Chapter 7

Quadratic Time-Frequency Analysis III: The Affine Class and Other Covariant Classes

Abstract: Affine time-frequency distributions appeared towards the middle of the 1980s with the emergence of wavelet theory. The affine class is built upon the principle of covariance of the affine group, i.e., contractions-dilations and translations in time. This group provides an interesting alternative to the group of translations in time and in frequency, which forms the basis for the conventional time-frequency distributions of Cohen's class. More precisely, as the Doppler effect on "broadband" signals is expressed in terms of contractions-dilations, it is for the analysis of this category of signals that the affine class is particularly destined. The objective of this chapter is to present the various approaches for constructing the affine class and the associated tools devised over the past years. We will demonstrate how the latter supported the introduction of new mathematical concepts in signal processing – group theory, operator theory – as well as of new classes of covariant time-frequency distributions.

Keywords: affine time-frequency analysis, wavelets, covariance principle, affine group, affine Wigner distributions, Bertrand distributions, unitary equivalence, hyperbolic class, power classes.

7.1. Introduction

The concept of *affine time-frequency representation* was introduced for the first time in 1985 [BER 85], while wavelet theory was simultaneously and independently being developed. Relying on the same formalism of signal deformation – contraction-dilation and temporal translation [GRO 84] – , it gave rise to what is now commonly

---

Chapter written by Paulo GONÇALVÈS, Jean-Philippe OVARLEZ and Richard BARANIUK.

referred to as the unitary affine Bertrand distribution. Originally derived by a tomographic construction, this distribution now plays a central role in the study of affine representations. In fact, the wavelet transform is directly connected to affine time-frequency representations by means of a smoothing operation in the time-frequency plane, which explains the great importance of affine time-frequency representations in signal analysis. Rigourous but not easily accessible, the first construction of these distributions was based on group theory [BER 92a, BER 92b], a powerful tool in signal analysis, which did not, however, stir up an enthusiasm comparable with that caused by the study of Cohen's class distributions [COH 66]. More accessible as it is more heuristic, a second approach [RIO 92] relies on the affine smoothing of certain distributions of Cohen's class.

This chapter presents these two different approaches, with their own advantages and drawbacks, which both prompted the emergence of new mathematical techniques in signal processing. At the end of the chapter, we will also illustrate the unitary equivalence principle by presenting new covariant classes, images of Cohen's class or of the affine class by the action of unitary operators.

7.2. General construction of the affine class

7.2.1. *Bilinearity of distributions*

Following the construction lines of Cohen's class introduced in Chapter 5, we describe here the principles which make it possible to define the affine class of bilinear time-frequency representations. As is the case for Cohen's class, we are interested in signal energy distributions. The arguments developed in Chapter 1 then justify the natural (although not necessary) choice of bilinear forms of the analyzed signal. We will briefly point out some of these arguments. Let $x(t)$ be a complex signal from $L^2(\mathbb{R})$ and $\widehat{x}(f)$ its Fourier transform. The squared norm of x or of its isometric $\widehat{x}$ (Parseval's theorem) defines the signal energy

$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |\widehat{x}(f)|^2 \, df .$$

The time-frequency energy representations, denoted by $\rho_x(t,f)$, stand as a hybrid intermediary between the two energy distributions constituted by the instantaneous power $|x(t)|^2$ and the spectral energy density $|\widehat{x}(f)|^2$. By combining the two variables t and f, $\rho_x(t,f)$ describes (topographically) how the energy in x is distributed over the time-frequency plane. Naturally, it is then expected that these joint distributions will preserve the total energy of the signal:

$$E_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_x(t,f) \, dt \, df , \tag{7.1}$$

without, however, imposing for their marginal distributions, $\int_{-\infty}^{\infty} \rho_x(t,f)\, df$ and $\int_{-\infty}^{\infty} \rho_x(t,f)\, dt$, to be systematically equal to the instantaneous power and to the spectral energy density, respectively.

In accordance with Chapters 1 and 5, we agree to write these bilinear energy distributions in a generic form involving indifferently the signal in time or in frequency according to

$$\rho_x(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(t_1,t_2;t,f)\,x(t_1)\,x^*(t_2)\,dt_1\,dt_2 \tag{7.2}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widehat{K}(f_1,f_2;t,f)\,\widehat{x}(f_1)\,\widehat{x}^*(f_2)\,df_1\,df_2\,. \tag{7.3}$$

In these expressions, the product $x(t_1)\,x^*(t_2)$ (respectively $\widehat{x}(f_1)\,\widehat{x}^*(f_2)$) is called the *useful part* of the signal and $K(t_1,t_2;t,f)$ (respectively $\widehat{K}(f_1,f_2;t,f)$) is an arbitrary parameterization kernel fully characterizing the properties of the distribution $\rho_x(t,f)$. The two forms $K(t_1,t_2;t,f)$ and $\widehat{K}(f_1,f_2;t,f)$ are connected by the double Fourier transformation:

$$K(t_1,t_2;t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widehat{K}(f_1,f_2;t,f)\,e^{-j2\pi(f_1t_1-f_2t_2)}\,df_1\,df_2\,.$$

The energy conservation constraint (7.1) adds a little more to the specification of the parameterization kernel K, which then has to satisfy the condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(t_1,t_2;t,f)\,dt\,df = \delta(t_1-t_2)\,, \tag{7.4}$$

where $\delta(\cdot)$ is the Dirac distribution.

A simple example of a bilinear time-frequency representation is provided by the squared modulus of a linear decomposition (projection onto the analyzing function $g(u;t,f)$ centered around time t and frequency f). In fact,

$$\left|\int_{-\infty}^{\infty} x(u)\,g^*(u;t,f)\,du\right|^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(u)\,x^*(u')\,\underbrace{g^*(u;t,f)\,g(u';t,f)}_{K(u,u';t,f)}\,du\,du'$$
$$= \rho_x(t,f)\,.$$

This, in particular, is the case of the *spectrogram* (squared modulus of the short-time Fourier transform), which, in fact, belongs to Cohen's class (see Chapters 1 and 5).

7.2.2. *Covariance principle*

To identify subclasses of specific solutions among the entirety of possible bilinear representations, we can constrain the structure of the arbitrary kernel K parameterizing the generic form (7.2), by requiring the distributions to satisfy certain analytical properties. For example, constraint (7.4) already guarantees that the associated distribution preserves the energy of the analyzed signal. Another constraint which appears natural in signal processing is that of covariance of the distributions to one or several transformations applied to the signal. Thus, if by $\boldsymbol{T}$ we designate an operator of

$L^2(\mathbb{R}) \to L^2(\mathbb{R})$ acting in the signal space and by $\widetilde{\boldsymbol{T}}$ the associated operator acting in the space $L^1(\mathbb{R}^2)$ of bilinear time-frequency distributions, imposing a principle of covariance on these operators amounts to the following identity:

$$\rho_{\boldsymbol{T}x}(t,f) = \widetilde{\boldsymbol{T}}\rho_x(t,f)\,. \tag{7.5}$$

Then, to a particular choice of the operator $\boldsymbol{T}$ (and of its dual[1] $\widetilde{\boldsymbol{T}}$), there corresponds a class of representations ρ_x, solutions of the associated covariance principle (7.5).

7.2.2.1. *Covariance to time and frequency shifts*

A natural first choice for the operator $\boldsymbol{T}$ in signal processing is that of translations in time and frequency, since these play a central role for linear time-invariant systems and for frequency modulations. This operator acts on the signal according to $\boldsymbol{T}_{t_0 f_0} x(t) = x(t-t_0)\, e^{j2\pi f_0 t}$, and its dual in the space of representations transforms ρ_x into $\widetilde{\boldsymbol{T}}_{t_0 f_0}\rho_x(t,f) = \rho_x(t-t_0, f-f_0)$. The only solutions ρ_x satisfying the associated covariance principle (7.5) are the time-frequency representations of *Cohen's class* [COH 66], whose kernels are reduced to the form $K(t_1,t_2;t,f) = K(t_1-t, t_2-t; 0,0)\, e^{-j2\pi f(t_1-t_2)}$ (see also Chapter 5). With this particular structure of kernels that depends only on two variables, the initial expression (7.2) can be rewritten in the form of a double linear filtering

$$C_x(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_{\text{t-f}}(t-t', f-f')\, W_x(t',f')\, dt'\, df' \tag{7.6}$$

of the *Wigner-Ville distribution*

$$W_x(t,f) := \int_{-\infty}^{\infty} x\Big(t+\frac{\tau}{2}\Big)\, x^*\Big(t-\frac{\tau}{2}\Big)\, e^{-j2\pi f\tau}\, d\tau\,, \tag{7.7}$$

by the reparametrized kernel $\phi_{\text{t-f}}(t,f) = \int_{-\infty}^{\infty} K(-t+\tau/2, -t-\tau/2; 0,0)\, e^{-j2\pi f\tau}\, d\tau$. Let us recall that this class admits several possible formulations, which are discussed in Chapter 5 and in [FLA 99].

7.2.2.2. *Covariance to the affine group*

Scale change (dilation/compression) is another operator at the core of an abundant mathematical literature, and which has been arousing a growing interest in signal analysis for the past 20 or so years. Initially used to model a certain number of natural physical signal transformations, like the Doppler effect, its use has then been extended with the development of wavelet techniques [GRO 84, DAU 92, MAL 89, MAL 99] and fractal analysis [MAN 68, ARN 95].

1. The concept of a dual operator is borrowed here from the work of J. and P. Bertrand. Thus, it must be understood as the image operator of $\boldsymbol{T}$ acting in the phase space of the signal representation, and not in its traditional sense.

The covariance (7.5), which we are now interested in, is related to time shifts and to scale changes (scalings). It is based on the affine group of two parameters, which is denoted by $A(a,b)$, where a, real positive, designates the scale parameter and b, real, represents the time shift. The group operation associated with the affine group is $(a,b)\,(a',b') = (a\,a', b + a\,b')$. It corresponds to a *clock change* on the time axis, $t \mapsto a\,t + b$, and is represented unitarily and irreducibly in the Hardy space $L^2(\mathbb{R}_*^+)$ of analytic signals according to

$$\begin{aligned} \boldsymbol{T}_{a,b} : \quad x(t) \quad &\longrightarrow \quad x_{a,b}(t) \;=\; \frac{1}{\sqrt{a}}\; x\!\left(\frac{t-b}{a}\right) \\ \widehat{x}(f) \quad &\longrightarrow \quad \widehat{x}_{a,b}(f) \;=\; \sqrt{a}\; e^{-j2\pi f b}\, \widehat{x}(af)\,. \end{aligned} \tag{7.8}$$

The covariance requirement (7.5) expressed with this new set of displacement operators of the time-frequency plane then becomes

$$\rho_{\boldsymbol{T}_{a,b}x}(t,f) \;=\; \rho_x\!\left(\frac{t-b}{a}, af\right). \tag{7.9}$$

Note 1: As mentioned previously, the operators of the affine group (scaling and time shift) make it possible to mathematically describe the Doppler effect due to the emission of a wave by a moving source, and the time of propagation of this wave to the target. In the borderline case of narrowband signals, we can reasonably approximate to the first order the effect of dilation (or of compression) by a simple frequency shift of the spectral content of the emitted wave. Using the same reasoning, we can then predict that the representations $\rho_x(t,f)$ that are solutions of the affine covariance principle (7.9) will behave in the narrowband limit as Cohen's class distributions.

Note 2: Let us apply the affine transformation (7.8) ($b = 0$) to a unimodal function $\Psi(f)$ localized in the Fourier space around an arbitrary frequency $f_0 > 0$. The dilated (or compressed) version $\sqrt{a}\;\Psi(af)$ is then localized around the frequency $f_a = f_0/a > 0$. Thus, when the scale parameter a varies continuously between zero and infinity, the modal frequency f_a explores the entire frequency axis $\mathbb{R}_*^+$. Actually, this example demonstrates that it is sometimes possible to relate the scale parameter a to a strictly positive characteristic frequency f_a, using the formal identification $a = f_0/f_a$. Under certain conditions [RIO 92], this identification will also apply to certain time-frequency representations $\rho_x(t,f)$ that are covariant to the affine group transformations (7.8), and that can be interpreted just as well as time-scale representations $\rho_x(t,a)$. We then pass from one of these writings to the respective other via the mapping $a = f_0/f$, possibly setting (without loss of generality) $f_0 = 1$, and without forgetting the associated measure $da = df/f^2$. We have to be careful though, for this identification is not without consequences for the definition (7.3) of affine bilinear forms, whose integration bounds should now be restricted to the quarter-plane $\mathbb{R}_*^+ \times \mathbb{R}_*^+$.

