SAPIENZA
UNIVERSITÀ DI ROMA

Department of Basic and Applied Sciences for Engineering

PhD thesis in Mathematical Models for Engineering, Electromagnetism and Nanosciences

Carlo Alberini

Finite difference methods for degenerate diffusion equations and fractional diffusion equations

Advisor: Prof. Maria Agostina Vivaldi

Academic Year 2019-2020

Contents

List of Figures

List of Tables

Introduction

This thesis focuses on the study of three mathematical models consisting of degenerate diffusion equations and fractional diffusion equations arising from different study needs: the first one is taken from the study of self-organized criticality phenomena; the second one is connected to the obstacle problem, while the third one, that is a time-fractional type model, can find a wider use, for example, from biology to mechanics, to superslow diffusion in porous media, till financial type phenomena.

More precisely, in the first chapter of the present work, it will be described a numerical implementation of a differential model for the simulation of self-organized criticality (SOC) phenomena arising from recent papers by Barbu [11, 12], i.e.

$$\begin{cases} u_t - \Delta H(u(t) - u^c) \ni 0 & \text{in } \Omega \times (0,\infty) \\ u(0) = u^0 > u^c & \text{in } \Omega \\ 0 \in H(u(t) - u^c) & \text{on } \partial\Omega \times (0,\infty) \end{cases} \tag{1}$$

where Ω is a bounded subset of $\mathbb{R}^2$, $u^c \in C^0(\overline{\Omega})$ a given target function (called *critical state*) and u^0 a supercritical initial datum, while H is the multivalued Heaviside function:

$$H(r) = \begin{cases} 1 & \text{if } r > 0 \\ [0,1] & \text{for } r = 0 \\ 0 & \text{if } r < 0. \end{cases} \tag{2}$$

We immediately state that the complete and appropriate conditions under which problem (1) is studied, through the aforementioned papers of Barbu, will be reported in the first chapter of the present work, together with a brief discussion that will be aimed to showing the origins of the problem itself and its intrinsic mathematical structure, (see Sections 1.2 and 1.3).

However we can argue by saying that in that singular nonlinear diffusion problem the initial supercritical state evolves in a finite time towards the given critical solution, progressively from the boundary towards the internal regions. The key elements are the Heaviside function which plays the role of a switch for the dynamics, and the initial boundary contact with the critical state.

A finite difference implicit scheme on a fixed grid will be proposed in Section 1.4.2.1 for a regularized version of the problem, with the Heaviside replaced by a C^1 function, showing the same behavior of the solution: convergence in finite time toward the critical state on every single node, up to any prescribed accuracy, remaining supercritical during all the process, (related results will be presented in Section 1.4.1).

We will also implement this regularized version of the problem (1) through the use of synchronized spatial-temporal grids with progressive refinements (in the spirit of [46], see Section 1.4.2.2) simulates the appearance of short-range interactions of an increasing number of particles, speeding up the convergence to the critical solution and allowing a strong reduction of computational cost. The results of some numerical simulations are discussed in one and two dimensions, and they give the evidence that the temporal evolution of the solution u is characterized just by a progressive alignment to the target function u^c starting from the boundary of Ω, and proceeding towards the more internal values.

Aim of the second chapter of the present work, instead, is to study the following problem:

$$\begin{cases} u_t - H\left(u-u^c\right)\left(\Delta u + f\right) = 0 & \text{a.e. in } \Omega, \text{ for all } t \in (0,T) \\ u\left(0\right) = u^0 & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega, \text{ for all } t \in (0,T) \end{cases} \tag{3}$$

with $T > 0$, where H is the extended Heaviside function such that $H(0) = 0$, that is

$$H(r) = \begin{cases} 1 & \text{for } r > 0 \\ 0 & \text{for } r \leq 0 \end{cases} \tag{4}$$

and Ω is a bounded domain in $\mathbb{R}^n$, $n = 1, 2$ with smooth boundary.

In this case, we will assume that the initial datum u^0, the (independent of time) source term f and the given target function (the obstacle) u^c satisfy the following conditions

$$u^0 \in H_0^1(\Omega)\,,\ f \in L^2(\Omega)\,,\ u^c \in H^2(\Omega)\,,\ u^c \leq 0 \text{ on } \partial\Omega. \tag{5}$$

and we will define as solution of problem (3) a function $u \in L^2\left(0, T; H_0^1(\Omega) \cap H^2(\Omega)\right)$, with $u_t \in L^2\left(0, T; L^2(\Omega)\right)$, which satisfies the equations in (3).

Under these assumptions and suitable conditions (see $\mathbf{H}_1$ and $\mathbf{H}_2$ explained in Section 2.1), we will able to show that the nonlinear degenerate parabolic problem (3), whose diffusion coefficient is now represented by the Heaviside function of the distance of the solution itself from a given target function, behaves as an evolutive variational inequality having the target as an obstacle: in other words, we will able to show that, under the aforementioned hypotheses, starting from an initial state above the target, the solution evolves in time towards an asymptotic solution, eventually getting in contact with part of the target itself. Finally, also for this model, through a finite difference approach, we will study the behavior of the solution of this problem, using in this case both the exact Heaviside function and a regular approximation of it, showing the results in some numerical tests, (see Section 2.2).

In the third chapter, at last, we will focus our attention on the following problem, which generalizes the one addressed in the second chapter, for $0 < \alpha < 1$:

$$\begin{cases} \partial_t^\alpha u - H(u - \psi)\,\Delta u = 0 & \text{a.e. in } \Omega, \text{ for all } t \in (0, T) \\ u(0) = u^0 & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega, \text{ for all } t \in (0, T) \end{cases} \tag{6}$$

with $T > 0$, where H is, still, the extended Heaviside function (4) such that $H(0) = 0$ and Ω is a bounded domain in $\mathbb{R}^n$, $n = 1, 2$ with smooth boundary. Only for simplicity, we omitted in this case the presence of a forcing term in the problem.

Let us also precise that here $\partial_t^\alpha u$ denotes the Caputo time-fractional derivative, that is,

$$\partial_t^\alpha u(x, t) := \frac{1}{\Gamma(1-\alpha)} \int_0^t (t - s)^{-\alpha} \partial_s u(x, s)\ \mathrm{d}s \tag{7}$$

where Γ is the Gamma function.

About the set of the assumptions adopted for this topic, we will define as solution of problem (6) a function $u \in L^2\left(0,T; H_0^1\left(\Omega\right) \cap H^2\left(\Omega\right)\right)$, with $\partial_t^\alpha u(x,t) \in L^2\left(0,T; L^2(\Omega)\right)$, which satisfies the equations in (6). Also for this model, it will be possible to prove the equivalence to the fractional parabolic obstacle problem, showing that its solution evolves for any $\alpha \in (0,1)$ to the same stationary state, the solution of the classic elliptic obstacle problem. The only thing which changes with α is the convergence speed (see Section 3.1).

Later, we will also study this problem from the numerical point of view, comparing some finite different approaches, and showing the results of some tests. We remark that these results extend what we will prove in the second chapter, that coincides with the case $\alpha = 1$, (see Section 3.2.1).

We finally point out that the common feature of these three models lies in the different use made of the Heaviside (multi)function: in the formulation of evolutive differential problems is useful when a discontinuous behavior can occur according to the specific values of the solution itself: such a function acts as a dynamic switch for this behavior. It can be applied to the differential operator itself (for example, here, the Laplacian), giving rise to nonlocal phenomena: examples of that kind can be found in papers related to self-organized criticality such as the sandpile model (see e.g.[11, 12], [46], [47]). In these cases the Heaviside function, calculated on the distance between the solution and an assigned critical state, is able to govern the spread of the problem on a global level: the initial data tend to the critical state progressing from the edge towards the interior of the domain during the time evolution, which stops when all the solution gets in contact with the threshold critical state.

At last, referring to this particular behavior, it is interesting to note that the three typical aspects that we can find in general in every sandpile model, i.e. local equilibrium, threshold activation and diffusive character of the updating rules which cause the solution of (1) in the above settings to evolve in time towards the critical state, can be also shared by simple biological models of contamination and epidemic spreading, where the Heaviside term acts, also in this case, within the Laplace operator, as a switch of the process itself (see e.g. [37]).

Moreover, in the second example of use of the Heaviside function, i.e. as a degenerate diffusion coefficient applied externally to the Laplace operator, calculated between the solution and the assigned critical state, it is able to locally control the diffusivity of the process in any areas where contact between solution and obstacle is reached.

Other examples of this approach can be found in the literature even without the Heaviside function, for example in cases where the model changes its behavior in a discontinuous way according to the values of the solution or of its partial derivatives: in these cases the Heaviside function could be replaced by the positive part function, $\sigma \to (\sigma)_+ := \max\{\sigma, 0\}$, for all $\sigma \in \mathbb{R}$, to describe non-reversible phenomena that can be still connected to the avalanche behavior in the sandpile models [40] and to damage mechanics models [1].

Chapter 1

A nonlinear diffusion model for self-organized criticality phenomena

In this chapter, the basis theory behind SOC models will be analyzed, i.e. we will focus our attention above all those physical dynamic systems that spontaneously are able to rearrange themselves, in a *finite* time, from an any unstable configuration to a stable time-independent one, through the so-called avalanches phenomena.

In the meanwhile, we will also compare the two aforementioned scientific articles: the first one by V. Barbu (*Self-organized criticality of cellular automata model; absorbtion in finite-time of supercritical region into the critical one*, (2013) Math. Methods Appl. Sci.), [12] and the latter one by U. Mosco (*Finite-time Self-Organized-Criticality on synchronized infinite grids*, (2018) SIAM J. Math. Anal., Vol. 50, No. 3), [46], since they underlie, as said, some of the research results presented in the following Sections of this first chapter.

1.1 Preliminaries

A nice and suitable point of view for introducing the concepts described below could be to focus our attention on that part of real dynamic physical phenomena which have, among their main features, that of being *dissipative*. If analyzed in depth, a physical system of this kind tends to minimize its dissipations in order to stabilize itself around what it could be defined as a *minimally stable state*, rather than around other absolute stability criteria (like thermal equilibrium, for example).

This characteristic becomes extremely interesting to analyze when the physical system is composed of a large number of particles. In this case, in order to reach this minimally stable state, entire areas of the system can be observed, even experimentally, influencing each other thanks to the interactions due to the mutual presence of the particles of the system itself. So, from this point of view, it is important to underlie that both the minimally stable state is continuously influenced by these (even minimal) perturbations occurring between the particles in question and that these perturbations, even observed in nature, always occur at a short distance between the particles, but at a high frequency between them.

Finally, again for such physical systems, these continuous movements of particles generate what it is defined as *avalanches*, which tend to change their distribution over time within the physical system itself, until a stable configuration (i.e. the minimally stable state) is reached in a finite time. These avalanches of particles, during the temporal evolution of the system, can change their size in such a way as to dissipate the internal energy of the system, but, at the same time, prevent this dissipation from happening *too quickly*. This dynamic continues until these avalanches end, and this happens when the whole physical system has reached its (new) minimally stable state that, now, we can define as a *critical state*. *Critical* here means that this state is time-independent, but any (even minimal) perturbation of particles from this state moves the system toward *another* critical configuration through the production of *other* avalanches, but always in a finite time.

Typical examples of these physical systems are the mathematical models that describe the evolution of sandpiles over time. So, these models became the prototype for the study of many complex system dynamics where the current state evolves *spontaneously* towards a critical state in a finite time.

As mentioned above, the grains of sand (identifiable with the *degrees of freedom* of the system) of these models do not interact with each other at great distances, but are able to create avalanches of any size whenever, through the evolution of the system, a critical slope is reached locally within the sandpile itself.

It is interesting to note that the distribution of these avalanches within the sandpile follows a power law, while the character that emerges spontaneously from the evolution mechanism of the system itself is called *self-organized criticality* (SOC below), given its peculiar characteristic of being able to rearrange itself in a time-dependet evolution that spontaneously brings it to a certain time-independet critical state. It is important to underlie that this evolution always last in a finite time. This critical state, that coincides with the minimally stable state defined above, can be thought as an *attractor* for all the dynamic. For a more detailed discussion, see [5, 6, 7] and [8].

The fundamentals of the SOC theory date back to the late 1980s, when Bak, Tang and Wiesenfeld [6, 7] introduced the sandpile cellular automata model in order to analyze the time behavior of avalanches on a $N \times N$ plane lattice. In the model when the local height h_{ij} of the sandpile at the ij-site reaches a prescribed critical threshold h^c it becomes unstable, yielding the toppling of grains on the four adjacent sites. This automata dynamics was formalized by Dahr [27], who introduced the topplig matrix D: after a toppling in the $k\ell$-site, the height h_{ij} of the sandpile at any site changes according to:

$$h_{ij}^{t+1} = h_{ij}^{t} - D_{ij,k\ell} \tag{1.1}$$

where

$$D_{ij,k\ell} = \begin{cases} 4 \text{ if } ij = k\ell \\ -1 \text{ if } ij \text{ and } k\ell \text{ adjacent sites} \\ 0 \text{ otherwise.} \end{cases} \tag{1.2}$$

By this rule subsequent topplings can occur, generating avalanches which end as soon as stability is reached again over all the lattice sites ($h_{ij} < h^c$ at any site). In compact form the toppling law can be interpreted as an implicit in time nonlinear finite difference system

$$h^{t+1} - h^{t} = DH(h^{t+1} - h^{c}) \tag{1.3}$$

for the the vector of the heights h and the matrix D, and H is the standard Heaviside function:

$$H(r) = \begin{cases} 1 & \text{if } r \geq 0 \\ 0 & \text{if } r < 0. \end{cases} \tag{1.4}$$

Remark 1.1.1. *The dynamic arising in* (1.1) *is directly connected to the equation* (1.3) *that describes the only way to have a sand transfer to a site (activated) to another. In this last equation we have solution not equal to zero if and only if we are in the critical region, or above it, i.e. in the supercritical region. In all other cases the site it must be considered unchanged (stable).*

Note that the matrix D recalls the well-known tridiagonal block matrix coming from the 5-point finite difference Laplacian approximation. It is henceforth not surprising that Carlson and Swindle [24] were able to characterize the continuum limit of the cellular automaton proposed by Bak, Tang and Wiesenfeld as the solution $u(t)$ of the following singular diffusion equation

$$u_t = \Delta H(u(t) - u^c) \tag{1.5}$$

with u^c a given critical state which plays the role of the threshold. Then the evolution of the system toward the equilibrium can be described by two distinct time scales, a low one far from the critical state and a fast one in its neighborhood (corresponding to the avalanche process).

Due to the discontinuity of the Heaviside function, the ordinary existence results are not applyable to an equation as (1.5). That is why the multivalued setting was necessary to prove in general that the solution of such a problem exists and evolves spontaneously in time towards the critical state u^c from above; in other words, according to this assumption, it will be possible to demonstrate that the supercritical region is absorbed into the critical one in a finite time, see [11, 12, 46].

Let us now introduce two articles on SOC phenomena, one by V. Barbu and one by U. Mosco, which represent the basis and starting point of this PhD thesis as well as some of the results produced. The articles are: *Self-organized criticality of cellular automata model; absorbtion in finite-time of supercritical region into the critical one* (2013) by V. Barbu and the article *Finite-time self-organized-criticality on synchronized infinite* *grids* (2018) by U. Mosco and they describe SOC models starting from very different perspectives: the first one makes a continuous analysis of the phenomenon connecting the problem to the theory of semigroups in order to give existence, uniqueness and finite time absorption results, while the second proposes a totally discrete interpretation and analysis of the problem, however reaching similar results.

1.2 The Barbu's continuous SOC model

Now, let's start with the analysis of the article [12].

Let be $u = u(x,t)$ the arbitrary state of the system and define it in a domain $\Omega \subset \mathbb{R}^n$, $n = 1, 2, 3$. Let us also consider the associated critical state $u^c = u^c(x), x \in \Omega$, time-independent, and the following partition of the domain Ω

$$\Omega_0^t = \{x \in \Omega; u(x,t) = u_\mathrm{c}(x)\} \text{ (critical region)}$$

$$\Omega_-^t = \{x \in \Omega; u(x,t) < u^c(x)\} \text{ (subcritical region), and}$$

$$\Omega_+^t = \{x \in \Omega; u(x,t) > u^c(x)\} \text{ (supercritical region).}$$

Let also the space of all Lebesgue measurable and p-integrable functions on Ω be indicated with $L^p(\Omega)$, as $1 \leq p \leq \infty$, and the corresponding norm with $|\cdot|_p$. Moreover, let us denote by $W^{k,p}(\Omega)$, $H^k(\Omega) = W^{k,2}(\Omega)$, $k = 1, 2$, the standard Sobolev spaces on Ω, and let us set $H_0^1(\Omega) = \{u \in H^1(\Omega); u = 0 \text{ on } \partial\Omega\}$ and $W_0^{1,p}(\Omega) = \{u \in W^{1,p}(\Omega); u = 0 \text{ on } \partial\Omega\}$, where $u = 0$ on $\partial\Omega$ is taken in sense of traces.

Finally let us denote by $H^{-1}(\Omega)$ the dual of $H_0^1(\Omega)$ in the pairing $\langle \cdot, \cdot \rangle$ with the pivot space $L^2(\Omega)$. Moreover, let Y be a Banach space, it will be denoted with $C([0,T]; Y)$ and $L^p(0,T;Y)$ the spaces of $Y-$valued continuous, respectively L^p-integrable functions, on $[0,T]$. At last, let $W^{1,p}(0,T;Y)$ be the infinite dimensional Sobolev space $\left\{ y \in L^p(0,T;Y); \frac{\mathrm{d}y}{\mathrm{d}t} \in L^p(0,T;Y) \right\}$ where $\frac{\mathrm{d}y}{\mathrm{d}t}$ is given in the sense of vectorial distributions. For more details, see [11] and [18].

In this setting, to well define the problem, the first important assumption to do in order to provide a continue analysis of the SOC model (1.5), since the equation holds true for all i, j of the discrete square lattice $N \times N$, is to substitute this finite discret 2-D domain by a continuous one, for instance $\Omega = (0,1) \times (0,1)$, or - more in general - with a subset $\Omega \subset \mathbb{R}^n$, and consider the generic point (i, j) as an element x in Ω, so to have the previous model similar to the nonlinear diffusion equation (1.5) on the region $\Omega \subset \mathbb{R}^n$ in the continuous time interval $[0, T]$.

Furthermore, the second important assumption, due to the discontinuity of the standard Heaviside function H, is to consider the multivalued Heaviside function $\widetilde{H}$ which enjoys the properties of maximal monotone graph in $\mathbb{R} \times \mathbb{R}$ (see [11]). This generalization allows us to set the model (1.5) in a more general environment, invoking the theory of semigroups and coming to demonstrate results of existence, uniqueness and extinction of the SOC phenomenon in a finite time, so let us consider the following multivalued Heaviside function:

$$\widetilde{H}(r) = \begin{cases} 1 & \text{for} \quad r > 0 \\ [0,1] & \text{for} \quad r = 0 \\ 0 & \text{for} \quad r < 0. \end{cases} \tag{1.6}$$

So, instead of equation (1.5), it will be considered the following multi-valued nonlinear diffusion problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} - \Delta \widetilde{H}\left(u\left(x,t\right) - u^c\left(x\right)\right) \ni 0 & \text{in } \Omega \times (0, \infty) \\ u\left(x, 0\right) = u_0(x), & x \in \Omega \\ 0 \in \widetilde{H}\left(u\left(x,t\right) - u^c\left(x\right)\right) & x \in \partial\Omega \text{ and } t \geq 0. \end{cases} \tag{1.7}$$

As shown in the following Section, it will be proved that the boundary value problem (1.7) is well posed and the *absorption* of the supercritical region Ω_+^t of a sandpile into its critical one Ω_0^t is reached in a finite time T not dependent of x, see [11, 12].

1.2.1 The nonlinear diffusion equation: existence and uniqueness of the solution

In this Section it will be discussed the problem (1.7) by setting, for simplicity and without any loss of generality, $w\left(x,t\right) = u\left(x,t\right) - u^c\left(x\right)$. So, the problem will be studied in the following form:

$$\begin{cases} \dfrac{\partial w}{\partial t} - \Delta\widetilde{H}(w\,(x,t)) \ni 0 & x \in \Omega, \quad t > 0 \\ w(0,x) = w_0(x) & x \in \Omega \\ \widetilde{H}(w) \ni 0 & \text{on } \mathbb{R}^+ \times \partial\Omega \end{cases} \tag{1.8}$$

with Ω bounded and open domain in $\mathbb{R}^n$ with smooth boundary $\partial\Omega$, $n \geq 1$ and $\widetilde{H}$ is the multivalued Heaviside function (1.6).

The definition of the problem that it is now arising is concerned about the theory behind the *porous media equation.* It is known that for all equation of the form

$$\frac{\partial w}{\partial t} - \Delta\psi(w) \ni 0 \tag{1.9}$$

with maximal monotone function $\psi : \mathbb{R} \to 2^{\mathbb{R}}$, the previuos problem (1.8) con be rewritten as an infinite dimensional Caucht problem in the space $L^1(\Omega)$.

So, it is possible to get:

$$\begin{cases} \dfrac{\mathrm{d}w}{\mathrm{d}t} + Aw(t) \ni 0 & t > 0 \\ w(0) = w_0 \end{cases} \tag{1.10}$$

where the operator $A : D(A) \subset L^1(\Omega) \to L^1(\Omega)$ is defined as:

$$Aw = \left\{-\Delta\eta; \eta \in W_0^{1,1}(\Omega), \eta(x) \in \widetilde{H}(w(x)),\ \text{a.e. } x \in \Omega\right\}. \tag{1.11}$$

Following this definition, it holds that the domain of the operator A, $D\,(A)$ is the set of all $w \in L^1(\Omega)$ for which such a Section $\eta \in W_0^{1,1}(\Omega)$ of $\widetilde{H}(w)$ exists.

In this case, it is possibile to proof that the operator A is also $m-$*accretive* in $L^1(\Omega)$ (see, [10], p.230); so, using the results of the classical Crandall - Liggett generation theorem (see, [10], p.131) follow that for all $w_0 \in \overline{D(A)}$ and $T > 0$[1], the infinite dimensional Cauchy problem (1.10) has a *unique mild* solution $w \in C\left([0,T]; L^1(\Omega)\right)$ expressed by:

$$w(t) = \lim_{n\to\infty} \left(I + \frac{t}{n}A\right)^{-n} w_0, \qquad t \geq 0. \tag{1.12}$$

[1] Here, $\overline{D(A)}$ denotes the closure of $D(A)$ in $L^1(\Omega)$.

Equivalently

$$w(t) = \lim_{\varepsilon \to 0} w_\varepsilon(t) \text{ uniformly in } t \tag{1.13}$$

where w_ε denotes the step function

$$w_\varepsilon(t) = w_i^\varepsilon \text{ for } t \in [i\varepsilon, (i+1)\varepsilon) \tag{1.14}$$

where w_i^ε is a solution of

$$\begin{aligned} & w_{i+1}^\varepsilon + \varepsilon A w_{i+1}^\varepsilon \ni w_i^\varepsilon, \quad i = 0, 1, 2, \ldots, N; N = \left[\frac{T}{\varepsilon}\right] \\ & w_0^\varepsilon = w_0. \end{aligned} \tag{1.15}$$

The main theorem in this Section is represented by the following result, that is concerned with the well-posedness of problem (1.8), also given in (1.10) and (1.11) formulation:

Theorem 1.2.1. *Assume that $w_0 \in L^2(\Omega)$. Then,* (1.8) *(or, more in general, the* (1.10) *formulation) has a unique solution w satisfying*

$$w \in C\left([0,T]; L^1(\Omega) \cap H^{-1}(\Omega)\right) \cap W^{1,2}\left(0,T; H^{-1}(\Omega)\right), \tag{1.16}$$

$$\begin{cases} \dfrac{\mathrm{d}w}{\mathrm{d}t}(t) - \Delta\eta(t) = 0, & a.e. \ in \ \Omega, \quad t \in (0,T) \\ \eta(x,t) \in \widetilde{H}(w(x,t)), & a.e. \ (x,t) \in \Omega \times (0,T) \end{cases} \tag{1.17}$$

where $\eta \in L^2\left(0,T; H_0^1(\Omega)\right)$. Moreover, $t \to \int_\Omega w(x,t)\mathrm{d}x$ is absolutely continuous. If $w_0 \in L^\infty(\Omega)$, then $w \in L^\infty(\Omega \times (0,T))$. If $w_0 \geq 0$, a.e. in Ω, then $w \geq 0$, a.e. in $\Omega \times (0,T)$.

Theorem and proof can be found in [12], pp. 1728–1730.

1.2.2 The absorbtion of supercritical region into the critical one

In this Section it will be set that $u_0, u^c \in L^\infty(\Omega)$ and that $u_0(x) \geq u^c(x),$ a.e. $x \in \Omega$.

Then, by Theorem 1.2.1, for all $T > 0$, there is a unique evolution in time of the function $u = u(x,t)$ satisfying problem (1.7) in the space (1.16). It means that

$$\begin{cases} u \in C\left([0,T];\ L^1(\Omega) \cap H^{-1}(\Omega)\right),\ \dfrac{\mathrm{d}u}{\mathrm{d}t} \in L^2\left(0,T; H^{-1}(\Omega)\right) \\ \eta \in \widetilde{H}\left(u - u^c\right),\ \eta \in L^2\left(0,T; H_0^1(\Omega)\right),\ u \in L^\infty(\Omega \times (0,T)) \\ u(x,t) \geq u^c(x),\ \text{a.e. } (x,t) \in \Omega \times (0,T) \\ \dfrac{\mathrm{d}u}{\mathrm{d}t} - \Delta\eta(t) = 0,\ \text{a.e. in } \Omega,\ t \in (0,T) \\ u(0) = u_0. \end{cases} \tag{1.18}$$

Remark 1.2.1. *It is posed that $\dfrac{\mathrm{d}}{\mathrm{d}t}$ is the strong derivative of u in the strong topology of $H^{-1}(\Omega)$.*

The main theorem in this Section is represented by the following result:

Theorem 1.2.2. *Holding all of the previous assumptions, we have*

$$u(x,t) - u^c(x) \equiv 0,\ \ a.e.\ x \in \Omega,\ \textit{for all } t \geq T^* \tag{1.19}$$

where

$$T^* = \frac{p^*}{p^* - 2}\gamma^2 \left|y_0\right|_\infty^{\frac{2}{p^*}} \left(\int_\Omega y_0 \mathrm{d}x\right)^{1 - \frac{2}{p^*}}. \tag{1.20}$$

Here, $p^* = \dfrac{2d}{d-2}$ for $d \geq 3, p^* > 2$ for $d = 1,2$ and γ is the Sobolev - Poincaré inequality constant.

This means that, at time $t = T^*$, the supercritical region Ω_+^t is completely absorbed into the critical zone Ω_0^t and remains there for all $t \geq T^*$.

Theorem and proof can be found in [12], pp. 1730–1732.

1.3 The Mosco's fully discrete SOC model

Let us continue our analysis by introducing [46].

In this second paper, there are several new developed points about SOC models. In order to fully understand the point of view expressed in this work, and the deep reasons in it, it is important to keep in mind that the models that describe the behavior of sandpiles are a representation of intrinsically discrete natural physical dynamic systems, as they consist of a large number of interacting particles.

For this reason, starting from this assumption, the SOC mathematical model described in this Section is entirely discrete, and, due to this feature, it is able to take into account the short-distance and the high-frequency interaction between the particles of the system itself. We observe that this particular feature was lost in the previous model one, as we passed from the discrete square lattice $N \times N$ to the continuous domain $\Omega \in \mathbb{R}^n$.

So, the novelty introduced in this article is just the presence of the numerical study of the solution set in a discrete square lattice $N \times N$, a priori infinite, domain whose number of the initially considered points increases in a synchronized way following a progress in time according to geometric growth laws. For this reason we speak about the evolutionary process of the sandpile as an impulsive one, defined on an infinite space-time synchronized grid, and the model is fully-discret.

1.3.1 Discretization of time and space

Let us start to explain the adopted space-time synchronization over a discrete square lattice $N \times N$.

1.3.1.1 Time discretization

Let us start by discretizing time over the real half-line $[0, +\infty)$.

With the help of each multi-integer index

$$ms\ell k \in (\mathbb{N} \cup \{0\})^4$$

it is associated the time $t_{m,s}^{\ell,k}$ set by:

$$\begin{cases} t^{\ell,k}_{m,s} = 4m + 4^{-m}\left(\ell + k4^{-s}\right) \\ 0 \leq m < +\infty \qquad\qquad 0 \leq s < +\infty \\ 0 \leq \ell \leq 4^{m+1} - 1 \qquad\qquad 0 \leq k \leq 4^{s} - 1 \\ m, s, \ell, k \in \mathbb{N}. \end{cases} \tag{1.21}$$

Note that, for all s, the set of the indices $mlk \in (\mathbb{N} \cup \{0\})^3$ follows a lexicographical order, as m, ℓ, k increase in their ranges. So, the immediate successive index that follows $ms\ell k$ is denoted by $m'sl'k'$. From this, two consecutive time instants

$$t^{\ell,k}_{m,s} < t^{\ell',k'}_{m',s}$$

are separated, on the positive real line, by quantity

$$t^{\ell',k'}_{m',s} - t^{\ell,k}_{m,s} = 4^{-(m+s)}$$

For given m and s, the set

$$\mathcal{T}_{m,s} := \left\{t^{\ell,k}_{m,s} : 0 \leq \ell \leq 4^{m+1} - 1, \quad 0 \leq k \leq 4^{s} - 1\right\} \tag{1.22}$$

divided into a partition the time interval $[4m, 4(m+1))$ into the subintervals $\left[t^{\ell,k}_{\overline{m},\overline{s}}, t^{\ell',k'}_{\overline{m},\overline{s}}\right)$. In accordance with the above, all of these subintervals have the same length, more precisely equal to $4^{-(m+s)}$. Moreover, it is introduced, for some $m \geq 0$ and $s \geq 0$ the following notation

$$t^{LAST}_{m,s} := t^{4^{m+1}-1, 4^{s}-1}_{m,s} \tag{1.23}$$

that simplify the recall of the last time instant in the corresponding subinterval, so it is possible to say that $t^{LAST}_{m,s} = 4(m+1) - 4^{-(m+s)}$. It is important to observe that $t^{LAST}_{m,s}$ converges to $4(m+1)$ from the left as $s \to +\infty$. It is also set

$$\mathcal{T}_{\infty} := \bigcup_{m=0}^{\infty} \bigcup_{s=0}^{\infty} \mathcal{T}_{m,s}. \tag{1.24}$$

1.3.1.2 Space discretization

Let us consider a $2L$ diameter discrete square lattice $N \times N$, where $L > 0$ is a given size parameter. It is defined also $\Omega = (-L, L) \times (-L, L)$, $\partial\Omega = (-L \times [-L, L]) \cup (L \times [-L, L]) \cup ([-L, L] \times -L) \cup ([-L, L] \times L)$ and $\overline{\Omega} = [-L, L] \times [-L, L]$.

In this way it has been substituted the continuous set Ω, $\overline{\Omega}$ and $[0, \infty)$ by the corespondent discrete grids of increasing cardinality.

Moreover, for every $m \in \mathbb{N}$ it has been defined the mesh size

$$h_m = \frac{2L}{2^m} \tag{1.25}$$

and set the discrete grids $G_m \subset \Omega$, $\partial G_m \subset \partial\Omega$, $\overline{G}_m \subset \overline{\Omega}$ as:

1. for $m = 0$, holds $G_0 = \emptyset$ and $\partial G_0 = \overline{G}_0$ as the set of the 4 vertices of Ω, while

2. for $m \in \mathbb{N}$, holds

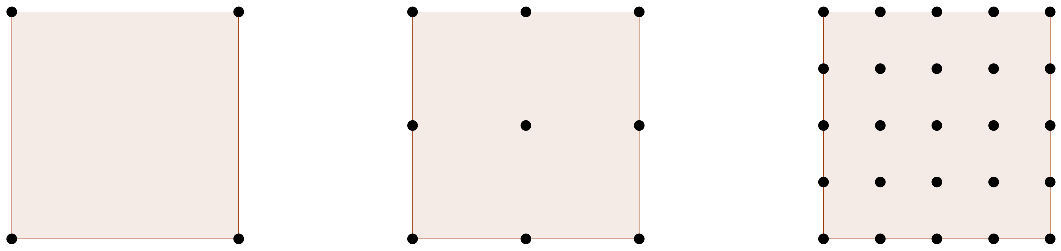
$$\begin{aligned}
G_m &:= \left\{\left(q_1\frac{2L}{2^m}, q_2\frac{2L}{2^m}\right) : (q_1, q_2) \in \mathbb{Z}^2, |q_1| \leq 2^{m-1} - 1, |q_2| \leq 2^{m-1} - 1\right\} \\
\partial G_m &:= \left\{\left(q_1\frac{2L}{2^m}, q_2\frac{2L}{2^m}\right) : (q_1, q_2) \in \mathbb{Z}^2, |q_1| = 2^{m-1}, |q_2| \leq 2^{m-1} - 1\right\} \\
&\cup \left\{\left(q_1\frac{2L}{2^m}, q_2\frac{2L}{2^m}\right) : (q_1, q_2) \in \mathbb{Z}^2, |q_1| \leq 2^{m-1} - 1, |q_2| = 2^{m-1}\right\} \\
&\cup \left\{\left(q_1\frac{2L}{2^m}, q_2\frac{2L}{2^m}\right) : (q_1, q_2) \in \mathbb{Z}^2, |q_1| = 2^{m-1}, |q_2| = 2^{m-1}\right\} \\
\overline{G}_m &:= G_m \cup \partial G_m.
\end{aligned}$$

Thanks to these positions, G_m is composed of $(2^m - 1) \times (2^m - 1)$ distinct sites in total, all in Ω while ∂G_m is composed of $2^m 4$ sites in total, all on $\partial\Omega$.

In the following it shall be considered

$$G_\infty := \bigcup_{m=0}^{\infty} G_m, \qquad \partial G_\infty := \bigcup_{m=0}^{\infty} \partial G_m, \qquad \overline{G}_\infty := G_\infty \bigcup \partial G_\infty \tag{1.26}$$

that are all countable space grids. It is set that $G_\infty \subset \Omega, \partial G_\infty \subset \partial\Omega, \overline{G}_\infty \subset \overline{\Omega}$. Figure (1.1) shows the progressive refinement of the numerical grids $\overline{G}_m$ and the corresponding growth of the number of the points inside them with the synchronized increase of the parameter m.



