OPTIMAL QUADRATIC DETECTION AND ESTIMATION USING GENERALIZED JOINT SIGNAL REPRESENTATIONS

Akbar M. Sayeed and Douglas L. Jones[*]

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

E-mail: akbar@csl.uiuc.edu
jones@csl.uiuc.edu

Tel: (217) 244-6384
Fax: (217) 244-1642

To appear in the *IEEE Transactions on Signal Processing*

Abstract

Time-frequency analysis has recently undergone significant advances in two main directions: statistically optimized methods that extend the scope of time-frequency-based techniques from merely exploratory data analysis to more quantitative application, and generalized joint signal representations that extend time-frequency-based methods to a richer class of nonstationary signals. This paper fuses the two advances by developing statistically optimal detection and estimation techniques based on generalized joint signal representations. By generalizing the statistical methods developed for time-frequency representations to arbitrary joint signal representations, this paper develops a unified theory applicable to a wide variety of problems in nonstationary statistical signal processing.

[*]This work was supported by the National Science Foundation under Grant No. MIP 90-12747, the Joint Services Electronics Program under Grant No. N00014-90-J-1270, and the Schlumberger Foundation.

1 Introduction

Many signal processing applications involve signals whose characteristics, deterministic or stochastic, change with time. Examples include radar, speech, communications, machine fault diagnostics, and geophysical and biomedical signal processing. Classical signal processing methods, based on stationary spectral analysis, are usually inadequate in such nonstationary scenarios.

Joint time-frequency representations (TFRs) provide a versatile set of nonparametric tools for the analysis and processing of nonstationary signals [1]. TFRs extend the fundamental concept of spectrum to nonstationary signals and facilitate a time-varying spectral analysis by representing signal characteristics jointly in terms of time and frequency. Although time-frequency-based techniques have shown a lot of promise in nonstationary signal processing, the use and development of new TFRs has been primarily geared towards exploratory data analysis: Using TFRs to provide a visual display of the time-varying spectral energy in the signal and then using this qualitative information as a starting point in further analysis/processing. Moreover, most techniques have been developed for deterministic, noise-free signal analysis. Most applications, on the other hand, involve noisy or random signals and often require detection, estimation or classification of certain nonstationary signal characteristics.

Recent work in statistically optimized time-frequency analysis has developed a promising theory for new techniques that go beyond exploratory data analysis and, for the first time, enable TFRs to be fully exploited in real applications. These techniques include optimal nonstationary spectral estimation using TFRs [2, 3, 4, 5] and optimal quadratic detection using bilinear TFRs and time-scale representations (TSRs) [6, 7, 8]. Such statistical time-frequency-based techniques, by taking into account the effect of random signals, noise and interference, have the potential of being successfully applied to real problems involving nonstationary signals. Some preliminary applications in machine fault diagnostics and biomedical engineering have been very encouraging [9, 10].

Another recent advance in time-frequency analysis has been the development of a theory of generalized joint signal representations (JSRs) that extend the scope of TFRs to a broader class of nonstationary signals [1], [11]–[19]. Generalized JSRs represent nonstationary signal characteristics in terms of quantities other than time and frequency (time and scale, for example) and have the ability to be "matched" to signals with radically different characteristics. For example, TSRs [20] such as the wavelet transform and the affine class perform a "constant-Q"[1] analysis, and the hyperbolic class [21] is useful for "Doppler-invariant" analysis. Despite the proliferation of generalized JSRs, their development has again been primarily directed towards exploratory data analysis. However, to exploit generalized JSRs fully, statistical techniques need to be developed for their successful use in real applications.

In this paper, we fuse the two recent advances by developing statistically optimal detection and estimation techniques based on generalized JSRs. In particular, we extend the minimum mean-squared-error (mmse) estimation framework of [3] and the quadratic detection framework of [6, 8] to any arbitrary class of generalized bilinear JSRs. The result is a unified theory of quadratic detection and estimation techniques based on generalized JSRs that can be applied in a wide variety of nonstationary statistical signal processing scenarios. Not surprisingly, the optimal JSR estimators, being quadratic in the signal, depend on certain fourth-order statistics; the optimal JSR-based detectors, on the other hand, are characterized by second-order statistics. We first provide a relevant description of the theory of generalized JSRs, and then develop the estimation theory followed by the detection framework.

