A Limitation of the Kernel Method for Joint Distributions of Arbitrary Variables

Richard G. Baraniuk, *Member, IEEE*

Abstract— **Recently, Cohen has proposed a construction for joint distributions of arbitrary physical quantities, in direct generalization of joint time-frequency representations. Actually, this method encompasses two approaches: one based on operator correspondences and one based on weighting kernels. The literature has emphasized the kernel method due to its ease of analysis; however, its simplicity comes at a price. In this letter, we use a simple example to demonstrate that the kernel method cannot generate all possible bilinear joint distributions. Our results suggest that the relationship between the operator method and the kernel method merits closer scrutiny.**

I. Introduction

By representing signals in terms of several physical quantities simultaneously, joint distribution functions can reveal signal features that remain hidden from other methods of analysis. Distributions measuring joint time-frequency content, such as the Wigner distribution and the spectrogram from Cohen's class [1], [2], have a long history and continue to play an important role in nonstationary signal analysis. More recently, distributions measuring joint time-scale [3], [4], scale-hyperbolic time [4]–[7], and warped time-frequency and warped time-scale [8], [9] content have been proposed for measuring quantities other than time-frequency.

With this proliferation of new distribution classes, it seems natural to search for general methods for generating all possible joint distributions. In the most successful effort to date, Cohen has extended the formulation of his class of time-frequency distributions to arbitrary variables [2], [10]. When interpreted appropriately, this method is quite general; however, it does have some heretofore undocumented limitations. In this note, we focus on one of them. We begin with a brief review of the general method.

II. Cohen's Kernel Method

The roots of joint distribution theory lie in quantum mechanics, where physical quantities are associated with operators on a Hilbert space (see [2] for more details).

The distribution of a single physical quantity is easily derived. Given the Hermitian operator $\mathcal{A}$ representing a quantity a of interest, the density of a in a signal s corresponds to the square of the projection of s onto the eigenfunctions $\mathbf{u}_a^{\mathcal{A}}$ of $\mathcal{A}$. We term this projection the *$\mathcal{A}$-Fourier transform* [8] and denote it by

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

An alternate route direct to the density employs the exponentiated operator $e^{j2\pi\alpha\mathcal{A}}$ in the *characteristic function*[1]

$$\left\langle s \right| e^{j2\pi\alpha\mathcal{A}} \left| s \right\rangle \equiv \left\langle s,\, e^{j2\pi\alpha\mathcal{A}} s \right\rangle = \int s^*(x) \left(e^{j2\pi\alpha\mathcal{A}} s\right)(x)\, dx.$$

Given the characteristic function of a, the density of a can be obtained by simple Fourier transformation [1], [2]

$$|(\mathbb{F}_{\mathcal{A}} s)(a)|^2 = \int \left\langle s \right| e^{j2\pi\alpha\mathcal{A}} \left| s \right\rangle e^{-j2\pi\alpha a}\, d\alpha. \tag{1}$$

The primary advantage of this second approach to densities is that it generalizes immediately to the multioperator case.

Cohen's pioneering construction for bilinear joint distributions generalizes the prescription (1) to two and higher dimensions. Given two operators $\mathcal{A}$ and $\mathcal{B}$ representing two physical quantities a and b, the joint distribution of a and b is formed as (see [2] and [10] for more details)

$$(\mathbf{C}s)(a,b) = \iint \left\langle s \right| e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})} \left| s \right\rangle e^{-j2\pi(\alpha a+\beta b)}\, d\alpha\, d\beta. \tag{2}$$

Like a true density, $(\mathbf{C}s)(a,b)$ marginalizes to the individual densities of a and b

$$\int (\mathbf{C}s)(a,b)\, db = |(\mathbb{F}_{\mathcal{A}} s)(a)|^2 \tag{3}$$

$$\int (\mathbf{C}s)(a,b)\, da = |(\mathbb{F}_{\mathcal{B}} s)(b)|^2. \tag{4}$$

Unlike single-variable distributions, joint distributions are not unique. In general, the operator representations $\mathcal{A}$ and $\mathcal{B}$ do not commute ($\mathcal{AB} \neq \mathcal{BA}$). Hence, the exponentiated

Manuscript received May 10, 1995. This work was supported by the National Science Foundation, grant no. MIP-9457438, the Office of Naval Research, grant no. N00014-95-1-0849, and the Texas Advanced Technology Program, grant no. TX-ATP 003604-002. The associate editor coordinating the review of this letter and approving it for publication was Prof. G. F. Boudreaux-Bartels.

The author is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251-1892, USA (email: richb@rice.edu).

Publisher Item Identifier S 1070-9908(96)01831-7.

