ACAP manuscript No.
(will be inserted by the editor)

A new nonlocal nonlinear diffusion equation for data analysis

Giacomo Aletti · **Monica Moroni** · **Giovanni Naldi**

Received: date / Accepted: date

Abstract In this paper we introduce and study a new feature-preserving nonlinear nonlocal diffusion equation for denoising signals. The proposed partial differential equation is based on a novel diffusivity coefficient that uses a nonlocal automatically detected parameter related to the local bounded variation and the local oscillating pattern of the noisy input signal. We provide a mathematical analysis of the existence of the solution in the two dimensional case, but easily extensible to the one-dimensional model. Finally, we show some numerical experiments, which demonstrate the effectiveness of the new approach.

Keywords nonlocal operators · nonlinear diffusion · signal processing

1 Introduction

Nonlinear partial differential equations (PDEs) can be used in the analysis and processing of digital images or image sequences, for example to filter out the noise, to produce higher quality image, to extract features and shapes (see e.g. [2,4,3,18,19] and the references herein). Perhaps, the main application of PDEs based methods in this field is smoothing and restoration of images. From the mathematical point of view, the input (gray scale) image can be modeled by a real function $u_0(x)$, $u_0 : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^d$, represents the spatial domain. Typically this domain Ω is rectangular and $d = 1,\ 2,\ 3$. The function u_0 is considered as an initial data for a suitable evolution equation with some kind of boundary conditions (usually homogeneous Neumann boundary conditions).

The simplest PDE method for smoothing images is based on linear diffusion process. The starting point is the simple observation that the so called Gauss function, with $\sigma > 0$ and where $|\cdot|$ is the Euclidean norm,

$$G_\sigma(x) = \frac{1}{(2\pi\sigma^2)^{d/2}}\, e^{-|x|^2/(2\sigma^2)}$$

This work was funded by ADAMSS Center, G. Naldi acknowledges the support of the Hausdorff Research Institute for Mathematics during the Special Trimester "Mathematics of Signal Processing". G.Aletti and G.Naldi are members of "Gruppo Nazionale per il Calcolo Scientifico (GNCS)" of the INDAM

G. Aletti
Department of Environmental Science and Policy & ADAMSS Center, Università di Milano, via Celoria 2 Milano, 20133, Italy
E-mail: giacomo.aletti@unimi.it

M. Moroni
IIT Genova, Via Morego, 30 Genova, 16163 Italy
E-mail: monica.moroni@iit.it

G. Naldi
Department of Environmental Science and Policy & ADAMSS Center, Università di Milano, via Celoria 2 Milano, 20133, Italy
E-mail: giovanni.naldi@unimi.it

is related to the fundamental solution of the linear diffusion (heat) equation. Then, it has been possible to reinterpret the classic smoothing operation of the convolution of an image with G_σ, with a given standard deviation σ, by solving the linear diffusion equation for a corresponding time $t = \sigma^2/2$ with initial condition given by the original image. For example, when $d = 2$, it is a classic result that for any bounded, continuous. and integrable $u_0(x)$, $x \in \mathbb{R}^2$, the linear diffusion equation on the whole space (here $\triangle$ denotes the Laplacian operator),

$$\frac{\partial u}{\partial t} = \triangle u,\ u(x,0) = u_0(x)$$

possesses the following solution

$$u(x,t) = \begin{cases} u_0(x),\ t = 0 \\ \\ (G_{\sqrt{2t}} * u_0)(x), t > 0 \end{cases}$$

where the convolution product $(g * f)(x)$ between the functions f and g is defined as

$$(g * f)(x) = \int_{\mathbb{R}^2} g(x-y) f(y) dy.$$

We point out that for different time (variance) t we obtain different levels of smoothing: this defines a *scale-space* for the image [12,20]. That is, we get copies of the image at different scales. Note, of course, that any scale t can be obtained from a scale τ, where $\tau < t$, as well as from the original images, this is usually denoted as the causality criteria for scale-spaces [2]. The solution of the above linear diffusion equation is unique, provided we restrict ourselves to functions satisfying some suitable grow conditions. Moreover, it depends continuously on the initial image u_0, and it fulfills the maximum/minimum principle

$$\inf_{x\in\mathbb{R}^2} u_0(x) \leq u(x,t) \leq \sup_{x\in\mathbb{R}^2} u_0(x) \text{ on } \mathbb{R}^2 \times [0,\infty).$$

In order to apply it to images processing we also need to consider appropriate boundary conditions: usually homogeneous Neumann conditions are used.

The flow produced by the linear diffusion equation is also denoted as isotropic diffusion, as it is diffusing the information equally in all directions. Then, the gray values of the initial image will spread, and, in the end, a uniform image, equal to the average of the initial gray values, is obtained. Although this property is good for local reducing noise (averaging is optimal for additive noise), this filtering operation also destroys the image content, that is, the boundaries of the objects and the subregions present in the image (the *edges*). This means that the Gaussian smoothing does not only smooth noise, but also blurs important features and it makes them harder to identify. Furthermore, linear diffusion filtering dislocates edges when moving from finer to coarser scales (see e.g. [20]). So, structures that are identified at a coarse scale do not give the right location and have to be traced back to the original image. Moreover, some smoothing properties of Gaussian scale-space do not carry over from the one-dimensional case to higher dimensions: it is generally not true that the number of local extrema, which are related to edges, is non-increasing. As suggested by Hummel [10] the linear diffusion is not the only PDE that can be used to enhance an image and that, in order to keep the scale-space property, we need only to make sure that the corresponding flow holds the maximum principle. Many approaches have been taken in the literature to implement this idea replacing the linear equation with a nonlinear PDE that does not diffuse the image in a uniform way: these flows are normally denoted as anisotropic diffusion. In particular, the diffusion coefficient is locally adapted, becoming negligible as object boundaries are approached. Noise is efficiently removed and object contours are strongly enhanced [19]. There is a vast literature concerning nonlinear anisotropic diffusions with application to image processing, which dates back to the seminal paper by Perona and Malik, who, in [16], consider a discrete version of the following equation

$$\begin{cases} \frac{\partial u}{\partial t} - \nabla \cdot (g(|\nabla u|)\nabla u) = 0, \text{ in } \ \Omega_T = (0,T) \times \Omega, \\ \\ u(x,0) = u_0(x) \qquad\qquad \text{on } \Omega \\ \\ \frac{\partial u}{\partial \mathbf{n}}(x,t) = 0, \qquad\qquad \text{on } \Gamma \times (0,T), \end{cases}$$

where $\Gamma = \partial\Omega$, the image domain $\Omega \subset \mathbb{R}^2$ is an open regular set (typically a rectangle), $\mathbf{n}$ denotes the unit outer normal to its boundary Γ, and $u(x,t)$ denotes the (scalar) image analysed at time (scale) t and point x. The initial condition $u_0(x)$ is, as in the linear case, the original image. In order to reduce smoothing at edges, the diffusivity g is chosen as a decreasing function of the "edge detector" $|\nabla u|$ (for a vector $V = (V_1, V_2) \in \mathbb{R}^2$, $|V|^2 = V_1^2 + V_2^2$). A typical choice is,

$$g(s) = \frac{1}{1+(s/\lambda)^2},\ s \geq 0,\ \lambda > 0.$$

Catté, Lions, Morel and Coll [8] showed that the continuous Perona-Malik model is ill posed, and hence very close pictures can produce divergent solutions and therefore very different edges. This is caused by the fact that the diffusivity g leads to flux $s \cdot g(s)$ decreasing for some s and the scheme may work locally like the inverse heat equation which is known to be ill posed. This possible misbehavior surely represents a severe drawback of the Perona-Malik model when applied to data affected by noise. However, discrete implementations work as a regularization factor by introducing implicit diffusion into the model, and the filter is usually observed to be stable (with staircasing effect as the only observable instability). Then, in the continuous settings, a new model has been proposed [8] with the only modification of replacing the gradient ∇u in the diffusivity by its spatial regularization $(G_\sigma * \nabla u)$, which is obtained by smoothing the argument by a convolution with a C^∞ kernel G_σ. Typically G_σ is a Gaussian function and σ determines the scale beyond which regularization occurs. The equation will now diffuse if and only if the gradient is estimated to be small. We point out that the spatial regularization leads to processes where the solution converges to a constant steady state. Then, in order to get nontrivial results, we have to specify a stopping time T. Sometimes it is attempted to circumvent this task by adding an additional reaction term, which keeps the steady state solution close to the original image u_0, for example

$$\frac{\partial u}{\partial t} - \nabla \cdot (g(|\nabla G_\sigma * u|)\nabla u) = f(u_0 - u),$$

where f is a Lipshitz continuous, non decreasing function such that $f(0) = 0$. During the last years, many other nonlinear parabolic equations have been proposed as an image analysis model. The common theme in this proliferation of models is the following, one attempts to fix one intrinsic diffusion direction and tunes the diffusion using the size of the gradient or the value of an estimate of the gradient. A few of the proposed models are even systems of PDEs, for example there exist reaction diffusion systems which have been applied to image restoration and which are connected to Perona-Malik idea or based on Turing's pattern formation model [19].

From the numerical point of view the approaches seen above for denoising images can be considered as a suitable spatial averaging of nearby values. However, although this method removes noise it creates blur. In order to improve their effectiveness neighborhood filters have recently been considered. These last filters perform an average of neighboring values of a signal, but only under the condition that these values are close enough to the one of the original in restoration. An example of this type of algorithm in a continuous form is the following [6],

$$u_F(\mathbf{x}) = \frac{1}{c(\mathbf{x})} \int_{B_\rho(\mathbf{x})} u(\mathbf{y}) e^{-\frac{|u(\mathbf{y})-u(\mathbf{x})|^2}{h^2}} d\mathbf{y},\ c(\mathbf{x}) = \int_{B_\rho(\mathbf{x})} e^{-\frac{|u(\mathbf{y})-u(\mathbf{x})|^2}{h^2}} d\mathbf{y},$$

where u_F is the filtered signal, $\mathbf{x}, \mathbf{y} \in \Omega$, $\Omega \subset \mathbb{R}^2$ is an open and bounded set, $c(\mathbf{x})$ is the normalization factor, and $B_\rho(\mathbf{x}) \subset \Omega$ is a ball of center $\mathbf{x}$ and radius ρ. Based on the ratio ρ/h this model get back to Gaussian filtering, or behaves as the Perona-Malik equation, or the intensity of the filtering tends to zero and the signal is hardly modified [6]. We point out that it has been observed that this method unfortunately could create shocks and staircasing effects. An evolution of the neighborhood filters is the non-local means (NLM) filter which averages similar values of the image according to their intensity distance [5]. The similarity between values is made robust with respect to the noise level by using region comparison rather than single values comparison. Moreover, the pattern redundancy is not restricted to be local (therefore, non-local). Then the values far from actual one being filtered are not penalized due to their distance to the current value. This approach is based on the redundancy of most natural images: for every small patch in a natural image one can find several similar

patches in the same image. In particular the algorithm makes use of the information encoded in the whole image. When modifying a value in the signal, the algorithm first computes the similarity between a fixed window centered on it and the windows centered on the other values in the whole signal, then it takes the similarity as a weight to adjust this value. Even if the NLM method has shown remarkable results, its efficiency is low due to windows matching.

