MARGINALS VS. COVARIANCE IN JOINT DISTRIBUTION THEORY

Richard G. Baraniuk *

Department of Electrical and Computer Engineering
Rice University
Houston, TX 77251–1892, USA

ABSTRACT

Recently, Cohen has proposed a method for constructing joint distributions of arbitrary physical quantities, in direct generalization of joint time-frequency representations. In this paper, we investigate the covariance properties of this procedure and caution that in its present form it cannot generate all possible distributions. Using group theory, we extend Cohen's construction to a more general form that can be customized to satisfy specific marginal and covariance requirements.

1. INTRODUCTION

Joint distributions of arbitrary variables extend the notion of time-frequency analysis to quantities such as scale, Mellin, chirp rate, and inverse frequency. Two complementary approaches to constructing distributions have been developed. Covariance-based methods [1–4] concentrate on certain canonical signal transformations that leave the form of the distribution unchanged, while marginal-based methods [5, 6] aim for the property that integrating out one variable leaves the valid density of the other. Both approaches have their merits, but also their limitations. Covariance approaches rely on a group structure that may not be present in general, while the generality of marginal methods makes characterizing the resulting distributions difficult. In particular, the difficulties introduced in dealing with the noncommuting operators that represent most physical quantities has lead to a simplified marginal method, the *kernel method* of Cohen and Scully-Cohen [5, 6], that if not used carefully can limit the range of possible joint distributions.

The case of time $|s(t)|^2$ and inverse frequency $|r^{-1}\,S(r^{-1})|^2$ distributions illustrates the essence of our discussion. The kernel method yields a class of distributions [6, p. 239]

$$(\mathbf{Z}s)(t,r) \;=\; \frac{1}{r^2}\iint \Phi(\alpha,\beta)\,A_s(\alpha,\beta)\,e^{-j2\pi(\alpha t+\beta f_0/r)}\,d\alpha\,d\beta.$$

parameterized by a kernel function Φ that is assumed to contain all distributions having time and inverse frequency marginals. (Here A_s is the narrow-band ambiguity function of s.) However, this class does not contain all of them, one notable outcast being [1], [6, p. 240] $(\mathbf{P}s)(t,r) = \frac{1}{r^3}\int S^*(\lambda(u)/r)\,S(\lambda(-u)/r)\,e^{j2\pi ut/r}\mu^2(u)\,du$, with $\lambda(u) = \frac{u}{1-e^{-u}}$ and $\mu^2(u) = \lambda(u)\lambda(-u)$. In contrast to the time and frequency shift covariance forced on distributions of the form $\mathbf{Z}s$, $\mathbf{P}s$ is covariant to time shifts and scale changes. This paper aims to understand the connections between marginals and covariance in the context of this powerful method for generating joint distribution classes.

2. ENERGY DENSITIES

An energy density measures the content of a physical quantity (time, frequency, scale, for example) in a signal s.[1]

2.1. Density of a Single Variable

Hermitian Representations. From quantum mechanics we appropriate the association of physical quantities with Hermitian operators on the vector space $L^2(\mathbb{R})$ of square integrable signals. We use script letters to represent these operators, with $\mathcal{A}$ and $\mathcal{B}$ corresponding to the arbitrary variables a and b. Examples include [6–8]: *Time* $(\mathcal{T}s)(x) = x\,s(x)$; *Frequency* $(\mathcal{F}s)(x) = \frac{1}{2\pi j}\dot{s}(x)$; *Time Scale* (Log Time) $(\mathcal{D}s)(x) = (\log\mathcal{T}s)(x) = \log(x)\,s(x)$, $x>0$; *Frequency Scale* (Log Frequency) $(\mathcal{D}'s)(x) = (\log\mathcal{F}s)(x) = \mathbb{F}^{-1}\log(f)\,S(f)$, $f>0$; *Mellin* $(\mathcal{H}s)(x) = \frac{1}{2}(\mathcal{T}\mathcal{F}+\mathcal{F}\mathcal{T})$; *Fourier-Mellin* $\mathcal{H}' = \mathbb{F}^{-1}\mathcal{H}\mathbb{F}$; and *Power Frequency* (Chirp) $\mathcal{F}^n$, including *Inverse Frequency* $\mathcal{R} = \mathcal{F}^{-1}$. (We use the symbol $\mathbb{F}$ to denote the usual Fourier transform $S(f) = (\mathbb{F}s)(f)$.)

