

Adaptive Diffusion as a Versatile Tool for Time-Frequency and Time-Scale Representations Processing: A Review

Julien Gosme, Cédric Richard, *Member, IEEE*, and Paulo Gonçalves

Abstract—Inspired by the work on image processing by Perona and Malik, diffusion-based models were first investigated by Gonçalves and Payot to improve the readability of Cohen class time-frequency representations. They rely on signal-dependent partial differential equations that yield adaptive smoothed representations with sharpened time-frequency components. Here, we demonstrate the versatility and utility of this family of methods, and we propose new adaptive diffusion processes to locally control both the orientation and the strength of smoothing. Our approach is an improvement on previous works as it provides a unified framework not only for the Cohen class but for the affine class as well. The latter is of particular interest because, except for some special techniques such as the reassignment method, no signal-dependent smoothing technique exists to process bilinear time-scale distributions, nor even a transposition of the adaptive optimal-kernel method proposed by Baraniuk and Jones.

Index Terms—Diffusion equations, interference terms, signal-dependent representations, time-frequency, time-scale.

I. INTRODUCTION

BILINEAR time-frequency and time-scale representations (TFRs and TSRs) provide a powerful tool for analyzing nonstationary signals. In particular, the Cohen class and the affine class offer a wide choice of distributions: In the case of the former, these are covariant with respect to time-frequency shifts and with the latter to time shifts and scale changes. All the elements of these two classes can be interpreted as generalized filtered versions of the Wigner distribution (WD) with two-dimensional (2-D) kernels. Although this central distribution has many theoretical properties, its use with real-world signals is unfortunately made difficult by the presence of interference terms that considerably limit its interpretability [1], [2]. In the area of signal analysis, one aims to preserve most of the attractive properties of the WD and also improve its readability. For example, the spectrogram and the scalogram are classical quasi-interference-free representations [3] of the Cohen and the affine classes, respectively, but both offer poor resolution. All this has prompted the development of specific task-oriented TFRs [4]–[7], as well as signal-dependent TFRs

[8]–[10]. Among the most notorious is the adaptive distribution by Baraniuk and Jones, whose radially Gaussian kernel is obtained by solving an energy-maximization problem subject to structural constraints. Despite its high efficiency on piecewise linear time-frequency components, however, the resulting kernel is optimum with respect to the entire signal [8] or to short-time segments of the signal [9]. Therefore, it does not locally adapt over time and frequency to account for dynamic changes in the signal to be analyzed. Moreover, this adaptive method is not applicable to TSRs of the affine class because of the scale-dependent convolution involved in their definition. In the recent past, a promising smoothing technique relying on a signal-dependent diffusion process was proposed [11]. It dynamically tailors the kernel shape over time and frequency to match the local representation characteristics. Minimization of an entropy function determines the tradeoff between excessive broadening and interference term suppression.

In this paper, we take an in-depth look at diffusion processes acting on bilinear TFRs and TSRs to demonstrate their versatility and utility for obtaining widely different signal representations. After a brief presentation of time-frequency processing and diffusion-based models, we show the link between the WD processed with the heat equation and the spectrogram computed with a Gaussian window; see [11]. We investigate several forms of variable-conductance diffusion processes to locally adjust diffusion strength and then perform an adaptive smoothing technique that matches the local TFR characteristics [11], [12]. Next, we propose a general diffusion scheme for tackling the problem of adaptive smoothing of TSRs of the affine class, and we show that resulting representations are covariant with respect to time shifts and scale changes [13]. Finally, we suggest the local control of the orientation and strength of smoothing via diffusion tensors. This refinement is applied to the TFRs of the Cohen class and the TSRs of the affine class.

II. TIME-FREQUENCY AND TIME-SCALE REPRESENTATIONS

Loosely speaking, the reason for time-frequency analysis is to give a mathematical core to the intuitive concept of time-varying Fourier spectrum for nonstationary signals. Over the years, substantial theoretical work has been carried out in this direction and has yielded many different classes of solutions: parametric or otherwise. Most of the parametric distributions of current interest belong to either the Cohen class [5] or the affine class [14]. In both cases, there exists a kernel formalism that allows

Manuscript received July 15, 2004; revised November 24, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Fredrik Gustafsson.

J. Gosme and C. Richard are with the Institut des Sciences et Technologies de l'Information de Troyes, Université de Technologie de Troyes (UTT), 10010 Troyes cedex, France (e-mail: julien.gosme@utt.fr; cedric.richard@utt.fr).

P. Gonçalves is with ITS-ISR, Lisbon-Portugal, on leave from INRIA, Rhône-Alpes, France (e-mail: paulo.goncalves@inrialpes.fr).

Digital Object Identifier 10.1109/TSP.2005.857048

us to interpret each element of these classes as the convolution product¹ of a central distribution (the WD) with a parameterizing 2-D kernel. In addition, these distributions suffer from a limitation inherent in the convolution product: The time and frequency resolutions are fixed, regardless of the local geometry of the signal signature. In practice, nevertheless, TFRs have proven to be useful in identifying nonstationarities in signals produced by a host of real-world applications such as speech analysis, narrowband radar, or machine monitoring.

A. Bilinear (Quadratic) Signal Distributions

The most general form of bilinear (quadratic) signal distribution is given by²

$$\rho_x(t, f; K) = \int \int x(u)x^*(v)K(u, v, t, f) du dv \quad (1)$$

where x is the finite energy signal to be analyzed and K an arbitrary parameterizing four-dimensional (4-D) kernel. The Cohen and the affine classes are two subsets of this large family. Both arise from specifying a structure of K in (1) so that the resulting distribution $\rho_x(t, f; K)$ satisfies required covariance properties. The Cohen class contains all the distributions C_x verifying the following commutativity diagram [5]:

$$\begin{array}{ccc} x(t) & \longrightarrow & C_x(t, f; K) \\ \downarrow & & \downarrow \\ \mathcal{W}x(t) \triangleq x(t - t_0)e^{j2\pi f_0 t} & \longrightarrow & C_x(t - t_0, f - f_0; K) \end{array} \quad (2)$$

i.e., distributions C_x that are covariant with respect to the application of the time-frequency shifts operator \mathcal{W} to the signal. These distributions can be written in the following parametric form [2], [5]:

$$C_x(t, f; \Pi) = \int \int W_x(u, v)\Pi(u - t, v - f) du dv \quad (3)$$

where W_x denotes the Wigner distribution

$$W_x(t, f) \triangleq \int x\left(t + \frac{\tau}{2}\right)x^*\left(t - \frac{\tau}{2}\right)e^{-j2\pi f\tau} d\tau \quad (4)$$

and Π stands for a simplified 2-D structure of the more general kernel K in (1). Similarly, the affine class contains all the distributions Ω_x that are covariant with respect to the affine operator \mathcal{A} of time shifts and scale changes [14], [15]

$$\begin{array}{ccc} x(t) & \longrightarrow & \Omega_x(t, a; K) \\ \downarrow & & \downarrow \\ \mathcal{A}x(t) \triangleq |a|^{-\frac{1}{2}}x\left(\frac{t-t_0}{a_0}\right) & \longrightarrow & \Omega_x\left(\frac{t-t_0}{a_0}, \frac{a}{a_0}; K\right). \end{array} \quad (5)$$

These distributions can be expressed as an *affine convolution* of the Wigner distribution W_x with a kernel Π as follows [16]:

$$\Omega_x(t, a; \Pi) = \int \int_0^{+\infty} W_x(u, v)\Pi\left(\frac{u-t}{a}, av\right) du dv. \quad (6)$$

In many cases, one can use the $a = f_0/f$ equivalence, where f_0 is a constant to relate frequency f and scale a [14].

¹For the affine class of TFRs, the kernel formalism involves a frequency-dependent convolution product.

²All integrals run from $-\infty$ to $+\infty$, unless otherwise noted.

B. Reduction of Interferences

The WD plays a key role in both the Cohen and the affine classes since it satisfies several fundamental properties. In particular, it offers optimal joint time-frequency resolutions that allow us to perfectly concentrate the signal components in the time-frequency domain [2]. However, due to its quadratic nature, the WD is affected by the presence of troublesome oscillating cross-terms, which are referred to as interference terms, coexisting with the signal auto terms and considerably limiting their expertise. To overcome this drawback, a number of reduced-interference distributions that exploit the convolutive forms in (3) and (6) have been proposed within both classes. Certainly, the most simple quasi-interference-free distribution of the Cohen class is the spectrogram [3]. Formally defined as the squared magnitude of a short-time Fourier transform with sliding time window w

$$SP_x(t, f) \triangleq \left| \int x(u)w^*(u - t)e^{-j2\pi fu} du \right|^2 \quad (7)$$

the spectrogram can be expressed in the formalism (3) with the smoothing kernel $\Pi(t, f) = W_w(t, f)$. While this smoothing operation removes the interference terms in W_x , it also spreads out the signal components. Refining this approach, the reduction of interference can be posed as a kernel Π design problem subject to the constraint of preserving some theoretical and/or practical properties of $C_x(t, f; \Pi)$ or $\Omega_x(t, a; \Pi)$. The optimal radially Gaussian kernel approach [8], [9] is such a tool. Although it performs extremely well for piecewise linear chirp signals, it is not always appropriate to process distributions with heterogeneous time-frequency patterns. Independently of the kernel shape, every distribution of the Cohen and the affine classes corresponds to a peculiar choice of Π that remains unchanged in the time-frequency domain. This means that the same processing is simultaneously applied to interference terms and signal components, removing the former ones and spreading out the latter.