In this paper we propose a new anisotropic diffusion equation introducing a nonlocal diffusive coefficient that takes into account the "monotonicity" and the oscillating pattern of the image. In other words, a high modulus of the gradient may lead to a small diffusion if the function is, for instance, locally monotone. Although the method proposed here considers some aspects of the methods described above, it represents a new approach and it does not fall within any of the previous frameworks.

1.1 A motivating $1D$ model

At present, the best view of the activity of a neural circuit is provided by multiple-electrode extracellular recording technologies, which allow us to simultaneously measure spike trains from up to a few hundred neurons in one or more brain areas during each trial. While the resulting data provide an extensive picture of neural spiking, their use in characterizing the fine timescale dynamics of a neural circuit is complicated by at least two factors. First, captured extracellular action potentials provide only an occasional view of the process from which they are generated, forcing us to interpolate the evolution of the circuit between the spikes. Second, the circuit activity may evolve quite differently on different trials that are otherwise experimentally identical. Experimental measurements are noisy. For neural recordings, the noise may arise from a multitude of sources, both intrinsic and extrinsic to the nervous system. Operationally, supposing that recorded data are composed of two parts, signal of interest and other processes unrelated to the experimental conditions. it is a challenge to preserve the essential signal features, such as suitable structures related to the neuronal activity, during the smoothing process. In Figure 1 we show an example of a noised signal of a neuron,

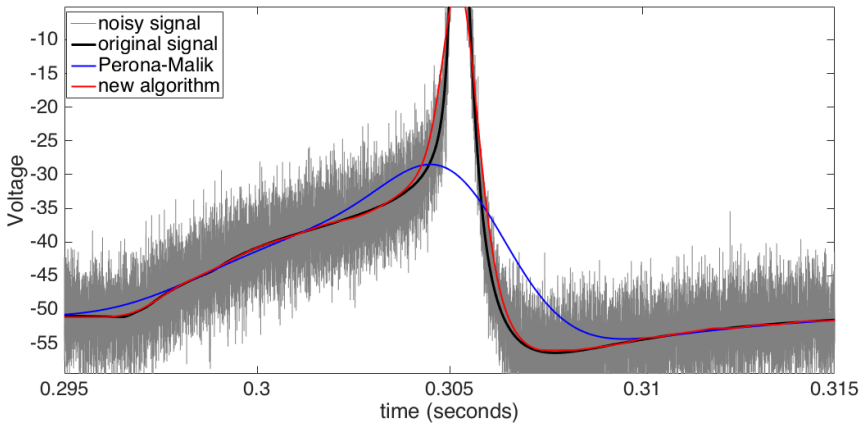

Fig. 1 $5s$ simulated membrane potential signals of a network of 20 neurons randomly connected (firing rate $= 5Hz$). In black, original signal of a neuron. White gaussian noise was added to the signals (gray). Blue: signal after denoising (Perona-Malik). Red: signal after denoising (new algorithm). Note that the event occurred between $t = 0.296$ and $t = 0.297$ is completely removed after Perona-Malik smoothing

where a white gaussian noise has been superposed to the original signal. These data were obtained from a simulation of potentials in a small neural network of cerebellar granule cell [1]. The peak in the central part

represents the firing of a neuron, it is of interest to reconstruct the behavior of the potential near this peak for the neurophysiological significance of the signal. Similar signals will be considered in the section dedicated to numerical tests. The method proposed in this paper is compared with the classic Perona-Malik algorithm. We point out that the diffusivity in the Perona-Malik model, or similar approaches, depends locally on the modulus of the gradient of the function. Instead, we introduce a nonlocal diffusive coefficient that takes into account of the "monotonicity" of the signal. In other words, a high modulus of the gradient may lead to a small diffusion if the function is also locally monotone. Motivated by this fact, we have developed the new approach presented in this paper. More precisely, the diffusion coefficient in a point x is based on the behavior of the

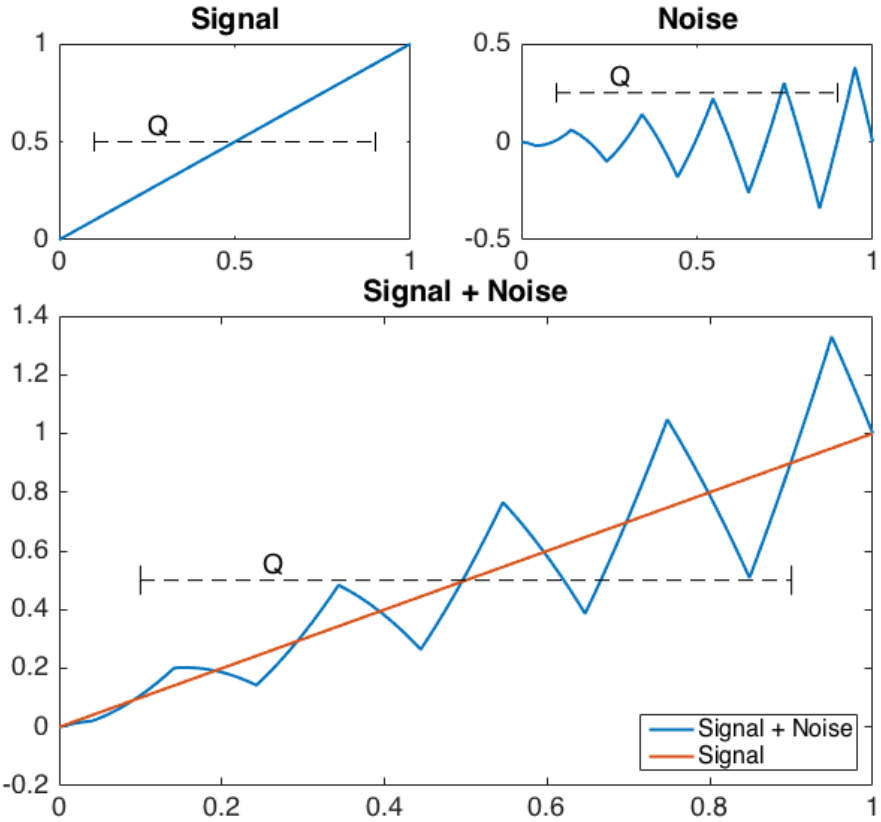

Fig. 2 Example of a nonlocal signal. The signal and the noise have the same total variation and the same modulus of the gradient. The global increment observed on Q is instead very different

function f in a interval $x+Q=(x-q,x+q)$, where $Q=(-q,+q)$, see Figure 2. Analytically, we compute the ratio between the variation $|f(x+q)-f(x-q)|$ and the total variation $\int_Q |\nabla f(s+x)|ds$ of the function in $x+Q$. A ratio close to 1 will imply a tiny noise in the signal, while a ratio close to 0 is related to a highly noised signal. As shown in Figure 2, a pure signal and a noised one may have the same total variation and the same modulus of gradient. Therefore, Perona-Malik like methods (and total variation based methods) treat the signals in the same way.

More precisely, for the one dimensional spatial case, let $u:[a,b]\to\mathbb{R}$ a real function defined on a bounded interval $[a,b]$, and a subinterval $[c,d]\subset[a,b]$. We define the *local variation* $LV_{[c,d]}(u)$ of u on the interval $[c,d]$

the value

$$LV_{[c,d]}(u) = |u(d) - u(c)|.$$

We also define the *total local variation* $TV_{[c,d]}(u)$ of u on the interval $[c,d]$ as follows

$$TV_{[c,d]}(u) = \sup_{\mathscr{P}} \sum_{i=0}^{n_P-1} |u(x_{i+1}) - u(x_i)|$$

where $\mathscr{P} = \{P = \{x_0, \dots, x_{n_P}\} \,|P \text{ is a partition of } [c,d]\}$ is the set of all possible finite partition of the interval $[c,d]$. It is easy to prove that if the function u is a monotone function on the interval $[c,d]$, then $LV_{[c,d]}(u) = TV_{[c,d]}(u)$. While, if the function u is not monotone, $LV_{[c,d]}(u) < TV_{[c,d]}(u)$. For the $1D$ signal, as the membrane potential of a neuron, where the independent variable has the dimension of a time, it is convenient to select instead of a symmetric window Q an asymmetric interval of a given length δ. Let $\varepsilon \in \mathbb{R}^+$ be a suitably "small number" and let $\delta \in \mathbb{R}^+$ a positive number. We define the ratio,

$$R_{\delta,u} = \frac{LV_{[x,x+\delta]}(u)}{\varepsilon + TV_{[x,x+\delta]}(u)} \tag{1}$$

If the parameter δ is chosen appropriately we can distinguish between oscillations caused by noise and by electrophysiological stimuli (in the following EPSP) contained in a range of amplitude δ. In the case of the membrane potential of a neuron the oscillations due to the noise and to EPSP occur on different time scales: it is possible to choose a value δ such that in a range of amplitude δ there is at least a full oscillation due to noise, but not to a complete EPSP. Then, there is an oscillation, the signal is not monotone and it is expected that the ratio $R_{\delta,u}$ is much less than one because $LV_{[x,x+\delta]}(u) << TV_{[x,x+\delta]}(u)$. On the contrary, if in the same time interval there is an EPSP, the ratio $R_{\delta,u}$ becomes close to one.