Averages of a physical quantity a can be computed using the operator representation $\mathcal{A}$ through $\langle s|\mathcal{A}|s\rangle = \langle s,\mathcal{A}s\rangle = \int s^*(x)\,(\mathcal{A}s)(x)\,dx$.[2] For example, the average (mean) frequency of a time signal s is given by $\langle s|\mathcal{F}|s\rangle = \int s^*(x)\,\dot{s}(x)\,dx/2\pi j = \int f\,|S(f)|^2\,df$. This generalizes in the obvious way to the average value of a function $g(a)$.

Densities. The square of the expansion onto the eigenfunctions $\mathbf{e}_a^{\mathcal{A}}(x)$ of the Hermitian representation measures the quantity of the associated concept in the signal. Define $(\mathbb{F}_{\mathcal{A}}s)(a) = \left\langle s, \mathbf{e}_a^{\mathcal{A}}\right\rangle$ for the variable a. The transforms measuring the quantities introduced above are [6, 7]:

*This work was supported by the National Science Foundation, grant no. MIP-9457438, and by the State of Texas, grant no. TX-ATP 003604-002. Email address: richb@rice.edu

[1] See [6–8] for detailed discussions on this topic.

[2] We adopt the physicists' notation of inner product linear in second element.

Time $(\mathbb{F}_{\mathcal{T}}s)(t) = s(t)$; *Frequency* $(\mathbb{F}_{\mathcal{F}}s)(f) = S(f)$; *Time Scale* $(\mathbb{F}_{\mathcal{D}}s)(d) = e^{d/2}s(e^d)$; *Frequency Scale* $(\mathbb{F}_{\mathcal{D}'}s)(d) = \mathbb{F}^{-1}e^{q/2}S(e^q)$; *Mellin* $(\mathbb{F}_{\mathcal{H}}s)(h) = \int s(x)\, e^{-j2\pi h \log x} \frac{dx}{\sqrt{x}}$; *Fourier-Mellin* $\mathbb{F}_{\mathcal{H}'} = \mathbb{F}_{\mathcal{H}}\mathbb{F}$ and *Power Frequency* $(\mathbb{F}_{\mathcal{F}^n}s)(r) = r^{1/n}S(r^{1/n})$.

Unitary Representations of physical quantities are obtained by exponentiating Hermitian representations, via the formal Taylor series $e^{j2\pi\alpha\mathcal{A}} = \sum_{n=0}^{\infty} \frac{(j2\pi\alpha\mathcal{A})^n}{n!}$. Note that $|\mathbb{F}_{\mathcal{A}}s|^2$ is *invariant* to the unitary operator $e^{j2\pi\alpha\mathcal{A}}$ — $|\mathbb{F}_{\mathcal{A}}\, e^{j2\pi\alpha\mathcal{A}}\, s|^2 = |\mathbb{F}_{\mathcal{A}}s|^2$ — and thus $\mathbf{e}^{j2\pi\alpha\mathcal{A}}$ cannot correspond directly to the physical quantity a. (In fact, it corresponds to some "orthogonal" concept [7].) To find the unitary operator representing the variable a, we solve the operator equation $\mathbb{F}_{\mathcal{A}}\mathbf{A}_{a_1} = \mathcal{T}^{+}_{a_1}\mathbb{F}_{\mathcal{A}}$, with $(\mathcal{T}^{+}_{k}s)(x) = s(x-k)$ the additive translation operator [8]. The density $\mathbb{F}_{\mathcal{A}}$ is *covariant* to $\mathbf{A}$, and thus a, $\mathcal{A}$, and $\mathbf{A}$ are equivalent representations of the same physical concept. Covariances other than additive can be handled by group methods [8]; covariance via multiplicative translation is defined using $(\mathcal{T}^{\times}_{k}s)(x) = s(x/k)/\sqrt{k}$ in the above. We will use boldface letters to denote unitary operators. Continuing the examples from above, we have for unitary representations: *Time* $(\mathbf{T}_t s)(x) = s(x-t) = (e^{-j2\pi t\mathcal{F}}s)(x)$; *Frequency* $(\mathbf{F}_f s)(x) = e^{j2\pi f x}s(x) = (e^{j2\pi f\mathcal{T}}s)(x)$; *Time Scale* $(\mathbf{D}_d s)(x) = e^{-d/2}s(xe^{-d}) = (e^{j2\pi d\mathcal{H}}s)(x)$; *Frequency Scale* $\mathbf{D}'_d = \mathbf{D}_{-d}$; *Mellin* $(\mathbf{H}_h s)(x) = e^{j2\pi h\log x}s(x) = (e^{j2\pi h\mathcal{D}}s)(x)$; *Fourier-Mellin* $\mathbf{H}'_h = \mathbb{F}^{-1}\mathbf{H}_h\mathbb{F}$; and *Power Frequency* $(\mathbf{R}^n_r s)(x) = \mathbb{F}^{-1}S\left[(f^n - r)^{1/n}\right](f^n - r)^{1/n} f^{1/n}$.