Similarly, based on the temporal formulation of the dilation operator (7.8) ($b = 0$), we could establish a formal correspondence between the scale parameter a and a characteristic time $t = a\,t_0 > 0$ (with an arbitrarily chosen reference time $t_0 > 0$).

7.2.3. *Affine class of time-frequency representations*

Once the affine covariance principle has been formulated, we now have to identify among all bilinear forms (7.2) or (7.3) the time-frequency representations that are covariant to time shifts and scale changes. The affine class of time-frequency representations resulting from this restriction has been studied at approximately the same period by two teams [BER 85, BER 92b] and [FLA 90, FLA 91, RIO 92], each following an original and independent approach. However, although the results obtained in either case proceed from relatively distant motivations, what distinguishes them is essentially an arbitrary initial choice K of kernel parameterization. We here decided to reproduce these reasonings within a unified framework, while not hesitating, whenever necessary, to adopt the parameterization that either lends itself better to result interpretation, or opens a more general perspective.

Let us reconsider the affine transformation (7.9), where we develop $\rho_x(t,f)$ in its generic frequency form (7.3). The covariance constraint directly translates to the kernel $\widehat{K}$ via the equation

$$\widehat{K}(f_1, f_2; t, f) \;=\; a\,\widehat{K}\left(af_1, af_2; \frac{t-b}{a}, af\right) e^{j2\pi b(f_1-f_2)}.$$

This equation must be satisfied for any value of the pair (a, b). Thus, in particular, for $b = t$ and $a = f_0/f > 0$, with $f_0 > 0$ fixed, we must have

$$\widehat{K}(f_1, f_2; t, f) \;=\; \frac{f_0}{f}\,\widehat{K}\left(\frac{f_1 f_0}{f}, \frac{f_2\, f_0}{f}; 0, f_0\right) e^{j2\pi t(f_1-f_2)}.$$

The kernel $\widehat{K}$ then reduces to a form only dependent on two variables (instead of four), which we shall hereafter denote more simply as $\widehat{K}(f_1, f_2; 0, f_0) =: \widehat{K}(f_1, f_2)$. Finally, inserting this kernel into the canonical form (7.3), we obtain the following result.

The affine class [BER 92a, RIO 92] comprising all bilinear time-frequency representations covariant to time shifts and scale changes is given by the following parametric form

$$\Omega_x(t,f) \;=\; \frac{f}{f_0}\int_0^{\infty}\!\!\int_0^{\infty} \widehat{K}(f_1, f_2)\; \widehat{x}\left(\frac{f_1 f}{f_0}\right) \widehat{x}^*\left(\frac{f_2\, f}{f_0}\right) \cdot\, e^{j2\pi f t\,(f_1-f_2)/f_0}\, df_1\, df_2\,, \qquad f > 0\,, \qquad (7.10)$$

where $\widehat{K}$ is a two-dimensional parameterization kernel and f_0 is an arbitrary positive reference frequency that we shall, without loss of generality, suppose to be equal to 1 Hz hereafter.

In accordance with the previous notes on the scaling operator, expression (7.10) involves only frequency components of $x(t)$ lying on the half-line $\mathbb{R}_*^+$. Defined only for positive frequencies, $\Omega_x(t,f)$ thus implicitly resorts to the analytic signal (obtained by canceling the frequency components $\widehat{x}(f)$, $\forall f<0$ of the real signal).

Moreover, should we wish the distributions $\Omega_x(t,f)$ to be real, the kernel $\widehat{K}(f_1,f_2)$ must then itself be real and symmetric (i.e., $\widehat{K}(f_1,f_2)=\widehat{K}(f_2,f_1)$). Other desirable properties of $\Omega_x(t,f)$ will be considered in Section 7.3.

Once again, let us recall that there are other possible parameterizations of this class. For example, if we introduce the "frequency-Doppler (f-D)" kernel

$$\phi_{\text{f-D}}(\nu,\xi) \;=\; \widehat{K}\left(\nu+\frac{\xi}{2},\nu-\frac{\xi}{2}\right), \qquad |\xi|<2\nu\,, \tag{7.11}$$

expression (7.10) is rewritten in a completely equivalent way:

$$\Omega_x(t,f) \;=\; f\int_0^\infty d\nu \int_{-2\nu}^{2\nu} \phi_{\text{f-D}}(\nu,\xi)\; \widehat{x}\left(f\left(\nu+\frac{\xi}{2}\right)\right)\, \widehat{x}^*\left(f\left(\nu-\frac{\xi}{2}\right)\right) \cdot e^{j2\pi\xi f t}\, d\xi\,, \quad f>0\,. \tag{7.12}$$

To this second formulation of the affine class, we can add three other alternative expressions, making use of partial Fourier transformations on the variables of the kernel $\phi_{\text{f-D}}(\nu,\xi)$ [FLA 99]

$$\phi_{\text{f-D}}(\nu,\xi) \;=\; \int_{-\infty}^{\infty} \phi_{\text{t-f}}(u,\nu)\, e^{j2\pi\xi u}\, du \tag{7.13}$$

$$=\; \int_{-\infty}^{\infty} \phi_{\text{d-D}}(\tau,\xi)\, e^{j2\pi\nu\tau}\, d\tau \tag{7.14}$$

$$=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_{\text{t-d}}(u,\tau)\, e^{j2\pi(\xi u+\nu\tau)}\, du\, d\tau\,. \tag{7.15}$$

Depending on the selected parameterization (*frequency-Doppler* (7.11), *time-frequency* (7.13), *delay-Doppler* (7.14) or *time-delay* (7.15)), we obtain different expressions of the affine class that are all equivalent to (7.12). In particular, the expression in terms of the *time-frequency* kernel $\phi_{\text{t-f}}(u,\nu)$ clarifies the central role of the Wigner-Ville distribution (7.7), since then

$$\Omega_x(t,f) \;=\; \int_0^\infty\int_{-\infty}^{\infty} W_x(u,\nu)\; \phi_{\text{t-f}}\left(f\,(u-t),\frac{\nu}{f}\right)\, du\, d\nu\,, \qquad f>0\,. \tag{7.16}$$

The operation of affine correlation appearing in this expression calls for several comments to be made. First of all, it shows that the Wigner-Ville distribution is itself an affine class distribution, for the particular kernel $\phi_{\text{t-f}}(u,\nu)\;=\;\delta(u)\,\delta(\nu-1)$. As it also belongs to Cohen's class (7.6), the Wigner-Ville distribution is covariant to an extension of the affine group to three parameters: translation in time, scale change and translation in frequency. A simple calculation then shows that only the affine covariance is naturally preserved by this affine correlation:

$$\begin{aligned}\Omega_{x_{a,b}}(t,f) &= \int_0^{\infty}\int_{-\infty}^{\infty} W_x\left(\frac{u-b}{a}, a\,\nu\right)\phi_{\text{t-f}}\left(f\,(u-t), \frac{\nu}{f}\right) du\,d\nu \\ &= \int_0^{\infty}\int_{-\infty}^{\infty} W_x(u',\nu')\;\phi_{\text{t-f}}\left(af\left(u'-\frac{t-b}{a}\right), \frac{\nu'}{af}\right) du'd\nu' \\ &= \Omega_x\left(\frac{t-b}{a}, af\right).\end{aligned}$$

It is therefore not surprising that all distributions Ω_x resulting from an *affine filtering* of the Wigner-Ville distribution with any parameterization function $\phi_{\text{t-f}}$ systematically belong to the affine class. However, the converse may be astonishing, i.e., that *any* distribution of the affine class can be interpreted as an *affine filtering* of the Wigner-Ville distribution. Two points should be underlined. On the one hand, the arbitrariness in the choice of the new parameterization (7.11) could just as easily lead to another *generating* distribution of the affine class. In this respect, not only does the Wigner-Ville distribution not play a unique role, but also it can be shown [FLA 99] that any affine distribution $\Omega_x(t,f)$ in bijective correspondence (up to some phase) with the analyzed signal can generate the affine class following a pattern similar to that of relation (7.16). On the other hand, there is no *a priori* restriction on the structure of the kernel $\phi_{\text{t-f}}$, which authorizes choices outpacing the framework of (low-pass type) smoothing kernels. In particular, we will see that certain *localized bi-frequency kernels* (oscillating structures) correspond to elements of the affine class that are very distant from the Wigner-Ville distribution.

The narrowband ambiguity function (or Woodward ambiguity function [WOO 53]) is related to the Wigner-Ville distribution by a double Fourier transformation:

$$\begin{aligned}A_x(\tau,\xi) &:= \int_{-\infty}^{\infty} x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi\xi t}\,dt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_x(t,f)\; e^{-j2\pi(\xi t - f\tau)}\,dt\,df\,.\end{aligned}$$

This other quadratic form of the signal leads to a new canonical expression of the affine class, dual to (7.16), and bringing into play the *delay-Doppler* version (7.14) of the parameterization kernel:

$$\Omega_x(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_x(\tau,\xi)\;\phi_{\text{d-D}}\left(f\tau, \frac{\xi}{f}\right) e^{j2\pi\xi t}\,d\tau\,d\xi\,, \qquad f>0\,. \tag{7.17}$$

Finally, let us mention a last expression equivalent to (7.16):

$$\Omega_x(t,f) = f\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_{\text{t-d}}\big(f\,(u-t), f\tau\big)\; x\left(u+\frac{\tau}{2}\right) x^*\left(u-\frac{\tau}{2}\right) du\,d\tau\,, \qquad f>0\,.$$

Based on the *time-delay* expression (7.15) of the parameterization kernel, it sheds light on the relationship existing between affine time-frequency representations and a local correlation function of the signal. In fact, if for the sake of concreteness we suppose the kernel $\phi_{\text{t-d}}(u, \tau)$ to be low-pass with respect to u and to oscillate with respect to τ, we can interpret the integral

$$r_x(t, \tau; f) \;=\; f \int_{-\infty}^{\infty} x\Big(u + \frac{\tau}{2}\Big)\, x^*\Big(u - \frac{\tau}{2}\Big)\, \phi_{\text{t-d}}\big(f(u-t), f\tau\big)\, du$$

as a sliding-window (or short-term) estimator of the autocorrelation function of the signal x (see Chapter 5). The oscillating integral with respect to variable τ then amounts to a band-pass filtering (similar to a Fourier transform) applied to this local autocorrelation function. The affine nature of these *evolutionary spectra*, and what distinguishes them from the time-frequency representations of Cohen's class, is due to parameter f appearing as a pre-factor of the variables of convolution kernel $\phi_{\text{t-d}}(u, \tau)$. The latter introduces a functional dependence between the characteristics of the estimator (resolutions in time and frequency) and the analyzing scale $1/f$.

7.3. Properties of the affine class

Regardless of the parametric form retained for the formulation of the affine class, we will show, using some examples, that imposing a property on Ω always implies the introduction of a corresponding structural constraint on the parameterization kernel. Thanks to this correspondence, it becomes relatively easy to isolate among all distributions of the affine class those satisfying a given combination of desired properties by resolving the system of associated constraints. However, as some of these admissibility conditions on the kernel may be mutually exclusive, the system might have no solution, meaning that there does not exist an affine time-frequency representation satisfying all the desired properties.

We shall now present some examples of properties and corresponding constraints, using in each case the parameterization most appropriate for interpreting the respective result. A more complete list of properties and constraints can be found in [RIO 92, FLA 99].

7.3.1. *Energy*

Constraint (7.4) guaranteeing that generic bilinear forms preserve signal energy can be specified within the affine framework. Thus, distributions Ω_x of the affine class satisfy the energy preservation property

$$\int_0^{\infty} \int_{-\infty}^{\infty} \Omega_x(t, f)\; dt\, df \;=\; E_x$$

if and only if the associated kernel satisfies the constraint

$$\int_0^{\infty} \frac{\phi_{\text{f-D}}(\nu, 0)}{\nu}\; d\nu \;=\; 1\,. \tag{7.18}$$

7.3.2. *Marginals*

Disregarding problems caused by the non-positivity of time-frequency representations, we can, as for Cohen's class, interpret the affine distributions satisfying (7.18) as energy densities (in the *probabilistic* sense). As the energy spectrum density $|\widehat{x}(f)|^2$ (or the instantaneous power $|x(t)|^2$) is the frequency (or time) distribution of the energy of x, we may desire that the integral of time-frequency distributions $\Omega_x(t,f)$ with respect to the time (or frequency) variable be equal to the marginal distribution $|\widehat{x}(f)|^2$ (or $|x(t)|^2$). In each one of these cases, there is then a structural constraint associated with the distribution kernel:

– *marginal distribution in frequency:*

$$\int_{-\infty}^{\infty} \Omega_x(t,f)\,dt \;=\; |\widehat{x}(f)|^2 \qquad \Longleftrightarrow \qquad \phi_{\text{f-D}}(\nu,0) \;=\; \delta(\nu-1)\,;$$

– *marginal distribution in time:*

$$\int_{0}^{\infty} \Omega_x(t,f)\,df \;=\; |x(t)|^2 \qquad \Longleftrightarrow \qquad \int_0^{\infty} \phi_{\text{d-D}}\left(f\tau, \frac{\xi}{f}\right) df \;=\; \delta(\tau)\,,\ \forall \xi\,.$$

Naturally, each of these marginals satisfied individually automatically implies the energy preservation property.