[1] The analysis-bandwidth is proportional to the center frequency.

Notation. All integrals are from $-\infty$ to ∞ unless explicitly labeled. All operators will be denoted by bold-faced letters. The only exception is the symbol ϕ that could represent either a kernel in Cohen's method or an operator in the covariance-based method.

2 A Brief Review of Generalized Theory

Throughout this section we assume that the signals of interest belong to $L^2(\mathbb{R})$, the Hilbert space of finite-energy signals, equipped with the usual inner product[2]

$$\langle x, y\rangle \equiv \int x(t)y^*(t)dt \ . \tag{1}$$

Joint (bilinear) TFRs analyze signals in terms of time and frequency content and ideally should provide a measure of the time-varying spectral energy in the signal. A widely-used class of TFRs is Cohen's class [1] which can be characterized as

$$(\mathbf{C}s)(t,f;\phi) = \int\int\int \phi(\theta,\tau)s\left(u+\frac{\tau}{2}\right)s^*\left(u-\frac{\tau}{2}\right)e^{-j2\pi\theta(u-t)}e^{-j2\pi\tau f}dud\theta d\tau \ , \tag{2}$$

where s is the analyzed signal, and the two-dimensional (2d) kernel ϕ completely characterizes a particular TFR represented by the operator $\mathbf{C} : L^2(\mathbb{R}) \to L^2(\mathbb{R}^2)$. If the kernel satisfies $\phi(\theta,0) = \phi(0,\tau) = 1$ for all θ,τ, the TFR satisfies the marginal relations [1]

$$\int (\mathbf{C}s)(t,f;\phi)dt = |(\mathbf{IF}s)(f)|^2 \quad , \quad \int (\mathbf{C}s)(t,f;\phi)df = |s(t)|^2 \tag{3}$$

where $\mathbf{IF}$ is the Fourier transform operator defined as

$$(\mathbf{IF}s)(f) \equiv \int s(t)e^{-j2\pi ft}dt \ . \tag{4}$$

Another property of Cohen's class is that the TFRs are covariant to time and frequency shifts; that is, if we define the time and frequency shift operators as $(\mathbf{T}_\mu s)(t) \equiv s(t-\mu)$ and $(\mathbf{F}_\nu s) \equiv e^{j2\pi\nu t}s(t)$, respectively, then

$$(\mathbf{C}\mathbf{T}_\mu\mathbf{F}_\nu s)(t,f;\phi) = (\mathbf{C}\mathbf{F}_\nu\mathbf{T}_\mu s)(t,f;\phi) = (\mathbf{C}s)(t-\mu, f-\nu) \ . \tag{5}$$

Cohen's class is special in the sense that both the marginal and covariance properties are easily characterized.

Generalized JSRs extend the scope of TFRs to a richer class of signals by representing signal characteristics in terms of variables other than time and frequency. The idea is to represent signal characteristics in terms of variables that are better matched to a given class of signals. In existing literature, there are two main approaches to such generalizations: Cohen's general distributional method [1], and arbitrary JSR constructions based on covariance arguments [17, 18, 19]. Cohen's recent generalization is an extension of his original method (for TFRs) to arbitrary variables [1, 11]. Baraniuk also proposed a generalization based on group theoretic arguments [13] which was shown to be equivalent to Cohen's approach [15]. Although marginals are trivially characterized in Cohen's recipe, the covariance properties are neither guaranteed nor easily characterized. As we will see, covariance properties are crucial in a JSR-based detection framework and thus for such applications a covariance-based generalization is needed (the affine class [20] was constructed using covariance arguments). In such generalizations, which have also been proposed recently [17, 18, 19],

[2]The discussion also applies to cases in which the signal space of interest is a closed (and hence complete) subspace of $L^2(\mathbb{R})$; for example, the space of analytic signals whose Fourier transform vanishes for negative frequencies.

the marginal properties become difficult to characterize except in well-known cases (such as the class of JSRs that are unitarily equivalent to Cohen's class of TFRs [12, 16, 18]). We now provide a brief description of the two methods. Although we restrict our discussion to JSRs with respect to two variables, the results can be readily extended to more variables.