Discret grid for $m = 0$, i.e. $\overline{G}_0 := G_0 \cup \partial G_0$.

Discret grid for $m = 1$, i.e. $\overline{G}_1 := G_1 \cup \partial G_1$.

Discret grid for $m = 2$, i.e. $\overline{G}_2 := G_2 \cup \partial G_2$.

Figure 1.1: Example of the structure of the grid $\overline{G}_0$, $\overline{G}_1$ and $\overline{G}_2$.

It is now clear as the space-time synchronization works: for all discrete grid G_m at a mesh size given by h_m we have a synchronized time interval of the type $[4m, 4(m+1))$. As the parameter m increases, the mesh of the corresponding grid G_m grows and the time ticking becomes correspondingly quicker.

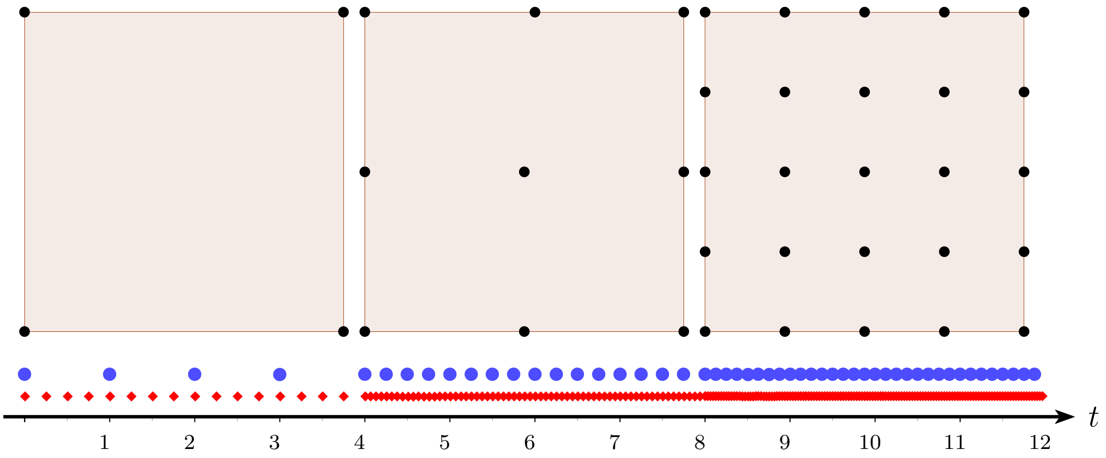


Figure 1.2: Discret grids for $m = 0$, $m = 1$ and $m = 2$, i.e. $\overline{G}_0$, $\overline{G}_1$ and $\overline{G}_2$ with the corresponding time ticking (in blue m parameter and in red s parameter).

A more simple notation is now introduced for grid sites and functions set on them: for every fixed $m \in \mathbb{N}$, it holds in an equivalent way that

$$ij = (ih_m, jh_m) = (q_1 h_m, q_2 h_m) \in \overline{G}_m$$

and for every function u on $\overline{G}_m$,

$$u_{ij} = u\left(ih_m, jh_m\right) = u\left(q_1 h_m, q_2 h_m\right).$$

1.3.1.3 Construction of the forward difference gradient and the difference Laplacian

With this notation, for a function u defined on G_m it shall be set the forward difference x_1-partial derivative

$$\left(D^+_{1m} u\right)_{ij} = \frac{1}{h_m}\left[u_{(i+1)h_m, jh_m} - u_{ih_m, jh_m}\right] = \frac{1}{h_m}\left[u_{i+1j} - u_{ij}\right] \tag{1.27}$$

at the sites $ij = (q_1 h_m, q_2 h_m) \in \overline{G}_m$ with $-2^{m-1} \leq q_1 \leq 2^{m-1} - 1$ and $-2^{m-1} \leq q_2 \leq 2^{m-1}$, the forward difference x_2-partial derivative

$$\left(D^+_{2m} u\right)_{ij} = \frac{1}{h_m}\left[u\left(ih_m, (j+1)h_m\right) - u\left(ih_m, jh_m\right)\right] = \frac{1}{h_m}\left[u_{ij+1} - u_{ij}\right] \tag{1.28}$$

at the sites $ij = (q_1 h_m, q_2 h_m) \in \overline{G}_m$ with $-2^{m-1} \leq q_1 \leq 2^{m-1}$ and $-2^{m-1} \leq q_2 \leq 2^{m-1} - 1$, the forward difference gradient

$$\left(\nabla^+_m u\right)_{ij} = \left(\left(D^+_{1m} u\right)_{ij}, \left(D^+_{2m} u\right)_{ij}\right) \tag{1.29}$$

at the sites $ij = (q_1 h_m, q_2 h_m) \in \overline{G}_m$ with $-2^{m-1} \leq q_1 \leq 2^{m-1} - 1$ and $-2^{m-1} \leq q_2 \leq 2^{m-1} - 1$.

Now, for every function u on $\overline{G}_m$ and at every site $ij \in G_m$, it is set the centered difference Laplacian Δ_m, as

$$\left(\Delta_m u\right)_{ij} = \frac{1}{h_m^2}\left[4u_{ij} - u_{i-1j} - u_{i+1j} - u_{ij-1} - u_{ij+1}\right] \quad \forall ij \in G_m. \tag{1.30}$$

Moreover, in the following, it will be useful the next equality that holds for every arbitrary vectors $a = (a_k)_{k=0}^{N+1}$ and $b = (b_k)_{k=0}^{N+1}$, known as the summation-by-parts identity:

$$\sum_{k=0}^{N}\left(a_{k+1}-a_{k}\right)\left(b_{k+1}-b_{k}\right)= \tag{1.31}$$
$$=\left(a_{N+1}-a_{N}\right)b_{N+1}-\left(a_{1}-a_{0}\right)b_{0}+\textstyle\sum_{k=1}^{N}\left(2a_{k}-a_{k-1}-a_{k+1}\right)b_{k}.$$

From (1.31) and by considering $a_k = u_{kj}$ and $b_k = v_{kj}$, for $-2^{m-1} \leq k = q_1 \leq 2^{m-1}-1$ and for every fixed j with $-2^{m-1} \leq j = q_2 \leq 2^{m-1}$, it follows that

$$\sum_{k=-2^{m-1}}^{2^{m-1}-1}\left(u_{k+1j}-u_{kj}\right)\left(v_{k+1j}-v_{kj}\right)=$$
$$=\left(u_{2^{m-1}j}-u_{(2^{m-1}-1)j}\right)v_{2^{m-1}j}-\left(u_{(-2^{m-1}+1)j}-u_{(-2^{m-1})j}\right)v_{(-2^{m-1})j}+$$
$$+\sum_{k=-2^{m-1}+1}^{2^{m-1}-1}\left(2u_{kj}-u_{(k-1)j}-u_{(k+1)j}\right)v_{kj}.$$

Moreover, if holds that $v = 0$ on ∂G_m, the previous inequality becomes

$$\sum_{k=-2^{m-1}}^{2^{m-1}-1}\left(u_{k+1j}-u_{kj}\right)\left(v_{k+1j}-v_{kj}\right)=\sum_{k=-2^{m-1}+1}^{2^{m-1}-1}\left(2u_{kj}-u_{(k-1)j}-u_{(k+1)j}\right)v_{kj} \tag{1.32}$$

considering the definition of D^{+}_{1m} (1.27), the equality (1.32) reduces to

$$\sum_{k=-2^{m-1}}^{2^{m-1}-1}\left(D^{+}_{1m}u\right)_{kj}\left(D^{+}_{1m}v\right)_{kj}=\sum_{k=-2^{m-1}+1}^{2^{m-1}-1}\frac{1}{h_m^2}\left(2u_{kj}-u_{(k-1)j}-u_{(k+1)j}\right)v_{kj}. \tag{1.33}$$

On the other hand, by summing both sides of (1.33) over $-2^{m-1} \leq j = q_2 \leq 2^{m-1}-1$ and noting that the right-hand side $v_{kj} = 0$ for $j = -2^{m-1}$ for every k, it is possible to write the identity

$$\sum_{j=-2^{m-1}}^{2^{m-1}-1}\sum_{k=-2^{m-1}}^{2^{m-1}-1}\left(D^{+}_{1m}u\right)_{kj}\left(D^{+}_{1m}v\right)_{kj} \tag{1.34}$$
$$=\sum_{j=-2^{m-1}-1}^{2^{m-1}-1}\sum_{k=-2^{m-1}+1}^{2^{m-1}-1}\frac{1}{h_m^2}\left(2u_{kj}-u_{(k-1)j}-u_{(k+1)j}\right)v_{kj}$$

for all functions u and v on $\overline{G}_m$ with $v = 0$ on ∂G_m.

Similarly, for all functions u and v on $\overline{G}_m$ with $v = 0$ on ∂G_m, holds

$$\begin{aligned}&\sum_{k=-2^{m-1}}^{2^{m-1}-1} \sum_{j=-2^{m-1}}^{2^{m-1}-1} \left(D_{2m}^{+} u\right)_{kj} \left(D_{2m}^{+} v\right)_{kj} \\ &= \sum_{k=-2^{m-1}+1}^{2^{m-1}-1} \sum_{j=-2^{m-1}+1}^{2^{m-1}-1} \frac{1}{h_m^2} \left(2u_{kj} - u_{k(j-1)} - u_{k(j+1)}\right) v_{kj}.\end{aligned} \tag{1.35}$$

Working with (1.34) and (1.35) together, it is possible to have the identity

$$\sum_{i,j=-2^{m-1}}^{2^{m-1}-1} \left[\left(D_{1m}^{+} u\right)_{ij} \left(D_{1m}^{+} v\right)_{ij} + \left(D_{2m}^{+} u\right)_{ij} \left(D_{2m}^{+} v\right)_{ij}\right] = \sum_{ij \in G_m} \left(\Delta_m u\right)_{ij} v_{ij} \tag{1.36}$$

that, as said before, is satisfied by all functions $u = (u_{ij})$ and $v = (v_{ij})$ on $\overline{G}_m$ with $v_{ij} = 0$ for every $ij \in \partial G_m$, where $\Delta_m u = \left(\Delta_m u\right)_{ij \in G_m}$ is the centered difference Laplacian defined in (1.30).

All the above defined tools find their natural mathematical collocation in the following Hilbert space, more precisely: for every $m \in \mathbb{N}$ it is set

$$Y_m = \left\{u = (u_{ij}) : ij \in \overline{G}_m\right\} \tag{1.37}$$

with inner product

$$\langle u, v\rangle_m = \sum_{ij \in \overline{G}_m} u_{ij} v_{ij} h_m^2, \quad u = (u_{ij}) \in Y_m, v = (v_{ij}) \in Y_m \tag{1.38}$$

and norm

$$\|u\|_m = \langle u, u\rangle_m^{1/2}. \tag{1.39}$$

Then, for vectors $\nabla_m^{+} u = \left(D_{1m}^{+} u, D_{2m}^{+} u\right)$ and $\nabla_m^{+} v = \left(D_{1m}^{+} v, D_{2m}^{+} v\right)$, holds

$$
\begin{aligned}
&\langle \nabla_m^+ u, \nabla_m^+ v \rangle_m = \left\langle D_{1m}^+ u, D_{1m}^+ v \right\rangle_m + \left\langle D_{2m}^+ u, D_{2m}^+ v \right\rangle_m \text{ with} \\
&\left\langle D_{1m}^+ u, D_{1m}^+ v \right\rangle_m = \textstyle\sum_{i,j=-2^{m-1}}^{2^{m-1}-1} \left(D_{1m}^+ u\right)_{ij} \left(D_{1m}^+ v\right)_{ij} h_m^2 \\
&\left\langle D_{2m}^+ u, D_{2m}^+ v \right\rangle_m = \textstyle\sum_{i,j=-2^{m-1}}^{2^{m-1}-1} \left(D_{2m}^+ u\right)_{ij} \left(D_{2m}^+ v\right)_{ij} h_m^2
\end{aligned}
\tag{1.40}
$$

and sets

$$\left\| \nabla_m^+ u \right\|_m = \left\langle \nabla_m^+ u, \nabla_m^+ u \right\rangle_m^{1/2}.$$

In this Hilbert-space, it is possible to write the identity (1.36) in the following way

$$\left\langle \nabla_m^+ u, \nabla_m^+ v \right\rangle_m = \langle \Delta_m u, v \rangle_m \tag{1.41}$$

$u = (u_{ij})$ and $v = (v_{ij})$ on $\overline{G}_m$ with $v_{ij} = 0$ for every $ij \in \partial G_m$.

Remark 1.3.1. *It is important to point out that* $\Delta_m u = (\Delta_m u)_{ij \in G_m}$ *is the centered difference Laplacian defined in* (1.30)*.*

Now, it has been introduced the following subspace:

$$X_m = \{u = (u_{ij}) \in Y_m : u_{ij} = 0 \quad \forall ij \in \partial G_m\} \tag{1.42}$$

of Y_m. This space is a Hilbert on itself and it is endowed with the following inner product

$$(u, v)_m = \langle u, v \rangle_m + \left\langle \nabla_m^+ u, \nabla_m^+ v \right\rangle_m, \quad u = (u_{ij}) \in Y_m, v = (v_{ij}) \in Y_m. \tag{1.43}$$

Moreover, in this contest it is possible to prove that on X_m holds the following discrete Poicaré inequality

$$\|u\|_m \leq c_P \left\| \nabla^+ u \right\|_m \quad \forall u \in X_m \tag{1.44}$$

with $c_P = L\sqrt{2}$.

To give existence and uniqueness results, it will be now built an equation (see (1.47)) that has the required features. In order to do this, it is now introduced on the subspace X_m of Y_m the bilinear form $E_m(u, v) = \langle \nabla_m^+ u, \nabla_m^+ v \rangle_m$ with $\langle \nabla_m^+ u, \nabla_m^+ v \rangle_m$ is given by (1.40); to go on, it is set

$$\begin{aligned} E_m(u,v) &= \langle \nabla_m^+ u, \nabla_m^+ v\rangle_m = \\ &= \sum_{i,j=-2^{m-1}}^{2^{m-1}-1} \left[\left(D_{1m}^+ u\right)_{ij} \left(D_{1m}^+ v\right)_{ij} + \left(D_{2m}^+ u\right)_{ij} \left(D_{2m}^+ v\right)_{ij} \right] h_m^2 \end{aligned} \tag{1.45}$$

in the domain $\mathcal{D}\left[E_m\right] = X_m$. By (1.44), it is possible to assert that the form $E_m\left(u,v\right)$ is coercive

$$\frac{1}{c_P^2+1}(u,u)_m^2 \leq E_m(u,u) \quad \forall u \in X_m \tag{1.46}$$

and that the inner product $\left(u,v\right)_m$, set in (1.42), is equivalent to the form $E_m\left(u,v\right)$. For every $m \in \mathbb{N} \cup \{0\}$, it is now defined the linear operator

$$\mathcal{G}_m : Y_m \mapsto Y_m$$

as follows. For every $f \in Y_m$ it is set $u = \mathcal{G}_m f$ as the solution of

$$u \in X_m : \quad E_m(u,v) = \langle f, v\rangle_m \quad \forall v \in X_m \tag{1.47}$$

Moreover, using (1.46), for every $f \in Y_m$ is possible to say that the solution u of (1.47) exists and is unique. For a more detailed discussion, see [46], pp. 2414 – 2417.

The most important point of this Subsection is that the operator $\mathcal{G}_m$ and the centered difference Laplacian $\Delta_m u = \left(\Delta_m u\right)_{ij \in G_m}$, previously defined in (1.30), are related on X_m by the identity

$$\mathcal{G}_m\left(\Delta_m u\right) = u \tag{1.48}$$

and that it is satisfied componentwise the relation

$$\left(\mathcal{G}_m\left(\Delta_m u\right)\right)_{ij} = u_{ij} \quad \forall ij \in \overline{G}_m \tag{1.49}$$

for every $u = \left(u_{ij}\right)_{ij \in \overline{G}_m}, u \in X_m$.

Now, to conclude this preparation of the discrete Mosco model for sandpiles, associated with the multivalued Heaviside function

$$\widetilde{H}(r) = \begin{cases} 1 & \text{for} \quad r > 0 \\ [0,1] & \text{for} \quad r = 0 \\ 0 & \text{for} \quad r < 0 \end{cases} \tag{1.50}$$

it is introduced another operator, namely

$$\mathbf{H} : Y_m \mapsto Y_m \tag{1.51}$$

in Y_m through the relation $\eta \in \mathbf{H}(z)$ for generic vector functions $z \in Y_m$, $z = (z_{ij})_{ij \in \overline{G}_m}$, $\eta \in Y_m$, $\eta = (\eta_{ij})_{ij \in \overline{G}_m}$ componentwise as

$$(\mathbf{H}(z))_{ij} = H(z_{ij}) \quad \forall ij \in \overline{G}_m \tag{1.52}$$

i.e.,

$$\eta_{ij} = 0 \text{ if } z_{ij} < 0; \quad 0 \leq \eta_{ij} \leq 1 \text{ if } z_{ij} = 0; \quad \eta_{ij} = 1 \text{ if } z_{ij} > 0 \quad \forall ij \in \overline{G}_m.$$

This multivalued operator $z \to \mathbf{H}(z)$ is

1. monotone in $X_m \times X_m$ because it satisfies (see [10], p.43)

$$\langle \eta_1 - \eta_2, z_1 - z_2 \rangle_m \geq 0 \quad \forall \eta_i \in \mathbf{H}(z_i), z_i, \eta_i \in X_m, i = 1, 2 \tag{1.53}$$

2. maximal monotone and coercive because the range of $z \to \lambda z + \mathbf{H}(z)$ is all of X_m for every $\lambda > 0$.

By this features it can be said that $\mathcal{G}_m + c\mathbf{H}$ is maximal monotone and coercive on $X_m \times X_m$ for every constant $c > 0$ and its range is all of X_m. For a more detailed discussion, see [46], p. 2417. This is the fundamental condition to have the existence proof of the following results.

1.3.2 The fully discrete SOC model

We are now ready to introduce the fully discret Mosco SOC model. The model is set on the countable grid $\overline{G}_\infty \times \mathcal{T}_\infty$ and we shall refer about notation to the previous Section.

Remark 1.3.2. *For a function u set in $\overline{G}_\infty$, the restriction of u to $\overline{G}_m$ is could be denoted by u_m.*

In this model, the given data are (given) functions $u^c = \left(u^c_{ij}\right)_{ij\in\overline{G}_\infty}$ (the meaning of these functions is to be the sandpile *critical state*) and $u^0 = \left(u^0_{ij}\right)_{ij\in\overline{G}_\infty}$ (this is, instead, the sandpile initial configuration).

Both functions are defined on $\overline{G}_\infty$ and satisfy the following conditions

$$\begin{aligned} & u^c = \left(u^c_{ij}\right)_{ij\in\overline{G}_\infty}, \quad u^0 = \left(u^0_{ij}\right)_{ij\in\overline{G}_\infty} \text{ satisfying} \\ & u^0_{ij} \geq u^c_{ij} \ \ \forall ij \in \overline{G}_\infty, \quad u^0_{ij} = u^c_{ij} \ \ \forall ij \in \partial G_\infty. \end{aligned} \tag{1.54}$$

That is, the model is based on two sets of equations:

1. the first set of equations define the solutions $u^{(s)}\left(t^{\ell,k}_{m,s}\right)$ for $\left(t^{\ell,k}_{m,s}, ij\right) \in \mathcal{T}_{m,s} \times \overline{G}_m$.

   (a) For fixed $ms \in \mathbb{N} \times \mathbb{N}$, they are iteratively get in ℓ, k by the following finite-difference scheme with given initial conditions u^*_m on the grid $\overline{G}_m$

$$\begin{aligned} & u^{(s)}\left(t^{0,0}_{m,s}\right) = u^*_m \quad \text{ on } \overline{G}_m \\ & u^{(s)}\left(t^{\ell',k'}_{m,s}\right) = u^{(s)}\left(t^{\ell,k}_{m,s}\right) - 4^{-(m+s)}\Delta_m\left(\eta^{\ell',k'}_{m,s}\right) \text{ with} \\ & \eta^{\ell',k'}_{m,s} \in \mathbf{H}\left(u^{(s)}\left(t^{\ell',k'}_{m,s}\right) - u^c_m\right) \text{ on } \overline{G}_m \\ & u^{(s)}\left(t^{\ell',k'}_{m,s}\right) = u^{(s)}\left(t^{\ell,k}_{m,s}\right) = u^c_m, \quad \eta^{\ell',k'}_{m,s} = 0 \quad \text{ on } \partial G_m \end{aligned} \tag{1.55}$$

   for $0 \leq \ell < 4^{m+1} - 1, 0 \leq k \leq 4^s - 1$. We note that the index ℓ', k' is the index-pair successive to ℓ, k in the lexicographic order. The times $t^{0,0}_{m,s}$ are independent of s; in fact, it holds that $t^{0,0}_{m,s} = 4m$ for all s.

2. the second set of equations is determined iteratively in $m \in \mathbb{N}\cup\{0\}$ and it is referred to the sequence of initial states u^*_m on the grids $\overline{G}_m$ in the preceding equations.

   (a) For $m = 0$, $u^*_m = u^*_0$ is assigned by setting

$$u^*_0 := u^0_0 \text{ on } \overline{G}_0 \tag{1.56}$$

(b) while, for given indices $1 \leq m \leq n, \quad n \in \mathbb{N}$, u^*_m is given on the grid G_m in three levels:

i. for every $s \in \{\mathbb{N}\}$, it is set that

$$\begin{aligned} &u^*_m(ij) = u^c(ij) + 4\left(u^{(s)}_{m-1}\left(t^{4^m-1,4^s-1}_{m-1,s}, ij\right) - u^c(ij)\right) \\ &\forall n \geq 1, \forall 1 \leq m \leq n, \forall ij \in G_m \cap G_{m-1} = G_{m-1}, \end{aligned} \tag{1.57}$$

ii. then, it is also set

$$\begin{aligned} &u^*_m(ij) := u^c(ij) + \tfrac{1}{4}4^{-(n-m)}\left(u^0(ij) - u^c(ij)\right) \\ &\forall n \geq 1, \forall 1 \leq m \leq n, \forall ij \in G_m - G_{m-1}, \end{aligned} \tag{1.58}$$

iii. in the end, it is defined

$$u^*_m(ij) = u^c(ij) \quad \forall ij \in \partial G_m. \tag{1.59}$$

In these three steps it has been defined the complete calculation of u^*_m on the grids $\overline{G}_m$, $\overline{G}_m = G_m \cup \partial G_m$, for all m, here considered.

Remark 1.3.3. *Note that for $m = 0, \overline{G}_0 = \partial G_0 = \{ij : i, j = \pm 2^m\}$ consists of the 4 vertices of the square $[-L, L] \times [-L, L]$. Moreover, $u^0_0 = u^c_0$ on $\overline{G}_0$.*

Remark 1.3.4. *The value $u^{(s)}_{m-1}\left(t^{4^m-1,4^s-1}_{m-1,s}, ij\right) = u^{(s)}_{m-1}\left(t^{LAST}_{m,s}\right)$ is the last value got by the solution on the grid G at the step $m-1$ in the time interval $\left[t^{0,0}_{m-1}, t_m\right)$.*

Now, we give the most relevant result in this Section:

Theorem 1.3.1. *We are given an initial state u^0 and a critical state u^c on $\overline{G}_\infty$ satisfying (1.54). We fix an arbitrary $s \in \mathbb{N}$. Then for all $m \in \mathbb{N}$, all $0 \leq \ell \leq 4^{m+1} - 1$, and all $0 \leq k \leq 4^s - 1$ and for all discrete times $t^{\ell,k}_{m,s} \geq 0$ as in (1.21), there exists a unique solution $u^{(s)}$, with values $u^{(s)}\left(t^{\ell,k}_{m,s}\right)$, of the impulsive system from (1.55) to (1.59). Moreover,*

$$u^c \leq u^{(s)}\left(t^{\ell,k}_{m,s}\right) \leq u^c + M_m \quad \text{on } \overline{G}_m \quad \forall t^{\ell,k}_{m,s} \geq 0 \tag{1.60}$$

where

$$M_m := \max_{ij\in\overline{G}_m} \left\{ u^0_{ij} - u^c_{ij} \right\}. \tag{1.61}$$

Theorem and proof can be found in [46], pp. 2425–2436.

Remark 1.3.5. *Analyzing the result of* Theorem 1.3.1*, which is given for a fixed grid* G_m *and a given time set* $[4m, 4(m+1))$*, it is possible to say that it does not get any information about the long-time behavior of the solutions. To give some information about it, it is necessary fix additional hypothesis on the data.*

1.3.3 Long-time behavior

This Section is based on the study of the discrete model considered in Theorem 1.3.1. The basic data of this model are two real-valued functions u^0 (initial sandpile state) and u^c (critical sandpile state) pointwise defined on $\overline{G}_\infty$ introduced in (1.26). For the following results, it shall be assumed that condition (1.54) holds and that there exist two constant M_∞ and M_1 s.t.

$$0 < M_\infty < +\infty, \quad M_1 < +\infty \tag{1.62}$$

defined as

$$M_\infty := M_\infty\left(u^0 - u^c\right) = \sup_{ij\in G_\infty} \left(u^0(ij) - u^c(ij)\right), \tag{1.63}$$

$$M_1 := M_1\left(u^0 - u^c\right) = \sup_{n\in\mathbb{N}} \sum_{ij\in\overline{G}_n} \left(u^0(ij) - u^c(ij)\right) h_n^2. \tag{1.64}$$

For $\tau = 4m + \ell 4^{-m}$, $m \geq 0$, $\ell \in \{0, \dots, 4^{m+1} - 1\}$, it is set

$$\Phi_\infty(\tau) := \liminf_{s\to+\infty} \sum_{ij\in\overline{G}_m} \left(u^{(s)}\left(t^{\ell,k}_{m,s}, ij\right) - (u^c)_{ij}\right) h_m^2 \tag{1.65}$$

where $u^{(s)}\left(t^{\ell,k}_{m,s}\right)$ are the solutions of the impulsive system from (1.55) to (1.59). It is also set

$$\begin{aligned} &T^* := \sup\{4n : n \in \mathbb{N},\ \text{such that } \Phi_\infty(\tau) > 0 \quad \forall 0 \leq m \leq n \\ &\forall \tau \in [4m, 4(m+1)), \tau = 4m + \ell 4^{-m}, \ell \in \{0, \ldots, 4^{m+1} - 1\}\} \in [0, +\infty] \end{aligned} \tag{1.66}$$

This Section is concluded by the following theorem:

Theorem 1.3.2. *Let u^0 and u^c be two functions defined on the infinite grid $\overline{G}_\infty$, satisfying the assumptions* (1.54) *and* (1.62)*. Then, there exist a grid $\overline{G}_{m^*}$ and a finite time $\tau^* = 4m^* + \ell^* 4^{-m^*}$, $m^* \geq 0$, $0 \leq \ell^* \leq 4^{m^*+1} - 1$, such that the solutions $u^{(s)}$ of the system from* (1.55) *to* (1.59) *satisfy the property*

$$\Phi_\infty(\tau^*) = 0 \tag{1.67}$$

implying

$$T^* < +\infty. \tag{1.68}$$

Theorem and proof can be found in [46], pp. 2436–2437.

1.3.4 A priori estimate

In this last Section, by strengthening the assumptions on the data functions u^0 and u^c, it will be given a more precise and explicit estimate about the finite equilibrium time T^* with a direct connection and dependence by the initial data.

About the following Section it shall be considered a coarses time grid. For every $m \in \mathbb{N}$ and $\ell = 0, \ldots, 4^m - 1$ hold the times

$$t_m^\ell = 4m + \ell 4^{-m} \tag{1.69}$$

The set

$$\mathcal{T}_m := \left\{t_m^\ell : 0 \leq \ell \leq 4^{m+1} - 1\right\} \tag{1.70}$$

has the usual lexicographer order in relationship with the index $m\ell$. We set

$$\mathcal{T}^\infty := \bigcup_{m=1}^{n} \mathcal{T}_m. \tag{1.71}$$

For $\tau \in \mathcal{T}^\infty$, there are a unique $m \in \mathbb{N} \cup \{0\}$ and a unique $\ell \in \{0, \dots, 4^m - 1\}$, s.t. $\tau = t_m^\ell \in \mathcal{T}_m$. At last, we define

$$\Phi^\infty(\tau) := \limsup_{s \to +\infty} \sum_{ij \in \overline{G}_m} \left(u^{(s)} \left(t_{m,s}^{\ell,k}, ij \right) - (u^c)_{ij} \right) h_m^2 \tag{1.72}$$

where $u^s = u^{(s)} \left(t_{m,s}^{\ell,k} \right)$ are the solutions of the impulsive system from (1.55) to (1.59).

Theorem 1.3.3. *Let u^0 and u^c be the restrictions to $\overline{G}_\infty$ of two continuous functions $u^0 \geq u^c$ on $\overline{\Omega}$ such that $u^0 - u^c$ is Lipschitz continuous on $\overline{\Omega}$ Then, we have*

$$\Phi^\infty(\tau) = 0 \tag{1.73}$$

for all $\tau \in \mathcal{T}^\infty$, $\tau \geq \tau^*$, where

$$\tau^* := A^{-2} \left[\left(\frac{|B|}{4} \right)^{\frac{1}{2}} + \left(\frac{C}{4} \right)^{\frac{1}{2}} \right]^2 < +\infty. \tag{1.74}$$

The constants in (1.74) are defined as follows:

$$A := 2 C_S^{-2} M_\infty^{-\frac{1}{2}} \tag{1.75}$$

where $M_\infty = M_\infty \left(u^0 - u^c \right)$ is given by (1.63) and C_S is a constant depending only on L and p;

$$B := \iint_{\overline{O}} \left(u^0 - u^c \right) \mathrm{d}x \mathrm{d}y \tag{1.76}$$

$$C := C_{Lip} 2L\sqrt{2} \tag{1.77}$$

where

$$C_{Lip} = C_{Lip} \left(u^0 - u^c \right) \tag{1.78}$$

is the Lipschitz seminorm of $u^0 - u^c$ on $\overline{\Omega}$.

Theorem and proof can be found in [46], pp. 2437–2438.

1.4 Numerical analysis

This Section is devoted to a numerical study of the evolutionary non singular problem (1.7) as presented in [12] through an implicit finite difference approach. As to do so, in a similar way to what is in (1.3) done, we have to replace the multivalued Heaviside function with a smooth approximation of it: this will be discussed in Subsection 1.4.1, together with some properties of this regularized problem.

The numerical scheme is implemented in Subsection 1.4.2.1 on a fixed space grid. Since the discrete settings cannot reproduce completely the fine spatial interactions of the continuous model, the finite time convergence corresponds to the time evolution of the numerical solution to the target on every single node up to any prescribed accuracy, globally with exponential rate, remaining supercritical during all the process.

In Subsection 1.4.2.2 we will adapt our scheme to a space-time synchronized family of grids in order to reduce the convergence time and the total computational cost. The idea, as said, is inspired by the recent paper of Mosco [46], above discussed, where a fully discrete analytic model allowing for infinite degrees of freedom is studied to preserve the physical aspect of SOC phenomena which are intrinsically discrete particle processes, better captured by discrete equations on an infinite spatial-temporal lattice incorporating arbitrary short-range and high-frequency particle interactions. The equations of the process on each finite spatial grid are coupled in time with an impulsive change of the initial data at each refinement, keeping constant the parabolic ratio between the discretization steps, up to the desired accuracy.

In Subsection 1.4.3 we finally present the results of some numerical simulations both in one and two dimensions.

1.4.1 The basic model

In view of the discretization of problem (1.7), we first of all introduce a regular approximation $\eta \in C^1(\mathbb{R})$ of H in a neighborhood of the origin, for example by means of the cubic polynomial:

$$\eta_1(r) = \begin{cases} 1 & \text{if } r > \dfrac{1}{n} \\ -\dfrac{n^3r^3}{4} + \dfrac{3}{4}nr + \dfrac{1}{2} & \text{if } -\dfrac{1}{n} \leq r \leq \dfrac{1}{n} \\ 0 & \text{if } r < -\dfrac{1}{n}. \end{cases} \tag{1.79}$$

With respect to the integer parameter $n \in \mathbb{N}$, we see that $\eta_1(r) = \widetilde{H}(r)$ for $|r| \geq \dfrac{1}{n}$, and that

$$\left\| \widetilde{H} - \eta_1 \right\|_{L^1(-1,1)} = \int_{-1}^{1} |\widetilde{H}(r) - \eta_1(r)| \, \mathrm{d}r = \int_{-1/n}^{1/n} |\widetilde{H}(r) - \eta_1(r)| \, \mathrm{d}r \to 0 \quad as\ n \to \infty. \tag{1.80}$$

Note that $\eta_1(0) = 0.5$, so that this will be the value taken at the contact points between the solution of problem (1.7) (with $\widetilde{H}$ replaced by η_1) and the critical one, in particular at the boundary. Moreover, the values of η_1 in $(0.5, 1]$ characterize the supercritical states of the solution, those in $[0, 0.5)$ the subcritical ones.

There are of course other possible candidates for the approximation of $\widetilde{H}$. For example the following monotone functions

$$\eta_2(r) = \frac{1}{(1 + e^{-nr})} \ , \ \text{or} \quad \eta_3(r) = \frac{1}{2} + \frac{1}{\pi} \arctan(nr) \tag{1.81}$$

have similar properties for a large n, even if in those cases the support of $(\widetilde{H} - \eta)$ is unbounded. The numerical tests show that the choice of η slightly influences the qualitative behavior of the solution, that is the way it converges to the target (more impulsive for η_1, more smooth for the other choices), but it is not relevant for the final time of convergence, which is essentially determined by the quantity $\eta'(0)$ (in the previous examples: $\eta_1'(0) = 3n/4$, $\eta_2'(0) = n/4$, $\eta_3'(0) = n/\pi$): the higher is this value, the closer we are to the real Heaviside, and the faster is the convergence to the target (see Section 1.4.3).

The problem to solve then becomes:

$$\begin{cases} u_t - \Delta\eta(u - u^c) = 0 & \text{in } \Omega \times (0, \infty) \\ u(0) = u^0 > u^c & \text{in } \Omega \\ u = u^c & \text{on } \partial\Omega \times [0, \infty), \end{cases} \tag{1.82}$$

where η denotes one of the approximate Heaviside functions previously defined, for a fixed large parameter n that for simplicity from now on we neglect.

