ON THE EQUIVALENCE OF THE OPERATOR AND KERNEL METHODS FOR JOINT DISTRIBUTIONS OF ARBITRARY VARIABLES

Akbar M. Sayeed*

Coordinated Science Laboratory
University of Illinois at Urbana–Champaign
1308 West Main Street
Urbana, IL 61801

E-mail: akbar@csl.uiuc.edu
Tel: (217) 244-6384
Fax: (217) 244-1642

Submitted to the *IEEE Transactions on Signal Processing*, November 10, 1995
EDICS Number SP 2.3 Time-Frequency Signal Analysis

Abstract

[1] Generalizing the concept of time-frequency representations, Cohen has recently proposed a general method, based on operator correspondence rules, for generating joint distributions of arbitrary variables. As an alternative to considering all such rules, which is a practical impossibility in general, Cohen has proposed the kernel method in which different distributions are generated from a fixed rule via an arbitrary kernel. In this paper, we derive a simple but rather stringent necessary condition, on the underlying operators, for the kernel method (with the kernel functionally independent of the variables) to generate *all* bilinear distributions. Of the specific pairs of variables that have been studied, essentially only time and frequency satisfy the condition; in particular, the important variables of time and scale do not. The results warrant further study for a systematic characterization of bilinear distributions in Cohen's method.

1 Introduction

Time-frequency representations (TFRs), such as the Wigner distribution and the short-time Fourier transform, represent signal characteristics jointly in terms of time and frequency, and are powerful tools for nonstationary signal analysis and processing [1]. However, due to their inherent structure, TFRs can accurately represent only a limited class of nonstationary signal characteristics. Recently, in an attempt to tailor joint signal representations to a broader class of signals, there has been substantial progress in the development of joint distributions of variables other than time and frequency [2]–[7]. Joint time-scale representations constituted first such generalizations [2, 3], spurred by the interest in the wavelet transform [8].

In view of this recent trend, general theories for joint distributions of arbitrary variables have been proposed by many authors [1, 5, 9, 10, 11]. The first such generalization was proposed by Scully and Cohen [12], and developed by Cohen [1, 5], in direct extension of his original method for generating joint TFRs

*This work was supported by the Joint Services Electronics Program under Grant No. N00014-90-J-1270.

[1]Permission to publish this abstract separately is granted.

[13]. Baraniuk proposed a general approach based on group theoretic arguments [9] which was shown by Sayeed and Jones [14] to be equivalent to Scully and Cohen's method. Other covariance-based generalizations have also been proposed [10, 11] which complement Cohen's distributional method by characterizing joint representations in terms of covariance properties. However, Cohen's method seems to be the most general approach to date, since no joint group structure is imposed on the variables as is done in [10, 11].

Fundamental to Cohen's method is the idea of associating variables with Hermitian (self-adjoint) operators [1]. For given variables, the entire class of joint distributions is generated by the infinitely many (in general) operator correspondence rules for an exponential function of the variables (the characteristic function *operator method* [1]). As an alternative to considering all possible correspondence rules, which is a practical impossibility in general, Cohen has proposed the *kernel method* in which a fixed operator correspondence is used and different joint distributions are generated via an arbitrary kernel. In this correspondence, we show that for two variables, say a and b, the corresponding Hermitian operators $\mathcal{A}$ and $\mathcal{B}$ must satisfy rather stringent conditions, such as

$$e^{j\alpha\mathcal{A}}e^{j\beta\mathcal{B}} = e^{jf(\alpha,\beta)}e^{j\beta\mathcal{B}}e^{j\alpha\mathcal{A}}, \quad \text{for all } (\alpha,\beta) \in \mathbb{R}^2 \text{ , and for some } f : \mathbb{R}^2 \to \mathbb{R} \text{ ,} \tag{1}$$

for the kernel method[2] to generate the entire class of bilinear distributions determined by the operator method. We also generalize the result to an arbitrary number of variables, and show that of the specific variables considered in the literature, essentially only time and frequency satisfy the conditions. We begin with a brief description of Cohen's method.