As in the Perona-Malik model, we adapt the diffusive coefficient by using the above ratio $R_{\delta,u}$. For small values of the latter we have to reduce the noise, while for values close to 1, the upper bound of $R_{\delta,u}$, we have to preserve the signal variation (as the edges in the image). The resulting non-local equation is the following,

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(g\left(\frac{LV_{[x,x+\delta]}(u)}{\varepsilon + TV_{[x,x+\delta]}(u)} \right) \frac{\partial u}{\partial x} \right) = 0, \tag{2}$$

where the function g has the same properties as in the Perona-Malik model and $\delta > 0$. If u is a differentiable function and u' is integrable, the total variation can be written as,

$$TV_{[x,x+\delta]}(u) = \int_x^{x+\delta} |u'(s)| ds,$$

and the non linear diffusion equation (2) can be stated as

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(g\left(\frac{\left| \int_x^{x+\delta} u'(s) ds \right|}{\varepsilon + \int_x^{x+\delta} |u'(s)| \, ds} \right) \frac{\partial u}{\partial x} \right) = 0.$$

for a function $u(x,t)$, $x \in (a,b)$, $t > 0$. As initial condition we take the original signal u_0 but with some regularization obtained with a standard smoothing filter, e.g. a Gaussian filter, and we assume homogeneous Neumann condition at the boundary, that is $\partial u/\partial x = 0$ for $x = a, b$ and $t > 0$.

1.2 The multidimensional case

In order to apply the new model to a multidimensional signal, in particular in the two-dimensional case (a gray level digital image), we have to generalize the ratio $R_{\delta,u}$, see (1). We must therefore consider the role of local oscillations and their relation to any noise. In the classic model $u+n$, images are assumed to be a sum of two components $u(x)$ and $n(x)$. The first component $u(x)$ is aimed at modeling the objects which are present in the given image. This $u(x)$ component should provide a good approximation to the image, while the $n(x)$ term is responsible for the noise. In the basic decompositions $u+n$ the u component is obtained through a low-pass filtering and the n component by a high-pass filtering. Then this n component is obviously oscillating. But u is modeling the objects that are contained in an image. These objects have edges, which also contain high frequencies. Then $u(x)$ and $v(x)$ should be coupled by some constraints. In the ideal case, the $u(x)$ component should remain untouched while the $v(x)$ might disappear. In some model the u component is assumed to belong to a specific ball of a functional Banach space B. Typically in image processing, we want to detect objects (planar sub-domains) delimited by contours (edges). In this model, the function $u(x)$ is assumed to be smooth in each planar domain with jump discontinuities across the boundaries. In any case we want to avoid to break an image into too many pieces and the penalty for a domain decomposition of a given image will be the sum of the lengths of the edges. But these lengths appear in the BV norm of $u(x)$. Moreover,the BV norm of a function $u(x)$ can be defined as the total mass of the distributional gradient of u: the L^1 norm of the gradient of the restriction of u to the interior of the domains (for a discussion about the functional space model see e.g. [14]). Therefore a possible extension of the one-dimensional approach to the two-dimensional case involves the behavior of the BV norm of u in appropriate subregions. We point out that this is not the only strategy to extend the model to the two-dimensional case.
Let $A \subseteq \mathbb{R}^d$ and $u : \Omega \to \mathbb{R}$ an integrable function smooth function, the total variation $TV(u)$ (or BV seminorm), can be computed as [21]

$$TV(u) = \int_A |\nabla u(x)|\, dx$$

where ∇u is the gradient of u. Here, we consider the anisotropic total variation,

$$TV_a(u) = \int_A |\nabla u(x)|_1\, dx$$

considering the l^1 norm, $|v|_1 = |v_1| + |v_2| + \ldots + |v_d|$, instead of the Euclidean norm. The usual total variation $TV(u)$ is invariant to rotations of the domain, but anisotropic TV_a is not. However, the latter allows for other approaches that do not apply with the usual TV, for example the graph-cuts algorithm [9]. Moreover, using the norm l_1 the scalar case is reduced to the one-dimensional model (2): this allows to theoretically treat both cases, $1D$ and $2D$ within the same framework. For the local variation term, the numerator of the ratio $R_{\delta,u}$, we have to compute the variation of the function u in a region A by taking into account the flux of u at the boundary ∂A of the same set A. Following the definition of the $BV-$seminorm [21], and the choice we propose the definition,

$$LV_A(u) = \sup\{\int_A \nabla u(x)\nabla h(x)\, dx,\ |\nabla h(x)|_1 \leq 1\ \forall x, h \text{ harmonic on } A\}.$$

In the above definition, due to the properties of the test function h, we have

$$\int_A \nabla u \nabla h\, dx = \int_{\partial A} u \nabla h \cdot \mathbf{n}_A\, ds - \int_A u\, div(\nabla h)\, dx$$

where div is the divergence operator, $\mathbf{n}_A$ denotes the unit outer normal to ∂A, and the last integral is equal to zero because h is harmonic. Then

$$\int_A \nabla u \nabla h\, dx = \int_{\partial A} u \nabla h \cdot \mathbf{n}_A\, ds,$$

and the supremum for the $LV_A(u)$ is taken considering all the possible orientations of the vector ∇h with respect to $\mathbf{n}_A$. Returning to the one-dimensional case, for $A = [c,d]$, we obtain,

$$LV_A(u) = \sup\{(u(c) - u(d)), (-u(c) + u(d))\} = |u(d) - u(c)|.$$

The remainder of this paper is organized as follows. In Section 2, we provide the mathematical analysis of the new non-linear and non-local diffusion equation in the two dimensional spatial case. Instead of analyzing only the one-dimensional case we consider the mathematical problem of the existence of the solution in the two dimensional case. It is straightforward to adapt the analysis to the one-dimensional model. On the other hand we will consider the application of our approach in the case of digital images with a comparison with other methods in a future paper. In particular, we show the existence of a solution for the model by using a suitable semidiscrete scheme under reasonable hypotheses for applications in image processing. In Section 3 we show some numerical experiments for one-dimensional signals.

Notation

In that follows, $\Omega \subset \mathbb{R}^2$ denotes a open bounded domain with Lipschitz continuous boundary $\Gamma = \partial\Omega$, and $\Omega_T = \Omega \times (0,T)$, with $T > 0$. We denote by $H^k(\Omega)$, k is a positive integer, the Sobolev space of all function u defined in Ω such that u and its distributional derivatives of order k all belong to $L^2(\Omega)$. Let D^s the distributional derivatives, H^k is a Hilbert space for the norm,

$$\|u\|_k = \|u\|_{H^k} = \left(\sum_{|s|\le k} \int_\Omega |D^s u(x)|^2 dx \right)^{1/2}, \qquad \|u\|_0 = \|u\|_{L^2}.$$

Let $V = H^1$, V^* stands for its dual space. We denote by $L^p(0,T;H^k(\Omega))$ the set of all functions u, such that, for almost every t in $(0,T)$, $u(t)$ belong to $H^k(\Omega)$, $L^p(0,T;H^k(\Omega))$ is a normed space for the norm

$$\|u\|_{L^p(0,T;H^k(\Omega))} = \left(\int_0^T \|u\|_k^p dt \right)^{1/p}$$

$p \ge 1$ and k a positive integer. We denote by $(\cdot,\cdot)$, the scalar product in $L^2(\Omega)$.

2 Analysis of the new nonlocal and nonlinear equation, $2D$ case

In this section we will consider the two-dimensional spatial case and we will prove the existence of a variational solution of the corresponding non-local diffusion equation. From the discussion in the subsections (1.1)-(1.2), given $U \in L^2(0,T;V)$ and $Q = (-q_1,+q_1) \times (-q_2,+q_2)$ (the local window), we can define the ratio coefficient R as the function

$$R_{Q,U}(x,t) = \begin{cases} \frac{\sup\left\{ \int_Q \nabla U(x+y,t)\nabla h(y)\,dy, |\nabla h(x)|_1 \le 1 \ \forall x,h \text{ harmonic on } Q \right\}}{\int_{x+Q} |\nabla U(y,t)|_1 dy}, & \\ & \text{if } \int_{x+Q} |\nabla U(y,t)|_1 dy > 0; \\ 0, & \text{otherwise;} \end{cases}$$

where $|\cdot|_1$ is the l^1 norm in $\mathbb{R}^2$. It is easy to verify that the function $R_{Q,U}(x,t)$ is measurable, and $0 \le R_{Q,U} \le 1$. Moreover, note that $\int_{x+Q} |\nabla U(y,t)|_1 dy$ is continuous in x since $U \in L^2(0,T;V)$.

Let $g : [0,+\infty) \longrightarrow \mathbb{R}$ be a Lipschitz continuous nonincreasing function such that $g(0) = 1$, $g(s) > 0$, $\forall s \ge 0$, $g(1) = \varepsilon > 0$. It follows that $1 \ge g(R_{Q,U}(x,t)) \ge \varepsilon$.

Let Q be the window that is used in the definition of the diffusive coefficient $R_{Q,u}$. We assume that

$$\text{Assumption 1} \quad \inf_{x\in\Omega} \frac{|\Omega \cap \{x+Q/3\}|}{|\{x+Q/3\}|} = q_\Omega > 0, \tag{3}$$

where if A is a measurable set, let $|A|$ be the Lebesque measure of A.

The smoothing process of the image u_I is obtained by the solution $u(x,t)$ of the following non-linear, non-local diffusion equation,

$$\begin{array}{ll} \frac{\partial u}{\partial t} - \operatorname{div}\left(g(R_{Q,u}(x,t))\nabla u\right) = 0, & \text{in } \Omega_T; \\ \frac{\partial u}{\partial \mathbf{n}} = 0, & \text{on } \Gamma \times (0,T); \\ u(x,0) = u_0(x) \in V; & \end{array} \tag{4}$$

with homogeneous Neumann boundary conditions for the normal derivative $\frac{\partial u}{\partial \mathbf{n}}$, and initial data $u_0 \in V$ which is a smoother version of the original image u_I.

Remark 1 We point out that the initial data is more regular with respect to classical parabolic theory but we need to ensure the well-posedness of the diffusion coefficients $R_{Q,u}$. In the numerical approximation of the equation (4) we obtain a suitable initial data from the original signal u_I by using a convolutional operator with a Gaussian filter.

2.1 Rothe method and a priori estimates

In order to prove the existence of a solution $u \in L^2(0,T;V) \bigcap C^0(0,T;L^2)$ we consider the so called Rothe-type approximation [11] of (4) which consists in using time discretization to approximate the evolution problem by a sequence of elliptic ones. To show the convergence of such a process, a common approach is to follow the following steps:

1. for each approximate problem, prove the existence of a solution, and derive a-priori estimates satisfied by any solution;
2. then use compactness arguments to show (up to the extraction of a subsequence) the existence of a limit;
3. Finally, prove that the previous limit satisfies the original problem.