2.1 Cohen's method

In Cohen's method, JSRs are interpreted as quasi-energy distributions as a function of the variables of interest, and a fundamental idea is to associate the variables with Hermitian operators [1, 16, 14]. Let a and b be two variables of interest, associated with the Hermitian operators $\mathbf{A}$ and $\mathbf{B}$, respectively. The projection onto the eigenfunctions of the operators defines two unitary signal transforms, $\mathbf{S_A}$ and $\mathbf{S_B}$, and the squared-magnitudes of the projections, $|(\mathbf{S_A}s)(a)|^2$ and $|(\mathbf{S_B}s)(b)|^2$, are the distribution of the signal energy with respect to the corresponding variables. A key role is played by the characteristic function, $\mathbf{M}: L^2(\mathbb{R}) \to L^2(\mathbb{R}^2)$, that is related to the JSR, $\mathbf{P}: L^2(\mathbb{R}) \to L^2(\mathbb{R}^2)$, via a 2d Fourier transform:

$$(\mathbf{M}s)(\alpha,\beta) \equiv \int\int (\mathbf{P}s)(a,b)e^{-j2\pi\alpha a}e^{j2\pi\beta b}da db \; , \tag{6}$$

$$(\mathbf{P}s)(a,b) = \int\int (\mathbf{M}s)(\alpha,\beta)e^{j2\pi\alpha a}e^{-j2\pi\beta b}d\alpha d\beta \; . \tag{7}$$

The characteristic function, $\mathbf{M}$, being the average (or expected value) of the function $e^{-j2\pi\alpha a}e^{j2\pi\beta b}$, can be computed directly from the signal using the quadratic form

$$(\mathbf{M}s)(\alpha,\beta) = \langle \mathbf{M}_{(\alpha,\beta)}s, s\rangle \tag{8}$$

where $\mathbf{M}_{(\alpha,\beta)}$ is a "characteristic function operator" corresponding to the function $e^{-j2\pi\alpha a}e^{j2\pi\beta b}$, an example being $\mathbf{M}_{(\alpha,\beta)} = e^{-j2\pi(\alpha\mathbf{A}-\beta\mathbf{B})}$. The JSR, $\mathbf{P}$, can then be recovered from (8) by using (7). Since the operators $\mathbf{A}$ and $\mathbf{B}$ do not commute in general, there are infinitely many operator correspondences to the function $e^{-j2\pi\alpha a}e^{j2\pi\beta b}$, and together they *define* the entire class of JSRs with respect to the variables a and b. Since it is impossible to explicitly characterize all such correspondences, Cohen's has proposed a *kernel method* [1] that generates a whole class of JSRs by weighting a particular characteristic function, say $\mathbf{M}_o$, by a 2d kernel ϕ[3]

$$(\mathbf{P}s)(a,b;\phi) = \int\int (\mathbf{M}_o s)(\alpha,\beta)\phi(\alpha,\beta)e^{j2\pi\alpha a}e^{-j2\pi\beta b}d\alpha d\beta \; , \tag{9}$$

where, given a fixed characteristic function $\mathbf{M}_o$, the kernel ϕ completely characterizes the class of JSRs.[4] If the kernel ϕ satisfies $\phi(\alpha,0) = \phi(0,\beta) = 1$ for all α,β, then the representation satisfies the marginal relations

$$\int (\mathbf{P}s)(a,b;\phi)db = |(\mathbf{S_A}s)(a)|^2 \quad , \quad \int (\mathbf{P}s)(a,b;\phi)da = |(\mathbf{S_B}s)(b)|^2 \; , \tag{10}$$

analogous to joint probability density functions. The covariance properties, however, are neither guaranteed nor easily characterizable except in special cases.

[3]The kernel may also depend on the variables (a,b) [1]. However, for most part, we will be interested in kernels that are independent of the variables.