2 Cohen's Method

We describe the method for two variables; extension to more variables will be obvious. We assume that all signals of interest belong to $L^2(\mathbb{R})$, the space of finite energy signals.

The characteristic function M of a joint a-b distribution P of signal s is defined as [1]

$$(Ms)(\alpha,\beta) = \int\int (Ps)(a,b)e^{j2\pi\alpha a}e^{j2\pi\beta b}dadb \tag{2}$$

and the distribution can be recovered from M as

$$(Ps)(a,b) = \int\int (Ms)(\alpha,\beta)e^{-j2\pi\alpha a}e^{-j2\pi\beta b}d\alpha d\beta \ . \tag{3}$$

The key observation is that the characteristic function can be directly computed from the signal by using a characteristic function operator $\mathbf{M}^{(\alpha,\beta)}$, corresponding to the function $e^{j2\pi\alpha a}e^{j2\pi\beta b}$, as

$$(Ms)(\alpha,\beta) = \left\langle \mathbf{M}^{(\alpha,\beta)}s, s \right\rangle \equiv \int (\mathbf{M}^{(\alpha,\beta)}s)(x)s^*(x)dx \ . \tag{4}$$

Since the operators $\mathcal{A}$ and $\mathcal{B}$ do not commute in general, there are infinitely many ways in which the function $e^{j2\pi\alpha a}e^{j2\pi\beta b}$ can be associated with an operator; three prominent examples are $e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$ (Weyl correspondence), $e^{j2\pi\alpha\mathcal{A}}e^{j2\pi\beta\mathcal{B}}$, and $e^{j2\pi\beta\mathcal{B}}e^{j2\pi\alpha\mathcal{A}}$, which we will use throughout the paper. The corresponding infinitely many joint distributions can then be recovered via (3), and they *define* the entire class of joint a-b distributions.

In order to characterize all the different correspondence rules, and hence the entire class of joint a-b distributions, Cohen has proposed the kernel method which assumes that all characteristic functions can be

[2]We restrict the discussion to kernels that are functionally independent of the signal and the variables. See remarks in footnote 3 on the issue of kernel dependence.

generated by weighting any one particular one with an arbitrary kernel [1, p. 229]. That is, given a particular characteristic function, say M_o, *all* the infinitely many characteristic functions can be generated as

$$(M(\phi))(s)(\alpha,\beta) = (M_o s)(\alpha,\beta)\phi(\alpha,\beta) \tag{5}$$

where ϕ is the weighting kernel.[3] The corresponding joint distributions $P(\phi)$ can then be recovered by using (5) in (3). In the case of time-frequency, fixing M_o to be the Weyl correspondence yields the following commonly used characterization of Cohen's class of TFRs first proposed in [13]

$$(C(\phi)s)(t,f) = \int\int\int \phi(\theta,\tau)s(u+\tau/2)s^*(u-\tau/2)e^{j2\pi\theta(u-t)}e^{-j2\pi\tau f}du d\theta d\tau\ . \tag{6}$$

3 Necessary Conditions for the Validity of the Kernel Method

According to Cohen's kernel method, any two characteristic functions, say M_1 and M_2, corresponding to two different operator correspondences, $\mathbf{M}_1^{(\alpha,\beta)}$ and $\mathbf{M}_2^{(\alpha,\beta)}$, must be related by

$$(M_1 s)(\alpha,\beta) \equiv \left\langle \mathbf{M}_1^{(\alpha,\beta)} s, s\right\rangle = \phi(\alpha,\beta)(M_2 s)(\alpha,\beta) \equiv \phi(\alpha,\beta)\left\langle \mathbf{M}_2^{(\alpha,\beta)} s, s\right\rangle\ , \quad \text{for all } s\in L^2(\mathbb{R})\ , \tag{7}$$