If we set $w\left(x,t\right) := u\left(x,t\right) - u^c\left(x\right)$, problem (1.82) can also be written as:

$$\begin{cases} w_t - \Delta\eta(w) = 0 & \text{in } \Omega \times (0,\infty) \\ w(0) = w_0 > 0 & \text{in } \Omega \\ w = 0 & \text{on } \partial\Omega \times [0,\infty). \end{cases} \tag{1.83}$$

Applying the divergence theorem, together with the boundary conditions on w, we get:

$$\int_\Omega \Delta\eta\left(w\right)\ \mathrm{d}x = \oint_{\partial\Omega} \nabla\eta\left(w\right)\cdot\vec{n}\ \mathrm{d}s = \oint_{\partial\Omega} \eta'(w)\nabla w\cdot\vec{n}\ \mathrm{d}s = \eta'(0)\oint_{\partial\Omega}\frac{\partial w}{\partial n}\ \mathrm{d}s \le 0, \tag{1.84}$$

since $\eta'(0) > 0$ (η is increasing at the origin) and the normal derivative of w on the boundary in the direction of the external normal vector is non positive.

From (1.83) and (1.84) then follows:

$$\int_\Omega w_t\ \mathrm{d}x = \int_\Omega \Delta\eta\left(w\right)\ \mathrm{d}x \le 0 \tag{1.85}$$

according to that, if we define the two quantities:

$$M(t) = \int_\Omega w(x,t)\ \mathrm{d}x, \quad E(t) = -\int_\Omega \Delta\eta(w)\ \mathrm{d}x \tag{1.86}$$

then we have $M'(t) \le 0$ and $E(t) \ge 0$.

In other words, from (1.86), $M(t)$, which can be interpreted as the total mass of the problem (that is the global distance from the target), always decreases with time. This fact does not imply that w decreases at any point of Ω: from (1.83) we see that this should be equivalent to say that the function $\eta(w)$ be globally superharmonic in Ω. Of course, according to the specific given initial data, this could not be the starting situation. In fact the second quantity $E(t)$, which represents a sort of distance from harmonicity for $\eta(w)$, might initially grow before it starts to decrease toward zero. We will find the same behavior even in the discrete settings.

For the original continuous model (1.7) it was proved in [12] that $u(t) \to u^c$ (and $M(t) \to 0$) in finite time. Asimptotically, model (1.82) shares the same properties. In particular at any point of Ω as t grows: $w(t) \to 0$, $\eta(w(t)) \to 0.5$, and $\Delta\eta(w(t)) \to 0$. We will prove these properties for the finite difference solution of an implicit scheme which discretizes (1.82).

1.4.2 Numerical approximations

We now present our numerical study of the problem (1.7) first on fixed grid in time, and then, in the spirit of [46], through the use of a space-time synchronized family of grids.

1.4.2.1 Fixed grid case

Let us assume for simplicity that $\Omega = [-1,1]^2$, that is the unit square in $\mathbb{R}^2$. On Ω we define for $m \in \mathbb{N}$ a sequence of uniform discrete grids G_m of increasing cardinality and size

$$h_m = 2^{1-m} \ .$$

Then G_m will have $M_m = (2^m - 1) \times (2^m - 1)$ internal nodes over a total number of $(2^m + 1) \times (2^m + 1)$. Note in particular that $G_m \subset G_{m+1}$, so that if a node is in G_m it will belong to every G_s with $s > m$. Concerning time discretization, on the grid G_m we adopted the time step

$$dt_m = 4^{-m},$$

in order to keep constant the parabolic ratio γ between the steps, for any m: $\gamma = \frac{dt_m}{h_m^2} = 1/4$. Different step ratios can be adopted, but over a certain threshold (for example if $\gamma \geq 1$) the solution starts to oscillate and the convergence to the target is lost.

We call x_{ij} the generic node belonging to the i-th row and the j-th column of the square mesh G_m, and $t_m^k = kdt_m$ the instant times where we compute the solution: then by $(u_m^k)_{ij}$ (or simply by u_{ij}^k when m is fixed over all the computation) we denote the discrete solution at time t_m^k in a node x_{ij} of G_m. The initial data will be given by

$$\begin{aligned} &u_{ij}^0 = u^0(x_{ij}),\ u_{ij}^c = u^c(x_{ij}),\ \text{with}\ \ u_{ij}^0 > u_{ij}^c \text{ and } x_{i,j} \in G_m, \\ &\text{while } u_{ij}^0 = u_{ij}^c \ \text{ on boundary nodes.} \end{aligned} \tag{1.87}$$

A possible numerical solution of (1.82) on the grid G_m for a fixed m is the solution of the following nonlinear implicit finite-difference scheme:

Scheme (S1): *Until* $\|u^k - u^c\|_\infty < tol,$ *for any* $k = 0, 1, ...,$ *for any* $x_{ij} \in G_m,$ *solve* :

$$\begin{cases} u_{ij}^{k+1} = u_{ij}^k + dt_m \Delta_m \left(\eta_{ij}^{k+1}\right) := u_{ij}^k + \gamma \left(\eta_{i+1,j}^{k+1} + \eta_{i-1,j}^{k+1} - 4\eta_{i,j}^{k+1} + \eta_{i,j+1}^{k+1} + \eta_{i,j-1}^{k+1}\right) \\ u_{ij}^{k+1} = u_{ij}^k = u_{ij}^c \qquad \text{at } x_{ij} \in \partial G_m \end{cases} \tag{1.88}$$

where *tol* is a given tolerance parameter, Δ_m denotes the usual 5-point discrete Laplacian operator over G_m, and of course $\eta_{ij}^k := \eta\left(u_{ij}^k - u_{ij}^c\right)$. Note in particular that $\eta_{ij}^k = 0.5$ on ∂G_m.

For any time step the implicit system (1.88) is solved by the vectorial Newton method, finding (more precisely approximating) a vector $z^* \in \mathbb{R}^{M_n}$, such that $F(z^*) = 0$, where $F : \mathbb{R}^{M_n} \to \mathbb{R}^{M_n}$ is defined by

$$F(z) = z - U^k + \gamma A\eta(z - U^c) - \gamma g \tag{1.89}$$

and U^k, U^c denote respectively the vectors of the values of u^k and u^c at the internal nodes of G_m (taken in the lexicographic order), $A = -\Delta_m$ is the block tridiagonal matrix associated to the 5-point discrete Laplacian on that mesh, and g is a correction vector which takes into account the boundary values of η where $u = u^c$. In practice for any internal node close to the square sides we have to set $g = 0.5$, $g = 1$ on the four vertices and $g = 0$ elsewhere. Since we have chosen a regular function η we may assume $F \in C^1$, so that its Jacobian matrix $J_F(z)$ is well defined, and the Newton method becomes:

$given\ z_0 = U^k, \quad for\ n = 0, 1, ..., itmax:$

1. *solve the linear system* : $J_F(z_n)\delta z_n = -F(z_n)$,
2. *set* : $z_{n+1} = z_n + \delta z_n$,
3. if $||z_{n+1} - z_n|| < \tau$ *STOP and set* : $U^{k+1} = z_{n+1}$.

Here $||.||$ denotes the euclidean norm in $\mathbb{R}^{M_n}$, τ is an assigned tolerance, and *itmax* is the maximum number of iterations.

We want to prove that the discrete solution of our scheme satisfies the same evolution properties of the solution of the continuous problem, in particular that it always remains over the target and asymptotically reaches it. For simplicity we will prove such a result in the one dimensional case. What we say can be easily extended to the general 2D case.

So let $\Omega = (-1, 1)$, and G_m the mesh over Ω with $2^m + 1$ equally spaced nodes x_j, with $x_0 = -1$ and $x_{2^m} = 1$. Then the 1D version of (1.88) reduces to

$$u_j^{k+1} = u_j^k + \gamma(\eta_{j-1}^{k+1} - 2\eta_j^{k+1} + \eta_{j+1}^{k+1}) = u_j^k + \gamma\ \delta\eta_j^{k+1} \tag{1.90}$$

for any internal node x_j, $j = 1, 2, ..., 2^m - 1$. For a vector $z = \{z_j\}$ defined over G_m we have set $\delta z_j := z_{j-1} - 2z_j + z_{j+1}$, so that the discrete Laplacian operator applied to z_j has now the form of the second order central difference

$$\Delta_m z_j = \frac{\delta z_j}{h_m^2} := \frac{z_{j-1} - 2z_j + z_{j+1}}{h_m^2}.$$

Of course $\eta_j^k := \eta(u_j^k - u_j^c)$, for the choosen relaxed Heaviside function η. Note that if we set $w_j^k := u_j^k - u_j^c$, by subtracting u_j^c from both sides of (1.90) such equation can equivalently be written as

$$w_j^{k+1} = w_j^k + \gamma\ \delta\eta_j^{k+1}. \tag{1.91}$$

The following result holds true:

Theorem 1.4.1. *Let m be fixed; then the solution $u^k = \{u_j^k\}_j$ of (1.90) on G_m satisfies the following properties:*

1. *for any k and j: $u_j^k \geq u_j^c$, $1 \geq \eta_j^k \geq 0.5$ (the solution remains always supercritical);*

2. *for any k: $\sum_j w_j^{k+1} \leq \sum_j w_j^k$ (the global distance from the target decreases);*

3. *as $k \to \infty$, for any j: $w_j^k \to 0$, $\eta_j^k \to 0.5$ (the solution converges to the target).*

Proof. Taking into account the boundary values for η, any iteration of scheme (1.90) corresponds to solve the following nonlinear system:

$$u^{k+1} = u^k - \gamma A\eta^{k+1} + \gamma g, \tag{1.92}$$

where the already defined correction vector g has in 1D all null components except the first and the last (equal to 0.5), and A reduces to the tridiagonal symmetric positive definite matrix with all 2 on the main diagonal and -1 on the sub and super diagonals. Note that $Af = g$, with f the constant vector with all the values equal to 0.5. Then (1.92) becomes

$$u^{k+1} = u^k - \gamma A(\eta^{k+1} - f), \tag{1.93}$$

showing that the evolution of the system could stop only when η^k tends to f (that is when u^k tends to u^c in all the considered nodes). Moreover, since A is a known monotone matrix, in the worst case, that is when the whole vector u^k decreases, then, from (1.93), $A(\eta^{k+1} - f) \geq 0$, yielding $\eta^{k+1} \geq f$: that is, the solution remains supercritical ($u^k \geq u^c$) since all the values of η^k belong to the interval $[0.5, 1]$. Of course, when some component of u^k grows, the corresponding value of η^k cannot decrease, for the monotonicity of η. This ends the proof of point *1.*

Adding up (1.91) for any index j we easily get

$$\sum_{j=1}^{2^m-1} (w_j^{k+1} - w_j^k) = \gamma \sum_{j=1}^{2^m-1} (\eta_{j-1}^{k+1} - 2\eta_j^{k+1} + \eta_{j+1}^{k+1}) = \gamma(1 - \eta_1^{k+1} - \eta_{2^m-1}^{k+1}) \leq 0, \tag{1.94}$$

since $\eta_1^k, \eta_{2^m-1}^k \geq 0.5$. This immediately proves point *2.* It also says that if the distance w^k from the target grows at some node it has to decrease at least at another one in such a way that the global distance cannot grow. Note that (1.94) can be interpreted as the 1D discrete version of (1.85): on the uniform grid G_m, it is enough to approximate the integrals by the repeated rectangle quadrature formula of step h_m. The quantity

$$M_m^k = h_m \sum_j w_j^k = h_m \sum_j (u_j^k - u_j^c) \tag{1.95}$$

then represents the discrete version of the mass $M(t)$ at the iteration k on the grid G_m. From points *1.* and *2.* we see that $M^k \to \overline{M} \geq 0$. To conclude the proof of the theorem it is then sufficient to prove that $\overline{M} = 0$. Such result would easily follow if we could prove that $u^k \to \tilde{u}$ for some vector $\tilde{u}$, because in that case $\eta^k = \eta(u^k) \to \eta(\tilde{u}) = f$, yielding $\tilde{u}_j = u_j^c$ for any j. But we only know that the sum of the distances converge. In principle we could have that single components of u^k do not have limit but tend to some loops which do not alter the sum. Such situation can be excluded as a consequence of (1.94): passing to the limit we see in fact that necessarily both η_1^k and $\eta_{2^m-1}^k$ have to converge to 0.5. This implies for example that the solution cannot oscillate in x_2 (otherwise it should do that also in x_1); then even $\eta_2^k \to 0.5$, and so on, with the same reasoning for any subsequent node. □

In the numerical simulations it is easy to see that the convergence of u^k to u^c is not in general monotone; but we can state the following proposition:

Proposition 1.4.1. *Assume that:*

(*) there exists an index k_0 such that $u_j^{k_0} \geq u_j^{k_0+1}$ holds true for any j;

then for any $k > k_0$ the solution of (1.90) satisfies:

1. $u_j^{k_0} \geq u_j^k \geq u_j^{k+1} \geq u_j^c, \quad 1 \geq \eta_j^{k_0} \geq \eta_j^k \geq \eta_j^{k+1} \geq 0.5, \quad \forall x_j \in G_m$;
2. $-\sum_j \delta\eta_j^k \geq -\sum_j \delta\eta_j^{k+1}$, $\quad -\sum_j \delta\eta_j^k \to 0$.

In other words after the iteration k_0 all the quantities u_j^k, η_j^k and even the discrete version of $E(t)$, that is $E_m^k = -\sum_j \delta\eta_j^k/h_m$ converge in a monotone way.

Proof. It is clear that if u^k decreases also w^k decreases, and η^k does the same, for the monotonicity of the η function. So, to claim *1.* we use an induction argument. Let us suppose that for an index $k > k_0$ the property $u_j^k \geq u_j^{k+1}$ holds true for any j; we want to prove (reasoning by contradiction) that also $u_j^{k+1} \geq u_j^{k+2}$ for any j. Such result together with (*) will prove *1.*

If it is false, it should exist at least an index j such that $u_j^{k+2} > u_j^{k+1}$; then $\eta_j^{k+2} \geq \eta_j^{k+1}$, and (from (1.90))

$$\delta\eta_j^{k+2} > 0;$$

let us show that in such a case the solution has to grow also at an adjacent node; by the induction assumption:

$$\delta\eta_j^{k+1} \leq 0 \quad (\text{since } u_j^{k+1} \leq u_j^k);$$

then

$$\eta_{j-1}^{k+1} + \eta_{j+1}^{k+1} \leq 2\eta_j^{k+1} \leq 2\eta_j^{k+2} < \eta_{j-1}^{k+2} + \eta_{j+1}^{k+2} \ ,$$

and η^{k+1} has to grow, at least at one of the two adjacent nodes; if it grows at both, let us consider anyway the one where the growth is larger, and assume for example it is x_{j+1}. The previous argument shows also that, if η^{k+1} grows of a positive quantity α_j at x_j, then it has to grow at x_{j+1} of a quantity $\alpha_{j+1} > \alpha_j$; in fact:

$$\delta\eta_j^{k+2} = -2(\eta_j^{k+1} + \alpha_j) + (\eta_{j-1}^{k+1} + \alpha_{j-1}) + (\eta_{j+1}^{k+1} + \alpha_{j+1}) =$$

$$= \delta\eta_j^{k+1} - 2\alpha_j + \alpha_{j-1} + \alpha_{j+1} > 0,$$

so that

$$\alpha_{j-1} + \alpha_{j+1} > 2\alpha_j - \delta\eta_j^{k+1} \geq 2\alpha_j \quad (\text{for the induction hypothesis}),$$

and we can deduce that $\alpha_{j+1} > \alpha_j$ (since $\alpha_{j+1} > \alpha_{j-1}$); but a similar argument can be repeated now at the node x_{j+1}, showing that necessarily η^{k+1} has to grow also at node x_{j+2} by a quantity $\alpha_{j+2} > \alpha_{j+1}$. In fact:

$$\delta\eta_{j+1}^{k+2} = -2(\eta_{j+1}^{k+1} + \alpha_{j+1}) + (\eta_j^{k+1} + \alpha_j) + (\eta_{j+2}^{k+1} + \alpha_{j+2}) =$$

$$= \delta\eta_{j+1}^{k+1} - 2\alpha_{j+1} + \alpha_j + \alpha_{j+2} > 0,$$

so that

$$\alpha_{j+2} > 2\alpha_{j+1} - \alpha_j - \delta\eta_{j+1}^{k+1} > \alpha_{j+1} + (\alpha_{j+1} - \alpha_j);$$

in such a way we should have that u^{k+2}, and then η^{k+2} has to grow at all the subsequent nodes up to the node x_{M-1} close to the right boundary (here $M = 2^m$), where

$$\delta\eta_{M-1}^{k+2} = -2(\eta_{M-1}^{k+1} + \alpha_{M-1}) + \frac{1}{2} + (\eta_{M-2}^{k+1} + \alpha_{M-2}) > 0\ ,$$

yielding $(2\alpha_{M-1} - \alpha_{M-2}) < 0$, which is impossible for what we have stated before. So our initial assumption has led us to a contradiction, and we can conclude by induction that $u^{k+1} \geq u^{k+2}$, as we wanted to prove.

As η_j^k decreases and tends to 0.5 for all j, point *2.* is an easy consequence of (1.94) and *1.*, since $-\sum_j \delta\eta_j^k = \eta_1^k + \eta_{2^m-1}^k - 1$. □

Note that (*) holds true for $k_0 = 0$ if, for example, function $\eta = \eta_1$ from (1.79) and $\eta_j^0 = 1$ for any internal node x_j: in that case at the first iteration the solution cannot grow. If it would happen at a node x_j then necessarily $\eta_j^1 = \eta_j^0 = 1$, but from (1.90)

$$\delta\eta_j^1 = \eta_{j-1}^1 - 2\eta_j^1 + \eta_{j+1}^1 > 0 \quad \Rightarrow \quad \eta_{j-1}^1 + \eta_{j+1}^1 > 2,$$

which is impossible since $\eta_j \leq 1$ for any j. Then $u_j^0 \geq u_j^1 \geq u_j^c$ for any j. In (1.82) the Laplacian of the Heaviside plays the role of a sort of switch for the dynamics, so it is not surprising that nothing changes in the discrete model at a node where $\delta\eta_j = 0$: this happens of course when η is constant on the three-point stencil (that is when $\eta_{j-1} = \eta_j = \eta_{j+1}$), or more generally when it is harmonic (linear, in 1D) on it. If $\eta = \eta_1$ and the starting values η_j^0 on the internal nodes of G_m are all equal to 1, the only values of the solution which initially decrease towards the target (by $\gamma/2$ at each iteration) are the ones at the two nodes closed to the boundary: for example at x_1 one has $\delta\eta_1 = 0.5 - 2 + 1 = -0.5 < 0$. At the more internal nodes the solution does not change. When the distance to the target becomes lower than $1/n$ the value η_1 starts to relax monotonically from 1 until it reaches 0.75 (since there $\delta\eta_1 = 0.5 - 1.5 + 1 = 0$). It is at that time that the dynamics is impulsively activated at x_2 and the value u_2^k starts to decrease towards u_2^c (where $\delta\eta_2 < 0$), while (η_1, η_2) relax from $(0.75, 1)$ towards the values $(0.667, 0.833)$ (so that again $\delta\eta_1 = \delta\eta_2 = 0$ and $\delta\eta_3 < 0$). Then u_3^k starts to move, and so on. One node at the time, from the boundary towards the interior of Ω, all the values of the solution tend to the corresponding values of u^c. When all the differences $u_j^k - u_j^c$ become smaller than $1/n$ the final relaxation of η rapidly starts, with all the values of η_j tending together to 0.5 and the dynamics stops. If one uses other approximations for the Heaviside function, as η_2 or η_3, the decay activation can be less sharp but the behavior remains essentially the same.

In analogy with what we said for the continuous problem, it follows also directly from (1.90) that (*) holds if at some iteration the vector η^k is such that $\delta\eta_j^k \leq 0$ for any j (a sort of discrete superharmonicity of η^k at any node). This is not true in general from the beginning: for some iterations it could happen that the distance of the solution from the target increases at certain nodes, according to the data of the problem. This is the case for example if $u_j^0 = u_j^c$ at an internal node: the solution detaches from the target for a while before going back. Anyway, we could say that in a finite number of iterations, due to the boundary conditions, condition (*) is automatically satisfied.

We conclude this Section with a couple of remarks. If the Heaviside under the laplacian has the role of a switch, its approximated versions (1.79) or (1.81) do much than this: through their relaxed values in $[0.5, 1]$ they are able to slow down the descent when the distance to the target tends to zero, avoiding jumps to subcritical values for the solution. Moreover, denoting by t^* the time where the solution reaches the target up to a prescribed tolerance on the given grid, we have that when the parameter n of the relaxed Heaviside funcion η grows, then t^* decreases. The closer is the η to the real Heaviside function, the faster is the convergence to the target. Unfortunately there are stability limits for this process, as we will see in the tests, and n cannot be choosen arbitrarily large.

1.4.2.2 Space-time synchronized grids case

In the previous Section we have seen that on a given (fixed) grid the scheme (1.88) (equivalently, (1.90) in the 1D case) is able to bring in time all the values of the initial solution towards the corresponding target values. If one desires great accuracy with respect to the target, then a large m has to be choosen, with a consequent higher computational cost. Moreover, since the convergence progresses from the boundary towards the interior, one layer at a time, the entire process is heavily slowed down. That is why we tested a modified approach (partially inspired by [46]), based on an increasing family of grids G_m, with a precise synchronization of space and time steps which keeps constant the parameter γ.

The idea is the following. Scheme (1.88) is activated on an initial coarse grid G_{m_0}, for a given time interval $(0, T]$. For example, if $m_0 = 2$ and $T = 1$, we start to compute the solution with initial datum $u^*_{m_0} = u^0$ on 25 nodes, 9 of which internal, for 16 instant times (since on G_2 $h_2 = 1/2$ and $dt_2 = 1/16$).

When the scheme has been completed for all the instants of the first time iterval the spatial grid is refined, from G_{m_0} to G_{m_1} (with $m_1 = m_0 + 1$), and we need to set a new initial value on it, the function $u^*_{m_1}$, in order to let the evolution start again on another time interval of length T. On the old nodes of G_{m_0}, which still belong to G_{m_1}, we simply use the values reached by the solution u_{m_0} at the last time iteration, but on the added nodes (those in $G_{m_1} - G_{m_0}$) we have to introduce new values. One cannot make a direct use of the initial function u_0: since some time has passed we need a sort of actualization of it, and different strategies are possible for such an update process, keeping in mind that we are running in direction of the target. Then the scheme starts again with a reduced time step. In the previous example, after the refinement to G_3, the solution is computed on 81 nodes, 49 of which internal, for 64 instant times in $(1, 2]$ (since on G_3 now $h_3 = 1/4$ and $dt_3 = 1/64$). In such a way the process is repeated on a sequence of synchronized increasing grids, up to the desired finest grid G_N, where a suitable stopping criterium is imposed for the desired tolerance. In order to accelerate the process, one could introduce such criterium even at any previous grid, anticipating the refinements when the tolerance is achieved on the relative nodes.

The simplest update strategy for u^*_m is to use a convex combination of the two initial data u^0 and u^c with a suitable actualization coefficient $\lambda(m)$:

$$(u_m^*)_{ij} = u_{ij}^c + \lambda(m)(u_{ij}^0 - u_{ij}^c); \tag{1.96}$$

we adopted for example $\lambda(m) = 1/4^{m-m_0}$, in such a way that the values of u_m^* tend towards those of u^c as m increases. Note that by this formula all the new values are naturally supercritical, property which is not guaranteed for example by the simple interpolation of the values of u at the adjacent old nodes. But by using the same parameter $\lambda(m)$ for any new node we do not take care of the different decay rates towards the target achieved on the previous grid. The consequence is that after the refinement some spurious oscillations can be introduced in the dynamics, at least for some iterations, before that the monotonicity behavior described in Proposition 1.4.1 comes back to act.

Such phenomenon can be reduced or even prevented by adopting other update strategies which make use of local values for $\lambda(m)$. Here we limit ourselves to a couple of them in the simple 1D case.

In the first one we set for each new node x_j of G_m:

$$(u_m^*)_j = u_j^c + \lambda(m)_j(u_j^0 - u_j^c), \quad \lambda(m)_j = \min_{s=j\pm1} \frac{(u_{m-1}^{last})_s - u_s^c}{u_s^0 - u_s^c}\ , \tag{1.97}$$

where the quotient represents the decay rate of the solution at a node x_s of G_{m-1}. In other words the best decay rate of the adjacent nodes is adopted.

Another idea is the following: in order to avoid oscillations after the update, we choose the values of η at a new node x_j in order to have a zero discrete Laplacian on it, that is

$$(u_m^*)_j = \eta^{-1}[(\eta_m^*)_j] + u_j^c, \quad \text{with} \quad (\eta_m^*)_j = \frac{(\eta_{m-1})_{j+1} + (\eta_{m-1})_{j-1}}{2}, \tag{1.98}$$

where η^{-1} denotes the inverse function of η in the range $[0.5, 1]$. With formula (1.98) the monotonicity property of Proposition 1.4.1 is preserved even after any refinement. As an example, when $\eta = \eta_3$ we easily get the explicit formula:

$$(u_m^*)_j = \frac{\tan(\pi((\eta_m^*)_j - 0.5))}{n} + u_j^c.$$

Summing up, the new scheme on the set of synchronized grids can be expressed by:

Scheme (S2): for $m = m_0, ..., N$

- compute $u_m^0 = \begin{cases} u_0 & \text{if } m = m_0 \\ u_m^* & \text{if } m > m_0 \end{cases}$
- while $k \leq T/dt_m$ on G_m, for any $x_{ij} \in G_m$

$$\begin{cases} (u_m^{k+1})_{ij} = (u_m^k)_{ij} + dt_m \Delta_m (\eta_m^{k+1})_{ij} \\ \\ (u_m^{k+1})_{ij} = (u_m^k)_{ij} = u_{ij}^c \quad \text{at } x_{ij} \in \partial G_m \end{cases} . \tag{1.99}$$

- if $\|u_N^k - u^c\|_\infty = \max_{ij} |(u_N^k)_{ij} - u_{ij}^c| < tol$ on $G_N \Rightarrow$ STOP.

As usual $(\eta_m^k)_{ij} = \eta((u_m^k)_{ij} - u_{ij}^c)$ for the corrent active m, with $(\eta_m^k)_{ij} = 0.5$ on ∂G_m. The scheme then works on any grid for a fixed interval of time T, and u_m^k represents the approximate solution on G_m at time $t = (m - m_0)T + kdt_m$ (or even earlier if the stopping criterium is imposed on any grid).

We will see in the next Section that Scheme (S2) usually speeds up the convergence towards the target, reducing at the same time the computational cost in a considerable way.

1.4.3 Numerical tests

In this Section we report the results of some numerical tests made applying schemes (S1) or (S2). We start with the simplified one dimensional case, in order to compare more easily the performance of the two approaches. Now $\Omega = (-1, 1)$, and problem (1.82) reduces to

$$u_t = (\eta(u - u^c))_{xx} \text{ in } \Omega \times (0, \infty), \quad u(0) = u^0 \text{ in } \Omega, \quad u = u^c \text{ on } \partial\Omega \times [0, \infty) ,$$

where η denotes the approximated Heaviside function (one of those defined in Section 1.4).

It means that for every internal node x_i of G_m we solve scheme (1.90), with, as usual, $\gamma = 1/4$. At any time step such nonlinear implicit system is solved via the Newton method (with a maximum number *itmax* of iterations and a given tolerance τ, usually $itmax = 5$ and $\tau = 10^{-5}$). We compare the behavior of solution of Scheme (S1) on the fixed grid G_N with the one of Scheme (S2) on the increasing grid sequence $\{G_2, \ldots, G_N\}$ up to the same final accuracy: the computation ends when the distance between u_N^k and u^c in the infinity norm becomes less than a given tolerance. For (S2) it means in particular that the scheme will run on each mesh for a fixed interval of time or until the given tolerance is achieved, followed by a suitable mesh refinement and an upgrade process, up to the final grid G_N.

TEST 1. $u^c(x) = 1 - |x|, \quad u^0(x) = 5(1 - |x|) \quad$ [see Fig.1.3].

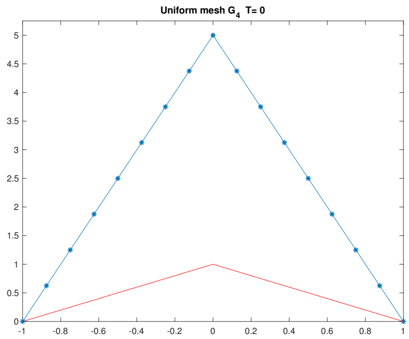
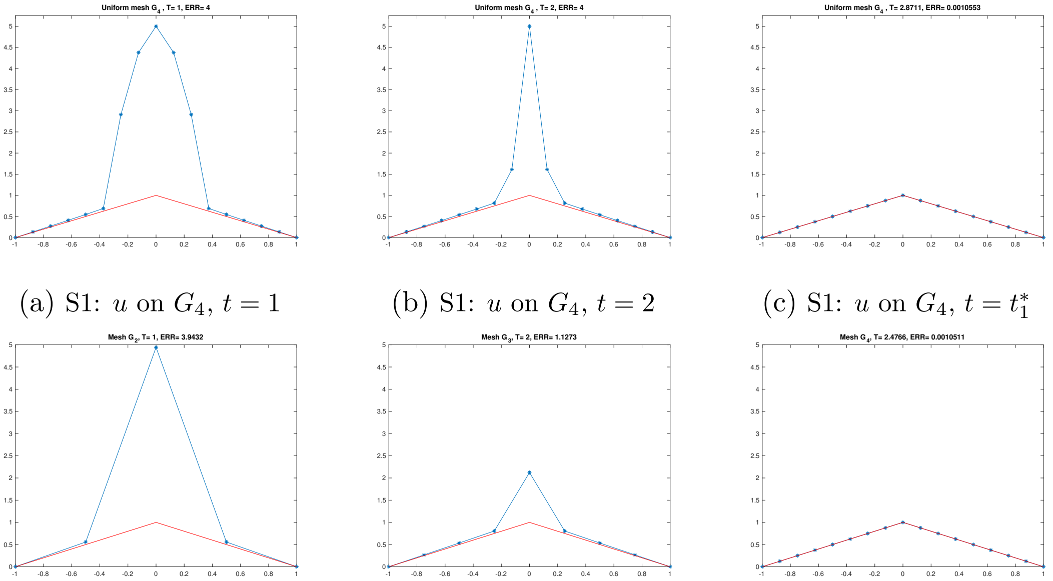


Figure 1.3: TEST 1. Initial datum (blue) and target function (red)

In Fig.1.4 we compare the evolution of the solution of (S1) on G_4 with the one of (S2) on $\{G_2, G_3, G_4\}$ at the same time values: in both cases it progressively gets closer (decreasing) to u^c, from the boundary towards the interior.

Since we adopted the function η_1 of (1.79) this essentially happens one node at a time. Each value of u_0 moves from its initial position only when the solution at an adjacent node becomes sufficiently close to the critical one. The convergence is monotone also in η^k as stated by Proposition 1.4.1: its values, which at the beginning are equal to 1 at any internal node (since u_0 was largely supercritical), slowly relaxe towards the value 0.5 which characterizes the contact points. It means that the solution remains supercritical along the whole process. In (S2) a mesh refinement is applied after the same fixed time interval (in this case equal to one).

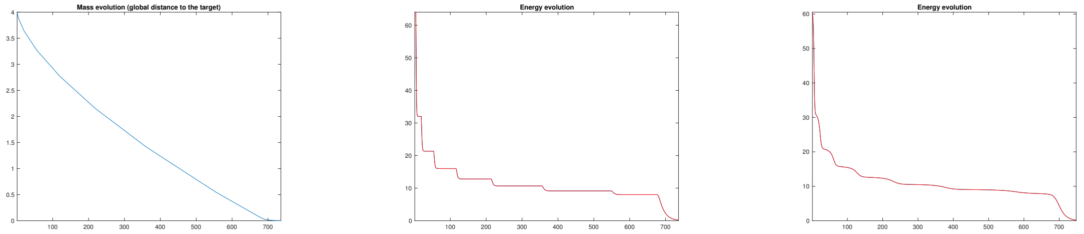
In Fig. 1.5 we show the evolution of the quantities M_m^k and E_m^k (discrete counterpart of (1.86)) for $t = t^k$ in G_m; look in particular to the impulsive character produced by the specific choice of η_1 in Fig.1.5(b). In this test the assumptions of Proposition 1.4.1 are satisfied with $k_0 = 0$, and both the quantities decrease in a monotone way.



(a) S1: u on G_4, $t = 1$ (b) S1: u on G_4, $t = 2$ (c) S1: u on G_4, $t = t_1^*$

(d) S2: u_2 on G_2, $t = 1$ (e) S2: u_3 on G_3, $t = 2$ (f) S2: u_4 on G_4, $t = t_2^* < t_1^*$

Figure 1.4: TEST 1. Evolution of u, (S1) versus (S2) (η_1, $tol = 10^{-3}$, update (1.96))



(a) M^k (b) E^k with η_1 (c) E^k with η_3

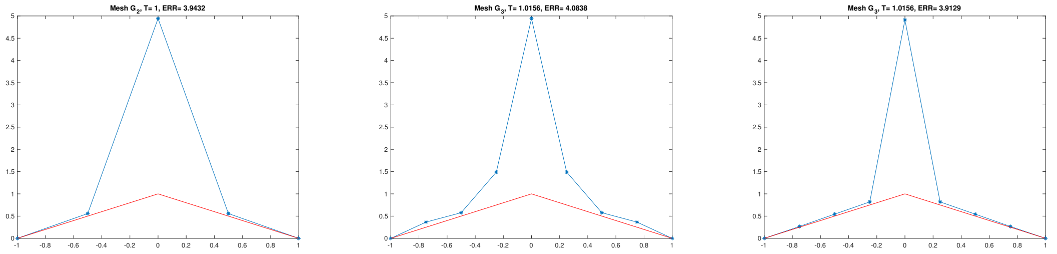
Figure 1.5: TEST 1. M^k and E^k evolution with (S1) on G_4.

