Image Enhancement and Denoising by Complex Diffusion Processes

Guy Gilboa, Nir Sochen and Yehoshua Y. Zeevi

G. Gilboa and Y.Y. Zeevi are with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion city, Haifa 32000, Israel. e-mail: gilboa@tx.technion.ac.il, zeevi@ee.technion.ac.il

N. Sochen is with the Department of Applied Mathematics, University of Tel-Aviv Ramat-Aviv, Tel-Aviv 69978, Israel e-mail: sochen@math.tau.ac.il

Abstract

The linear and nonlinear scale spaces, generated by the inherently real-valued diffusion equation, are generalized to complex diffusion processes, by incorporating the free Schrödinger equation. A fundamental solution for the linear case of the complex diffusion equation is developed. Analysis of its behavior shows that the generalized diffusion process combines properties of both forward and inverse diffusion. We prove that the imaginary part is a smoothed second derivative, scaled by time, when the complex diffusion coefficient approaches the real axis. Based on this observation, we develop two examples of nonlinear complex processes, useful in image processing: a regularized shock filter for image enhancement and a ramp preserving denoising process.

Keywords: *scale-space, image filtering, image denoising , image enhancement, nonlinear diffusion, complex diffusion, edge detection, shock filters.*

I. INTRODUCTION

Two basic approaches to the implementation of PDEs in low level vision have been previously presented. The first is the axiomatic approach, formalized in the fundamental contribution of Alvarez et al [1]. In this and related studies (see for example [22] and references therein) a set of assumptions about the nature of the image and the required filtering is incorporated at the axiomatic level. The second approach is based on variational calculus, whereby the diffusion-like PDE is derived by a functional minimization process (see [3] and references therein).

In our study we take a fresh look at the application of PDEs in image processing and computer vision and propose a new, more general, framework. In various areas of physics and engineering, it was realized that extending the analysis from the real axis to the complex domain is very useful, even if the variables and/or quantities of interest are real. The analytical structure, in many cases, reveals important features of the system, which are difficult to account for by different means. Examples to this effect can be found in such unrelated subjects as the S-matrix elements in high-energy physics, and in the bread and butter of signal processing - the Fourier transform. Similarly the Gabor [8] the Gabor-wavelets [37], and Morlet-wavelet [16], [17] are complex-valued transforms. The latter is relevant to our study in that it incorporates a discrete set of scaled Gaussian filters and a set of scaled approximations of the Gaussian second derivative. All of these are examples of complex filters used in the processing of real signals.

Diffusion-like PDEs generalize filter theory. Relations between PDEs and filters, morphological and others, were pointed out recently by several groups [18], [23]. We follow in this paper

the idea of complexification and generalize it from filters to PDEs.

We generalize the linear scale spaces in the complex domain, by combining the diffusion equation with the free Schrödinger equation. A fundamental solution for the linear case is developed. Analysis of the linear complex diffusion shows that the generalized diffusion has properties of both forward and inverse diffusion. We thus obtain a *stable* flow that violates the maximum principle, while preserving other desirable mathematical and perceptual properties. The example of this flow may pave the way to a new class of diffusion like processes.

An important observation, supported theoretically and numerically, is that the imaginary part can serve as an edge detector (smoothed second derivative scaled by time), when the complex diffusion coefficient approaches the real axis. Based on this observation, we develop two examples of nonlinear complex processes for the denoising and the enhancement of images.

The paper is organized as follows: Various scientific domains that deal with, or are based on, complex-valued PDEs, are presented in Section 2 as motivating examples and as special cases of our framework. Previous studies in vision and signal processing using variants of complex-valued PDEs are reviewed in this section as well. The analysis of linear complex diffusion is presented in Section 3. Two non-linear cases are studied in Sections 4 and 5, taking advantage of the intuition gained from the analysis of the linear case, to construct nonlinear complex schemes for denoising of ramps and for enhancement by regularized shock filters. The advantages over known real-valued PDE-based algorithms, such as [2] and [21] are demonstrated in one- and two-dimensional examples.

II. Previous related studies

I this section we give some brief background on motivation for adopting the PDE-based approach to image processing, indicating how our new approach fits within this framework.

According to the axiomatic approach, a set of proper axioms leads to a PDE which is used, in turn, in denoising or for obtaining a scale-space. The axioms are of structural and morphological type. An important additional axiom is the maximum principle. Roughly speaking, the structural axioms ensure that the process be a sufficiently smooth semi-group. The maximum principle is the non-linear analog of filter positivity in linear filtering theory. The morphological axioms require commutativity of the semi-group flow with respect to a Lie group of interest. In their

original paper the authors analyze monotone transformations of gray value, and spatial affine transformations (and their sub-groups) [1].

The second method originates from the variational calculus approach to image processing and computer vision [3]. The diffusion-like PDE is accordingly derived using a functional minimization process. The canonical examples are the L_2 norm of the gradient, that leads to linear scale-space, and the L_1 norm of the gradient which leads to the total variation flow [34]. Whereas one cannot give a full account of the equations that results from this approach, it has been applied successfully in denoising, segmentation, stereo, optical flow, to name a few important applications [3]. This approach offers several important advantages. In particular, one can usually prove convergence as well as a relation to a systematic probabilistic approach, mainly to the Bayesian framework. The disadvantages are the possibility of convergence to a local minimum, and the fact that the set of PDEs, derived from a functional, is much smaller than the set of all possible diffusion- like PDEs. It is also difficult to consider the maximum principle by adopting this approach. We note that the equations introduced by us cannot be derived from a real valued functional.

The canonical, and most simple, example for both approaches is the linear scale-space equation

$$I_t = c\Delta I, \quad I|_{t=0} = I_0, \quad 0 < c \in \mathbb{R}. \tag{1}$$

The derivation of the linear scale space from the axiomatic approach can be found, for example, in [22]. But, it can be derived as a gradient descent of the L_2 norm of the gradient of the image.

The scale-space approach is by now a well established multi-resolution technique for image structure analysis (see for example [36],[20],[30]). The original Gaussian representation, introduced a scale dimension by convolving the original image with a Gaussian of a standard deviation $\sigma = \sqrt{2t}$. This is analogous to solving the linear diffusion equation with a constant diffusion coefficient $c = 1$.

The information distributed over all scales is generated as the solution $I(x, y, t)$ of the linear heat equation. Important cues, such as edges and critical points, are gathered from all scales in order to analyze the scene as a whole. One of the problems associated with this approach is that important structural features such as edges are smoothed and blurred along the flow, as the processed image evolves in time. As a consequence, the trajectories of zero crossings of the

second derivative, which indicate the locations of edges, vary from scale to scale.

To overcome this problem, Perona and Malik (P-M) [28] proposed a nonlinear adaptive diffusion process, termed anisotropic diffusion, wherein diffusion takes place according to a variable, adaptive, diffusion coefficient, to reduce the smoothing effect near edges. The same generic equation was proposed and analyzed by Rosenau [32], in the context of mathematical physics. The P-M nonlinear diffusion equation is of the form:

$$I_t = \nabla \cdot (c(|\nabla I|)\nabla I), \quad c(\cdot) > 0, \tag{2}$$

where c is a decreasing function of the gradient such as $(1+(|\nabla I|/k_{PM})^2)^{-1}$. It implies a direct relation between the image smoothness at a point and the image gradient. We further generalize the linear and nonlinear scale-spaces, and view them as special cases of a more general theory of complex diffusion-type processes.

Complex diffusion-type processes are commonly encountered, for example, in quantum physics and electro-optics [7], [26]. The time dependent *Schrödinger equation* is the fundamental equation of quantum mechanics. In the simplest case for a particle without spin, subjected to an external field it has the form

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi + V(x)\psi \quad , \tag{3}$$

where $\psi = \psi(t,x)$ is the wave function of a quantum particle, m is its mass, $\hbar$ is Planck's constant, $V(x)$ is the external field potential, Δ is the Laplacian and $i \doteq \sqrt{-1}$. With an initial condition $\psi|_{t=0} = \psi_0(x)$, requiring that $\psi(t,\cdot) \in L_2$ for each fixed t, the solution is $\psi(t,\cdot) = e^{-\frac{i}{\hbar}tH}\psi_0$, where the exponent is a shorthand for the corresponding power series, and the higher order terms are defined recursively by $H^n\Psi = H(H^{n-1}\Psi)$. The operator

$$H = -\frac{\hbar^2}{2m}\Delta + V(x), \tag{4}$$

called the *Schrödinger operator*, is interpreted as the energy operator of the particle under consideration. The first term is the kinetic energy and the second is the potential energy. The duality relations that exist between the Schrödinger equation and diffusion theory have been studied and are considered, for example, in [25].

It is very revealing to study the basic solution of the free, i.e. $V = 0$, "particle". Using separation of variables $\Psi(x,y,t) = \phi(t)\Phi(x,y)$, and simple manipulation of the equation, we

get

$$i\hbar\frac{\phi_t}{\phi} = -\frac{\hbar^2}{2m}\frac{\Delta\Phi}{\Phi} = E \ .$$

Since this equation is valid for all x, y and t, it is clear that E is a constant. The basic solution is therefore $\phi = \exp(-i\frac{E}{\hbar}t)$ and $\Phi = exp(\pm i\frac{\sqrt{2m}}{\hbar}k \cdot x)$, where $k \cdot k = E$. This implies that the basic solution is a plane wave! We will encounter this 'wavy behavior' in our complex flow.

Another important complex PDE in the field of phase transitions in traveling wave systems is the *complex Ginzburg-Landau equation* (CGL) [15]: $u_t = (1 + i\nu)u_{xx} + Ru - (1 + i\mu)|u|^2u$. Note that although these flows have a structure of a diffusion process, because of the complex coefficient, they retain also wave propagation properties.

In both cases of complex diffusion a non-linearity is introduced by adding a potential term, while the kinetic energy remains linear. In this study we employ the equation with zero potential (no external field) but with complex and non-linear "kinetic energy".

There are several examples of diffusion of complex-valued features in low-level vision (e.g. Whitaker and Gerig [35]; Barbaresco [4]; Kimmel, Malladi and Sochen [19]). Whitaker and Gerig generate a collection of band-passed images by means of Gabor filtering with specific set of frequencies. This vector-valued feature space was then smoothed linearly and in an anisotropic way. It is important to note that only the coefficient of the drift term (the first derivatives) becomes complex; This is a basic difference from our approach, as the qualitative behavior of a diffusion equation depends primarily on the coefficient (or tensor in the general case) of the diffusion term. It follows that the complex scale-space equation(s) that we present in this study are extremely different from the Whitaker and Gerig equations. A similar argument is relevant in reference to the approach presented by Kimmel,, Malladi and Sochen. In their study the coefficients of the Gabor-Morlet wavelet transform are smoothed by the Beltrami flow. Although the values of these filters are complex, the diffusion tensor is real and the behavior of the Beltrami flow is different from the one described in this paper. Another interesting work that study the diffusion of complex-valued functions is the one presented by Barbaresco [4]. This study is concerned, however, with complex curves using a variational technique.

III. Linear Complex Diffusion

A. Problem Definition

We consider the following initial value problem:

$$\begin{aligned} I_t &= cI_{xx}, \qquad t>0, \qquad x\in\mathbb{R} \\ I(x;0) &= I_0\in\mathbb{R}, \quad c, I\in\mathbb{C}. \end{aligned} \tag{5}$$

This equation unifies the linear *diffusion* equation (1) for $c\in\mathbb{R}$ and the free *Schrödinger* equation, i.e. $c\in\mathbb{I}$ and $V(x)\equiv 0$. When $c\in\mathbb{R}$ there are two cases: for $c>0$ the process constitutes a well-posed forward diffusion, whereas for $c<0$ an ill-posed inverse diffusion process is obtained. In the general case the initial condition I_0 is complex. In this paper we discuss the particular case of real initial conditions, where I_0 is the original image.