Let $0 = t_0 < t_1 < \ldots < t_N = T$ denote the time discretization with $t_{i+1} = t_i + \tau$, where τ is the time step. Let u_i be the solution of linear equation,

$$\frac{u_i - u_{i-1}}{\tau} - \operatorname{div}\left(g(R_{Q,u_{i-1}}(x,t))\nabla u_i\right) = 0, \tag{5}$$

with $u_0 = u(x,0)$, and homogeneous Neumann boundary conditions. Let $\delta u_i = (u_i - u_{i-1})/\tau$ the backward difference at time t_i, we understand the solution of (5) in the variational sense, i.e., we look for $u_i \in V$, for $i = 1,\ldots,N$ satisfies the identity

$$(\delta u_i, v) + \left(g(R_{Q,u_{i-1}})\nabla u_i, \nabla v\right) = 0, \ \forall v \in V, \tag{6}$$

where $u_0 \in V$ is given. By introducing the bilinear form $a_{\tau,w}$, on $V \times V$,

$$a_{\tau,w} = (u,v) + \tau a_w(u,v), \qquad a_w(u,v) = (g(R_{Q,w})\nabla u, \nabla v), \tag{7}$$

for a given $w \in V$, we can rewrite the previous identity as,

$$a_{\tau,u_{i-1}}(u_i, v) = (u_{i-1}, v), \forall v \in V. \tag{8}$$

The term $a_w(u,v)$ in (7) is weakly coercive, i.e., there exist two constants $c_1 > 0$ and $c_2 > 0$ such that

$$a_w(u,u) + c_2\|u\|_0^2 \geq c_1\|u\|_1^2, \forall u \in V. \tag{9}$$

Furthermore, the form $a_{\tau,w}$ is continuous and it verifies

$$a_{\tau,w}(u,u) \geq \tau c_1\|u\|_1^2 + (1 - \tau c_2)\|u\|_0^2,$$

then it is $V-$elliptic if $\tau c_2 \leq 1$. Under this coercivity condition, the existence and the uniqueness of $u_i \in V$, $i = 1,\dots,N$ from (6) is guaranteed by the Lax-Milgram Theorem [17]. Now, we introduce the so-called Rothe function

$$u^{(N)}(t) = u_{i-1} + (t - t_{i-1})\delta u_i, \text{ for } t_{i-1} \leq t \leq t_i, i = 1,\dots,N \tag{10}$$

which we consider as a linear piecewise approximation of the problem (4). Together with $u^{(N)}$ we consider the step function

$$\bar{u}^{(N)}(t) = u_i, \text{ for } t_{i-1} < t \leq t_i, i = 1,\dots,N \tag{11}$$

with $\bar{u}^{(N)}(0) = u_0$. In that follows C denotes the generic positive constant.

Lemma 1 *Let u_i, $i = 1,\dots,N$, be the solution of problem (8), then*

$$\max_{1\leq i\leq N} \|u_i\|_0 \leq C, \tag{12}$$

hold uniformly for N, and $\bar{u}^{(N)}(t)$, $u^{(N)}(t) \in L^\infty(0,T;L^2)$.

Proof First we test (6) at time t_{k+1} by $v = \tau u_{k+1}$ and sum over $k = 0,\dots,p \geq 1$, we obtain (let us define $a_{u_k}(u,v) = a_k(u,v)$),

$$\sum_{k=0}^{p} (u_{k+1} - u_k, u_{k+1}) + \tau \sum_{k=0}^{p} a_k(u_{k+1}, u_{k+1}) = 0, \qquad p = 1,\dots(n-1).$$

By using the identity $2(u-v,u) = (u-v,u-v) + (u,u) - (v,v)$, we have

$$\sum_{k=0}^{p} \|u_{k+1} - u_k\|_0^2 + \|u_{p+1}\|_0^2 - \|u_0\|_0^2 + 2\tau \sum_{k=0}^{p} a_k(u_{k+1}, u_{k+1}) = 0,$$

Then, from (9),

$$\sum_{k=0}^{p} \|u_{k+1} - u_k\|_0^2 + \|u_{p+1}\|_0^2 - \|u_0\|_0^2 + 2\tau c_1 \sum_{k=0}^{p} \|u_{k+1}\|_1^2 - 2\tau c_2 \sum_{k=0}^{p} \|u_{k+1}\|_0^2 \leq 0,$$

which can be rewritten as follows

$$\sum_{k=0}^{p} \|u_{k+1} - u_k\|_0^2 + \|u_{p+1}\|_0^2 + 2\tau c_1 \sum_{k=0}^{p} \|u_{k+1}\|_1^2 \leq 2\tau c_2 \sum_{k=0}^{p} \|u_{k+1}\|_0^2 + \|u_0\|_0^2,$$

that is

$$s_{p+1} \leq \|u_0\|_0^2 + 2\tau c_2 \sum_{k=1}^{p} \|u_k\|_1^2,$$

where

$$s_p = \sum_{k=0}^{p-1} \|u_{k+1} - u_k\|_0^2 + \frac{1}{2}\|u_p\|_0^2 + 2\tau c_1 \sum_{k=0}^{p-1} \|u_{k+1}\|_1^2.$$

By the inequality $2s_p \geq \|u_p\|_0^2$ and by the definition of s_p the following estimate is obtained

$$s_{p+1} \leq \|u_0\|_0^2 + 4\tau c_2 \sum_{k=1}^{p} s_k.$$

Applying the discrete Gronwall lemma we have the following inequalities,

$$s_p \leq \|u_0\|_0^2 (1 + 4\tau c_2)^{p-1} \text{ for } p = 1,\dots,N.$$

Then

$$s_p \leq \|u_0\|_0^2 (1 + 4\tau c_2)^N = \|u_0\|_0^2 (1 + 4\tau c_2)^{T/\tau} \text{ for } p = 1,\dots,N.$$

The function $\tau \to (1 + 4\tau c_1)^{T/\tau}$ is bounded for $\tau \in (0, +\infty)$, then, for a suitable constant $\bar{C}$,

$$\max_p \frac{\|u_p\|_0^2}{2} \leq \max_p s_p \leq \|u_0\|_0^2 \bar{C} = C, \tag{13}$$

which leads to the a-priori estimates in (12).

Lemma 2 *The estimates*

$$\tau \sum_{k=1}^{N} \|\nabla u_k\|_0^2 \leq C, \qquad \sum_{k=1}^{N} \|u_k - u_{k-1}\|_0^2 \leq C \tag{14}$$

hold uniformly with respect to N. Furthermore, there is a constant C such that

$$\max_{0\leq t\leq T} \|u^{(N)}(t)\|_0^2 + \int_0^T \|\bar{u}^{(N)}(t)\|_1^2 dt \leq C\|u_0\|_0^2, \tag{15}$$

$$\int_0^T \|\bar{u}^{(N)}(t) - u^{(N)}(t)\|_0^2 dt \leq C\tau\|u_0\|_0^2, \tag{16}$$

where $u^{(N)}(t)$ and $\bar{u}^{(N)}(t)$ are, respectively, the Rothe function, see (10), and the step function, see (11).

Proof The estimates followed by the upper bound obtained in the Lemma 1, see (13),

$$\sum_{k=0}^{p} \|u_{k+1} - u_k\|_0^2 + \frac{1}{2}\|u_{p+1}\|_0^2 + 2\tau c_1 \sum_{k=0}^{p} \|u_{k+1}\|_1^2 \leq C\|u_0\|_0^2, \tag{17}$$

and from the properties of the bilinear form $a_k(u,v)$.
We have

$$\max_{0\leq t\leq T} \|u^{(N)}(t)\|_0^2 = \max_{0\leq n\leq N} \|u_n\|_0^2 \text{ and } \int_0^T \|\bar{u}^{(N)}(t)\|_1^2 dt = \tau \sum_{n=1}^{N} \|u_n\|_1^2,$$

and, from (13),

$$\max_{0\leq t\leq T} \|u^{(N)}(t)\|_0^2 + \int_0^T \|\bar{u}^{(N)}(t)\|_1^2 dt \leq C\|u_0\|_0^2.$$

Now,

$$\int_0^T \|\bar{u}^{(N)}(t) - u^{(N)}(t)\|_0^2 dt = \tau \sum_{n=0}^{N-1} \int_0^1 \|\bar{u}^{(N)}((n+s)\tau) - u^{(N)}((n+s)\tau)\|_0^2 ds =$$
$$= \tau \sum_{n=0}^{N-1} \int_0^1 (1-s)^2 \|u_{n+1} - u_n\|_0^2 ds = \frac{\tau}{3} \sum_{n=0}^{N-1} \|u_{n+1} - u_n\|_0^2.$$

Using the inequality (17) we can state

$$\int_0^T \|\bar{u}^{(N)}(t) - u^{(N)}(t)\|_0^2 dt \leq \tau C\|u_0\|_0^2.$$

2.2 Compactness and passage to the limit

Lemma 3 *There exists $u \in L^2(0,T;V)$ such that (in the sense of subsequences)*

$$\bar{u}^{(N)} \rightharpoonup u \quad in\ L^2(0,T;V);$$
$$u^{(N)} \rightharpoonup u \quad in\ L^2(0,T;V).$$

Proof From the estimates of Lemma 2 we can deduce

$$\begin{gathered} \|u^{(N)}\|_{L^\infty(0,T;L^2)} + \|\bar{u}^{(N)}\|_{L^\infty(0,T;V)} \leq C \\ \|\bar{u}^{(N)} - u^{(N)}\|_{L^2(0,T;L^2)} \leq C\sqrt{\tau}. \end{gathered} \tag{18}$$

It follows that, at least for a subsequence,

$$\begin{gathered} u^{(N)} \overset{*}{\rightharpoonup} u \ \text{ in } L^\infty(0,T;L^2) \\ \bar{u}^{(N)} \rightharpoonup z \ \text{ in } L^2(0,T;V). \end{gathered} \tag{19}$$

But, the (18) imply that

$$\bar{u}^{(N)} - u^{(N)} \to 0 \text{ in } L^2(0,T;L^2),$$

then $z = u$ in (19) and

$$\bar{u}^{(N)} \rightharpoonup u \text{ in } L^2(0,T;V). \tag{20}$$

We recall that $\Omega_T = \Omega \times (0,T)$, the estimates (12) and (14) imply the equicontinuity of the Rothe approximation $u^{(N)}$, together with the step function $\overline{u}(N)$.