In Table 1.1 it is possible to compare the cost performances of the two schemes for this test: for any scheme we reported the stopping time t^*, the number of internal nodes involved for any grid and the time iterations executed over each of them. This gives (in the last column) a total number of solution evaluations up to the stop. The stopping time is much shorter for the second scheme, in particular when the anisotropic update (1.97) is used, with an evaluation reduction of more than the 80% !

Finally in Fig. 1.6 we compare the different effects of the update formulas (1.96) and (1.97) on u_3^* after the refinement from G_2 to G_3, showing the advantage of the second approach.

Table 1.1: TEST 1. Performance comparison of schemes (S1) and (S2).

scheme	t^*	grid	nodes	time it.	evals	tot. evals
S1	2.8711	G_4	15	736	11040	11040
		G_2	3	16	48	
		G_3	7	64	448	
S2 + (1.96)	2.4766	G_4	15	123	1845	2341
S2 + (1.97)	2.2695	G_4	15	70	1050	1546



(a) u_2 on G_2, $t = 1$ (b) u_3^* with (1.96) (c) u_3^* with (1.97)

Figure 1.6: TEST 1. Scheme (S2): the update process after $G_2 - G_3$ refinement

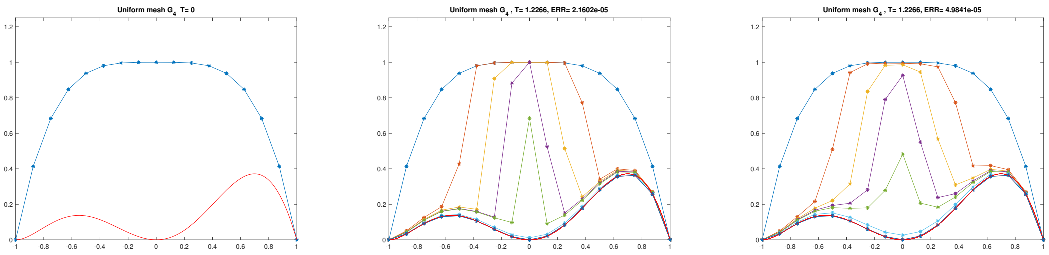
We remember that the stopping time does not depend on the specific choice of the approximated Heaviside function η, but only on the value of $\eta'(0)$, which in some sense determines the accuracy of the approximation. The following Table 1.2 shows, for equal values of $\eta'(0)$ in the three considered cases for η, how the stopping time on G_4 decreases as n grows, in this case with a tolerance of 10^{-3}. Since when n grows η approaches the real Heaviside function, one could imagine that t^* would converge to the finite time of convergence of the continuous model (1.7). Unfortunately, due to stability problems, when n becomes too large, the discrete solution can easily overcome the target in some nodes, causing oscillations and loops, so that the convergence gets lost.

Table 1.2: TEST 1. Stopping time t^* for different η functions on G_4 when n grows.

$\eta'(0)$	η_1	η_2	η_3
7.5	2.8711	2.8789	2.918
15	2.7656	2.7656	2.7891
30	2.7188	2.7188	2.7305

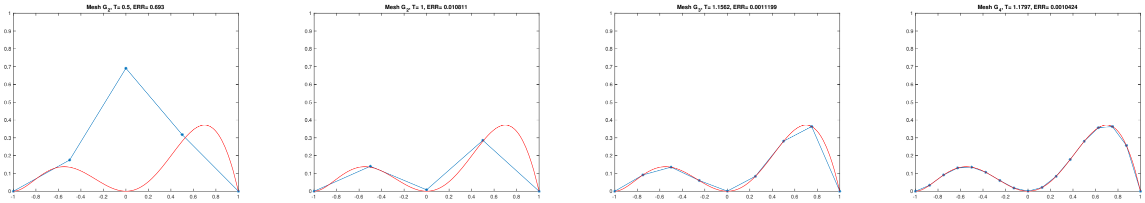
TEST 2. $u^c(x) = 1.5x^2(1-x^2)\sin(1+x)$, $\quad u^0(x) = (1-x^4)$ $\quad$ [see Fig.1.7 (a)].

In this example the target is not symmetric, and the initial datum has no relation with it at all. Even in this case we have that the solution evolution with scheme (S1) is monotone (see Fig.1.7(b-c)). Note the little different behavior induced by the choosen η function, even if the final stopping time is essentially the same. In Fig.1.8 we see how the scheme (S2) works in this case: the solution gets close to the target already on the coarsest grid, so that the refinements only act to extend the accuracy to the whole increasing set of nodes.



(a) Test 2: initial data (b) (S1) on G_4 with η_1 (c) (S1) on G_4 with η_3

Figure 1.7: TEST 2. (in b) and c) the solution of (S1) is plotted every 45 iterations.



(a) G_2, $t = 0.5$ (b) G_2, $t = 1$ (c) G_3, $t = 1.15$ (d) G_4, $t = 1.175$

Figure 1.8: TEST 2. Solution evolution with scheme (S2)

On this example we tested the time of convergence of the solution to the target, which of course depends on the choosen tolerance for the stopping criterium. Experiments show that its rate is exponential (see Table 1.3.a below). We could say that in the discrete case the finite time of convergence to the target is the one corresponding to the machine precision (of order 10^{-16}). On the contrary, we see that the final time does not grow with the choosen order of the grid G_N of interest (Table 1.3.b below).

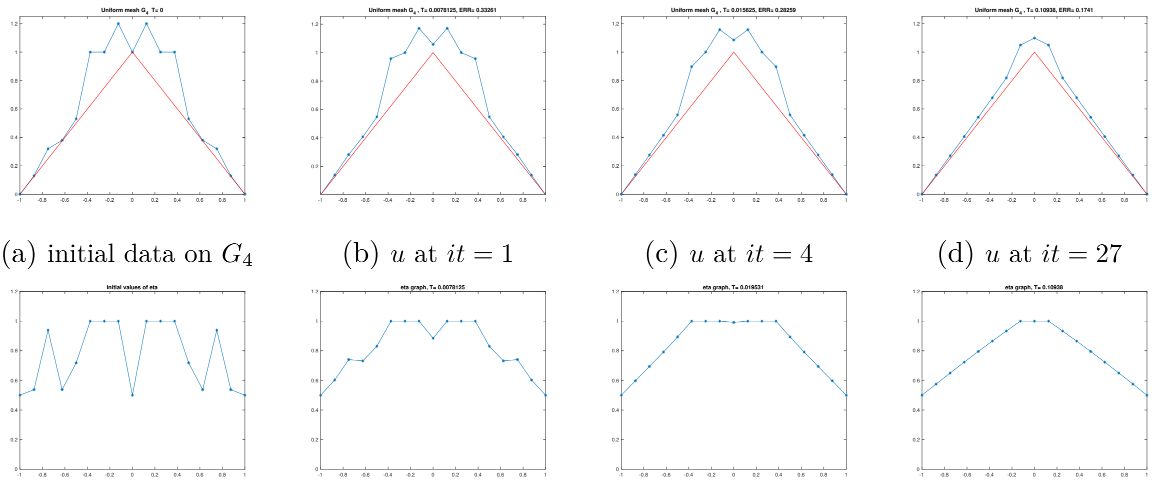
Table 1.3: TEST 2. (η_1, $n = 10$) Stopping time for (S1) versus a) tolerance, b) grid order.

tol	stopping time (on G_4)
10^{-4}	1.1367
10^{-8}	1.6562
10^{-12}	2.1719
10^{-16}	2.7007

grid	stopping time ($tol = 10^{-5}$)
3	1.3281
4	1.2695
5	1.2539
6	1.2507

TEST 3. $u^c(x) = 1 - |x|, \quad u^0$ given vector [Fig.1.9 (a)]

In this case the initial datum is a vector of values on G_4 very close to the target at some nodes, and even equal to that at $x = 0$. Looking at the graph of η^0 (Fig.1.9 (e)) we can see that $\delta\eta^0_j > 0$ at some nodes: there the solution initially grows (in particular it detaches from the target at the center).



(a) initial data on G_4 (b) u at $it = 1$ (c) u at $it = 4$ (d) u at $it = 27$

(e) initial η on G_4 (f) η at $it = 1$ (g) η at $it = 4$ (h) η at $it = 27$

Figure 1.9: Test 3. Scheme (S1), u versus η

But after a few iterations $\delta\eta^k_j \leq 0$ at any node (Fig.1.9 (h)) and it starts to decrease everywhere, entering in the 'monotone regime'. This behavior is confirmed from the corresponding evolution of E^k (Fig.1.10) which starts decreasing only after some iterations.

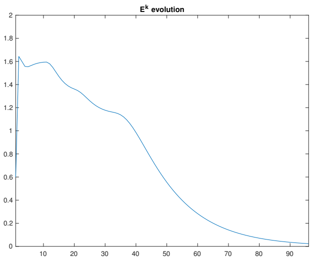


Figure 1.10: TEST 3. E^k evolution

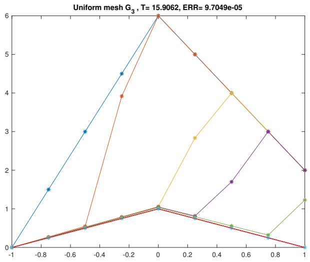


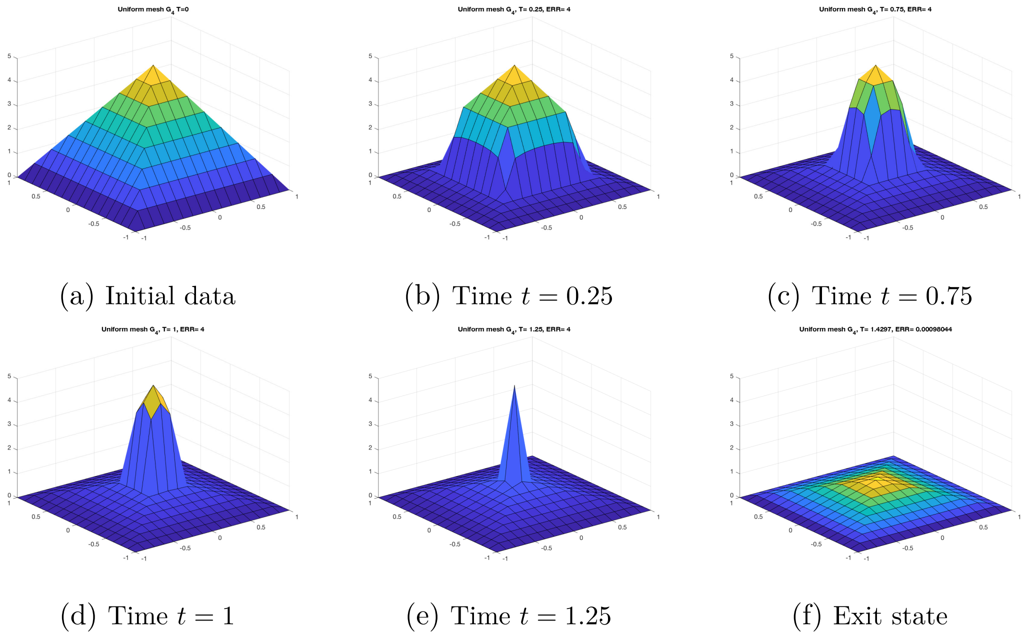
Figure 1.11: TEST 4. Solution evolution with (S1) on G_3, times $t = 1, 6, 9, 13$

TEST 4. $u^c(x) = 1 - |x|, \quad u^0 = 5(1 - |x|) + 1 + x \quad$ [Fig.1.11]

In this example the contact between u^0 and u^c is only at the left boundary point. Nevertheless this is enough to produce the transmission of information up to the opposite side, so that all the solution is again assorbed by the target. It is sufficient to impose a homogeneous Neumann boundary condition on the opposite side (see Fig.1.11). In other words what is requested in order to start the global process is a boundary contact of the initial datum with the target.

TEST 5. $u^c(x, y) = 1 - \frac{|x+y|+|y-x|}{2}, \quad u^0(x, y) = 5\left(1 - \frac{|x+y|+|y-x|}{2}\right) \quad$ [Fig.1.12 (a)]

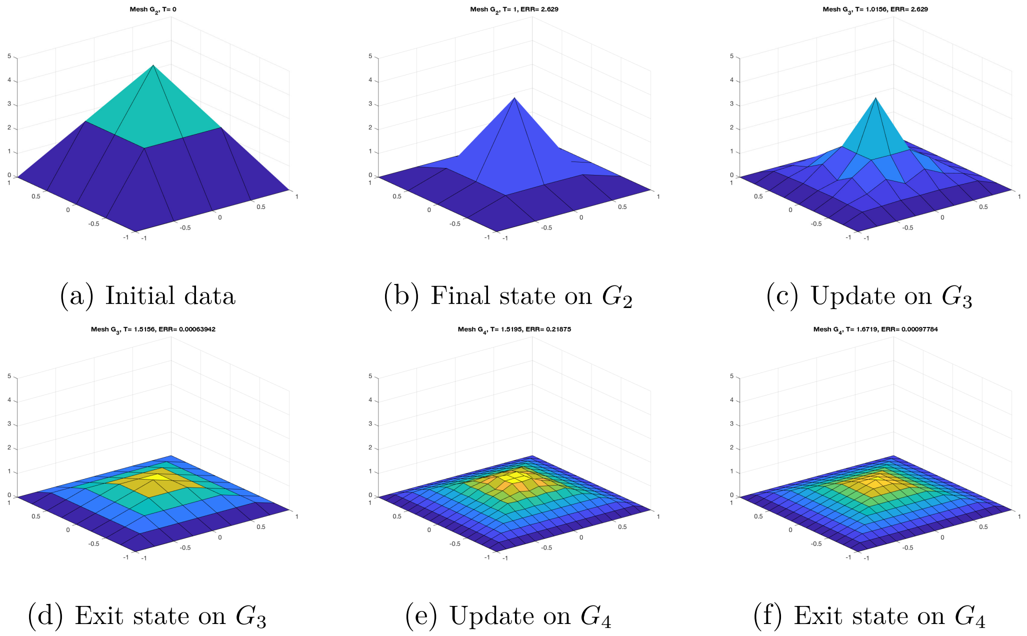
We end with a test on the more general 2D case, where all the previous behaviors are confirmed. We generalized the initial data choice of TEST 1 assuming as data two square pyramids with different height and common basis the square $\Omega = (-1, 1)^2$, and using η_1. With scheme (S1) on G_4 the highest pyramid collapses towards the smallest one as time grows, from the boundary of the square towards its center (see Fig.1.12). With scheme (S2) (Fig.1.13) this happens already on a reduced number of nodes (on G_3), then with much less computations. The update process at any change of mesh (here given by formula (1.96)) introduces some little oscillations which only for a very short time disturb the monotonic decay of the solution. Unfortunately the other update strategies (1.97) and (1.98) do not extend trivially to the 2D settings, so that the best strategy for the update remains an open problem in this case.



(a) Initial data (b) Time $t = 0.25$ (c) Time $t = 0.75$

(d) Time $t = 1$ (e) Time $t = 1.25$ (f) Exit state

Figure 1.12: TEST 5. 2D-evolution of solution (scheme S1 on G_4)



(a) Initial data (b) Final state on G_2 (c) Update on G_3

(d) Exit state on G_3 (e) Update on G_4 (f) Exit state on G_4

Figure 1.13: TEST 5. 2D-evolution of solution (scheme S2 on $G2 - G4$)

Chapter 2

An Heaviside function driven degenerate diffusion model

Aim of this chapter is to study the following problem

$$\begin{cases} u_t - H\left(u - u^c\right)\left(\Delta u + f\right) = 0 & \text{a.e. in } \Omega,\ \text{for all } t \in (0,T) \\ u\left(0\right) = u^0 & \text{a.e. in } \Omega \\ u = 0 & \text{on } \partial\Omega,\ \text{for all } t \in (0,T) \end{cases} \tag{2.1}$$

with $T > 0$, where H is the extended Heaviside function such that $H(0) = 0$, that is

$$H(r) = \begin{cases} 1 & \text{for } r > 0 \\ 0 & \text{for } r \leq 0 \end{cases} \tag{2.2}$$

and Ω is a bounded domain in $\mathbb{R}^n$, $n \in \mathbb{N}$, with smooth boundary.

We will refer to [14, 16, 17, 18, 19, 25, 26, 32, 49, 51, 52] for further details about both obstacle problem and about the results that we will point out in order to build ours. An overview about these topics, through the basic definitions and properties, will be also given in Appendix A, page 85.

We remark that the initial datum u^0, the (independent of time) source term f and the given target function (the obstacle) u^c satisfy the following conditions

$$u^0 \in H_0^1\left(\Omega\right),\ f \in L^2\left(\Omega\right),\ u^c \in H^2\left(\Omega\right),\ u^c \leq 0 \text{ on } \partial\Omega. \tag{2.3}$$

We define as solution of problem (2.1) a function $u \in L^2\left(0,T; H_0^1\left(\Omega\right) \cap H^2\left(\Omega\right)\right)$ with $u_t \in L^2\left(0,T; L^2(\Omega)\right)$ which satisfies the conditions of the problem (2.1).

Problem (2.1) fits into the typology of degenerate parabolic problems, since the diffusion coefficient is a discontinuous function which vanishes where the solution touches the obstacle u^c.

In Section 2.1 we prove under suitable conditions the equivalence of problem (2.1) with a parabolic obstacle problem, that is with a variational inequality on the convex set of the functions above the target. As a consequence, it will be possible to characterize the asymptotic solution of the problem (2.1) as the solution of the corresponding stationary (elliptic) obstacle problem.

In Section 2.2 we discuss both a direct numerical approximation of problem (2.1) in one and two dimensions through a semi implicit finite difference scheme, either through the use of the exact Heaviside function, or of a C^1 approximation of it and we present some numerical tests both in one and two dimensions. An efficient variable time step strategy will be here also described in order to reduce the computational costs of the method.

2.1 The model

In the present Section we analyze problem (2.1), showing that, under suitable conditions, it is equivalent to a parabolic variational inequality with obstacle u^c. In such a way we will be able to prove the existence of an asymptotic solution for such a problem, and to characterize it as the solution of the corresponding stationary obstacle problem.

2.1.1 Equivalence with the obstacle problem

We start by recalling the classic parabolic obstacle problem:

$$\begin{cases} w(t) \in \mathcal{K}, \quad \displaystyle\int_\Omega (w_t - \Delta w - f)(\varphi - w)\, dx \geq 0 \quad \forall \varphi \in \mathcal{K}, \forall t \in (0,T) \\ w\,(0) = u^0 \in \mathcal{K} \end{cases} \tag{2.4}$$

where $\mathcal{K}$ denotes the convex set

$$\mathcal{K} = \left\{\varphi \in H_0^1\,(\Omega)\,, \varphi \geq u^c \text{ in } \Omega\right\} \tag{2.5}$$

and u^0, u^c, f are the same data used in (2.1). It is well known (see [17], pp. 99-102) that under the assumptions (2.3), there exists a unique solution $w = w(x,t)$ for problem (2.4), with $w \in L^2(0,T;\mathcal{K} \cap H^2(\Omega))$ and $w_t \in L^2\left(0,T;L^2(\Omega)\right)$ (see also [25] and [32]). Moreover (2.4) can be written in the equivalent formulation of a complementarity system, that holds for all $t > 0$:

$$\begin{cases} w(x,t) \geq u^c(x) & \text{a.e. in } \Omega \\ w_t \geq \Delta w + f & \text{a.e. in } \Omega \\ (w - u^c)(w_t - \Delta w - f) = 0 & \text{a.e. in } \Omega \\ w\,(x,0) = u^0 & \text{a.e. in } \Omega \\ w(x,t) = 0 & \text{on } \partial\Omega. \end{cases} \tag{2.6}$$

We point out that a fundamental result for this problem is given by the parabolic version of the Lewy - Stampacchia inequality (see e.g. [25], inequality (31) p. 119)

$$f \leq w_t - \Delta w \leq \sup(0, -\Delta u^c - f) + f. \tag{2.7}$$

We are able to prove the equivalence of problems (2.1) and (2.4) under the following assumptions:

H$_1$**:** $u^0 > u^c$ a.e. in Ω;

H$_2$**:** $\Delta u^c + f \leq 0$ a.e. in Ω.

We point out that under conditions (2.3) and **H**$_2$, if w solves problem (2.4) (hence (2.6)), then w solves (2.1). In fact, initial and boundary conditions are the same in (2.1) and (2.6). If $w > u^c$, we obtain $w_t - \Delta w - f = 0$ a.e. If $w = u^c$, then by (2.7) and **H**$_2$, $w_t - \Delta u^c \leq \sup(0, -\Delta u^c - f) + f = -\Delta u^c$ so that $w_t \leq 0$: as $w \geq u^c$, we obtain $w_t = 0$.

Then under such conditions we obtain that a solution of problem (2.1) exists. In the following Proposition 2.1.1, we prove that under the further condition **H**$_1$, problems (2.4) and (2.1) are equivalent.

Proposition 2.1.1. *Assume that u solves problem* (2.1)*, and that conditions* (2.3)*,* **H**$_1$ *and* **H**$_2$ *hold, then u coincides for any time with the unique solution w of* (2.4)*, and hence of* (2.6)*.*

Proof. Initial and boundary conditions are the same in (2.1) and (2.6).

Let us now prove that if $u(x,t)$ is a solution of (2.1) then necessarily $u(.,t) \geq u^c(.)$ in Ω for any time t (that is $u(t) \in \mathcal{K}$). It is true for $t = 0$ thanks to $\mathbf{H}_1$. Suppose that for a given $x \in \Omega$ there exists a first time t^* such that $u(x,t^*) = u^c(x)$ from above. Then, from (2.1) $u_t(x,t^*) = 0$, so that $u(x,t) \geq u^c(x)$ for any $t > t^*$. Then the first inequality of (2.6) is satisfied by u.

The equation in the third line of (2.6) is trivially satisfied where $u(x,t) = u^c(x)$; where $u(x,t) > u^c(x)$ then from (2.1) $u_t - \Delta u - f = 0$, so it is always true.

Concerning the second inequality of (2.6), we have already seen that it is satisfied (with the equal sign) when $u > u^c$. But when $u(x,t) = u^c(x)$, (2.1) and assumption $\mathbf{H}_2$ imply that

$$u_t - \Delta u - f = -\Delta u^c - f \geq 0.$$

Then u coincides with the solution of (2.6); this also proves its uniqueness. □

Remark 2.1.1. *We point out that the hypothesis* $\mathbf{H}_1$ *is crucial in order to guarantee the equivalence of problem* (2.1) *and* (2.4)*. In fact, if* $u^0 = u^c$ *in some region* $D \subset \Omega$*, then* $u(t) = u^c$ *in* D *for any* $t > 0$*, differently from what would happen for the solution* $w(t)$ *of* (2.4)*. Then the entire evolution of the solution and hence the asymptotic solution of the problem change: see the next Subsection discussion and an example (see* Test 3*) in Section* 2.2*, page* 61*.*

Remark 2.1.2. *Assumption* $\mathbf{H}_2$ *is a natural condition for the contact region* $C(t) = \{x \in \Omega : u(x,t) = u^c(x)\}$ *(the place where it is used inside the proof), at least when the obstacle is sufficiently smooth. If not satisfied by the data,* $C(t)$ *remains empty for any time, and the two problems* (2.1) *and* (2.4) *are trivially equivalent. On the other side, it is possible to verify (for example, in the case with no source term,* $f = 0$*) that the contact is possible from above only at regions of* Ω *where the obstacle is superharmonic (*$-\Delta u^c \geq 0$*); in order for the solution to reach regions of the obstacle where it is subharmonic one needs a sufficiently negative source term* f *to balance the positivity of* Δu^c*. Then the assumption of* $\mathbf{H}_2$ *in all of* Ω *is too restrictive, but we left it here in this form, since the contact region itself is an unknown of the problem. We will show an example (see* Test 4*) in Section* 2.2*, page* 61*.*

2.1.2 Asymptotic solution of the problem

Aim of this Subsection is to study the asymptotic behavior in time of the solution of problem (2.1). Using the result of Proposition 2.1.1, we will deduce that it evolves towards the unique solution $\overline{u}$ of the corresponding stationary (i.e. elliptic) obstacle problem

$$\overline{u} \in H_0^1(\Omega), \quad \overline{u} \geq u^c, \quad -\Delta\overline{u} \geq f, \quad (\overline{u} - u^c)(\Delta\overline{u} + f) = 0 \text{ a.e. in } \Omega. \tag{2.8}$$

Remark 2.1.3. *We remark that the stationary problem corresponding to* (2.1)*, that is*

$$\begin{cases} H(\tilde{u} - u^c)(\Delta\tilde{u} + f) = 0 & \text{in } \Omega \\ \tilde{u} = 0 & \text{on } \partial\Omega \end{cases} \tag{2.9}$$

is not well posed, since uniqueness fails. For example, in one dimension, if $f = 0$, any function in $H^2(\Omega)$ which coincides with the obstacle in a subset of Ω and reaches zero on the boundary in a linear way outside of that is clearly a solution of $H(\tilde{u} - u^c)\Delta\tilde{u} = 0$ and a possible asymptotic solution for problem (2.1)*.*

Let us consider, for example, $\Omega = [-1, 1]$, $f = 0$ and the obstacle given by

$$u^c(x) = \frac{1}{2} - \left(2x^2 - \frac{1}{2}\right)^2 \tag{2.10}$$

then the following functions $u_1(x)$ and $u_2(x)$ are both solutions of problem (2.9)*, but only the first is solution of problem* (2.8)*:*

$$u_1(x) = \begin{cases} a\,(1+x) & \text{in } -1 \leq x < -b \\ u^c(x) & \text{in } -b < x \leq -0.5 \\ 0.5 & \text{in } -0.5 < x < 0.5 \\ u^c(x) & \text{in } 0.5 \leq x < b \\ a\,(1-x) & \text{in } b \leq x \leq 1 \end{cases}, \quad u_2(x) = \begin{cases} a\,(1+x) & \text{in } -1 \leq x \leq -b \\ u^c(x) & \text{in } -b < x < b \\ a\,(1-x) & \text{in } b \leq x \leq 1 \end{cases} \tag{2.11}$$

with $a, b \in \mathbb{R}$ such that

$$a(1-b) = u^c(b) = u^c(-b), \quad a = (u^c)'(-b) = -(u^c)'(b).$$

(see Fig. 2.1*). Note that $u^c(x), u_1(x), u_2(x) \in C^1\left(\overline{\Omega}\right) \cap H^2\left(\Omega\right)$. If we use u_2 as initial datum for both problems* (2.1) *and* (2.8) *(then with $\mathbf{H}_1$ violated), we see that $w(t)$ evolves in time towards u_1, while $u(t)$ remains equal to u_2 for any time.*

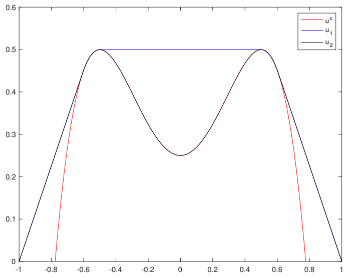


Figure 2.1: Graph of the functions $u^c(x)$, see 2.10, and $u_1(x), u_2(x)$.

Theorem 2.1.1. *Assume conditions* (2.3)*,* $\mathbf{H}_1$ *and* $\mathbf{H}_2$*. Let* $u(t)$ *be the global (in time) solution to the degenerate parabolic problem* (2.1) *and let* $\overline{u}$ *be the unique solution of the obstacle problem* (2.8)*. Then,* $u(t)$ *converges to* $\overline{u}$ *strongly in* $H_0^1(\Omega)$ *for* $t \to \infty$*, and there is a constant* $C > 0$ *such that, for every* $t \geq 1$*,*

$$\|u(t) - \overline{u}\|_{H^1(\Omega)} \leq e^{-Ct} \ . \tag{2.12}$$

Proof. Let us make a change of variables: if we set $v(x,t) = u(x,t) - u^c(x)$, then it is easy to see that v solves, for all $t \in (0,T)$, the problem

$$\begin{cases} v_t - H\left(v\right)\left(\Delta v + F\right) = 0 & \text{a.e. in } \Omega \\ v\left(0\right) = v^0 & \text{in } \Omega \\ v = g & \text{on } \partial\Omega \end{cases} \tag{2.13}$$

with $F = \Delta u^c + f$, $v^0 = u^0 - u^c > 0$ and $g = -u^c \geq 0$ on $\partial\Omega$. For Proposition 2.1.1 we know that v also solves the variational inequality

$$\begin{cases} v(t) \in \mathcal{K}_g, \quad \displaystyle\int_\Omega (v_t - \Delta v - F)(\varphi - v)\, dx \geq 0 \quad \forall \varphi \in \mathcal{K}_g, \forall t \in (0,T) \\ v\left(0\right) = v^0 \in \mathcal{K}_g \end{cases} \tag{2.14}$$

where

$$\mathcal{K}_g = \left\{\varphi \in H^1\left(\Omega\right), \varphi - g \in H_0^1\left(\Omega\right), \varphi \geq 0 \text{ in } \Omega\right\}. \tag{2.15}$$

The corresponding elliptic obstacle problem is then

$$\bar{v} \in \mathcal{K}_g, \quad \int_\Omega (-\Delta \bar{v} - F)(\varphi - \bar{v})\, dx \geq 0 \quad \text{for all } \varphi \in \mathcal{K}_g \ , \tag{2.16}$$

and $\bar{v}$ minimizes in $\mathcal{K}_g$ the following functional

$$\mathcal{F}(\varphi) = \frac{1}{2}\int_\Omega |\nabla\varphi|^2 \, \mathrm{d}x - \int_\Omega F\varphi \, \mathrm{d}x.$$

We note that $-\nabla\mathcal{F}(\varphi) = \Delta\varphi + F$.

The proof follows the idea of Theorem 1.7 of [26]. The main difference is that in [26] it is considered $F = -1$, while here we will consider a generic datum F with $F \leq 0$ by $\mathbf{H}_2$.

First of all we prove the following constrained Lojasiewicz inequality for the obstacle problem (see Proposition 4.1 of [26]): there is a positive constant $C_d > 0$ such that

$$\left(\mathcal{F}\left(v\right) - \mathcal{F}\left(\overline{v}\right)\right)_+^{\frac{1}{2}} \leq C_d \left\|\nabla\mathcal{F}\left(v\right)\right\|_{\mathcal{K}_g} \tag{2.17}$$

for every $v \in H^2(\Omega) \cap \mathcal{K}_g$ where

$$\left\|\nabla\mathcal{F}\left(v\right)\right\|_{\mathcal{K}_g} := \sup\left\{0, \sup_{\varphi\in\mathcal{K}_g\setminus\{v\}} \frac{-\displaystyle\int_\Omega \left(\varphi - v\right)\nabla\mathcal{F}\left(v\right) \, \mathrm{d}x}{\left\|\varphi - v\right\|_{L^2}}\right\}. \tag{2.18}$$

We point out that the unique solution $\overline{v}$ of the obstacle problem (2.16) solves $-\Delta\overline{v} = F\chi_{\{\overline{v}>0\}}$ in Ω and $\overline{v} = g$ on $\partial\Omega$. Then, taking $\varphi = \overline{v}$,

$$\begin{aligned}
\left\|\nabla\mathcal{F}\left(v\right)\right\|_{\mathcal{K}} &\geq -\frac{\displaystyle\int_\Omega \left(\overline{v} - v\right)\nabla\mathcal{F}\left(v\right) \, \mathrm{d}x}{\left\|\overline{v} - v\right\|_{L^2}} = \\
&= -\frac{1}{\left\|\overline{v} - v\right\|_{L^2}} \int_\Omega \left(v - \overline{v}\right)\left(\Delta v + F\right)\mathrm{d}x = \\
&= -\frac{1}{\left\|\overline{v} - v\right\|_{L^2}} \int_\Omega \left(v - \overline{v}\right)\left(\Delta v - \Delta\overline{v}\right)\mathrm{d}x + \\
&\quad -\frac{1}{\left\|\overline{v} - v\right\|_{L^2}} \int\limits_{\Omega\cap\{\overline{v}=0\}} \left(v - \overline{v}\right)F\mathrm{d}x = \\
&= \frac{1}{\left\|\overline{v} - v\right\|_{L^2}} \left(\frac{1}{2}\int_\Omega \left|\nabla\left(v - \overline{v}\right)\right|^2 \mathrm{d}x - \int\limits_{\Omega\cap\{\overline{v}=0\}} F\left(v - \overline{v}\right)\mathrm{d}x\right).
\end{aligned} \tag{2.19}$$

As

$$\begin{aligned}\int_\Omega |\nabla (v-\overline{v})|^2 \,\mathrm{d}x &= \int_\Omega |\nabla v|^2 \ \mathrm{d}x + \int_\Omega |\nabla \overline{v}|^2 \,\mathrm{d}x - 2\int_\Omega \nabla v \nabla \overline{v} \mathrm{d}x = \\ &= \int_\Omega |\nabla v|^2 \,\mathrm{d}x + 2\int_\Omega |\nabla \overline{v}|^2 \,\mathrm{d}x - 2\int_\Omega \nabla v \nabla \overline{v} \ \mathrm{d}x - \int_\Omega |\nabla \overline{v}|^2 \,\mathrm{d}x = \\ &= \int_\Omega |\nabla v|^2 \,\mathrm{d}x + 2\int_\Omega \nabla \overline{v} \cdot (\nabla \overline{v} \ - \nabla v) \,\mathrm{d}x - \int_\Omega |\nabla \overline{v}|^2 \,\mathrm{d}x = \\ &= \int_\Omega |\nabla v|^2 \,\mathrm{d}x - \int_\Omega |\nabla \overline{v}|^2 \,\mathrm{d}x - 2\int_\Omega \Delta \overline{v} \,(\overline{v}-v) \,\mathrm{d}x = \\ &= \int_\Omega |\nabla v|^2 \,\mathrm{d}x - \int_\Omega |\nabla \overline{v}|^2 \,\mathrm{d}x - 2 \int\limits_{\Omega\cap\{\overline{v}>0\}} F\,(v-\overline{v}) \,\mathrm{d}x \end{aligned} \tag{2.20}$$

estimate (2.19) becomes

$$\begin{aligned}\|\nabla \mathcal{F}(v)\|_{\mathcal{K}_g} &\geq \frac{1}{\|\overline{v}-v\|_{L^2}} \left(\frac{1}{2}\int\limits_\Omega |\nabla v|^2 \,\mathrm{d}x - \frac{1}{2}\int\limits_\Omega |\nabla \overline{v}|^2 \,\mathrm{d}x + \right. \\ &\left. - \int\limits_{\Omega\cap\{\overline{v}>0\}} F\,(v-\overline{v}) \,\mathrm{d}x - \int\limits_{\Omega\cap\{\overline{v}=0\}} F\,(v-\overline{v}) \,\mathrm{d}x \right) = \\ &= \frac{1}{\|\overline{v}-v\|_{L^2}} \left(\mathcal{F}(v) - \mathcal{F}(\overline{v})\right) \ . \end{aligned} \tag{2.21}$$

By Poincaré inequality and by $\mathbf{H}_2$, we obtain

$$\begin{aligned}\|v-\overline{v}\|^2_{L^2(\Omega)} &\leq C_p \|\nabla (v-\overline{v})\|^2_{L^2(\Omega)} = \\ &= C_p \left(\int\limits_\Omega |\nabla v|^2 \,\mathrm{d}x - 2 \int\limits_{\Omega\cap\{\overline{v}>0\}} vF \mathrm{d}x - \int\limits_\Omega |\nabla \overline{v}|^2 \,\mathrm{d}x + 2 \int\limits_{\Omega\cap\{\overline{v}>0\}} \overline{v} F \mathrm{d}x \right) \\ &\leq 2C_p \left(\mathcal{F}(v) - \mathcal{F}(\overline{v})\right)\end{aligned} \tag{2.22}$$

and so

$$\|\nabla \mathcal{F}(v)\|_{\mathcal{K}_g} \geq \frac{1}{\|v-\overline{v}\|_{L^2}} \left(\mathcal{F}\left(v\right)-\mathcal{F}\left(\overline{v}\right)\right) \geq \frac{1}{\sqrt{2C_p}} \frac{\mathcal{F}\left(v\right)-\mathcal{F}\left(\overline{v}\right)}{\left(\mathcal{F}\left(v\right)-\mathcal{F}\left(\overline{v}\right)\right)^{\frac{1}{2}}} \ , \tag{2.23}$$

that is the following constrained Lojasiewicz inequality holds

$$\left(\mathcal{F}\left(v\right)-\mathcal{F}\left(\overline{v}\right)\right)^{\frac{1}{2}} \leq \sqrt{2C_p}\, \|\nabla \mathcal{F}(v)\|_{\mathcal{K}_g} . \tag{2.24}$$

Then, by using Proposition 2.10 in [26], we conclude the proof, since

$$\|u-\overline{u}\|_{H^1(\Omega)} = \|v-\overline{v}\|_{H^1(\Omega)} \leq e^{-Ct} \ . \tag{2.25}$$

□

Remark 2.1.4. *Let us consider the following quantities:*

$$M(t) = \int_\Omega (u(t)-u^c)\mathrm{d}x, \quad I(t) = \int_\Omega H(u(t)-u^c)(\Delta u(t)+f)\mathrm{d}x.$$

The first one measures the global distance in time of the solution from the obstacle. Theorem 2.1.1 *and stationary problem* (2.9)*, as* $t \to \infty$*, imply that*

$$M(t) \to \overline{M} = \int_\Omega (\bar{u}-u^c)\mathrm{d}x, \qquad I(t) \to 0.$$

If we integrate the equation in (2.1) *we get*

$$\frac{d}{dt}M(t) = \int_\Omega u_t(t)\mathrm{d}x = \int_\Omega H(u(t)-u^c)(\Delta u(t)+f)\mathrm{d}x = I(t). \tag{2.26}$$

When $I(t)$ *does not change its sign in time, then the convergence of* $M(t)$ *is monotone. To look at the time profiles of* $M(t)$ *and* $I(t)$ *is interesting, since it gives some informations on the global evolution of the solution (for example, it reveals the contact times with the obstacle). We will look at their corresponding discrete quantities in the numerical simulations of the last Section.*

2.2 Numerical analysis

In the following, it will be given the numerical analysis of the problem (2.1).