7.3.3. *Unitarity*

Affine class distributions are bilinear applications of the one-dimensional vector space $L^2(\mathbb{R}, dt)$ of analytic signals with finite energy to the two-dimensional vector space $L^1(\mathbb{R}\times\mathbb{R}_*^+, dt\,df)$. Each of these spaces has a scalar product (and an associated Haar measure). Although there cannot be a one-to-one topological mapping between these two spaces, it is, nevertheless, possible to find distributions of the affine class that leave distances invariant while preserving the scalar products of each of the spaces. For these distributions, known as *unitary* or *isometric*, we have the identity given by Moyal's formula [MOY 49]

$$\int_0^{\infty}\int_{-\infty}^{\infty} \Omega_{x_1}(t,f)\;\Omega_{x_2}^*(t,f)\;dt\,df \;=\; \left|\int_{-\infty}^{\infty} x_1(t)\,x_2^*(t)\,dt\right|^2. \tag{7.19}$$

We note that this identity does not necessarily imply energy preservation, but places a condition on the quadratic norm of the distribution, since then $\int_0^{\infty}\int_{-\infty}^{\infty}|\Omega_x(t,f)|^2\,dt\,df = E_x^2$. The converse implication is clearly false. Then, developing the affine class formulation (7.17) in (7.19), it is seen that Moyal's formula imposes a particularly constraining condition on the parameterization kernel [RIO 92, FLA 99, HLA 93]:

$$\int_0^{\infty} \phi_{\text{d-D}}\left(f\tau, \frac{\xi}{f}\right) \phi_{\text{d-D}}^*\left(f\tau', \frac{\xi}{f}\right) df \;=\; \delta(\tau-\tau') \qquad \forall\,\xi\,.$$

We finally note that the Wigner-Ville distribution, calculated for analytic signals, satisfies the unitarity property (7.19).

The isometry (7.19) is an interesting mathematical bridge that makes it straightforward to transpose results established in linear filter theory to the context of time-frequency representations. As an example of great importance in signal processing, let us consider the *continuous wavelet decomposition* (see Chapter 4)

$$\mathrm{CWT}_x^{\psi}(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(u)\; \psi^*\left(\frac{u-t}{a}\right) du\,, \qquad (t,a) \in \mathbb{R}\times\mathbb{R}_*^+\,. \tag{7.20}$$

In this expression, ψ is a zero-mean oscillating function, $\int_{-\infty}^{\infty} \psi(t)\, dt = 0$, that is reasonably localized in time and in frequency (within the limits of the Gabor-Heisenberg uncertainty principle). We will demonstrate that the instantaneous power of the output of the filter defined by (7.20), commonly termed *scalogram*, belongs to the affine class of time-frequency representations. To that end, setting $a = 1/f$, let us reformulate the scalar product (7.20) using Moyal's formula (7.19) applied to the Wigner-Ville distributions of x and ψ (or to any other unitary distribution of the affine class):

$$\begin{aligned}\left|\mathrm{CWT}_x^{\psi}\left(t,\frac{1}{f}\right)\right|^2 &= \left|\int_{-\infty}^{\infty} x(u)\, \psi^*_{t,1/f}(u)\, du\right|^2 \\ &= \int_0^{\infty}\int_{-\infty}^{\infty} W_x(u,\nu)\; W_{\psi_{t,1/f}}(u,\nu)\; du\, d\nu\,.\end{aligned}$$

By applying the affine covariance principle (7.9) to $W_{\psi_{t,1/f}}$, we finally obtain

$$\left|\mathrm{CWT}_x^{\psi}\left(t,\frac{1}{f}\right)\right|^2 = \int_0^{\infty}\int_{-\infty}^{\infty} W_x(u,\nu)\; W_{\psi}\left(f\,(u-t), \frac{\nu}{f}\right) du\, d\nu\,, \tag{7.21}$$

wherein we easily recognize a specific case of the parametric form (7.16) with $\phi_{\text{t-f}} = W_{\psi}$.

Finally, we stress that the property of unitarity is interesting from an algebraic point of view, but is also very restrictive in terms of the structure of the kernels. Consequently, it is a property which is often paid for by the exclusion of other theoretical advantages, and often also by a lack of legibility and interpretability of the associated representations.

7.3.4. *Localization*

The study of transients and, more generally, the identification of structures strongly localized in the time-frequency plane has without doubt been a driving element for the development of time-frequency analysis techniques. In this sense, the derivation of representations that respect the temporal (or frequency) localization of a temporal impulse[2] (or of a pure harmonic signal) is a central motivation. Each of these two types of localization corresponds to a specific condition on the parameterization kernel $\widehat{K}$ of (7.10):

– *localization in time*:

$$\widehat{z}_{t_0}(f) = \frac{\mathrm{U}(f)}{\sqrt{f}}\, e^{-j2\pi f t_0} \quad \Rightarrow \quad \Omega_{z_{t_0}}(t,f) = \frac{1}{f}\,\delta(t-t_0)\,, \quad f>0 \tag{7.22}$$

$$\Longleftrightarrow \quad \int_{-\infty}^{\infty} \widehat{K}(f_1, f_1-f_2)\, df_1 = 1\,, \tag{7.23}$$

where $\mathrm{U}(\cdot)$ is the unit step function;

– *localization in frequency*:

$$\widehat{x}_{f_0}(f) = \delta(f-f_0) \quad \Rightarrow \quad \Omega_{x_{f_0}}(t,f) = \delta(f-f_0)\,, \quad f>0 \tag{7.24}$$

$$\Longleftrightarrow \quad \widehat{K}(f,f) = \delta(f-1)\,. \tag{7.25}$$

The Wigner-Ville distribution (7.7) does not comply with the property of temporal localization for the *analytic impulse* as defined in (7.22), but does so for its real counterpart $x_{t_0}(t) = \delta(t-t_0)$. On the other hand, it does satisfy the property of frequency localization (7.24). What is remarkable, however, is that the Wigner-Ville distribution also has a third localization property for another category of signals: *linear chirps*. These are complex signals given by $\widehat{x}(f) = \exp\{j\,\psi_x(f)\}$, $f \in \mathbb{R}$ with linear group delay $t_x(f) := -\frac{1}{2\pi}\frac{d\psi_x(f)}{df} = t_0 + \alpha f$ and for which, in fact, $W_x(t,f) = \delta\big(t-(t_0+\alpha f)\big)$ (see Chapters 1 and 5).

Inspired by this very particular case of the Wigner-Ville distribution, we may then wonder if in the affine class, this principle of triple localization can be extended to more general expressions of the group delay $t_x(f)$. A complete answer to this question is given in [BER 92b], by a rigorous theoretical derivation where the algebra used is that of the affine group restricted to the Hardy space of analytic signals: $\{x \in L^2(\mathbb{R}) : \widehat{x}(f) \equiv 0,\ \forall\, f<0\}$. *Strictly speaking*, this analysis framework excludes the Wigner-Ville distribution from the set of possible solutions, because neither the linear chirp nor the Dirac impulse is an analytic signal.

The challenge raised in [BER 92b] is thus to characterize the pair formed by the *phase spectrum* $\psi_x(f)$ and the *distribution* $\Omega_x(t,f)$, such that

$$\widehat{x}(f) = \frac{\mathrm{U}(f)}{\sqrt{f}}\,\exp\{j\psi_x(f)\} \quad \Longrightarrow \quad \Omega_x(t,f) = \frac{1}{f}\,\delta\big(t-t_x(f)\big)\,, \quad f>0\,.$$

2. The temporal impulse $x_{t_0}(t) = \delta(t-t_0)$ does not belong to the Hardy space of analytic functions $\{x \in L^2(\mathbb{R}) : \widehat{x}(f) \equiv 0,\ \forall f<0\}$. Here, we consider the analytic extension of x_{t_0} to the complex plane, which is denoted z_{t_0}.

A direct calculation provides a functional equation that binds the structure of kernel $\widehat{K}(f_1, f_2)$ (or in an equivalent way that of $\phi_{\text{f-D}}(\nu, \xi)$) to the expression of the phase spectrum $\psi_x(f)$ according to

$$\left\{ \begin{array}{l} \displaystyle\int_0^\infty \sqrt{f_1(f_1-f_2)}\, \widehat{K}(f_1, f_1-f_2)\, e^{j[\psi_x(ff_1) - \psi_x(f(f_1-f_2))]}\, df_1 \;=\; e^{-jff_2\frac{d\psi_x(f)}{df}}, \\ \displaystyle\int_{\frac{|\xi|}{2}}^\infty \sqrt{\nu^2 - \frac{\xi^2}{4}}\, \phi_{\text{f-D}}(\nu, \xi)\, e^{j[\psi_x(f(\nu-\xi/2)) - \psi_x(f(\nu-\xi/2))]}\, d\nu \;=\; e^{-j\xi f\frac{d\psi_x(f)}{df}}. \end{array} \right. \tag{7.26}$$

Resolving these equations is difficult and does not systematically yield a solution $\widehat{K}$ (or $\phi_{\text{f-D}}$) for an arbitrary choice of the phase spectrum $\psi_x(f)$. The thorough analytical study of this problem, performed in [BER 92b], shows in fact that only a certain parameterized family of phase spectra $\{\psi_k\}_{k\in\mathbb{Z}}$ leads to a subclass of localized solutions $\Omega_x \in \{P_x^{(k)}\}_{k\in\mathbb{Z}}$. Before broadly reproducing the derivation of this result, let us specify a little more the context in which it is placed.

Note 3: The localization of $\Omega_x(t, f)$ on a given group delay $t_x(f)$ can be seen as the by-product of coupling a localization property (7.22) or (7.24) with the property of covariance to an additional displacement operator. However, this third operator, defined by the warping that would transform the impulse $z_{t_0}(t)$ or $\delta(f-f_0)$ into a signal with group delay $t_x(f)$, only has mathematical interest if the resulting extension of the affine group to three parameters (translation in time + scaling + third operator) preserves a Lie structure [BER 92b]. The existence and the identification of localized distributions $P_x^{(k)}$ are, thus, connected to the study of affine three-parameter Lie groups and to their representations in the associated phase space.

Note 4: The set of solutions $\widehat{K}$ of (7.26) constitutes a subclass of functions characterized by a diagonal form of the kernel:

$$\widehat{K}(f_1, f_2) \;=\; \Gamma(f_1)\, \delta\big(f_2 - \Upsilon(f_1)\big)\,, \tag{7.27}$$

where Γ and Υ are real functions that determine the theoretical properties of the corresponding distributions $\Omega_x(t, f)$ and, in particular, those of localization on a given group delay.

Note 5: We may indifferently choose to solve the problem of localization on nonlinear group delays $t_x(f)$ by solving the functional equation (7.26) for $\phi_{\text{f-D}}(\nu, \xi)$ and, thus, follow the approach of [RIO 92]. The obtained solutions are then all expressed in the form

$$\phi_{\text{f-D}}(\nu, \xi) \;=\; G(\xi)\, \delta\big(\nu - H(\xi)\big)\,, \tag{7.28}$$

an expression which, by its diagonal structure, is completely comparable with (7.27). However, analytical expressions of the real functions G and H are only known for a restricted number of choices of ψ_k corresponding to the cases $k = -1$, 0, $1/2$, and 2 [RIO 92, FLA 96]. We will consider these particular cases at a later point.

7.4. Affine Wigner distributions

Independently of the property of localization on nonlinear group delays, we may be interested in the subclass of affine distributions generated by the diagonal form of kernels (7.27) or (7.28).