Lemma 4 *There exists $u \in L^2(0,T;V)$ with $\partial u/\partial t \in L^2(0,T;V^*)$ such that (in the sense of subsequences)*

$$u^{(N)} \to u, \quad \overline{u}^{(N)} \to u \text{ in } L^2(\Omega_T);$$
$$\frac{\partial u^{(N)}}{\partial t} \rightharpoonup \frac{\partial u}{\partial t} \text{ in } L^2(0,T;V^*).$$

Proof Let $s \in (0,T)$, we consider the time translate variation of the approximation $\overline{u}^{(N)}(t)$,

$$J_s = \int_0^{T-s} \|\overline{u}^{(N)}(t+s) - \overline{u}^{(N)}(t)\|_0^2 dt.$$

There exists an integer k such that $k\tau \le s \le (k+1)\tau$, then, by the definition of the step function $\overline{u}^{(N)}$,

$$J_s = \tau \sum_{l=0}^{N-k} \|u_{l+k} - u_l\|_0^2.$$

From the Lemma 1 and 2 it follows that

$$J_s \le Ck\tau,$$

for a suitable constant C, and the time translate estimate,

$$\int_0^{T-s} \|\overline{u}^{(N)}(t+s) - \overline{u}^{(N)}(t)\|_0^2 dt \le C(s+\tau). \tag{21}$$

By using again the estimates (12), (14), and the definition of the step approximation $\overline{u}^{(N)}$, it is easy to show that

$$\int_0^T \|\overline{u}^{(N)}(t)\|_1^2 dt \le C_u. \tag{22}$$

Given a vector $\xi \in \mathbb{R}^2$, let $\Omega_\xi = \{x \in \Omega : x+\xi \in \Omega\}$ and $\Omega_{\xi,T} = \Omega_\xi \times (0,T)$. From the inequality (22) we have the following space translate estimate,

$$\int_{\Omega_{\xi,T}} \|\overline{u}^{(N)}(x+\xi,t) - \overline{u}^{(N)}(x,t)\|_0^2 \, dxdt \le C_\xi |\xi|^2, \quad \text{for } |\xi| \text{ sufficiently small.} \tag{23}$$

Due to the time and translate estimates, respectively (21) and (23), the set $\{\overline{u}^{(N)}\}_N$ is compact in $L^2(\Omega_T)$ because of Kolmogorov's relative compactness criterion [7,13]. The, we can conclude $\overline{u}^{(N)} \longrightarrow u$ in $L^2(\Omega_T)$ (and also pointwise in Ω_T),and $u \in L^2(0,T;V)$. From the definition of $\overline{u}^{(N)}$ and $u^{(N)}$, and from Lemma 2 it follows the estimate

$$\int_0^T \|\overline{u}^{(N)}(t) - u^{(N)}(t)\|_0^2 dt \le \frac{C_d}{N},$$

then $u^{(N)} \longrightarrow u$ in $L^2(\Omega_T)$.
Observing that $\partial u^{(N)}/\partial t = (u_i - u_{i-1})/\tau$, for $t \in (t_{i-1}, t_i)$, we can compute

$$\left\| \frac{\partial u^{(N)}}{\partial t} \right\|_* = \sup_{v \in V,\ \|v\| \le 1} |((u_i - u_{i-1})/\tau, v)|.$$

Then, the following estimate holds uniformly for N,

$$\int_0^T \left\| \frac{\partial u^{(N)}}{\partial t} \right\|_*^2 dt \le C_*,$$

and we can deduce the weak convergence of the time derivative of the Rothe approximations $u^{(N)}$.

Lemma 5 *With the notation of Lemma 4,*

$$u^{(N)} \to u, \quad \bar{u}^{(N)} \to u \qquad in\ L^2(0,T;V);$$

Proof Now we shall prove the $L^2(0,T;V)$ convergence of $\bar{u}^{(N)}$ to u (which belongs to the space $L^2(0,T;V)$). So, let us test (6) by $v = \bar{u}^{(N)} - u$ and integrate it over the time interval $(0,T)$ by using the partition from the subinterval (t_{i-1},t_i),

$$\sum_i^N \int_{t_{i-1}}^{t_i} \left[((u_i - u_{i-1})/\tau, u_i - u) + \left(g(R_{Q,u_{i-1}})\nabla u_i, \nabla u_i - \nabla u\right)\right] dt = 0. \tag{24}$$

From the definition of the Rothe function $u^{(N)}$ we observe that for $t \in [t_{i-1},t_i]$,

$$\frac{\partial u^{(N)}}{\partial t} = \frac{u_i - u_{i-1}}{\tau} = \delta_i.$$

Moreover, see Lemma 4, $\partial u^{(N)}/\partial t \rightharpoonup \partial u/\partial t$, then, in the limit $N \to \infty$,

$$\int_0^{t_i} < \frac{\partial u^{(N)}}{\partial t}, u > dt \ \to \ \int_0^{t_i} < \frac{\partial u}{\partial t}, u > dt, \tag{25}$$

and from the integration by part formula,

$$\int_0^{t_i} < \frac{\partial u^{(N)}}{\partial t}, u > dt \ \to \ \frac{1}{2}\|u(t_i)\|_0^2 - \frac{1}{2}\|u_0\|_0^2. \tag{26}$$

In order to get an estimate for the backward difference δ_i we consider the variational equation (6) with $v = (u_i - u_{i-1})$, adding for $i = 1,2,\ldots,N$, using (9) and assuming $2\tau c_2 \le 1$, we state

$$\frac{1}{\tau}\sum_{i=1}^N \|u_i - u_{i-1}\|_0^2 \le C\|u_0\|_0^2. \tag{27}$$

Moreover, using a discrete analog of Leibniz' product rule we have

$$\sum_{i=1}^N (u_i - u_{i-1}, u_i) = \frac{1}{2}\|u_N\|_0^2 - \frac{1}{2}\|u_0\|_0^2 + \frac{1}{2}\sum_{i=1}^N \|u_i - u_{i-1}\|_0^2. \tag{28}$$

From the Lemma 4, the identity (28), and the inequality (27) we deduce that

$$\lim_{N\to\infty} \sum_i^N \int_{t_{i-1}}^{t_i} ((u_i - u_{i-1})/\tau, u_i - u)\, dt = 0. \tag{29}$$

We recall that $1 \ge g(R_{Q,u_{i-1}}) \ge g(1) = \varepsilon$, then we obtain

$$\varepsilon \int_0^T \|\nabla \bar{u}^{(N)} - \nabla u\|_0^2 dt \le \sum_i^N \int_{t_{i-1}}^{t_i} \left(g(R_{Q,u_{i-1}})(\nabla u_i - \nabla u), \nabla u_i - \nabla u\right) dt.$$

From (24), and the above inequality, we have

$$\begin{aligned} \varepsilon \int_0^T \|\nabla \bar{u}^{(N)} - \nabla u\|_0^2 dt + \sum_i^N \int_{t_{i-1}}^{t_i} \left(g(R_{Q,u_{i-1}})\nabla u, \nabla u_i - \nabla u\right) dt \\ \le \sum_i^N \int_{t_{i-1}}^{t_i} ((u_i - u_{i-1})/\tau, u - u_i)\, dt. \end{aligned}$$

Now, from the (29), the weak convergence as in Lemma 3, we can find a vanishing sequence $\{C_N\}$, $C_N \in \mathbb{R}$, $\lim_{N\to\infty} C_N = 0$ such that

$$\varepsilon \int_0^T |\nabla \bar{u}^{(N)} - \nabla u|^2 dt \le C_N,$$

which implies $\bar{u}^{(N)} \to u$ in $L^2(0,T;V)$. Using Lemma 2 it is also possible to prove the convergence $u^{(N)} \to u$ in $L^2(0,T;V)$.

2.3 Existence of a variational solution

In order to prove that the limit u is a variational solution of (4) we have to consider the property of the stability of the kernel $g(R_{Q,u})$ with respect a variation in the space V. First, we introduce some results from the measure theory.

Lemma 6 (Vitali covering) *Let $\cup_{i=1}^{n} x_i + Q/3$ be a finite cover of a set $\tilde{\Omega} \subseteq \Omega$. Then there exists a finite sub-cover $\cup_{j=1}^{m}\{x_j + Q\}$ such that $\{x_j + Q/3, j = 1,\dots,m\}$ are disjoint.*

Corollary 1 *Let $\tilde{\Omega} \subseteq \Omega$. Then there exists $x_0,\dots,x_{N_0} \in \tilde{\Omega}$ such that $\tilde{\Omega} \subseteq \cup_{j=1}^{N_0}\{x_j + Q\}$ and $N_0 \leq \frac{3^2|\Omega|}{|Q|q_\Omega}$.*

Proof Denote by $\hat{\Omega}$ the closure of $\tilde{\Omega}$ in Ω. Take the open cover of $\hat{\Omega}$ made by $\mathscr{C} = \{x + Q/3, x \in \tilde{\Omega}\}$. Since $\overline{\Omega}$ is compact, then $\hat{\Omega}$ is compact and hence there exists a finite cover of $\hat{\Omega}$ made by $\{\cup_{i=1}^{n}\{x_i + Q/3\}\}$. By the Vitali covering Lemma, there exists a finite sub-cover $\cup_{j=1}^{N_0}\{x_j + Q\}$ such that $\{x_j + Q/3, j = 1,\dots,N_0\}$ are disjoint. Moreover,

$$|\Omega| \geq |\Omega \cap (\cup_{j=1}^{N_0}\{x_j + Q/3\})| = \sum_{j=1}^{N_0} |\Omega \cap \{x_j + Q/3\})| \geq q_\Omega N_0 \frac{|Q|}{3^2},$$

the last inequality being a consequence of (3), and since $|\{x_j + Q/3\}| = |Q/3| = |Q|/3^2$.

The following result shows the stability of the kernel $g(R_{Q,u})$ when the limiting function is not locally constant.