2.2.1 Numerical approximation

For simplicity, let us start with the one dimensional setting, and $\Omega = (-1,1)$. Then the problem to solve becomes:

$$\begin{cases} u_t - H(u-u^c)(u_{xx}+f) = 0 & \text{in } \Omega \times (0,T) \\ u(0) = u^0 & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \times (0,T), \end{cases} \tag{2.27}$$

where T is a sufficiently large time, $u^0 > u^c$ and H denotes the Heaviside function defined in (2.2), or eventually a regular approximation of it in a small right neighborhood of the origin, for example the C^1 function η_n given by:

$$\eta_n(r) = \begin{cases} 1 & \text{if } r > \dfrac{1}{n} \\ -2n^3r^3 + 3n^2r^2 & \text{if } 0 \le r \le \dfrac{1}{n} \\ 0 & \text{if } r < 0 \end{cases} . \tag{2.28}$$

According to the fixed integer parameter $n \in \mathbb{N}$, we see that $\eta_n(r) = H(r)$ for $r \geq \frac{1}{n}$ and $r \leq 0$, and that

$$\|H - \eta_n\|_{L^1(-1,1)} = \int_{-1}^{1} |H(r) - \eta_n(r)|\, dr = \int_{0}^{1/n} |H(r) - \eta_n(r)|\, dr \to 0 \quad as\ n \to \infty. \tag{2.29}$$

Note that, as happens for H, $\eta_n(0) = 0$, so that even in this case the diffusion coefficient vanishes at the contact points between the solution of problem (2.1) (with H replaced by η_n) and the obstacle. But now also all the values of η_n in $(0,1]$ characterize supercritical states of the solution close to the obstacle itself.

On Ω we define for a given $N \in \mathbb{N}$ a uniform grid G_h of size $h = 2/N$. Then G_h will have $N-1$ internal nodes $x_j = -1 + jh$ $(j = 1,..,N-1)$ over a total number of $(N+1)$.

Concerning time discretization, we adopted a uniform time step $\Delta t = T/M$, for a given $M \in \mathbb{N}$, so that the solution is computed at any time $t_k = k\Delta t$ $(k = 1,..,M)$: by

u_j^k we denote the discrete solution at time t_k in a node x_j of G_h. The initial data will be given by

$$u_j^0 = u^0(x_j),\ u_j^c = u^c(x_j),\ \text{with}\ \ u_j^0 > u_j^c \text{ for any } j = 1,..,N-1,\ u_0^0 = u_N^0 = 0. \tag{2.30}$$

We are interested in the numerical solution of (2.27) on the grid G_h; to avoid stability problems without heavy restrictions on the parabolic step ratio $\gamma = \Delta t/h^2$ we adopted a semi implicit finite difference scheme:

(S): $for\ any\ k = 0, 1, ..., M-1\ solve\ for\ any\ internal\ node\ x_j$:

$$\begin{cases} u_j^{k+1} = u_j^k + \Delta t\ z_j^k\ (\delta_h u_j^{k+1} + f_j) := u_j^k + \gamma\ z_j^k\ \left(u_{j-1}^{k+1} - 2u_j^{k+1} + u_{j+1}^{k+1}\right) + \Delta t\ z_j^k\ f_j\ , \\ \text{if } u_j^{k+1} < u_j^c \Rightarrow u_j^{k+1} = u_j^c\ , \\ u_0^{k+1} = u_N^{k+1} = 0 \quad \text{(boundary values);} \end{cases} \tag{2.31}$$

with δ_h we have indicated the usual 3-point second order finite difference operator over G_h; we will talk of scheme (S_H) when $z_j^k = H_j^k := H\left(u_j^k - u_j^c\right)$ (that is when we use the sharp Heaviside values), of scheme (S_η) when $z_j^k = \eta_j^k := \eta\left(u_j^k - u_j^c\right)$ (that is when we use its approximated values given by (2.28)). Then at any time iteration k one has to solve the following linear system:

$$B^k u^{k+1} := (I + \gamma z^k * A)u^{k+1} = u^k + \Delta t z^k F$$

where u^k, z^k and $F = (f_j)$ are column vectors of dimension $(N-1)$, A is the tridiagonal $(N-1) \times (N-1)$ matrix associated to the discrete Laplacian in one dimension, and by $v * M$ we mean the vector matrix product in which each line j of M is multiplied for the $j-$th component of v.

Remark 2.2.1. *Let us explain the second line of scheme* (2.31)*. We proved in the previous Section that the solution of* (2.1) *always remains over the obstacle. In the discrete settings with scheme* (2.31) *anyway, the impact with the obstacle happens at a certain time iteration, with a thrust which depends on the parameter* γ *and which can cause the overcoming of the obstacle before the Heaviside term can stop the diffusion. So it is necessary to force the discrete solution to coincide with the obstacle where it has gone over. When* γ *is large, anyway, the solution can overcome the obstacle in many adjacent nodes* *at a single instant time, yielding an overestimation of the contact set which the subsequent iterations are no more able to correct. That is why, even if the scheme has no stability constraints, a reduced value of γ (hence of Δt) should be necessary in order to evolve towards the correct stationary solution, with a consequent grow of computational costs.*

To face such a problem we have experimented some variants of our approach. The first one consists in the use of the approximate Heaviside function η_n of (2.28) to determine the diffusion coefficient: when the solution gets closed to the obstacle, it has the effect to reduce progressively the thrust and even to prevent the overcome of the obstacle (if a suitable value of n is chosen).

Another idea is to use a variable discretization time step, reducing it only when it is necessary. We tested two ways for that. The first one measures the impact thrust of each Δt in terms of the number of nodes involved in the contact at a single iteration time, halving it until this number remains large but resetting it at the initial value when the contact with the obstacle becomes sufficiently stable. It works well, but this "trial and error" process is still too expensive. The second way comes directly from the scheme. Assume for simplicity $f = 0$; if $u_j^k \geq u_j^c$ for any j, in order to remain over the obstacle everywhere at the $k+1$ iteration we should have

$$u_j^{k+1} = u_j^k + \Delta t z_j^k \delta_h u_j^{k+1} \geq u_j^c, \quad \forall j \ ;$$

where there is already a contact ($z_j^k = 0$) there is nothing to prove; elsewhere $z_j^k = 1$ and if the solution decreases at a node x_j then necessarily $\delta_h u_j^{k+1} < 0$, so that the previous inequality is equivalent to ask

$$\Delta t \leq D_j := \frac{u_j^k - u_j^c}{-\delta_h u_j^{k+1}} \ ; \tag{2.32}$$

then the estimate of the smallest positive value of D_j (with $\delta_h u_j^{k+1}$ replaced by $\delta_h u_j^k$) gives at any iteration a sufficiently small time step in order to reach the obstacle with the right thrust. We have tested all these approaches in the experiments of the next Section, trying a comparison evaluation.

In order to emphasize the convergence of the solutions to the stationary state, as discussed in the previous Section, we adopted for scheme (2.31) the following stopping criterium:

$$\max_j \left[(u_j^k - u_j^c) |\delta_h u_j^k + f_j| \right] < tol \tag{2.33}$$

where *tol* indicates a prescribed small tolerance. In other words the scheme stops before the final time T if the limit problem is sufficiently solved.

For sake of comparison, we have also implemented a numerical scheme for the corresponding parabolic and elliptic obstacle problems (respectively (2.6) and (2.8)), with the same discretization parameters, showing even at a discrete level the essential coincidence of the solutions of the two evolutive problems (if γ is not too large) and their convergence to the same asymptotic solution. Many algorithms can be found in the literature for the obstacle problem: among them we have choosen the ones presented in [20], based on the iterative solutions of piecewise linear systems. The discrete version of the equation in (2.6) becomes

$$(w^{k+1} - u^c)^T (w^{k+1} + \gamma A w^{k+1} - w^k - \Delta t f) = 0. \tag{2.34}$$

Setting $y = w^{k+1} - u^c \geq 0$, then y has to solve

$$y^T (y + \gamma A y - b) = 0$$

with $b = w^k - u^c + \Delta t f - \gamma A u^c$. In [20] it is proved that $y = \max(x, 0)$ is a solution of the previous equation if x solves

$$[I + \gamma A P(x)] x = b \tag{2.35}$$

where $P(x)$ is the diagonal matrix with $p_{jj} = H(x_j)$, and H is the Heaviside (sign) function (2.2). In order to solve the last implicit equation a quasi-Newton method is implemented which needs a certain number of linear system solutions (Picard iterations) for any discrete time step:

$$P^0 = O \text{ (null matrix)}, \quad (I + \gamma A P^n) x^{n+1} = b, \quad for\ n = 0, 1, ... \text{ until } \ P^n = P^{n+1} \ ;$$

then $x = x^{n+1}$ is the solution of (2.35); hence $w^k = y + u^c$ solves (2.34) and evolves in time towards the solution $\bar{u}$ of the corresponding stationary obstacle problem (2.8) on the grid G_h.

The extension of scheme (2.31) to the two-dimensional case is straightforward, at least when Ω is a rectangular open set $(a, b) \times (c, d)$. Using equal space steps $\Delta x = \Delta y = h$, the discrete solution u_{ij}^k will denote the approximated value of u in the node x_{ij} at time $t^k = k\Delta t$. It is then sufficient to replace the finite difference operator δ_h with the usual five-point Laplacian approximation scheme:

$$\delta_h^2 u_{ij}^k = \frac{u_{i+1j}^k + u_{i-1j}^k - 4u_{ij}^k + u_{ij+1}^k + u_{ij-1}^k}{h^2}.$$

All the previous considerations remain unchanged.

2.2.2 Numerical tests

We have tested scheme (2.31) with the stopping criterium (2.33), for $tol = 10^{-4}$ and different initial data and obstacles in one dimension on $\Omega = (-1, 1)$ and in two dimensions on square regions. Here we discuss the results of these experiments.

Test 1. $u^0 = 0.7 - 0.7x^2$, $u^c = 0.5 - 2x^2$ (inverted parabola, with negative values at $\partial\Omega$); when $f = 0$, the solution decreases in time until it touches the obstacle from the top; the two lateral branches (in the detachment region) then rapidly become linear, that is harmonic, until nothing changes anymore (Fig.2.2 a). The discrete contact region (with $N = 101$ nodes) is the set $C = [-0.14, 0.14]$. In Fig.2.2 b) the plots are reported of the discrete quantities corresponding to $M(t)$ and $I(t)$, which in this case are monotone in time. The impact time of the solution with the obstacle is highlighted by the change of slope in the second plot. The addition of a constant negative source term ($f = -1.5$) correctly increases the contact set (now $C = [-0.26, 0.26]$) and reduces the final stopping time (Fig.2.2 c).

On this example we tried a comparison, in terms of precision and computational costs, of the different approaches introduced in the previous Section. In Table 2.1 the first column indicates the type of Heaviside function (H=exact, η_n=approximated), the second one if a fixed (F) or variable (V) time step approach (the one based on the time step estimate (2.32)) is adopted during the evolution; T^* denotes the exit time reached applying criterium (2.33), C_{bound} the right extremum of the detected symmetric contact set with the obstacle (which as we said should be in this case 0.14), $\|u-w\|_\infty$ the maximum norm of the difference in time between the discrete solutions of schemes (2.31) and (2.34), that is:

$$\|u - w\|_\infty = \max_k \|u^k - w^k\|_\infty.$$

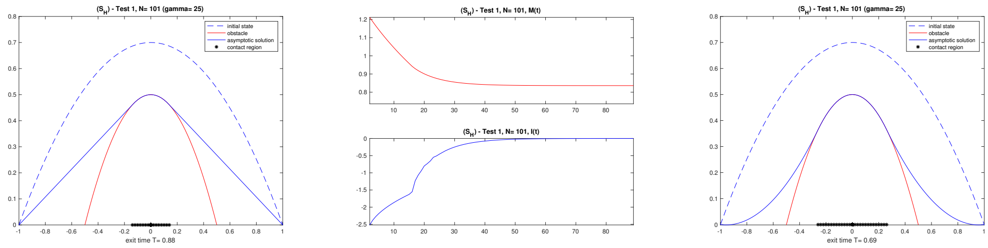


Figure 2.2: Test 1. a) $f = 0$, b) discrete $M(t)$ and $I(t)$ evolution; c) $f = -1.5$.

Table 2.1 values show with a certain evidence some aspects of the different approaches:

- if γ is too high, the contact set can be overestimated, and the asymptotic solution is incorrect (see also Fig.2.3);

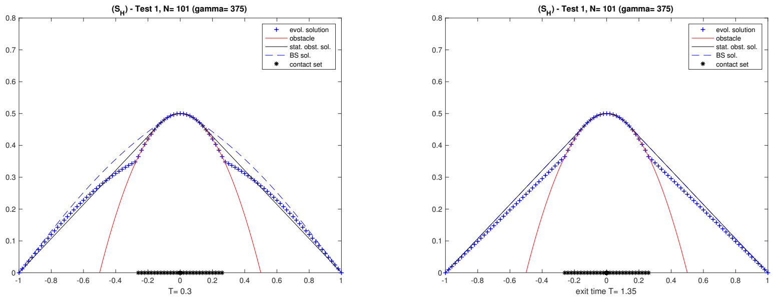


Figure 2.3: Test 1. Overestimation of the contact set for large γ: a) first time impact; b) final uncorrect solution.

- in order to have a good coincidence between u and w, a low value of γ is necessary, that is a little Δt and many time iterations;

- the use of an approximated Heaviside function with a sufficiently low parameter n helps a little, since the right contact set can be found, and a better coincidence between the two solutions in time. But the evolution is slowed down in an artificial way, and the contact is less sharp;

- a better performance comes from the variable step approach, where, without a significative change in the exit time, the correct solution and contact set are recovered. A higher number of time iterations is needed, but much less than the one needed (with a consistent reduction of γ) in order to get the same precision.

- the table also allows a cost comparison between our semi implicit approach to the obstacle problem and the implicit one of (2.34): while in the first one for any time step a single linear system has to be solved, in the second one a certain number of linear system solutions is needed. For example, with $\gamma = 75$ at the end the total number of these resolutions is of the order of 400, much more than the total time iterations of scheme (2.31), even in its variable time step version. Consequently, our approach to the numerical resolution of problem (2.1) can be considered as a competitive algorithm for the approximation of the parabolic variational inequality (2.4).

Table 2.1: Test 1. Performance comparison of scheme (S) with exact (H) or approximated (η_n) Heaviside function, fixed (F) or variable (V) time step.

Heav	time step	γ	T^*	time iter.	C_{bound}	$\|u-w\|_\infty$
H	F	375	1.35	10	**0.26**	$6\ 10^{-2}$
H	V	375	1.35	28	0.14	$1.2\ 10^{-3}$
H	F	187.5	1.05	15	**0.2**	$3.4\ 10^{-2}$
η_{20}	F	187.5	1.275	18	0.14	$1.25\ 10^{-2}$
H	V	187.5	1.12	34	0.14	$6\ 10^{-4}$
H	F	150	1.08	19	0.14	$1.4\ 10^{-2}$
H	F	75	0.96	33	0.14	$1.4\ 10^{-2}$
η_{50}	F	75	1.56	53	0.14	$4.1\ 10^{-3}$
H	V	75	0.96	50	0.14	$2.3\ 10^{-4}$
H	F	37.5	0.9	61	0.14	$1.8\ 10^{-4}$
H	F	18.75	0.86	116	0.14	$4.4\ 10^{-4}$
H	F	9.37	0.84	226	0.14	$6.6\ 10^{-4}$

Test 2. $u^0 = \frac{1}{(1+10x^2)} - \frac{1}{11}$ (partially convex initial state), same obstacle and source term of Test 1; we get the same stationary solution of Test 1, but a different evolution (Fig.2.4). Note that now the solution initially grows in regions where it is convex and decreases where it is concave: despite of that, the total mass $M(t)$ decreases for any time. On the contrary, the quantity $I(t)$ decreases during the first part of evolution, before increasing towards zero, remaining all the time negative.

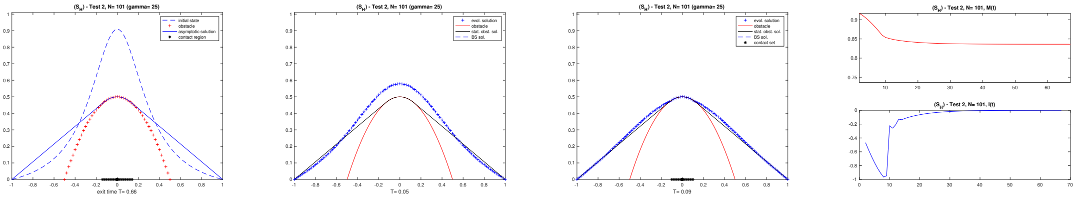


Figure 2.4: Test 2. a) initial datum and final solution, b) t=0.05, c) t=0.09, d) discrete $M(t)$ and $I(t)$ evolution.

Test 3. $u^0 = (1 - x^2)(1 + x^2)^3$, $u^c = 1 - 2x^2$ (initial contact point with the obstacle at the origin), $f = 0$; this example shows that the assumption $u^0 > u^c$ is essential in order to have the same evolution of the corresponding parabolic obstacle problem (see Remark 2.1.1). Here the asymptotic solution is the same for the two problems, and even the final contact set is the same ($C = [-0.3, 0.3]$), but the evolution is completely different: in the contact point the solution of (2.31) (++) cannot detach anymore from the obstacle, differently to what happens to the other one (dotted), see Fig.2.5.

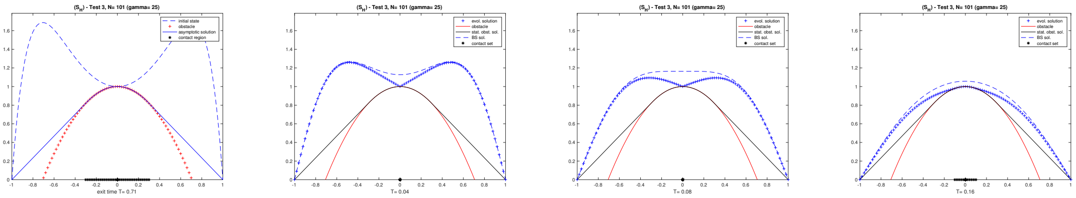


Figure 2.5: Test 3. a) initial datum and final solution, b) t=0.04, c) t=0.08, d) t=0.16.

Test 4. $u^0 = 1 - x^2$, $u^c = 0.5 - (2x^2 - 0.5)^2$ (two equal hills with a valley in the middle): it is the example of Remark 2.1.3. When $f = 0$ the solution leans on the hills and remains stretched over the valley (Fig.2.6 a). The final contact region is now given by $C = (-b, -0.5) \cup (0.5, b)$, with $b \simeq 0.6054$. Note that assumption $\mathbf{H}_2$ in this case is not satisfied in a small neighborhood of the origin which does not belong to the contact set. Even in this case $M(t)$ and $I(t)$ are monotone (Fig.2.6 b). In order to push the solution in contact with the whole convex region of the obstacle a sufficiently negative source term has to be added: in this case $f = -4$ is necessary to make $\Delta u^c + f \leq 0$, so that $\mathbf{H}_2$ holds in all Ω, and in particular in the whole connected contact region $C = (-0.66, 0.66)$ (Fig.2.6 c).

In Fig.2.6 d-e we show what happens if we start with a different initial datum very close to the obstacle:

$$u^0 = \max(0, 0.5 - (2x^2 - 0.5)^2 + 0.1);$$

the solution converges towards the same asymptotic solution, but now essentially from below; then $I(t)$ tends to zero from positive values and $M(t)$ is monotone increasing.

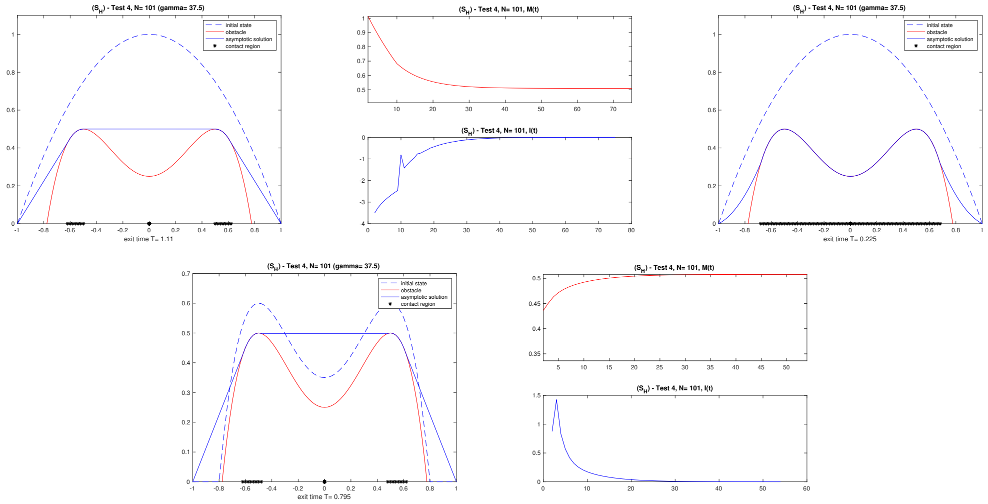


Figure 2.6: Test 4. a) $f = 0$, initial datum and final solution, b) $M(t)$ and $I(t)$ evolution, c) $f = -4$, d-e) $f = 0$ but initial datum close to the obstacle.

In the next two examples we considered less regular obstacles, not differentiable or even discontinuous. The experiments show that model (2.1) still works also in these cases and that the scheme (2.31) behaves correctly.

Test 5. $u^0 = 1.6 - 1.6x^2$, $u^c = \max(1 - 3|x|, 0.5 - 4|x + 0.7|, 0.4 - 8|x - 0.8|)$ (three peaks), $f = 3x$; the contact set consists of three distinct points (Fig.2.7 a).

Test 6. $u^0 = 2 - 2x^2$, $u^c = x + 0.5$ for $x < 0$, $u^c = 1 - x$ for $x \geq 0$, $f = 0$; $C = [0, 1]$ (Fig.2.7 b).

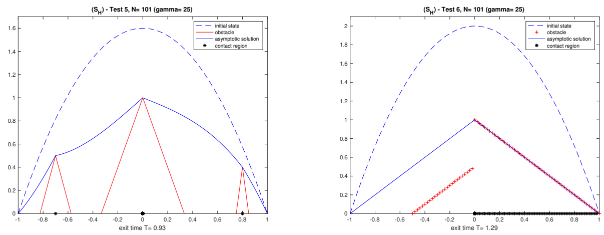


Figure 2.7: a) Test 5, b) Test 6.

Finally we report the results of some 2D tests in the square region $\Omega = (-1,1)^2$. For any example we show the final situation, with the surface contact evidence, and explicitely (in blue) the contact area, that is the nodes of the mesh where the solution touches the obstacle.

Test 7. $u^0 = 2(1-x^2)(1-y^2)$; $u^c = 1 - 2(x^2+y^2)$ (a reversed paraboloid); $f = -1$. The contact set is a disk (Fig.2.8 a).

Test 8. $u^0 = 4(1-x^2)(1-y^2)$; $u^c = 1 - (3.5(x^2+y^2) - 2)^2$ (a sort of crater of a volcano); $f = 0$. The contact set is a circular crown (Fig.2.8 b).

Test 9. $u^0 = 2(2 - |x+y| - |y-x|)$; $u^c = (2 - |x+y|) - |y-x|) - 1$ (a central pyramid); $f = 0$. The contact set is made by two crossing lines (Fig.2.8 c).

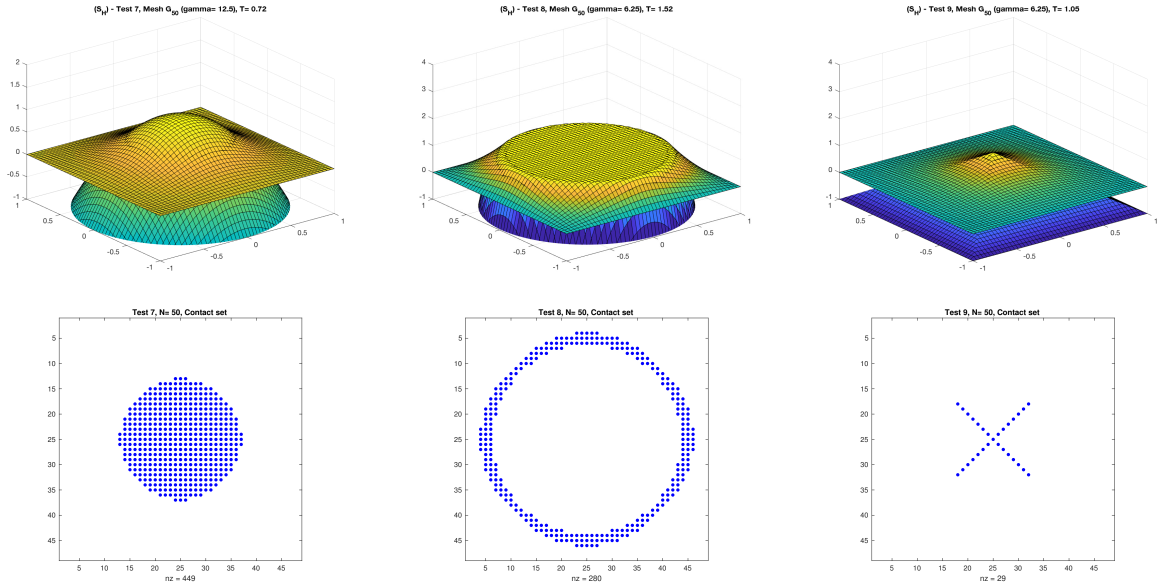


Figure 2.8: a) Test 7, b) Test 8, c) Test 9.

Test 10. $u^0 = (2-0.5x^2)(2-0.5y^2)$; $u^c = 1+x^2+2y^2-x^4-y^4$ (a sort of landscape with hills and valleys); we compare the final results for $f = 0$ and $f = -2$, respectively, with disconnected and connected contact sets (Fig.2.9).

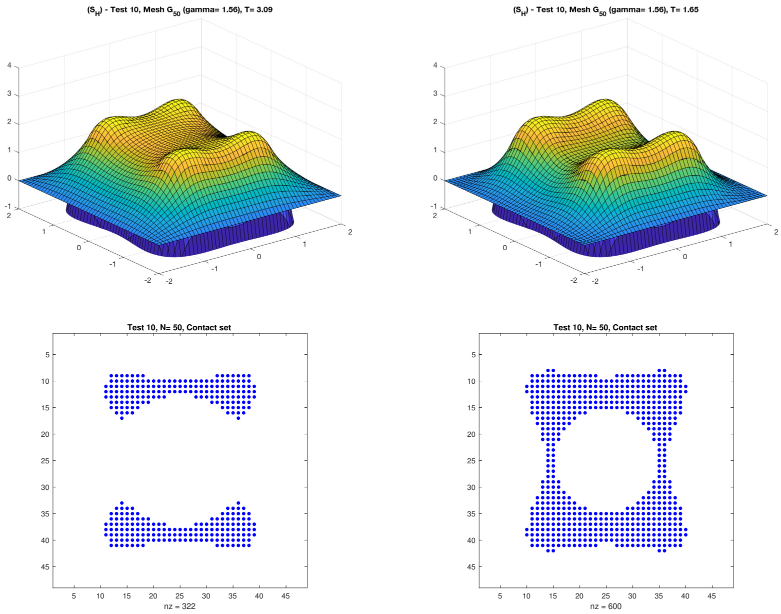


Figure 2.9: Test 10. a) $f = 0$, b) $f = -2$.

Chapter 3

An Heaviside function driven degenerate diffusion model with Caputo time fractional derivative

3.1 The model

Having established the importance of fractional calculus in recent years as a mathematical tool capable of describing the most disparate scientific phenomena, the purpose of this chapter is to study the following problem, which generalizes the one addressed in previous chapter, for $0 < \alpha < 1$:

$$\begin{cases} \partial_t^\alpha u - H\left(u - u^c\right) \Delta u = 0 & \text{a.e. in } \Omega, \text{ for all } t \in (0,T) \\ u\left(0\right) = u^0 & \text{a.e. in } \Omega \\ u = 0 & \text{on } \partial\Omega, \text{ for all } t \in (0,T) \end{cases} \tag{3.1}$$

with $T > 0$, where H is the extended Heaviside function such that $H(0) = 0$, that is

$$H(r) = \begin{cases} 1 & \text{for } r > 0 \\ 0 & \text{for } r \leq 0 \end{cases} \tag{3.2}$$

and Ω is a bounded domain in $\mathbb{R}^n$, $n \in \mathbb{N}$, with smooth boundary. For simplicity, we omitted the presence of a forcing term in the problem.

We will refer to [13, 21, 28, 29, 30, 31, 34, 35, 39, 41, 42, 43, 44, 45, 48, 50] for further details on fractional derivatives and about the results that we will point out in order to build ours. An overview about these topics, through the basic definitions and properties, will be also given in Appendix B, page 91. Moreover, here $\partial_t^\alpha u$ denotes the Caputo fractional derivative, that is,

$$\partial_t^\alpha u(x,t) := \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} \partial_s u(x,s) \mathrm{d}s, \tag{3.3}$$

where Γ is the Gamma function, (about definition and basic properties, see Appendix B).

The Riemann-Liouville derivative, (about definition and basic properties, see Appendix B), is defined as

$$^R\partial_t^\alpha u(x,t) := \frac{\partial}{\partial t} \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha}\, u(x,s) \mathrm{d}s. \tag{3.4}$$

We observe that

$$^R\partial_t^\alpha C = \frac{1}{\Gamma(1-\alpha)} \partial_t \int_0^t \frac{C}{(t-s)^\alpha} \mathrm{d}s = C \frac{t^{-\alpha}}{\Gamma(1-\alpha)}$$

and the following relation between derivatives holds

$$\partial_t^\alpha u(x,t) = {}^R\partial_t^\alpha \big(u(x,t) - u(x,0)\big) = {}^R\partial_t^\alpha u(x,t) - u(x,0) \frac{t^{-\alpha}}{\Gamma(1-\alpha)}. \tag{3.5}$$

For a more detailed discussion about the previous relation, see [28], Lemma 3.4 and Lemma 3.5, p. 53.