[4]In general, such a weighing by a 2d kernel (independent of the variables) does not generate all possible correspondences [22, 23]. Starting with a particular characteristic function, $\mathbf{M}_o$, the kernel method (9) generates a *subclass* of the entire class. Thus, our discussion would implicitly be based on the subclass generated by the particular characteristic function $\mathbf{M}_o$ used in (9).

2.2 Covariance-based method

Although Cohen's method associates each variable with a Hermitian operator, variables can equivalently be associated with families of unitary operators [12, 16] as well. In a covariance-based approach, the latter correspondence is more attractive because the transformation of the signal by the unitary operator effectively corresponds to a "shift" in the corresponding variable [16]. A general framework for covariance-based JSRs has been developed in [17, 19, 18]. However, we follow the recipe of [19] because of its simplicity and adequacy for the detection framework.[5]

Consider a parameter set $G \subset \mathbb{R}^2$ and a family of unitary operators $\{\mathbf{U}_{(a,b)}\}$, $(a,b) \in \mathbb{R}^2$, defined on $L^2(\mathbb{R})$; that is, $\langle \mathbf{U}_{(a,b)} s, \mathbf{U}_{(a,b)} s \rangle = \langle s, s \rangle$ for all $s \in L^2(\mathbb{R})$ and for all $(a,b) \in \mathbb{R}^2$. The two "coordinates" a and b represent our variables of interest. We note that in most cases the operator $\mathbf{U}_{(a,b)}$ is a composition of two unitary operators, say $\mathbf{A}_a$ and $\mathbf{B}_b$, that represent individual variables of interest [19, 14, 18].

In a covariance-based approach we are interested in bilinear JSRs, with respect to the variables a and b, which are covariant to the unitary operator $\mathbf{U}_{(a,b)}$ in the sense that transformation of the signal by $\mathbf{U}_{(a,b)}$ produces a "shift" in the coordinates of the JSR, $(\mathbf{P}s)(a,b)$. For example, TFRs from Cohen's class are covariant to the time-frequency shift operator $\mathbf{T}_\mu \mathbf{F}_\nu$ as evident from (5), and the TSRs from the affine class are covariant to the time-scale shift operator $\mathbf{T}_\mu \mathbf{D}_c$, where $(\mathbf{D}_c s)(t) = \frac{1}{\sqrt{c}} s(t/c)$ (see the example in Section 3.2). It can be shown that under fairly natural assumptions regarding the covariance properties, the family $\mathbf{U}_{(a,b)}$ must form a group under composition [24, 14, 19, 17]; that is, the parameter set G is a group with the group operation $\bullet$ defined by [19, 17, 18]

$$\mathbf{U}_{(a,b)} \mathbf{U}_{(a',b')} = \mathbf{U}_{(a,b)\bullet(a',b')} \ \ \text{(within a phase factor)} \ . \tag{11}$$

The appropriate covariance relation is the group operation; that is, the JSRs satisfy [19, 17, 18]

$$\left(\mathbf{P}\mathbf{U}_{(a',b')} s\right)(a,b) = (\mathbf{P}s)\left((a',b')^{-1} \bullet (a,b)\right) \ , \ \text{for all } (a,b),\ (a',b') \in G \ . \tag{12}$$

All bilinear JSRs that satisfy (12) can be simply characterized as [19, 17, 18]

$$(\mathbf{P}s)(a,b;\phi) = \left\langle \phi \mathbf{U}_{(a,b)^{-1}} s, \mathbf{U}_{(a,b)^{-1}} s \right\rangle \ , \ (a,b) \in G \ , \tag{13}$$

where $\phi : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is a linear operator characterizing the JSRs from the class.

We note that Cohen's class of TFRs[6] is special in that it can be derived using either of the two methods.

3 Optimum Estimation Framework

As mentioned in the Introduction, generalized JSRs can provide an accurate representation of a wide variety of nonstationary signal characteristics. Such information about the signal structure can be used for understanding the underlying physical mechanisms, or for further analysis or processing. However, in many applications the signal of interest is best modeled as random, and the signal measurements are often corrupted by noise or interference. In such practical nonstationary scenarios, the desired JSR of the signal of interest has to be estimated from corrupted observations. In this section, we develop optimal estimation techniques relevant in this context.