for some $\phi : \mathbb{R}^2 \to \mathbb{C}$. It follows that a necessary and sufficient condition for the kernel method to hold is that *any* two operator correspondences must be related by[4]

$$\mathbf{M}_1^{(\alpha,\beta)} = \phi(\alpha,\beta)\mathbf{M}_2^{(\alpha,\beta)} \tag{8}$$

for some ϕ. In particular, the above relationship must hold in the case when both $\mathbf{M}_1^{(\alpha,\beta)}$ and $\mathbf{M}_2^{(\alpha,\beta)}$ are unitary operators,[5] in which case it can be easily verified that $|\phi(\alpha,\beta)| = 1$, for all (α,β). For example, all characteristic function operators of the following form are unitary[6]

$$\mathbf{M}^{(\alpha,\beta)} = \prod_k e^{j2\pi\gamma_k\mathcal{C}_k} \quad \text{where} \tag{9}$$

$$\mathcal{C}_k = \mathcal{A} \text{ or } \mathcal{B}\ ,\ \gamma_k = \begin{cases} \alpha_k & \text{if } \mathcal{C}_k = \mathcal{A} \\ \beta_k & \text{if } \mathcal{C}_k = \mathcal{B} \end{cases}\ , \text{ and } \sum_k \alpha_k = \alpha\ ,\ \sum_k \beta_k = \beta\ . \tag{10}$$

Two specific cases are the correspondences $\mathbf{M}_1^{(\alpha,\beta)} = e^{j2\pi\alpha\mathcal{A}}e^{j2\pi\beta\mathcal{B}}$ and $\mathbf{M}_2^{(\alpha,\beta)} = e^{j2\pi\beta\mathcal{B}}e^{j2\pi\alpha\mathcal{A}}$ which result in the relationship (1). Extension to more than two variables immediately follows, and we have the following general result.

Proposition. Let $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_N$ be the Hermitian operators corresponding to N variables $a_1, a_2, \cdots, a_N$ in Cohen's method. Then, a necessary condition for the kernel method to generate *all* bilinear joint distributions of $a_1, a_2, \cdots, a_N$ is that for *any* two *unitary* characteristic function operator correspondences, $\mathbf{M}_1^{(\alpha_1,\alpha_2,\cdots,\alpha_N)}$ and $\mathbf{M}_2^{(\alpha_1,\alpha_2,\cdots,\alpha_N)}$, the following relationship must hold for all $(\alpha_1,\alpha_2,\cdots,\alpha_N) \in \mathbb{R}^N$:

$$\mathbf{M}_1^{(\alpha_1,\alpha_2,\cdots,\alpha_N)} = e^{jf(\alpha_1,\alpha_2,\cdots,\alpha_N)}\mathbf{M}_2^{(\alpha_1,\alpha_2,\cdots,\alpha_N)}\ , \quad \text{for some } f:\mathbb{R}^N \to \mathbb{R}. \tag{11}$$

[3] Cohen does not preclude the possibility of functional dependence of the kernel on the variables and the signal [1, p. 140]. However, we restrict the discussion to the important case of bilinear distributions which precludes signal-dependent kernels. Moreover, we are interested in a characterization of bilinear distributions in which the kernel is not a function of the variables, as is true for all covariance-based generalizations [10, 11], and for Cohen's class of bilinear TFRs [1] and the affine class of bilinear time-scale representations [3, 2], in particular.

[4] We use the fact that if $\mathbf{A}$ is a linear operator on a complex inner product space $\mathcal{H}$, then $\langle \mathbf{A}s, s\rangle = 0$ for all $s \in \mathcal{H} \Leftrightarrow \mathbf{A} \equiv \mathbf{0}$; see, for example, [15, p. 374].

[5] An operator $\mathbf{U}$ is unitary if $\langle \mathbf{U}s, \mathbf{U}s\rangle = \langle s, s\rangle$ for all s.

[6] Which follows from the fact that $e^{j\mathcal{A}}$ is a unitary operator if $\mathcal{A}$ is Hermitian, and that the composition of unitary operators is unitary.

