Joint Distributions of Arbitrary Variables Made Easy

Richard G. Baraniuk*

Department of Electrical and Computer Engineering Rice University Houston, TX 77005, USA

ABSTRACT

In this paper, we propose a new framework for studying joint 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, our method can generate many distributions previously attainable only by the general method of Cohen.

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 $|(\mathbb{F}_{\mathcal{A}}s)(a)|^2$ measuring the "*a* content" of the signal *s* simply by squaring the projection of *s* onto the eigenfunctions $u_a^{\mathcal{A}}$ of \mathcal{A} [1]

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

Classical examples of single variable densities include the time density $|(\mathbb{F}_{\mathcal{T}}s)(t)|^2 = |s(t)|^2$ and frequency density $|(\mathbb{F}_{\mathcal{T}}s)(f)|^2 = |(S)(f)|^2$ 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 $(\mathcal{F}s)(x) \equiv \frac{1}{i^2\pi}\dot{s}(x)$.

Email: richb@rice.edu. URL: http://www-dsp.rice.edu.

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

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

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 flexibility for simplicity by bootstrapping existing distributions into new contexts using signal or axis transformations [5].

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,

Proc. 1996 IEEE Digital Signal Processing Workshop

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

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

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 A and 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)

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

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

2.2. Axis transformation method

Joint a-b distributions are easily obtained when we can 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]

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

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

Example: Distributions $P_{t,rs}$ 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]. The resulting class of distributions, obtained as

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

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

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], with

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

and U a unitary transformation. In this case, a joint a-b distribution can be obtained by preprocessing a time-frequency distribution with the transformation U

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

The signal transformation U can be interpreted as "rotating" the coordinates of the time-frequency distribution to the new variables.

All $\mathbf{P}_{a,bs}$ obtained in this way correctly marginalize to $|(\mathbb{F}_{\mathcal{A}}s)(a)|^2 = |(\mathbb{F}_{\mathcal{T}}\mathbf{U}s)(a)|^2 =$ $|(\mathbf{U}_s)(a)|^2$ and $|(\mathbb{F}_{\mathcal{B}}s)(b)|^2 = |(\mathbb{F}_{\mathcal{F}}\mathbf{U}s)(b)|^2 =$ $|(\mathbb{F}\mathbf{U}s)(b)|^2$, with $\mathbb{F} \equiv \mathbb{F}_{\mathcal{F}}$ the usual Fourier transform operator [5].

Example: Define the logarithmic time operator $(\mathcal{L}s)(x) \equiv \log(x) s(x)$, and define the Mellin operator $\mathcal{H} = \frac{1}{2}(\mathcal{TF} + \mathcal{FT})$. (Cohen refers to \mathcal{H} as the "scale" operator [1].) These operators are unitarily equivalent to \mathcal{T} and \mathcal{F} as in (1) 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

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

simply by preprocessing the signal by U before computing a valid time-frequency distribution [5, 6].

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

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

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

Conventional wisdom has maintained that, owing 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 A and B satisfy the following conditions:

1. A and B can be related to the time and frequency operators as

$$\mathcal{A} = \mathbf{Z}^{-1} \mathcal{T} \mathbf{Z}, \qquad \mathcal{B} = \mathbf{U}^{-1} \mathcal{F} \mathbf{U}, \quad (2)$$

with Z and U unitary transformations.

2. The composition $V \equiv ZU^{-1}$ is an axis warping operator of the form

$$(\nabla g)(x) = |\dot{v}(x)|^{1/2} g[v(x)].$$
 (3)

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

$$(\mathbf{P}_{a,b} s)(a,b) = |\dot{v}(a)| (\mathbf{P}_{t,f} \mathbf{U} s)[v(a),b].$$
 (4)

Proof: Direct evaluation shows that each distribution of the form (4) marginalizes to $|(\mathbb{F}_{\mathcal{A}}s)(a)|^2 = |(\mathbb{Z}s)(a)|^2$ and $|(\mathbb{F}_{\mathcal{B}}s)(b)|^2 = |(\mathbb{F}Us)(b)|^2$.

The transformation U rotates the (t, f) coordinates of the time-frequency distribution to new coordinates (u, b). The transformation 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 (2) and (3), then we can obtain a large

class of a-b distributions using a simple three-step procedure:³

- 1. transform the signal $s \mapsto Us$:
- 2. compute a conventional time-frequency distribution of the transformed signal;
- 3. warp the remapped time axis of the resulting distribution.

A numerical algorithm for warping time-frequency distributions suitable for use on a digital computer is readily programmed.

The advantage of the double transformation procedure 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 U and Z, we can derive a much larger class of distributions. However, completely free choice is still not possible, because U and Z must have the structural property of cancelling modulo the warping operator V.

Several extensions to the theorem are straightforward. Reversing the rôles of time and frequency moves the warping transformation in (4) into the second argument. We also do not have to start from time-frequency distributions. We can map distributions based on arbitrary operators C, Dto distributions based on A, B so long as all four operators obey a relation analogous to (2). The advantage of starting from time-frequency distributions is simply that they are well understood and therefore aid in understanding the new a-bdistributions.

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 Z to the identity operator, choose U to be a warping operator based around the axis warping function v^{-1} , and set $V = U^{-1}$ to warp the rotated time axis back the time variable. The resulting distributions marginalize to time and the transform

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

³If we define the 2-d transformation \overline{V} as the areapreserving change of variables

 $^{(\}overline{\mathbf{V}G})(x,y) \equiv |\dot{v}(x)| G[v(a),b],$ then we can write (4) 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 $\overline{\mathbf{V}}$.

Each different 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 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 (timescale in his terminology) [1]. This class contains the Marinovich-Altes (warped Wigner) distribution [6, 7]. It is important to note that this class is unattainable by either axis or signal 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. CONCLUSIONS

In this paper, we have proposed a new framework for studying joint distributions of arbitrary variables. While our method is straightforward and easy to understand, it can generate many distributions previously attainable only by the (more complicated) general method of Cohen, including time versus frequency, 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, our warp-based constructions lead to efficient implementations for use in applications.⁴

⁴Thanks to Leon Cohen for suggesting this more indepth 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.





Figure 1. (a) Wigner distribution of test signal. (b) 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.

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. A. Altes, "Wideband, proportional-bandwidth Wigner-Ville analysis," *IEEE Trans. Acoust.*, Speech, Signal Processing, vol. 38, pp. 1005–1012, June 1990.
- [7] G. Eichmann and N. M. Marinovich, "Scaleinvariant Wigner distribution," in Proc. SPIE Int. Soc. Opt. Eng., vol. 519, pp. 18-25, 1984.