Let us consider the complementary system, for all $t > 0$,

$$\begin{cases} w(x,t) \geq u^c(x) & \text{a.e. in } \Omega \\ \partial_t^\alpha w \geq \Delta w & \text{a.e. in } \Omega \\ (w - u^c)(\partial_t^\alpha w - \Delta w) = 0 & \text{a.e. in } \Omega \\ w\,(x,0) = u^0 & \text{a.e. in } \Omega \\ w(x,t) = 0 & \text{on } \partial\Omega. \end{cases} \tag{3.6}$$

In Sections 3.1.1 and 3.1.2, we will prove that:

- by analogy with the classical case, (3.1) is equivalent to the complementarity system (3.6);

- asymptotically, the solution evolves for each α towards the same steady state of the classical problem, that is

  $$u \geq u^c, \quad -\Delta u \geq 0, \quad (u - u^c)\Delta u = 0 \quad \text{in } \Omega \times (0,T) \tag{3.7}$$

  with different convergence speed (for $\alpha = 1$ exponential, for $\alpha \in (0,1)$ polynomial).

In Section 3.2.1 we present three possible finite difference schemes for the numerical approximation of problem (3.1) or its equivalent form given by the complementary system (3.6). The Caputo derivative has been discretized using standard methods found in the literature, the so-called L1 or Convolution Quadrature (CQ) approaches (see e. g. [33] or [38]). The space discretization has been carried out through the semi implicit finite differences scheme introduced in [3] for problem (3.1), or through the implicit scheme of [20] for the evolutive obstacle problem (3.6). Note that for $\alpha \to 1^-$ all these schemes give back the known results of the classic heat equation with obstacle.

Our aim was not to find an optimal strategy of approximation for the problem, but only to derive working schemes in orders to confirm through explicit simulations the behavior of the solution as characterized by the results of Section 3.1.1. For that, in Section 3.2.2, we have tested the previous schemes on a couple of one-dimensional examples. Such simulations also allow to compare their reliability and computational cost. The semi implicit approach requires time step restrictions (strongly increasing as $\alpha \to 0^+$) in order to detect the correct contact set with the obstacle, restrictions unnecessary for the implicit approach. On the other hand, in the first case each time iteration is much less expensive, since it reduces to a single linear system solution. So, after all, all the proposed schemes appear to be competitive.

At last, for the following, we assume that the initial datum u^0 and the (independent of time) obstacle u^c satisfy the following conditions

$$u^0 \in H_0^1(\Omega), \, u^c \in H^2(\Omega), \; u^c \leq 0 \text{ on } \partial\Omega \tag{3.8}$$

and we define as solution of problem (3.1) a function $u \in L^2(0,T; H_0^1(\Omega) \cap H^2(\Omega))$ with $\partial_t^\alpha u(x,t) \in L^2(0,T; L^2(\Omega))$ which satisfies the conditions of the problem (3.1).

3.1.1 Equivalence with the obstacle problem

In the present Section we analyze problem (3.1), showing that, under suitable conditions, it is equivalent to the complementarity system (3.6). Let us introduce the following hypothesis:

$\mathbf{H}_1$: $u^0 > u^c$ a.e. in Ω;

$\mathbf{H}_2$: $\Delta u^c \leq 0$ a.e. in Ω.

Proposition 3.1.1. *Assume that conditions* (2.3)*,* $\mathbf{H}_1$ *and* $\mathbf{H}_2$ *hold. Then problems* (3.1) *and* (3.6) *are equivalent.*

Proof.

First we note that if w solves problem (3.6), then w solves (3.1). In fact, initial and boundary conditions are the same in (3.1) and (3.6). Since $(w - u^c)(\partial_t^\alpha w - \Delta w) = 0$ a.e. where $w > u^c$, then we obtain $\partial_t^\alpha w - \Delta w = 0$ a.e whereas, if $w(x,t) = u^c(x)$ the problem is no longer an evolutive one, then we have $\partial_t w = 0$ which implies $\partial_t^\alpha w = 0$.

Now we prove that if u solves problem (3.1) then u coincides with solution w of (3.6). Initial and boundary conditions are the same in (3.1) and (3.6). Let us now prove that if $u(x,t)$ is a solution of (3.1) then necessarily $u(\cdot,t) \geq u^c(\cdot)$ in Ω for any time t. Assume that $u < u^c$. Then $\partial_t^\alpha u = 0$ from (3.1). From $\mathbf{H}_1$, $u^0 > u^c$. Moreover, due to the fact that $u < u^c < u^0$, we obtain from (3.5)

$$\partial_t^\alpha u = {}^R\partial_t^\alpha (u - u^0) < 0$$

which does not agree with $\partial_t^\alpha u = 0$. Then $u \geq u^c$ by contradiction and the first inequality of (3.6) holds. The equation in the third line of (3.6) is trivially satisfied where $u(x,t) = u^c(x)$; where $u(x,t) > u^c(x)$, from (3.1) we obtain $\partial_t^\alpha u - \Delta u = 0$, so it is always true. Concerning the second inequality of (3.6), we have already seen that it is satisfied (with the equal sign) when $u > u^c$. But when $u(x,t) = u^c(x)$, (3.1) and assumption $\mathbf{H}_2$ imply that

$$\partial_t^\alpha u - \Delta u = -\Delta u^c \geq 0.$$

Then u solves problem (3.6). □

We recall that the same problem for $\alpha = 1$ has been faced in the second chapter of this thesis, proving in particular the equivalence of problem (3.1) with a parabolic obstacle problem.

Let v be the unique solution of problem (3.6) with $\alpha = 1$, that is

$$\begin{cases} v(x,t) \geq u^c(x) & \text{in } \Omega \\ \partial_t v \geq \Delta v & \text{a.e. in } \Omega \\ (v-u^c)(\partial_t v - \Delta v) = 0 & \text{a.e. in } \Omega \\ v\,(x,0) = u^0 & \text{in } \Omega \\ v(x,t) = 0. & \text{on } \partial\Omega \end{cases} \tag{3.9}$$

Let us introduce

$$u(x,t) = \int_0^\infty v(x,s)\,\ell(s,t)\,ds \tag{3.10}$$

where the probability density $\ell : [0,\infty) \times [0,\infty) \to [0,\infty)$ satisfies (in particular we are referring to [31])

$$\begin{cases} \partial_t^\alpha \ell = -\dfrac{\partial \ell}{\partial s}, & (s,t) \in (0,\infty) \times (0,\infty) \\ \ell(s,0) = \delta(s), & s > 0. \end{cases} \quad \text{and} \quad \begin{cases} {}^R\partial_t^\alpha \ell = -\dfrac{\partial \ell}{\partial s}, & (s,t) \in (0,\infty) \times (0,\infty) \\ \ell(s,0) = \delta(s), & s > 0 \\ \ell(0,t) = \dfrac{t^{-\alpha}}{\Gamma(1-\alpha)}, & t > 0. \end{cases}$$

In both cases, we also set that $\ell(s,t) = 0$ for all $s < 0$.

Remark 3.1.1. *As a probability density, we recall that*

$$\int_{-\infty}^{+\infty} \ell(s,t)\mathrm{d}s = \int_0^{+\infty} \ell(s,t)\mathrm{d}s = 1.$$

Moreover, for all $t > 0$ we have that $\ell\,(\cdot,t) \in L^1(\mathbb{R})$.

Proposition 3.1.2. *The function u defined in* (3.10) *solves* (3.6)*.*

Proof.

As $v \geq u^c$ and $\ell(s,t) = 0$ for $s < 0$, then

$$\int_0^\infty v(x,s)\,\ell(s,t)\mathrm{d}s \geq \int_0^\infty u^c(x)\,\ell(s,t)\mathrm{d}s = u^c(x)$$

and we obtain that $u \geq u^c$.

Now we prove that $(u-u^c)(\partial_t^\alpha u - \Delta u) = 0$. For $u > u^c$, it holds the classical theory about fractional Cauchy problems (see [13], [21, Theorem 5.2] and [39]) and therefore we have that, in $L^2(\Omega)$

$$\partial_t^\alpha u(x,t) = \Delta u(x,t), \quad u(x,0) = u^0(x).$$

Concerning the second inequality of (3.6), we observe that if $\partial_t v - \Delta v \geq 0$, then

$$\begin{aligned}\Delta \int_0^\infty v(x,s)\,\ell(s,t)\mathrm{d}s \leq & \int_0^\infty \partial_s v(x,s)\,\ell(s,t)\mathrm{d}s \\ = & - v(x,0)\frac{t^{-\alpha}}{\Gamma(1-\alpha)} - \int_0^\infty v(x,s)\,\partial_s \ell(s,t)\mathrm{d}s \\ = & - v(x,0)\frac{t^{-\alpha}}{\Gamma(1-\alpha)} + {}^R\partial_t^\alpha u(x,t).\end{aligned}$$

Since $u(x,0) = v(x,0)$ from the relation (3.5) we obtain $\partial_t^\alpha u - \Delta u \geq 0$.

The initial condition and the boundary condition for the function u follow trivially from the initial condition and the boundary condition for the function v. □

3.1.2 Asymptotic solution of the problem

In chapter 2, the asymptotic solution of (3.9) has been characterized as the solution of the corresponding stationary (elliptic) obstacle problem. In particular, it has been proved that under conditions (3.8), $\mathbf{H}_1$ and $\mathbf{H}_2$, the solution $v(t)$ to the parabolic obstacle problem (3.9) converges strongly in $H_0^1(\Omega)$, for $t \to \infty$, to the unique solution $\overline{u}$ of the corresponding stationary obstacle problem

$$\overline{u} \in H_0^1(\Omega), \quad \overline{u} \geq u^c, \quad -\Delta\overline{u} \geq 0, \quad (\overline{u} - u^c)(\Delta\overline{u}) = 0 \;\text{ a.e. in } \Omega. \tag{3.11}$$

Moreover, it has been proved that there is a constant $C > 0$ such that, for every $t \geq 1$,

$$\|v(t) - \overline{u}\|_{H^1(\Omega)} \leq e^{-Ct}\ . \tag{3.12}$$

In the next Theorem we prove that similar result holds for any $\alpha \in (0,1)$.

Theorem 3.1.1. *Assume conditions* (3.8), $\mathbf{H}_1$ *and* $\mathbf{H}_2$ *hold. Let* $u(t)$ *be the function defined in* (3.10) *solution of* (3.6). *Let* $\overline{u}$ *be the unique solution of the obstacle problem*

(3.11). *Then, $u(t)$ converges to $\overline{u}$ for $t \to \infty$, and there is a constant $C > 0$ such that, for every $t \geq 1$,*

$$\|u(t) - \overline{u}\|_{L^1(\Omega)} \leq \kappa_\Omega \, E_\alpha(-Ct^\alpha) \tag{3.13}$$

for any $\alpha \in (0,1)$, where $\kappa_\Omega = \left(\int_\Omega \mathrm{d}x\right)^{1/2}$ and

$$\frac{E_\alpha(-Ct^\alpha)}{J(t)} \to 1, \quad \textit{as} \quad t \to \infty \tag{3.14}$$

with

$$J(t) = \frac{1}{C} \frac{t^{-\alpha}}{\Gamma(1-\alpha)}. \tag{3.15}$$

Proof. Since

$$\|v - \bar{u}\|_{L^1(\Omega)} \leq \kappa_\Omega \; \|v - \bar{u}\|_{L^2(\Omega)}$$

we have

$$\begin{aligned}
\|u(t) - \bar{u}\|_{L^1(\Omega)} &= \int_\Omega \left| \int_0^\infty v(x,s)\, \ell(s,t) \mathrm{d}s - \bar{u}(x) \right| \mathrm{d}x \\
&= \int_\Omega \left| \int_0^\infty (v(x,s) - \bar{u}(x))\, \ell(s,t) \mathrm{d}s \right| \mathrm{d}x \\
&\leq \int_\Omega \int_0^\infty \left| v(x,s) - \bar{u}(x) \right| \ell(s,t) \mathrm{d}s \mathrm{d}x \\
&= \int_0^\infty \|v(x,s) - \bar{u}(x)\|_{L^1(\Omega)}\, \ell(s,t) \mathrm{d}s \\
&\leq \kappa_\Omega \int_0^\infty e^{-Cs}\, \ell(s,t)\, ds = \kappa_\Omega \, E_\alpha(-Ct^\alpha), \quad \forall \alpha \in (0,1)
\end{aligned}$$

where, in the last step, we used the fact that E_α is the Laplace transform of the density ℓ. For the asymptotic behavior of the Mittag-Leffler, consult the book [34, formula (4.4.17)].

□

3.2 Numerical analysis

Let us now conclude this chapter with the numerical study of problem (3.1).

3.2.1 Numerical approximation

The starting idea for a numerical approximation of the problem (3.1) has been to combine classical time-stepping discretization schemes for the Caputo derivative, such as the convolution quadrature (CQ) or the finite difference (L1) scheme, see e.g. [38], with schemes usually working for the parabolic obstacle problem (3.6), such as the one proposed in [20], but also the semi-implicit f.d. scheme tested in [3] for the equivalent Heaviside function formulation (3.1) of the problem .

For sake of simplicity we start from the one-dimensional case, with $\Omega = (a, b)$. Let us call $h > 0$ the space discretization step ($h = (b-a)/N$, so that we have $(N-1)$ internal nodes in Ω, $x_i = a + ih$, for $i = 1, ..., N-1$) and $\tau = T/M$ the time discretization step (with M time instants $t^m = m\tau$, for $m = 1, ..., M$); with $\gamma_\alpha = \tau^\alpha/h^2$ we denote the parabolic ratio between the steps related to a specific α. Note that for fixed steps h and τ, if α decreases to zero then γ_α quickly grows: in other words for small α, in order to keep γ_α small on a fixed mesh, the step τ has to be considerably reduced, with a significant increase of computational costs.

Here we analyze three possible approaches:

1. **Scheme S1: solves problem (3.1) with L1 for the Caputo derivative and the semi implicit f.d. scheme of [3] in space**

   The time discretization of the Caputo derivative by the L1 scheme leeds to the formula (see [38]):

$$\partial_t^\alpha u(x, t^m) \simeq \frac{1}{\Gamma(2-\alpha)\tau^\alpha} \left\{ u(x, t^m) - \sum_{k=0}^{m-1} C_{m,k} u(x, t^k) \right\},$$

   where

$$C_{m,0} := f(m), \quad C_{m,k} := f(m-k) - f(m-(k-1)) \quad \text{for } k = 1, ..., m-1,$$

   and

$$f(r) := r^{1-\alpha} - (r-1)^{1-\alpha}, \text{ for } r \geq 1.$$

Then, after semidiscretization in time, we need to solve for any instant t^m the equation

$$\frac{1}{\Gamma(2-\alpha)\tau^\alpha}\left\{u(x,t^m)-\sum_{k=0}^{m-1}C_{m,k}u(x,t^k)\right\}-H(u-u^c)u_{xx}=0 \quad \text{in } \Omega, \tag{3.16}$$

with the same boundary conditions of (3.1). Since Ω was splitted in N subintervals through the nodes $x=\{x_i\}_i$, the initial data will be the vector $u^0\in\mathbb{R}^{N-1}$, with $u_i^0=u^0(x_i)$; applying a semi implicit finite difference scheme in space, the solution u^1 at the first discrete time $t^1=\tau$ will solve at any node the relation:

$$\frac{1}{g\tau^\alpha}\left(u_i^1-C_{1,0}u_i^0\right)=H(u_i^0-u_i^c)\delta^2u_i^1:=H(u_i^0-u_i^c)\frac{u_{i-1}^1-2u_i^1+u_{i+1}^1}{h^2},$$

with $g=\Gamma(2-\alpha)$ (note that $0.8862\le g\le 1, \forall\alpha\in[0,1]$, so that this term will be negligible with respect to γ_α). If we set $v_i^k=u_i^k-u_i^c$, redistributing all the terms between the two members, it is equivalent to solve

$$u_i^1-g\tau^\alpha H(v_i^0)\delta^2u_i^1=C_{1,0}u_i^0 \quad \text{for every } i;$$

with vector notations it means that u^1 solves the linear system:

$$B^0u^1:=(I+g\gamma_\alpha H(v^0)*A)u^1=C_{1,0}u^0,$$

where A is the usual tridiagonal matrix $(N-1)\times(N-1)$ with values 2 on the main diagonal and -1 on the two adjacent diagonals, and we denoted

$$\{(H(v)*A)u\}_i:=H(v_i)(Au)_i=H(v_i)\sum_{j=1}^{N-1}A_{i,j}u_j.$$

Since the discrete solution could overstep the obstacle at some nodes, in particular when a large value of γ_α is used, the following correction is needed at any iteration:

$$u_i^1=\max(u_i^1,u^c(x_i)).$$

In the same way we see that u^2 solves for any i

$$\frac{1}{g\tau^\alpha}\left(u_i^2-C_{2,0}u_i^0-C_{2,1}u_i^1\right)=H(u_i^1-u_i^c)\delta^2u_i^2,$$

that is the system

$$B^1u^2 = (I + g\gamma_\alpha H(v^1) * A)u^2 = C_{2,0}u^0 + C_{2,1}u^1,$$

with the same matrix A and the subsequent correction. In general at any time step u^m solves the linear system

$$B^{m-1}u^m = b^m, \tag{3.17}$$

where we have set

$$B^{m-1} = I + g\gamma_\alpha H(v^{m-1}) * A, \qquad b^m = \sum_{k=0}^{m-1} C_{m,k}u^k \ , \tag{3.18}$$

followed by the correction

$$u^m = \max(u^m, u^c(x)). \tag{3.19}$$

Note that all the matrices B^m are symmetric positive definite (and M-matrices), since so it is A, while $H(v) \geq 0$. Then all the previous linear systems are well posed.

With respect to the classic parabolic obstacle problem approach discussed in [3] there is here an important difference. In that case (which corresponds to the case $\alpha = 1$), when the solution at time t^{m-1} touches the obstacle at node x_i, then $v_i^{m-1} = 0$, $H(v_i^{m-1}) = 0$, and (3.17) trivially yields:

$$u_i^m = u_i^{m-1}.$$

In other words, once touched the obstacle at a particular node the solution does not change there anymore, but only at the remaining free nodes. In the general case of $\alpha \in (0,1)$ on the contrary, at the contact time system (3.17) immediately yields for the i-th component:

$$u_i^m = \sum_{k=0}^{m-1} C_{m,k}u_i^k = C_{m,0}u_i^0 + ... + C_{m,m-1}u_i^c > u_i^c \sum_{k=0}^{m-1} C_{m,k} = u_i^c,$$

since at least $u_i^0 > u_i^c$ and $\sum_{k=0}^{m-1} C_{m,k} = 1$. It follows that the solution has a little rebound at x_i which detaches it again from the obstacle, and produces an (innatural) oscillating evolution from that time on. The width of such rebound depends on the size of γ_α. Then, even if in principle the semi implicit scheme does not not require stability restrictions on the discretization steps, a small value of γ_α will be necessary to reduce the oscillations (they would vanish for $\tau \to 0$, since in the continuous setting a null Caputo derivative implies $u_t = 0$, then a constant solution in time). As a consequence of that, the scheme would become enormously expensive, the more the more α is close to zero. The way to solve this difficulty is to remove the memory effect at the contact nodes of the solution with the obstacle, that is where the obstacle retains the solution. This suggests to modify the scheme replacing the vector b^m of (3.18) by the vector $\hat{b}^m$ defined by

$$\hat{b}^m = \max(H(v^{m-1}) * b^m, u^{m-1}); \tag{3.20}$$

if $H(v_i^{m-1}) = 0$, that is $u_i^{m-1} = u_i^c$, then $\hat{b}_i^m = u_i^{m-1}$ and $u_i^m = u_i^{m-1}$; otherwise $\hat{b}_i^m = b_i^m$, since $b_i^m \geq u_i^{m-1}$, and all remains as before. No rebound is still possible after a contact.

2. **Scheme S2: solves problem (3.1) with CQ for the Caputo derivative and the semi implicit f.d. scheme of [3] in space**

In this case the Caputo derivative is approximated through the so-called *convolution quadrature* (CQ) method, proposed by Lubich for the discretization of Volterra integral equations. In particular, if we consider the Riemann-Liouville derivative

$$^R\partial_t^\alpha \varphi := \frac{d}{dt}\frac{1}{\Gamma(1-\alpha)}\int_0^t (t-s)^{-\alpha}\varphi(s)\mathrm{d}s,$$

(with $\varphi(0) = 0$), it can be approximated by the discrete convolution:

$$^R\partial_\tau^\alpha \varphi^m := \frac{1}{\tau^\alpha}\sum_{j=0}^{m} c_j \varphi^{m-j},$$

where $\varphi^m = \varphi(t_m)$, and the coefficients $\{c_j\}$ are obtained from a suitable power series expansion, connected to a specific approximation method for the ODE (see [38]). In the case of the Euler backward method, it is known as the Grunwald-Letnikov approximation, and provides the following recursive formula for the coefficients:

$$c_0 = 1, \quad c_j = -\frac{\alpha - j + 1}{j} c_{j-1}.$$

Then, using the relation (3.5) between the Caputo and the Riemann-Liouville derivatives we can rewrite the initial problem as

$$^{R}\partial_t^\alpha(u-u^0) - H(u-\psi)\Delta u = 0,$$

which discretized in time and space (with the same notations of S1) becomes:

$$\frac{1}{\tau^\alpha}\sum_{j=0}^{m} c_j(u^{m-j}-u^0) + \frac{1}{h^2}H(u^{m-1}-u^c)Au^m = 0,$$

equivalent to the solution at any iteration of the linear system

$$B^{m-1}u^m = b^m, \tag{3.21}$$

where this time we have set

$$B^{m-1} = I + \gamma_\alpha H(u^{m-1}-u^c) * A, \qquad b^m = u^0 - \sum_{j=1}^{m-1} c_j(u^{m-j}-u^0), \tag{3.22}$$

followed again by the correction

$$u^m = \max(u^m, u^c(x)). \tag{3.23}$$

Even in this case the vector b^m has to be modified in the contact set in order to remove the memory effect and prevent rebounds, as done in (3.20). In fact when $u_i^{m-1} = u_i^c$ then again $b_i^m = 0$, and from (3.22) we get

$$u_i^m = u_i^0 - \sum_{j=1}^{m-1} c_j(u_i^{m-j}-u_i^0) > u_i^0 - (u_i^c - u_i^0)\sum_{j=1}^{m-1} c_j > u_i^c\ ,$$

since $u_i^{m-j} \geq u_i^c$ for any j, and $\sum_{j=1}^{m-1} c_j \geq -1$.

3. **Scheme S3: solves problem (3.6) with L1 for the Caputo derivative and the scheme of [20] for the evolutive obstacle problem.**

   If we discretize the equation of system (3.6) through finite differences, using the L1 scheme for the Caputo derivative, we get the equation

$$(u^m - u^c)^T\left(\frac{1}{g\tau^\alpha}(u^m - \sum_{k=0}^{m-1} C_{m,k}u^k) + \frac{1}{h^2}Au^m\right) = 0\ .$$

Setting $y^m = u^m - u^c$, and remembering that $\sum_{k=0}^{m-1} C_{m,k} = 1$, it is equivalent to

$$y^m \left(y^m + g\gamma_\alpha A(y^m + u^c) - \sum_{k=0}^{m-1} C_{m,k} y^k \right) = 0 \ .$$

Then $y^m = \max(0, x^m)$ is solution of the previous equation if x^m solves

$$(I + g\gamma_\alpha AP(x))x = b^m, \tag{3.24}$$

where now

$$b^m = \sum_{k=0}^{m-1} C_{m,k} y^k - g\gamma_\alpha A u^c = \sum_{k=0}^{m-1} C_{m,k} u^k - u^c - g\gamma_\alpha A u^c, \tag{3.25}$$

while $P(x) = diag(p(x_i))$ denotes the diagonal matrix with $p(x_i) = 1$ if $x_i > 0$ and $p(x_i) = 0$ otherwise. As seen in [20], (3.24) can be solved by the so-called Picard iterations:

$$P^0 = O, \quad (I + g\gamma_\alpha AP^n)x^{n+1} = b^m, \quad P^{n+1} = diag(p(x^{n+1})) \quad \text{per } n = 0, 1, ...$$

until $P^{n+1} = P^n$ (O is the null matrix); at that point x^{n+1} is the sought solution.

Of course other schemes could be obtained by different combinations of specific numerical approaches, but for our purposes the three previous schemes were sufficient to perform explicit simulations of the problem (see next Section).

3.2.2 Numerical tests

We have applied the schemes described in Section 3.2.1 to some specific examples, for different values of α. We have choosen a sufficiently large final time T, but also added a stopping time criterium in order to put in evidence the convergence towards the asymptotic solution. Since this convergence corresponds to the stabilization of the solution vector and to the satisfaction of the asymptotic complementarity relation

$$(u - u^c)\Delta u = 0,$$

which means u harmonic (linear in 1D) outside the contact set, we have used the criterium

$$STOP \ when \quad \max(\|u^m - u^{m-1}\|_\infty, \| (u^m - u^c)Au^m\|_\infty) < tol, \tag{3.26}$$

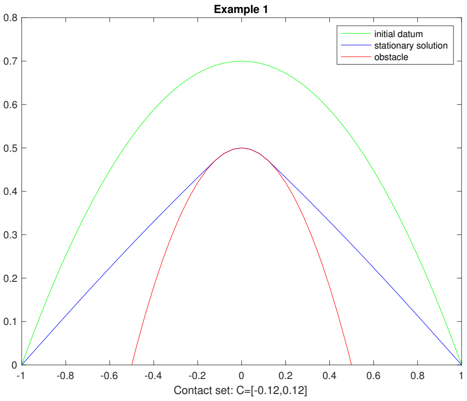


Figure 3.1: Data of Example 1.

for a given tolerance parameter tol.

Example 1. $\Omega = (-1, 1)$, $u^0(x) = 0.7 - 0.7x^2$, $u^c(x) = 0.5 - 2x^2$ (see Fig.3.1).

In the following Table 3.1 we reported some results obtained by the simulations with schemes S1, S2 and S3, for different values of α, N and γ_α, with $tol = 10^{-4}$. We adopted the following notations:

- FC time = full contact time, that is the first time at which the solution has reached the whole contact set (no more changing in the successive iterations);
- STOP time = the exit time according to criterium (3.26);
- # iter. = final number of time iterations;
- # Pic. = average number of Picard iterations for each time step in scheme S3;
- # LS = approximate number of linear systems to be solved (essentially the product of the previous two)
- S = the working schemes (the ones detecting the correct contact set).

Looking at the table 3.1 some remarks and comments are possible:

- The stationary solution is the same for any α, as stated by Theorem 3.1.1, and corresponds to the one of the stationary problem (3.7). Different is only the speed of convergence. The detected right extremum of the contact set C on the used meshes is 0.125 (the continuous value should be approximately 0.132).

Table 3.1

α	N	γ_α	τ	FC time	STOP time	# iter.	# Pic.	# LS	S
0	32	256	any	first it.	no conv.	1	14		S3
	64	1024	any	first it.	no conv.	1	29		S3
	128	4096	any	first it.	second it.	1	57		S3
0.3	32	60	0.0079	0.10	93.16	11739	12	140868	all
	32	75	0.016	0.11	93.15	5580	13	72540	S3
	64	220	0.0059	0.49	1.33	226	21	4746	all
	128	850	0.005	0.2	1.66	315	42	13230	all
0.5	32	25	0.009	0.16	8.38	880	8	7040	all
	32	50	0.038	0.19	8.39	221	11	2431	S3
	64	100	0.009	0.39	2.13	225	15	3375	all
	128	400	0.009	0.8	2.67	282	28	7896	all
0.7	32	50	0.097	0.29	5.81	61	11	671	all
	32	100	0.026	0.52	10.44	41	14	574	S3
	64	200	0.097	0.38	7.08	74	21	1554	all
	128	400	0.036	0.5	4.82	135	29	3915	all
1	32	15	0.058	0.27	0.99	18	7	126	all
	32	20	0.078	0.31	1.09	15	8	120	S3
	64	60	0.058	0.35	1.05	19	13	247	all
	128	240	0.058	0.64	1.05	19	23	437	all

- Schemes S1 and S2 have a very similar behaviour; in both cases for large values of γ_α the contact set can be overestimated, a problem not present for S3, due to the implicit nature of the quasi-Newton scheme of [20]. The semi implicit schemes S1 and S2 pay for the delay with which the contact information with the obstacle is achieved, allowing an uncorrect evolution of the solution. To avoid that, strong τ-step restrictions are necessary: experiments show that all the schemes work correctly for example if $\tau^\alpha < 0.1$, a bound which becomes particularly heavy when α is small. In the Table we reported approximately for any number of nodes the largest values of γ_α for which all the three schemes give the same correct solution.

- Computational costs: the previous remark suggests that S3 is the more reliable and even the less expensive of the three schemes, allowing larger time steps and hence less iterations. Anyway, any single time iteration of S3 is much more expensive, many Picard iterations (growing with N) with respect to a single linear system necessary to be solved for S1 and S2. Then, comparing the computational cost of the schemes, even these two schemes reveal competitive in terms of the total number of linear system to solve.

- For a fixed α the full contact time grows with the number of nodes, and does not seem to depend from γ_α. On the contrary the stabilization time grows with γ_α but decreases with the number of nodes.

- All the schemes correctly work also for the case $\alpha = 1$: it easy to see that in that case both the used approximations of the Caputo derivative reduce to the standard incremental ratio in time.

- The case $\alpha = 0$ is a sort of control test: since in that case $\partial_t^\alpha u = u - u^0$, in absence of an obstacle the equation (2.1) would reduce to the stationary equation

  $$-\Delta\overline{u} + \overline{u} = u^0. \tag{3.27}$$

  Then in presence of an obstacle we expect that the discrete solution satisfy (3.27) but only outside the contact set, and from the first iteration. It is in fact clear from (3.18) and (3.22) that since $H(u^0 - u^c) = 1$ and $\gamma_\alpha = 1/h^2$, the first iteration of all the schemes becomes

  $$B^0 u^1 = \left(I + \frac{1}{h^2}A\right) u^1 = u^0 \ ,$$

  which is essentially the discrete version of (3.27). If u^1 goes over the obstacle, such identity will be satisfied only where $u^1 > u^c$. In Figure 1.4 it can be seen what happens in our example with $N = 128$ nodes and scheme S3: the solution u is plotted in blue, the obstacle in red, the initial datum u^0 in black and the quantity $-\Delta u + u$ through asterisks: the last two quantities coincide in the detachment set, with a natural discontinuity at the boundary of such a region. Since the time step τ has no effect on the solution, the semi implicit schemes S1 and S2 for this example always overestimate the contact set.

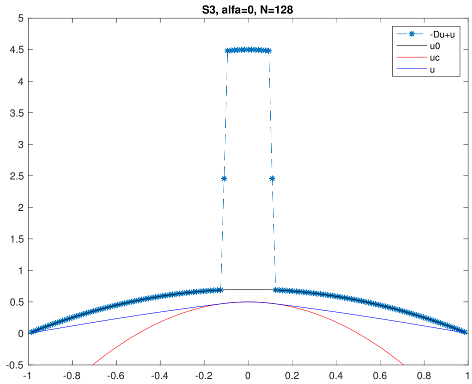


Figure 3.2: Solution for $\alpha = 0$ and 128 nodes, scheme S3.

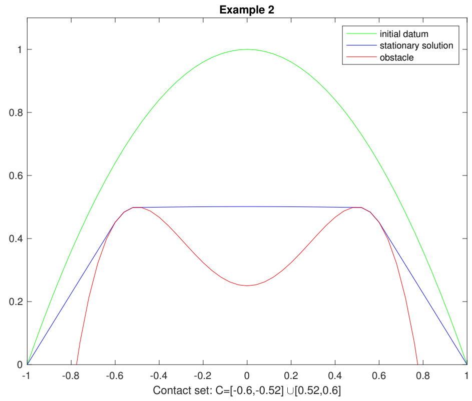


Figure 3.3: Data of Example 2.

Example 2. $\Omega = (-1, 1)$, $u^0(x) = 1 - x^2$, $u^c(x) = 0.5 - (2x^2 - 0.5)^2$ (see Fig.3.3).

On this example we tested numerically the estimate (3.13) of Theorem 3.1.1, using scheme S3. The L^1 norm of the error at time t^m on the given mesh was approximated by a natural quadrature formula, that is

$$\|u(t^m) - \bar{u}\|_{L^1(\Omega)} \simeq h \sum_i |u_i^m - \bar{u}_i|$$

(the vector $\bar{u}$ on the mesh was computed in advance with a sufficiently high precision). Such an error was computed for different values of α and of the time t^m. In Table 3.2, page 82, we reported (in the third column) the discrete L^1 error at $T = 10$ with $N = 32$ nodes and the same $\gamma_\alpha = 50$ for any α. In the fourth column the corresponding values of quantity $J(T) = \dfrac{1}{C}\dfrac{T^{-\alpha}}{\Gamma(1-\alpha)}$ of (3.15 in Theorem (3.1.1), page 70) are shown; for the constant C we adopted a computed numerical estimate ($C = 26$). It is evident that the two quantities decay at the same rate. In the second column it is also possible to see the number of instant times M needed for each α to keep the same γ_α, and the consequent increasing complexity of the computation. In order to confirm the order of decay of the error, polynomial for $\alpha \in (0, 1)$ and exponential (up to the machine precision) for $\alpha = 1$, we illustrate in Figure 3.4 such behavior plotting the quantities $J(t)$ in the time interval $(0, 20)$ for $\alpha = 0.3, 0.6$ and 0.9, and the computed errors on ten different discrete times (marked by asterisks).

We close this chapter by arguing that 2D examples are not interesting in this case.

Table 3.2: Example 2: $N = 32$, $\gamma_\alpha = 50$, scheme $S3$.

α	M	L^1 error at $T = 10$	$T^{-\alpha}/(C\Gamma(1-\alpha))$
1	51	$8.07\ 10^{-6}$	$\simeq 0$
0.9	61	$5.27\ 10^{-4}$	$5.09\ 10^{-4}$
0.8	77	$1.37\ 10^{-3}$	$1.32\ 10^{-3}$
0.7	103	$2.62\ 10^{-3}$	$2.56\ 10^{-3}$
0.6	152	$4.41\ 10^{-3}$	$4.35\ 10^{-3}$
0.5	262	$6.87\ 10^{-3}$	$6.86\ 10^{-3}$
0.4	593	$1.01\ 10^{-2}$	$1.02\ 10^{-2}$
0.3	2313	$1.44\ 10^{-2}$	$1.48\ 10^{-2}$
0.2	35184	$1.98\ 10^{-2}$	$2.08\ 10^{-2}$
0.1	123000000	$2.7\ 10^{-2}$ (est.)	$2.85\ 10^{-2}$

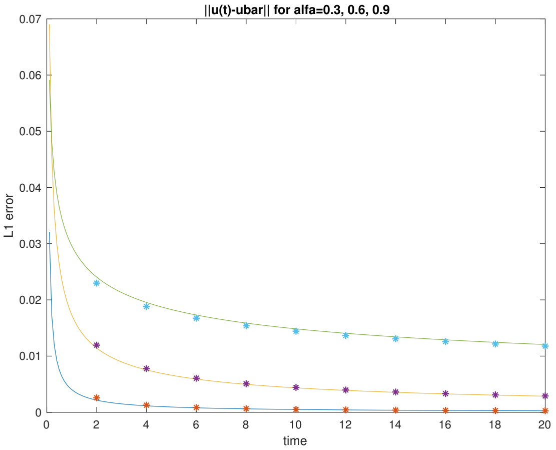


Figure 3.4: Example 2: polynomial error decay for $\alpha = 0.3, 0.6$ and 0.9.