[5]The method described in [17, 18] involves a remapping of coordinates to time-frequency that is unnecessary from a detection viewpoint, and our interest in the covariance-based approach is primarily for the detection framework. However, we will briefly discuss the issue of remapping of coordinates as well.

[6]And other classes that are unitary equivalent to Cohen's class of TFRs [12, 16].

The theory that we develop is an extension of the minimum-mean-squared-error (mmse) estimation framework developed in [3] for Cohen's class of TFRs. As for notation in this section, we will denote random processes and random variables by uppercase letters, and deterministic signals, constants and particular realizations of random processes by lowercase letters. All integrals involving random functions will be interpreted as mean-square integrals [25].

Let $X(t)$ and $Y(t)$, $t \in T$, be two random processes that we assume to possess finite fourth-order moments. In addition, we make the mild assumption that they have finite energy:

$$\int_T E|X(t)|^2 dt < \infty \ , \tag{14}$$

and similarly for Y, where E denotes the expectation operator. Y denotes the underlying "desired" process whose nonstationary characteristics, as determined by a particular JSR, we are interested in estimating from the observed process X that is correlated in some way to Y; for example, X could be a noise/interference corrupted version of Y.

We could be interested in two types of characteristics: *Average* characteristics that are the expected value of a particular JSR of Y, or *realization-based* characteristics that are the value of a particular JSR for *each* realization of the random process Y. Average characteristics measure the average distribution of signal energy as a function of the variables of interest, whereas realization-based characteristics measure the energy distribution of individual realizations of the process. Both types of characteristics expose the nonstationary structure of the process that is particularly matched to the variables of a given class of JSRs.

We now formulate the two estimation problems. Assume that we have chosen a class C of JSRs, corresponding to variables a and b, that is well-suited for processing the random processes X and Y. For example, for processes exhibiting a self-similar structure, the affine class of TSRs may be used.

Average characteristics. The average characteristics of the desired process Y that we want to estimate are determined by the JSR, $\mathbf{P}(\phi_o)$, where ϕ_o is the characterizing kernel (Cohen's method) or operator (covariance-based method) that is chosen *a priori*[7]; that is, $E\{(\mathbf{P}Y)(a,b;\phi_o)\}$ needs to be estimated from a realization of the observed process X. The class of estimators is the same class, C, of JSRs. The objective is to find an estimator, equivalently a kernel or operator ϕ_{opt}, that is optimal in the following mmse sense:

$$\phi_{opt} \equiv \arg\inf_{\phi} E\left\{\int\int |E(\mathbf{P}Y)(a,b;\phi_o) - (\mathbf{P}X)(a,b;\phi)|^2 \, d\mu(a,b)\right\} \ , \tag{15}$$

where $d\mu(a,b) = da \, db$ in Cohen's method and μ is the left Haar measure [26, 24] for the underlying group in the covariance-based method. Thus, the optimal estimator minimizes the mean-squared-error (mse), integrated over the a-b plane, between the desired expected JSR, $E\{(\mathbf{P}Y)(\phi_o)\}$, and the estimate $(\mathbf{P}X)(\phi)$. For example, the Wigner-Ville spectrum (WVS) [27, 28] (the expected value of the Wigner distribution) of the process Y may be estimated from a realization that is corrupted by $1/f$-type noise, by using an optimal estimator from the affine class of TSRs that perform a "constant-Q" processing appropriate for $1/f$-type noise. Optimal WVS estimation using Cohen's class of TFRs was addressed in [3, 2].

Realization-based characteristics. Suppose that a JSR characterized by the kernel/operator ϕ_r, the "reference" kernel, that we assume to have been chosen *a priori*[8], generates useful JSRs for *all* realizations of the desired process Y. In this case, for each realization, y, of Y, we are interested in estimating the

[7]The choice of ϕ_o is, of course, a nontrivial problem, but it is tangential to the scope of the estimation framework. Our focus, in this section, is on optimally estimating a *given* JSR from corrupted observations.