Lemma 7 *The function $R_{Q,U}(x,t)$ is continuous at $U \in V$ on the set*

$$\Big\{(x,t)\colon \int_{x+Q} |\nabla U(y,t)|_1 dy > 0\Big\}.$$

Proof Denote by

$$N_U(x,t) = \sup\Big\{\int_Q \nabla U(x+y,t)\nabla h(y)\,dy,\ |\nabla h(x)|_1 \leq 1\ \forall x, h \text{ harmonic in } Q\Big\},$$
$$D_U(x,t) = \int_{x+Q} |\nabla U(y,t)|_1 dy,$$

and by

$$|\nabla u|_1 = \Big|\frac{\partial u}{\partial x_1}\Big| + \Big|\frac{\partial u}{\partial x_2}\Big|, \qquad |\nabla u|_2 = \sqrt{\Big(\frac{\partial u}{\partial x_1}\Big)^2 + \Big(\frac{\partial u}{\partial x_2}\Big)^2},$$

the seminorm $|\cdot|_1$ and, respectively, $|\cdot|_2$, then $\frac{|\nabla u|_1}{\sqrt{2}} \leq |\nabla u|_2 \leq |\nabla u|_1$. Let $U_n \to U$ in $L^2(0,T;V)$. By definition, for a.e. $t \in (0,T)$,

$$\int_\Omega |\nabla U - \nabla U_n|_2^2 dy \to 0,$$

and hence

$$\frac{1}{2\cdot|\Omega|}\Big(\int_\Omega |\nabla U - \nabla U_n|_1\,dy\Big)^2 \leq \int_\Omega \Big(\frac{|\nabla U - \nabla U_n|_1}{\sqrt{2}}\Big)^2 dy \leq \int_\Omega |\nabla U - \nabla U_n|_2^2\,dy \to 0. \tag{30}$$

As a direct consequence,

$$D_{U_n}(x,t) \to D_U(x,t) > 0, \qquad \text{for a.e. } (x,t)\colon \int_{x+Q} |\nabla U(y,t)|_1 dy > 0. \tag{31}$$

For what concerns N_U, if $|\nabla h(x)|_1 \le 1$ then we get $|\nabla h(x)|_\infty \le 1$. Then

$$\begin{aligned}\Big|\int_Q \nabla U_1(x+y,t)\nabla h(y)\,dy - \int_Q \nabla U_2(x+y,t)\nabla h(y)\,dy\Big| \\ \le \textstyle\int_Q |\nabla U_1(x+y,t) - \nabla U_2(x+y,t)\,\nabla h(y)|\,dy \\ \le \|\nabla U_1(x+\cdot,t) - \nabla U_2(x+\cdot,t)\|_1\,\|\nabla h\|_\infty \\ \le \textstyle\int_{x+Q} |\nabla U_1(y,t) - \nabla U_2(y,t)|_1\,dy\end{aligned}$$

and hence, for a.e. t,

$$\begin{aligned}&\Big|\sup\Big\{\int_Q \nabla U_1(x+y,t)\nabla h(y)\,dy,\ |\nabla h(x)|_1 \le 1\ \forall x, h \text{ harmonic in } Q\Big\} \\ &\qquad - \sup\Big\{\int_Q \nabla U_2(x+y,t)\nabla h(y)\,dy,\ |\nabla h(x)|_1 \le 1\ \forall x, h \text{ harmonic in } Q\Big\}\Big| \\ &\le \sup\Big\{\Big|\int_Q \nabla U_1(x+y,t)\nabla h(y)\,dy - \int_Q \nabla U_2(x+y,t)\nabla h(y)\,dy\Big|,\ |\nabla h(x)|_1 \le 1\ \forall x, h \text{ harmonic in } Q\Big\} \\ &\le \int_{x+Q} |\nabla U_1(y,t) - \nabla U_2(y,t)|_1\,dy.\end{aligned}$$

By (30), $N_{U_n}(x,t) \to N_U(x,t)$, which concludes the proof together with (31).

Remark. In the one-dimensional case the continuity of the function $R_{Q,U}(x,t)$ with respect to the $U \in V$ is much simpler. In fact, in this case, we have

$$\begin{aligned}N_U(x,t) &= \Big|\int_x^{x+\delta} u'(s)ds\Big| \\ D_U(x,t) &= \varepsilon + \int_x^{x+\delta} \big|u'(s)\big|\,ds.\end{aligned}$$

Lemma 8 *Let $u_n \to u$ in V. For any $w \in C^1(\Omega;\mathbb{R})$ and $t \in (0,T)$,*

$$\lim_{n\to\infty}\Big(\int_\Omega g(R_{Q,u_n})\nabla u_n \nabla w\,dx - \int_\Omega g(R_{Q,u})\nabla u \nabla w\,dx\Big) = 0.$$

Proof Note that

$$\begin{aligned}&\Big|\int_\Omega g(R_{Q,u_n})\nabla u_n\nabla w\,dx - \int_\Omega g(R_{Q,u})\nabla u\nabla w\,dx\Big| \\ &\le \int_\Omega |g(R_{Q,u_n})|\,|(\nabla u_n - \nabla u)\nabla w|\,dx + \int_\Omega |g(R_{Q,u_n}) - g(R_{Q,u})|\,|\nabla u\nabla w|\,dx.\end{aligned} \tag{32}$$

For what concerns the RHS of (32), the first term vanishes as n goes to infinity since g is bounded. For the second term of the RHS of (32), we get

$$\begin{aligned}\int_\Omega |g(R_{Q,u_n}) - g(R_{Q,u})|\,|\nabla u\nabla w|\,dx &= \int_{\widetilde{\Omega}^\varepsilon} |g(R_{Q,u_n}) - g(R_{Q,u})|\,|\nabla u\nabla w|\,dx \\ &+ \int_{\Omega\setminus\widetilde{\Omega}^\varepsilon} |g(R_{Q,u_n}) - g(R_{Q,u})|\,|\nabla u\nabla w|\,dx\end{aligned} \tag{33}$$

where,

$$\widetilde{\Omega}^\varepsilon = \Big\{x\colon \int_{x+Q} |\nabla u(y)|_1 dy < \frac{\varepsilon|Q|q_\Omega}{g(0)K3^2|\Omega|} \le \frac{\varepsilon}{g(0)KN_0}\Big\},$$

K is such that $\max(|\frac{\partial w}{x_1}|, |\frac{\partial w}{x_2}|) \leq K$, and N_0 is defined in Corollary 1. The first term of the RHS of (33) is uniformly bounded (in n) by ε: by Corollary 1

$$\begin{aligned}\int_{\widetilde{\Omega}^\varepsilon} |g(R_{Q,u_n}) - g(R_{Q,u})|\,|\nabla u \nabla w|\,dy &\leq g(0)K \int_{\widetilde{\Omega}^\varepsilon} |\nabla u(y)|_1\,dy \\ &\leq g(0)K \int_{\cup_{j=1}^{N_0}\{x_j+Q\}} |\nabla u(y)|_1\,dy \\ &\leq g(0)K \sum_{j=1}^{N_0} \int_{x_j+Q} |\nabla u(y)|_1\,dy \\ &\leq g(0)KN_0 \int_{x_j+Q} |\nabla u(y)|_1\,dy \leq \varepsilon.\end{aligned}$$

The second term of the RHS of (33) vanishes as a consequence of the Dominated Convergence Theorem. In fact, $|g(R_{Q,u_n}) - g(R_{Q,u})\nabla u \nabla w| \leq g(0)|\nabla u \nabla w| \in L^1$; and $g(R_{Q,u_n}) \to g(R_{Q,u})$ on $\Omega \setminus \widetilde{\Omega}^\varepsilon$ by Lemma 7.

Now we prove that the limit u is a variational solution of (4).

Lemma 9 (Existence) *For any $u_0 \in V$, there exists $u \in L^2(0,T;V)$ with $\frac{\partial u}{\partial t} \in L^2(0,T;V^*)$ such that $u(x,0) = u_0(x)$ on Ω, $\frac{\partial u}{\partial \mathbf{n}} = 0$ on $\Gamma \times (0,T)$ and*

$$\int_\Omega \frac{\partial u}{\partial t} w\,dx = \int_\Omega \operatorname{div}(g(R_{Q,u})\nabla u) w\,dx, \qquad \forall w \in C_0^1(\Omega).$$

Proof By Lemma 4 and Lemma 5, there exists a sequence $u^{(N)}$ such that

$$u^{(N)} \to u \text{ in } L^2(0,T;V), \qquad \frac{\partial u^{(N)}}{\partial t} \rightharpoonup \frac{\partial u}{\partial t} \text{ in } L^2(0,T;V^*).$$

Let $\phi \in C_c^\infty(0,T)$ be a real-valued test function and $w \in C_0^1(\Omega)$. Taking $v(x,t) = \phi(t)w(x)$ as a test function and integrating the result with respect to t, we find that

$$\int_0^T \Big(\int_\Omega \frac{\partial u^{(N)}}{\partial t} v(x,t)\,dx + \int_\Omega g(R_{Q,u_N})\nabla u^{(N)} \nabla v(x,t)\,dx \Big)\,dt = 0.$$

We take the limit of this equation as $N \to \infty$. Since the function $t \to \phi(t)w$ belongs to $L^2(0,T;V)$, we have

$$\int_0^T \int_\Omega \frac{\partial u^{(N)}}{\partial t} v(x,t)\,dx\,dt \longrightarrow \int_0^T \int_\Omega \frac{\partial u}{\partial t} v(x,t)\,dx\,dt.$$

Moreover, Lemma 8 shows that

$$\int_0^T \Big(\int_\Omega g(R_{Q,u_N})\nabla u^{(N)} \nabla w\,dx \Big)\phi(t)\,dt \to \int_0^T \Big(\int_\Omega g(R_{Q,u})\nabla u \nabla w\,dx \Big)\phi(t)\,dt$$

It therefore follows that u satisfies

$$\int_0^T \phi(t) \Big(\int_\Omega \frac{\partial u^{(N)}}{\partial t} w\,dx + \int_\Omega g(R_{Q,u})\nabla u \nabla w\,dx \Big)\,dt = 0, \qquad \forall \phi \in C_c^\infty(0,T),$$

and hence, for almost every $t \in (0,T)$,

$$\int_\Omega \frac{\partial u}{\partial t} w\,dx = \int_\Omega \operatorname{div}(g(R_{Q,u})\nabla u) w\,dx, \qquad \forall w \in C_0^1(\Omega).$$

3 Numerical experiments

We report here some numerical test for the one-dimensional case that is for $1D$ signal. We point out that a digital signal/image is usually defined on a uniform subdivision of an interval or a rectangular domain. For the one-dimensional case we consider a finite differences approach. In that follows we intend to show the good properties of the new method while a comparison with other approaches and the numerical analysis, both for 1D and 2D cases, of the considered approximation will be considered in a future paper.

