

Joint Distributions of Arbitrary Variables Made Easy

Richard G. Baraniuk *

Department of Electrical and Computer Engineering
Rice University
6100 South Main Street
Houston, TX 77005, USA
Email: richb@rice.edu
URL: <http://www-dsp.rice.edu>

Submitted to *Multidimensional Systems and Signal Processing*, August 1996

Abstract

In this paper, we propose a simple framework for studying certain distributions of variables beyond time-frequency and time-scale. When applicable, our results turn the theory of joint distributions of arbitrary variables into an easy exercise of coordinate transformation. While straightforward, the method can generate many distributions previously attainable only by the general construction of Cohen, including time versus inverse frequency, time versus Mellin transform (scale), and time versus chirp distributions. In addition to providing insight into these new signal analysis tools, warp-based distributions have efficient implementations for potential use in applications.

*This work was supported by the National Science Foundation, grant no. MIP-9457438, and by the Office of Naval Research, grant no. N00014-95-1-0849.

1 Introduction

The successful application of joint time-frequency distributions to problems in time-varying spectral analysis has stimulated considerable recent interest in distributions of other variables for use when a strict time-frequency analysis is not appropriate.

Joint distributions generalise single variable distributions that measure the energy content of some physical quantity in a signal. Given a quantity a represented by the Hermitian (symmetric) operator $\mathcal{A}$, we obtain the density $|(\mathbf{F}_{\mathcal{A}}s)(a)|^2$ measuring the “ a content” of the signal s simply by squaring the projection of s onto the eigenfunctions $\mathbf{u}_a^{\mathcal{A}}$ of $\mathcal{A}$ [1]

$$|(\mathbf{F}_{\mathcal{A}}s)(a)|^2 \equiv \left| \int s(x) [\mathbf{u}_a^{\mathcal{A}}(x)]^* dx \right|^2. \quad (1)$$

Classical examples of single variable densities include the time density $|(\mathbf{F}_{\mathcal{T}}s)(t)|^2 = |s(t)|^2$ and frequency density $|(\mathbf{F}_{\mathcal{F}}s)(f)|^2 = |S(f)|^2$ obtained by projecting onto the Dirac eigenfunctions of the time operator $(\mathcal{T}s)(x) \equiv x s(x)$ and the sinusoidal eigenfunctions of the frequency operator $(\mathcal{F}s)(x) \equiv \frac{1}{j2\pi} \dot{s}(x)$. (We will use both S and $\mathbf{F}s$ to denote the Fourier transform of the signal s .)

Joint distributions attempt to measure the simultaneous signal energy content of multiple physical quantities. Given two quantities a and b , a joint distribution $(\mathbf{P}_{a,b}s)(a,b)$ measuring the joint a - b content in the signal s has as marginals the respective $\mathcal{A}$ and $\mathcal{B}$ energy densities¹

$$\int (\mathbf{P}_{a,b}s)(a,b) db = |(\mathbf{F}_{\mathcal{A}}s)(a)|^2, \quad \int (\mathbf{P}_{a,b}s)(a,b) da = |(\mathbf{F}_{\mathcal{B}}s)(b)|^2. \quad (2)$$

The Wigner distribution from Cohen’s class of time-frequency distributions [1] supplies a classical example of a joint distribution that marginalizes to the time and frequency densities.

Many different constructions have been proposed for generating joint distributions. The various approaches fall into two broad categories: *general methods* (difficult) and *coordinate change methods* (easy). General methods can create distributions for every possible pairing of physical quantities by working from first principles [1, 4]. Coordinate change methods, on the contrary, sacrifice some

¹Alternatively, we can define joint distributions in terms of their covariance properties under certain unitary transformations. For more details, see [2, 3].

flexibility for simplicity by bootstrapping existing distributions into new contexts using signal or axis transformations [4–8].

While not general purpose, coordinate change approaches to joint distributions remain attractive, because they provide a straightforward interpretation of the distributions they construct. Individually, each transformation method has severe limitations, however, as each technique can generate distributions for only a very restricted set of physical quantity pairs. The purpose of this paper is to demonstrate that a novel combination of two different coordinate change procedures can manufacture joint distributions for a much larger number of physical quantity pairs. In particular, our new procedure can generate most of the distributions considered as examples of the general method of Cohen in [1]. Thus, when applicable, our results turn the difficult theory of joint distributions of arbitrary variables into an easy exercise of coordinate transformation.

