk. Mat. Zh.* **61** (2020), no. 2, 453–475; translation in *Sib. Math. J.* **61** (2020), no. 2, 359–378.

[22] Zh. D. Totieva, Linearized two-dimensional inverse problem of determining the kernel of the viscoelasticity equation. (Russian) *Vladikavkaz. Mat. Zh.* **23** (2021), no. 2, 87–103.

[23] L. von Wolfersdorf, On identification of memory kernels in linear theory of heat conduction. *Math. Methods Appl. Sci.* **17** (1994), no. 12, 919–932.

[24] M. F. Zedan and C. Dalton, Higher-order axial singularity distributions for potential flow about bodies of revolution. *Comput. Methods Appl. Mech. Eng.* **21** (1980), no. 3, 295–314.

(Received 11.05.2022; revised 13.11.2022; accepted 15.11.2022)

Authors' addresses:

Soufiane Abid
Laboratoire de Mathématiques, Informatique et Sciences de l'ingénieur (MISI), Université Hassan 1, Settat 26000, Morocco.
E-mail: s.abid@uhp.ac.ma

Khalid Atifi
Laboratoire de Mathématiques, Informatique et Sciences de l'ingénieur (MISI), Université Hassan 1, Settat 26000, Morocco.
E-mail: k.atifi.uhp@gmail.com

El-Hassan Essoufi
Laboratoire de Mathématiques, Informatique et Sciences de l'ingénieur (MISI), Université Hassan 1, Settat 26000, Morocco.
E-mail: e.h.essoufi@gmail.com