B. Fundamental Solution

We seek the complex fundamental solution $h(x;t)$ that satisfies the relation:

$$I(x;t) = I_0 * h(x;t), \tag{6}$$

where $*$ denotes convolution. Rewriting the complex diffusion coefficient as $c \doteq re^{i\theta}$, since there does not exist a stable fundamental solution of the inverse diffusion process, restrict the analysis to a positive real value of c, that is $\theta\in(-\frac{\pi}{2},\frac{\pi}{2})$. Replacing the real time variable t by the complex time $\tau=ct$, yields $I_\tau = I_{xx}$, $I(x;0)=I_0$. This is identical in its form to the linear real-valued diffusion equation. Its fundamental solution, therefore, is in a Gaussian form. In order to satisfy the initial condition $I(x;0)=I_0$ we require

$$\begin{aligned} &(a) \quad \textstyle\int_{-\infty}^{\infty} h(x;t\to 0)dx = 1, \\ &(b) \quad \textstyle\int_{|x|>\epsilon} |h(x;t\to 0)|dx \to 0, \quad \text{where } \epsilon=\epsilon(t),\ \lim_{t\to 0}\epsilon(t)=0. \end{aligned} \tag{7}$$

This leads to the following fundamental solution:

$$h(x;t) = Ag_\sigma(x;t)e^{i\alpha(x;t)}, \tag{8}$$

where $g_\sigma(x;t)=\frac{1}{\sqrt{2\pi}\sigma(t)}e^{-x^2/2\sigma^2(t)}$, and

$$A=\frac{1}{\sqrt{\cos\theta}}, \quad \alpha(x;t)=\frac{x^2\sin\theta}{4tr}-\frac{\theta}{2}, \quad \sigma(t)=\sqrt{\frac{2tr}{\cos\theta}}. \tag{9}$$

For more details see Appendix F.

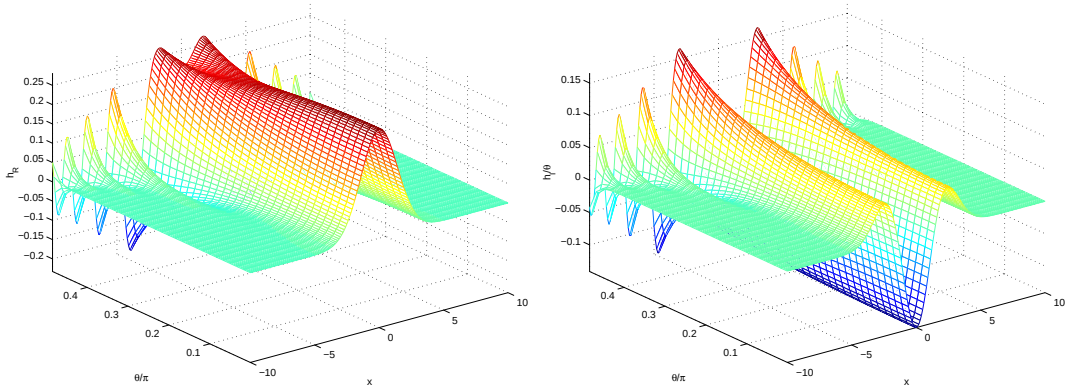

Fig. 1. Fundamental solution $h_\theta(x;t)$ as a function of x and θ ($t = 1$). Left - real part (h_R), right - imaginary part normalized by θ (h_I/θ).

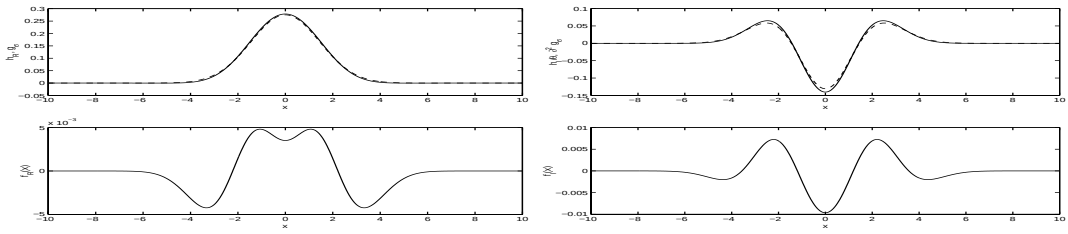

Fig. 2. Top: right - h_R (full line) and g_σ (dashed line) as a function of x, left - h_I/θ (full line) and $\frac{\partial^2}{\partial x^2} g_\sigma t$ (dashed line) as a function of x. Bottom: left - the difference function $h_R - g_\sigma$, right - the difference function $h_I/\theta - \frac{\partial^2}{\partial x^2} g_\sigma t$. $\theta = \frac{\pi}{10}$.

C. *Approximate Solution for Small Theta*

We will now show that as $\theta \to 0$, the imaginary part can be regarded as a smoothed second derivative of the initial signal, factored by θ and the time t. Generalizing the solution to any dimension in Cartesian coordinates $\mathbf{x} \doteq (x_1, x_2, ...x_N) \in \mathbb{R}^N$, $I(\mathbf{x};t) \in \mathbb{C}^N$, and denoting that in this coordinate system $\mathbf{g}_\sigma(\mathbf{x};t) \doteq \prod_i^N g_\sigma(x_i;t)$, we show that:

$$\lim_{\theta \to 0} \frac{Im(I)}{\theta} = t\Delta \mathbf{g}_{\tilde{\sigma}} * I_0, \tag{10}$$

where $Im(\cdot)$ is the imaginary value and $\tilde{\sigma} = lim_{\theta \to 0}\sigma = \sqrt{2t}$. Restricting the analysis, for convenience, to a unitary complex diffusion coefficient $c = e^{i\theta}$, utilizing the approximation $cos\theta = 1 + O(\theta^2)$ and $sin\theta = \theta + O(\theta^3)$ for small θ, and introducing an operator $\tilde{H} = c\Delta$, equation (5) can be written (for higher dimensional systems too) as: $I_t = \tilde{H}I$; $I|_{t=0} = I_0$. The

solution $I = e^{t\tilde{H}} I_0$, is the equivalent of (6) and (8). The above approximations yield:

$$\begin{aligned} I(\mathbf{x}, t) &= e^{ct\Delta} I_0 = e^{e^{i\theta} t\Delta} I_0 \\ &\approx e^{(1+i\theta)t\Delta} I_0 = e^{t\Delta} e^{i\theta t\Delta} I_0 \\ &\approx e^{t\Delta}(1 + i\theta t\Delta) I_0 = (1 + i\theta t\Delta)\mathbf{g}_{\tilde{\sigma}} * I_0. \end{aligned}$$

Further insight into the behavior of the small-theta approximation can be gained by separating the real and imaginary parts of the signal, $I = I_R + iI_I$, and diffusion coefficient, $c = c_R + ic_I$, into a set of two equations:

$$\begin{cases} I_{Rt} &= c_R I_{Rxx} - c_I I_{Ixx} \quad , I_R|_{t=0} = I_0 \\ I_{It} &= c_I I_{Rxx} + c_R I_{Ixx} \quad , I_I|_{t=0} = 0, \end{cases} \tag{11}$$

where $c_R = \cos\theta$, $c_I = sin\theta$. The relation $I_{Rxx} \gg \theta I_{Ixx}$ holds for small enough θ, which allows us to omit the second term on the r.h.s. of the first equation, to get the small theta approximation:

$$I_{Rt} \approx I_{Rxx} \quad ; \qquad I_{It} \approx I_{Ixx} + \theta I_{Rxx}. \tag{12}$$

In (12) I_R is controlled by a linear forward diffusion equation, whereas I_I is affected by both the real and imaginary equations. We can regard the imaginary part as $I_{It} \approx \theta I_{Rxx}+$("a smoothing process"). Note that since the initial condition is real valued, the term θI_{Rxx} is dominant and cannot be omitted even for very small θ (at $t = 0$ it is infinitely larger than I_{Ixx} as $I_I|_{t=0} \equiv 0$).

D. Analysis of the Fundamental Solution

We consider a few properties with reference to the fundamental solution, and derive bounds on error under the small θ approximation. The approximation of the real part to a Gaussian, and of the imaginary part to its second derivative scaled by time, obtained for small θ is of the order $O(\theta^2)$. Proofs and calculations of the bounds are provided in the Appendix. Whereas more technical details can be found in our report [10]. Here we limit the presentation to the summary of our results.

The kernel can be separated into its real and imaginary parts. As the initial condition I_0 is real-valued in this study, the real part of $I(x;t)$ is affected only by the real kernel and the imaginary part of $I(x;t)$ is affected only by the imaginary kernel:

$$I(x;t) = I_R + iI_I = I_0 * h = I_0 * h_R + iI_0 * h_I, \tag{13}$$

where $h = h_R + ih_I$. (we get $I_R = I_0 * h_R$, $I_I = I_0 * h_I$).

The nature of the complex kernel does not change through the evolution; the kernel is basically rescaled according the time t (or to σ). Therefore we can analyze a few characteristics of the kernel as a function of σ for different values of θ. In the sequel we present some of the major characteristics of the real and imaginary kernels.

1) Properties of the real kernel h_R:

1) **Kernel formulation**

$$h_R(x;t) = Ag_\sigma(x;t)\cos\alpha(x;t). \tag{14}$$

2) **Maximal amplification**

$$\frac{\max_{x,t}|I_0 * h_R|}{\max_x|I_0|} \leq A. \tag{15}$$

For small theta we get $A = \cos^{-1/2}\theta = 1 + \frac{\theta^2}{4} + O(\theta^4)$.

3) **Effectively positive kernels**. One requirement of the linear scale-space is to avoid creation of new local extrema along the scale-space in 1D. Kernels obeying this requirement should be positive everywhere. In 1D this is equivalent to the requirement that the operator be causal [22]. As this kernel is not positive everywhere, we check how close is it to a positive kernel. Let us define a positivity measure $-1 \leq P_h \leq 1$ of a kernel h as follows:

$$P_h \doteq \frac{\int_{-\infty}^{\infty} h(x)dx}{\int_{-\infty}^{\infty}|h(x)|dx}. \tag{16}$$

We regard a kernel h as effectively positive with the measure $\epsilon \ll 1$, if $P_h \geq 1 - \epsilon$. For h_R we get the following bound

$$P_{h_R} \geq \frac{1 - 8\Phi(-x_1)}{1 + 8\Phi(-x_1)}, \tag{17}$$

where

$$\Phi(x) = \int_{-\infty}^{x} g_{\sigma=1}(s)ds \tag{18}$$

and

$$x_1 = \sqrt{(\frac{2}{3}\pi + \theta)\cot\theta}. \tag{19}$$

This bound is valid for $1.28_{rad} = 73^\circ \geq \theta \geq 0$. Example (a): for $\theta = 1^\circ = \frac{\pi}{180}$ we get $x_1 \approx 11$, and $P_{h_R} \geq \frac{1-8\Phi(-11)}{1+8\Phi(-11)} = 1 - 2 * 10^{-27}$. Example (b): for $P_{h_R} > 0.99999, \quad (\epsilon < 10^{-5})$ we require $\theta < 5^\circ$.