3.1 $1D$ problem

We approximate the $1D$ non linear diffusion equation (2), for $x \in [0,L]$, $t \in [0,T]$, coupled with homogeneous Neumann conditions for $x = 0$, $x = L$. We introduce the lattice of coordinates $(ih, k\tau)$ where $h = L/N$, $N > 1$ integer, $0 \leq i \leq N$; $\tau = T/M$, $M > 1$ integer, $0 \leq k \leq M$. We denote u_i^k an approximation of $u(ih, k\tau)$. For simplicity in numerical tests we will choose the parameter δ (the length of the window for the non local term) as $\delta = lh$, where $l \in \mathbb{N}$. We approximate the local variation $LV_{[x_i, x_i+\delta]}\left(u^k\right)$ by $LV_i = |u_{i+l}^k - u_i^k|$, and we define $TV_i^k = \sum_{j=0}^{l-1} |u_{i+j+1}^k - u_{i+j}^k|$ as a discretization of the total variation $TV_{[x_i, x_i+\delta]}\left(u^k\right)$. We also define the following quantities (where $g(s)$ is the edge-stopping function),

$$g_i^k = g\left(\frac{LV_i}{\varepsilon + TV_i}\right) \qquad g_{i+\frac{1}{2}}^k = \frac{g_i^k + g_{i+1}^k}{2}$$
$$\partial_x u_i^{k+1} = \frac{u_{i+1}^{k+1} - u_i^{k+1}}{h} \qquad \phi_{i+\frac{1}{2}}^k = g_{i+\frac{1}{2}}^k \partial_x u_i^{k+1}$$

$$\begin{aligned}
[\partial_x (g \partial_x u)]_i^k &= \frac{\phi_{i+\frac{1}{2}}^k - \phi_{i-\frac{1}{2}}^k}{h} \\
&= \frac{g_{i-1}^k + g_i^k}{2h^2} u_{i-1}^{k+1} - \frac{g_{i-1}^k + 2g_i^k + g_{i+1}^k}{2h^2} u_i^{k+1} + \frac{g_i^k + g_{i+1}^k}{2h^2} u_{i+1}^{k+1} \\
&= \beta_i^k u_{i-1}^{k+1} - \alpha_i^k u_i^{k+1} + \gamma_i^k u_{i+1}^{k+1}
\end{aligned}$$

Then we can state our semi-implicit numerical scheme

$$\frac{U^{k+1} - U^k}{\tau} = A\left(U^k\right) U^{k+1}$$

where U^k is the vector of the values u_i^k, and the matrix $A\left(U^k\right)$ is defined as

$$A(U^k) = \begin{pmatrix}
\alpha_1^k & \gamma_1^k & & & \\
\beta_2^k & \alpha_2^k & \gamma_2^k & & \\
& \ddots & \ddots & \ddots & \\
& & \beta_{N-1}^k & \alpha_{N-1}^k & \gamma_{N-1}^k \\
& & & \beta_N^k & \alpha_N^k
\end{pmatrix}$$

Then at each time step we have to numerically solve the following linear system,

$$\left(I - \tau A\left(U^k\right)\right) U^{k+1} = U^k$$

Let $B\left(U^k\right)$ the matrix $\left(I - \tau A\left(u^k\right)\right)$, it is easy to show that B is a strictly diagonally dominant matrix, and then it is non singular.

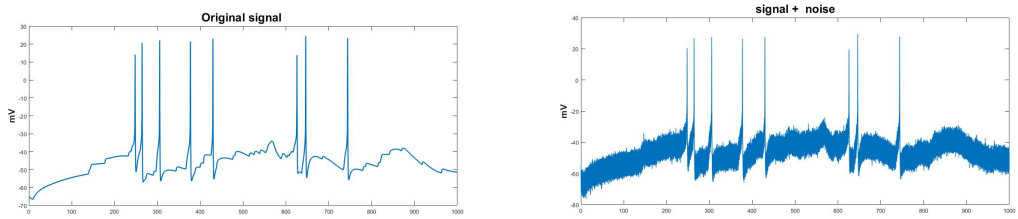

Fig. 3 On the left: $5s$ simulated membrane potential signals of a neuron (firing rate $= 5Hz$). On the right: Gaussian noise was added to the original signal.

3.2 $1D$ Test

In this section we present numerical experiments demonstrating some features of the model (2). The signals considered here are biological signals related to simulations and experiments in neurophysiology [1,15]. In Figure 3 we show a signal that represents the action potential (in milliVolt=mV) of a neuron within a small neural network. Then, we added a Gaussian noise to the original signal and applied our method for the reconstruction of the signal. The mesh size $h \sim 10^{-4}$ depends on the sampling rate of the potential and it cannot change. The parameter ε has the same order of magnitude as h.
In Figure 4 we have reported the numerical results for different choices of the parameter δ; (A) $\delta = 50h$, (B) $\delta = 20h$, (C) $\delta = 10h$, (D) $\delta = 5h$. In these numerical test the final (pseudo-)time is $T = 2$. There are few and

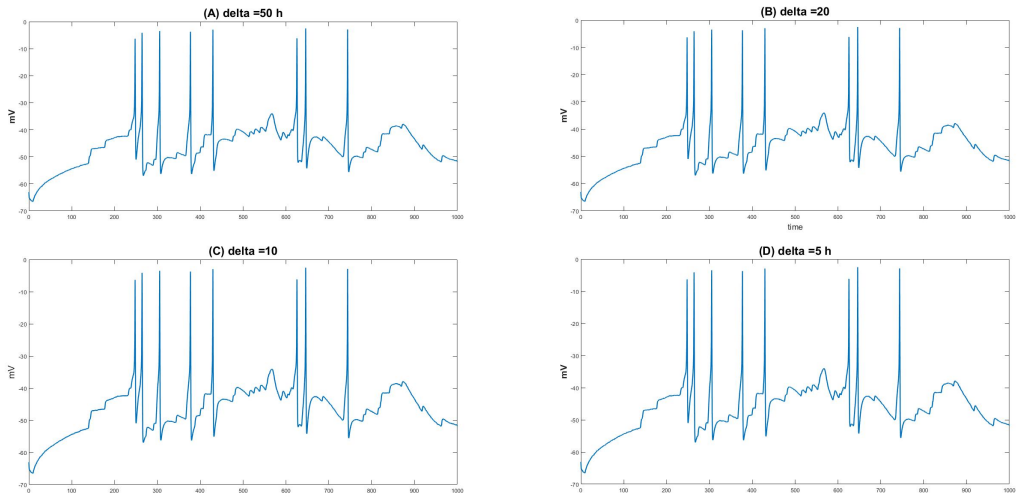

Fig. 4 The reconstructed signal for different values of the parameter δ (window width in which to evaluate the relationship between total and local variation).

negligible differences between the various reconstruction signals. The only difference concerns the height of the peaks of the reconstructed signal, which are lower than the original signal.
In the second numerical test we considered different types of noise. In particular we added colored noise signal with a power spectral density of $1/|f|^{\alpha}$ over its entire frequency range. In Figure 5 we show a signal with different noise: pink-noise ($\alpha = 1$), red-noise or Brownian noise ($\alpha = 2$), and the white noise or Gaussian noise ($\alpha = 0$). All these types of noise can appear in electrophysiological signals. In Figure 6 it is possible to observe the reconstructed signals, for each test we fixed both the $\delta = 20h$ parameter and $T = 2$. The worst

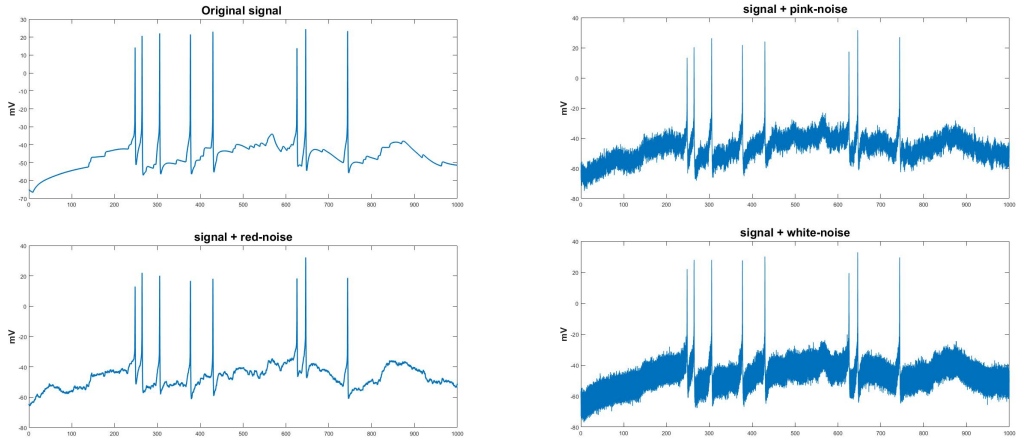

Fig. 5 A signal with different types of noise: pink-noise, red-noise, white-noise.

reconstruction seems to be in correspondence with the pink-noise but it must be emphasized that a more in depth experimentation would be needed depending on the specific characteristics of the noise. Finally we con-

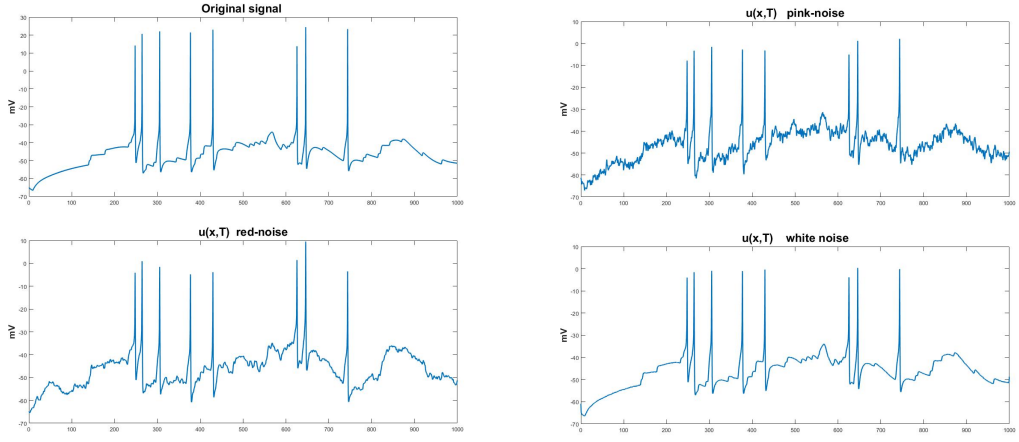

Fig. 6 Reconstructed signals in the presence of different types of noise: pink-noise, red-noise, white-noise.

sider a different signal that does not present too pronounced peaks. For this we have selected a row from a digital image (which we can consider as a "typical row", or a column, with different patterns and a Gaussian noise). Figure 7 shows the behavior of the signal evolution for different final times T. The numerical solution U seems to reach a steady state in which the monotonicity regions of the signal are preserved while the noise level is reduced.