7.4.1. *Diagonal form of kernels*

This section only considers parameterization kernels of the affine class that can be written in a diagonal form. The arbitrariness of this choice may be surprising, but according to the logic guiding the original procedure in [BER 92b], which we do not reproduce in full detail here, diagonal structures are the unique non-trivial solutions guaranteeing an algebraic topology for the extension of the affine group to three parameters.

7.4.1.1. *Diagonal form of the kernel* $\widehat{K}$

Let us then consider kernels $\widehat{K}$ of diagonal expression (7.27), which we develop into the equivalent integral form

$$\widehat{K}(f_1, f_2) = \int_{-\infty}^{\infty} \mu(u)\, \delta\big(f_1 - \lambda(u)\big)\, \delta\big(f_2 - \lambda(-u)\big)\, du\,.$$

The function λ is continuous and bijective from $\mathbb{R}$ to $\mathbb{R}_*^+$, and the function μ ensures the real and symmetric nature of the kernel $\widehat{K}$. Using this kernel in (7.10) and proceeding by successive integrations with respect to f_1 and then f_2, we define the class of *affine Wigner (time-frequency) distributions*, whose canonical expression is

$$P_x(t, f) = f \int_{-\infty}^{\infty} \mu(u)\, \widehat{x}\big(f\lambda(u)\big)\, \widehat{x}^*\big(f\lambda(-u)\big)\, e^{j2\pi f t\,[\lambda(u) - \lambda(-u)]}\, du\,, \quad f > 0\,. \tag{7.29}$$

7.4.1.2. *Properties of diagonal kernel distributions*

It is possible to specify a little further the constraints on the diagonal kernel $\widehat{K}$ so that the associated distributions (7.29) satisfy the properties listed in Section 7.3. Thus:

– the frequency marginal property given by the equation $\int_{-\infty}^{\infty} P_x(t, f)\, dt = |\widehat{x}(f)|^2$ implies the two following conditions simultaneously:

$$\frac{\mu(0)}{2\lambda'(0)} = 1\,, \qquad \lambda(0) = 1\,; \tag{7.30}$$

– the unitarity property (7.19) does not concern the choice of function λ, but restricts the range of possible functions μ to the single quantity

$$\mu(u) = \sqrt{\big|\lambda'(u) + \lambda'(-u)\big|}\, \sqrt{\big|\lambda'(-u)\, \lambda(u) + \lambda'(u)\, \lambda(-u)\big|}\,; \tag{7.31}$$

– the condition for the frequency localization property (7.24) coincides with the pair of conditions (7.30) ensuring the frequency marginalization property;

– the time localization property (7.22) corresponds to the following pair of conditions:

$$\begin{aligned} &\text{– the function } \zeta(u) := \lambda(u) - \lambda(-u) \text{ must be bijective from } \mathbb{R} \text{ to } \mathbb{R}, \\ &– \mu(u) = \sqrt{\lambda(u)\,\lambda(-u)}\,\big|\lambda'(u) + \lambda'(-u)\big|. \end{aligned} \tag{7.32}$$

In a second step, we may look if there exist distributions of the form (7.29) that simultaneously meet several of these properties, by solving the system formed by the corresponding constraints. For example, affine distributions that are unitary and simultaneously time-localized should correspond to parameterization functions $\lambda(\cdot)$ that are solutions to the nonlinear differential equation obtained by merging the two constraints (7.31) and (7.32):

$$\frac{d}{du}\big(\lambda(u) - \lambda(-u)\big) = \frac{d}{du}\ln\!\left(\frac{\lambda(u)}{\lambda(-u)}\right).$$

Remarkably, there is a unique solution $\lambda(\cdot)$; we obtain [BER 92b]

$$\left\{\begin{aligned} \lambda(u) &= \frac{u\,e^{u/2}}{2\,\sinh(u/2)} \\ \mu(u) &= \sqrt{\lambda(u)\,\lambda(-u)}\,. \end{aligned}\right. \tag{7.33}$$

Moreover, the function $\zeta(u) := \lambda(u) - \lambda(-u)$ is bijective from $\mathbb{R}$ to $\mathbb{R}$. The affine Wigner distribution resulting from the specific parameterization functions (7.33) is commonly called the P_0 *distribution* or *unitary Bertrand distribution*; its expression reads

$$P_x^{(0)}(t,f) = f\int_{-\infty}^{\infty} \frac{u}{2\,\sinh(u/2)}\,\widehat{x}\!\left(\frac{fu\;e^{u/2}}{2\,\sinh(u/2)}\right)\widehat{x}^*\!\left(\frac{fu\;e^{-u/2}}{2\,\sinh(u/2)}\right) e^{j2\pi ftu}\,du\,,\quad f>0\,. \tag{7.34}$$

Within the class of affine Wigner distributions, the unitary Bertrand distribution plays a role comparable to that played by the Wigner-Ville distribution within Cohen's class.

7.4.1.3. *Diagonal forms of the kernel* $\phi_{\text{f-D}}$

Based on the alternative canonical form (7.12), it is naturally possible to find a counterpart formulation for the affine Wigner distributions with diagonal kernel $\widehat{K}$. Similarly to our development above, we need to constrain the corresponding version $\phi_{\text{f-D}}(\nu,\xi)$ of the parameterization kernel to a diagonal structure whose support is localized on a curve of the type $\nu = H(\xi)$. The resulting *localized bi-frequency* kernel, and its dual form $\phi_{\text{d-D}}(\tau,\xi)$, read[3] [RIO 92]

$$\phi_{\text{f-D}}(\nu,\xi) \;=\; G(\xi)\,\delta\big(\nu - H(\xi)\big) \qquad \Longleftrightarrow \qquad \phi_{\text{d-D}}(\tau,\xi) \;=\; G(\xi)\, e^{-j2\pi H(\xi)\tau},$$

where $G(\cdot)$ and $H(\cdot)$ are two real functions. This other diagonal kernel expression thus allows a reformulation of the subclass of *affine Wigner (time-frequency) distributions* (7.29) according to the companion parametric definition

$$P_x(t,f) \;=\; f\int_{-\infty}^{\infty} G(\xi)\,\widehat{x}\bigg(f\bigg(H(\xi)+\frac{\xi}{2}\bigg)\bigg)\,\widehat{x}^*\bigg(f\bigg(H(\xi)-\frac{\xi}{2}\bigg)\bigg)\, e^{j2\pi\xi f t}\, d\xi\,, \quad f>0\,. \tag{7.35}$$

As in the quadratic form (7.29), the product $\widehat{x}\big(f\big(H(\xi)-\frac{\xi}{2}\big)\big)\,\widehat{x}^*\big(f\big(H(\xi)+\frac{\xi}{2}\big)\big)$ only induces interaction between components of $\widehat{x}$ located on the same positive axis $\mathbb{R}_*^+$. Whereas this constraint resulted in $\lambda(u)>0$, $\forall u\in\mathbb{R}$ in definition (7.29), it now implies the condition $H(\xi)\geq\left|\frac{\xi}{2}\right|$ in the alternative definition (7.35).

7.4.1.4. *Properties of distributions with diagonal kernel* $\phi_{\text{f-D}}$

We have seen above that imposing theoretical properties on the affine Wigner distributions amounts to structurally constraining the functions $\lambda(\cdot)$ and $\mu(\cdot)$ in (7.29). These constraints can also be expressed in terms of the parameterization functions $H(\cdot)$ and $G(\cdot)$ in (7.35). For example, the constraints for the particularly interesting properties of unitarity (7.19) and temporal localization (7.22) are given by the functional relations

$$G^2(\xi) \;=\; H(\xi) - \xi\,\frac{dH(\xi)}{d\xi} \tag{7.36}$$

and

$$G^2(\xi) \;=\; H^2(\xi) - \left(\frac{\xi}{2}\right)^2, \tag{7.37}$$

respectively. Thus, a distribution (7.35) satisfying the unitarity and temporal localization properties is defined by the pair (H,G) that solves the differential equation obtained by identifying expressions (7.36) and (7.37). Such a solution exists; it is given by

$$H(\xi) \;=\; \frac{\xi}{2}\coth\left(\frac{\xi}{2}\right) \qquad \text{and} \qquad G(\xi) \;=\; \frac{\xi/2}{\sinh(\xi/2)}\,.$$

Inserting this solution into the canonical form (7.35), we reobtain the unitary Bertrand distribution $P_x^{(0)}(t,f)$ already encountered in (7.34).

3. It is formally possible to write down the $\phi_{\text{t-f}}$ and the $\phi_{\text{t-d}}$ versions corresponding to the localized bi-frequency kernels, but the resulting Fourier-integral forms do not reduce to simple closed-form expressions.

7.4.2. *Covariance to the three-parameter affine group*

We may further reduce the class of affine distributions provided in (7.10) by extending the principle of affine covariance to an extension of the affine group A to three parameters [BER 92a, BER 92b][4]. These groups, denoted by G_k, are indexed by a real parameter k and characterized by the multiplication law acting on the triplets $g = (a, b, c)$ and $g' = (a', b', c')$, elements of the group, as

$$g\,g' = \begin{cases} (aa', b + ab', c + a^k c') & \text{for } G_k,\ k \neq 0, 1 \\ (aa', b + ab', c + c') & \text{for } G_0 \\ (aa', b + ab' + a \ln a c', c + ac') & \text{for } G_1\,. \end{cases} \tag{7.38}$$

By construction, the distributions (7.10) already satisfy the covariance property with respect to the affine group A of two parameters (a, b). We therefore have to identify among the distributions (7.10) the subclass of distributions which, in addition, are covariant to the transformation related to the third parameter c. The group G_k with $g = (1, 0, c)$ acts on the analytic signals $\widehat{x}(f)$ according to

$$\widehat{x}^g(f) = \begin{cases} e^{-j2\pi c f^k}\, \widehat{x}(f) & \text{for } k \neq 0, 1 \\ e^{-j2\pi c \ln f}\, \widehat{x}(f) & \text{for } k = 0 \\ e^{-j2\pi c f \ln f}\, \widehat{x}(f) & \text{for } k = 1\,, \end{cases} \tag{7.39}$$

and on the sought covariant affine time-frequency representations (7.10) according to

$$\Omega_{x^g}(t, f) = \begin{cases} \Omega_x\big(t - c\,k f^{k-1}, f\big) & \text{for } k \neq 0, 1 \\ \Omega_x\big(t - \frac{c}{f}, f\big) & \text{for } k = 0 \\ \Omega_x\big(t - c\,(1 + \ln f), f\big) & \text{for } k = 1\,. \end{cases} \tag{7.40}$$

7.4.2.1. *Construction of P_k distributions*

Based on the principle of covariance with respect to this extended three-parameter affine group G_k, we construct a subclass of affine distributions, which, quite remarkably, corresponds to a specific parametric form of the distributions (7.29) with diagonal kernel $\widehat{K}$:

$$P_x^{(k)}(t, f) = f \int_{-\infty}^{\infty} \mu(u)\, \widehat{x}\big(f \lambda_k(u)\big)\, \widehat{x}^*\big(f \lambda_k(-u)\big) \cdot e^{j2\pi f t\,[\lambda_k(u) - \lambda_k(-u)]}\, du\,, \quad f > 0\,. \tag{7.41}$$

---

4. To preserve the properties associated with Lie groups and algebras, the extension of the affine group to three parameters is not arbitrary. These are all the possible extensions (and unitary equivalences) in this specific topological context that are studied in [BER 92a, BER 92b], and there are no others.

These particular affine Wigner distributions, usually called P_k *distributions* or *Bertrand distributions*, are characterized by a family $(\lambda_k)_{k\in\mathbb{R}}$ of parameterizing functions whose expression is

$$\lambda_k(u) = \begin{cases} \left[k\frac{e^{-u}-1}{e^{-ku}-1}\right]^{\frac{1}{k-1}} & \text{for } k \neq 0,1 \\ \frac{u\,e^{u/2}}{2\sinh(u/2)} & \text{for } k = 0 \\ \exp\left(1 - \frac{u}{e^u-1}\right) & \text{for } k = 1, \end{cases}$$

and which satisfies the following properties:

$$\lambda_k(u) = e^u\,\lambda_k(-u), \quad \lambda_k(0) = 1, \quad \text{and } \lambda_k'(0) = \frac{1}{2}, \quad \forall\, k\in\mathbb{R}.$$

As for the function μ, it remains arbitrary but must be real and positive. If we require the corresponding distribution to be covariant to temporal inversion, μ must also be an even function.