4) **Small theta approximation**. We define the distance between convolution kernels as the norm of their difference operator: $d(h,g) \doteq ||T_{h-g}||_\infty$, where the norm of a linear operator (using the ∞ norm) is

$$||T_h||_\infty \doteq \sup_{||f||\neq 0} \frac{||T_h f||_\infty}{||f||_\infty} = \sup_{||f||\neq 0} \frac{||h * f||_\infty}{||f||_\infty}. \tag{20}$$

The distance between the real kernel $h_R(x;t)$ and a Gaussian $g_\sigma(x;t)$ is

$$d(h_R, g_\sigma) = O(\theta^2), \tag{21}$$

and for small values of theta the distance is bounded by:

$$d(h_R, g_\sigma) < 0.5\theta^2, \quad \forall\theta \in [0, \frac{\pi}{10}], \forall t \geq 0 \tag{22}$$

5) **Definite integral**

$$\int_{-\infty}^{\infty} h_R(x;t)dx = 1 \quad . \tag{23}$$

2) Properties of the imaginary kernel h_I:

1) **Kernel formulation**

$$h_I(x;t) = A g_\sigma(x;t) \sin\alpha(x;t). \tag{24}$$

2) **Maximal amplification**

$$\frac{\max_{x,t}|I_0 * h_I|}{\max_x |I_0|} \leq A. \tag{25}$$

3) **Small theta approximation**. The distance between the imaginary kernel, normalized by θ, and a Gaussian's second derivative scaled by time, is

$$d(h_I/\theta, \frac{\partial^2}{\partial x^2} g_\sigma t) = O(\theta^2), \tag{26}$$

and for small values of theta the distance is bounded by:

$$d(h_I/\theta, \frac{\partial^2}{\partial x^2} g_\sigma t) < 0.5\theta^2, \quad \forall\theta \in [0, \frac{\pi}{10}], \forall t \geq 0 \tag{27}$$

4) **Definite integral**

$$\int_{-\infty}^{\infty} h_I(x;t)dx = 0 \quad . \tag{28}$$

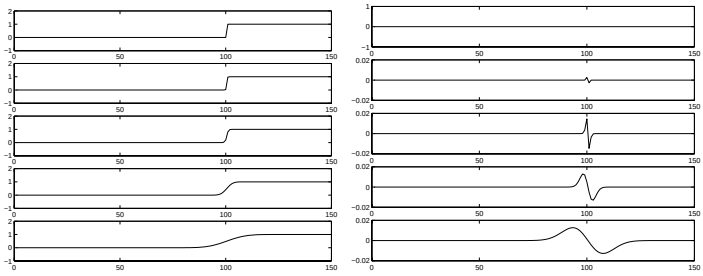

Fig. 3. Complex diffusion of a small theta, $\theta = \pi/30$, applied to a step signal. Left frame - real values, right frame - imaginary values. Each frame depicts from top to bottom: original step, diffused signal after times: 0.025, 0.25, 2.5, 25.

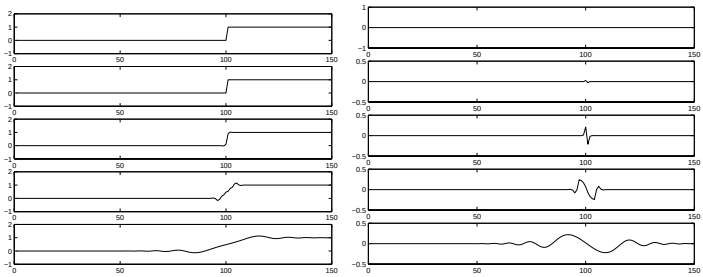

Fig. 4. Complex diffusion of a large theta, $\theta = 14\pi/30$, applied to a step signal. Left frame - real values, right frame - imaginary values. Each frame depicts from top to bottom: original step, diffused signal after times: 0.025, 0.25, 2.5, 25.

E. Examples

We present examples of 1D and 2D signal processing with linear complex diffusion, characterized by small and large values of θ. In Fig.'s (3) and (4), we depict the evolution of a unit step, processed by a complex diffusion of small and large θ ($\frac{\pi}{30}, \frac{14\pi}{30}$), respectively. The same θ values are used in the processing of the the camera-man image (Fig.'s. (5) and (6), respectively). The qualitative properties of the edge detection (smoothed second derivative) are clearly apparent in the imaginary part of the signals, for the small θ value, whereas the real value depicts the properties of ordinary Gaussian scale-space. For large θ, however, the imaginary part feeds back into the real part significantly, creating wave-like ringing. In addition, the signal overshoots and undershoots, exceeding the original maximum and minimum values and thereby violating the "Maximum-minimum" principle – a property suitable for sharpening purposes, similar to the Mach Bands characteristic of vision [31].

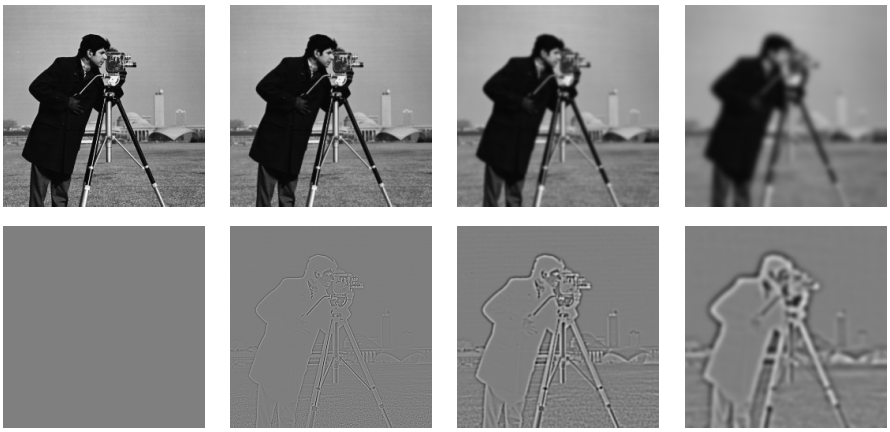

Fig. 5. Complex diffusion with small theta ($\theta = \pi/30$), applied to the camera-man image. Top images – real values, bottom images – imaginary values (factored by 20). Each frame (from left to right): original image, result obtained after processing time 0.25, 2.5, 25, respectively.

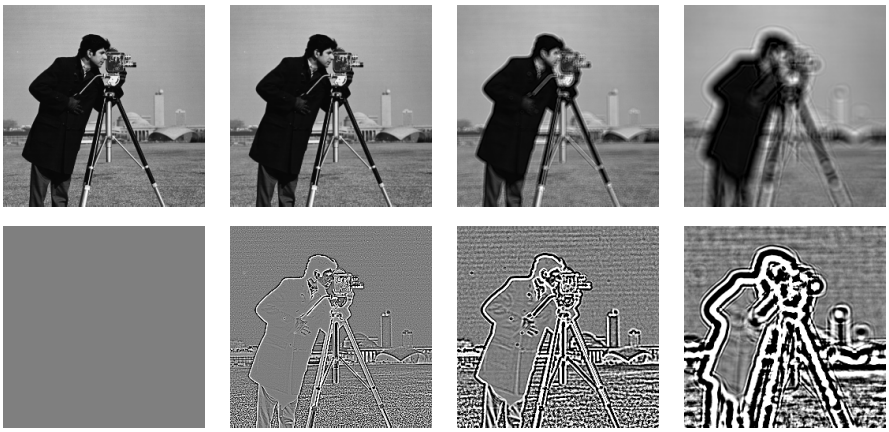

Fig. 6. Complex diffusion with large theta ($\theta = 14\pi/30$), applied to the cameraman image. Top sequence of images – real values, bottom sequence – imaginary values (factored by 20). Each sequence depicts from left to right the original image and the results of the processing after t=0.25, 2.5, and 25, respectively.

F. Generalization to Nonlinear Complex Diffusion

Nonlinear complex processes can be derived, based on the properties of the linear complex diffusion, to match the requirements of specific applications. We present two such nonlinear schemes, developed for application in image denoising and enhancement.

IV. Ramp Preserving Denoising

Ramp functions can be used as a model of the basic structure of edges in images or their equivalent 1D functions. Step-type (singular) edges are a limiting case of ramp functions. Visual responses to ramp functions have been widely investigated both psychophysically and physiologically. In particular they are known for the Mach bands associated with their perception [31]. Ramp-type edges are generic fundamental components of images, and as such, are extracted in the raw primal sketch of images [24]. It is therefore of special interest and importance to compare the action of a nonlinear complex-diffusion equation on a ramp function with those of real non-linear diffusion and other previously proposed operators.

We are looking for a general nonlinear diffusion equation

$$I_t = \frac{\partial}{\partial x}\left(c(\cdot)I_x\right) \tag{29}$$

that preserves smoothed ramps.

As was the case with real nonlinear diffusion processes, we search here too for a suitable differential operator $\mathcal{D}$ for ramp edges. Eq. (29) with a diffusion coefficient $c(|\mathcal{D}I|)$, which is a decreasing function of $|\mathcal{D}I|$, can be regarded as a ramp preserving process. Examining the gradient, as a possible candidate, leads to the conclusion that it is not a proper measure for two reasons: The gradient does not detect the ramp main features - namely its endpoints; Moreover, it has a nearly uniform value across the whole smoothed ramp, causing a nonlinear gradient-dependent diffusion to slow the diffusion process in that region and therefore being less effective in noise reduction within the ramp-edge. The second derivative (Laplacian in more than one dimension) is a more suitable choice: It has a high magnitude near the endpoints and low magnitude elsewhere, and thus enables the nonlinear diffusion process to reduce noise over the ramp.

We formulate $c(s)$ as a decreasing function of s:

$$c(s) = \frac{1}{1+s^2}, \text{ where } c(s) = c(|I_{xx}|), \tag{30}$$

and apply it in (29) to yield:

$$I_t = \frac{\partial}{\partial x}\left(\frac{I_x}{1+I_{xx}^2}\right) = \frac{1+I_{xx}^2-2I_xI_{xxx}}{(1+I_{xx}^2)^2}I_{xx}. \tag{31}$$

Two main problems are associated with this scheme. The first and more severe one is the fact that noise has very large (theoretically unbounded) second derivatives. Secondly, a numerical problem arises when third order derivatives are computed, with large numerical support and noisier derivative estimations. These two problems are solved by using this nonlinear complex diffusion.

Following the results of the linear complex diffusion (Eq. 10), we implement the imaginary value of the signal (divided by θ) in controlling the diffusion process.

Whereas for small t this term vanishes, allowing stronger diffusion to reduce the noise, with time its influence increases and preserves the ramp features of the signal.

The equation for the multidimensional process is

$$\begin{aligned} I_t &= \nabla \cdot (c(Im(I))\nabla I), \\ c(Im(I)) &= \frac{e^{i\theta}}{1+\left(\frac{Im(I)}{k\theta}\right)^2} \end{aligned} \tag{32}$$

where k is a threshold parameter. For the same reasons discussed in the linear case, here too the phase angle θ should be small ($\theta << 1$). Since the imaginary part is normalized by θ, the process is almost not affected by changing the value of θ, as long as it stays small ($\theta < 5^\circ$). The discrete implementation details are in Appendix I.

In Figs. 7 and 8 we compare denoising of a one dimensional ramp signal by a P-M process (Eq. 2), with the performance of the above process (Eq. 32). This example illustrates that the staircasing effect, characteristic of the P-M process, does not occur in processing by our nonlinear complex scheme. In Fig. 9 a two dimensional box with varying illumination was processed. We tried to demonstrate how denoising by our process can perform better in changing illumination conditions. We also show how the imaginary value can be of use for segmentation in such circumstances. With regard to the P-M and similar gradient-controlled processes, it is demonstrated that overcoming the staircasing effects by increasing the threshold k_{PM} (thus causing the gradient magnitude of the illumination to be in the convex regime of the process) comes at a cost of severely degrading the edges. In Figs. 10 and 11 two face images were processed (where there are typically shadings and illumination changes). To the first figure (part of the Barbara image) additive white Gaussian noise was added synthetically (SNR=20dB) whereas the Mona Lisa image is a low quality JPEG image with visible artifacts. The ramp-

preserving process smoothes gradual changes well, yet preserves edges (though not as strongly as the P-M). The staircasing effect of the P-M process (that can create false edges) are apparent. Note also the JPEG artifacts near the eyes (Fig. 11) that were removed by our process.

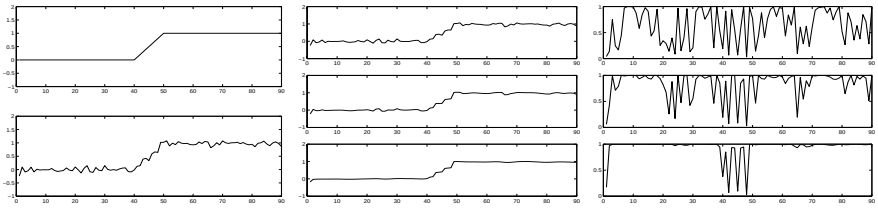

Fig. 7. Perona-Malik nonlinear diffusion process applied to a ramp-type soft edge ($k_{PM} = 0.1$). Left - original (top) and noisy ramp signal (white Gaussian, SNR=15dB) . Middle - denoised signal at times $0.25, 1, 2.5$, from top to bottom, respectively. Right - respective values of c coefficient.