Characteristic Functions provide a direct route to densities circumventing eigenanalysis [5,6]. Given a density $|(\mathbb{F}_{\mathcal{A}}s)(a)|^2$, its inverse Fourier transform defines a characteristic (ambiguity) function $(\mathbf{M}s)(\alpha) = \int |(\mathbb{F}_{\mathcal{A}}s)(a)|^2\, e^{j2\pi\alpha a}\, da$. Since the right side of this expression corresponds to an average of the quantity $g(a) = e^{j2\pi\alpha a}$ with respect to the density of a, it can be computed directly from the signal via (see "Averages" above) $(\mathbf{M}s)(\alpha) = \left\langle s \,\middle|\, e^{j2\pi\alpha\mathcal{A}} \,\middle|\, s\right\rangle$. Combining these two equations yields $|(\mathbb{F}_{\mathcal{A}}s)(a)|^2 = \int \left\langle s \,\middle|\, e^{j2\pi\alpha\mathcal{A}} \,\middle|\, s\right\rangle\, e^{-j2\pi\alpha a}\, d\alpha$.

2.2. Joint Densities Via Characteristic Functions

Joint densities attempt to indicate the simultaneous content of two (or more) physical quantities in a signal. Scully and Cohen, guided by the one-dimensional characteristic function method, derived a formula for the joint density of a and b [6]

$$(\mathbf{C}s)(a,b) = \iint \left\langle s \,\middle|\, e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})} \,\middle|\, s\right\rangle e^{-j2\pi(\alpha a+\beta b)}\, d\alpha\, d\beta. \tag{1}$$

Like a true density, this functional marginalizes to the individual densities of a and b; that is, integration over b yields

$$\int (\mathbf{C}s)(a,b)\, db = \int \left\langle s \,\middle|\, e^{j2\pi\alpha\mathcal{A}} \,\middle|\, s\right\rangle e^{-j2\pi\alpha a}\, d\alpha = |(\mathbb{F}_{\mathcal{A}}s)(a)|^2,$$

with a similar result for integration over a.

Joint distributions are not unique. Since in general $\mathcal{A}$ and $\mathcal{B}$ do not commute ($\mathcal{AB} \neq \mathcal{BA}$), the exponential $e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$ in (1) can be evaluated in many ways, giving different distributions satisfying the same marginals. The three simplest *correspondence rules* are [5, 6]: the symmetric *Weyl correspondence* $e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$, where the sum $\mathcal{A}+\mathcal{B}$ is exponentiated ensemble, the distributed correspondence $e^{j2\pi\frac{\alpha}{2}\mathcal{A}}\, e^{j2\pi\beta\mathcal{B}}\, e^{j2\pi\frac{\alpha}{2}\mathcal{A}}$, where $\mathcal{A}$ and $\mathcal{B}$ are exponentiated separately and then composed, and the similar simpler correspondence $e^{j2\pi\alpha\mathcal{A}}\, e^{j2\pi\beta\mathcal{B}}$. Despite the ordering differences, every correspondence rule yields a distribution marginalizing to $|\mathbb{F}_{\mathcal{A}}s|^2$ and $|\mathbb{F}_{\mathcal{B}}s|^2$.

Since keeping track of all possible correspondence rules is an arduous task, Cohen fixes a single correspondence and then inserts a fixed kernel function[3] $\Phi(\alpha,\beta)$ in (1) to take care of the other possibilities [5, 6]. Constraining $\Phi(\alpha, 0) = \Phi(0, \beta) = 1$ $\forall\, \alpha, \beta$ generates distributions with correct marginals. Leaving Φ unconstrained generalizes the concept of distribution to representations that may not have correct marginals yet in some sense still measure joint a-b energy content.

Example. *Time and Frequency:* Employing the operator pair $\mathcal{T}, \mathcal{F}$ in (1) yields the classical Cohen's class of time-frequency representations having time and frequency marginals [5, 6]. Included among the generalized distributions that do not marginalize are such workhorse representations as the spectrogram and the smoothed pseudo Wigner distribution.