Corollary. A particular necessary condition for the validity of the kernel method is

$$e^{j\mathcal{A}_1} e^{j\mathcal{A}_2} \cdots e^{j\mathcal{A}_N} = e^{jc} e^{j\mathcal{A}_N} e^{j\mathcal{A}_{N-1}} \cdots e^{j\mathcal{A}_1} \text{ , for some } c \in \mathbb{R} \text{ .} \tag{12}$$

4 Examples

Time and Frequency. Defining the time and frequency Hermitian operators as $(\mathcal{T}s)(t) = ts(t)$ and $(\mathcal{F}s)(t) = -\frac{j}{2\pi}\dot{s}(t)$ [1][7], respectively, we have $(e^{j2\pi\theta\mathcal{T}}s)(t) = e^{j2\pi\theta t}s(t)$ and $(e^{j2\pi\tau\mathcal{F}}s)(t) = s(t+\tau)$ [1]. The following relationships hold between the three main correspondences [1, p. 155]

$$e^{j2\pi(\theta\mathcal{T}+\tau\mathcal{F})} = e^{-j\pi\theta\tau} e^{j2\pi\tau\mathcal{F}} e^{j2\pi\theta\mathcal{T}} = e^{j\pi\theta\tau} e^{j2\pi\theta\mathcal{T}} e^{j2\pi\tau\mathcal{F}} \tag{13}$$

from which it can be easily verified that all the three correspondences satisfy (11) pairwise. In fact, the relationships (13) can be used to show that the necessary and sufficient condition (8) is satisfied for all pairs of orderings, and thus the kernel-based characterization (6) does indeed generate all possible bilinear joint time-frequency distributions.

Time and Scale. Define the operator $\mathcal{C} = \frac{1}{2}(\mathcal{T}\mathcal{F} + \mathcal{F}\mathcal{T})$ which is associated with scale in [1].[8] The corresponding exponential operator is the scaling operator $(e^{j2\pi\sigma\mathcal{C}}s)(t) = e^{\sigma/2}s(e^{\sigma}t)$ [1]. The three main characteristic function operators for $\mathcal{T}$ and $\mathcal{C}$ are related as [1]

$$e^{j2\pi(\theta\mathcal{T}+\sigma\mathcal{C})} = e^{j2\pi\theta\left(\frac{e^{\sigma}-1}{\sigma}-e^{\sigma}\right)\mathcal{T}} e^{j2\pi\sigma\mathcal{C}} e^{j2\pi\theta\mathcal{T}} = e^{j2\pi\theta\left(\frac{e^{\sigma}-\sigma-1}{\sigma}\right)\mathcal{T}} e^{j2\pi\theta\mathcal{T}} e^{j2\pi\sigma\mathcal{C}}, \text{ or} \tag{14}$$

$$e^{j2\pi(\theta\mathcal{T}+\sigma\mathcal{C})} = e^{j2\pi\sigma\mathcal{C}} e^{j2\pi\theta\mathcal{T}} e^{j2\pi\theta\left(\frac{1-e^{-\sigma}}{\sigma}-e^{-2\sigma}\right)\mathcal{T}} = e^{j2\pi\theta\mathcal{T}} e^{j2\pi\sigma\mathcal{C}} e^{j2\pi\theta\left(\frac{(1-e^{-\sigma})(\sigma+1)}{\sigma}-e^{-2\sigma}\right)\mathcal{T}}, \text{ and} \tag{15}$$

$$e^{j2\pi\theta\mathcal{T}} e^{j2\pi\sigma\mathcal{C}} = e^{j2\pi\sigma\mathcal{C}} e^{j2\pi\theta\mathcal{T}} e^{j2\pi\theta(e^{-\sigma}-1)\mathcal{T}} = e^{j2\pi\theta(1-e^{\sigma})\mathcal{T}} e^{j2\pi\sigma\mathcal{C}} e^{j2\pi\theta\mathcal{T}} \text{ .} \tag{16}$$

We note that none of the unitary characteristic function operators is simply a scalar multiple of the others for arbitrary values of the parameters; instead of a weighting function, an *operator* relates pairs of correspondences. Thus, the condition of the Proposition (and in particular (1)) is violated and hence the kernel method does not generate all joint $\mathcal{T}$-$\mathcal{C}$ distributions. Indeed, a specific counter example is constructed in [16] to show that the characteristic functions corresponding to the two correspondences in (16) are not related by a weighting kernel.