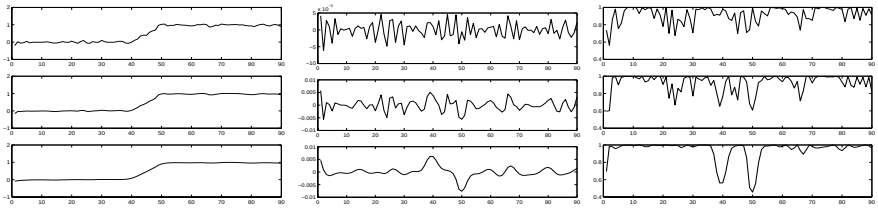

Fig. 8. Nonlinear complex diffusion process applied to a ramp-type soft edge ($\theta = \pi/30$, $k = 0.07$). Left - real values of denoised signal at times $0.25, 1, 2.5$, from top to bottom, respectively. Middle - respective imaginary values, right - respective real values of c.

V. Regularized Shock Filters

Most of the research concerning the application of partial differential equations in the fields of computer vision and image processing focused on parabolic (diffusion-type) equations. In [27] Osher and Rudin proposed a hyperbolic equation, called shock filter, that can serve as a stable deblurring algorithm approximating deconvolution (Fig. 12).

A. Problem Statement

The formulation of the shock filter equation is:

$$I_t = -|I_x| F(I_{xx}), \tag{33}$$

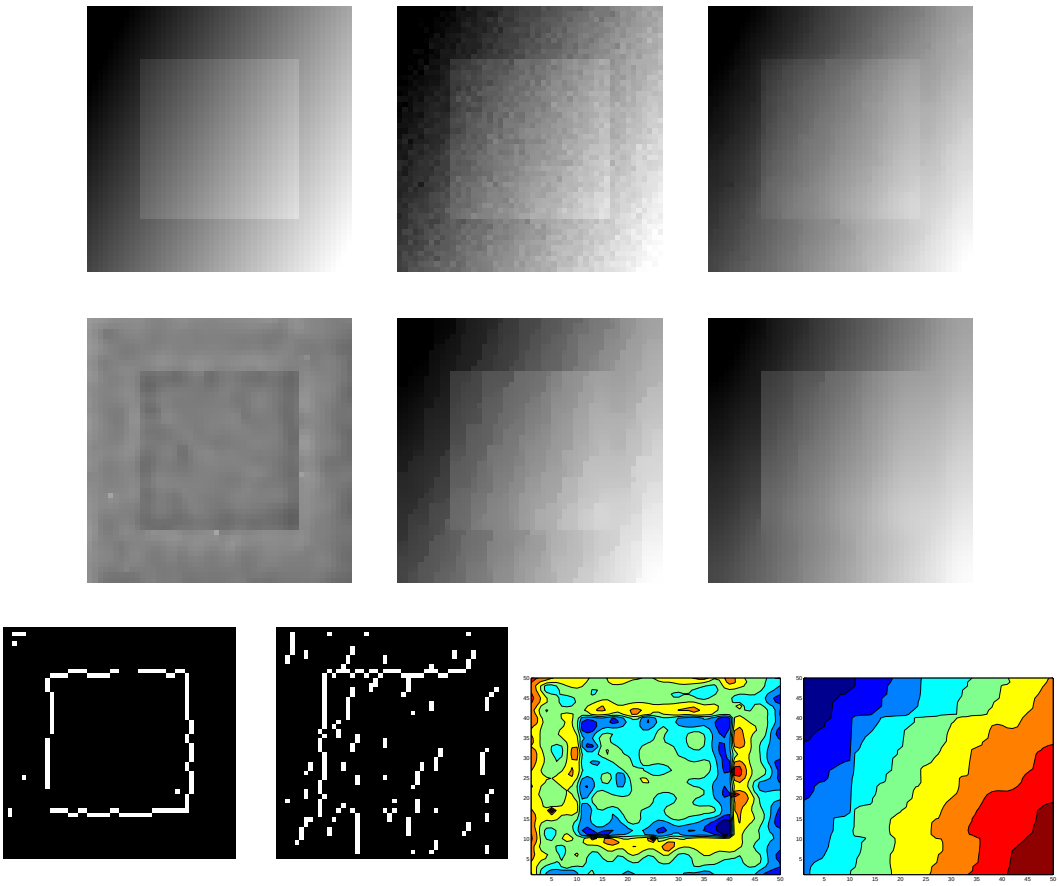

Fig. 9. Filtering of a synthetic test image of ramp varying illumination. Top row (from left): Original image, with noise (SNR=5dB) and linearly varying illumination ($I = I + 4x + 2y$), ramp-preserving denoising - real part ($k = 2$); Middle row (from left): ramp-preserving denoising - imaginary part, Perona-Malik with lower threshold ($k_{PM} = 3$), Perona-Malik with higher threshold ($k_{PM} = 5$); Bottom row (from left): Thresholded gradient of imaginary part ramp preserving, thresholded gradient of P-M (low threshold), level-sets of imaginary part and P-M.

where F should satisfy $F(0) = 0$, and $F(s)\text{sign}(s) \geq 0$. [1]

Choosing $F(s) = \text{sign}(s)$ yields the classical shock filter equation:

$$I_t = -\text{sign}(I_{xx})|I_x|, \tag{34}$$

generalized in the 2D case to:

$$I_t = -\text{sign}(I_{\eta\eta})|\nabla I|, \tag{35}$$

where η is the direction of the gradient.

[1] Note that the above equation and all other evolutionary equations in this section have initial conditions $I(x, 0) = I_0(x)$ and Neumann boundary conditions ($\frac{\partial I}{\partial n} = 0$ where n is the orientation perpendicular to the boundary).

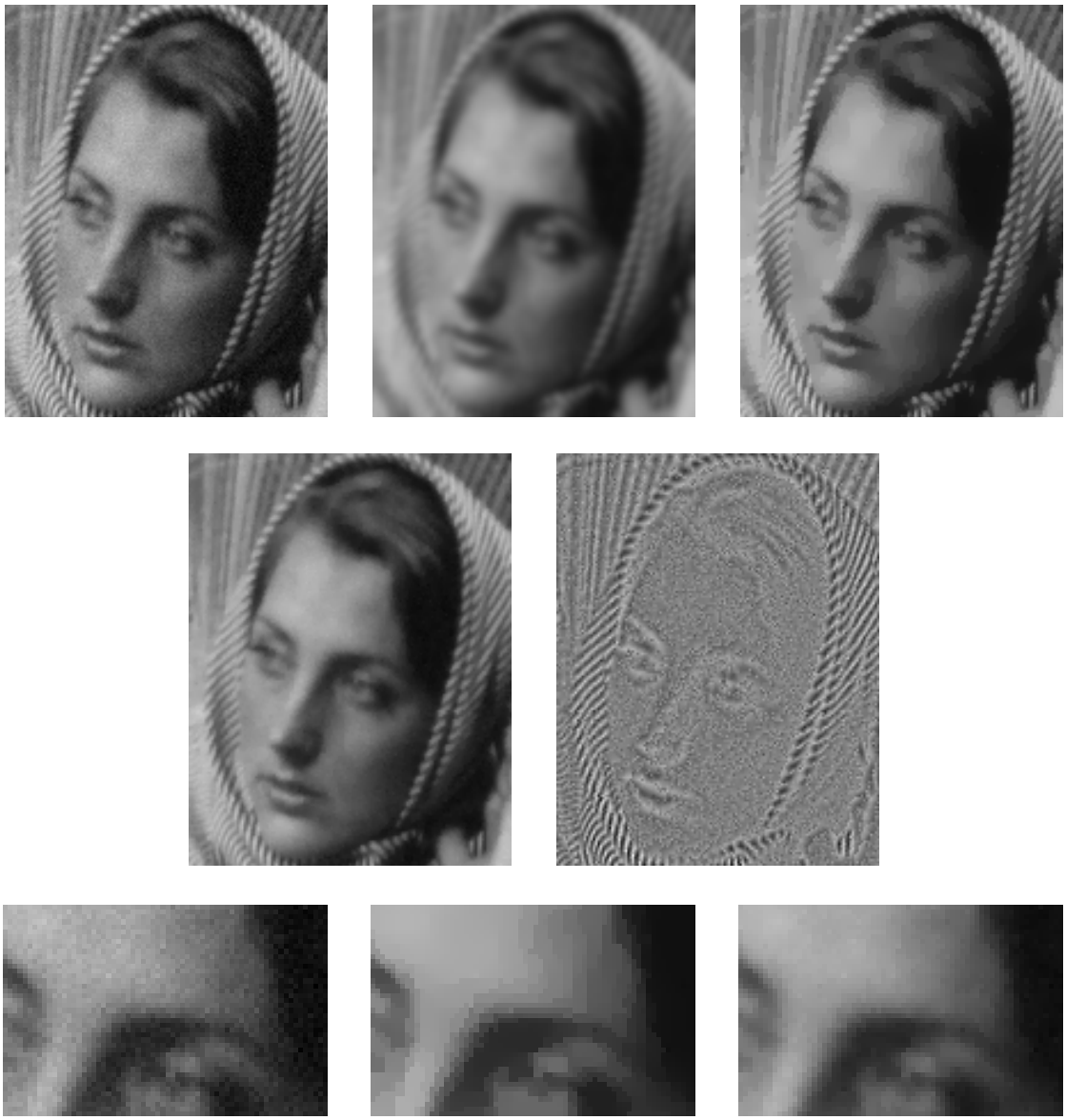

Fig. 10. Filtering of a face (Barbara image). Top row (from left): Noisy image (SNR=20dB), result of filtering with a linear diffusion process, filtering with Perona-Malik diffusion process; Middle row (from left): Ramp-preserving denoising - real part, imaginary part; Bottom row: Enlargement of forehead (from left) - original, P-M, ramp-preserving real part.