[1] Here, we adopt the physicists' notation of inner products linear in their second element. The exponentiated operator $e^{j2\pi\alpha\mathcal{A}}$ can be evaluated formally using the Taylor series expansion $e^{j2\pi\alpha\mathcal{A}} = \sum_{n=0}^{\infty} \frac{(j2\pi\alpha\mathcal{A})^n}{n!}$. Exponentiated Hermitian operators are unitary [2].

operator $e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$ in the characteristic function of (2) can be evaluated in many ways to obtain different distributions that satisfy the same marginals. The three simplest evaluation or *correspondence rules* are the following [1], [2]: the *Weyl* correspondence $e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$, where the sum $\mathcal{A}+\mathcal{B}$ is exponentiated ensemble, the *normal* correspondence $e^{j2\pi\alpha\mathcal{A}}\,e^{j2\pi\beta\mathcal{B}}$, where $\mathcal{A}$ and $\mathcal{B}$ are exponentiated separately and then composed, and the *antinormal* correspondence $e^{j2\pi\beta\mathcal{B}}\,e^{j2\pi\alpha\mathcal{A}}$, where the order is reversed. Despite the ordering differences, every correspondence rule yields a distribution that marginalizes as in (3) and (4).

Since keeping track of all possible correspondence rules is an arduous task, Cohen has developed a simplified characteristic function method known as the *kernel method* [1], [2], [10]. This method fixes a single correspondence rule and then introduces a kernel function $\phi(\alpha,\beta)$ in (2) to take care of the other possible orderings. The resulting class of bilinear distributions is given by[2]

$$(\mathbf{C}'s)(a,b) = \iint \left\langle s\middle| e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})_{\text{fixed}}}\middle| s\right\rangle \phi(\alpha,\beta)e^{-j2\pi(\alpha a+\beta b)}\,d\alpha\,d\beta \tag{5}$$

where the "fixed" in the characteristic function reminds us that the correspondence rule used to evaluate the exponential remains fixed. Each kernel generates a different a-b distribution; kernels obeying the constraint $\phi(\alpha,0)=\phi(0,\beta)=1\forall\alpha,\beta$ correspond to distributions satisfying the marginals. Other kernels generate generalized distributions that do not satisfy the marginals yet, in some sense, still measure joint a-b energy content.

Cohen originally developed his kernel-based characteristic function method to study distributions of time and frequency (or position and momentum in a quantum mechanics setting) [1], [2]. The Hermitian operators representing time and frequency $(\mathcal{T}s)(x)\equiv x\,s(x)$ and $(\mathcal{F}s)(x)\equiv\frac{1}{j2\pi}\dot{s}(x)$ exponentiate to the time shift and frequency shift operators $\left(e^{j2\pi t\mathcal{F}}s\right)(x)=s(x+t)$ and $\left(e^{j2\pi f\mathcal{T}}s\right)(x)=e^{j2\pi fx}\,s(x)$ [2]. These operators almost commute, with $\mathcal{T}\mathcal{F}-\mathcal{F}\mathcal{T}=\frac{j}{2\pi}$. Fixing the Weyl correspondence in (5) yields the classical formulation of Cohen's class of time-frequency distributions [1], [2]

$$(\mathbf{P}s)(t,f)\equiv\iiint s^*\!\left(u-\frac{\tau}{2}\right)s\!\left(u+\frac{\tau}{2}\right)\phi(\theta,\tau)\;e^{j2\pi(\theta u-\theta t-\tau f)}\,du\,d\theta\,d\tau \tag{6}$$

whose marginals of time $|s(t)|^2$ and frequency $|S(f)|^2$ correspond to signal projections onto the impulse and sinusoid eigenfunctions of $\mathcal{T}$ and $\mathcal{F}$.

III. A Limitation of the Kernel Method

While the kernel method is simple, it has its limitations. In particular, it cannot generate all possible bilinear joint distributions. To show this, we now present a simple counterexample to demonstrate that a single fixed correspondence rule coupled with a kernel weighting cannot simulate all possible rules.

Consider joint distributions of time and Mellin (called "scale" by Cohen [2], [11]) generated by the operators $\mathcal{T}$ and $\mathcal{H}$, where

$$(\mathcal{H}s)(x)\equiv\left(\frac{\mathcal{T}\mathcal{F}+\mathcal{F}\mathcal{T}}{2}s\right)(x).$$

For single-sided signals $s(x)$ defined on $x>0$, the $\mathsf{F}_{\mathcal{H}}$ transform corresponds to the Mellin transform $M(\gamma)\equiv\int s(x)\,e^{-j2\pi\gamma\ln x}x^{-1/2}\,dx$, hence our terminology. Exponentiating $\mathcal{H}$ yields the unitary scaling operator $\left(e^{j2\pi\sigma\mathcal{H}}s\right)(x)=e^{\sigma/2}s(e^{\sigma}x)$.