7.4.2.2. *Properties of P_k distributions*

Since P_k distributions are particular cases of affine distributions with diagonal kernel, it suffices to adapt the various constraints established in Section 7.4.1.2 to the specific expressions of the λ_k functions. We leave this exercise to the reader, and instead elaborate on certain remarkable values of the index k. Thus:

– for $k = 2$, we recognize the Wigner-Ville distribution but restricted to analytic signals:

$$W_x(t,f) = \int_{-2f}^{2f} \widehat{x}\left(f+\frac{\xi}{2}\right)\,\widehat{x}^*\left(f-\frac{\xi}{2}\right)\,e^{j2\pi t\xi}\,d\xi, \quad f > 0.$$

As already mentioned, the Wigner-Ville distribution belongs simultaneously to Cohen's class and to the affine class, and thus it satisfies a principle of affine covariance extended to three parameters: scale change, translation in time, and translation in frequency;

– the case $k = 0$ corresponds to the unitary Bertrand distribution $P_x^{(0)}(t,f)$ already seen in (7.34). In addition to the fact that this is the only distribution (7.29) both unitary and localized in time, $P_x^{(0)}$ is covariant to affine changes and to dispersion along hyperbolic group delays (see operators (7.39) and (7.40) for $k = 0$);

– the case $k = -1$, associated with the choice of a function μ satisfying (7.32), corresponds to the active form of the Unterberger distribution [UNT 84]

$$U_x(t,f) = f\int_{-\infty}^{\infty}\cosh\left(\frac{u}{2}\right)\,\widehat{x}\left(f\,e^{u/2}\right)\,\widehat{x}^*\left(f\,e^{-u/2}\right)\,e^{j4\pi ft\,\sinh(u/2)}\,du, \quad f > 0.$$

Although it was originally suggested in a very different context, the Unterberger distribution thus proves to be covariant to the three-parameter affine group G_{-1}. Remarkably, it also has an interesting property of localization in the time-frequency plane on group delay trajectories $t_x(f) = t_0 - c/f^2$;

– more generally, for any $k \leq 0$, the corresponding P_k distributions with μ fixed by the condition of temporal localization (7.32) perfectly localize on group delay trajectories described by power-law functions:

$$\widehat{x}(f) = \frac{\mathrm{U}(f)}{\sqrt{f}}\, e^{-j2\pi(t_0 f + c f^k)} \quad \overset{k<0}{\Longrightarrow} \quad P_x^{(k)}(t,f) = \frac{1}{f}\,\delta\big(t - (t_0 + c\,k f^{k-1})\big)\,,$$

$$\widehat{x}(f) = \frac{\mathrm{U}(f)}{\sqrt{f}}\, e^{-j2\pi(t_0 f + c\ln f)} \quad \overset{k=0}{\Longrightarrow} \quad P_x^{(0)}(t,f) = \frac{1}{f}\,\delta\left(t - \left(t_0 + \frac{c}{f}\right)\right). \tag{7.42}$$

7.4.3. *Smoothed affine pseudo-Wigner distributions*

7.4.3.1. *Limits of affine Wigner distributions*

The affine time-frequency representations defined by the quadratic form (7.41) offer a very complete range of interesting theoretical properties. For example, with respect to joint resolution in the time-frequency plane, they attain degrees of energy concentration (notably on their matched group delays (7.42)) which largely exceed the potential offered by the scalogram (square modulus of the continuous wavelet transform (7.20)). Exploited in a *time-frequency based detection scheme*, this concentration strength leads to simple and very flexible detector structures, whose performance is comparable to that of adaptive filtering [GON 97].

This localization capacity notwithstanding, a certain number of theoretical and numerical difficulties have constituted a barrier to the natural expansion and diffusion of these tools. From a theoretical point of view first, like all quadratic representations, the bilinear form (7.41) does not abide by the principle of superposition. More precisely, the affine Wigner distribution $P_x^{(k)}$ of a sum of signals reveals, in addition to auto-terms, some cross terms resulting from the pairwise interaction of the signal components:

$$x(t) = \sum_{i=1}^{N} c_i\, x_i(t) \quad \Longrightarrow$$

$$P_x^{(k)}(t,f) = \underbrace{\sum_{i=1}^{N} |c_i|^2\, P_{x_i}^{(k)}(t,f)}_{\text{auto-terms}} + \underbrace{2\,\mathrm{Re}\left\{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} c_i\, c_j^*\, P_{x_i,x_j}^{(k)}(t,f)\right\}}_{\text{cross terms}}.$$

The inevitable presence of these interference terms, although essential for certain theoretical properties to hold, very often proves critical when we have to interpret and analyze non-trivial signals (noisy observations, multi-component signals, etc.). In [FLA 96], a detailed study of interference terms has enabled the identification of the geometrical principles underlying their construction, as well as a detailed characterization of their oscillating structure. In particular, it was demonstrated that the geometry underlying affine Wigner distributions is entirely ruled by generalized means or Stolarsky-type symmetries, and that group delays of the form (7.42), $t_x^{(k)}(f) = t_0 + cf^{k-1}$, are precisely the locus of the points globally invariant under these symmetries.

Turning now to the numerical implementation of the forms (7.41), we face two additional difficulties. First of all, the integral defining $P_x^{(k)}(t,f)$ at each moment t (or at each frequency f) corresponds to a global measure involving the entire signal $\widehat{x}(f)$ (or $x(t)$). As the limited memory of computing devices may not allow the processing of "long signals", a *short-time* approximation of (7.41) is thus necessary if we want to remove the "short signal" limitation.

The other numerical difficulty, which is less penalizing however, is related to the analytical reversibility of the function $\xi_k(u) := \lambda_k(u) - \lambda_k(-u)$. In fact, to reduce computational costs, we interpret the oscillating integral (7.41) defining $P_x^{(k)}$ as a simple Fourier transform between the dual variables $\alpha = t \cdot f$ and $v = \xi_k(u)$. Then, to benefit from the performance of discrete Fourier transforms (e.g., FFT algorithm), it is necessary for the reciprocal function $\xi_k^{-1}(\cdot)$ to admit a closed-form expression[5], which exists only for $k \in \{-1, 0, 1/2, 2, \pm\infty\}$ [BER 92b, FLA 96].

7.4.3.2. *Definition*

The approach proposed in [GON 96, GON 98] defining the *smoothed affine pseudo-Wigner distributions* largely benefits from the analogy existing between Cohen's class (see Chapter 5) and the affine class. In particular, the principle of short-term windowing on which this approach is based is an affine adaptation of the procedure leading to the *smoothed pseudo-Wigner-Ville distributions* of Cohen's class [CLA 80, FLA 84, STA 94].

Thus, let us rewrite expression (7.41) providing the affine Wigner distribution such that the temporal expression of the analyzed signal is made explicit:

$$P_x^{(k)}(t,f) = f \int_{-\infty}^{\infty} \mu(u) \left[\int_{-\infty}^{\infty} x(\tau)\, e^{-j2\pi\lambda_k(u) f(\tau - t)}\, d\tau \right] \times \left[\int_{-\infty}^{\infty} x(\tau')\, e^{-j2\pi\lambda_k(-u) f(\tau' - t)}\, d\tau' \right]^* du\,.$$

5. When ξ_k^{-1} does not admit a closed-form expression, we can use a correspondence table, at the cost of additional approximations.

In each of the two Fourier integrals brought into play, we limit the temporal extension of the signal by pre-windowing it with a real short-term function h:

$$\widetilde{P}_x^{(k)}(t,f) := f\int_{-\infty}^{\infty}\mu(u)\left[\int_{-\infty}^{\infty}x(\tau)\;h\big(f\lambda_k(u)(\tau-t)\big)\;e^{-j2\pi\lambda_k(u)f(\tau-t)}\,d\tau\right]$$
$$\times\left[\int_{-\infty}^{\infty}x(\tau')\;h\big(f\lambda_k(-u)(\tau'-t)\big)\;e^{-j2\pi\lambda_k(-u)f(\tau'-t)}\,d\tau'\right]^* du\,.$$

We note that in this expression, the temporal support of the window $h\big(f\lambda_k(\pm u)\,\tau\big)$ is directly related to the frequency of analysis f, which guarantees that the distribution $\widetilde{P}_x^{(k)}$ thus defined is truly covariant to scale changes. Then, introducing the oscillating function $\psi(\tau) := h(\tau)\,e^{j2\pi\tau}$, the *affine pseudo-Wigner distributions* are written more simply as

$$\widetilde{P}_x^{(k)}(t,f) = \int_{-\infty}^{\infty}\frac{\mu(u)}{\sqrt{\lambda_k(u)\,\lambda_k(-u)}}\,\widetilde{\mathrm{CWT}}_x^{\psi}(t,\lambda_k(u)f)$$
$$\times\left[\widetilde{\mathrm{CWT}}_x^{\psi}(t,\lambda_k(-u)f)\right]^* du\,, \qquad (7.43)$$

where $\widetilde{\mathrm{CWT}}_x^{\psi}(t,f) := \mathrm{CWT}_x^{\psi}(t,a)\big|_{a=1/f}$ is the time-frequency version of the wavelet transform defined in (7.20). By performing the preprocessing $x(t) \to \widetilde{\mathrm{CWT}}_x^{\psi}(t,f)$ before calculating the quadratic form, we simultaneously act on the three weaknesses that plague affine Wigner distributions:

– the short time support of the apodization window ψ limits long-range interactions in the signal and thus reduces the interferences between temporally distant components;

– acting as a sliding window, the convolution (7.20) plays in the expression (7.43) of affine pseudo-Wigner distributions the role played by the Fourier transform in the original expression (7.41) of affine Wigner distributions. However, unlike the latter, limiting the integration support amounts to working with a local measure which can now simply be implemented with an *on-line* algorithm, without a restriction to "short" signals;

– the structure of wavelet ψ directly integrates the oscillating exponential characterized by the expression $\exp\big(j2\pi\lambda_k(u)ft\big)$. Thus, the non-existence of a closed-form expression for $\xi_k^{-1}(\cdot)$ is no longer an obstacle to the fast computation of $\widetilde{P}_x^{(k)}$, whatever the value of parameter k.

Note 6: Whereas the scalogram (7.21) is obtained simply by squaring the modulus of a wavelet transform, the affine pseudo-Wigner distributions extend this principle to a generalized affine auto-correlation of the wavelet transform along the frequency axis.

The window h, as a pre-factor of the analyzed signal x, suppresses the interference terms oscillating in the frequency direction by means of a smoothing effect (with constant Q-factor[6]). The interference terms resulting from the interaction between components that are disjoint in frequency oscillate in the direction of the time axis and are thus not affected by the wavelet transform. To attenuate these components, it is necessary to introduce a window that will perform a smoothing in the time direction. The product variable $\alpha = t \cdot f$ is indirectly dual[7] to the variable u in (7.41). Therefore, limiting the integration support of u in (7.43) amounts to applying an *affine* time smoothing to $\widetilde{P}_x^{(k)}$. The *smoothed affine pseudo-Wigner distributions* are thus defined by the following generalized affine convolution:

$$\widetilde{P}_x^{(k)}(t,f) = \int_{-\infty}^{\infty} G(u) \frac{\mu(u)}{\sqrt{\lambda_k(u)\,\lambda_k(-u)}} \widetilde{\mathrm{CWT}}_x^{\psi}(t, \lambda_k(u)f) \times \left[\widetilde{\mathrm{CWT}}_x^{\psi}(t, \lambda_k(-u)f)\right]^* du\,, \qquad (7.44)$$