2.3. Extended Covariance

The a–indicating transform $\mathbb{F}_{\mathcal{A}}$ can be covariant to *multiple* operators in addition to $\mathbf{A}$. For instance, we might have $\mathbb{F}_{\mathcal{A}}\mathbf{A}'_{a'} = \mathcal{T}^{G}_{a'}\mathbb{F}_{\mathcal{A}}$ with $\mathbf{A}' \neq \mathbf{A}$ and $\mathcal{T}^{G}$ a unitary group translation different from addition. Thus, in addition to $\mathbf{A}$ content, $\mathbb{F}_{\mathcal{A}}$ also measures $\mathbf{A}'$ content, in a subtlely different way. In certain applications, this *extended covariance* of the transform $\mathbb{F}_{\mathcal{A}}$ can be more important than $\mathbf{A}$ covariance.

Time and frequency provide an excellent example of this behavior, since both the signal $s(t)$ and its Fourier transform $S(f)$ are covariant to scale changes $\mathbf{D}_d$ by multiplicative translation. Thus, $s(t)$ and $S(f)$ can each be interpreted as in some sense measuring the "scale content" of the signal in the time and frequency domains [9].[4] The scale covariance of the power and inverse frequency transform $\mathbb{F}_{\mathcal{F}^n}$ proves far more useful than its complicated additive covariance to $\mathbf{R}^n_r$.

3. COVARIANCE OF DISTRIBUTIONS

In many applications, the covariance of a distribution rivals the marginals in importance. A joint distribution $\mathbf{C}s$ is covariant to a unitary operator $\mathbf{X}_x$ if $(\mathbf{C}\mathbf{X}_x s)(a,b) = (\mathbf{C}s)(a', b')$, where a', b' are functions of a, b, and x, preferably a group law.

[3] We consider only signal-independent kernels in this paper.

[4] *Scale: Covariance and Confusion.* Since the Fourier transform $\mathbb{F}_{\mathcal{F}}$ is invariant to time shifts $\mathbf{T}_t$ in a signal, we know that it does not measure time content. Similarly, since the Mellin transform $\mathbb{F}_{\mathcal{H}}$ is invariant to scale changes $\mathbf{D}_d$, it does not and *cannot* measure the scale content of signals, contrary to popular belief. The density that *is* covariant to $\mathbf{D}_d$ — and thus measures scale content — was derived in [9] as precisely the transform $|(\mathbb{F}_{\mathcal{D}}s)(x) = e^x\, |s(e^x)|^2$ indicating the amount of signal stretching required to move the point x to the reference point $x_0 = 1$.

With joint distributions, it seems reasonable to expect covariance to two operators $\mathbf{X}$ and $\mathbf{Y}$, one relating to each variable. Since $\mathbf{C}s$ simply "shifts" in response to $s \mapsto \mathbf{X}_x \mathbf{Y}_y s$, it can be interpreted as somehow measuring the $\mathbf{X}$-$\mathbf{Y}$ content of signals in addition to the a-b content. Options for the pair $\mathbf{X}, \mathbf{Y}$ include various combinations of $e^{j2\pi\alpha\mathcal{A}}$, $e^{j2\pi\beta\mathcal{B}}$, $\mathbf{A}$, $\mathbf{B}$, and any extended covariance operators for $\mathbb{F}_\mathcal{A}$ and $\mathbb{F}_\mathcal{B}$. Cohen's class of time-frequency distributions provides a fine example of covariance; since $(\mathbf{C}\mathbf{T}_x\mathbf{F}_y s)(t,f) = (\mathbf{C}s)(t-x, f-y)$, changes in the time-frequency origin do not affect the properties of the distributions.

Covariance is deemphasized in the characteristic function method of Scully and Cohen (Section 2.2.). While (1) can construct all distributions with correct marginals irrespective of the choice of variables, the covariance properties of the results are neither predicted nor ensured. Furthermore, recall from the Introduction that there exist variables a, b (time and inverse frequency, for example) for which a simple weighting of one fixed correspondence rule with a kernel $\Phi(\alpha,\beta)$ cannot mimic all possible rules. In order to explain this apparent inconsistency, we now undertake a study of the relationship between the marginal and covariance properties of the characteristic function method.