The main properties of the shock filter are:

- Shocks develop at inflection points (zero crossings of second derivative).
- Local extrema remain unchanged in time. No erroneous local extrema are created. The scheme is total-variation-preserving (TVP).
- The steady state (weak) solution is piecewise constant (with discontinuities at the inflection

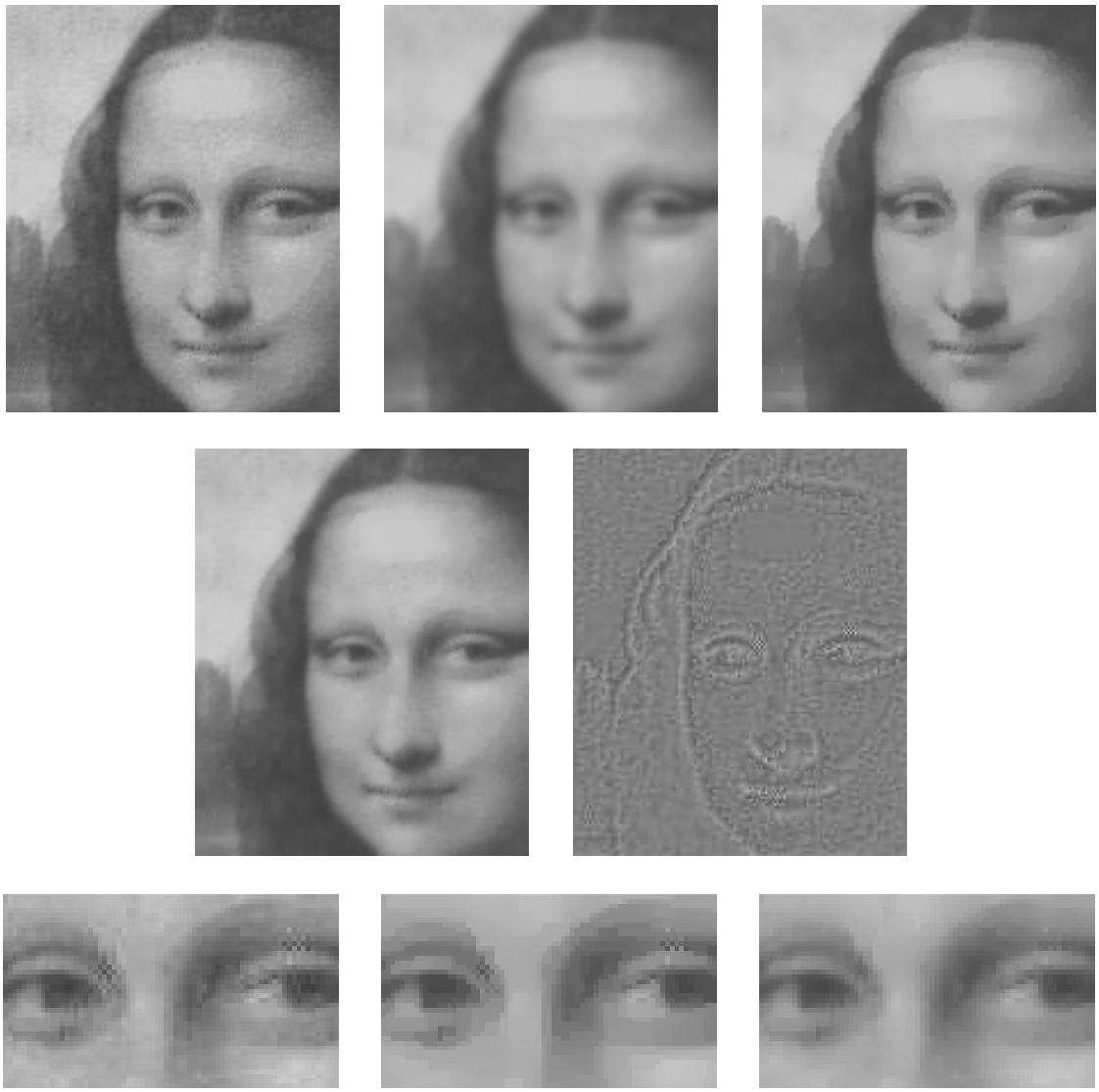

Fig. 11. Filtering a low quality JPEG of the Mona Lisa image. Top row (from left): Original image, result of filtering with a linear diffusion process, filtering with Perona-Malik diffusion process; Middle row: Ramp-preserving denoising - real part, imaginary part, Bottom row: Enlargement of the eyes - original, P-M, ramp-preserving real part.

points of I_0).

- The process approximates deconvolution.

As noted already in the original paper, any noise in the blurred signal will also be enhanced. As a matter of fact, this process is extremely sensitive to noise. Theoretically, in the continuous domain, any white noise added to the signal may add an infinite number of inflection points, disrupting the process completely. Discretization may help somewhat, but the basic sensitivity

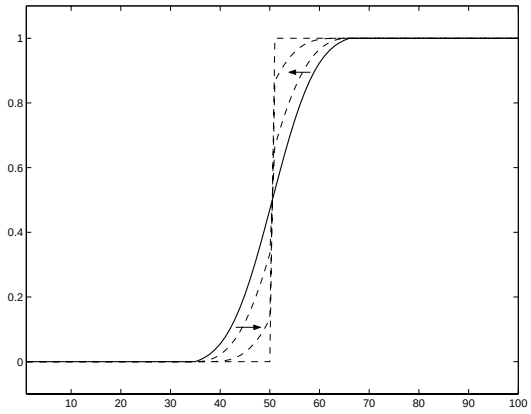

Fig. 12. Shock filter operation: deblurring a blurred step edge. Solid line - blurred step edge. Dashed lines - 3 steps in the evolution of the PDE towards formation of a shock in the location of the inflection point.

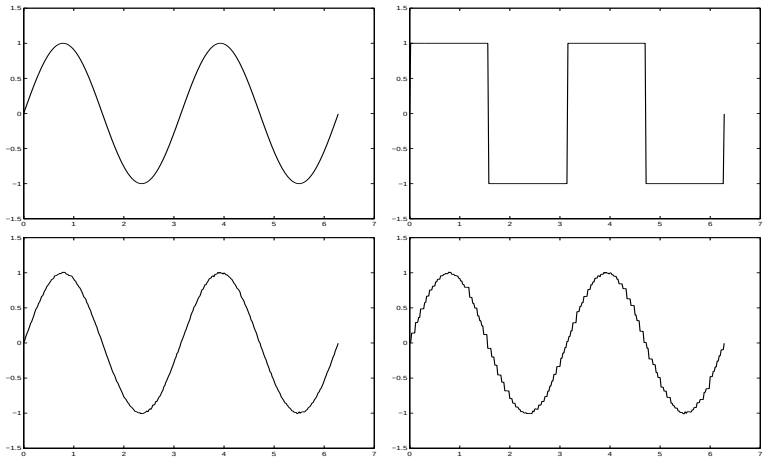

Fig. 13. A noiseless sinewave signal (top left) and the steady state of its processing by a shock filter (top right) are compared with the processing of a noisy signal generated by adding low level of white Gaussian noise– SNR=40dB (bottom left). The steady state of the processed noisy signal does not depict any enhancement and the only result is noise amplification (bottom right).

to noise persists. This is illustrated by comparison of the processing of a noiseless and a noisy sine wave signals (Fig. (13)). Whereas in the case of a noiseless signal the shock filter well enhances the edges, turning a sine wave into a square-wave signal, in the noisy case - the shock filter does not enhance the edges at all, and the primary result of the processing is amplification of noise, although only a very low level of white Gaussian noise was added to the input signal (SNR=40dB).

B. Previous Related Studies

The noise sensitivity problem is critical and, unless properly solved, will continue to hinder most practical applications of shock filters. Previous studies addressing this issue came up with several plausible solutions. The common approach to increase robustness ([2], [6], [21], [33]), is to convolve the signal's second derivative with a lowpass filter, such as a Gaussian:

$$I_t = -\text{sign}(G_\sigma * I_{xx})|I_x|, \tag{36}$$

where G_σ is a Gaussian of standard deviation σ.

This is generally not sufficient to overcome the noise problem: convolving the signal with a Gaussian of moderate width, does not cancel in many cases the inflection points produced by the noise; Their magnitude becomes considerably lower, but there is still a change of sign at these points, which induces flow in opposite direction on each side of the inflection point. For very wide (large scale) Gaussians, most inflection points produced by the noise are diminished but at a cost: the location of the signal's inflection points become less accurate. Moreover, the effective Gaussian's width σ often exceeds the signal's extent, thus causing the boundary conditions imposed on the process to strongly affect the solution. Lastly, from a computational viewpoint, the convolution process in each iteration is costly.

A more complex approach, is to address the issue as an enhancing-denoising problem: smoother parts are denoised, whereas edges are enhanced and sharpened. The main idea is to add some sort of anisotropic diffusion term with an adaptive weights of the shock and the diffusion processes. Alvarez and Mazorra were the first to couple shock and diffusion, proposing an equation of the form [2]:

$$I_t = -\text{sign}(G_\sigma * I_{\eta\eta})|\nabla I| + cI_{\xi\xi}, \tag{37}$$

where c is a positive constant and ξ is the direction perpendicular to the gradient ∇I. This equation, though, degenerates to (36) in the 1D case and the contribution of the diffusion to the combined process is lost.

A more advanced scheme, was proposed by Kornprobst et al. [21]:

$$I_t = \alpha_r(h_\tau I_{\eta\eta} + I_{\xi\xi}) - \alpha_e(1 - h_\tau)\text{sign}(G_\sigma * I_{\eta\eta})|\nabla I|, \tag{38}$$

where $h_\tau = h_\tau(|G_{\tilde{\sigma}} * \nabla I|) = 1$ if $|G_{\tilde{\sigma}} * \nabla I| < \tau$, and 0 otherwise. In [21] the scheme includes

another fidelity term $\alpha_f(I - I_0)$ that is omitted here (since such a term can be added to any scheme).

Another modern scheme was proposed by Coulon and Arridge [6],

$$I_t = \mathrm{div}(c\nabla I) - (1-c)^{\alpha}\mathrm{sign}(G_\sigma * I_{\eta\eta})|\nabla I|, \tag{39}$$

where $c = \exp(-\frac{|G_{\tilde{\sigma}} * \nabla I|^2}{k})$, was originally used for classification, based on a probabilistic framework. Eq. (39) is the adaptation of the original process for the task of image processing.

The performance of the last two schemes will be later compared with that of the process proposed in the present study.

C. The magnitude of the second derivative

To account for the magnitude of the second derivative controlling the flow, we return to the original shock filter formulation of (33) and employ $F(s) = \frac{2}{\pi}\arctan(as)$, where a is a parameter that controls the sharpness of the slope near zero. With this modification $F(s)$, Eq. (33), becomes:

$$I_t = -\frac{2}{\pi}\arctan(aI_{xx})|I_x| + \lambda I_{xx}. \tag{40}$$

Consequently, the inflection points are not of equal weight any longer; regions near edges, with large magnitude of the second derivative near the zero crossing, are sharpened much faster than relatively smooth regions. This type of process is implemented in the sequel in a new formulation of a complex PDE.

D. Complex Shock Filters

From (40) and (10) we derive the complex shock filter formulation for small θ:

$$I_t = -\frac{2}{\pi}\arctan(a\mathrm{Im}(\frac{I}{\theta}))|I_x| + \lambda I_{xx}, \tag{41}$$

where $\lambda = re^{i\theta}$is a complex scalar.

Generalization of the complex shock filter to 2D yields:

$$I_t = -\frac{2}{\pi}\arctan(a\mathrm{Im}(\frac{I}{\theta}))|\nabla I| + \lambda I_{\eta\eta} + \tilde{\lambda} I_{\xi\xi}, \tag{42}$$

where $\tilde{\lambda}$ is a real scalar.

The complex filter provides an original way to avoid the need for convolving the signal in each iteration, and still get smoothed estimations. The inherent time dependency contributes to the robustness of the process. Moreover, the imaginary value is affected by the shock and real directional diffusion processes – it is *enhanced* at sharp transitions by the shock and becomes more *coherent* in the level-set direction, thus, can better control the process than a simple second derivative. Note that here the imaginary value approximates a directional second derivative in the gradient direction (and not a Laplacian), since the complex diffusion term here is directional.

Process	Slope	Slope's variance	Shock success	Stability in time	Shock dislocation	Location variance	Location success	Location bias	SNR
Ideal	1	0	100%	1	0	0	100%	0	∞
Korn.	0.57	0.031	65%	0.73	2.6	14.3	93%	-0.5	8.7
C & A	0.76	0.192	72%	0.82	3.9	86.7	94%	-2.1	7.6
Ours	0.78	0.006	99%	0.99	1.7	4.7	99%	0.3	10.7
Ours - 0dB	0.62	0.024	81%	0.99	2.4	8.7	92%	0.6	8.8

TABLE I

EXPERIMENT RESULTS COMPARING THREE ROBUST SHOCK FILTERS PROCESSING A BLURRED NOISY STEP (SNR=5DB).

The performance of our complex valued shock filter (Eq. 41) is compared with the most advanced real valued robust shock filters, described earlier, of Kornprobst et al. (Eq. 38) and of Coulon and Arridge (Eq. 39). All three filters are designed to perform in a noisy environment, to produce shocks of important edges while simultaneously denoise fluctuations (of noise and/or texture). Trying to obtain objective quantitative measures to evaluate these filters we conducted a representative experiment of processing a blurred and noisy step edge.