2 Three Approaches to Joint Distributions

2.1 Cohen’s general method: Given two operators $\mathcal{A}$ and $\mathcal{B}$ representing two arbitrary physical quantities a and b , Cohen forms the joint distribution of a and b as (see [1, 4] for more details)

$$(\mathbf{P}_{a,b} s)(a, b) = \iiint s^*(x) (e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})} s)(x) e^{-j2\pi(\alpha a+\beta b)} dx d\alpha d\beta. \quad (3)$$

Cohen’s construction is general, but it requires that we solve a sometimes complicated operator equation to express the exponentiated operator $e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$. The time-frequency case is well understood; using the time and frequency operators $\mathcal{T}$ and $\mathcal{F}$ yields Cohen’s class of time-frequency distributions [1, p. 136].

2.2 Axis transformation method: Joint a – b distributions are easily obtained when we can functionally relate these variables to time and frequency by $a = \alpha^{-1}(t)$ and $b = \beta^{-1}(f)$. In this special case, we can derive an a – b distribution simply by warping the coordinates of a time-frequency distribution [4]

$$(\mathbf{P}_{a,b} s)(a, b) = \left| \dot{\alpha}(a) \dot{\beta}(b) \right| (\mathbf{P}_{t,f} s)[\alpha(a), \beta(b)]. \quad (4)$$

It is easily verified that all $\mathbf{P}_{a,b}s$ obtained in this way correctly marginalize to $|(\mathbf{F}_{\mathcal{A}}s)(a)|^2 = |\dot{\alpha}(a)| |(\mathbf{F}_{\mathcal{T}}s)(\alpha(a))|^2 = |\dot{\alpha}(a)| |s(\alpha(a))|^2$ and $|(\mathbf{F}_{\mathcal{B}}s)(b)|^2 = |\dot{\beta}(b)| |(\mathbf{F}_{\mathcal{F}}s)(\beta(b))|^2 = |\dot{\beta}(b)| |S(\beta(b))|^2$.

Example: Distributions $\mathbf{P}_{t,r}s$ of time t and “inverse frequency” r (represented by the operator $\mathcal{R} = \frac{f_0}{\mathcal{F}}$) can be constructed from Cohen’s class time-frequency distributions through the change of variable $r = \frac{f_0}{f}$ [1, p. 238]. The resulting class of distributions, obtained as

$$(\mathbf{P}_{t,r}s)(t, r) = \frac{f_0}{r^2} (\mathbf{P}_{t,f}s)\left(t, \frac{f_0}{r}\right), \quad (5)$$

marginalizes to time $|s(t)|^2$ and inverse frequency $|(\mathbf{F}_{\mathcal{R}}s)(r)|^2 = \frac{f_0}{r^2} |S(f_0/r)|^2$.

2.3 Signal transformation method: Joint a – b distributions are also easily generated when the quantities a and b are *unitarily equivalent* to time and frequency [5, 6], with

$$\mathcal{A} = \mathbf{U}^{-1}\mathcal{T}\mathbf{U}, \quad \mathcal{B} = \mathbf{U}^{-1}\mathcal{F}\mathbf{U} \quad (6)$$

and $\mathbf{U}$ a unitary transformation. In this case, a joint a – b distribution can be obtained by preprocessing a time-frequency distribution with $\mathbf{U}$

$$(\mathbf{P}_{a,b}s)(a, b) = (\mathbf{P}_{t,f}\mathbf{U}s)(a, b). \quad (7)$$

The signal transformation $\mathbf{U}$ can be interpreted as “rotating” the coordinates of the time-frequency distribution to the new variables.

All $\mathbf{P}_{a,b}s$ obtained in this way correctly marginalize to $|(\mathbf{F}_{\mathcal{A}}s)(a)|^2 = |(\mathbf{F}_{\mathcal{T}}\mathbf{U}s)(a)|^2 = |(\mathbf{U}s)(a)|^2$ and $|(\mathbf{F}_{\mathcal{B}}s)(b)|^2 = |(\mathbf{F}_{\mathcal{F}}\mathbf{U}s)(b)|^2 = |(\mathbf{F}\mathbf{U}s)(b)|^2$ [5].

Example: Define the logarithmic time operator $(\mathcal{L}s)(x) \equiv \log(x) s(x)$, and define the Mellin operator $\mathcal{H} \equiv \frac{1}{2}(\mathcal{T}\mathcal{F} + \mathcal{F}\mathcal{T})$. (Cohen refers to $\mathcal{H}$ as the “scale” operator [1, 9].) These operators are unitarily equivalent to $\mathcal{T}$ and $\mathcal{F}$ as in (6) with $(\mathbf{U}s)(x) = e^{x/2} s(e^x)$. Therefore, we can construct distributions marginalizing to log time $|(\mathbf{F}_{\mathcal{L}}s)(l)|^2 = |e^{l/2} s(e^l)|^2$ and Mellin transform

$$|(\mathbf{F}_{\mathcal{H}}s)(m)|^2 = \left| \int s(x) e^{-j2\pi m \log x} |x|^{-1/2} dx \right|^2 \quad (8)$$

simply by preprocessing the signal by $\mathbf{U}$ before computing a valid time-frequency distribution [5–8].