The Hermitian operators $\mathcal{A}, \mathcal{B}$ that determine the marginal properties of the distributions constructed by the characteristic function method also control the covariance properties, through the unitary operator $\mathbf{G}(\alpha,\beta) = e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$ that anchors the characteristic function in (1) [3, 8]. Covariance depends almost entirely on whether this operator is a representation of some 2-d group [10]; that is, on whether $\mathbf{G}(\alpha_1,\beta_1)\,\mathbf{G}(\alpha_2,\beta_2) = \mathbf{G}[(\alpha_1,\beta_1) \bullet_G (\alpha_2,\beta_2)]$ with group law $\bullet_G$. Equivalently, covariance demands that the sum $\mathcal{A}+\mathcal{B}$ generate a 2-d group representation.[5]

Examples. The CBH expansions (see footnote 5 below) of the pairs $\mathcal{T}+\mathcal{F}$, $\log\mathcal{T}+\mathcal{H}$, and $\log\mathcal{F}+\mathcal{H}'$ all terminate after three distinct terms, with the third a simple phase factor. The group action, $(\alpha_1,\beta_1)\bullet(\alpha_2,\beta_2) = (\alpha_1+\alpha_2, \beta_1+\beta_2)$ neglecting the phase, is that of the Weyl-Heisenberg group. The pairs $\mathcal{T}+\mathcal{H}$ and $\mathcal{F}+\mathcal{H}'$ each generate the "$ax+b$" affine group, with action $(\alpha_1, e^{\beta_1}) \bullet (\alpha_2, e^{\beta_2}) = (\alpha_1 + e^{\beta_1}\alpha_2, e^{\beta_1}\, e^{\beta_2})$. The CBH expansion of the pair $\mathcal{T}+\mathcal{F}^2$ terminates after three distinct terms in $\mathcal{T}$, $\mathcal{F}^2$, and $\mathcal{F}$. Thus, while these three operators generate a 3-d group, $\mathcal{T}+\mathcal{F}^2$ does not generate a 2-d group on its own. Similarly, the expansion of $\mathcal{T}+\mathcal{R}$ comprises terms in $\mathcal{T}$, $\mathcal{R}$, $\mathcal{F}^{-2}$, and $\mathcal{F}^{-3}$. The expansion of $\log\mathcal{T}+\mathcal{F}$ becomes a complicated mess after three terms.

[5] A useful test of whether $\mathcal{A}+\mathcal{B}$ generates a group entails expanding $\mathbf{G}(\alpha,\beta)$ using the Campbell-Baker-Hausdorff (CBH) Theorem [10], which states that

$$e^{\mathcal{A}+\mathcal{B}} = e^{\mathcal{A}}\, e^{\mathcal{B}}\, e^{\frac{1}{2}[\mathcal{A},\mathcal{B}]}\, e^{\frac{1}{12}[\mathcal{A},[\mathcal{A},\mathcal{B}]]}\, e^{\frac{-1}{12}[\mathcal{B},[\mathcal{A},\mathcal{B}]]}\, e^{\frac{1}{48}[\mathcal{B},[\mathcal{A},[\mathcal{A},\mathcal{B}]]]} \ldots$$

If this expansion terminates after N distinct terms, then $\mathcal{A}+\mathcal{B}$ generates a subset of an N-parameter group.

4. COVARIANCE VS. MARGINALS

We are now in a position to classify the covariance properties of the characteristic function method. Our discussion partitions naturally into three cases, depending on what type of group structure the operator representations $\mathcal{A}$ and $\mathcal{B}$ can support. In addition to providing new insight into Scully and Cohen's construction, our analysis indicates its limitations and suggests modifications and extensions.

4.1. Best Case: $\mathcal{A}+\mathcal{B}$ Generates a 2-d Group

In this, the best of all worlds, the characteristic function method coincides with the group-based approach of Shenoy and Parks [3] and succeeds both on the marginal and covariance fronts. All distributions constructed from the various correspondences boast covariance to the operator $\mathbf{G}$ appearing in the characteristic function. The covariance is of the form $(\mathbf{C}\mathbf{G}(u,v)s)(a,b) = (\mathbf{C}s)\big[(a,b) \bullet_{\widehat{G}} (u,v)\big]$, with $\bullet_{\widehat{G}}$ the coadjoint group action [3]. It is interesting to note that the $\mathbf{C}s$ are covariant to the operators $e^{j2\pi\alpha\mathcal{A}}$ and $e^{j2\pi\beta\mathcal{B}}$ to which the marginal distributions $|\mathbb{F}_\mathcal{A} s|^2$ and $|\mathbb{F}_\mathcal{B} a|^2$ are *invariant*, rather than covariant to $\mathbf{A}$ and $\mathbf{B}$. In fact, $\mathbf{A}_a\mathbf{B}_b$ is not even a valid group representation in general,[6] which has some interesting ramifications for operator pairs not unitarily equivalent to $\mathcal{T}$ and $\mathcal{F}$ [7]. The group structure of $\mathbf{G}$ elicits an extension of the characteristic function method that will generate not only all distributions with correct marginals but also all distributions covariant to the signal transformation $s \mapsto \mathbf{G}s$:

Step 1. The symmetrical Weyl correspondence rule $e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$ yields a unitary distribution [3]. (Note that some correspondences yield nonunitary distributions, so this choice is nonarbitrary.) To evaluate $e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$ in terms of $e^{j2\pi\alpha\mathcal{A}}$ and $e^{j2\pi\beta\mathcal{B}}$, either employ the group symmetrization procedure of Kirillov [3], solve an eigenequation [6], or use the CBH Theorem. Because of the strong parallel to the time-frequency case, we will call this special distribution the *a-b Wigner distribution* $(\mathbf{W}s)(a,b)$.

Step 2. To generate a class of distributions covariant to $\mathbf{G}$, "smooth" the Wigner distribution. The coadjoint group law supports two types of smoothed distributions. The first,

$$(\mathbf{C}_1' s)(a,b) \;=\; \iint (\mathbf{W}s)\big[(a,b) \bullet_{\widehat{G}} (u,v)\big]\, \phi_1(u,v)\, d(u,v),$$

with $\phi_1 \in L^1(\mathbb{R}^2)$ and $d(u,v)$ the invariant measure, remains in the a-b domain, while the second,

$$(\mathbf{C}_2' s)(u,v) \;=\; \iint (\mathbf{W}s)\big[(a,b) \bullet_{\widehat{G}} (u,v)\big]\, \phi_2(a,b)\, d(a,b),$$

lives in the domain of the parameters α, β of the covariance operators $e^{j2\pi\alpha\mathcal{A}}$ and $e^{j2\pi\beta\mathcal{B}}$. With $\mathbf{C}_2' s$, the marginal distributions are obtained by integrating along curves in the (u,v) plane. Since $\mathbf{W}$ is a unitary map, these constructions reach

[6] For example, $\mathcal{T}+\mathcal{H}$ generates a representation $e^{j2\pi f\mathcal{T}}\, e^{j2\pi d\mathcal{H}} = \mathbf{F}_f\mathbf{D}_d$ of the affine group, but the operators $\mathbf{T}_t$ and $\mathbf{H}_h$ to which the $\mathbb{F}_\mathcal{T}$ and $\mathbb{F}_\mathcal{H}$ marginals are covariant do not represent any group.

all distributions having $\mathbf{G}$ covariance, including all those with $\mathbb{F}_{\mathcal{A}}$ and $\mathbb{F}_{\mathcal{B}}$ marginals.

Alternative Step 2. In many cases, the group convolutions of Step 2 can be rewritten using the 2-d Fourier transform in terms of a kernel $\Phi(\alpha,\beta;\, a, b)$ weighting the characteristic function in (1).[7] In this case, the functional form of $\Phi(\alpha,\beta;\, a, b)$ can be determined directly by imposing the requisite covariance constraints on this generalized version of (1).

Analogous to the spectrogram in the time-frequency case, this class of distributions contains the squared magnitude of a linear "group transform" $(\mathbf{L}a)(u,v) = \langle s\,|\,\mathbf{G}(u,v)\,|\,g\rangle$ that corresponds to $(\mathbf{C}_2's)(u,v)$ from above with $\phi_2 = \mathbf{W}g$.

Examples. *Time and Frequency*: $\mathcal{T}+\mathcal{F}$ generates the Weyl-Heisenberg group, the usual Wigner distribution, and the classical Cohen's class of time and frequency covariant distributions. In this case, the two smoothing methods coincide, with $\mathbf{T}$ and $\mathbf{F}$ serving simultaneously as invariant *and* covariant transformations for the marginals. This follows because time and frequency are in a sense orthogonal concepts [7]. Kernels of the form $\Phi(\alpha,\beta)$ can reach all distributions in this class.