Using $\mathcal{T}$ and $\mathcal{H}$, we now construct two bilinear characteristic functions that are *not* equivalent modulo a multiplicative kernel. Form the characteristic function $\mathbf{N}s$ using the normal correspondence $e^{j2\pi\theta\mathcal{T}}\,e^{j2\pi\sigma\mathcal{H}}$

$$\begin{aligned}(\mathbf{N}s)(\theta,\sigma)&\equiv\left\langle s\middle|\,e^{j2\pi\theta\mathcal{T}}\,e^{j2\pi\sigma\mathcal{H}}\middle|s\right\rangle\\&=e^{\sigma/2}\int s^*(x)\,s(e^{\sigma}x)\,e^{j2\pi\theta x}\,dx.\end{aligned}$$

Form the characteristic function $\mathbf{A}s$ using the antinormal correspondence $e^{j2\pi\sigma\mathcal{H}}\,e^{j2\pi\theta\mathcal{T}}$

$$\begin{aligned}(\mathbf{A}s)(\theta,\sigma)&\equiv\left\langle s\middle|e^{j2\pi\sigma\mathcal{H}}e^{j2\pi\theta\mathcal{T}}\middle|s\right\rangle\\&=e^{\sigma/2}\int s^*(x)s(e^{\sigma}x)\,e^{j2\pi\theta e^{\sigma}x}dx.\end{aligned}$$

Taking bidimensional Fourier transforms of $(\mathbf{N}s)(\theta,\sigma)$ and $(\mathbf{A}s)(\theta,\sigma)$ with respect to θ and σ yields corresponding distributions of time t and Mellin variable γ. For the normal correspondence, we have

$$(\mathbf{C}^{\mathbf{N}}s)(t,\gamma)=M(\gamma)\,s^*(t)\,t^{-1/2}e^{j2\pi\gamma\ln t},\quad t>0$$

whereas for the antinormal correspondence, we have

$$(\mathbf{C}^{\mathbf{A}}s)(t,\gamma)=M^*(\gamma)s(t)\;t^{-1/2}\,e^{-j2\pi\gamma\ln t},\quad t>0.$$

A simple computation verifies that both of these bilinear distributions possess the desired time $|s(t)|^2$ and Mellin $|M(\gamma)|^2$ marginals.

The foundation of the kernel method rests on the assumption that $(\mathbf{N}s)(\theta,\sigma)$ and $(\mathbf{A}s)(\theta,\sigma)$ are equivalent up to weighting by a simple signal-independent kernel function

$$(\mathbf{A}s)(\theta,\sigma)=\phi(\theta,\sigma)\,(\mathbf{N}s)(\theta,\sigma)\quad\forall s.$$

Equivalently, the quotient $\mathbf{A}s/\mathbf{N}s$ must be constant and *signal independent*

$$\frac{(\mathbf{A}s)(\theta,\sigma)}{(\mathbf{N}s)(\theta,\sigma)}=\phi(\theta,\sigma)\quad\forall s. \tag{7}$$

Now, consider two rectangular pulse signals: $s_1(x)=1$ for $1\le x\le 2$ and $s_2(x)=\sqrt{2}$ for $1/2\le x\le 1$. (Both signals are zero outside their range of definition.) A straightforward calculation gives

$$(\mathbf{N}s_1)(\theta,\sigma)=\begin{cases}\frac{e^{\sigma/2}}{j2\pi\theta}\left(e^{j4\pi\theta e^{\sigma}}-e^{j2\pi\theta}\right), & -\ln 2\le\sigma\le 0\\ \frac{e^{\sigma/2}}{j2\pi\theta}\left(e^{j4\pi\theta}-e^{j2\pi\theta e^{\sigma}}\right), & 0<\sigma\le\ln 2\\ 0, & \text{otherwise}\end{cases}$$

[2] While Cohen has left open the possibility of signal-dependent kernels from the beginning [1], [2], restricting our scope to bilinear distributions limits us to signal-independent kernels.

$$(\mathbf{A}s_1)(\theta,\sigma)=\begin{cases}\frac{e^{-\sigma/2}}{j2\pi\theta}\left(e^{j4\pi\theta e^{2\sigma}}-e^{j2\pi\theta e^{\sigma}}\right), & -\ln 2\le\sigma\le 0\\ \frac{e^{-\sigma/2}}{j2\pi\theta}\left(e^{j4\pi\theta e^{\sigma}}-e^{j2\pi\theta e^{2\sigma}}\right), & 0<\sigma\le\ln 2\\ 0, & \text{otherwise}\end{cases}$$

and

$$(\mathbf{N}s_1)(\theta,\sigma)=(\mathbf{N}s_2)(2\theta,\sigma),$$
$$(\mathbf{A}s_1)(\theta,\sigma)=(\mathbf{A}s_2)(2\theta,\sigma).$$