Alternative solutions to the Cohen class and the affine class exist. Some of them consist of a nonuniform post-processing of representations guided by local signal characteristics. A typical example is the reassignment method, which is a point-wise transformation of regular TFRs [17]. It consists of moving each coefficient $C_x(t, f; \Pi)$ from its current location (t, f) , e.g., the geometric centroid of the kernel W_w in the case of the spectrogram, to another location (\hat{t}, \hat{f}) that depends on the local underlying geometry of the WD. This results in *squeezed representations* with sharp concentration on locally linear frequency modulations and no interference terms. However, an important proviso in using this technique is a reasonably high signal-to-noise ratio, as the coefficients $C_x(t, f; \Pi)$ may be locally reassigned onto random noise patterns instead of the signal components. Note that the reassignment method can also be applied to scalogram, yielding affine covariant distributions. Qian *et al.* have introduced another kind of approach that involves an explicit and controllable tradeoff between linear and bilinear distributions [18], [19]. This hybrid technique consists of expanding the signal to be analyzed in terms of a linear combination of elementary atoms, using a linear Gabor transform, for instance. Interference terms are then tapered as the distance between interacting atoms increases. It is possible

to locally control the degree of nonlinearity of this hybrid TFR and tune it for maximum concentration with minimum interference terms. This technique has also been transposed to the affine class in [20], which is a property shared by relatively few other methods. Finally, another group of methods that have been used with success involves the use of a local signal dependency in the kernel Π . Although, strictly speaking, the resulting distributions

$$C_x(t, f; \Pi_x) = \int \int W_x(u, v) \Pi_x(u - t, v - f) du dv \quad (8)$$

$$\Omega_x(t, a; \Pi_x) = \int \int_0^{+\infty} W_x(u, v) \Pi_x\left(\frac{u - t}{a}, av\right) du dv \quad (9)$$

are not members of the class of bilinear (quadratic) signal distributions, it is easy to check that C_x and Ω_x remain covariant with respect to time shifts, frequency shifts, or scale changes as long as Π_x is. Within this extended acceptance of both the Cohen class and the affine class, we now investigate a signal-dependent smoothing of the WD that adapts over time and frequency. It is based on diffusion processes such as those set out below.

III. DIFFUSION: BRIEF INTRODUCTION AND THE MAIN DEFINITIONS

Diffusion is the process by which matter is transported from areas of high concentration to areas of lower concentration as a result of the movement of an ensemble of molecules inside a region. This is mathematically formulated by a basic equation referred to as Fick's law [21]:

$$J = -D \nabla U \quad (10)$$

where J is the flux of molecules, U their concentration, D the diffusion tensor, and ∇ the gradient operator. The negative sign indicates that the diffusing mass flows in the direction of decreasing concentration. If D is a scalar-valued conductance function, which implies that J and ∇U are collinear, the diffusion process is called *isotropic*. Otherwise, it is called *anisotropic*. It is said that the diffusion process is *homogeneous* if the diffusion tensor D is constant over the region of interest. Location-dependent diffusion is called *nonhomogeneous* or *inhomogeneous*. The law of conservation of mass is expressed as

$$\frac{\partial U}{\partial \tau} = -\text{div}(J) \quad (11)$$

where τ is the diffusion time, and div the divergence operator. Combining this relationship with Fick's law produces the law of diffusion, which states

$$\frac{\partial U}{\partial \tau} = -\text{div}(-D \nabla U). \quad (12)$$

In the case where D is a positive constant, the resulting isotropic and homogeneous process (12) is often referred to as *heat diffusion equation* since it describes the evolution of temperature within a finite homogeneous continuum with no internal sources of heat. In this paper, this diffusion is said to be *linear* because

the tensor D does not vary with τ . Otherwise, it would be called *nonlinear*.³ Let us now restrict our discussion to the partial differential equations (PDEs) in the two-plus-one dimension

$$\begin{cases} U(v_1, v_2; \tau = 0) = U_0(v_1, v_2) \\ \frac{\partial U}{\partial \tau} = \text{div}(\nabla U) \end{cases} \quad (13)$$

where $U_0 \in \mathcal{L}^1(\mathbb{R}^2)$ denotes the initial spatial condition. It is well known that the solution to (13) is

$$U(v_1, v_2; \tau) = \begin{cases} (G * U_0)(v_1, v_2) & (\tau > 0) \\ U_0(v_1, v_2) & (\tau = 0) \end{cases} \quad (14)$$

where $*$ is the usual 2-D convolution, and $G(v_1, v_2; \tau) = (4\pi\tau)^{-1} \exp(-[v_1^2 + v_2^2]/4\tau)$ an isotropic Gaussian kernel, which is referred to as Green's function of the PDE given above. This means that the solution $U(v_1, v_2; \tau)$ of the heat diffusion equation (13) at each time instant τ can be simply obtained by convolution of the initial spatial condition $U_0(v_1, v_2)$ with Green's function $G(v_1, v_2; \tau)$. This result led Koenderink to use the basic PDE framework (13) for image processing, bringing a new insight into image restoration techniques viewed as homogeneous diffusion processes [22]. Note that earlier work was done by Iijima and published in Japanese [23], as pointed out in [24]. Thereafter, Perona and Malik proposed a nonlinear inhomogeneous generalization of this approach⁴ based on a real-valued conductance function $\eta(\cdot)$ of spatial coordinates (v_1, v_2) and time τ [25], [26]:

$$\frac{\partial U(v_1, v_2; \tau)}{\partial \tau} = \text{div}(\eta(|\nabla U|^2) \nabla U). \quad (15)$$

They chose to make the conductance function a decreasing function of the gradient magnitude in order to inhibit the diffusion process in areas of high contrast whereas areas of low contrast undergo faster diffusion. Systems such as (15) are intended to smooth out noise in images while sharp edges and contours are preserved. Significant advancements have been made since then, e.g., with diffusion processes acting differently along the gradient direction ∇U and the normal direction to ∇U . This last principle has been successfully applied by authors such as Weickert [29] and Alvarez *et al.* [30]. It has also been used by Rudin *et al.* [31] in the form of a backward diffusion along the gradient direction to sharpen edges. More recently, Gilboa *et al.* [32] and Smolka [33] have proposed forward-and-backward diffusion processes, where different regions of the image are either forward or backward diffused, according to the geometry within a neighborhood.

In the light of these considerations, we will now investigate new signal-dependent smoothing techniques of bilinear TFRs and TSRs that adapt over time, frequency, and scale to remove interference terms and noise while emphasizing signal components. Special efforts will be made to preserve the covariance properties of distributions with respect to time shifts, frequency shifts, or scale changes.

³In the theory of partial differential equations, a diffusion process is called *nonlinear* if D is a function of U or its derivatives.

⁴In [25]–[27], this diffusion process is referred to as *anisotropic*. However, according to the convention given in [28], this scheme is *isotropic* since its flux J defined in (10) is collinear to ∇U .

IV. ISOTROPIC DIFFUSION OF TFRs AND TSRS

Distributions in the Cohen class involves convolution of the WD with a 2-D kernel Π . This interpretation has led many authors to investigate design aspects of Π for reducing noise and interference terms in TFRs. Unfortunately, reduced interference distributions with fixed kernel necessarily spread out and delocalize signal components as well. In order to overcome this incompatibility, Gonçalves and Payot first revisited standard 2-D smoothing related to the Cohen class in terms of diffusion processes of the form (13) and (15) [11]. Here, we start by briefly sketching the different steps of this approach, and we check that resulting distributions are covariant with respect to time-frequency shifts. Next, we extend this framework for tackling the problem of adaptive smoothing of TSRS of the affine class.

A. Isotropic Diffusion in the Cohen Class

The Gaussian-windowed spectrogram (7) with variance σ_w^2 illustrates the connection between smoothed TFRs and PDEs. Indeed, its kernel $\Pi(t, f) = W_w(t, f)$ is the Gaussian function $\Pi(t, f) = 2 \exp(-[t^2 + 4\pi^2 \sigma_w^4 f^2]/\sigma_w^2)$. It corresponds to Green's function $G(t, f; \tau) = (4\pi\tau)^{-1} \exp(-[t^2 + f^2]/4\tau)$ of the heat equation

$$\begin{cases} \frac{\partial C_x(t, f; \tau)}{\partial \tau} = \text{div}(\nabla C_x(t, f; \tau)) \\ C_x(t, f; \tau = 0) = W_x(t, f) \end{cases} \quad (16)$$

when $\tau = (4\pi)^{-1}$ if, as the isotropy of $G(t, f; \tau)$ implies, $\sigma_w^2 = (2\pi)^{-1}$. At this instant of diffusion, the spectrogram $SP_x(t, f)$ mentioned above is then the solution of (16). In a more general way, $C_x(t, f; \tau)$ is a member of the Cohen class for all positive τ since it can be expressed as the 2-D convolution of $W_x(t, f)$ with $G(t, f; \tau)$.