[8]See footnote 7. Same remarks apply to ϕ_r.

JSR, $(\mathbf{P}y)(a,b;\phi_r)$, from the corresponding realization, x, of X, via an estimate $(\mathbf{P}x)(a,b;\phi_{opt})$ that again minimizes the integrated mse

$$\phi_{opt} \equiv \arg\inf_{\phi} E\left\{\int\int |(\mathbf{P}Y)(a,b;\phi_r) - (\mathbf{P}X)(a,b;\phi)|^2 \, d\mu(a,b)\right\} . \tag{16}$$

Note that in this case we are interested in estimating the nonstationary characteristics of *each* realization of Y, individually, as opposed to the average characteristics over *all* realizations in the previous case. For example, in biological signals such as the electrocardiogram, each realization may carry useful information that has to be extracted from its nonstationary characteristics.

Having formulated the two estimation problems, we now characterize the optimal solutions in the two frameworks for generalized JSRs.

3.1 Optimal kernel solution in Cohen's method

The relevant characterization of JSRs in Cohen's method is (9). We start by deriving the optimal kernel for estimating the average (expected) JSRs.

Average characteristics. Using the expression (9) for the JSRs in (15), and using Parseval's relation, we get the following equivalent formulation for (15):

$$\phi_{opt} = \arg\inf_{\phi} \int\int E\left\{\left|E\left\{(\mathbf{M}_o Y)(\alpha,\beta)\right\}\phi_o(\alpha,\beta) - (\mathbf{M}_o X)(\alpha,\beta)\phi(\alpha,\beta)\right|^2\right\} d\alpha d\beta \ , \tag{17}$$

where we have interchanged the order of expectation and integration, which is justified if the integral in (17) is finite.[9] Since the integrand in (17) is nonnegative everywhere, the optimization problem reduces to

$$\phi_{opt}(\alpha,\beta) = \arg\inf_{\phi(\alpha,\beta)} E\left\{|E\{(\mathbf{M}_o Y)(\alpha,\beta)\}\phi_o(\alpha,\beta) - (\mathbf{M}_o X)(\alpha,\beta)\phi(\alpha,\beta)|^2\right\} \ , \ (\alpha,\beta) \in \mathbb{R}^2 \ , \tag{18}$$

which is a simple linear mmse estimation problem. By the orthogonality principle[10] [25], the optimal kernel is characterized as

$$\phi_{opt}(\alpha,\beta) = \begin{cases} \frac{B_{YX}(\alpha,\beta)}{B_X(\alpha,\beta)}\phi_o(\alpha,\beta) & \text{if } (\alpha,\beta) \in S_{YX} \\ 0 & \text{otherwise} \end{cases} \ , \tag{19}$$

where

$$B_{YX}(\alpha,\beta) = E\{(\mathbf{M}_o Y)(\alpha,\beta)\}E\left\{(\mathbf{M}_o X)^*(\alpha,\beta)\right\} \ , \tag{20}$$

$$B_X(\alpha,\beta) = E\{|(\mathbf{M}_o X)(\alpha,\beta)|^2\} \ , \tag{21}$$

and S_{YX} is the support of B_{YX}; that is, the set of points where $|B_{YX}|$ is nonzero.

Realization-based characteristics. Using similar arguments as above, the optimal kernel for realization-based estimation, solving (16), is given by

$$\phi_{opt}(\alpha,\beta) = \begin{cases} \frac{\widehat{B}_{YX}(\alpha,\beta)}{B_X(\alpha,\beta)}\phi_r(\alpha,\beta) & \text{if } (\alpha,\beta) \in \widehat{S}_{YX} \\ 0 & \text{otherwise} \end{cases} \ , \tag{22}$$

where

$$\widehat{B}_{YX}(\alpha,\beta) = E\left\{(\mathbf{M}_o Y)(\alpha,\beta)(\mathbf{M}_o X)^*(\alpha,\beta)\right\} \ , \tag{23}$$

[9] We restrict the choice of ϕ_{opt} over a class for which the integral in (17) is finite.