Frequency Scale and Fourier-Mellin: $\log\mathcal{F}+\mathcal{H}'$ generates the Weyl-Heisenberg group, the Altes Q distribution, and the prehyperbolic class [4, 7] of scale and hyperbolic time-shift ($\mathbf{H}'$) covariant distributions. Kernels of the form $\Phi(\alpha,\beta)$ can reach all distributions in this class.

Frequency and Fourier-Mellin: $\mathcal{F}+\mathcal{H}'$ generates the affine group, the Bertrand distribution [1], and two classes of time and scale covariant distributions. Distributions $\mathbf{C}_1's$ have frequency and Fourier-Mellin coordinates. Distributions $\mathbf{C}_2's$ have time and scale coordinates, and hence inhabit the "affine class" defined in [1, 2]. Both parameterizations require kernels more general than $\Phi(\alpha,\beta)$; for example, Φ_2 must be of the form $\Phi_2(\alpha e^{-d}, \beta e^{d})$, with d the scale variable.

4.2. Worst Case: $\mathcal{A}+\mathcal{B}$ Does Not Generate a Group

Unfortunately, $\mathcal{A}$ and $\mathcal{B}$ must be very well matched in order to generate a 2-d group, as only *two* 2-parameter groups appear in nature: $\mathbb{R}^2$ (essentially the Weyl-Heisenberg group) and the affine group. Thus, group theoretic approaches break down in the general case for arbitrary variables, meaning that the characteristic function has no guaranteed covariance properties. In most cases, the best we can do is either keep track of all correspondence rules or introduce a generalized kernel function $\Phi(\alpha,\beta;a,b)$ in (1) and hope for the best.

Examples. Since the pair $\mathcal{T}+\mathcal{F}^2$ generates a nongroup subset of a three-parameter group requiring $\mathcal{F}$, we cannot expect covariance from time and chirp distributions without also including frequency shift as a (third) variable. The pair $\log\mathcal{T}+\mathcal{F}$ likewise does not appear to support any useful covariances.

[7] In general the kernel will *not* take the form $\Phi(\alpha,\beta)$. While Cohen anticipated the value of a-b-dependent kernels in even his original paper [5], such kernels have not been identified as critical for some choices of A, B until now.

4.3. Intermediate Case: $\mathcal{A}+\mathcal{B}$ Salvageable

In some cases, we can rescue variables that do not generate groups by dividing their associated distributions into subclasses that obey certain alternative covariances. Although the characteristic function operator $\mathbf{G}$ is not a group representation in general, it is potentially compatible with another unitary operator pair $\mathbf{X}_x\mathbf{Y}_y$ (there might exist several) that simultaneously: (1) represents a 2-d group, and (2) is physically reasonable in that it acts as an invariance, covariance or extended covariance operator for the marginal distributions. In this case, given a unitary distribution having both the marginals and this alternative covariance, smoothing via the group convolution induced by $\mathbf{X}_x\mathbf{Y}_y$ will generate a useful class of distributions. Each such class will contain *all* distributions covariant to $\mathbf{X}_x\mathbf{Y}_y$ as well as *some* of the distributions having a, b marginals. Different operator pairs will generate different, potentially nonoverlapping, distribution classes. A procedure for generating a covariance-based subclass of distributions runs as follows:

Step 1. Construct an a-b Wigner distribution using the Weyl or symmetrical correspondence. The eigenanalysis method of Cohen [6] can be applied to reduce the characteristic function operator $\mathbf{G}$ into manageable components. The resulting distribution is unitary and has correct marginals.

Step 2. Find a 2-d unitary group representation $\mathbf{X}_x\mathbf{Y}_y$ such that $\mathbf{X}$ and $\mathbf{Y}$ transform the marginals in meaningful ways (this is where the art comes in). Verify that the Wigner distribution shares these covariances. (If not, recompute using another correspondence, being sure to verify unitarity.) Smooth $\mathbf{W}s$ over the corresponding group as in Section 4.1. In certain cases, the smoothing operation can be implemented in the characteristic function domain by inserting a kernel $\Phi(\alpha,\beta;a,b)$ in (1).

The resulting class contains distributions covariant to $\mathbf{XY}$, *some* of which will have the correct $\mathbb{F}_{\mathcal{A}}$ and $\mathbb{F}_{\mathcal{B}}$ marginals.