These relations form the basis for a contradiction: While (7) requires that $\mathbf{A}s_1/\mathbf{N}s_1=\phi=\mathbf{A}s_2/\mathbf{N}s_2$, we have instead that

$$\frac{(\mathbf{A}s_1)(\theta,\sigma)}{(\mathbf{N}s_1)(\theta,\sigma)}=\frac{(\mathbf{A}s_2)(2\theta,\sigma)}{(\mathbf{N}s_2)(2\theta,\sigma)}\neq\frac{(\mathbf{A}s_2)(\theta,\sigma)}{(\mathbf{N}s_2)(\theta,\sigma)}.$$

Therefore $\mathbf{N}s$ and $\mathbf{A}s$ cannot be related by a single signal-independent kernel function.

The blame for this contradiction lies with the operators $\mathcal{T}$ and $\mathcal{H}$; since they do not commute, the antinormal correspondence operator $e^{j2\pi\sigma\mathcal{H}}e^{j2\pi\theta\mathcal{T}}$ cannot be expressed as a weighted version $\phi(\theta,\sigma)e^{j2\pi\theta\mathcal{T}}e^{j2\pi\sigma\mathcal{H}}$ of the normal correspondence operator. In contrast, since the time and frequency operators almost commute, we have $e^{j2\pi\theta\mathcal{T}}e^{j2\pi\tau\mathcal{F}}=e^{-j2\pi\theta\tau}e^{j2\pi\tau\mathcal{F}}e^{j2\pi\theta\mathcal{T}}$ [2]. Similar counterexamples can be constructed for many operator pairs besides $(\mathcal{T},\mathcal{H})$ (try $(\mathcal{T},\mathcal{F}^2)$, for example).

IV. Discussion and Conclusions

Although it plays a deservedly dominant role in the theory of joint distributions of arbitrary variables, the kernel method does have its limitations. Our counterexample demonstrates that in general, the functional relationship between different correspondence rules cannot be captured by a simple kernel weighting.

Our results indicate that in order to realize the full potential of the characteristic function method, we must either develop a more complete theory for signal-dependent and nonstationary kernels (as foreshadowed by Cohen [1], [2]) or forgo kernels entirely and work directly with the operator representations and correspondences themselves. Given the difficulty of both of these solutions, we may well ask for which operator pairs is this complicated machinery unnecessary; that is, for which operator pairs is the kernel method sufficient?

Unfortunately, the answer turns out to be *not many*. In [12], Sayeed extends the result of this paper by showing that the kernel method is sufficient only for the time-frequency pairing $(\mathcal{T},\mathcal{F})$ and the warped time-frequency pairing $(\mathbf{U}^{-1}\mathcal{T}\mathbf{U},\mathbf{U}^{-1}\mathcal{F}\mathbf{U})$, with $\mathbf{U}$ a unitary transformation [8], [9], [13].

Acknowledgment

The author would like to thank P. Gonçalvès, A. Sayeed, and R. Shenoy for insightful discussions regarding this material.

References

[1] L. Cohen, "Generalized phase space distribution functions," *J. Math. Phys.*, vol. 7, pp. 781–786, 1966.
[2] ———, *Time-Frequency Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1995.
[3] 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.
[4] 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.
[5] R. G. Shenoy and T. W. Parks, "Wide-band ambiguity functions and affine Wigner distributions," *Signal Processing*, vol. 41, no. 3, pp. 339–363, 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] 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.
[8] 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.
[9] R. G. Baraniuk and L. Cohen, "On joint distributions of arbitrary variables," *IEEE Signal Processing Lett.*, vol. 2, pp. 10–12, Jan. 1995.
[10] M. O. Scully and L. Cohen, "Quasi-probability distributions for arbitrary operators," in *Proc. Nonlinear Dynamics Chaos, Geometric Quantization, Wigner Functions*, Y. S. Kim and W. W. Zachary, Eds. Berlin: Springer-Verlag, 1987, pp. 253–263.
[11] R. G. Baraniuk, "A signal transform covariant to scale changes," *IEE Electron. Lett.*, vol. 29, pp. 1675–1676, Sept. 17, 1993.
[12] A. M. Sayeed, "On the equivalence of the operator and kernel methods for joint distributions of arbitrary variables," *IEEE Trans. Signal Processing*, 1995 (preprint).
[13] R. G. Baraniuk, "Marginals vs. covariance in joint distribution theory," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, ICASSP'95*, vol. 2, 1995, pp. 1021–1024.