Following the lines of [25]–[27] in an image processing context, a possible refinement of (16) consists of introducing inhomogeneity into the heat equation to perform time- and frequency-dependent smoothing [11]:

$$\begin{cases} \frac{\partial C_x(t, f; \tau)}{\partial \tau} = \text{div}(\eta_x(t, f) \nabla C_x(t, f; \tau)) \\ C_x(t, f; \tau = 0) = W_x(t, f) \end{cases} \quad (17)$$

where $\eta_x(t, f)$ is a conductance function depending both on the analyzed signal x and time-frequency location (t, f) . In their primary work, Perona and Malik chose to use a decreasing conductance function of the gradient magnitude, as mentioned in Section III. However, this solution is not adequate in a time-frequency context since the largest variations of the WD coincide with oscillating interference terms we want to remove. The conductance function η_x can be chosen in many different ways, depending on the information we want to emphasize in TFRs. Several solutions have been proposed in [11] and [12] to improve the readability of the WD. They involve an auxiliary TFR of the Cohen class, which is denoted by ρ_x , in order to combine the localization properties of the WD with some desired characteristics of ρ_x . For instance, Gonçalves and Payot have suggested using

$$\eta_{W_x}(t, f) = \mathbb{1}_{[-\infty, 0]}(W_x(t, f)) \quad (18)$$

where $\mathbb{1}_{[-\infty, 0]}(\cdot)$ equals 1 if its argument is negative and 0 otherwise. This diffusion function is intended to smooth out interference terms, which are mostly located in the neighborhood of

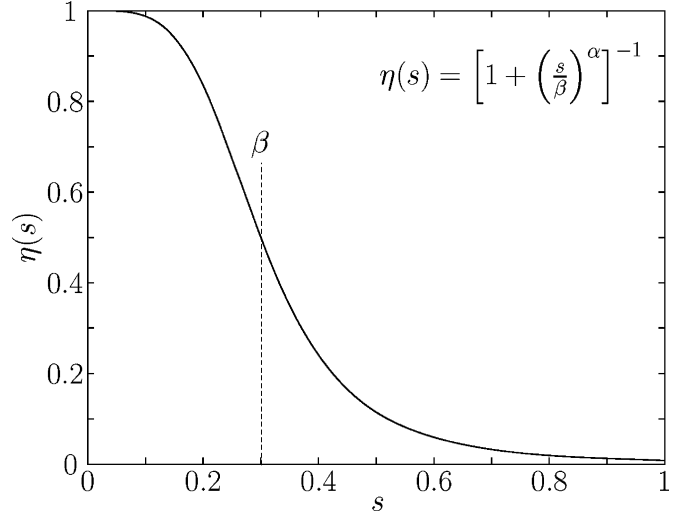


Fig. 1. Conductance function (19) with $\alpha = 4$ and $\beta = 0.3$.

time-frequency areas where $W_x(t, f) < 0$, due to their oscillating nature. Note that (18) can be made to be dependent on the diffusion time τ by using the evolving distribution $C_x(t, f; \tau)$ instead of $W_x(t, f)$. Such a feedback leads to nonlinear diffusion filters. Another intuitive diffusion-based approach for obtaining interference-free distributions relies on the spectrogram. Although this distribution spreads out signal components in the time-frequency domain, it is approximately equal to zero over nonenergetic areas, where the interference terms of the WD are likely to be situated. Thus, as illustrated in Fig. 1, making the conductance function a decreasing function of SP_x such as

$$\eta_{SP_x}(t, f) = \left[1 + \left(\frac{SP_x(t, f)}{\beta} \right)^\alpha \right]^{-1}, \quad (\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \quad (19)$$

allows for the tuning of the diffusion rate over the time-frequency domain. As a last example, the conductance function

$$\eta_{C_x}(t, f; \tau) = \left[1 + \left(\frac{C_x(t, f; \tau)}{\beta} \right)^\alpha \right]^{-1}, \quad (\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \quad (20)$$

assumes that signal components are more *structured* than interference terms. Hence, it inhibits the diffusion process in the neighborhood of structured signal components, whereas thin components undergo faster diffusion. Fig. 2 illustrates the diffusion process (17) associated with the spectrogram-based diffusion function (19). Compared to the WD, only the signal components are left unchanged, whereas the interference terms are progressively smoothed out. As shown in Fig. 3, this smoothing process still performs fairly well in a noisy environment. Because the WD maps a white noise to a 2-D random field in the time-frequency domain, positive and negative coefficients are locally balanced and partly cancelled out by the diffusion process. The robustness of the spectrogram with respect to noise also explains this interesting result.

We will now prove that $C_x(t, f; \tau)$, under mild conditions on the initial distribution and the conductance function, is covariant with respect to time-frequency shifts for all positive τ . Let us first rewrite the diffusion process (17) in the form $\tilde{\mathcal{D}}_\tau[C_x(t, f; \tau = 0)] = C_x(t, f; \tau)$, where $\tilde{\mathcal{D}}_\tau$ is the operator

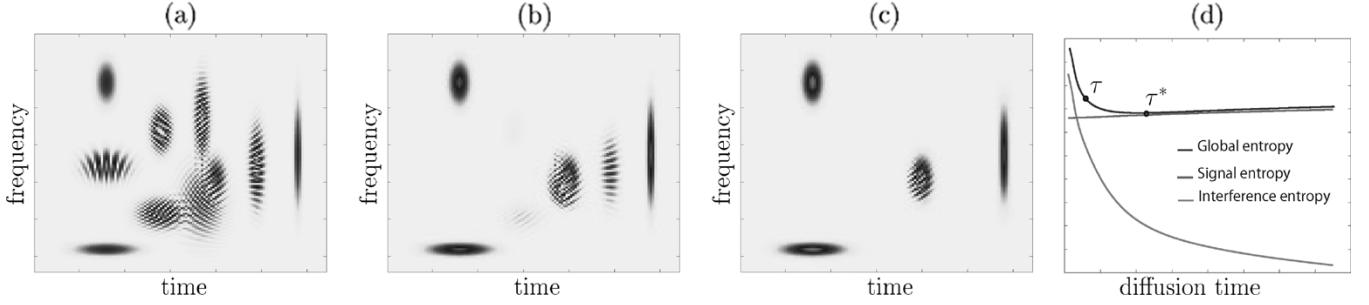


Fig. 2. Adaptive diffusion process with spectrogram-based conductance function applied to the 256-by-256 WD of a 256-sample four-component signal. (a) WD. (b) Diffusion until time instant τ . (c) Diffusion until the optimum stopping time instant τ^* . (d) Evolution of the entropy with contributions of signal components and interference terms. Note that these two contributions are not additive because of the superimposition of the TF supports of interference terms and signal components and the nonlinearity of entropy.

describing the evolution of $C_x(t, f; \tau)$. Let us also define the translation operator \hat{W}_{t_0, f_0} , which, applied to any distribution $C_x(t, f)$, gives $\hat{W}_{t_0, f_0}[C_x(t, f)] = C_x(t - t_0, f - f_0)$. The simple change of variables $t \leftarrow (t - t_0)$ and $f \leftarrow (f - f_0)$ in (17) allows us to show that

$$\hat{W}_{t_0, f_0} \tilde{D}_\tau [C_x(t, f; \tau)] = \tilde{D}_\tau \hat{W}_{t_0, f_0} [C_x(t, f; \tau)] \quad (21)$$

meaning that diffusing a shifted TFR is the same as shifting the diffused TFR. This property is referred to as translation invariance in [28]. On the one hand, we verify that $\hat{W}_{t_0, f_0} \tilde{D}_\tau [C_x(t, f; \tau = 0)] = C_x(t - t_0, f - f_0; \tau)$; on the other hand, we see that $\tilde{D}_\tau \hat{W}_{t_0, f_0} [C_x(t, f; \tau = 0)] = \tilde{D}_\tau [C_{Wx}(t, f; \tau = 0)]$ if the initial distribution $C_x(t, f; \tau = 0)$ is covariant with respect to time-frequency shifts, and $\tilde{D}_\tau [C_{Wx}(t, f; \tau = 0)] = C_{Wx}(t, f; \tau)$ if the conductance function $\eta_x(t, f)$ also verifies this covariance property, i.e., $\eta_{Wx}(t, f) = \eta_x(t - t_0, f - f_0)$. Under these two conditions, we can conclude with (21) that

$$C_{Wx}(t, f; \tau) = C_x(t - t_0, f - f_0; \tau) \quad (22)$$

which means that $C_x(t, f; \tau)$ is covariant with respect to time-frequency shifts, for all positive τ . In addition, note that (17) conserves energy, i.e., $\int \int C_x(t, f; \tau) dt df = \int |x(t)|^2 dt$, since (17) is a particular case of (12), which follows directly from the law of conservation of mass (11).