2.4 Shortcomings of transformation methods: While simple, both the axis and signal transformation methods place rather severe restrictions on the pairs of operators for which they can generate joint distributions. In the axis transformation method, we require the variables a and b to be warped versions of time and frequency.² In the signal transformation method, we require the operators $\mathcal{A}$ and $\mathcal{B}$ to be unitarily equivalent to time and frequency. Note, for instance, that in warping the frequency marginal to Mellin transform as above, we must accept log time for the other marginal.

3 Linking Signal and Axis Transformations

Due to the individual limitations of the axis and signal transformation methods, only general operator methods can generate joint distributions for all possible operator pairs. However, when used in *tandem*, the axis and signal transformation methods yield a powerful method for generating a large number of joint distributions. By executing first a unitary preprocessing transformation on the signal and then an axis warping transformation on the distribution, we can remap time-frequency distributions to a large class of different joint distributions.

Theorem: Let a and b be two variables whose Hermitian operator representations $\mathcal{A}$ and $\mathcal{B}$ satisfy the following conditions: (i) $\mathcal{A}$ and $\mathcal{B}$ can be related to the time and frequency operators as

$$\mathcal{A} = \mathbf{Z}^{-1} \mathcal{T} \mathbf{Z}, \quad \mathcal{B} = \mathbf{U}^{-1} \mathcal{F} \mathbf{U}, \quad (9)$$

with $\mathbf{Z}$ and $\mathbf{U}$ unitary transformations. (ii) The composition $\mathbf{V} \equiv \mathbf{Z} \mathbf{U}^{-1}$ is an axis warping operator of the form

$$(\mathbf{V}g)(x) = |\dot{v}(x)|^{1/2} g[v(x)] \quad (10)$$

with v a smooth, 1–1 function. Then a joint a – b distribution $(\mathbf{P}_{a,b} s)(a, b)$ can be constructed from a time-frequency distribution $(\mathbf{P}_{t,f} s)(t, f)$ through

$$(\mathbf{P}_{a,b} s)(a, b) = |\dot{v}(a)| (\mathbf{P}_{t,f} \mathbf{U} s)[v(a), b]. \quad (11)$$

²In fact, it was precisely this shortcoming that lead Scully and Cohen to develop the general approach of Section 2.1.

Proof: Direct evaluation shows that each distribution of the form (11) marginalizes to $|(\mathbf{F}_{\mathcal{A}}s)(a)|^2 = |(\mathbf{Z}s)(a)|^2$ and $|(\mathbf{F}_{\mathcal{B}}s)(b)|^2 = |(\mathbf{F}\mathbf{U}s)(b)|^2$.

The transformation $\mathbf{U}$ rotates the (t, f) coordinates of the time-frequency distribution to new coordinates (u, b) . The transformation $\mathbf{V}$ then warps the rotated time axis u to align it with the quantity a .

In other words, if $\mathcal{A}$ and $\mathcal{B}$ relate to $\mathcal{T}$ and $\mathcal{F}$ as in (9) and (10), then we can obtain a large class of a - b distributions using a simple three-step procedure:³ (i) transform the signal $s \mapsto \mathbf{U}s$; (ii) compute a conventional time-frequency distribution of the transformed signal; (iii) warp the remapped time axis of the resulting distribution. A numerical algorithm for warping time-frequency distributions suitable for use on the digital computer is readily programmed.

The advantage of the double transformation method is that it breaks the severe restrictions placed on the quantities a and b by both the axis and signal transformation methods described in Section 2. By allowing the choice of both $\mathbf{U}$ and $\mathbf{Z}$, we can derive a much larger class of distributions. However, completely free choice is still not possible, because $\mathbf{U}$ and $\mathbf{Z}$ must have the structural property of cancelling modulo the warping operator $\mathbf{V}$.