[10] Which states that the error must be orthogonal to the "observations"; that is, $E\{[E\{(\mathbf{M}_o Y)(\alpha,\beta)\}\phi_o(\alpha,\beta)$ - $(\mathbf{M}_o X)(\alpha,\beta)\phi(\alpha,\beta)](\mathbf{M}_o X)^*(\alpha,\beta)\} = 0$ for all (θ,τ).

and $\widehat{S}_{YX}$ is the support of $\widehat{B}_{YX}$.

We note from (19) and (22) that the optimal kernel depends on certain second- and fourth-order statistics of the processes: $\widehat{B}_{YX}$ and B_X defined in (23) and (21) depend on fourth-order statistics and B_{YX} defined in (20) depends on second-order statistics. Also, the optimal estimator for average statistics effectively depends on the statistics to be estimated.[11] However, the solution yields useful information about the structure of the optimal estimator that can be used to design estimators when only partial information, such as the supports of the expected ambiguity functions, is available [5]. It is also worth noting the similarity of the optimal kernel solution to the spectral domain Wiener filter solution for mmse estimation of a desired stationary process Y from a correlated stationary process X; B_{YX} and $\widehat{B}_{YX}$ replace the cross-spectral density of Y and X, and B_X replaces the auto-spectral density of X. Finally, we note that the results presented in [3] (global estimation) are a special case of the above general results for Cohen's method.

3.2 Solutions for the covariance-based method

Characterization of the optimal estimators in the covariance-based method is slightly more involved compared to that for Cohen's method. For notational simplicity we define

$$s_{(a,b)}(t) \equiv (\mathbf{U}_{(a,b)^{-1}} s)(t) \ , \ t \in \mathbb{R} \ , \ (a,b) \in G \ , \tag{24}$$

for any signal transformed by $\mathbf{U}_{(a,b)^{-1}}$. The relevant equation in this method is (13).

Average characteristics. Using (13) and the definition (24), we can express (15) more explicitly as

$$\phi_{opt} = \arg\inf_{\phi} E\left\{ \int\int \left| E\left\{ \langle \phi_o Y_{(a,b)}, Y_{(a,b)} \rangle \right\} - \langle \phi X_{(a,b)}, X_{(a,b)} \rangle \right|^2 d\mu(a,b) \right\} \ , \tag{25}$$

and $(\mathbf{P}X)(a,b;\phi)$ can be expressed more explicitly as

$$(\mathbf{P}X)(a,b;\phi) = \langle \phi X_{(a,b)}, X_{(a,b)} \rangle = \int\int K_\phi(u_1,u_2) X_{(a,b)}(u_2) X^*_{(a,b)}(u_1) du_1 du_2 \ , \tag{26}$$

and similarly for $(\mathbf{P}Y)(\phi_o)$, where K_ϕ represents the kernel of the operator ϕ:

$$(\phi s)(t) \equiv \int K_\phi(t,u) s(u) du \ . \tag{27}$$

Both $(\mathbf{P}X)(\phi)$ and $(\mathbf{P}Y)(\phi_o)$ belong to an appropriately defined Hilbert space, $\mathcal{H}$, of second-order random functions equipped with the inner product

$$\langle A, B \rangle_{\mathcal{H}} \equiv E\left\{ \int A(a,b) B^*(a,b) d\mu(a,b) \right\} \ , \ A, B \in \mathcal{H} \ , \tag{28}$$

and for fixed $(u_1,u_2) \in \mathbb{R}^2$, the random functions $Q_X^{(u_1,u_2)}(a,b) = X_{(a,b)}(u_2) X^*_{(a,b)}(u_1)$ and $Q_Y^{(u_1,u_2)}(a,b)$ also belong to $\mathcal{H}$. From (26) we note that, as an element of $\mathcal{H}$, $(\mathbf{P}X)(\phi)$ is in the linear span of the family of random functions $\{Q_X^{(u_1,u_2)} : (u_1,u_2) \in \mathbb{R}^2\}$, and thus (25) is simply a problem of linear mmse estimation in $\mathcal{H}$. Using the orthogonality principle [25], the optimal operator, ϕ_{opt}, for estimating the expected value of $(\mathbf{P}Y)(\phi_o)$ is characterized by the linear equation

$$\left\langle (\mathbf{P}X)(a,b;\phi_{opt}) - E\left\{ (\mathbf{P}Y)(a,b;\phi_o) \right\} \ , \ Q_X^{(u_1,u_2)} \right\rangle_{\mathcal{H}} = 0 \ , \ \text{ for all } (u_1,u_2) \in \mathbb{R}^2 \ , \tag{29}$$