Let us now present a discrete framework for the diffusion process (17). We suppose that TFRs are sampled on a finite grid $(n, m) \in [0, N] \times [0, M]$, where δt and δf are the sampling period in time and frequency [11], [25], [26]. Using a simple first-order Euler approximation gives the following approximate discrete evolution equation:

$$\begin{aligned} C_x[n, m; \tau_{k+1}] \\ = C_x[n, m; \tau_k] + \frac{1}{4} \sum_{i \in I} \eta_x^{(i)}[n, m] \nabla^{(i)} C_x[n, m; \tau_k] \end{aligned} \quad (23)$$

where $i \in I$ refers to the four neighboring points of $C_x[n, m; \tau_k]$, and $C_x[n, m; \tau_0] = W_x[n, m]$. In (23), $\eta_x^{(i)}$ denotes the average value of the conductance function along the direction i , e.g., for the upper neighbor

$$\eta_x^{(u)}[n, m] = \frac{\eta_x[n, m] + \eta_x[n, m + 1]}{2} \quad (24)$$

and $\nabla^{(i)} C_x[n, m; \tau_k]$ is the first-order derivative of $C_x[n, m; \tau_k]$ along the direction i , e.g., for the upper neighbor

$$\nabla^{(u)} C_x[n, m; \tau_k] = C_x[n, m + 1; \tau_k] - C_x[n, m; \tau_k]. \quad (25)$$

In the remainder of this paper, we use either straightforward adaptations of this numerical scheme or those proposed in [34] and [35].

Without a constraint on the diffusion time, C_x would converge to a uniform distribution over the time-frequency domain, regardless of the analyzed signal. Hence, the diffusion process is to be stopped when some criterion is achieved. The Shannon entropy, which is defined as [36]

$$H_C(\tau_k) = - \sum_n \sum_m |C_x[n, m; \tau_k]| \log |C_x[n, m; \tau_k]| \quad (26)$$

is a natural candidate for measuring the complexity of TFRs.⁵ Indeed, the TFRs of signals consisting of small numbers of components would lead to small entropy values, whereas the TFRs of information-rich signals would lead to large ones. A simple way to justify the above stopping criterion is to consider an example, where the overall integrals can be decomposed into two distinct region integrals: the first one over the domain of interference terms and the second one over the domain of signal components. The first integral tends to zero as the interference terms are smoothed out, whereas the second one increases at a slower rate due to the action of the conductance function. When these antagonistic effects balance one another, the Shannon entropy reaches a minimum that causes the diffusion process to stop. This behavior is illustrated in Figs. 2(d) and 3(f), where H_C rapidly decreases toward a minimum and then slowly increases beyond this optimum stopping time τ^* . In an image processing context, note that it was shown that the Shannon entropy monotonically increases with respect to the diffusion time for most commonly used diffusion processes [28], [37], [38]. This property is based on the convexity of the function $v \log v$, $v \in \mathbb{R}_+^*$ in the definition of the Shannon entropy. It does not hold for (26) when dealing with nonpositive distributions, as in the case of TFRs affected by the presence of interference terms. As a final remark, it should be noted that we could also consider Rényi's entropies instead of the Shannon entropy, as suggested in [36].

⁵We suggest referring to [36] for detailed study of the Rényi's entropies as time-frequency information measures. We also invite him to consult [37] and [38] for full details on the Rényi's entropies as a means of extracting information from images during diffusion.

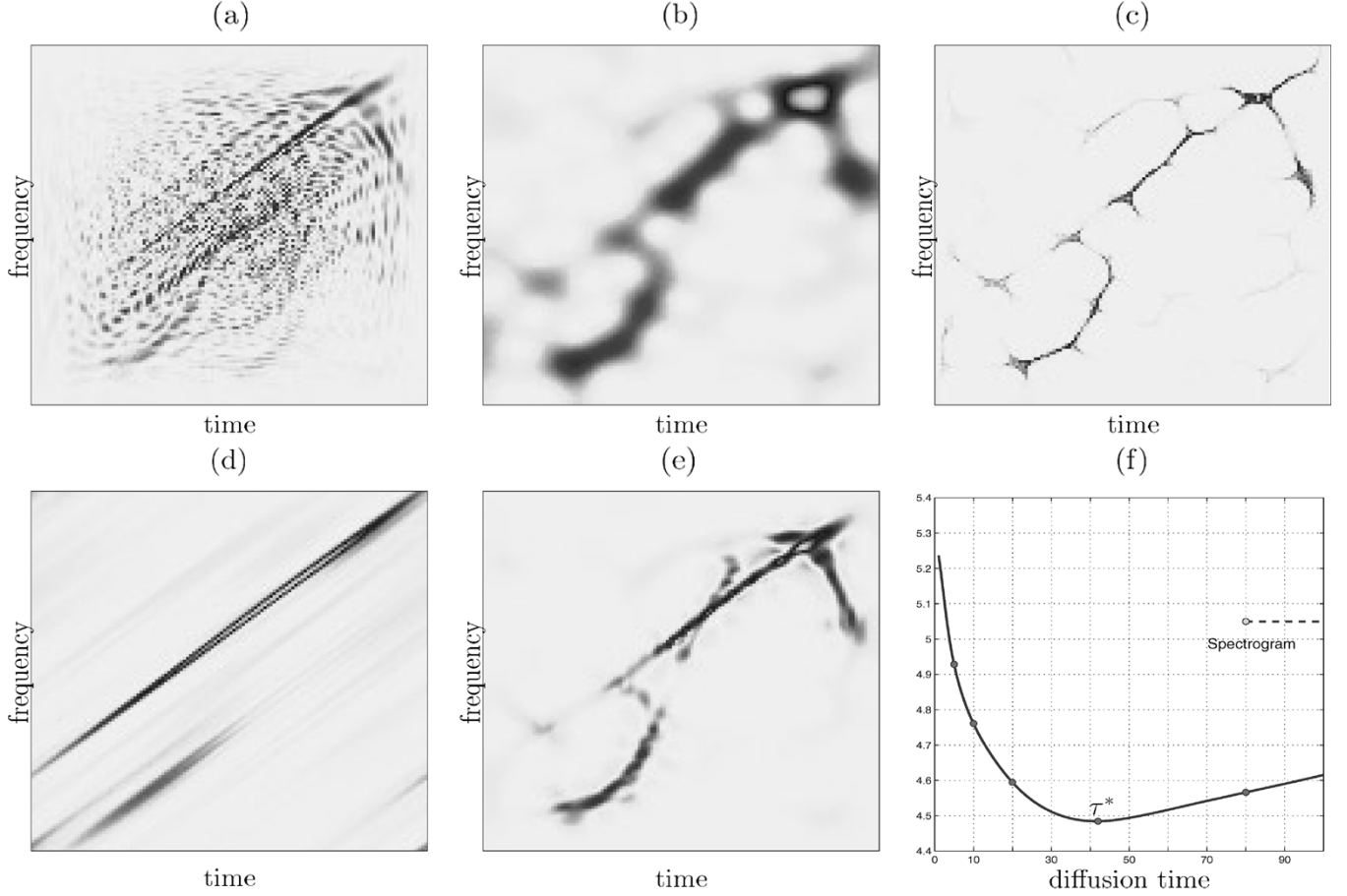


Fig. 3. Comparison of the adaptive diffusion process with well-known TFRs. (a) WD. (b) Spectrogram. (c) Spectrogram modified by the reassignment method. (d) TFR with adapted radially Gaussian kernel. (e) Diffusion until the optimum stopping time τ^* using the spectrogram-based conductance function. (f) Evolution of the entropy. These 256-by-256 TFRs are those of a 256-sample signal embedded in noise, which consists of two frequency-modulated waves with linear and sinusoidal modulations.

This would lead to different TFRs optimizing other criteria related to the compromise “resolution versus interference terms removal”. Throughout the remainder of this paper, we will restrict ourselves to the stopping criterion (26).

In this section, we have investigated several forms of variable-conductance diffusion processes in order to locally adjust diffusion strength and then perform adaptive smoothing that matches local characteristics of TFRs. We will now extend this approach to tackle the problem of adaptive smoothing of TSR’s of the affine class.