Several extensions to the theorem are straightforward. Reversing the rôles of time and frequency moves the warping transformation in (11) into the second argument. We also do not have to start from time-frequency distributions. We can map existing distributions based on arbitrary operators $\mathcal{C}$, $\mathcal{D}$ to distributions based on $\mathcal{A}$, $\mathcal{B}$ so long as all four operators obey a relation analogous to (9). The advantage of starting from time-frequency distributions is simply that they are well understood and therefore aid in understanding the new a - b distributions.

³If we define the 2-d transformation $\overline{\mathbf{V}}$ as the area-preserving change of variables

$$(\overline{\mathbf{V}}G)(x, y) \equiv |\dot{v}(x)| G[v(a), b],$$

then we can write (11) in the standard form $\mathbf{P}_{a,b} = \overline{\mathbf{V}}\mathbf{P}_{t,f}\mathbf{U}$ found in [5]. However, whereas [5] emphasized using $\overline{\mathbf{V}}$ only to warp both axes of $\mathbf{P}_{a,b}s$ back to indicate time and frequency, in this paper, we exploit a range of different $\overline{\mathbf{V}}$.

4 Examples

Linked signal/axis transformations are especially useful for transforming time-frequency distributions to distributions of time versus a “warped frequency” variable. In this case, we set $\mathbf{Z}$ to the identity operator, choose $\mathbf{U}$ to be a warping operator based around the axis warping function v^{-1} , and set $\mathbf{V} = \mathbf{U}^{-1}$ to warp the rotated time axis back the time variable. The resulting distributions marginalize to time and the integral transform

$$\int (\mathbf{P}_{t,b} s)(t, b) dt = \left| \int s(x) e^{-j2\pi b v(x)} |\dot{v}(x)|^{1/2} dx \right|^2. \quad (12)$$

Each choice of v results in a different warped frequency transform matched to a distinct class of instantaneous frequency characteristics. The Fourier, Mellin, and chirp transforms result from the choices $v(x) = x$, $v(x) = \log x$, and $v(x) = |x|^c \operatorname{sgn}(x)$, respectively.

To continue the example of Section 2.3, applying the warp $\mathbf{U}^{-1}$ to the log time axis of the log time versus Mellin distribution remaps that axis back to true time. The resulting distributions lie in Cohen’s class of time-Mellin distributions (time-scale in his terminology) [1, 9]. This class contains the Marinovich-Altes (warped Wigner) distribution [7, 10]. It is important to note that this class is unattainable by either signal or axis warping alone.

In Figure 1 we show two distributions of a signal consisting of two components concentrated along composite linear/sinusoidal instantaneous frequencies. Since the Wigner time-frequency distribution does not match signals of this type as well as sinusoids, impulses, and linear chirps, it exhibits copious cross-components. Prewarping the signal to account for the form of the signal yields a postwarped distribution that marginalizes to time and “composite linear/sinusoidal instantaneous frequency” content and therefore better matches the signal.

Reversing the rôles of time and frequency in the warping procedure will yield frequency versus warped time distributions that match different classes of group delay (dispersion) characteristics.

5 Going Further

While our approach here has been based on satisfying marginal distributions, analysis of the localization and covariance properties of warped time-frequency distributions is simple. An analysis similar to that in [5] shows that warped Cohen’s class time-frequency distributions localize along warped impulses, sinusoids, and linear chirps. Unfortunately, except for the special case $\mathbf{Z} = \mathbf{U}$ in the Theorem, most warped distributions lack any useful covariance to unitary transformations.⁴ In fact, the time versus “warped frequency” distributions of Section 4 will not even be covariant to time shifts in general. While somewhat disappointing, we note that lack of covariance is a limitation of joint distributions of arbitrary variables in general and not just of this particular approach [11].

In certain situations we may desire simpler (yet non-orthogonal) functions of the form $e^{-j2\pi bv(x)}$ rather than $|\dot{v}(x)|^{1/2} e^{-j2\pi bv(x)}$ in the “warped frequency” integral transform on the right side of (12). Marginals of this type are not supported by the general construction of Section 2.1, since non-orthogonal eigenfunctions dictate non-Hermitian operators which in turn lead to complex-valued distributions. However, using non-unitary warping operators, we can construct distributions with these streamlined marginals by migrating the measure $|\dot{v}(x)|$ into the left side of the calculation (2).

The signal/axis warping procedure can be applied to higher-order distributions such as the positive distributions of Cohen [1]. Positive distributions of arbitrary variables can also be set up directly, as

$$(\mathbf{P}_{a,b}^{\mathbf{P}}s)(a, b) = |(\mathbf{F}_{\mathcal{A}}s)(a)|^2 |(\mathbf{F}_{\mathcal{B}}s)(b)|^2 \Omega(a, b) \quad (13)$$

with Ω defined as in the time-frequency case described in [1, p. 198].