In the experiment 100 blurred and noisy step edges (white Gaussian noise, SNR=5dB) were processed by each filter. The summary of the results is shown in Table I. The discrete signal I is comprised of N grid points (I_i, $i = 1, 2.., N$). In this context, the gradient is a simple grid point difference $DI_i = I_{i+1} - I_i$, where the largest gradient was considered as the place of the shock. We will now explain each column of the table: "Slope" - slope of largest gradient $s(I) = \max_i |DI_i|$. "Slope's variance" - variance of $s(I)$ over 100 trials. "Shock success" - we regarded a successful shock creation if the shock's slope was at least half of the original magnitude $s(I) \geq$

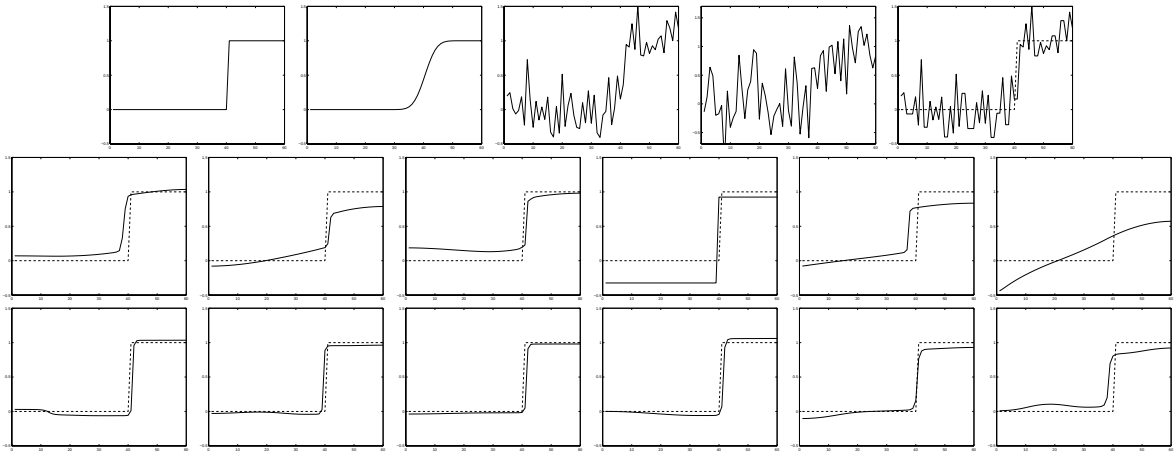

Fig. 14. Shock filters comparison experiment. Top row (from left): Original step, blurred step (Gaussian blur $\sigma_b = 3$), example of noisy signal with 5dB SNR, example of noisy signal with 0dB SNR, example of one result from Alvarez-Mazorra process (Eq. 37). Middle and bottom row shows some examples of processed signals from the experiment (5dB SNR), the result (solid) is superimposed on the ideal response (dashed): Middle row, 3 left figures: Kornprobst et al. (Eq. 38), 3 right figures: Coulon-Arridge (Eq. 39). Bottom row, 3 left figures: Our scheme (Eq. 41), 3 right figures: Our scheme processing noisier signals of 0dB SNR.

0.5. "Stability in time" - indicates how sensitive the result is to the stopping time. We computed the relative shock's slope after 10% more time: $s(I(1.1T))/s(I(T))$. "Shock dislocation" - average distance of the produced shock from the original shock location in terms of grid points, $E[|i_s - i_{orig}|]$, where $i_s = \text{argmax}_i |DI_i|$ and i_{orig} is the original shock point. "Location variance" - $Var[i_s]$. "Location success" - we regarded a success in terms of location accuracy if the distance of the formed shock was no more than 5 grid points from the original location, $|i_s - i_{orig}| \leq 5$. "Location bias" - $E[i_s - i_{orig}]$ (negative values means bias toward the center). The expected value of the shock location of an unbiased process is at the original location. "SNR" - average SNR of $I(T)$ with respect to the original unit step.

In this evaluation, for each process the parameters were first tuned to give good results and were kept constant in the experiment itself. The stopping time T was chosen automatically in order to produce a non oscillatory signal with a sharp and clear shock. For the experiment to be reproducible, all the parameters and exact criteria are listed in Appendix H. Some examples of processed outputs are shown in Fig. 14.

This experiment gives quantitative indications of the advantages of the complex shock filter with regard to the above criteria. The considerably lower variance in the results (sharpness

and location) accounts for the process reliability. The stability of the shock over evolution time indicates that a proper stopping time can be selected also in the enhancement of more compound signals with several blurred steps of different sizes and locations. Also, from our experience, it is far less sensitive to parameter tuning. Trying our process with noisier inputs of 0dB SNR gives comparable results to the other processes at 5dB SNR.

In Fig. 15 a blurred and noisy image was processed. In the case of two-dimensional signals, only the scheme of Kornprobst et at. and our complex scheme produce acceptable results at this levels of noise (SNR=15dB). Processing with the complex process results however, in sharper edges and is closer to the shock process, as can be observed in a comparison to an ideal shock response to a blurred image without noise (top-right image of Fig. 15). The combined enhancement-denoising properties of the complex scheme are highlighted by the display of one horizontal line of the image (bottom right of Fig. 15).

VI. Concluding remarks and discussion

Generalization of the linear and nonlinear scale spaces to the complex domain, by combining the diffusion and the free Schrödinger equations, further enhances the theoretical framework of the diffusion-type PDE approach to image processing.

The following advantages are afforded by the complexification of the diffusion equation according to the approach introduced in the present study.

1) The fundamental solution of the linear complex diffusion indicates that there exists a *stable process* over the wide range of the angular orientation of the complex diffusion coefficient, $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, that restricts the real value of the coefficient to be positive. [Issues related to aspects of inverse diffusion in image processing, i.e. negative real-valued diffusion coefficient, are dealt with elsewhere [13].]
2) In the case of small θ, two observations concerning the properties of the real and imaginary components of the complex diffusion process are relevant with regard to the application of this process in image processing: The real function is effectively decoupled from the imaginary one, and behaves like a real linear diffusion process, whereas the imaginary part approximates a smoothed second derivative of the real part, and can therefore well serve as an edge detector. In other words, the single complex diffusion process generates

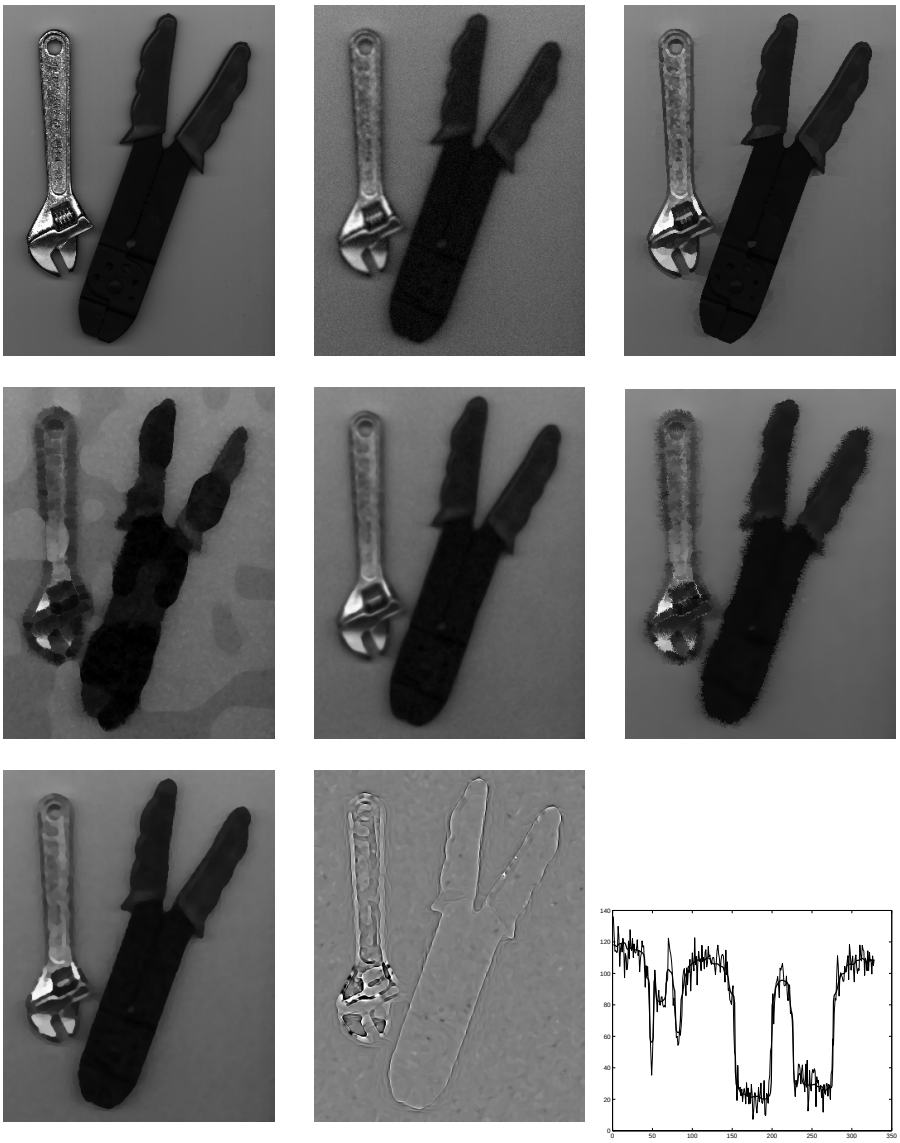

Fig. 15. Top row (from left): Original tools image, Gaussian blurred ($\sigma = 2$) with added white Gaussian noise (SNR=15dB), ideal shock response (of blurred image without the noise); middle row: evolutions of Eq. (37) - Alvarez-Mazorra ($\sigma = 10$), Eq. (38) - Kornprobst et al. ($\alpha_r = 0.2, \alpha_e = 0.1, \tau = 0.2, \sigma = 10, \tilde{\sigma} = 1$), Eq. (39) - Coulon-Arridge ($k = 5, \alpha = 1, \sigma = 10, \tilde{\sigma} = 1$); bottom: evolution of Eq. (41) - complex process, left: real values, middle: imaginary values ($|\lambda| = 0.1, \tilde{\lambda} = 0.5, a = 0.5$), right: grey level values generated along a horizontal line in the course of complex evolution of the process (thin line 1 iteration; bold line 100 iterations). All of the image evolution results are presented for 100 iterations (dt=0.1).

simultaneously an approximation of both the Gaussian and Laplacian pyramids [5] (at discrete set of temporal sampling points), i.e. the scale-space.

3) It paves the way to a more complete scale-space analysis. Further, the complex field is complete and brings along with it powerful tools for dealing with critical values.
4) In the linear case, the imaginary part is a bounded operator (and hence well-posed). Therefore, small perturbations in the data can not cause divergence of the results. This is unlike first or second order derivatives which are ill-posed operators, and are generally used for edge detection (pre-convolving the signal with a Gaussian still produces unstable results as $\sigma \rightarrow 0$, scaling by time is imperative). One may, therefore, conclude that the imaginary part can serve better as it is a "well-posed edge detector" for any $t \geq 0$. Its stability is inherent and does not depend on discretization effects or on the numerical schemes employed in the computations.
5) In many cases it is advantageous to switch on the non-linearity in an adiabatic way, such that over short time (small scale) the flow is mostly smoothing and as time progresses the interaction of the smoothing with the image's features take more important place and dominates the flow in large times. Explicit time dependency of the P-M coefficient and its benefits was demonstrated in [12]. In the complex framework, presented here, time dependency of the anisotropic case is inherent.
6) Complex diffusion enables better performance in different nonlinear tasks such as ramp denoising and regularization of shock filters.