B. Isotropic Diffusion in the Affine Class

Every bilinear TSR in the affine class is characterized by an *affine convolution* of the WD, as can be seen in (6) (see also Fig. 4). Just like the spectrogram within the Cohen class, the scalogram plays a key role within the affine class. It is defined as the square magnitude of the continuous wavelet transform

$$SC_x(t, a) \triangleq \left| \frac{1}{\sqrt{a}} \int x(u) w^* \left(\frac{u-t}{a} \right) du \right|^2 \quad (27)$$

where a is the analyzing scale, and $w(t)$ is an admissible wavelet, i.e., $\int w(t) dt = 0$ [39]. By analogy with the connection (16) established between the heat equation and smoothed

TFRs of the Cohen class, we propose to adapt the diffusion strength to the scale a as follows [13]:

$$\begin{cases} \frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = \text{div} (a^2 \nabla \Omega_x(t, a; \tau)) \\ \Omega_x(t, a; \tau = 0) = W_x(t, a) \end{cases} \quad (28)$$

where $\Omega_x(t, a; \tau = 0)$ is the TSR to be processed, e.g., the WD in the present case. Note that (28) differs from (16) by the presence of the factor a^2 in the divergence term. As can be shown by direct calculations based on a change of variables in (28), this modification is crucial in obtaining the following property:⁶

$$\tilde{\mathcal{A}}_{t_0, a_0} \tilde{\mathcal{D}}_\tau [\Omega_x(t, a; \tau)] = \tilde{\mathcal{D}}_\tau \tilde{\mathcal{A}}_{t_0, a_0} [\Omega_x(t, a; \tau)] \quad (29)$$

which plays the same role as (21) in the case of distributions of the Cohen class modified by the diffusion process (17). Here, $\tilde{\mathcal{A}}_{t_0, a_0}$ is the operator defined by $\tilde{\mathcal{A}}_{t_0, a_0} [\Omega_x(t, a)] = \Omega_x((t - t_0)/a_0, a/a_0)$, and $\tilde{\mathcal{D}}_\tau$ is the diffusion operator associated with (28). We can also check that $\tilde{\mathcal{A}}_{t_0, a_0} \tilde{\mathcal{D}}_\tau [\Omega_x(t, a; \tau = 0)] = \Omega_x((t - t_0)/a_0, a/a_0; \tau)$. In addition, if $\Omega_x(t, a; \tau = 0)$ is covariant with respect to the application of the affine operator \mathcal{A} , which is defined in (5), we have $\tilde{\mathcal{D}}_\tau \tilde{\mathcal{A}}_{t_0, a_0} [\Omega_x(t, a; \tau = 0)] =$

⁶Replace variables t and a in (28) by $(t - t_0)/a_0$ and a/a_0 , respectively, and calculate the divergence term to get (29).

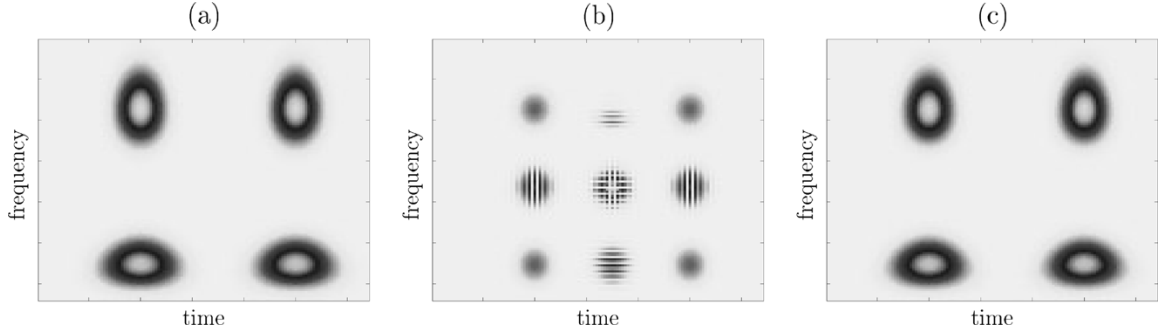


Fig. 4. (a) Affine isotropic diffusion applied to (b) the WD of a 256-sample four-component signal. (c) Gaussian-windowed scalogram. These 256-by-256 distributions are represented in a time-frequency form to clearly illustrate the effect of affine smoothing.

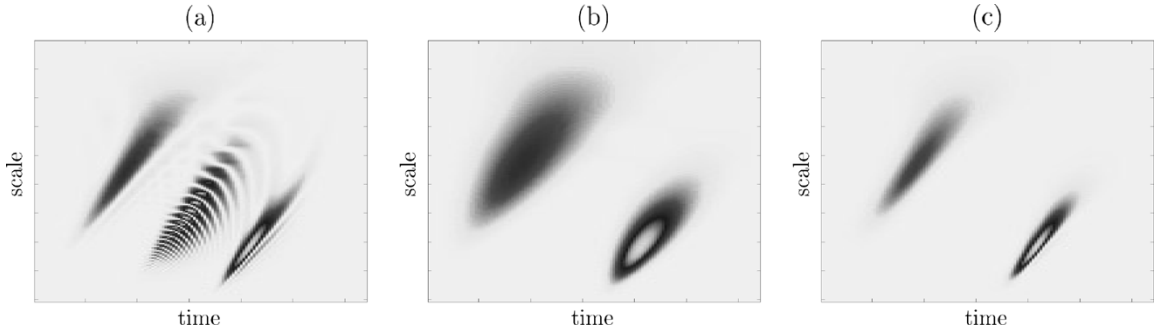


Fig. 5. Affine diffusion applied to two 256-sample Altes chirps. (a) WD. (b) Isotropic diffusion. (c) Adaptive isotropic diffusion with scalogram-based conductance function.

$\Omega_{Ax}(t, a; \tau)$. Combining these two results via (29), we then conclude that

$$\Omega_{Ax}(t, a; \tau) = \Omega_x\left(\frac{t - t_0}{a_0}, \frac{a}{a_0}; \tau\right), \quad \forall \tau \geq 0 \quad (30)$$

which means that $\Omega_x(t, a; \tau)$ is covariant with respect to time shifts and scale changes if $\Omega_x(t, a; \tau = 0)$ is. In addition, $\Omega_x(t, a; \tau)$ is a bilinear distribution because of the linearity of the differential operators in (28).⁷ We therefore conclude that $\Omega_x(t, a; \tau)$ is a member of the affine class. Then, according to the canonical definition (6) of this class, some kernel $\Pi(t, a; \tau)$ associated with $\Omega_x(t, a; \tau)$ may well be found. However, its identification is a nontrivial issue because it deals with an inhomogeneous diffusion scheme. To the best of our knowledge, this problem remains an open question. Finally, let us remark that (28) is not the only diffusion process that produces TSRs of the affine class. For instance, it shares this property with

$$\frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = a^2 \operatorname{div}(\nabla \Omega_x(t, a; \tau)) \quad (31)$$

$$\frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = a \operatorname{div}(a \nabla \Omega_x(t, a; \tau)). \quad (32)$$

Throughout the remainder of this section, we will restrict ourselves to the diffusion process (28), which conserves energy since it has the same form as (12). To tackle the problem of adaptive smoothing of TSRs, we now propose to extend (28) using the following isotropic inhomogeneous scheme:

$$\begin{cases} \frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = \operatorname{div}(a^2 \eta_x(t, a) \nabla \Omega_x(t, a; \tau)) \\ \Omega_x(t, a; \tau = 0) = W_x(t, a). \end{cases} \quad (33)$$

⁷Affine diffusion processes have been proposed in an image processing context [30]. Nevertheless, they do not preserve the bilinearity.

A similar proof to those given before shows that the covariance of $\eta_x(t, a)$ with respect to time shifts and scale changes allows this diffusion process to generate affine-covariant distributions $\Omega_x(t, a; \tau)$ for all positive τ . As already pointed out in the previous subsection, many strategies may be developed to design specific conductance functions. For instance, following the intuition that has led to (19), an approach for obtaining interference-free TSRs relies on the conductance function

$$\eta_{SC_x}(t, a) = \left[1 + \left(\frac{SC_x(t, a)}{\beta} \right)^\alpha \right]^{-1}, \quad (\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \quad (34)$$

calculated from a scalogram. Like the spectrogram, this distribution is strictly positive and does not suffer much from interference terms [3]. Since it is covariant with respect to time shifts and scale changes, it also ensures that the resulting distributions $\Omega_x(t, a; \tau)$ are members of the affine class. Fig. 5 clearly shows the benefit of using the inhomogeneous model (33) with this conductance function. It also illustrates the homogeneous model (28) tendency to simultaneously remove interference terms and smooth out signal components in the WD.