Similarly, it can be easily verified by using the Corollary to the Proposition that the joint frequency-scale and time-frequency-scale distributions discussed in [1, p. 258–259] are not completely characterized by the kernel method.

5 Discussion

The necessary condition stated in the Proposition is rather stringent. To appreciate this, we use the Baker-Campbell-Hausdorff formula which can be stated, to third order, as [17]

$$e^{j\mathcal{A}} e^{j\mathcal{B}} = e^{j(\mathcal{A}+\mathcal{B}+\frac{j}{2}\mathcal{D}-\frac{1}{12}[\mathcal{A},\mathcal{D}]-\frac{1}{12}[\mathcal{D},\mathcal{B}]\cdots)} \tag{17}$$

where $\mathcal{D} = [\mathcal{A},\mathcal{B}] \equiv \mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A}$ is the commutator operator. One of the simplest nontrivial special cases is when the commutator commutes with both the operators, in which case we have the following relationships

[7] Cohen uses the radian frequency operator $(\mathcal{W}s)(t) = -j\dot{s}(t)$.

[8] A different correspondence for scale is argued in [7, 19]. However, even for that correspondence, the kernel method does not hold for time and scale.

[1, 17]

$$e^{j\mathcal{A}}e^{j\mathcal{B}} = e^{j(\mathcal{A}+\mathcal{B})}e^{-\frac{1}{2}\mathcal{D}} = e^{-\frac{1}{2}\mathcal{D}}e^{j(\mathcal{A}+\mathcal{B})} = e^{j\mathcal{B}}e^{j\mathcal{A}}e^{-\mathcal{D}} = e^{-\mathcal{D}}e^{j\mathcal{B}}e^{j\mathcal{A}} \; . \tag{18}$$

Even in this case, unless $\mathcal{D} = jc\mathbf{I}$, $c \in \mathbb{R}$, the conditions of the Proposition are violated and the kernel method does not hold. Note that for time and frequency, $\mathcal{D} = \frac{j}{2\pi}\mathbf{I}$, and (18) yields (13), making the kernel method work. Moreover, since this commutator relationship does not change for operators which are unitarily equivalent to time and frequency [18, 19], joint distributions of variables which are unitarily equivalent to time and frequency [7, 18, 19] are also completely characterized by the kernel method.

Another special case studied in [1] regarding joint distributions involving scale is

$$\mathcal{D} = [\mathcal{A}, \mathcal{B}] = c_1\mathbf{I} + c_2\mathcal{A} \tag{19}$$

which results in the relationship [1, p. 228]

$$e^{j\alpha\mathcal{A}+j\beta\mathcal{B}} = e^{j\mu\alpha c_1/c_2}e^{j\alpha\mu\mathcal{A}}e^{j\beta\mathcal{B}}e^{j\alpha\mathcal{A}} \tag{20}$$

where $c_1, c_2 \in \mathbb{C}$, and $\mu = \frac{1}{j\beta c_2}\left[1 - (1 + j\beta c_2)e^{-j\beta c_2})\right]$. Again, (20) implies that the condition (11) is violated for the correspondences $\mathbf{M}_1^{(\alpha,\beta)} = e^{j\alpha\mathcal{A}+j\beta\mathcal{B}}$ and $\mathbf{M}_2^{(\alpha,\beta)} = e^{j\beta\mathcal{B}}e^{j\alpha\mathcal{A}}$, and as a specific example of this case we showed in the last section that the kernel method does not hold for joint $\mathcal{T}$-$\mathcal{C}$ distributions.