Naturally, the numerical schemes proposed in this chapter also work adequately in situations similar to those already explained in chapter 2, but since the final results are the same already shown in the same chapter and since it is possible to provide in the figure only the initial and final state of the simulations, in the end we decided not to report them.

Conclusions and future developments

The issues faced in this thesis started from a numerical study referred to two recent works about SOC by V. Barbu and U. Mosco, to arrive at the theoretical and numerical study of two Heaviside function driven degenerate diffusion models. This study made it possible to highlight some structural aspects of these mathematical models (such as the asymptotic behavior of their solution over time) and to be able to connect them, in some cases, to known results linked both to the theory of obstacle problems and fractional calculus. The common link between these mathematical models is, as mentioned, a different use of the Heaviside function which, depending on its use within the Laplace operator (see results presented in chapter 1) or as a degenerate diffusion coefficient applied externally to the same operator (see results presented in chapters 2 and 3), was able to highlight the following results:

1. the numerical study of the problem (1.7), page 6, was made possible thanks to Theorem (1.4.1), page 30, and Proposition (1.4.1), page 31. These results have shown that the behavior of the solution of the problem (1.7) is not affected by the use of functions approximating the multivalued Heaviside function, w. r. t. the theoretical results obtained by V. Barbu and U. Mosco. Clearly the meaning of the approximation of these functions is given in the sense of the estimate (1.80), page 26. The content of chapter 1 is taken from [2].

2. the study of the problem (2.1), page 45, has shown that this model behaves as an evolutive variational inequality having the target as an obstacle: under suitable hypotheses, starting from an initial state above the target the solution evolves in time towards an asymptotic solution, eventually getting in contact with part of the target itself. This result was made possible through Proposition (3.1.1), page 68,

and Theorem (3.1.1), page 70. The content of chapter 2 is taken from [3].

3. the study of the problem (3.1), page 65, new in fractional calculus field, has allowed us to bring to light some interesting aspects that connect it to the related fractional obstacle problem, as well as being a generalization of the problem (2.1), page 45, presented in chapter 2 of the present work that it is given only for $\alpha = 1$. More precisely, by analogy with the classical case, problem (3.1) is equivalent for all $\alpha \in (0, 1)$ to the complementarity system joined under the same initials and at the edge conditions and, asymptotically, the solution evolves for each α towards the same steady state of the classical problem with different convergence speed (i. e. for $\alpha = 1$ exponential, for $\alpha \in (0, 1)$ polynomial). The content of chapter 3 is taken from [4].

On the other hands, about possible future perspectives and further developments, it might be interesting

1. to investigate these problems in a more complex and general contest, like e.g. pre-fractal or fractal domains, as to extend our results;
2. referring to the obstacle problem equivalences about the models discussed in chapter 2 and chapter 3, it would be interesting to extend the validity of our results also for much more irregular obstacle functions, at least non-continuous, for example;
3. to give a probabilistic view of the problem (3.1), as to deeply study the behavior of its solution in a more general and more developed fractional calculus theory.
4. to study in deep the adopted numerical approaches as to determine the most efficient one from the point of view of the computational cost and to study in more detail the error trend and improve the definition of the stopping criteria.

Appendix A

Some theoretical aspects behind Chapter 2 - The obstacle problem

The literature behind obstacle problems and, more generally, behind variational inequalities is very large; mainly defined in the seventies/eightees of the last century, see for example [17, 25] and [32] et a., it has undergone great evolutions and generalizations for the following decades up to the most recent applications in mathematical modelling and general interests still present in the last years, like [1, 36] and [51] *et al.*

So, due to the vastness of the topic, in this Section we will refer to the works cited in chapter 2 and we will recall only the results that are directly linked to the results reported in chapter 2.

A.1 Preliminaries

Let us start by denoting with Ω a bounded open subset of $\mathbb{R}^n$, with $\partial\Omega$ its Lipschitz continuous boundary, in which Ω being locally on one side of $\partial\Omega$, and by $\delta_1\Omega$ an open subset of $\partial\Omega$.

Definition A.1.1. *Let $k \in \mathbb{N} \cup \{0\}$ and $1 \leq p$.*

$$L^p(\Omega) := \left\{ f : \Omega \to \mathbb{R}; f \text{ is measurable on } \Omega \text{ and } \int_\Omega |f(x)|^p \, \mathrm{d}x < \infty \right\},$$

$$L^\infty(\Omega) := \{ f : \Omega \to \mathbb{R}; f \text{ is measurable and essentially bounded on } \Omega \},$$

$$C^k(\overline{\Omega}) := \left\{ f : \overline{\Omega} \to \mathbb{R}; f \text{ has a continuous } k\text{-th derivative} \right\},$$

$C(\overline{\Omega}) := C^0(\overline{\Omega})$,

$H^k(\Omega) := W^{k,2}(\Omega)$, *i.e. the Sobolev space of order k on Ω, in our case $k = 1, 2$, when $p = 2$.*

Remark A.1.1. *Of course we will intend that the spaces L^p are provided with the usual Banach norm*

$$\|f\|_{L^p(\Omega)} = \left(\int_\Omega f^p \mathrm{d}x\right)^{\frac{1}{p}}$$

while for the space $H^1(\Omega)$ it will hold

$$\|f\|_{H^1(\Omega)} = \left(\int_\Omega f^2 \mathrm{d}x + \sum_{i=1}^{n} \int_\Omega f_{x_i}^2 \mathrm{d}x\right)^{\frac{1}{2}}$$

with the usual notation $f|_{\partial_1\Omega}$ for the trace on $\partial_1\Omega$ of $f \in H^1(\Omega)$.

In the case that $f \in H^1(\Omega)$ and $\partial_1\Omega \neq \emptyset$ it is possible to say that $f|_{\partial_1\Omega}$ is a square integrable function on $\partial_1\Omega$ (for a more detailed discussion, see [14] and [19]) and it is possible to define a closed subspace V of $H^1(\Omega)$ by setting

$$V = \left\{v \in H^1(\Omega) \text{ s.t. } v|_{\partial_1\Omega} = 0\right\}.$$

Remark A.1.2. *If $\partial_1\Omega = \partial\Omega$, V coincides with the Sobolev space $H_0^1(\Omega)$, i.e. with the closure of $H^1(\Omega)$ of the subspace of smooth functions with compact support in Ω. Moreover, if $\partial_1\Omega = \emptyset$, the equality $V = H^1(\Omega)$ holds. Anyway we have*

$$H_0^1(\Omega) \subseteq V \subseteq H^1(\Omega).$$

It will be identified $L^2(\Omega)$ with its dual, and it will be denoted by V^* the dual of V. The following inclusions hold

$$V \subset L^2(\Omega) \subset V^*.$$

Finally, let T be a positive real numeber, and let $Q = \Omega \times]0, T[$, $\Sigma = \partial\Omega \times]0, T[$ and $\Sigma_1 = \partial_1\Omega \times]0, T[$. Under these assumption, it will be set $\mathcal{V} = L^2(0, T; V)$ and $\mathcal{V}^* = L^2(0, T; V^*)$, where, if X is a Hilbert space endowed with the norm $|\cdot|_X$, then $L^2(=, T; X)$ is the space of all functions f from $]0, T[$ onto X s.t. $|f(t)|_X$ is square-integrable on $]0, T[$. This space will be naturally endowed with the following Hilbert norm

$$\|f\|_{L^2(0,T;X)} = \left(\int_0^T |f(t)|_X^2 \,\mathrm{d}t\right)^{\frac{1}{2}}.$$

Remark A.1.3. *Under these assumptions, we have that $\mathcal{V}$ turns out to be the closed subset of $L^2(0,T;H^1(\Omega))$ that contains all those functions whose traces on Σ_1 are zero; while the space $L^2(Q) = L^2(0,T;L^2(\Omega))$ is identified with its dual; then, the following continuous and dense embeddings hold*

$$\mathcal{V} \subset L^2(Q) \subset \mathcal{V}^*.$$

Now, with the symbol $(\cdot,\cdot)$ it will be denoted both the duality pairing between elements of V^* and elements of v, and the corresponding scalar product in $L^2(Q)$. Moreover, a functional $F \in \mathcal{V}^*$ it will be defined as *positive*, i.e. $F \geq 0$, if $(F,f) \geq 0$ for any $f \in \mathcal{V}$, $f \geq 0$ a.e. in Q.

We are finally ready to define the following

$$\begin{aligned} \mathcal{W} &= \{f \in \mathcal{V} \text{ s.t. } \partial_t f \in \mathcal{V}^*\}, \\ \widetilde{\mathcal{W}} &= \{f \in L^2(0,T;H^1(\Omega)) \text{ s.t. } \partial_t f \in \mathcal{V}^*\} \end{aligned}$$

in which we have set $\partial_t f$ as the derivative in the sense of the distributions on $]0,T[$ with range in $\mathcal{V}^*$.

Remark A.1.4. *Both $\mathcal{W}$ and $\widetilde{\mathcal{W}}$, endowed with their respective graph norms, are continuously imbedded into the space of continuous functions from $[0,T]$ into $L^2(\Omega)$, endowed with the maximum norm. For a more detailed discussion about this topic, see also* [14], Chapter I, Theorem 3.1.

Let us now define the following bilinear form on $L^2(0,T;H^1(\Omega))$:

$$a(f,g) = \int_Q \left[\sum_{i,j}^{n} a_{i,j} f_{x_i} g_{x_j} + \sum_{i=1}^{n} b_i f_{x_1} g + cfg \right] \mathrm{d}x \ \mathrm{d}t$$

and, of course, let us assume that the coefficients be essentially bounded and measurable real functions on Q. Moreover, let us assume that there exists a real number $\delta > 0$ s.t.

$$\sum_{i,j=1}^{n} a_{ij}\xi_i\xi_j \geq \delta\,|\xi|^2, \qquad \text{for all } \xi \in \mathbb{R}^n, \text{a.e. in } Q;$$

and that there exist real numbers $\alpha > 0$ and $\lambda \geq 0$ s.t.

$$a(f,f) \geq \alpha \left\| f \right\|_{\mathcal{V}}^2 - \lambda \left\| f \right\|_{L^2(Q)}^2 \quad \text{for all } f \in \mathcal{V}. \tag{A.1}$$

In order to define a bounded linear operator $\mathcal{A}$ (that will be, in our case, the Laplacian operator), we have to define it through the following identity

$$\mathcal{A} : L^2(0,T;H^1(\Omega)) \to \mathcal{V}^* \text{ s.t.}$$
$$g \in L^2(0,T;H^1(\Omega)), \quad (\mathcal{A}f, g) = a(f,g) \quad \text{for all } g \in \mathcal{V}.$$

Now, let us define $\psi \in \widetilde{\mathcal{W}}$ s.t.

$$\psi|_{\Sigma_1} \leq 0 \tag{A.2}$$

in the case $\partial_1\Omega \neq \emptyset$, or, equivalently, $\Sigma_1 \neq \emptyset$, and

$$\psi(0) \leq 0 \tag{A.3}$$

assume that a $h \in \mathcal{V}^*$, defined as

$$h = \partial_t\psi + \mathcal{A}\psi \tag{A.4}$$

can be decomposed as

$$h = h^+ - h^- \tag{A.5}$$

with h^+ and h^- positive elements of $\mathcal{V}^*$.

Finally, let us define a closed convex subset $\mathcal{K} \subseteq \mathcal{V}$ s.t.

$$\mathcal{K} = \left\{ f \in \mathcal{V} \text{ s.t. } f \geq \psi \text{ a.e. in } Q \right\}.$$

Remark A.1.5. *$\mathcal{K}$ is a not empty subset of $\mathcal{V}$ since at least it contains $\psi \vee 0$, i.e. the supremum ($\vee$) between ψ and 0. To assert this property it has been used the* lattice property *of $L^2(0,T;H_1(\Omega))$, which follows from the similar property of $H^1(\Omega)$ proven in* [52].

A.2 Main results

In this Section let us recall the main results that are at the base of Chapter 2 of this thesis.

Theorem A.2.1. *Under assumptions* (A.1)-(A.3) *and* (A.5) *there exists a unique solution to the problem*

$$u \in \mathcal{K} \cap \mathcal{W}, \quad (\partial_t u + \mathcal{A}u, v - u) \geq 0 \quad \textit{for all } v \in \mathcal{K} \tag{A.6}$$

with the initial condition

$$u(0) = 0. \tag{A.7}$$

Moreover, the following a priori estimates hold

$$0 \leq \partial_t u + \mathcal{A}u \leq h^+. \tag{A.8}$$

Remark A.2.1. (A.6) *is an example of the so-called* variational inequality*. The a priori estimates proposed in the above theorem is a Lewy-Stampacchia type inequality for obstacle problems. The hypothesis set in* Chapter 2 *of this thesis (*$\mathbf{H}_2$*) is given just in this sense.*

Remark A.2.2. *The case of a parabolic variational inequality s.t.*

$$u \in \mathcal{K} \cap \mathcal{W}, \quad (\partial_t u + \mathcal{A}u, v - u) \geq (f, v - u) \quad \textit{for all } v \in \mathcal{K} \tag{A.9}$$

$f \in \mathcal{V}^*$*, with the initial condition*

$$u(0) = u_0 \tag{A.10}$$

$u_0 \in L^2(\Omega)$*, is essentially the same as that of* (A.6) *and* (A.7)*, provided* $\psi(0)$ *is s.t.*

$$\psi(0) \leq u_0$$

instead of (A.3) *and* (A.5) *is assumed to hold for*

$$h = \partial_t \psi + \mathcal{A}\psi - f$$

instead of (A.4)*.*

Of course, let $\hat{u}$ be the solution of the following problem

$$\hat{u} \in \mathcal{W}, \quad \partial_t \hat{u} + \mathcal{A}\hat{u} = f, \quad \hat{u}(0) = u_0.$$

Then, the problem with obstacle $\psi - \hat{u}$ can be solved through the use of the Theorem (A.2.1); in fact, writing the solution corresponding to $\psi - \hat{u}$ as $u - \hat{u}$, it is possible to verify that the function u so obtained is the solution (necessarily unique) to (A.9) and (A.10); moreover,

$$f \leq \partial_t u + \mathcal{A}u \leq h^+ + f.$$

Let us conclude this Section by reporting an extract from the table taken from [17], p. 100 set in the case that $\mathcal{V} = H^1(\Omega)$, and $\mathcal{V}^* = H^{-1}(\Omega)$, recalling that $u^c \in H^2(\Omega)$.

Hypothesis	Conclusions
(a) $f \in L^2(0,T;H^{-1}(\Omega))$ $u_0 \in L^2(\Omega)$	$u \in C([0,T];L^2(\Omega)) \cap L^2(0,T;H_0^1(\Omega))$
(b) $f \in L^2(0,T;L^2(\Omega))$ $u_0 \in L^2(\Omega)$	$u \in C(]0,T];H_0^1(\Omega)) \cap L^2(\delta,T;H^2(\Omega))$ $\partial_t u \in L^2(\delta,T;L^2(\Omega))$ $\partial_t u - \Delta u \in L^2(0,T;L^2(\Omega))$ for all $\delta \in]0,T[$
(c) $f \in L^2(0,T;L^2(\Omega))$ $u_0 \in H_0^1(\Omega)$	$u \in C([0,T];H_0^1(\Omega)) \cap L^2(0,T;H^2(\Omega))$ $\partial_t u - \Delta u \in L^2(0,T;L^2(\Omega))$

For a more detailed discussion about the proofs of these result by Brezis, let see [17], for (a) Corollary II.1 and for (b) and (c), Theorem II.9 and Corollary II.2.

Appendix B

Some theoretical aspects behind Chapter 3 - Fractional calculus

The first papers relating to the fractional calculus are dated back from the end of the seventeenth century, starting from a correspondence between Leibnitz and de l'Hôpital in 1695. Since then, a huge literature on the subject has been produced within which it is significant to remember, for example, the contribution of Abel in 1823. Starting from the problem of the tautochrona, that is, that curve for which the necessary time taken by a particle sliding on it, without any friction forces and under the sole (uniform) gravity force is independent of its starting point on it, problem, however, already solved by Huygens in 1659, Abel showed that the total time required by the particle to slide on it, under these assumptions, is given by

$$t(x) = C \int_0^x \frac{s'(y)}{\sqrt{x-y}} \mathrm{d}y$$

where s is the arclenght of the curve. Thus, it was shown that the satisfactory curve to these characteristics was the cycloid, but, starting, from the point of view of the fractional calculus, this result can be interpreted as a Caputo-Djrbashian derivative of s, or the fractional integral of $v = s'$, both of order $\frac{1}{2}$.

As seen before, in this Section we will recall only the results that are directly linked to the results reported in chapter 3.

B.1 Preliminaries

Let us start with the basics of fractional calculus. In the following Sections, we will refer to the works cited in chapter 3, in particular to [28, 29] for the Paragraph about the Laplace transform and to [15] whenever we will recall the generalized Newton's binomial formula (see e.g., Sections B.1, B.3 and B.4). Moreover, about the notation of the functional spaces we will need to recall, we will refer to Definition (A.1.1).

Theorem B.1.1. (Fundamental Theorem of Classical Calculus) *Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and let $F : [a, b] \to \mathbb{R}$ s.t.*

$$F(x) := \int_a^x f(t)\mathrm{d}t.$$

Then, F is differentiable and

$$F' = f.$$

Definition B.1.1. *Let us also define the following operators:*

1. *By ∂, it will be denoted the operator that maps a differentiable function onto its derivative, i.e.*

$$\partial f(x) := f'(x).$$

2. *Assuming that f is a Riemann integrable function on the compact set $[a, b]$, by J_a, it will be denoted the operator that maps f onto its primitive centered at a, i.e.*

$$J_a f(x) := \int_a^x f(t)\mathrm{d}t$$

for $a \le x \le b$.

3. *Let $n \in \mathbb{N}$. It will be used the symbol ∂^n and the symbol J_a^n to denote the n-th iteration of ∂ and J_a, respectively, i.e. $\partial^1 := \partial$, $J^1 := J, \dots, \partial^n := \partial\partial^{n-1}$ and $J^n := JJ^{n-1}$, with $n \ge 2$.*

About the operator J_a^n, in the case $n \in \mathbb{N}$, the following formula holds:

Lemma B.1.1. *Let $n \in \mathbb{N}$ and f be a Riemann integrable function on the compact set $[a, b]$. Then, for $a \le x \le b$, holds*

$$J_a^n f(x) = \frac{1}{(n-1)!}\int_a^x (x-t)^{n-1} f(t)\mathrm{d}t. \tag{B.1}$$

The proof of this Lemma, given by induction, can be found in [50], see eq. (2.16). The formentioned formula is also known as *Cauchy formula for repeated integration.*

Moreover, an immediate consequence of Theorem (B.1.1), for the operators ∂ and J_a, is the following:

Lemma B.1.2. *Let $m, n \in \mathbb{N}$ s. t. $m > n$, and let f be a function having a continuous n-th derivative on the interval $[a, b]$. Then,*

$$\partial^n f = \partial^m J_a^{m-n} f.$$

The proof of this Lemma can be found in [28], p. 8.

As one of the main purpose of the basis of the fractional calculation is to give meaning to the exponents of the operators ∂^n and J_a^n when $n \notin \mathbb{N}$, the following definition generalizes the factorial term in (B.1) to non-integer arguments, i.e. when $n \notin \mathbb{N}$. Let us introduce the Euler's Gamma function.

Definition B.1.2. *Let be $z \in \mathbb{C}$. The function $\Gamma(z)$, defined by*

$$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} \mathrm{d}t,$$

is called Euler's Gamma function. This integral converges in the right half of the complex plane $\Re(z) > 0$.

Remark B.1.1. *However, as the identity $\Gamma(z) = \dfrac{\Gamma(z+1)}{z}$ holds, and the analytic continuation can be used, one can uniquely extend the integral formulation for $\Gamma(z)$ to a meromorphic function defined for all complex numbers z, except integers less than or equal to zero, where the function has simple poles. In the following Sections, we will alway refer to a Gamma function extended in this way.*

Theorem B.1.2. *For $n \in \mathbb{N}$ it holds true that*

$$(n-1)! = \Gamma(n).$$

Theorem B.1.3. (Fundamental Theorem in Lebesgue Spaces). *Let $f \in L^1[a, b]$. Then, $J_a f$ is differentiable almost everywhere in $[a, b]$, and $\partial J_a f = f$ also holds almost everywhere on $[a, b]$.*

A proof of this theorem can be found in [49], §23.

Finally, the following definition holds:

Definition B.1.3. *It will be denoted by A^n or $A^n[a,b]$ the set of functions with absolutely continuous $(n-1)$-st derivative, i.e. the functions f for which there exists (almost everywhere) a function $g \in L^1[a,b]$ s. t.*

$$f^{(n-1)}(x) = f^{(n-1)}(a) + \int_a^x g(t)\mathrm{d}t.$$

In this case, g will be called as the (generalized) n-th derivative of f, and it will be simply written $g = f^{(n)}$.

The Laplace transform. Now, let us recall the definition and main properties of the Laplace transform here given for a real variable $s > 0$. In particular, this method of transformation from differential equations to algebraic equations is extremely useful and efficient also in the fractional calculus context.

Definition B.1.4. *Let $f : [0,\infty) \to \mathbb{R}$ be a given function. The function F defined by*

$$F(s) := \mathcal{L}f(s) := \int_0^\infty f(x)e^{-sx}\mathrm{d}x$$

is called the Laplace transform *of f whenever the integral exists.*

Theorem B.1.4. *Let us assume that the functions f_1, f_2 and f_3 to be given on $[0,\infty)$ and to be such that their Laplace transforms exist for all $s \geq s_0$ with some suitable $s_0 \in \mathbb{R}$. Then, the following holds:*

1. *if $f_3 = a_1 f_1 + a_2 f_2$ with arbitrary real constants a_1 and a_2 then*

$$\mathcal{L}f_3(s) = a_1\mathcal{L}f_1(s) + a_2\mathcal{L}f_2(s);$$

2. *if f_3 is the convolution of f_1 and f_2, i.e. if*

$$f_3(x) = \int_0^x f_1(x-t)f_2(t)\mathrm{d}t,$$

then

$$\mathcal{L}f_3(s) = \mathcal{L}f_1(s) \cdot \mathcal{L}f_2(s);$$

3. *if* $f_3(x) = \int_0^x f_1(t)\mathrm{d}t$, *then it holds for* $s > \max\{0, s_0\}$
$$\mathcal{L}f_3(s) = \frac{1}{s}\mathcal{L}f_1(s);$$

4. *let* $m \in \mathbb{N}$. *If* $f_3 = \partial^m f_1$ *is the* m^{th} *derivative of* f_1 *then*
$$\mathcal{L}f_3(s) = s^m\mathcal{L}f_1(s) - \sum_{k=1}^{m} s^{m-k} f_1^{(k-1)}(0);$$

5. *let* $a > 0$ *and* $f_3(x) = f_1(ax)$. *Then*
$$\mathcal{L}f_3(s) = \frac{1}{a}\mathcal{L}f_1(s/a);$$

6. *let* $a \in \mathbb{R}$ *and* $f_3(x) = e^{-ax}f_1(x)$. *Then*
$$\mathcal{L}f_3(s) = \mathcal{L}f_1(s+a);$$

7. *let* $m \in \mathbb{N}$ *and* $f_3(x) = x^m f_1(x)$. *Then*
$$\mathcal{L}f_3(s) = (-1)^m \frac{\mathrm{d}^m}{\mathrm{d}s^m}\mathcal{L}f_1(s);$$

8. *let* $f_3(x) = f_1(x)/x$. *Then*
$$\mathcal{L}f_3(s) = \int_s^\infty \mathcal{L}f_1(\sigma)\mathrm{d}\sigma;$$

9. *let* $a \in \mathbb{R}$ *and*
$$f_3(x) = \begin{cases} 0 & \textit{for } x < a \\ f_1(x-a) & \textit{for } x \geq a \end{cases}$$
then
$$\mathcal{L}f_3(s) = e^{-as}\mathcal{L}f_1(s).$$

Hadamard's finite-part integral. Let us conclude this Section by reporting the definition of the so-called *Hadamard's finite-part integral.* This definition is useful to better understand the nature of the thesis of the Lemma B.3.2, at the end of Section B.3, in which the integral in (B.5) is not a convergent one.

In general, integrals of type $\int_a^b (x-a)^{-\eta} f(x)\mathrm{d}x$ are not convergent for $\eta \geq 1$ as $f(a) \neq 0$. However, it is useful to set a finite value to these integrals. This idea was conceived by Hadamard to solve some problems related to the solution methods of partial differential equations, and passed under the name of *finite part-integral.* This short paragraph is based on the assumption that $\eta \notin \mathbb{N}$, while considering integer values for η requires some modifications.

Before proceeding, let us recall the following definition, useful both in the Hadamard's finite-part inegral concept and in the definition of the Riemann-Liouville fractional differential operator (see Section B.3, p. 98).

Definition B.1.5. *Let be $x \in \mathbb{R}$, then the ceiling function is defined as follows*

$$\lceil x \rceil := \min \{m \in \mathbb{Z} : m \geq x\}$$

while the floor function as

$$\lfloor x \rfloor := \max \{m \in \mathbb{Z} : m \leq x\}$$

At a general level, the Hadamard finite-part of an integral is defined as a Taylor expansion of f at $x = a$ in which the resulting singular integrals are defined as

$$\int_a^b (x-a)^{-\eta} \mathrm{d}x = \frac{1}{1-\eta} (b-a)^{1-\eta} \quad (\eta > 1) . \tag{B.2}$$

In this way, it is possible to replace the divergent integral $\int_a^b (x-a)^{-\eta} f(x) \mathrm{d}x$ with the following convergent one $\int_{a+\delta}^b (x-a)^{-\eta} f(x) \mathrm{d}x$, for $\delta > 0$. So we have:

$$\int_{a+\delta}^b (x-a)^{-\eta} f(x) \mathrm{d}x = \frac{1}{1-\eta} \left[(b-a)^{1-\eta} - \delta^{1-\eta} \right] .$$

Passing to the limit for $\delta \to 0$, although the latter does not exist, Hadamard proposed not to consider the quantity $\lim_{\delta \to 0} \frac{\delta^{1-\eta}}{1-\eta}$ (which is divergent), but only the quantity $\frac{(b-a)^{1-\eta}}{1-\eta}$ that instead is finite.

Moreover, it is possible to define this concept more precisely, considering for $\eta \notin \mathbb{N}$, as:

$$\int_a^b (x-a)^{-\eta} f(x) \mathrm{d}x := \sum_{k=0}^{\lfloor \eta \rfloor - 1} \frac{f^{(k)}(a)(b-a)^{k+1-\eta}}{(k+1-\eta)k!} + \int_a^b (a-b)^{-\eta} \mathbf{R}_{\lfloor \eta \rfloor - 1}(x, a) \mathrm{d}x \tag{B.3}$$

in which the quantity

$$\mathbf{R}_p(x,a) := \frac{1}{p!}\int_a^x (x-y)^p \, f^{(p+1)}(y)\mathrm{d}y \tag{B.4}$$

is the remainder of the pth degree Taylor polynomial of f with expansion point a. We remark that a sufficient condition that guarantees the existence of the integral in (B.3) is that $f \in C^s[a,b]$ if $\eta - 1 < s \in \mathbb{N}$.

Most important properties of the Hadamard's finite-part integral are the following:

1. While Riemann and Lebesgue integrals are a positive functionals, the Hadamrd's finite-part integral not. This means that the inequality
$$\left| \int_a^b (x-a)^{-\eta} f(x)\mathrm{d}x \right| \leq \int_a^b (x-a)^{-\eta} \left| f(x) \right| \mathrm{d}x$$
is not true in general;
2. the finite-part integral is a consistent extension of the concept of regular integrals;
3. the finite-part integral is additive w.r.t. the union of integration intervals and invariant w.r.t. translation;
4. the finite-part integral is linear;
5. the standard change of variable rule still holds even if $\eta \notin \mathbb{N}$.

B.2 Riemann-Liouville integrals

Let us define the following operator:

Definition B.2.1. *Let $n \in \mathbb{R}^+$. The operator ${}^R J_a^n$, defined on $L^1[a,b]$ by*

$${}^R J_a^n f(x) := \frac{1}{\Gamma(n)}\int_a^x (x-t)^{n-1} f(t)\mathrm{d}t$$

for $a \leq x \leq b$, is called the Riemann-Liouville fractional integral operator of order n.

Remark B.2.1. *For $n = 0$, it will be set ${}^R J_a^0 := I$, i.e. the identity operator.*

Theorem B.2.1. *Let $f \in L^1[a,b]$ and $n > 0$. Then, the integral ${}^R J_a^n f(x)$ exists for almost every $x \in [a,b]$. Moreover, the function ${}^R J_a^n f$ itself is also an element of $L^1[a,b]$.*

The proof of the Theorem can be find in [28], p. 13.

Let us now recall one of the main property of integer-order integral operators

Theorem B.2.2. *Let $m, n \geq 0$ and $f \in L^1[a,b]$. Then,*

$$ {}^R J_a^m \, {}^R J_a^n f = {}^R J_a^{m+n} f $$

holds almost everywhere on $[a,b]$. If additionally $f \in C[a,b]$ or $m+n \geq 1$, then the identity holds everywhere on $[a,b]$.

The proof of the Theorem can be find in [28], p. 14.

Corollary B.2.1. *Under the assumptions of* Theorem (B.2.2),

$$ {}^R J_a^m \, {}^R J_a^n f = {}^R J_a^n \, {}^R J_a^m f $$

The proof of the Theorem can be find in [28], p. 14.

B.3 Riemann-Liouville derivaties

Recalling Lemma B.1.2, condition m and n are integers s.t. $m > n$ is now generalized here to the following assumption: let us assume that n is not an integer: however is still possible to continue to choose an integer m, s. t. $m > n$, but now there is an important difference between the classical case (i.e. m and n integers) and the present situation (i.e. n is not an integer): the operator defined in this way depends on the choice of the point a. Let us give the following definitions:

Definition B.3.1. *Let $n \in \mathbb{R}^+$ and $m = \lceil n \rceil$. The operator ${}^R\partial_a^n$, s.t.*

$$ {}^R\partial_a^n f := {}^R\partial_a^m \, {}^R J_a^{m-n} f $$

is called the Riemann-Liouville fractional differential operator of order *n.*

Remark B.3.1. *For $n = 0$, it will be set ${}^R\partial_a^0 := I$, i.e. the identity operator.*

Lemma B.3.1. *Let $f \in A^1[a,b]$ and $0 < n < 1$. Then ${}^R\partial_a^n f$ exists almost everywhere in $[a,b]$. Moreover ${}^R\partial_a^n f \in L^p[a,b]$ for $1 \leq p < \frac{1}{n}$ and*

$$ {}^R\partial_a^n f(x) = \frac{1}{\Gamma(1-n)} \left(\frac{f(a)}{(x-a)^n} + \int_a^x f'(t)(x-t)^{-n} \mathrm{d}t \right). $$

The proof of the Lemma can be find in [28], p. 27.

Theorem B.3.1. *Let us assume that* $n_1, n_2 \geq 0$. *Moreover let* $\varphi \in L^1[a,b]$ *and* $f = {}^R J_a^{n_1+n_2}\varphi$. *Then,*

$$ {}^R\partial_a^{n_1}\, {}^R\partial_a^{n_2} f = {}^R\partial_a^{n_1+n_2} f. $$

The proof of the Theorem can be find in [28], p. 27.

Theorem B.3.2. *Let* $n \geq 0$. *Then, for every* $f \in L^1[a,b]$,

$$ {}^R\partial_a^{n}\, {}^R J_a^{n} f = f. $$

The proof of the Theorem can be find in [28], p. 30.

Theorem B.3.3. *Let* f_1 *and* f_2 *be two functions defined on* $[a,b]$ *s. t.* ${}^R\partial_a^n f_1$ *and* ${}^R\partial_a^n f_2$ *exist almost everywhere. Moreover, let* $c_1, c_2 \in \mathbb{R}$. *Then,* ${}^R\partial_a^n(c_1 f_1 + c_2 f_2)$ *exists almost everywhere, and*

$$ {}^R\partial_a^n(c_1 f_1 + c_2 f_2) = c_1\, {}^R\partial_a^n f_1 + c_2\, {}^R\partial_a^n f_2. $$

The proof of the Theorem can be find in [28], p. 32. Let us recall the following classical result, known as Leibnitz' formula:

Theorem B.3.4. *Let* $n \in \mathbb{N}$, *and let* $f, g \in C^n[a,b]$. *Then,*

$$ \partial^n[fg] = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) (\partial^k f)(\partial^{n-k} g). $$

Theorem B.3.5. (Leibniz' formula for Riemann-Liouville operators) *Let* $n > 0$, *and assume that* f *and* g *are analytic on* $(a-h, a+h)$ *with some* $h > 0$.*Then,*

$$ {}^R\partial_a^n[fg](x) = \sum_{k=0}^{\lfloor n \rfloor} \left(\begin{array}{c} n \\ k \end{array} \right) ({}^R\partial_a^k f)(x)({}^R\partial_a^{n-k} g)(x) + \sum_{k=\lfloor n \rfloor+1}^{\infty} \left(\begin{array}{c} n \\ k \end{array} \right) ({}^R\partial_a^k f)(x)({}^R J_a^{k-n} g)(x) $$

for $a < x < a + \frac{h}{2}$.

The proof of the Theorem can be find in [28], p. 33.