Although nonlinear schemes remain to be further analyzed and better understood, nonlinear complex diffusion-type processes can be derived from the properties of the complex linear diffusion, and applied in image processing and enhancement. Such are the two schemes developed for denoising of ramp edges and for regularization of shock filters. In the first scheme, a nonlinear complex diffusion process controlled by the signal's imaginary value avoids the staircasing effect that is characteristic of gradient-controlled nonlinear processes such as the P-M process [28] (See Figures 7, 8, 9, 10 and 11). The second proposed scheme, presents a complex shock filter that overcomes problems inherent in the enhancement of noisy signals and images by the shock filters [27] and outperforms its various variants [2], [6], [21], [33]. In order for the reader to experience from first hand the performance and characteristics of our proposed linear and

nonlinear processes, we published our Matlab code implementation on the web [14].

Acknowledgments

This research has been supported in part by the Ollendorf Minerva Center, the Fund for the Promotion of Research at the Technion, Israel Academy of Science, Tel-Aviv University fund, the Adams Center and the Israeli Ministry of Science, the HASSIP Research Network Program HPRN-CT-2002-00285, sponsored by the European Commission and the BME Imaging Group of Columbia University and the ONR-MURI Grant N000M-01-1-0625, administered by Columbia University.

APPENDIX

In the following sections the analysis is presented in one dimension. As the kernel is separable, generalization to any number of dimensions is straightforward. More detailed proofs can be seen in our technical report [10].

A. *Maximal amplification*

$$\begin{aligned} \|T_{h_R}\|_\infty &\leq \|h_R\|_1 \text{ see section G (norm of a convolution operator)} \\ &= \int_{-\infty}^{\infty} |A g_\sigma(x)\cos\alpha(x)|dx \\ &\leq A\int_{-\infty}^{\infty} |g_\sigma(x)|dx \\ &= A = (\cos\theta)^{-1/2} \\ &= 1+\tfrac{\theta^2}{4}+O(\theta^4). \end{aligned} \tag{43}$$

$\|T_{h_I}\|$ can be considered in a similar manner.

B. *Effectively positive kernels*

In order to find a bound on P_{h_R}, we choose a point $x_1 > 0$ where $\cos\alpha(x_1) = 0.5$:

$$\begin{aligned} x_1 &= \sqrt{\tfrac{2}{3}(2\pi+3\theta)tr/\sin\theta} \\ &= \sigma\sqrt{(\tfrac{2}{3}\pi+\theta)\cot\theta}, \qquad \big(\text{using (9)}\big). \end{aligned} \tag{44}$$

For this bound we assume $x_1 > \sigma$, so it is valid for $1.28_{rad} = 73^\circ \geq \theta \geq 0$. We use the relations:

$$\begin{aligned} &(I) \quad -\int_{x_1}^{\infty} g_\sigma(x)dx \leq \int_{x_1}^{\infty} g_\sigma(x)\cos\alpha(x)dx \leq \int_{x_1}^{\infty} g_\sigma(x)dx \\ &\qquad (\text{as } -1 \leq \cos\alpha(x) \leq 1, \quad g_\sigma > 0), \\ &(II) \quad \tfrac{1}{8} \leq \tfrac{1}{2}\int_0^{\sigma} g_\sigma(x)dx \leq \tfrac{1}{2}\int_0^{x_1} g_\sigma(x)dx \leq \int_0^{x_1} g_\sigma(x)\cos\alpha(x)dx \\ &\qquad (\text{as } \cos\alpha(x) \geq \tfrac{1}{2} \text{ for all } x \in [0,x_1] \text{ , and } x_1 > \sigma). \end{aligned}$$

From the above relations and the fact that $h_R(x)$ is a symmetric function we can derive the bound of (17).

C. The distance between h_R and a Gaussian

We show the derivation of the bound on the distance between h_R and a Gaussian. The same arguments and very similar calculations are used in determining the bound on the h_I approximation.

Let us denote f_R as the difference function between h_R and a Gaussian at time t (see Fig. 2):

$$f_R(x;t) = h_R(x;t) - g_\sigma(x;t). \tag{45}$$

$$\begin{aligned} d(h_R, g_\sigma) &= \|T_{f_R}\|_\infty \le \|f_R\|_1 = \int_{-\infty}^{\infty} \left|h_R(x) - g_\sigma(x)\right| dx \\ &< 2\int_0^a \left|A\cos(\tfrac{x^2\tan\theta}{2\sigma^2} - \tfrac{\theta}{2}) - 1\right| g_\sigma(x)dx + 2(A+1)\int_a^\infty g_\sigma(x)dx \\ &< 2\max_{x\in[0,a]}\left[\left|A\cos(\tfrac{x^2\tan\theta}{2\sigma^2} - \tfrac{\theta}{2}) - 1\right|\right] + 4A\Phi(-a/\sigma) \end{aligned} \tag{46}$$

Let us choose $a = \sqrt{n}\sigma$ such that $n > 1$ and $\alpha(a) \ll \frac{\pi}{2}$ (therefore $\cos(\alpha(x)) > 0$ for any $x \in [0,a]$). Recalling that $\tan\theta = \theta + O(\theta^3)$ and $A = 1 + O(\theta^2)$ we get

$$\begin{aligned} &\max_{x\in[0,a]}\left[\left|A\cos(\tfrac{x^2\tan\theta}{2\sigma^2} - \tfrac{\theta}{2}) - 1\right|\right] \\ &= \max\left\{\left(A-1\right), \left(1 - (A\cos(\tfrac{a^2\tan\theta}{2\sigma^2} - \tfrac{\theta}{2}))\right)\right\} \\ &= \max\left\{\left(A-1\right), \left(1 - A\cos(\tilde{n}\theta + O(\theta^3))\right)\right\} \\ &= O(\theta^2). \end{aligned} \tag{47}$$

The term $\Phi(-a/\sigma) = \Phi(-\sqrt{n})$ decays exponentially with the growth of n. Therefore we conclude that

$$d_R = O(\theta^2) \tag{48}$$

The bound for the range $\theta \in (0, \frac{\pi}{10})$ was calculated numerically, see [10].

D. Definite integrals of h_R, h_I

Follows directly from section F.

E. Small theta approximation from the fundamental solution

We show here a little less elegant but more straightforward way for the small theta approximation in one dimension of h_I. The second derivative of a Gaussian (in 1D) is:

$$\frac{\partial^2}{\partial x^2} g_\sigma(x) = C\frac{x^2-\sigma^2}{\sigma^4} e^{-x^2/2\sigma^2}, \text{ where } C \doteq \frac{1}{\sqrt{2\pi}\sigma(t)}. \tag{49}$$

Small theta approximation of h_I:

$$\begin{aligned} h_I &= \tfrac{1}{\sqrt{\cos\theta}} C e^{-x^2/2\sigma^2} \sin\left(\tfrac{x^2\sin\theta}{4tr} - \tfrac{\theta}{2}\right) \\ &\approx Ce^{-x^2/2\sigma^2}\sin\left(\tfrac{x^2\theta}{4tr} - \tfrac{\theta}{2}\right) \qquad \left(\sqrt{\cos\theta} \approx 1 \quad , \sin\theta \approx \theta\right) \\ &\approx Ce^{-x^2/2\sigma^2}\left(\tfrac{\theta(x^2-2tr)}{4tr}\right) \qquad \left(\sin(\tfrac{x^2\theta}{4tr} - \tfrac{\theta}{2}) \approx (\tfrac{x^2\theta}{4tr} - \tfrac{\theta}{2})\right) \\ &\approx Ce^{-x^2/2\sigma^2}\left(\tfrac{\theta(x^2-\sigma^2)}{2\sigma^2}\right) \qquad \left(\sigma^2 = 2tr/\cos\theta \approx 2tr\right) \\ &= \tfrac{\partial^2}{\partial x^2} g_\sigma(x)\theta tr. \end{aligned} \tag{50}$$

F. Constant of the fundamental solution

The basic fundamental solution is

$$\begin{aligned} h(x;t) &= Ke^{-\frac{x^2\cos\theta}{4tr}}e^{i\frac{x^2\sin\theta}{4tr}} \\ &= |K|e^{-\frac{x^2}{2\sigma^2}}e^{i\left(\frac{x^2\tan\theta}{2\sigma^2}+\varphi\right)} \end{aligned} \tag{51}$$

As the constant K is complex, there are two degrees of freedom: the magnitude $|K|$ and the phase φ. The first requirement (7.a) dictated by the initial condition translates to

$$\begin{aligned} &(a.I) \quad \int_{-\infty}^{\infty} h_R(x;t\to 0)dx = 1 \\ &(a.II) \quad \int_{-\infty}^{\infty} h_I(x;t\to 0)dx = 0. \end{aligned} \tag{52}$$

We will see that (7.b) is satisfied based on the properties of the Gaussian.

First, let us find the definite integral of h_R. We use the following definite integral formula (taken from [29] p. 459, Eq. 16):

$$\begin{aligned} &\int_{-\infty}^{\infty} e^{-ax^2+bx+c}\{\sin|\cos\}(px^2+qx+r)dx = \\ &\frac{\sqrt{\pi}}{(a^2+p^2)^{1/4}}\exp\left[\frac{a(b^2-4ac)-(aq^2-2bpq+4cp^2}{4(a^2+p^2)}\right]\{\sin|\cos\}\left[\frac{1}{2}\arctan\frac{p}{a}-\frac{p(q^2-4pr)-(b^2p-2abq+4a^2r)}{4(a^2+p^2)}\right] \end{aligned} \tag{53}$$

From (51) we write h_R as

$$\begin{aligned} h_R &= |K|e^{-\frac{x^2}{2\sigma^2}}\cos\left(\frac{x^2\tan\theta}{2\sigma^2}+\varphi\right) \\ &= |K|e^{-ax^2}\cos\left(px^2+\varphi\right), \\ &\quad \text{where } a=\tfrac{1}{2\sigma^2}, \quad p=\tfrac{\tan\theta}{2\sigma^2}. \end{aligned} \tag{54}$$

From (53) and (54) we get:

$$\begin{aligned} \int_{-\infty}^{\infty} h_R(x)dx &= |K|\int_{-\infty}^{\infty} e^{-ax^2}\cos\left(px^2+\varphi\right)dx \\ &= |K|\sqrt{2\pi\cos\theta}\sigma\cos\left(\tfrac{1}{2}\theta+\varphi\right). \end{aligned} \tag{55}$$

Similarly,

$$\int_{-\infty}^{\infty} h_I(x)dx = |K|\sqrt{2\pi\cos\theta}\sigma\sin\left(\tfrac{1}{2}\theta+\varphi\right). \tag{56}$$

From (52.$a.II$) and (56) we get

$$\varphi = -\frac{1}{2}\theta \quad , \tag{57}$$

and from (52.$a.I$) and (55) we get

$$|K| = \frac{1}{\sqrt{2\pi\cos\theta}\sigma} \quad , \tag{58}$$

hence the constant K is

$$K = \frac{e^{-i\frac{1}{2}\theta}}{\sqrt{2\pi\cos\theta}\sigma}. \tag{59}$$

Requirement (7.b) is retained due to the characteristics of the Gaussian function. Let us choose $\epsilon = \sqrt{\sigma}$. Therefore $\epsilon(t) = (2tr/\cos\theta)^{1/4} \to_{t\to 0} 0$ for any $|\theta| < \frac{\pi}{2}$. And we get

$$\int_{|x|>\epsilon} |h(x)|dx = \frac{2}{\sqrt{\cos\theta}}\Phi(-\frac{1}{\sqrt{\sigma}}) \to_{\sigma\to 0} 0 \quad . \tag{60}$$

In the concise writing of the fundamental solution (8), K is actually separated into 3 multiplicative parts in the expressions of A, g_σ and α.