Acknowledgements: Thanks to Leon Cohen for suggesting this more in-depth analysis of warped time-frequency distributions. Thanks also to Faye Boudreaux-Bartels for valuable discussions and to Doug Jones for suggesting the use of the JAM in this context.

⁴By contrast, Cohen’s class time-frequency distributions [1] are covariant to time and frequency shifts and affine class distributions [2, 3] are covariant to time shifts and scale changes.

References

- [1] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [2] O. Rioul and P. Flandrin, "Time-scale energy distributions: A general class extending wavelet transforms," *IEEE Trans. Signal Processing*, vol. 40, pp. 1746–1757, July 1992.
- [3] J. Bertrand and P. Bertrand, "A class of affine Wigner functions with extended covariance properties," *J. Math. Phys.*, vol. 33, pp. 2515–2527, July 1992.
- [4] M. O. Scully and L. Cohen, "Quasi-probability distributions for arbitrary operators," in *Proceedings of Nonlinear Dynamics and Chaos, Geometric Quantization, and Wigner Functions* (Y. S. Kim and W. W. Zachary, eds.), (Berlin), pp. 253–263, Springer-Verlag, 1987.
- [5] R. G. Baraniuk and D. L. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. Signal Processing*, vol. 43, pp. 2269–2282, Oct. 1995.
- [6] R. G. Baraniuk and L. Cohen, "On joint distributions of arbitrary variables," *IEEE Signal Processing Letters*, vol. 2, pp. 10–12, Jan. 1995.
- [7] R. A. Altes, "Wideband, proportional-bandwidth Wigner-Ville analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1005–1012, June 1990.
- [8] A. Papandreou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "The hyperbolic class of quadratic time-frequency representations. Part I: Constant-Q warping, the hyperbolic paradigm, properties, and members," *IEEE Trans. Signal Processing*, vol. 41, pp. 3425–3444, Dec. 1993.
- [9] L. Cohen, "The scale representation," *IEEE Trans. Signal Processing*, vol. 41, pp. 3275–3292, Dec. 1993.
- [10] G. Eichmann and N. M. Marinovich, "Scale-invariant Wigner distribution," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 519, pp. 18–25, 1984.
- [11] R. G. Baraniuk, "Covariant time-frequency representations through unitary equivalence," *IEEE Signal Processing Letters*, vol. 3, pp. 79–81, Mar. 1996.

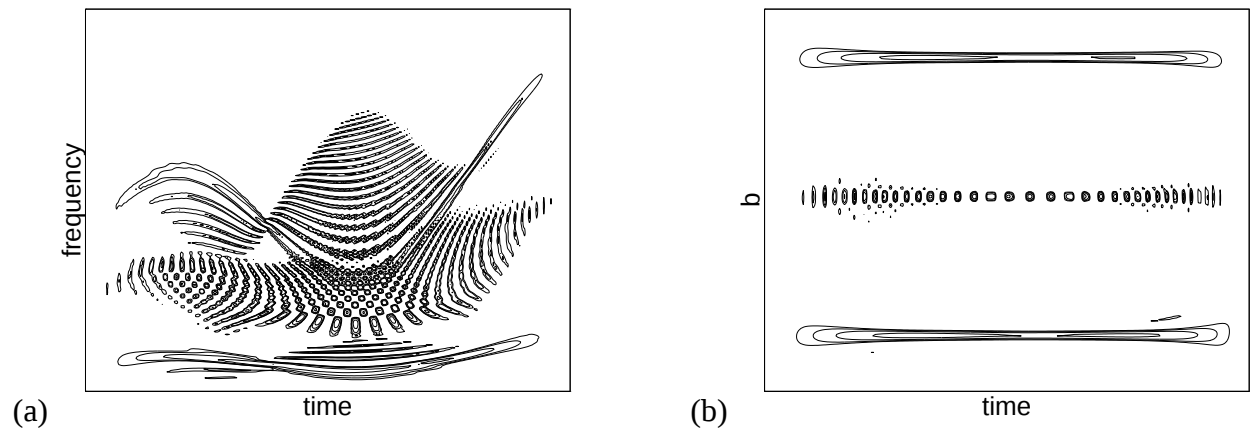


Figure 1: (a) Wigner time-frequency distribution of a synthetic test signal. (b) Warped Wigner distribution with time and “composite linear/sinusoidal instantaneous frequency” (variable “ b ”) marginals. The variable b measures the variation of the linear/sinusoidal instantaneous frequency path in time-frequency.