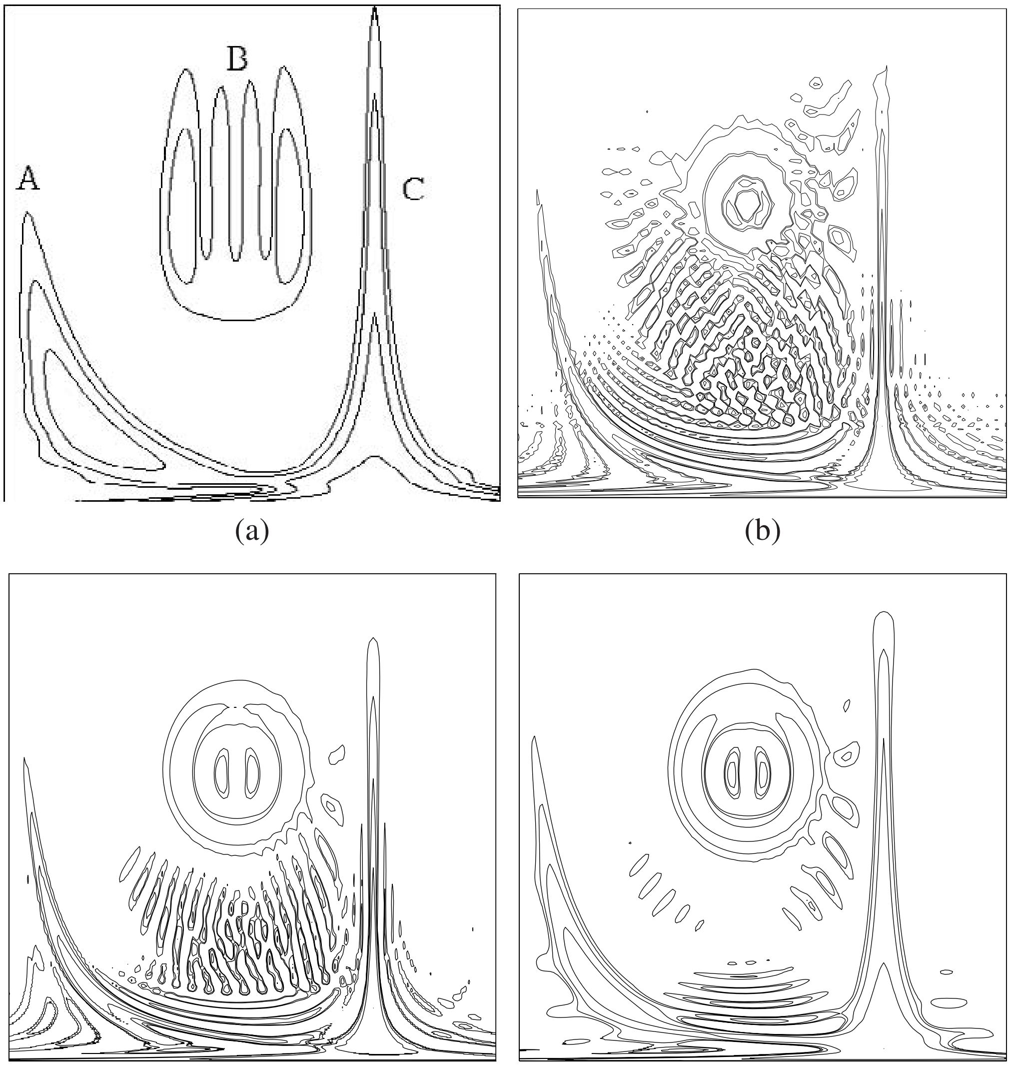
where $G(u)$ denotes a real window. Thanks to this separation between the smoothing induced by G and the smoothing induced by ψ, we can independently control the time and frequency resolutions of $\widetilde{P}^{(k)}$, which is obviously impossible with the scalogram. However, this advantage is not new. In fact, in [RIO 92], the authors had first introduced an alternative to the scalogram, the *affine smoothed pseudo-Wigner-Ville distributions with separable kernels*, which already allowed a flexible, independent choice of the time and frequency resolutions. The representations defined in (7.44) generalize this principle to all values of parameter k, as the case treated in [RIO 92] equals the specific smoothed affine pseudo-Wigner distribution $\widetilde{P}^{(k)}$ for which $k = 2$.

Finally, we stress that the suppression of interferences comes at a cost. In fact, by suppressing the oscillating terms (interferences) of bilinear distributions, *strictly speaking* we lose a certain number of theoretical properties which, paradoxically, constitute the principal attraction of the affine Wigner distributions (7.41). First of all, strict localization on power-law trajectories in the time-frequency plane, as expressed by equation (7.42), cannot be maintained. However, the affine pseudo-Wigner distributions (7.43) converge exactly towards the affine Wigner distributions (7.41) when the Q-factor of wavelet ψ tends towards infinity. Thus, we can approach this concentration property, like all others, with arbitrarily large degrees of precision.

Figure 7.1 presents a toy example illustrating the differences between an affine Wigner distribution $P_x^{(k)}$, the corresponding affine pseudo-Wigner and smoothed affine pseudo-Wigner distributions $\widetilde{P}_x^{(k)}$, and the scalogram.

6. For a wavelet ψ, we define the quality factor or briefly Q-factor as the mean frequency of its spectrum $\left|\widehat{\psi}(f)\right|^2$, normalized by its equivalent bandwidth.

7. To be more precise, the variable $\alpha = t \cdot f$ is the Fourier dual of the variable $v = \xi_k(u)$. However, since the function $\xi_k(\cdot)$ is strictly monotonous, limiting the extension of the variable u is equivalent to limiting the extension of $\xi_k(u)$.



(a) (b)

(c) (d)

Figure 7.1. *Affine time-frequency representations of a synthetic signal consisting of a hyperbolic modulation* $\widehat{x}_{\mathrm{A}}(f) = e^{j2\pi\alpha \ln f}$ *(component marked* A*), a* 3*rd degree Hermite function (component* B*), and a Lipschitz singularity* $x_{\mathrm{C}}(t) = |t - t_0|^{-0.1}$ *(component* C*). The representations are displayed by iso-contour lines; the abscissa axis corresponds to time and the ordinate axis to frequency. (a) Scalogram* $\left|\widetilde{\mathrm{CWT}}_x^{\psi}(t,f)\right|^2$ *using a Morlet wavelet* ψ *of quality factor* $Q = 2$*; (b) unitary Bertrand distribution* $P_x^{(0)}(t,f)$*; (c) affine pseudo-Wigner distribution (7.43),* $\widetilde{P}_x^{(0)}(t,f)$*, using a Morlet wavelet with* $Q = 8$*; (d) smoothed affine pseudo-Wigner distribution (7.44) using the same Morlet wavelet (*$Q = 8$*) and a Gaussian window* G*. The smoothed affine pseudo-Wigner distributions (7.44) enable a continuous transition between the scalogram (with poor resolution but reduced interferences) and the affine Wigner distributions (7.41) (with high resolution but large interferences)*

For practical aspects *of numerical implementation* and more theoretical considerations such as *distributions generating the affine class*, see [GON 96, GON 98], where a detailed study of these subjects is provided.

7.5. Advanced considerations

7.5.1. *Principle of tomography*

Among all desirable theoretical properties, positivity of time-frequency distributions is certainly the most exclusive one. For this reason, the identification, from a theoretical point of view, of interesting distributions with true energy densities is very often contested and lies at the origin of controversial debates. Nevertheless, when a distribution has positive marginal distributions in time and frequency, it is possible to demonstrate that it also guarantees positive marginals by integration along other time-frequency paths. To that end, it suffices to transform the time and frequency axes by means of the displacement operators corresponding to the group of translations considered: time shift and frequency shift of the Weyl-Heisenberg group for Cohen's class; time shift and scale change of the affine group for the affine class.

7.5.1.1. *Wigner-Ville tomography*

The Wigner-Ville distribution is known to positively integrate along any straight line of the time-frequency plane. This can be easily demonstrated using the property of unitarity:

$$\left|\int_{-\infty}^{\infty} x(t)\, y^*(t)\, dt\right|^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_x(t,f)\, W_y(t,f)\, dt\, df\ ,$$

and by setting $y(t) = \exp\big(j\pi(at^2+2bt)\big)$, which is a linearly frequency-modulated signal (or *linear chirp*). Since $y(t)$ is represented in phase space by a delta function on the straight line $f = at + b$, it immediately follows that

$$\begin{aligned}\left|\int_{-\infty}^{\infty} x(t)\, e^{-j\pi(at^2+2bt)}\, dt\right|^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_x(t,f)\ \delta\big(f-(at+b)\big)\ dt\ df \\ &= \int_{-\infty}^{\infty} W_x(t, at+b)\ dt\ \geq 0\,.\end{aligned}$$

The integral of a two-dimensional function over the set of all straight lines is known as the Radon transform [LUD 66]. More specifically, applied to the Wigner-Ville distribution, it also corresponds to the inner product of x with the set of linear chirps, an operation that is also termed *angular Fourier transform*. The numerical implementation of the Radon transform resorts to the Hough transform, a transformation that is commonly used in image analysis to detect the presence of edges in an image.

7.5.1.2. *Tomography of the affine Wigner distribution*

Using the same arguments as previously, we can show that the unitary affine Wigner distribution $P_x^{(0)}(t, f)$ defined by (7.34) yields a positive marginal density when integrated over a set of hyperbolae in the time-frequency plane [BER 85]. Again, to demonstrate this, we use the unitarity property of $P_x^{(0)}(t, f)$:

$$\left|\int_0^{\infty} \widehat{x}(f)\, \widehat{y}^*(f)\, df\right|^2 = \int_0^{\infty}\int_{-\infty}^{\infty} P_x^{(0)}(t, f)\, P_y^{(0)}(t, f)\; dt\, df\,,$$

and set $\widehat{y}(f) = \frac{\mathrm{U}(f)}{\sqrt{f}}\, e^{-j2\pi(\tau f\, +\, \beta\, \ln f)}$, which is a family of signals represented by delta functions on hyperbolae in the phase space. Recalling the localization property (7.42), the above equation becomes

$$\begin{aligned}\left|\int_0^{\infty} \widehat{x}(f)\, \frac{1}{\sqrt{f}}\; e^{j2\pi(\tau f\, +\, \beta\, \ln f)}\, df\right|^2 &= \int_0^{\infty}\int_{-\infty}^{\infty} P_x^{(0)}(t, f)\; \delta\Big(t - \Big(\tau + \frac{\beta}{f}\Big)\Big)\, dt\, df \\ &= \int_0^{\infty} P_x^{(0)}\Big(\tau + \frac{\beta}{f}, f\Big)\, df\,.\end{aligned}$$

In this path integral of the unitary Bertrand distribution, we recognize a *Radon transform* with respect to the set of time-frequency hyperbolae $t = \tau + \beta/f$. As for the left hand side of the above equation, it coincides with the squared modulus of the Mellin transform of $\widehat{x}$. This linear transformation, which will be denoted as $M_x^{\tau}(\beta)$, has a scaling-invariant magnitude:

$$\begin{aligned}\widehat{x}(f) \quad &\Longrightarrow \quad M_x^{\tau}(\beta) \\ a^{1/2}\, e^{-j2\pi\tau(1-a)f}\; \widehat{x}(af) \quad &\Longrightarrow \quad a^{-j2\pi\beta}\, M_x^{\tau}(\beta)\,,\end{aligned}$$

and it constitutes the pivot of efficient numerical algorithms for calculating affine Wigner distributions [OVA 92a, OVA 92b, BER 90, BER 95].

Similarly, for any negative value of parameter k, it can be shown that there exists a unitary affine Wigner distribution $P_x^{(k)}$ whose integral over paths defined by a power law of degree $k-1$ yields a positive marginal distribution. Then, with the appropriate parameterization function $\mu(u) = \sqrt{\lambda_k(u)\, \lambda_k(-u)}$ and $k \leq 0$, the associated distribution $P_x^{(k)}$ satisfies

$$\int_0^{\infty} P_x^{(k)}\big(\tau + k\, \beta f^{k-1}, f\big)\, df \;=\; \left|\int_0^{\infty} \widehat{x}(f)\; e^{j2\pi(\tau f\, +\, \beta f^k)}\, \frac{df}{\sqrt{f}}\right|^2 \geq 0\,, \quad \forall\, \beta, \tau\,.$$

7.5.2. *Operators and groups*

Despite the broad applicability of time-frequency and time-scale representations, there exist situations where a strict time-frequency or time-scale analysis is not the most appropriate type of analysis. These problems require joint representations in terms of other variables, in order to redeploy the signal content in a more advantageous descriptive plane.

There are two main approaches to constructing joint distributions of arbitrary variables: the method of *covariance*, which leads to joint distributions adapted to a specific type of group transformations, and the method of *marginals*, which, by analogy with probability densities, provides representations whose marginal distributions (obtained by integrating along one variable) coincide with specific signal energy distributions.

In this section, we briefly recall the principles of these two constructions. We also describe a simple method, based on the property of unitary equivalence, which makes it possible to derive novel distributions from more conventional ones.

7.5.2.1. *Method of covariance*

A representation ρ_x is *covariant* to a signal transformation $\boldsymbol{T}$ if $\rho_{\boldsymbol{T}x}$ entirely derives from ρ_x by means of a coordinate warping – a property that is algebraically expressed by the equality (7.5). Let us illustrate this principle with Cohen's class and the affine class of time-frequency representations, and use these to introduce some specific notations.