V. DIFFUSION OF TFRS AND TSRs: EXTENSION TO ANISOTROPIC PROCESSES

Thus far, we have investigated diffusion-based models with scalar-valued conductance functions to locally adjust diffusion strength. Here, we propose to control both diffusion strength and orientation with the tensor D involved in the general diffusion model (12). In the spirit of the work by Baraniuk and Jones on signal-dependent time-frequency analysis [9], the anisotropic diffusion schemes presented below allow the

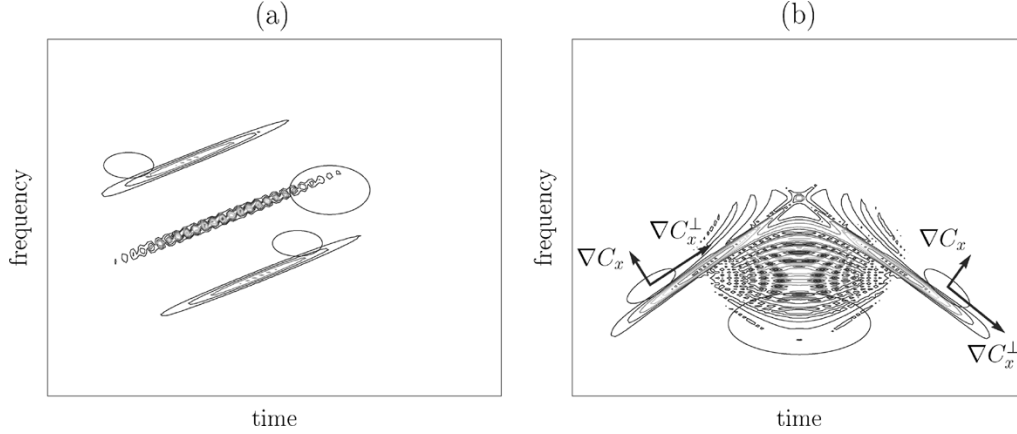


Fig. 6. (a) Isotropic and (b) anisotropic diffusion processes, with local kernels schematized by ellipses. Anisotropy allows the kernel orientation to be usefully adapted to the distribution.

direction and amount of smoothing to be adjusted. The cases of TFRs of the Cohen class and TSRs of the affine class are considered.

A. Anisotropic Diffusion in the Cohen Class

In contrast to the isotropic process (16), where the diffusion is solely governed by the geometry of the distribution to be processed, we propose a more flexible scheme where the diffusion model is also controlled by a tensor D_{tf} built into the diffusion equation:

$$\begin{cases} \frac{\partial C_x(t, f; \tau)}{\partial \tau} = \text{div}(D_{tf} \nabla C_x(t, f; \tau)) \\ C_x(t, f; \tau = 0) = W_x(t, f) \end{cases} \quad (35)$$

where D_{tf} is a (2×2) symmetric positive definite matrix in order to ensure the stability of the system [28]. Note that (16) can be recovered by using the unit matrix for D_{tf} . For a fixed matrix D_{tf} , which is denoted below by D , it is a well-known result that the solution $C_x(t, f; \tau)$ to (35) at each diffusion instant τ can be simply obtained by convolving the initial distribution $W_x(t, f)$ with the following 2-D Gaussian kernel [28]:

$$\Pi(t, f; \tau) = \frac{1}{4\pi\tau|D|^{\frac{1}{2}}} \exp\left(-\frac{(t, f)D^{-1}(t, f)^t}{4\tau}\right). \quad (36)$$

The eigenvectors and the eigenvalues of the diffusion tensor then define the two orthogonal directions of smoothing, as well as the smoothing strength in these directions.

Some signals, such as linearly modulated ones, may consist of time-frequency components having a dominant orientation. Incorporating such information as prior knowledge in the diffusion process should enhance its performance, in particular if diffusion along signal components is preferred to diffusion perpendicular to them. This can be achieved by adapting the Gaussian kernel orientation to the distribution to be processed. Here, we suggest choosing D_{tf} in (35) as a function of the gradient ∇C_x within the vicinity of any investigated location (t, f) , as shown in Fig. 6. To this end, the orthonormal system of eigenvectors $\{v_1, v_2\}$ of D_{tf} is fixed such that v_1 and v_2 are collinear to ∇C_x and ∇C_x^\perp , respectively:

$$D_{tf} = (v_1 \ v_2) \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} (v_1 \ v_2)^t \quad (37)$$

where the eigenvalue λ has to be chosen greater than 1 in order to produce more aggressive smoothing along signal components than across them. A convenient way to estimate local orientation is to introduce the gradient-square tensor, which is defined as [40]

$$J_x(t, f) = \nabla C_x (\nabla C_x)^t = \begin{pmatrix} \left[\frac{\partial C_x}{\partial t}\right]^2 & \frac{\partial C_x}{\partial t} \frac{\partial C_x}{\partial f} \\ \frac{\partial C_x}{\partial t} \frac{\partial C_x}{\partial f} & \left[\frac{\partial C_x}{\partial f}\right]^2 \end{pmatrix}. \quad (38)$$

The eigenvectors of $J_x(t, f)$ are collinear to ∇C_x and ∇C_x^\perp . The interest in using the gradient-square tensor is that it gives the same result for gradients with opposite signs. As a consequence, this matrix describes the local orientation of the structure (not its direction). In the presence of noise, it is well-known that the gradient does not give reliable information about local orientation. In practical applications, a smoothing filter is used to produce a gradient-square tensor with average orientation information from an arbitrary weighted neighborhood [28], [40]. This operation can be performed componentwise by any low-pass filter, which is usually a Gaussian window or a rectangular window filter L_σ of width σ :

$$\bar{J}_x(t, f; \sigma) = \begin{pmatrix} L_\sigma * \left[\frac{\partial C_x}{\partial t}\right]^2 & L_\sigma * \left[\frac{\partial C_x}{\partial t} \frac{\partial C_x}{\partial f}\right] \\ L_\sigma * \left[\frac{\partial C_x}{\partial t} \frac{\partial C_x}{\partial f}\right] & L_\sigma * \left[\frac{\partial C_x}{\partial f}\right]^2 \end{pmatrix} \quad (39)$$

where $*$ is the usual 1-D convolution. This makes the gradient-square tensor insensitive to noise and irrelevant artifacts of scale smaller than $o(\sigma)$. Note that smoothing is also a requirement from a theoretical point of view to get an anisotropic diffusion process: Without it, ∇C_x is an eigenvector of the gradient-square tensor, and the flux is then collinear to the gradient. It can be shown that the matrix $\bar{J}_x(t, f; \sigma)$ is symmetric and positive semidefinite. Hence, there exists an orthonormal basis of eigenvectors v_1 and v_2 that can be used in (37) to determine D_{tf} . Now, we are in a position to propose an inhomogeneous anisotropic diffusion scheme to adapt the orientation and the strength of the smoothing operator to the TFR under analysis:

$$\begin{cases} \frac{\partial C_x(t, f; \tau)}{\partial \tau} = \text{div}(\eta_x(t, f) D_{tf} \nabla C_x(t, f; \tau)) \\ C_x(t, f; \tau = 0) = W_x(t, f) \end{cases} \quad (40)$$

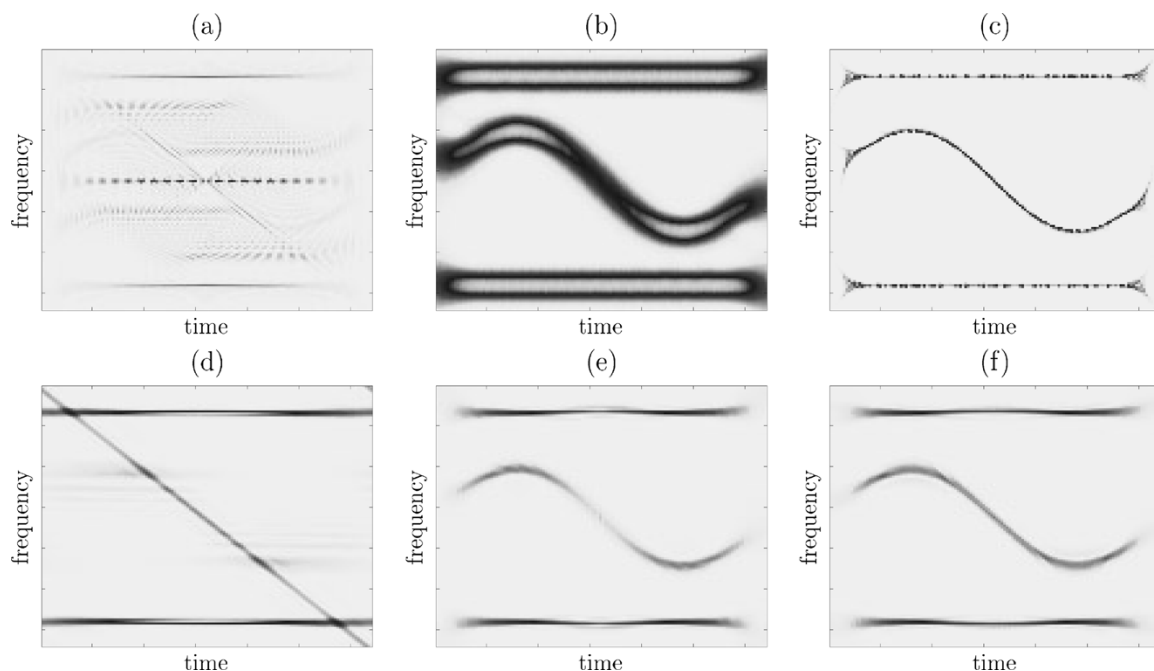


Fig. 7. Comparison of different TFRs. (a) WD. (b) Homogeneous diffusion. (c) Spectrogram modified by the reassignment method. (d) TFR with signal-dependent radially Gaussian kernel. (e) Adaptive diffusion. (f) Anisotropic adaptive diffusion. These 256-by-256 TFRs are those of a 256-sample signal that consists of a sinusoidally modulated signal and two waves.