Lemma B.3.2. *Let* $n > 0$, $n \notin \mathbb{N}$, *and* $m = \lceil n \rceil$. *Assume that* $f \in C^m[a,b]$ *and* $x \in [a,b]$. *Then,*

$$ {}^R\partial_a^n f(x) = \frac{1}{\Gamma(-n)} \int_a^x (x-t)^{-n-1} f(t)\mathrm{d}t. \qquad \text{(B.5)} $$

The proof of the Lemma can be find in [28], p. 38.

Remark B.3.2. *It is important to note that the integrand in* (B.5) *shows a singularity of order* $n+1$ *that is strictly grater than* 1. *So, in general, the integrals exist neither in the proper nor in the improper sense. Therefore it is defined according to Hadamard's finite-part integral concept. See* [28], Appendix D.4.

B.4 Grünwald-Letnikov operators

This Subsection is devoted to present the basic definition and properties of the so-called *Grünwald-Letnikov operators* and their links to the fractional calculus. Such a results are also the basis for some numerical approximations and implementations of fractional PDEs. As done before, let us recall a classical calculus result, remembering that derivatives can be defined as differential quotients, i.e. as limits of difference quotients. Using, for example, backward differences of order n with step size h it is possible to write

$$\Delta_h^n f(x) := \sum_{k=0}^{n} (-1)^k \left(\begin{array}{c} n \\ k \end{array} \right) f(x-kh) \tag{B.6}$$

and the following classical result holds:

Theorem B.4.1. *Let* $n \in \mathbb{N}$, $f \in C^n[a,b]$ *and* $a < x \leq b$.*Then*

$$\partial^n f(x) = \lim_{h \to 0} \frac{\Delta_h^n f(x)}{h^n}.$$

So, let us give the following

Definition B.4.1. *Let* $n > 0$, $f \in C^{\lceil n \rceil}[a,b]$ *and* $a < x \leq b$.*Then*

$$\tilde{\partial}_a^n f(x) = \lim_{N \to \infty} \frac{\Delta_{h_N}^n f(x)}{h_N^n} = \lim_{N \to \infty} \frac{1}{h_N^n} \sum_{k=0}^{N} (-1)^k \left(\begin{array}{c} n \\ k \end{array} \right) f(x - kh_N)$$

with $h_N = \frac{(x-a)}{N}$ *is called the* Grünwald-Letnikov fractional derivative of order n *of the function* f.

Theorem B.4.2. *Let* $n > 0$, $m = \lceil n \rceil$ *and* $f \in C^m[a,b]$. *Then, for* $x \in (a,b]$,

$$\tilde{\partial}_a^n f(x) = {}^R\partial_a^n f(x).$$

The proof of the Theorem can be find in [28], p. 43.

Theorem B.4.3. *Let $n > 0$, $f \in C[a,b]$ and $a \leq x \leq b$.Then,with $h_N = \frac{(x-a)}{N}$, it holds*

$$^{R}J^{n}_{a}f(x) = \lim_{N\to\infty} h^{n}_{N} \sum_{k=0}^{N} (-1)^k \left(\begin{array}{c} -n \\ k \end{array} \right) f(x - kh_N).$$

The proof of the Theorem can be find in [28], p. 45.

Definition B.4.2. *Let $n > 0$, $f \in C[a,b]$ and $a < x \leq b$.Then*

$$\tilde{J}^{n}_{a}f(x) := \frac{1}{\Gamma(n)} \lim_{N\to\infty} h^{n}_{N} \sum_{k=0}^{N} \frac{\Gamma(n+k)}{\Gamma(k+1)} f(x - kh_N)$$

with $h_N = \dfrac{(x-a)}{N}$ is called the Grünwald-Letnikov fractional integral of order n *of the function f.*

For many other details about Grünwald-Letnikov differential and integral operators, see [50], §20.

B.5 Caputo-Djrbashian derivative

In the last decades there has been a rapid increase of works, especially related to the theory of viscoelasticity and in the field of hereditary solid mechanics, in which the theory of fractional derivatives has been used to better describe the observed phenomena and the properties of the studied materials. The resulting mathematical modeling essentially leads to fractional PDEs which need suitable initial conditions to be solved. In other words, applied mathematics problems require definitions of fractional derivatives that allow to have physically interpretable initial conditions, i.e. containing values such as $f(a)$, $f'(a)$, etc. However, Riemann-Liouville type fractional derivatives do not have this approach, since they lead to initial conditions containing the limit values of their fractional derivatives at the lower terminal $x = a$.

A possible solution to this problem was given by M. Caputo, see [22, 23]. Caputo's derivative definition can be written as

$$^{C}_{a}\partial^{\alpha}_{x}f(x) = \frac{1}{\Gamma(\alpha - n)} \int_{a}^{x} \frac{f^{(n)}(y)}{(x-y)^{\alpha+1-n}} \mathrm{d}y.$$

Clearly, under smoothness natural assumptions about the function $f(x)$, as $\alpha \to n$, the Caputo derivative coincides with the classical n-th derivative of the same function, in fact, for all $n \geq 1$, we have:

$$\begin{aligned} {}_a^C\partial_x^\alpha f(x) &= \lim_{\alpha \to n^-} \left(\frac{f^{(n)}(a)(x-a)^{n-\alpha}}{\Gamma(n-\alpha+1)} + \right. \\ &\left. + \frac{1}{\Gamma(n-\alpha+1)} \int_a^x \frac{f^{(n+1)}(y)}{(x-y)^{\alpha-n}} \mathrm{d}y \right) = f^{(n)}(a) + \int_a^x f^{(n+1)}(y) \mathrm{d}y. \end{aligned}$$

This statement suggests that also for the Caputo derivative, as well as in the Riemann-Liouville and Gründwall-Letnivkov approaches, the interpolation between derivatives of integer order still holds. For a more detailed discussion on this topic, see [48], pp. 79-80.

Moreover, now referring to the results presented in Chapter 3 of this thesis, it is used the following definition of the Caputo derivative, that states setting $n = 1$, so $\alpha \in (0, 1)$, and $a = 0$:

$$\partial_x^\alpha f(x) = \frac{1}{\Gamma(1-\alpha)} \int_0^x \frac{f'(y)}{(x-y)^\alpha} \mathrm{d}y.$$

Finally, also for $n - 1 < \alpha < n$, the following relation holds true:

$$\partial_x^\alpha u(x) = J_x^{\alpha-n} \frac{\mathrm{d}^n}{\mathrm{d}x^n} u(x).$$

Let us conclude this Section recalling the following basic relations that also hold true about the Caputo derivative:

1. $\partial_x^\alpha \partial_x^\beta \neq \partial_x^{\alpha+\beta}$;

2. $\partial_x^\alpha C = 0$ when C is an arbitrary constant;

3. $\partial_x^\alpha f \to f'$, as $\alpha \uparrow 1$;

4. $\partial_x^\alpha f \to f' - f'(0)$ as $\alpha \downarrow 1$ (only left-continuous);

5. $\partial_x^\alpha f \to f$ as $\alpha \to 0$.

The following theorem holds:

Theorem B.5.1. *Given $\alpha \in (0,1)$, the unique solution to*

$$\partial_x^\alpha f(x) = cf(x), \quad x \in (0,\infty), \quad c \in \mathbb{R}, \quad f(0) = 1$$

is the Mittag-Leffler function

$$f(x) = E_\alpha(cx^\alpha) = \sum_{k\geq 0} \frac{(cx^\alpha)^k}{\Gamma(\alpha k + 1)}.$$

The Mittag-Leffler function is defined in the following Subsection, as a particular case of Bernstein functions.

The proof can be found in [30].

Remark B.5.1. *Setting ${}^R\partial_x^\alpha u$ the Riemann-Liuoville derivative and $\partial_x^\alpha u$ the Caputo derivative, as an immediate consequence of Theorem* B.3.3*, through the definition of Riemann-Liuoville and Caputo type derivatives, the following relation between this two kind of derivatives holds:*

$$\partial_x^\alpha f = {}^R\partial_x^\alpha (f - f(0)) = {}^R\partial_x^\alpha f - \frac{x^{-\alpha}}{\Gamma(1-\alpha)} f(0)$$

where, for the constant function $\mathbf{1}(x) = 1, x \in \mathbb{R}$,

$${}^R\partial_x^\alpha \mathbf{1}(x) = \frac{x^{-\alpha}}{\Gamma(1-\alpha)}.$$

This result is also successfully used in the numerical implementation of the Caputo derivative.

B.6 Bernstein functions

A function $\Phi : (0,\infty) \mapsto \mathbb{R}$ is a Bernstein function if

1. $\Phi(z) \geq 0$ for all $z > 0$;

2. $(-1)^k \frac{\mathrm{d}^k \Phi}{\mathrm{d}z^k}(z) \leq 0, \qquad$ for all $k \in \mathbb{N}, z > 0$.

Through the Bernstein's representation theorem, the function Φ is a Bernstein function if and only if Φ admits the following representation

$$\Phi(z) = a + bz + \int_0^\infty (1 - e^{-zy})\varphi(\mathrm{d}y)$$

where $a, b \geq 0$ and φ on $(0, \infty)$ is a measure satisfyng

$$\int_0^\infty (1 \wedge y)\varphi(\mathrm{d}y) < \infty.$$

We also underline the fact that the composition of two or more Bernstein functions (i.e. Φ_1, Φ_2, $\dots$) is still a Bernstein function. Moreover, considering the case $a = 0$ and $b = 0$ just only for simplicity, some examples of Bernstein function are:

1. $\Phi(z) = z^\alpha, \varphi(\mathrm{d}y) = \dfrac{1}{\Gamma(1-\alpha)} y^{-\alpha-1}\mathrm{d}y;$
2. $\Phi(z) = (\eta + z)^\alpha - \eta^\alpha, \varphi(z) = \dfrac{\alpha}{\Gamma(1-\alpha)} y^{-\alpha-1}e^{-\eta y}$, $\eta \geq 0$;
3. $\Phi(z) = \ln(1 + z^\alpha), \varphi(\mathrm{d}y) = \alpha y^{-1} E_\alpha(-y)$ where

$$E_\alpha(-y) = \sum_{k \geq 0} \frac{(-y)^k}{\Gamma(\alpha k + 1)}$$

is the Mittag-Leffler function.

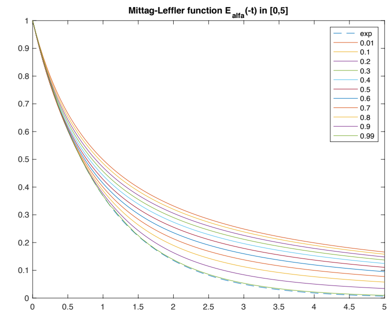


Figure B.1: The Mittag-Leffler fun-ction as a function of α.

B.7 A simple case

Behind the degenerate diffusion model studied in the Chapter 3 of this thesis, it has been studied first, at a numerical level, the behavior of the solutions of the more simple case of a Cauchy problem with time fractional (in the sense of Caputo) heat equation in a bounded domain Ω, when the parameter $\alpha \in (0, 1)$ changes. More precisely, we numerically studied the following problem:

$$\begin{cases} \partial_t^\alpha u - \Delta u = 0 & \text{in } \Omega \times (0,T) \\ u(x,0) = u^0(x) & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \times (0,T) \end{cases} \tag{B.7}$$

where Ω is a bounded domain of $\mathbb{R}^2$, $u^0 \in H_0^1(\Omega)$ is the (positive) initial data (which satisfies the Dirichlet conditions at the edge), $\alpha \in (0,1)$ and the fractional temporal derivative is given, as said, in the sense of Caputo.

One of the reasons of this preliminary study was first to test, from a numerical point of view, the finite difference schemes proposed in [33] and in [38] which give back the known results of the classic heat equation with obstacle for $\alpha \to 1^-$.

After this numerical study, below reported, it was also possible to verify the regularity with which the solution of the fractional heat equation evolved up to 0 over time, as $\alpha \in (0,1)$, finding analogue results as in [35]. However, on the other hand, this problem could be trivially considered as an obstacle problem in which the obstacle function could be considered identically zero, that is coincident with the same stationary solution of the problem itself. Of course, referring to this point of view, assumptions $\mathbf{H}_1$ and $\mathbf{H}_2$, mentioned in chapter 2, become obvious and trivially satisfied. Thus, under these hypothesis, the evolution of the solution of the fractional heat equation also seemed to verify an estimate like Theorem 2.1.1 type.

All these considerations have led to the possibility that, even in the fractional case with Caputo's time fractional derivative, the degenerate problem studied in Chapter 2 could have some connection with the corresponding time fractional one, and the related time fractional obstacle problem, thus generalizing the results obtained in Chapter 2 not only for $\alpha = 1$, but for all $\alpha \in (0,1)$.

Turning back to the fractional heat equation, in order to study the corresponding behavior of the solution, for simplicity, we started from the one-dimensional case, with $\Omega = (0,1)$ and $u^0(x) = \sin(\pi x)$.

It is well known, see [45], that

$$u(x,t) = \sum_{k \geq 0} E_\alpha(-k^2\pi^2 t^\alpha) \, \sin(k\pi x) \, u_{0,k}$$

is the absolutely and uniformly convergent solution of the fractionl Cauchy problem (B.7) (with $u_0 \in D(\Delta_{Dirichlet})$) where

$$u_{0,k} = \int_0^\pi u_0(x)\, \sin(k\pi x)\, dx, \quad k \in \mathbb{N}$$

and the Mittag-Leffler function can be written as follows

$$E_\alpha(-k^2\pi^2 t^\alpha) = \sum_{i\geq 0} \frac{(-k^2\pi^2 t^\alpha)^i}{\Gamma(\alpha i + 1)}.$$

Remark B.7.1. *Referring in particular to* [50]*, given* $\alpha \in (0,1)$*, we observe that*

$$E_\alpha(0) = 1, \qquad 0 \leq E_\alpha(-z^\alpha) \leq \frac{1}{1+z^\alpha},\ z \geq 0;$$

$$E_\alpha(-z^\alpha) \approx \exp -\frac{z^\alpha}{\Gamma(1+\alpha)} \approx 1 - \frac{z^\alpha}{\Gamma(1+\alpha)}, \quad z \ll 1;$$

$$E_\alpha(-z^\alpha) \approx \frac{z^{-\alpha}}{\Gamma(1-\alpha)} - \frac{z^{-2\alpha}}{\Gamma(1-2\alpha)} \ldots, \qquad z \gg 1.$$

Considering the initial condition $g(x) = \sin(\pi x)$, thus

$$g_k = \begin{cases} 1, & k = 1 \\ 0, & k \neq 1 \end{cases}$$

and the solution of the problem (B.7) becomes:

$$u(x,t) = E_\alpha(-\pi^2 t^\alpha) \sin(\pi x). \tag{B.8}$$

Of course, the solution will have a limited time regularity for every α, as for $t \to 0$ $E_\alpha(-\pi^2 t^\alpha) \simeq 1 - \frac{\pi^2}{\Gamma(\alpha+1)} t^\alpha$, which is continuous for $t = 0$ but has first derivative unbounded.

B.7.1 Numerical approximation

The following numerical schemes are given for $\alpha \in (0,1)$.

- **Scheme (S1)**, from [33, 38]:

  If $\tau > 0$ indicates the time discretization step, from the definition of the fractional Caputo derivative, dividing the integral into subintervals of amplitude τ and approximating on each of them the derivative with the incremental ratio, one gets (for calculations see [33, 38]):

$$\partial_t^\alpha u(x, m\tau) \simeq \frac{1}{\Gamma(2-\alpha)\tau^\alpha} \left\{ u(x, m\tau) - \sum_{k=0}^{m-1} C_{m,k} u(x, k\tau) \right\}, \tag{B.9}$$

  where $k = 1, ..., m-1$, and

$$C_{m,0} := f(m),\ C_{m,k} := f(m-k) - f(m-(k-1)),\ f(r) := r^{1-\alpha} - (r-1)^{1-\alpha}. \tag{B.10}$$

  Since it is $C_{m,k} \geq 0$ for every k, it follows that the pattern will be monotonous. We then define by induction a sequence of functions $\{U^\tau(., M\tau)\}_{m\in\mathbb{N}\cup\{0\}}$; be

$$U^\tau(.,0) := u_0^\tau, \quad \text{with} \quad \sup_\Omega |u_0^\tau - u_0| \to 0 \text{ as } \tau \to 0.$$

  We then define $U^\tau(., m\tau) \in C(\Omega)$ for $m \geq 1$ as the solution of:

$$\frac{1}{\Gamma(2-\alpha)\tau^\alpha} \left\{ u(x) - \sum_{k=0}^{m-1} C_{m,k} U^\tau(x, k\tau) \right\} - u_{xx} = 0 \ \text{ in } \Omega \tag{B.11}$$

  which meets the conditions at the edge. Then the piecewise linear solution in time is defined by

$$u^\tau(x,t) := U^\tau(x, m\tau) \text{ for each } x \in \Omega,\ t \in [m\tau, (m+1)\tau),\ m \in \mathbb{N} \cup \{0\}.$$

  To obtain a complete numerical solution it is therefore necessary to discretize the space (3.16). Let us introduce the partition of $\Omega = (c,d)$ into N subintervals by means of the nodes $x_i = c + ih$, $i = 0, 1, .., N$ (with $h = (d-c)/N$ spatial discretization step). The initial data will then be the vector $u_i^0 = u_0(x_i)$ of the values on the mesh; then using an implicit scheme and the second finite differences centered for the second derivative, the solution u^1 at the first time instant $h = T/M$ will solve the relationship in each node:

$$\frac{1}{g\tau^\alpha}\left(u_i^1 - C_{1,0}u_i^0\right) = \delta^2 u_i^1 := \frac{u_{i-1}^1 - 2u_i^1 + u_{i+1}^1}{h^2}$$

where $g = \Gamma(2-\alpha)$ is placed. Redistributing the terms between first and second member, this is equivalent to

$$u_i^1 + g\frac{\tau^\alpha}{h^2}(-u_{i-1}^1 + 2u_i^1 - u_{i+1}^1) = C_{1,0}u_i^0 \quad \text{for any} \quad i;$$

with vector notations is equivalent to saying that u^1 solves the linear system

$$Bu^1 := (I + g\gamma A)u^1 = C_{1,0}u_i^0,$$

where $\gamma = \frac{\tau^\alpha}{h^2}$, while A is the tridiagonal matrix $(N-1) \times (N-1)$, symmetric and positive defined associated to the discrete Laplacian, with values (2 on the main diagonal and -1 above and below diagonal. Similarly, u^2 solves

$$\frac{1}{g\tau^\alpha}\left(u_i^2 - C_{2,0}u_i^0 - C_{2,1}u_i^1\right) = \delta^2 u_i^2,$$

i.e. the system

$$Bu^2 = C_{2,0}u^0 + C_{2,1}u^1,$$

with the same B matrix. In general, at each time step u^m solves

$$Bu^m = b^m := \sum_{k=0}^{m-1} C_{m,k}u^k. \tag{B.12}$$

- **Scheme (S2)**: convolution quadrature type, from [38]

In this case, the Caputo derivative is approximated with a numerical time-stepping method, by means of *convolution quadrature.* This method was proposed by Lubich, as before mentioned, for the discretization of the Volterra integral equation. In particular, if we consider the Riemann-Liouville derivative

$${}^R D_x^\alpha f := \frac{1}{\Gamma(1-\alpha)}\frac{\mathrm{d}}{\mathrm{d}x}\int_0^x (x-s)^{-\alpha} f(s)\mathrm{d}s,$$

(with $f(0) = 0$), this can be approximated by the discrete convolution:

$$D_\tau^\alpha f^m := \frac{1}{\tau^\alpha} \sum_{j=0}^{m} b_j f^{m-j},$$

where $f^m = f(t_m)$, and the coefficients $\{b_j\}$ are obtained from an appropriate expansion in power series, linked to a specific method of approximation of ODE (for more details, see [38]). The special case related to the backward Euler method, also known as the Grunwald-Letnikov approximation, gives the recurrence relationshi

$$b_0 = 1, \quad b_j = -\frac{\alpha - j + 1}{j} b_{j-1}.$$

Then, using the relationship between the Caputo derivative and the Riemann-Liouville derivative, see Remark (B.5.1), one gets:

$$\partial_t^\alpha \varphi(t) = {}^R\partial_t^\alpha(\varphi(t) - \varphi(0)),$$

and it is possible to rewrite the initial problem as

$${}^R\partial_t^\alpha(u - u^0) - \Delta u = 0,$$

that discretized in time, is equivalent to find U^m, approximation of $u(t_m)$, which solves:

$$\partial_\tau^\alpha(U - u^0)^m - \Delta U^m = 0, \quad m = 1, 2, ..., M; \quad U^0 = u^0.$$

By adding the discretization in space with the same notations as in the previous case, we obtain:

$$\frac{1}{\tau^\alpha} \sum_{j=0}^{m} b_j(u^{m-j} - u^0) + \frac{1}{h^2} A u^m = 0.$$

Rearranging the terms (and remembering that $b_0 = 1$ and $U^0 = u^0$), this is equivalent to solving each step the linear system:

$$Qu^m := (I + \gamma A)u^m = q^m := u^0 - \sum_{j=1}^{m-1} b_j(u^{m-j} - u^0). \tag{B.13}$$

B.7.2 Numerical tests

It is worth noting that since the scheme is implicit, there seem to be no stability problems, so any choice of steps (any ratio $\gamma = \tau^\alpha/h^2$ provides reliable results), except to reduce the accuracy of the result (see next table). For the model problem, the exact solution (B.8) is calculated by means of the function Matlab mlf.m which calculates the Mittag-Leffler function with the desired precision (we assumed 10^{-10}). In Fig (B.2) it is possible to see the progress of the numerical solution in the central point $x = 0.5$ over time for a set of values of α between 0.01 and 0.99. In particular, the last graph clearly shows how steep the solution is at the initial moment, especially for small values of α.

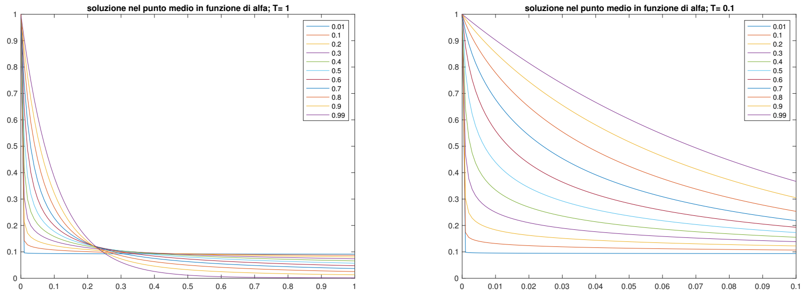


Figure B.2: Solution at the midpoint as a function of α and the final time $T = 1$ (left) and $T = 0.1$ (right).

In order to give a possible estimate of the error of the method in function of N and M, calculating the relative error to the final time $T = 1$ in the norm L^∞ between the numerical solution u and the exact one sol in the case $\alpha = 0.5$, i.e.:

$$err = \frac{\|u(T) - sol(T)\|_\infty}{\|sol(T)\|_\infty}.$$

From the simulations it is clear that the error obviously falls to the trend of τ and h at zero; but also that the error decreases with increasing time: in other words it is maximum at the first iteration and then it drops, in coherence with the lack of regularity of the Mittag Leffler function for $t = 0$. To reduce this error you need a very small τ step, and this proves to be a significant computational cost by using uniform time grids. It therefore seems reasonable to use adaptive time grids, with an initially small step that then grows as time increases.

From the Table B.1 it can also be seen that for a sufficient number of spatial nodes the error decreases linearly as a function of τ. E. g. for $N = 100$ the error is halved with each halving of the time step, in line with the theoretical order of the difference method.

Table B.1: Relative error for the solution of the problem (B.7) for $\alpha = 0.5$ and $T = 1$, in function of N and M, scheme $S1$.

N	M	τ/h^2	err
20	80	5	0.0052
	200	2	0.0033
	400	1	0.0027
	800	0.5	0.0023
40	200	8	0.0018
	400	4	0.0011
100	100	100	0.0026
	200	50	0.0013
	400	25	0.0007
	800	12.5	0.00039
200	200	200	0.0013
	400	100	0.00064
	800	50	0.00033

Short comparison between the two schemes

Both schemes, proposed in [33] and [38], work in the sense that, from the first iterations, graphically they provide solutions that can be superimposed on the exact one (see Fig. B.3), with a very slight greater precision for the scheme proposed in [33] and as γ decreases (see Table B.2). At last both schemes show a continuity for $\alpha \to 1$.

Table B.2: Error for problem (B.7) for $N = 32$, $T = 0.3$, in function of α e γ, schemes $S1$ and $S2$..

α	γ	err_{S1}	err_{S2}
0.001	1017	0.006	0.006
0.1	513	0.00087	0.00088
0.5	32	0.0016	0.0019
0.5	10	0.00086	0.00089
0.9	2	0.0064	0.0082
1	3	1	0.046

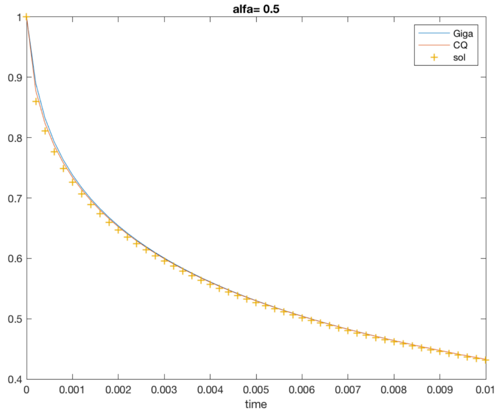


Figure B.3: Exact solution and numerical ones in the middle point for $\alpha = 0.5$ and $T = 0.01$.

Acknowledgements

I wish to thank Professors Maria Agostina Vivaldi, Raffaela Capitanelli, Mirko D'Ovidio and Stefano Finzi Vita for their helpful scientific support during these three years, and for giving me the opportunity to reach this important university goal.

Bibliography

[1] G. Akagi, and M. Kimura, *Unidirectional evolution equations of diffusion type*, J. Differential Equations, 266 (2019), 1, 1–43.

[2] C. Alberini and S. Finzi Vita, *A numerical approach to a nonlinear diffusion model for self-organized criticality phenomena*, (ICIAM 2019), FRACTALS (Fractals in engineering: Theoretical aspects and Numerical approximation), M. Lancia and A. Rozanova, eds, (2020), to appear.

[3] C. Alberini, R. Capitanelli and S. Finzi Vita, *A numerical study of an Heaviside function driven degenerate diffusion equation*, preprint 2020, submitted.

[4] C. Alberini, R. Capitanelli, M. D'Ovidio and S. Finzi Vita, *On the time fractional heat equation with obstacle*, preprint 2020, submitted.

[5] P. Bak and M. Creutz, *Fractals and self-organized criticality*, in Fractals in Science, A. Bunde and S. Havlin, eds., Springer-Verlag, Berlin, Heidelberg, (1994), 27–48.

[6] P. Bak, C. Tang, and K. Wiesenfeld, *Self-organized criticality: An explanation of the 1/f noise*, Phys. Rev. Lett., 59 (1987), 381–394.

[7] P. Bak, C. Tang. and K. Wiesenfeld, *Self-organized criticality*, (1988) Phys. Rev. A (3), 38, 364–374.

[8] P. Bántay and M. Jánosi, *Self-organization and anomalous diffusion*, (1992) Phys. A, 185, 11–18.

[9] V. Barbu, *Nonlinear semigroups and differential equations in Banach spaces*, 1976, Noordhoff: Leyden.

[10] V. Barbu, *Nonlinear Differential Equations of Monotone Type in Banach Spaces*, 2010, Springer, New York.

[11] V. Barbu, *Self-organized criticality and convergence to equilibrium of solutions to nonlinear diffusion equations*, (2010) Annual Reviews in Control 34, pp. 52–61.

[12] V. Barbu, *Self-organized criticality of cellular automata model; absorbtion in finite-time of supercritical region into the critical one*, (2013) Math. Methods Appl. Sci., 36, pp. 1726–1733.

[13] E.G. Bazhlekova, *Subordination principle for fractional evolution equations*, (2000), Fractional Calculus and Applied Analysis, 3, 3, 213–230.

[14] A. Bensoussan, J. Frehse and U. Mosco, *A stochastic impulse control problem with quadratic growth Hamiltonian and the coresponding quasi variationalin inequality*, (1982) J. Reine Angew. Math. 331, 124–145.

[15] N. Bourbaki, *Elements of the history of mathematics*, 1994, Springer.

[16] H. Brezis, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, 1971, in E. Zarantonello (Ed.), *Contributions to nonlinear functional analysis. New York: Academic Press.*.

[17] H. Brezis, *Problemes unilateraux*, (1972), J. Math. Pures Appl. 51, 9, 1–168.

[18] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, 1973, North-Holland: Amsterdam.

[19] H. Brezis, *Analyse Fonctionnelle. Théorie et Applications.*, 1983, Paris: Masson.

[20] L. Brugnano and A. Sestini, *Iterative solution of piecewise linear systems for the numerical solution of obstacle problems*, (2011), J. Num. Anal. Ind. Appl. Math. (JNAIAM) 6, 3-4, 67–82.

[21] R. Capitanelli and M. D'Ovidio, *Fractional equations via convergence of forms*, (2019) Fract. Calc. Appl. Anal., vol. 22, 4, 844–870.

[22] M. Caputo, *Linear model of dissipation whose Q is almost frequency independet* - II, (1967) Geophys. J. R. Astr. Soc., vol. 13, 529–539.

[23] M. Caputo, *Elasticità e dissipazione*, 1969, Zanichelli, Bologna.

[24] J.M. Carlson and G. H. Swindle, *Self-organized criticality: Sandpiles, singularities, and scaling*, (1995) Proc. Natl. Acad. Sci. USA, 92, 6712–6719.

[25] P. Charrier and G.M. Troianiello, *On strong solutions of parabolic unilateral problems with obstacle dependent on time*, (1978), J. Math. Anal. Appl. 65, 110–125.

[26] M. Colombo, L. Spolaor and B. Velichkov, *On the asymptotical behavior of the solutions to parabolic variational inequalities*, (2018) arXiv:1809.06075v1.

[27] D. Dahr, *Self-organized critical state of sandpile automaton models*, (1990) Phys. Rev. Lett., 64, 1613.

[28] K. Diethelm, *The analysis of fractional differential equations. An application-oriented exposition*, 2010, Springer.

[29] G. Doetsch, *Anleitung zum praktischen Gebrauch der Laplace-Transformation und der Z-Transformation*, 1989, 6^{th} edn. Oldenbourg, München

[30] M. D'Ovidio, *Fractional Calculus and Singular Equations*, (2019) SBAI Department internal note for PhD courses.

[31] M. D'Ovidio, *On the fractional counterpart of the higher-order equations*, (2011), Statistics and Probability Letters, 81, 1929 –1939.

[32] A. Friedman, *Variational principles and free-boundary problems*, 1982, John Wiley & Sons.

[33] Y. Giga, Q. Liu and H. Mitake, *On a discrete scheme for time fractional fully nonlinear evolution equations*, (2020) Asymptotic Analysis, 120, 1-2, 151–162.

[34] R. Gorenflo, A.A. Kilbas, F. Mainardi, S.V. Rogosin, *Mittag-Leffler Functions, Related Topics and Applications*, 2014, Springer Monographs in Mathematics, Springer-Verlag Berlin Heidelberg.

[35] R. Gorenflo and F. Mainardi, *Fractional Calculus: Integral and Differential Equations of Fractional Order*, (2008) arXiv:0805.3823v1.

[36] L. Hue and X.L. Cheng, *An algorithm for solving the obstacle problem*, (2004) Computers and Mathematics with Applications, 48, 1651–1657.

[37] S. Ion and G. Marinoschi, *A self-organizing criticality mathematical model for contamination and epidemic spreading*, (2017) Discrete and continuous dynamical systems, Series B, Vol. 22, No. 2, pp. 383–405.

[38] B. Jin, R. Lazarov and Z. Zhou, *Numerical methods for time-fractional evolution equations with nonsmooth data: a concise overview*, (2019) Comput. Methods Appl. Mech. Engrg. 346, 332-358.

[39] A. N. Kochubei, *The Cauchy problem for evolution equations of fractional order*, (1989) Differential Equations, 25, 967–974.

[40] A. Lo Giudice, G. Giammanco, D. Fransos and L. Preziosi, *Modelling Sand Slides by a Mechanics-Based Degenerate Parabolic Equation*, (2019), Mathematics and Mechanics of Solids 24, 8, 2558–2575.

[41] F. Mainardi, Y. Luchko and G. Pagnini, *The fundamental solution of the space-time fractional diffusion equation.* (2001), Fract. Calc. Appl. Anal. 4, 2, 153–192.

[42] F. Mainardi, A. Mura and G. Pagnini, *The M-Wright function in time- fractional diffusion processes: A tutorial survey*, 2010, Int. J. Differ. Equ.

[43] F. Mainardi, G. Pagnini and R. Gorenflo, *Some aspects of fractional diffusion equations of single and distributed order.* (2007) Appl. Math. and Computing 187, 295–305.

[44] F. Mainardi, G. Pagnini and R.K. Saxena, *Fox H-functions in fractional diffusion.* (2005) J. of Comput. and Appl. Math. 178, 321–331.

[45] M.M. Meerschaert, E. Nane and P. Vellaisamy, *Fractional Cauchy Problems on Bounded Domains* (2009) Ann. Probab. 37, 3, 979–1007.

[46] U. Mosco, *Finite-time Self-Organized-Criticality on synchronized infinite grids*, (2018) SIAM J. Math. Anal., 50, 3, 2409–2440.

[47] U. Mosco and M.A. Vivaldi, *On a discrete self-organized-criticality finite time result*, (2020), Discrete and Continuous Dynamical Systems, 40, 8, 5079–5103.

[48] I. Podlubny, *Fractional Differential Equations*, 1999, Academic Press.

[49] F. Riesz, B. Sz.-Nagy, B, *Vorlesungen über Funktionalanalysis*, 1956, Deutscher Verlag der Wissenschaften, Berlin.

[50] S.G. Samko, A.A. Kilbas, O.I. Marikev, *Fractional integrals and derivatives*, 1993, Gordon and Breach Science Publishers.

[51] A.H. Siddiqi, P. Manchanda, R. Bhardwaj, *Mathematical models, methods and applications*, 2015, Industrial and Applied Mathematics, Springer.

[52] G. Stampacchia, *Le problème de Dirichlet pour les équations du 2^{nd} ordre à coefficients discontinus*, (1965) Ann. Inst. Fourier, 15, 189.