6 Conclusions

Cohen's general method for generating distributions of arbitrary variables, when viewed from the perspective of operator correspondences, is a powerful and versatile tool. However, characterizing all the different operator correspondences is nontrivial, and the simple kernel method proposed by Cohen does not encompass all possible correspondences in general. In fact, the necessary conditions derived in this paper for the validity of the kernel method are rather stringent and, in the case of two variables, seem to hold only for time and frequency and variables that are unitarily equivalent to time and frequency.

Thus, in general, applying the kernel method to a particular correspondence rule generates a proper subset of the entire class of joint distributions. However, it is conceivable that the families of joint distributions generated by a finite set of correspondence rules, via the kernel method, may cover the entire class of joint distributions.

It is worth noting that covariance-based generalizations of joint distributions [10, 11], which necessarily impose a joint group structure on the variables, naturally yield a kernel method which generates all the joint distributions in the class. Thus, it might be fruitful to study the relationship between the two approaches for arbitrary joint distributions in order to develop a systematic characterization of all the correspondence rules in Cohen's general method.

Acknowledgements

The author would like to thank R. Baraniuk and L. Cohen and for insightful discussions regarding this material.

References

[1] L. Cohen, *Time-Frequency Analysis*. Prentice Hall, 1995.

[2] J. Bertrand and P. Bertrand, "A class of affine Wigner distributions with extended covariance properties," *J. Math. Phys.*, vol. 33, no. 7, pp. 2515–2527, 1992.

[3] O. Rioul and P. Flandrin, "Time-scale distributions: A general class extending the wavelet transform," *IEEE Trans. Signal Processing*, vol. 46, pp. 1746–1757, May 1992.

[4] A. Papandreou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "The hyperbolic class of time-frequency representations part I: Constant-Q warping, the hyperbolic paradigm, properties, and members," *IEEE Trans. Signal Processing*, vol. 41, pp. 3425–3444, December 1993.

[5] L. Cohen, "A general approach for obtaining joint representations in signal analysis and an application to scale," in *Proc. SPIE 1566*, (San Diego), July 1991.

[6] 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.

[7] R. G. Baraniuk and D. L. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. Signal Processing*, October 1995.

[8] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, October 1991.

[9] R. G. Baraniuk, "Beyond time-frequency analysis: energy densities in one and many dimensions," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '94*, 1994.

[10] F. Hlawatsch and H. Bölcskei, "Displacement-covariant time-frequency energy distributions," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '95*, pp. 1025–1028, 1995.

[11] A. M. Sayeed and D. L. Jones, "A canonical covariance-based method for generalized joint signal representations," *To appear in the IEEE Signal Processing Letters.*

[12] M. Scully and L. Cohen, "Quasi-probability distributions for arbitrary operators," in *The Physics of Phase Space*, (Springer Verlag), (Y.S. Kim and W.W. Zachary Eds.), 1987.

[13] L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7, pp. 781–786, 1966.

[14] A. M. Sayeed and D. L. Jones, "On the equivalence of generalized joint signal representations," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '95*, pp. 1533–1536, 1995.

[15] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. Springer-Verlag, 1982.

[16] R. G. Baraniuk, "A limitation of the kernel method for joint distributions of arbitrary variables," *Submitted to IEEE Signal Processing Letters*, 1995.

[17] R. M. Wilcox, "Exponential operators and parameter differentiation in quantum physics," *J. Math. Phys.*, vol. 8, pp. 962–982, 1967.

[18] R. G. Baraniuk and L. Cohen, "On joint distributions of arbitrary variables," *IEEE Signal Processing Letters*, vol. 2, pp. 10–12, January 1995.

[19] A. M. Sayeed and D. L. Jones, "Integral transforms covariant to unitary operators and their implications for joint signal representations," *To appear in the IEEE Trans. Signal Processing.*