The *Weyl-Heisenberg group* is an algebraic group of transformations whose unitary representation on the space of finite-energy signals is the time-frequency shift operator, $x(t) \to x(t-t_0)\, e^{j2\pi f_0 t}$. Concomitantly, its action on the time-frequency plane (considered as an orbit of it with respect to its coadjoint representation) is given by the coordinate translation $(t, f) \to (t-t_0, f-f_0)$. *Cohen's class* is the set of phase-space representations that ensure a covariant mapping between this particular displacement operator and this particular phase-space coordinate warping (see Section 7.2.2.1 and Chapter 5).

The same reasoning holds between the *affine group* and the *affine class* of time-frequency representations. The affine group, whose unitary representation in the space $L^2(\mathbb{R}_*^+)$ reads $\widehat{x}(f) \to a\, e^{-i2\pi f t_0}\, \widehat{x}(af)$, acts on the time-frequency half-plane (its physical phase space) by the natural representation $(t, f) \to \left(\frac{t-t_0}{a}, af\right)$. Imposing a covariant mapping between these two representations leads to the affine class of time-frequency representations (see Section 7.2.2.2).

As we can see, the generalization of this construction principle requires the prior choice of a pair of displacement operators in a given signal space. It is then a matter of imposing the covariance equation (7.5) on energy distributions expressed generically as (7.2) or (7.3), in order to obtain a covariant mapping between these operators and the corresponding phase-space coordinate warping [HLA 94, BAR 95a, BAR 96a, SAY 96b, SAY 96a, HLA 03b, HLA 03a].

This method is not very constrained; it allows an infinity of possible choices for the displacement operators, leading to as many classes of covariant distributions (as illustration, two specific examples of classes will be considered in Section 7.5.2.3).

However, it should be noted that all these pairs of operators, without exception, are reducible through unitary equivalences to either one of the two fundamental pairs of displacement operators related to the Weyl-Heisenberg group and to the affine group, respectively [BAR 96b].

7.5.2.2. *Method of marginals*

The principle of tomography discussed in Section 7.5.1 makes it possible to obtain signal energy densities by path-integration of certain time-frequency or time-scale representations. We can then reverse the perspective and pose this marginal property as the generating principle underlying the definition of new classes of signal representations. Without going into detail about their construction, which is clearly explained in [COH 95], we just state that representations $\rho_x(a,b)$ obtained by application of this principle have positive marginal distributions:

$$\int_{-\infty}^{\infty} \rho_x(a,b)\, db \;=\; \left|F_x^{(1)}(a)\right|^2, \qquad \int_{-\infty}^{\infty} \rho_x(a,b)\, da \;=\; \left|F_x^{(2)}(b)\right|^2.$$

Here, $F_x^{(1)}(a)$ and $F_x^{(2)}(b)$ are the so-called group Fourier transforms of x in the description variables a and b, respectively.

For example, certain Cohen's class representations $C_x(t,f)$ have the usual Fourier spectrum as frequency marginal:

$$\int_{-\infty}^{\infty} C_x(t,f)\, dt \;=\; |\widehat{x}(f)|^2 \;=\; \left|\int_{-\infty}^{\infty} x(u)\, e^{-j2\pi uf}\, du\right|^2.$$

Furthermore, there exist "time-Mellin" representations $H_x(t,\beta)$ having a marginal in the Mellin variable $\beta = (t-t_0)f$ that coincides with the Fourier-Mellin spectrum:

$$\int_{-\infty}^{\infty} H_x(t,\beta)\, dt \;=\; \left|M_x^{t_0}(\beta)\right|^2 \;=\; \left|\int_{0}^{\infty} \widehat{x}(f)\, \frac{1}{\sqrt{f}}\, e^{j2\pi(t_0 f + \beta \ln f)}\, df\right|^2.$$

Unlike the covariance principle, the marginal approach allows the construction of joint signal representations in a large variety of cases [SCU 87, COH 95, COH 96, BAR 98a, BAR 98b, BAR 96a, HLA 97b]. Technically, it suffices that the operators associated with the chosen pair of description variables satisfy a non-commutativity relation [TWA 99]. It is, however, not surprising that except for some specific cases, which include the Wigner-Ville distribution and the unitary Bertrand distribution, most of the representations constructed according to this principle do not have interesting properties from the covariance point of view. On the other hand, there are rare cases, and once again the Wigner-Ville and unitary Bertrand distributions are examples, where the two construction methods yield the same result. The theoretical framework that allows us to formalize this intersection relies on the concept of *unitary equivalence*.

7.5.2.3. *Unitarily equivalent distributions*

Two operators $\boldsymbol{A}$ and $\boldsymbol{B}$ are unitarily equivalent if there exists a unitary operator $\boldsymbol{U}$ such that [BAR 95b, BAR 98b]

$$\boldsymbol{B} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}\,.$$

Let a time-frequency representation $C_x(t,f)$ be covariant to the operators of time-shift ($\boldsymbol{T}$) and of frequency-shift ($\boldsymbol{F}$), and simultaneously let $C_x(t,f)$ satisfy the marginal properties in time and in frequency. Let $\boldsymbol{U}$ be a unitary operator. Then, calculating the distribution C of $\boldsymbol{U}x$ amounts to defining a new representation P that is covariant to the unitarily transformed operators

$$\boldsymbol{T}' = \boldsymbol{U}^{-1}\boldsymbol{T}\boldsymbol{U} \qquad \text{and} \qquad \boldsymbol{F}' = \boldsymbol{U}^{-1}\boldsymbol{F}\boldsymbol{U}\,.$$

Moreover, because C_x satisfies the marginal properties, P_x satisfies the following integral relations:

$$\int P_x(t',f')\,df' = |\boldsymbol{U}x(t')|^2\,,$$

$$\int P_x(t',f')\,dt' = \left|\int x(t)\,\left(\boldsymbol{U}^{-1}\left\{e^{j2\pi f't}\right\}\right)^*\,dt\right|^2,$$

where the warped plane (t',f') is defined by the physical variables of the translation operators $\boldsymbol{T}'$ and $\boldsymbol{F}'$ defined above.

In other words, any distribution that is covariant to a group of transformations unitarily equivalent to the Weyl-Heisenberg group of translations is, up to an appropriate axis warping, the image of a specific time-frequency representation of Cohen's class [BAR 95b]. Therefore, all properties of these new distributions can be deduced from the standard properties of Cohen's class via the action of the unitary operator $\boldsymbol{U}$ [BAR 95b]. The remarkable point to retain here is that in many cases, this match by unitary equivalence reduces the arduous problem of constructing covariant distributions to the potentially simpler problem of finding an appropriate axis transformation.

Unitary equivalences of comparable nature exist between the affine class of time-scale distributions and other classes of distributions. However, it is important to note that time-frequency translations (Weyl-Heisenberg group) and time-scale translations (affine group) are not representations of the same algebraic group. For this reason, there does not exist a unitary operator $\boldsymbol{U}$ mapping these two pairs of displacement operators, and hence Cohen's class and the affine class cannot be deduced from one another by means of unitary equivalences.

To illustrate this principle of unitary equivalence, we first consider the *hyperbolic class* [PAP 93, HLA 97a] whose elements are given by the following parametric form:

$$Y_x(t,f) = \frac{f}{f_0} \int_0^\infty \int_0^\infty \widehat{K}(f_1,f_2)\, \widehat{x}\Big(\frac{f_1 f}{f_0}\Big)\, \widehat{x}^*\Big(\frac{f_2\, f}{f_0}\Big)$$

$$\cdot\; e^{j2\pi f t\,[\ln(f_1/f_0)-\ln(f_2/f_0)]}\, df_1\, df_2\,, \qquad f>0\,,$$

where $\widehat{K}$ is a kernel and f_0 is an arbitrarily chosen positive reference frequency. The hyperbolic class is unitarily equivalent to Cohen's class because all its elements $Y_x(t,f)$ can be deduced from a corresponding element $C_x(t,f)$ of Cohen's class using the unitary transformation

$$Y_x(t,f) = C_{\boldsymbol{U}x}\bigg(\frac{tf}{f_0}, f_0 \ln\Big(\frac{f}{f_0}\Big)\bigg), \qquad f>0\,. \tag{7.45}$$

Here, the action of the unitary operator $\boldsymbol{U}$ is described in the Fourier domain by $\big(\widehat{\boldsymbol{U}}\widehat{x}\big)(f) = \sqrt{e^{f/f_0}}\, \widehat{x}\big(f_0\, e^{f/f_0}\big)$, $f \in \mathbb{R}$. This is a logarithmic frequency compression which, combined with the exponential frequency *warping* (7.45), converts the "constant-bandwidth" analysis performed by $C_x(t,f)$ into a "constant-Q" analysis [PAP 93].

The hyperbolic representations satisfy a double covariance property to scale changes and to dispersive time shifts with logarithmic phase (see equation (7.39) with $k\!=\!0$, corresponding to a hyperbolic group delay $\tau(f) \propto 1/f$):

$$Y_{\boldsymbol{T}_{a,b}x}(t,f) = Y_x\bigg(\frac{t-b/f}{a}, af\bigg) \quad \text{with} \quad \big(\widehat{\boldsymbol{T}}_{a,b}\,\widehat{x}\big)(f) = \sqrt{a}\; e^{-j2\pi b \ln(f/f_0)}\, \widehat{x}(af)\,. \tag{7.46}$$

Although the definition of the hyperbolic class differs from that of the affine class by the absence of time shift covariance, the structure of covariance (7.46) is somewhat reminiscent of the one in equations (7.9) and (7.40) with $k=0$. In fact, as we saw in Section 7.4.2, the intersection between the hyperbolic class and the affine class is not empty. It contains the family of affine Wigner distributions P_0 (equation (7.41) with $k=0$), which are covariant by G_0, the extension of the affine group to three parameters (equation (7.38) with $k=0$).

Another interesting illustration of the unitary equivalence principle is provided by the *power classes* [HLA 99, PAP 96]. The power class with power parameter k (where $k \in \mathbb{R}\backslash\{0\}$) is given by

$$\Omega_x^{(k)}(t,f) = \frac{f}{f_0} \int_0^\infty \int_0^\infty \widehat{K}(f_1,f_2)\, \widehat{x}\Big(\frac{f_1 f}{f_0}\Big)\, \widehat{x}^*\Big(\frac{f_2\, f}{f_0}\Big)$$

$$\cdot\; \exp\bigg(j2\pi \frac{ft}{k}\bigg[\Big(\frac{f_1}{f_0}\Big)^k - \Big(\frac{f_2}{f_0}\Big)^k\bigg]\bigg)\, df_1\, df_2\,, \qquad f>0\,.$$

It is easily verified that the affine class (see (7.10)) is the power class with $k\!=\!1$. More remarkably, all power classes are unitarily equivalent to the affine class. In fact, each element $\Omega_x^{(k)}(t,f)$ of the power class with parameter k corresponds to an element $\Omega_x(t,f) = \Omega_x^{(1)}(t,f)$ of the affine class via the unitary transformation

$$\Omega_x^{(k)}(t,f) \;=\; \Omega_{\boldsymbol{U}_k x}\left(\frac{t}{k\,(f/f_0)^{k-1}},\, f_0\left(\frac{f}{f_0}\right)^k\right), \qquad f > 0\,, \tag{7.47}$$

where the unitary operator $\big(\widehat{\boldsymbol{U}}_k \widehat{x}\big)(f) = \sqrt{\frac{1}{|k|}\left(\frac{f}{f_0}\right)^{\frac{1}{k}-1}}\;\widehat{x}\Big(f_0\left(\frac{f}{f_0}\right)^{\frac{1}{k}}\Big)$, $f > 0$, performs a power-law frequency *warping*.

Thus, in addition to being covariant to scale changes, the representations of the power class with parameter k are also covariant to dispersive temporal translations with power-law group delay ($\tau(f) \propto k\, f^{k-1}$):

$$\Omega^{(k)}_{\boldsymbol{T}_{a,b}x}(t,f) \;=\; \Omega_x^{(k)}\left(\frac{t - b\,k\,(f/f_0)^{k-1}/f_0}{a},\, af\right),$$

where $\big(\widehat{\boldsymbol{T}}_{a,b}\,\widehat{x}\big)(f) = \sqrt{a}\; e^{-j2\pi b\,(f/f_0)^k}\;\widehat{x}(af)$. For $k \neq 0, 1$, we here recognize the covariance properties (7.39) and (7.40) of the affine Wigner distributions P_k (associated with the same parameter k). As the affine Wigner distributions P_k additionally satisfy the time-shift covariance property (see Section 7.4.2), they stand at the intersection of the power classes with the affine class of time-frequency representations [BER 92b, HLA 99].