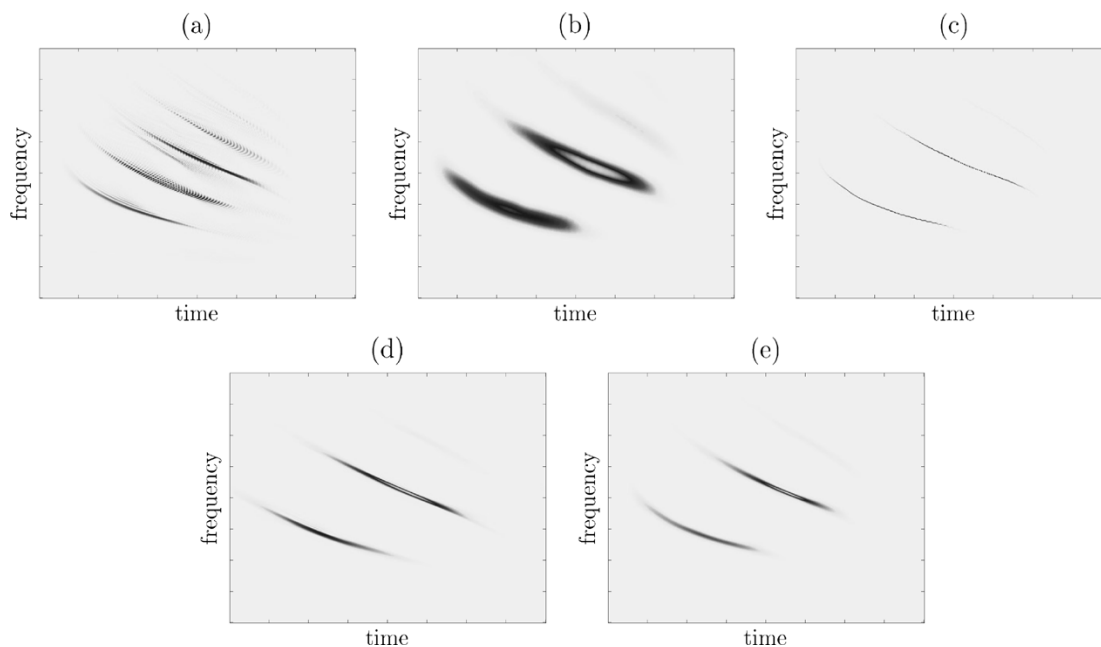


Fig. 8. Comparison of different TFRs applied to a 500-sample bat echolocation signal. (a) WD. (b) Spectrogram. (c) Spectrogram modified by the reassignment method. (d) TFR with signal-dependent radially Gaussian kernel. (e) Anisotropic adaptive diffusion. Signal kindly provided by C. Condon, K. White, and A. Feng from Beckman Institute, University of Illinois.

where $\eta_x(t, f)$ is a conductance function that can be chosen among those listed in Section IV-A. Just like the isotropic diffusion process (17), and because D_{tf} only involves derivatives, this model provides shift-covariant distributions if the conductance function is itself covariant with respect to time-frequency shifts. In addition, it conserves energy since it is of the form (12). Fig. 7 compares anisotropic adaptive diffusion with other approaches for improving the readability of TFRs of a synthetic signal. In Fig. 8, the same methods are used to represent

a natural signal of echolocation of a bat. The favorable effects of smoothing along signal components clearly appears on both examples.

B. Anisotropic Diffusion in the Affine Class

It is interesting to see that the anisotropic process (35) can also be used to process distributions of the affine class, provided that the scaling variable a is replaced by the frequency variable

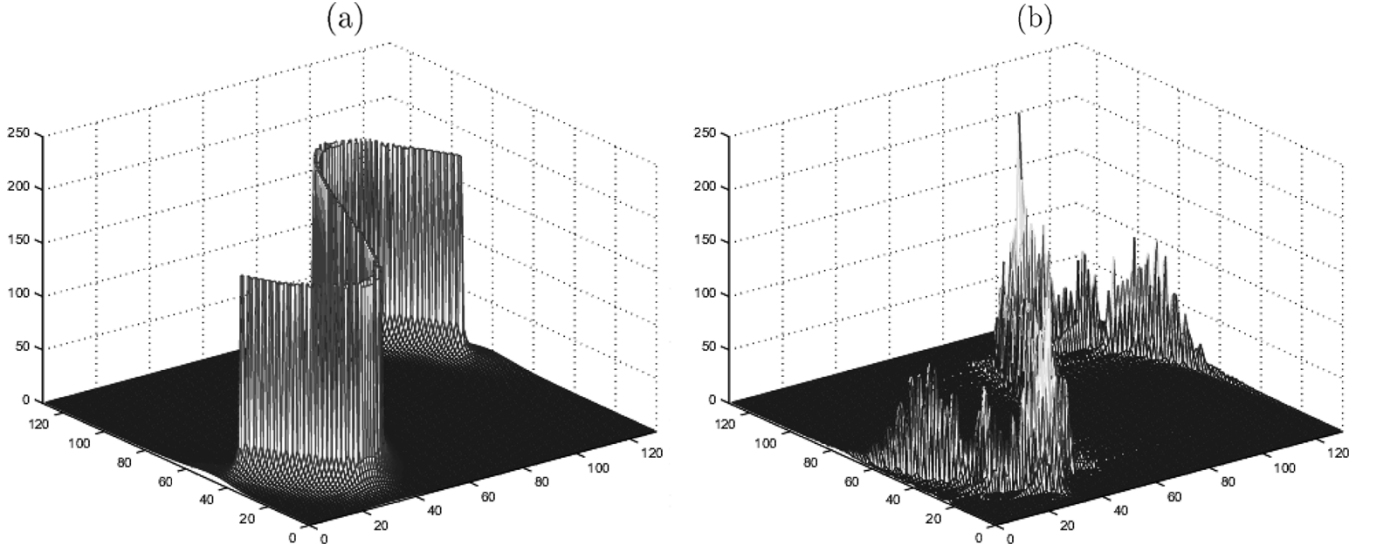


Fig. 9. Comparison of (a) a backward-and-forward diffusion with (b) the reassignment method on a sinusoidally-modulated signal.

f according to the correspondence rule $f \propto (1/a)$. It can easily be shown that

$$\begin{cases} \frac{\partial \Omega_x(t, f; \tau)}{\partial \tau} = \text{div}(D_f \nabla \Omega_x(t, f; \tau)) \\ \Omega_x(t, f; 0) = W_x(t, f) \end{cases} \quad \text{with } D_f = \begin{pmatrix} f^{-2} & 0 \\ 0 & f^2 \end{pmatrix} \quad (41)$$

produces distributions that satisfy $\Omega_{Ax}(t, f; \tau) = \Omega_x((t - t_0)/a_0, a_0 f; \tau)$. Note that a straightforward signal-dependent extension of (41) is obtained by adjusting the smoothing strength with a conductance function $\eta_x(t, f)$ built into the diffusion equation. Nevertheless, this approach is far from satisfactory as the smoothing orientation, which is given by D_f , does not depend on the signal to be analyzed. An intuitive solution to this problem would consist of substituting D_f with a gradient-square tensor based on an affine covariant gradient. However, such a measure of local orientation does not exist since angles in the time-frequency domain are not covariant with respect to the application of the affine operator.⁸ The time-frequency parameterization in (41) is responsible for this failure, although it is of current use in time-scale analysis. However, a close look at the definition of the gradient-square tensor applied to $\Omega_{Ax}(t, a; \tau)$ reveals that

$$\begin{aligned} J_{Ax}(t, a) &= \begin{pmatrix} \left[\frac{\partial \Omega_{Ax}}{\partial t} \right]^2 & \frac{\partial \Omega_{Ax}}{\partial t} \frac{\partial \Omega_{Ax}}{\partial a} \\ \frac{\partial \Omega_{Ax}}{\partial t} \frac{\partial \Omega_{Ax}}{\partial a} & \left[\frac{\partial \Omega_{Ax}}{\partial a} \right]^2 \end{pmatrix} \\ &= \frac{1}{a_0^2} J_x \left(\frac{t - t_0}{a_0}, \frac{a}{a_0} \right). \end{aligned}$$

A similar result also holds good for the smoothed gradient-square tensor. As a result, information on local orientation provided by eigenvectors of $J_x(t, a)$ or $\bar{J}_x(t, a; \sigma)$ is covariant with respect to time shifts and scale changes. It can be used in (37) to

determine the diffusion tensor D_{ta} of the following anisotropic inhomogeneous diffusion process:

$$\begin{cases} \frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = \text{div}(a^2 \eta_x(t, a) D_{ta} \nabla \Omega_x(t, a; \tau)) \\ \Omega_x(t, a; \tau = 0) = W_x(t, a). \end{cases} \quad (42)$$

Just like (33), this process is covariant with respect to the application of the affine operator \mathcal{A} , provided that the conductance function $\eta_x(t, a)$ is also covariant under its action. In addition, as a particular case of (12), it conserves energy.