G. Norm of a convolution operator

Let f be a bounded function $\max(f) < M$, ($f \in L^\infty$). Let T_h be the convolution operator with the kernel h ($h \in L^1$): $T_h f = h * f$. We want to prove the relation

$$||T_h||_\infty \leq ||h||_1. \tag{61}$$

Let us first find a bound on $||T_h f||_\infty$. We use Young's inequality

$$\begin{aligned} &||f * h||_r \leq ||f||_p ||h||_q, \\ &f \in L^p, g \in L^q, \quad r^{-1} = p^{-1} + q^{-1} - 1, \quad (1 \leq p, q, r \leq \infty). \end{aligned} \tag{62}$$

Setting $p = \infty, q = 1, r = \infty$ we get

$$||T_h f||_\infty \leq ||f||_\infty ||h||_1. \tag{63}$$

We let

$$f_{sup} = \text{argsup}_{||f|| \neq 0} \frac{||T_h f||_\infty}{||f||_\infty}$$

From the definition in (20) it follows that $||T_h f_{sup}||_\infty = ||T_h||_\infty ||f_{sup}||_\infty$. Using the relation of (63) we get

$$||T_h||_\infty ||f_{sup}||_\infty \leq ||h||_1 ||f_{sup}||_\infty.$$

H. Noisy step experiment description

A unit step was discretized by 60 grid points where the step is off-center (in our case between points 40 and 41). The step was blurred by a Gaussian kernel of $\sigma = 3$. 100 instances with additive white Gaussian noise were produced ($\sigma_n = 0.25$, SNR with respect to the blurred signal is 5dB). Each instance was processed by the three robust shock filters (Kornprobst et al., Coulon-Arridge, ours). Stopping time was chosen automatically according to two criteria: In order to choose a sufficiently denoised signal the process was first evolved until the TV norm was less than 20% above that of a unit step: $\int |\nabla I| dx < 1.2$. Then we required that the largest gradient is monotonically increasing in time (stopped the process when the maximal absolute gradient was smaller than the previous check). This way we got a sharp gradient (presumably where the shock is) of a relatively smooth signal outside the vicinity of the shock. If the conditions were not met, the process was stopped after 10,000 iterations. Parameters: Eq. (38), $\alpha_r = 1, \alpha_e = 0.5, \tau = 0.03, \sigma = 1, \tilde{\sigma} = 2$. Eq. (39), $k = 0.02, \alpha = 1, \sigma = 1, \tilde{\sigma} = 1$. Eq. (41), $|\lambda| = 0.2, a = 8$ (in the 0dB experiments $a = 2$), $\theta = \frac{\pi}{1000}$.

I. Numerical implementation

We give the explicit 2D schemes for the three processes proposed by us in this paper. These are iterative schemes, where the value of each complex valued pixel $I_{i,j}^{n+1}$ (row i, column j) at iteration $n+1$ is updated according to the values of the neighboring pixels at iteration n: $I_{i,j}^{n+1} = I_{i,j}^n + \mu D_t I_{i,j}^n$. The value of μ is discussed below. [Note that here the addition and multiplication operators are complex valued.]

1) Linear diffusion :

$$D_t I^n_{i,j} = e^{i\theta}(I^n_{i+1,j} + I^n_{i-1,j} + I^n_{i,j+1} + I^n_{i,j-1} - 4I^n_{i,j}).$$

2) Ramp preserving diffusion : We use the direct complex generalization of the discretization scheme of [28]. Let us denote the four neighboring pixels as: $I_N = I_{i-1,j}$, $I_S = I_{i+1,j}$, $I_W = I_{i,j-1}$, $I_E = I_{i,j+1}$. The respective directional derivatives estimates are: $D(I_Z) \equiv I_Z - I_{i,j}$, $Z \in \{N, S, W, E\}$. The directional diffusion coefficient is $C(I_Z) = \frac{e^{i\theta}}{1+(Im(I_Z)/(k\theta))^2}$. The update is according to

$$D_t I^n_{i,j} = \Sigma_{Z\in\{N,S,W,E\}} D(I^n_Z) C(I^n_Z).$$

3) Complex shock filter:

$$D_t I^n_{i,j} = -\frac{2}{\pi}\arctan(a\mathrm{Im}(I^n_{i,j})/\theta)\sqrt{|\tilde{D}_x I^n_{i,j}|^2 + |\tilde{D}_y I^n_{i,j}|^2} + \lambda D^2_\eta I^n_{i,j} + \tilde{\lambda} D^2_\xi I^n_{i,j},$$

where $\tilde{D}_{\{x/y\}}$ is a symmetric first order approximation in the x or y direction defined by the *minmod* function, D_η approximates derivative in the gradient direction and D_ξ approximates derivative perpendicular to the gradient. We refer the reader to [27], [2] for specific implementation of these operators.

Time step bound: In explicit schemes the time step between each iteration is bounded for numerical stability by the CFL bound. As discussed earlier, the fundamental solution includes a Gaussian-type kernel of variance $\sigma^2 = \frac{2tr}{\cos\theta}$. Implementing Gaussian convolution of time τ, by incremental time steps where $\sigma^2 = 2\tau$, requires the time step bound to be: $\Delta\tau \leq 0.25h^2$ (in 2 dimensions, where h is the spatial step). In our case we have $\tau = \frac{tr}{\cos\theta}$ and hence in the general case we require: $\Delta t \leq 0.25h^2\frac{\cos\theta}{r}$. This applies for both the linear and the nonlinear cases presented above. For $r = 1$, $h = 1$ we get $\mu = \Delta t \leq 0.25\cos\theta$. Thus, when θ approaches $\pi/2$ it becomes very inefficient to implement complex diffusion with incremental time-steps. For small θ, however, there is essentially no difference from the case of real diffusion. In order to model more closely the analytic PDEs, one may use smaller time steps than the CFL bound at the beginning of the evolution.

For our implementation example code see [14].

REFERENCES

[1] L. Alvarez, F. Guichard, P.L. Lions, J.M. Morel, "Axioms and fundamental equations of image processing", Arch. Ration. Mech. Anal., 123(3): 199-257, 1993.

[2] L. Alvarez, L. Mazorra, "Signal and image restoration using shock filters and anisotropic diffusion", SIAM J. Numer. Anal. Vol. 31, No. 2, pp. 590-605, 1994.

[3] G. Aubert, P. Kornprobst, Mathematical Problems in Image Processing, Applied Mathematical Sciences 147, Springer-Verlag New-York, 2002.

[4] F. Barbaresco, "Calcul des variations et analyse spectrale: equations de Fourier et de Burgers pour modeles autoregressifs regularises", Traitement du Signal, vol. 17, No. 5/6, 2000.

[5] P.J. Burt, E.H. Adelson, "The Laplacian pyramid as a compact image code", IEEE Trans. Commun., vol. COM-31, No. 4, pp.532-540, 1983.

[6] O. Coulon, S.R. Arridge, "Dual echo MR image processing using multi-spectral probabilistic diffusion coupled with shock filters", MIUA'2000, British Conference on Medical Image Understanding and Analysis, London, United-Kingdom, 2000.

[7] M.C. Cross, P.C. Hohenberg, "Pattern formation outside of equilibrium", Rev. Modern Phys. 65, 854-1090 (1993).

[8] D. Gabor, "Theory of communication", Journal of the Institution of Electrical Engineers, 93(III):429–457, 1946.

[9] G. Gilboa, Y.Y. Zeevi, N. Sochen, "Complex diffusion processes for image filtering", Michael Kerckhove (Ed.): Scale-Space 2001, LNCS 2106, pp. 299-307, Springer-Verlag 2001.

[10] G. Gilboa, Y.Y. Zeevi, N. Sochen, "Image enhancement and denoising by complex diffusion processes", CCIT report 404, Technion, Israel, November 2002.

[11] G. Gilboa, N. Sochen, Y.Y. Zeevi, "Regularized shock filters and complex diffusion", ECCV-'02, LNCS 2350, pp. 399-313, Springer-Verlag 2002.

[12] G. Gilboa, N. Sochen, Y.Y. Zeevi, "Image enhancement segmentation and denoising by time dependent nonlinear diffusion processes", Proc. Int. Conf. on Image Processing (ICIP) 2001, vol. 3, pp. 134-137, Greece 2001.

[13] G. Gilboa, N. Sochen, Y.Y. Zeevi, "A Forward-and-Backward Diffusion Process for Adaptive Image Enhancement and Denoising", IEEE Trans. on Image Processing, Vol. 11, No. 7, pp. 689-703, 2002.

[14] G. Gilboa, Matlab code link: http://tiger.technion.ac.il/∼gilboa/PDE-based_image_filtering.html

[15] V. L. Ginzburg and L. D. Landau, Zh. V. L. Ginzburg and L. D. Landau, Zh. Eksp. Teor. Fiz. 20, 1064 (1950); English translation: see Men of Physics: Landau, Vol. II, edited by D. ter Haar, Pergamon, New York, 1965, pp. 546-568.

[16] P. Goupillaud, A. Grossmann, J. Morlet, "Cycle-octave and related transforms in seismic signal analysis", Geoexploration, 23, 85-102, 1984-1985.

[17] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape", SIAM J. Math. Anal, Vol. 15, pp 723-736, 1984.

[18] F. Guichard, J-M. Morel, "Image Analysis and PDE's" (new book in preparation).

[19] R Kimmel, R Malladi and N Sochen, "Images as Embedding Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images", Int. J. of Computer Vision, 39(2):111-129, Sept. 2000.

[20] J.J. Koenderink, "The structure of images", *Biol. Cybern.*, 50: 363-370, 1984.

[21] P. Kornprobst, R. Deriche, G. Aubert, "Image coupling, restoration and enhancement via PDE's", Proc. Int. Conf. on Image Processing 1997, pp. 458–461, Santa-Barbara (USA), 1997.

[22] T. Lindeberg, B. ter Haar Romeny, "Linear scale-space: (I) Basic theory and (II) Early visual operations.", in Geometry-Driven Diffusion, B. ter Haar Romeny (Ed.). pp. 1-77, 1994. Petros

[23] P. Maragos and F. Meyer, several articles in the Proceedings of Scale-Space Theories in Computer Vision, Lecture Notes in Computer Science, v.1682, 1999.

[24] D. Marr, *V*ision, Freeman & Co., 1982.

[25] M. Nagasawa, *S*chrödinger equations and diffusion theory, Monographs in mathematics, vol. 86, Birkhäuser Verlag, Basel, Switzerland 1993.

[26] A.C. Newell, "Envelope equations", Lect. in Appl. Math. 15, pp. 157-163, 1974.

[27] S.J. Osher and L. I. Rudin, "Feature-oriented image enhancement using shock filters", SIAM J. Numer. Anal. 27, pp. 919-940, 1990.

[28] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", IEEE Trans. Pat. Anal. Machine Intel., vol. PAMI-12,no. 7, pp. 629-639, 1990.

[29] A.P. Prudnikov, Y.A. Brychkov, O.I. Marichev, "Integrals and series", (English translation by N.M Queen), Vol. I, Gordon and Breach, 1986.

[30] B M ter Haar Romeny Ed., *Geometry Driven Diffusion in Computer Vision*, Kluwer Academic Publishers, 1994.

[31] F. Ratliff, *Mach bands: quantitative studies on neural networks in the retina*, Holden-Day, San Francisco (1965).

[32] P. Rosenau, “Free energy functionals at the high gradient limit”, Phys. Rev. A Vol. 41, pp 2227-2230, 1990.

[33] N.Rougon, F.Preteux, “Controlled anisotropic diffusion”, Proc. SPIE Conf. on Nonlinear Image Processing VI - IS&T / SPIE Symp. on Electronic Imaging, Science and Technology ’95, San Jose, CA, Vol. 2424, 1995, pp. 329-340.

[34] L. Rudin, S. Osher, E. Fatemi, “Nonlinear Total Variation based noise removal algorithms”, Physica D 60 259-268,‘92.

[35] R. Whitaker , G. Gerig, “Vector-valued diffusion”, in Geometry-Driven Diffusion. B. ter Haar Romeny (Ed.). pp. 93-134, 1994.

[36] A. P. Witkin, “Scale space filtering”, Proc. Int. Joint Conf. On Artificial Intelligence, pp. 1019-1023, 1983.

[37] M. Zibulski and Y. Y. Zeevi, ”Analysis of multi-window Gabor-type schemes by frame methods”, J. Applied and Computational Harmonic Analyisis, Vol. 4, pp. 188-221, 1997.