In addition to the examples just considered, there is an infinity of other classes of bilinear representations that are unitarily equivalent to Cohen's class or to the affine class. All of them satisfy a certain twofold covariance property (which, moreover, constitutes an axiomatic definition) and all are associated with a specific "time-frequency geometry" [BAR 95b, BAR 96a, FLA 96, PAP 01, HLA 03b].

7.6. Conclusions

As we have seen, principles of the same nature govern the constructions of energy time-frequency representations that belong either to the affine class or to Cohen's class. However, the essential difference between these two classes comes from the algebraic structures of their underlying groups – the affine group in the first case and the Weyl-Heisenberg group in the second. In addition, precisely because there is no unitary equivalence between these two groups, affine class and Cohen's class constitute specific analysis tools that are non-interchangeable and, indeed, complementary. While the harmonic analysis of non-stationary signals remains the privileged field of Cohen's class, time-frequency distributions of the affine class will certainly be preferred when it comes to highlighting the presence or the absence of characteristic scales in certain signals or systems. Similarly, when it is desired to locally (or globally) characterize scaling structures, such as singularities, long-range dependence, self-similarities, etc., resorting to elements of the affine class will definitely be advantageous.

With these two classes of distributions, and all those which may be derived from them by unitary equivalence, we have at our disposal a very rich arsenal of analysis tools for non-stationary signals. Curiously, however, a difficulty often encountered in the application of these objects lies in the diversity of the choice offered! For this versatility not to constitute a barrier to the diffusion and to the development of these methods, it is important to make the choice of a distribution dependent on the targeted theoretical properties, and not, for example, systematically consider the absence of interference terms as an absolute necessity for the analysis of complex or broadband signals.

Finally, we stress that the algorithmic cost of energy time-frequency distributions has for a long time penalized them with respect to some less complex representations based on linear decompositions (e.g., short-term Fourier transform and wavelet transform). Today, this imbalance is less pronounced. In particular, since in many cases the calculation cost of affine Wigner distributions is no longer a problem, let us recall that the scalogram (squared modulus of a wavelet transform) is merely one specific element of the affine class. In view of the many desirable properties offered by time-frequency representations of the affine class and related classes, we may count on the potential of these tools for signal processing and, perhaps one day, even for image processing.

7.7. Bibliography

[ARN 95] ARNÉODO A., BACRY E., MUZY J. F., "The thermodynamic of fractals revisited with wavelets", *Physica A*, vol. 213, pp. 232–275, 1995.

[BAR 95a] BARANIUK R. G., COHEN L., "On joint distributions for arbitrary variables", *IEEE Signal Process. Lett.*, vol. 2, no. 1, pp. 10–12, Jan. 1995.

[BAR 95b] BARANIUK R. G., JONES D. L., "Unitary equivalence: A new twist on signal processing", *IEEE Trans. Signal Process.*, vol. 43, no. 10, pp. 2269–2282, Oct. 1995.

[BAR 96a] BARANIUK R. G., "Covariant time-frequency representations through unitary equivalence", *IEEE Signal Process. Lett.*, vol. 3, no. 3, pp. 79–81, Mar. 1996.

[BAR 96b] BARANIUK R. G., "A limitation of the kernel method for joint distributions of arbitrary variables", *IEEE Signal Process. Lett.*, vol. 3, no. 2, pp. 51–53, Feb. 1996.

[BAR 98a] BARANIUK R. G., "Beyond time-frequency analysis: Energy densities in one and many dimensions", *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2305–2314, Sep. 1998.

[BAR 98b] BARANIUK R. G., "Joint distributions of arbitrary variables made easy", *Multidim. Syst. Signal Process.*, vol. 9, no. 4, pp. 341–348, Oct. 1998.

[BER 85] BERTRAND J., BERTRAND P., "Représentation des signaux à large bande", *La Recherche Aérospatiale*, vol. 5, pp. 277–283, Sep.–Oct. 1985.

[BER 90] BERTRAND J., BERTRAND P., OVARLEZ J. P., "Discrete Mellin transform for signal analysis", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Albuquerque (NM), pp. 1603–1606, Apr. 1990.

[BER 92a] BERTRAND J., BERTRAND P., "Affine time-frequency distributions", BOASHASH B., Ed., *Time-Frequency Signal Analysis – Methods and Applications*, pp. 118–140, Longman Cheshire, Melbourne (Australia), 1992.

[BER 92b] BERTRAND J., BERTRAND P., "A class of affine Wigner functions with extended covariance properties", *J. Math. Phys.*, vol. 33, no. 7, pp. 2515–2527, Jul. 1992.

[BER 95] BERTRAND J., BERTRAND P., OVARLEZ J.-P., "The Mellin transform", POULARIKAS A. D., Ed., *The Transforms and Applications Handbook*, pp. 829–885, CRC Press, Boca Raton (FL), 1995.

[CLA 80] CLAASEN T. A. C. M., MECKLENBRÄUKER W. F. G., "The Wigner distribution – A tool for time-frequency signal analysis, Part I : Continuous-time signals", *Philips J. Res.*, vol. 35, no. 3, pp. 217–250, 1980.

[COH 66] COHEN L., "Generalized phase-space distribution functions", *J. Math. Phys.*, vol. 7, no. 5, pp. 781–786, May 1966.

[COH 95] COHEN L., *Time-Frequency Analysis*, Prentice-Hall, Englewood Cliffs (NJ), 1995.

[COH 96] COHEN L., "A general approach to obtaining joint representations in signal analysis – Part I : Characteristic function operator method", *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1080–1090, May 1996.

[DAU 92] DAUBECHIES I., *Ten Lectures on Wavelets*, SIAM, 1992.

[FLA 84] FLANDRIN P., "Some features of time-frequency representations of multi-component signals", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, San Diego (CA), pp. 41.B.4.1–4, 1984.

[FLA 90] FLANDRIN P., RIOUL O., "Affine smoothing of the Wigner-Ville distribution", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Albuquerque (NM), pp. 2455–2458, Apr. 1990.

[FLA 91] FLANDRIN P., "Sur une classe générale d'extensions affines de la distribution de Wigner-Ville", *13ème Coll. GRETSI*, Juan-les-Pins (France), pp. 17–20, 1991.

[FLA 96] FLANDRIN P., GONÇALVÈS P., "Geometry of affine time-frequency distributions", *Appl. Comput. Harm. Anal.*, vol. 3, pp. 10–39, 1996.

[FLA 99] FLANDRIN P., *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego, CA, 1999.

[GON 96] GONÇALVÈS P., BARANIUK R. G., "A pseudo-Bertrand distribution for time-scale analysis", *IEEE Signal Process. Lett.*, vol. 3, no. 3, pp. 82–84, Mar. 1996.

[GON 97] GONÇALVÈS P., FLANDRIN P., CHASSANDE-MOTTIN E., "Time-frequency methods in time-series data analysis", *Second Workshop on Gravitational Wave Data Analysis*, pp. 35–46, 1997.

[GON 98] GONÇALVÈS P., BARANIUK R. G., "Pseudo affine Wigner distributions: Definition and kernel formulation", *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1505–1516, Jun. 1998.

[GRO 84] GROSSMANN A., MORLET J., "Decomposition of Hardy functions into square integrable wavelets of constant shape", *SIAM J. Math. Anal.*, vol. 15, no. 4, pp. 723–736, 1984.

[HLA 93] HLAWATSCH F., PAPANDREOU A., BOUDREAUX-BARTELS G. F., "Regularity and unitarity of affine and hyperbolic time-frequency representations", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Minneapolis (MN), vol. 3, pp. 245–248, Apr. 1993.

[HLA 94] HLAWATSCH F., BÖLCSKEI H., "Unified theory of displacement-covariant time-frequency analysis", *Proc. IEEE Int. Sympos. Time-Frequency Time-Scale Anal.*, Philadelphia (PA), pp. 524–527, Oct. 1994.

[HLA 97a] HLAWATSCH F., PAPANDREOU-SUPPAPPOLA A., BOUDREAUX-BARTELS G. F., "The hyperbolic class of quadratic time-frequency representations – Part II: Subclasses, intersection with the affine and power classes, regularity, and unitarity", *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 303–315, Feb. 1997.

[HLA 97b] HLAWATSCH F., TWAROCH T., "Extending the characteristic function method for joint *a–b* and time-frequency analysis", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Munich (Germany), pp. 2049–2052, Apr. 1997.

[HLA 99] HLAWATSCH F., PAPANDREOU-SUPPAPPOLA A., BOUDREAUX-BARTELS G. F., "The power classes – Quadratic time-frequency representations with scale covariance and dispersive time-shift covariance", *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3067–3083, Nov. 1999.

[HLA 03a] HLAWATSCH F., TAUBÖCK G., "The covariance theory of time-frequency analysis", BOASHASH B., Ed., *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, pp. 102–113, Elsevier, Oxford (UK), 2003.

[HLA 03b] HLAWATSCH F., TAUBÖCK G., TWAROCH T., "Covariant time-frequency analysis", DEBNATH L., Ed., *Wavelets and Signal Processing*, Chapter 7, pp. 177–231, Birkhäuser, Boston (MA), 2003.

[LUD 66] LUDWIG D., "The Radon transform on Euclidean space", *Comm. Pure Appl. Math.*, vol. XIX, pp. 49–81, 1966.

[MAL 89] MALLAT S. G., "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[MAL 99] MALLAT S. G., *A Wavelet Tour of Signal Processing*, Academic Press, 2nd edition, 1999.

[MAN 68] MANDELBROT B., VAN NESS J. W., "Fractional Brownian motions, fractional noises and applications", *SIAM Rev.*, vol. 10, no. 4, pp. 422–437, 1968.

[MOY 49] MOYAL J. E., "Quantum mechanics as a statistical theory", *Proc. Cambridge Phil. Soc.*, vol. 45, pp. 99–124, 1949.

[OVA 92a] OVARLEZ J.-P., La Transformation de Mellin : un outil pour l'analyse des signaux à large bande, Doctoral dissertation, University Paris 6, Paris (France), 1992.

[OVA 92b] OVARLEZ J.-P., BERTRAND J., BERTRAND P., "Computation of affine time-frequency distributions using the fast Mellin transform", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, San Francisco (CA), vol. 5, pp. 117–120, Mar. 1992.

[PAP 93] PAPANDREOU A., HLAWATSCH F., BOUDREAUX-BARTELS G. F., "The hyperbolic class of quadratic time-frequency representations – Part I: Constant-Q warping, the hyperbolic paradigm, properties, and members", *IEEE Trans. Signal Process., Special Issue on Wavelets and Signal Processing*, vol. 41, no. 12, pp. 3425–3444, Dec. 1993.

[PAP 96] PAPANDREOU-SUPPAPPOLA A., HLAWATSCH F., BOUDREAUX-BARTELS G. F., "Power class time-frequency representations: Interference geometry, smoothing, and implementation", *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Anal.*, Paris (France), pp. 193–196, Jun. 1996.

[PAP 01] PAPANDREOU-SUPPAPPOLA A., MURRAY R. L., BYEONG-GWAN IEM, BOUDREAUX-BARTELS G. F., "Group delay shift covariant quadratic time-frequency representations", *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2549–2564, Nov. 2001.

[RIO 92] RIOUL O., FLANDRIN P., "Time-scale energy distributions : A general class extending wavelet transforms", *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1746–1757, Jul. 1992.

[SAY 96a] SAYEED A. M., JONES D. L., "A canonical covariance-based method for generalized joint signal representations", *IEEE Signal Process. Lett.*, vol. 3, no. 4, pp. 121–123, Apr. 1996.

[SAY 96b] SAYEED A. M., JONES D. L., "Integral transforms covariant to unitary operators and their implications for joint signal representations", *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1365–1377, Jun. 1996.

[SCU 87] SCULLY M. O., COHEN L., "Quasi-probability distributions for arbitrary operators", KIM Y. S., ZACHARY W. W., Eds., *Proceedings of Nonlinear Dynamics and Chaos, Geometric Quantization, and Wigner Functions*, pp. 253–263, Springer-Verlag, 1987, Berlin (Germany).

[STA 94] STANKOVIĆ L., "A method for time-frequency analysis", *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 225–229, Jan. 1994.

[TWA 99] TWAROCH T., Signal representations and group theory, Doctoral dissertation, Vienna University of Technology, Vienna (Austria), 1999.

[UNT 84] UNTERBERGER A., "The calculus of pseudo-differential operators of Fuchs type", *Comm. Part. Diff. Eq.*, vol. 9, pp. 1179–1236, 1984.

[WOO 53] WOODWARD P., *Probability and Information Theory with Applications to Radar*, Pergamon, London (UK), 1953.