VI. CONCLUSION

In this paper, we have focused our attention on adaptive diffusion for processing joint signal representations. We have emphasized the versatility of diffusion schemes as smoothing operators to improve the readability of bilinear time-frequency and time-scale representations. As a starting point, we revisited some standard representations, such as the spectrogram and the scalogram, in terms of diffusion processes. Next, we capitalized on recent results in image processing to design fully adaptive smoothed representations with enhanced localization capabilities and preserved covariance properties. Our approach improves on previous works by providing a unified framework not only for the Cohen class of TFRs but for the affine class of TSRs as well. The latter is of particular interest because, to the best of our knowledge, no signal-dependent smoothing solution exists to process distributions of the affine class, nor even a transposition of optimal radially Gaussian kernels of Baraniuk and Jones.

In an ongoing study, we are investigating selective forward-and-backward diffusion processes [32], [33] in order to simultaneously reach a high concentration in signal components and a cross-term removal in the same spirit of using distributions modified by the reassignment method [17]. Fig. 9 is a promising illustration of the potential interest of such a method when applying a negative-valued conductance function.

REFERENCES

⁸We suggest referring to [41] and references therein for a discussion on first- and second-order affine invariant/covariant descriptors.

- [1] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

- [2] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. San Diego, CA: Academic, 1999.
- [3] S. Kadambe and G. F. Boudreaux-Bartels, "A comparison of the existence of "cross terms" in the Wigner distribution and the squared magnitude of the wavelet transform and the short time Fourier transform," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2498–2517, Oct. 1992.
- [4] H. Choi and W. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 6, pp. 862–871, June 1989.
- [5] L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7, no. 5, pp. 781–786, 1966.
- [6] W. Williams and J. Jeong, *Time-Frequency Signal Analysis—Methods and Applications*, Longman Cheshire, London, U.K., pp. 74–97, 1992.
- [7] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1084–1091, Jul. 1990.
- [8] R. Baraniuk and D. Jones, "Signal-dependent time-frequency analysis using a radially Gaussian kernel," *Signal Process.*, vol. 2, no. 3, pp. 263–284, 1993.
- [9] D. Jones and R. Baraniuk, "An adaptive optimal-kernel time-frequency representation," *IEEE Trans. Signal Process.*, vol. 43, no. 10, pp. 2361–2371, Oct. 1995.
- [10] D. Brie, "La classe des distributions temps-fréquence locales: définition et propriétés," in *Proc. XIX Colloque Gretsi*, 2001.
- [11] P. Gonçalves and E. Payot, "Adaptive diffusion equation for time-frequency representations," in *Proc. IEEE Digital Signal Process. Workshop*, 1998.
- [12] J. Gosme, C. Richard, and P. Gonçalves, "Anisotropic diffusion equations for adaptive quadratic representations," in *Proc. XII Eur. Signal Process. Conf.*, 2004.
- [13] —, "Diffusion equations for adaptive affine distributions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2004.
- [14] O. Rioul and P. Flandrin, "Time-scale energy distributions: a general class extending wavelet transform," *IEEE Trans. Signal Process.*, vol. 40, no. 11, pp. 1746–1757, Nov. 1992.
- [15] J. Bertrand and P. Bertrand, "A class of affine Wigner functions with extended covariance properties," *J. Math. Phys.*, vol. 33, no. 7, pp. 2515–2527, 1992.
- [16] P. Flandrin and O. Rioul, "Affine smoothing of the Wigner–Ville distribution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1990, pp. 2455–2458.
- [17] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by reassignment methods," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May 1995.
- [18] S. Qian and J. Morris, "Wigner distribution decomposition and cross-terms deleted representation," *Signal Process.*, vol. 27, pp. 125–144, 1992.
- [19] S. Qian and D. Chen, "Decomposition of the Wigner–Ville distribution and time-frequency distribution series," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2836–2842, Oct. 1994.
- [20] M. Pasquier, P. Gonçalves, and R. Baraniuk, "Hybrid linear/bilinear time-scale analysis," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 254–259, Jan. 1998.
- [21] J. Crank, *The Mathematics of Diffusion*. London, U.K.: Oxford Univ. Press, 1957.
- [22] J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, pp. 363–370, 1984.
- [23] T. Iijima, "Basic equation of figure and observational transformation," *Syst., Comput., Control*, vol. 2, no. 4, pp. 70–77, 1971.
- [24] J. Weickert, S. Ishikawa, and A. Imiya, "Linear scale-space has first been proposed in Japan," *J. Math. Imaging Vision*, vol. 10, pp. 237–252, 1999.
- [25] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," in *Proc. IEEE Comput. Soc. Workshop Comput. Vision*, 1987, pp. 16–27.
- [26] —, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [27] F. Voci, S. Eiho, N. Sugimoto, and H. Sekibuchi, "Estimating the gradient in the Perona–Malik equation," *IEEE Signal Process. Mag.*, vol. 21, no. 3, pp. 39–65, Mar. 2004.
- [28] J. Weickert, *Anisotropic Diffusion in Image Processing*. Stuttgart, Germany: B. G. Teubner, 1998.
- [29] —, "A review of nonlinear diffusion filtering," in *Proc. First Int. Conf. Scale-Space Theories Comput. Vision*, vol. 1252, 1997, pp. 3–28.
- [30] L. Alvarez, F. Guichard, P. Lions, and J. Morel, "Axioms and fundamental equations of image processing," *Arch. Rational Mechanics Anal.*, vol. 123, no. 3, pp. 199–257, 1993.
- [31] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
- [32] G. Gilboa, N. Sochen, and Y. Zeevi, "Forward-and-backward diffusion processes for adaptive image enhancement and denoising," *IEEE Trans. Image Process.*, vol. 11, no. 7, pp. 689–703, Jul. 2002.
- [33] B. Smolka, "Combined forward and backward anisotropic diffusion filtering of color images," in *Proc. 24th DAGM Symp. Pattern Recogn.*, 2003, pp. 314–321.
- [34] J. Weickert, B. ter Haar Romeny, and M. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Trans. Image Process.*, vol. 7, pp. 398–410, Jul. 1998.
- [35] J. W. H. Scharf, "A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance," *J. Visual Commun. Image Represent.*, vol. 13, pp. 103–118, 2002.
- [36] R. Baraniuk, P. Flandrin, M. Jansen, and O. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1391–1409, May 2001.
- [37] M. Ferraro, G. Boccignone, and T. Caelli, "On the representation of image structures via scale space entropy conditions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 11, pp. 1199–1203, Nov. 1999.
- [38] J. Sporring and J. Weickert, "Information measures in scale-spaces," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1051–1058, Apr. 1999.
- [39] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
- [40] W. Förstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," in *Proc. ISPRS Intercommis. Conf. Fast Process. Photogrammetric Data*, Interlaken, Switzerland, 1987, pp. 281–305.
- [41] P. Olver, G. Sapiro, and A. Tannenbaum, "Affine invariant detection: edge maps, anisotropic diffusion, and active contours," *Acta Applicandae Mathematicae*, vol. 59, pp. 45–77, 1999.



Julien Gosme received the Dipl.-Ing. degree in information and decision making systems in 2000 and the M.S. degree in system optimization and reliability in 2001 from Troyes University of Technology, Troyes, France, where he is pursuing the Ph.D. degree with the ISTIT Laboratory (FRE CNRS 2732).

His scientific interests are in the fields of adaptive joint signal processing and kernel methods.



Cédric Richard (M'01) was born in Sarrebourg, France, on January 24, 1970. He received the Dipl.-Ing. and the M.S. degrees in 1994 and the Ph.D. degree in 1998 from Compiègne University of Technology, Compiègne, France, all in electrical and computer engineering.

From 1999 to 2003, he was an Associate Professor with Troyes University of Technology, Troyes, France, where, since 2003, he has been a Professor at the ISTIT Laboratory (FRE CNRS 2732). His current research interests involve time-frequency analysis, statistical estimation and decision theories, and pattern recognition.

Paulo Gonçalves graduated from the Signal Processing Department of ICPI, Lyon, France in 1993. He received the Masters (DEA) and Ph.D. degrees in signal processing from the Institut National Polytechnique, Grenoble, France, in 1990 and 1993, respectively. While working toward the Ph.D. degree, he was with École Normale Supérieure, Lyon.

From 1994 to 1996, he was a Postdoctoral Fellow at Rice University, Houston, TX. He then joined the research group FRACTALES at INRIA, Rocquencourt, France, as researcher associate. From 2000 to 2003, he was with group IS2 at INRIA Rhône-Alpes, France, and since 2003, he has been on leave with IST-ISR, Lisbon, Portugal. His research interests are in multiscale signal and image analysis and in wavelet-based statistical inference.