Examples. *Time and Inverse Frequency* illustrate the trade-off of marginals for covariance. Since the operator pair $\mathcal{T}+\mathcal{R}$ does not generate a 2-d group, the covariance of t-r distributions rests on shaky ground. However, in addition to representing the affine group, the time-scale operator $\mathbf{T}_t\mathbf{D}_d$ performs a natural transformation in (t, r) coordinates, with $\mathbb{F}_{\mathcal{T}}$ covariant to $\mathbf{T}$ and $\mathbf{D}$ and $\mathbb{F}_{\mathcal{R}}$ invariant to $\mathbf{T}$ and covariant to $\mathbf{D}$. Smoothing over the affine group induces the affine class [1, 2] centered about the distribution $\mathbf{P}s$ from the Introduction. This class also contains the popular scalogram [2], the squared magnitude of the continuous wavelet transform $(\mathbf{L}_1 s)(t,r) = \left|\int s(x)\, h^*\left(\frac{x-t}{r}\right) dx\right|^2$.

The time-frequency operator $\mathbf{T}_t\mathbf{F}_f$ performs similar duties for the operator pair $\mathcal{T}+\mathcal{R}$. Smoothing over the Weyl-Heisenberg group induces the reparameterized Cohen's class $\mathbf{Z}s$ from the Introduction. Included in this class is the remapped spectrogram $(\mathbf{L}_2 s)(t,r) = \left|\int s(x)\, g^*(x-t)\, e^{-j2\pi f_0/r}\, dx\right|^2$.

5. CONCLUSIONS

By ranging through every possible correspondence rule representing $\mathbf{G}(\alpha, \beta) = e^{j2\pi(\alpha\mathcal{A}+\beta\mathcal{B})}$ in (1), the characteristic function method of Scully and Cohen can generate every possible distribution having a and b marginals. While certainly general enough, this approach lacks practicality, however, because studying distribution classes through their correspondence rules would be exasperating. Furthermore, this approach can construct *only* distributions satisfying the marginals, leaving external to the theory most of the useful distributions (spectrogram, scalogram, pseudo Wigner distribution, etc.) employed in practice. Cohen's elegant kernel method takes care of many of these practical issues, but since it introduces limitations of its own, it must be approached with caution.

The three-way classification of the operators $\mathcal{A}$ and $\mathcal{B}$ representing the variables has significant ramifications for joint distribution design. When $\mathcal{A}+\mathcal{B}$ generates a 2-d group, then a generalized version of the kernel method will succeed and generate a single unique class containing marginalizing and covariant distributions. Since there are only two 2-d groups in nature, most of the work for this case has already been completed: Cohen's class represents the Weyl-Heisenberg group [5] and the affine class represents the affine group [1–3]. Simple coordinate changes (unitary equivalence) can generate all other distribution classes resembling these two [7].

When $\mathcal{A}+\mathcal{B}$ fails to generate a group, the kernel method does not appear foolproof. The approach of imposing order on the chaos of correspondences through alternative "pseudo-covariances" appears promising, because it has generated the affine class, an important representation class that has remained outside the marginal-based theory until this time.[8]

REFERENCES

[1] J. Bertrand, P. Bertrand, "A class of affine Wigner functions," *J. Math. Phys.*, vol. 33, pp. 2515–2527, July 1992.

[2] O. Rioul, P. Flandrin, "Time-scale energy distributions," *IEEE Trans. Signal Proc.*, vol. 40, pp. 1746–1757, July 1992.

[3] R. G. Shenoy, T. W. Parks, "Wide-band ambiguity and affine Wigner distributions," *Signal Proc.*, vol. 41, no. 3, 1995.

[4] A. Papandreou, et al, "The hyperbolic class of quadratic time-frequency representations," *IEEE Trans. Signal Proc.*, vol. 41, pp. 3425–3444, Dec. 1993.

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

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

[7] R. Baraniuk, D. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. Signal Proc.*, 1993. Preprint.

[8] R. Baraniuk, "Beyond time-frequency analysis: Energy densities in one and many dimensions," in *Proc. ICASSP '94*, vol. III, pp. 357–360, 1994.

[9] R. Baraniuk, "A signal transform covariant to scale changes," *Electronics Letters*, vol. 29, pp. 1675–1676, Sept. 17, 1993.

[10] L. Corwin, F. Greenleaf, *Representations of Nilpotent Lie Groups and Their Applications, Part I*, 1990.

[8] Thanks to J. Bertrand, L. Cohen, P. Gonçalvès, A. Sayeed, R. Shenoy, and R. Wells for insightful discussions regarding this material.