3.3 1D Application

In this numerical experiment we consider a recorded calcium imaging data from a 3D cultures of cortical neurons. The sampling rate was 65Hz and the sampling time interval was about 8 seconds. The data was

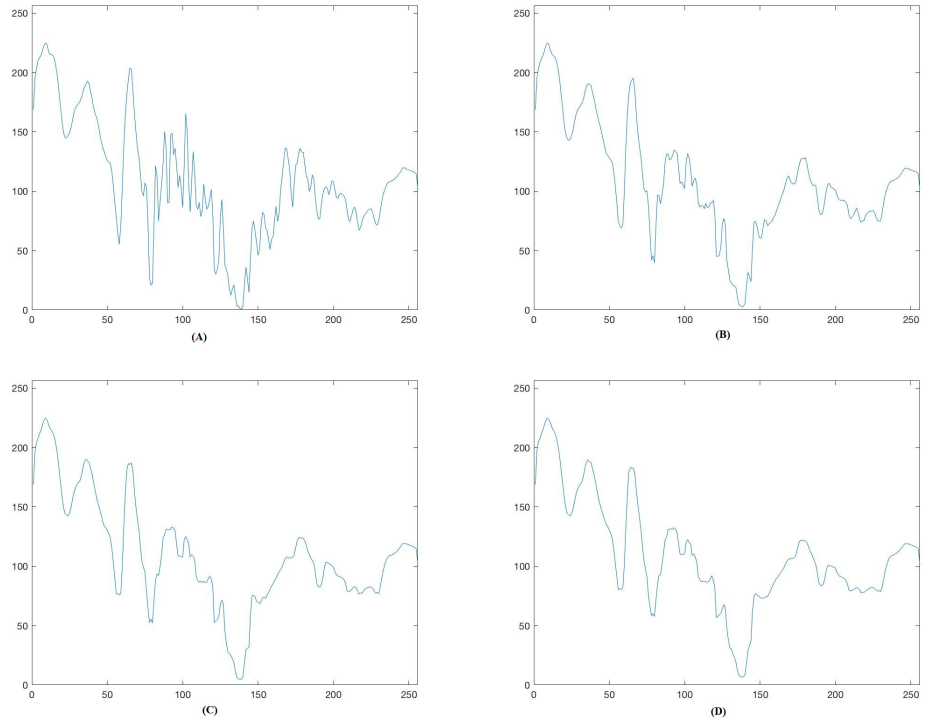

Fig. 7 Example of the numerical behavior of a row of a digital image: (A) the initial row, (B) the row after 100 iterations, (C) the numerical result after 300 iterations, (D) the row at the final time (500 iterations).

collected in the Department of Neuroscience and Brain Technologies of the Fondazione Istituto Italiano di Tecnologia. In Figure 8 we show a typical trace of the calcium signal together with different smoothed signals at different time T. In the test we used $\tau = 0.1$, $\delta = 20$, and $h = 1/65$, the initial data U^0 is obtained by the convolution of the original signal with a Gaussian filter with $\sigma = 0.01$.

4 Conclusion

Image denoising/smoothing is one of basic issues in image processing. It plays a key preliminary step in many computer based vision systems, but it is also a starting point towards more complex tasks. Since image noise removal represents a relevant issue in various image analysis and computer vision problems, it is a challenge to preserve the essential image features, such as edges and other sharp structures during the smoothing process. The feature preserving image noise reduction still represents a challenging image processing task. In this paper we propose a new method based on a nonlinear and nonlocal diffusion equation. The new approach has already been successfully applied in the analysis of membrane potentials in a neural network [1], and for data analysis of recorded calcium signals in a $3D$ culture cells [15]: these denoising experiments provided very encouraging results. Here we focused on the mathematical analysis of the model and its numerical approximation. In particular, we provided an existence theorem for the variational solution for the $1D$ and $2D$ case and a numerical scheme for the one-dimensional model.
We observe that the uniqueness of the solution of the novel equation remains an open problem for the $2D$ case (while can be guaranteed for the $1D$ case). Also the analysis of suitable numerical schemes should be completed. We have already developed some preliminary results that will be reported in a forthcoming paper. In particular, it is possible to show that the semi-implicit numerical scheme satisfies the same discrete scale-space

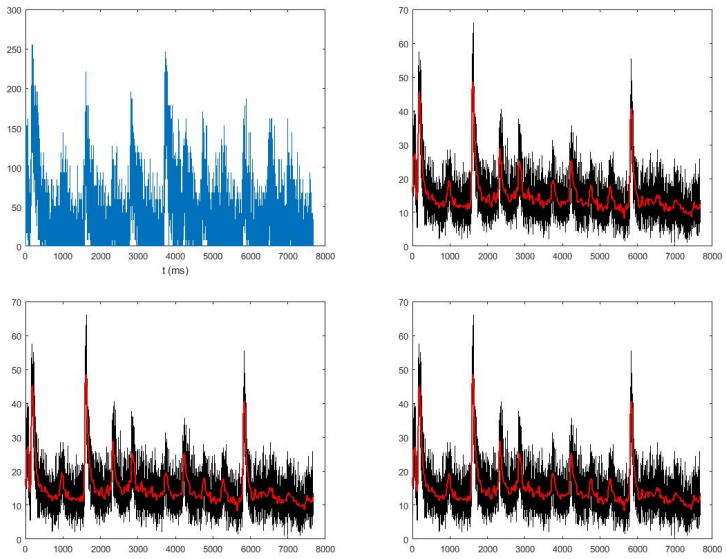

Fig. 8 Reconstruction with the non linear, non local diffusion equation (2) of a Calcium trace. The original signal was obtained in the IIT lab based in Genova (Italy). First line, on the left the original signal, on the right the solution for $T = 300$. Second line, on the left the solution for $T = 400$, on the right the solution for $T = 500$. In all numerical experiments, $\delta = 20$, $\tau = 0.1$.

properties as for the Perona-Malik method. Finally, a more complete comparison with other methods has to be done, but this goes beyond the aims of this paper.

Acknowledgments

The authors are also extremely grateful to T. Nieus (UniMI) for providing them with data of simulated membrane potentials, and to F. Difato (IIT-Ge) for providing the recorded calcium signals.

References

1. Aletti, G., Lonardoni, D., Naldi, G., Nieus, T.: From dynamics to links: a sparse reconstruction of the topology of a neural network. Commun. Appl. Ind. Math. **10**(2),2–11 (2019)
2. Alvarez, L., Guichard, F., Lions, P.-L., Morel, J.-M.: Axioms and fundamental equations of image processing. Arch. Rational Mech. Anal. **123**, 199–257 (1993)
3. Angenent, S., Pichon, E., Tannenbaum A.: Mathematical methods in medical image processing. Bull. Amer. Math. Soc. (N.S.) **43**, 365–396 (2006)
4. Aubert, G., Kornprobst, P.: Mathematical problems in image processing. vol. 147 of Applied Mathematical Sciences, Springer New York (2006)
5. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. Proc. IEEE CVPR **2**, 60–65 (2005)
6. Buades, A., Coll, B., Morel, J.M.: Neighborhood filters and PDE's. Numer. Math. **105**, 1–34 (2006)
7. Brezis, H.: Functional analysis, Sobolev spaces and partial differential equations. Universitext Springer-New York (2011)
8. Catté, F., Lions, P.-L., Morel, J.-M., Coll T.: Image selective smoothing and edge detection by nonlinear diffusion. SIAM J. Numer. Anal. **29**, 182–193 (1992)
9. Chambolle, A., Darbon J.: On total variation minimization and surface evolution using parametric maximum flows. International Journal of Computer Vision **84** 288–307 (2009)
10. Hummel, R. A.: Representations based on zero-crossings in scale-space, in Proceedings IEEE CVPR '86, 204–209 (1986)
11. Kačur, J.: Method of Rothe in evolution equations. vol. 80 of Teubner-Texte zur Mathematik Leipzig (1985)
12. Koenderink, J.: The structure of images, Biological Cybernetics **50**, 363–370 (1984)
13. McOwen, R.: Partial Differential Equations: Methods and Applications. Featured Titles for Partial Differential Equations Series, Prentice Hall (2003)

14. Meyer, Y.: Oscillating Patterns in Image Processing and in Some Nonlinear Evolution Equations. The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures, Univ. Lecture Ser. 22, AMS, Providence, RI, (2001)
15. Palazzolo, G., Moroni, M., Soloperto, A., Aletti, G., Naldi, G., Vassalli, M., Nieus, T., Difato, F.: Fast wide-volume functional imaging of engineered in vitro brain tissues, Scientific Reports **7**, Article number: 8499 (2017)
16. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **12**, 629–639 (1990)
17. Rabier, P., Thomas J.-M.: Exercices d'analyse numérique des équations aux dérivées partielles. Collection Mathématiques Appliquées pour la Maîtrise, Masson Paris, (1985)
18. G. Sapiro, G.: Geometric partial differential equations and image analysis. Cambridge University Press, Cambridge (2006)
19. Weickert, J.: Anisotropic diffusion in image processing. ECMI series, B. G. Teubner Stuttgart (1998)
20. Witkin, A.P.: Scale-space filtering. In Proceedings of the 8th International Joint Conference on Artificial Intelligence, vol. 2, William Kaufmann Inc, 1019–1022 (1983)
21. W. P. Ziemer, W. P.: Weakly differentiable functions. Vol. 120 of Graduate Texts in Mathematics, Springer-Verlag New York (1989)