[11] Just as the mmse-optimal stationary spectrum estimator depends on the spectrum itself [29, pp. 275], [3].

which can be expressed more explicitly as

$$\begin{aligned}
& \int\!\!\int\!\!\int\!\!\int K_{\phi_{opt}}(t_1,t_2) E\left\{X_{(a,b)}(t_2)X^*_{(a,b)}(t_1)X_{(a,b)}(u_1)X^*_{(a,b)}(u_2)\right\} d\mu(a,b)dt_1dt_2 \\
= & \int\!\!\int\!\!\int\!\!\int K_{\phi_o}(t_1,t_2) E\left\{Y_{(a,b)}(t_2)Y^*_{(a,b)}(t_1)\right\} E\left\{X_{(a,b)}(u_1)X^*_{(a,b)}(u_2)\right\} d\mu(a,b)dt_1dt_2 \qquad (30)
\end{aligned}$$

for all $(u_1,u_2) \in \mathbb{R}^2$. The characterizing equation (30) can be written more compactly as

$$\int\!\!\int K_{\phi_{opt}}(t_1,t_2) G_X(t_1,t_2;u_1,u_2) dt_1 dt_2 = \int\!\!\int K_{\phi_o}(t_1,t_2) G_{YX}(t_1,t_2;u_1,u_2) dt_1 dt_2 \ , \qquad (31)$$

for all $(u_1,u_2) \in \mathbb{R}^2$, where

$$\begin{aligned}
G_X(t_1,t_2;u_1,u_2) &\equiv \int\!\!\int E\left\{X_{(a,b)}(t_2)X^*_{(a,b)}(t_1)X_{(a,b)}(u_1)X^*_{(a,b)}(u_2)\right\} d\mu(a,b) \ \text{ and} \qquad (32)\\
G_{YX}(t_1,t_2;u_1,u_2) &\equiv \int\!\!\int E\left\{Y_{(a,b)}(t_2)Y^*_{(a,b)}(t_1)\right\} E\left\{X_{(a,b)}(u_1)X^*_{(a,b)}(u_2)\right\} d\mu(a,b) \ . \qquad (33)
\end{aligned}$$

Recall that K_ϕ is the kernel of the operator ϕ (see (27)).

Realization-based characteristics. Using similar arguments as above, the optimal operator ϕ_{opt} for realization-based estimation, solving (16), is characterized by the linear equation

$$\int\!\!\int K_{\phi_{opt}}(t_1,t_2) G_X(t_1,t_2;u_1,u_2) dt_1 dt_2 = \int\!\!\int K_{\phi_r}(t_1,t_2) \widehat{G}_{YX}(t_1,t_2;u_1,u_2) dt_1 dt_2 \ , \qquad (34)$$

for all $(u_1,u_2) \in \mathbb{R}^2$, where

$$\widehat{G}_{YX}(t_1,t_2;u_1,u_2) \equiv \int\!\!\int E\left\{Y_{(a,b)}(t_2)Y^*_{(a,b)}(t_1)X_{(a,b)}(u_1)X^*_{(a,b)}(u_2)\right\} d\mu(a,b) \ . \qquad (35)$$

We again note from (31) and (34) that the optimal operator depends on certain second- and fourth-order statistics of the processes. Also, the characterization of the optimal operator in the covariance-based framework is a level more complex than that in Cohen's framework: the kernel of the operator is characterized via a linear integral equation as compared to the simple pointwise characterization in Cohen's method.[12]

Example: Estimation using the affine class. Let $G = \mathbb{R} \times (0,\infty)$, with the group operation $(t_1,c_1)\bullet(t_2,c_2) = (t_1+c_1t_2, c_1c_2)$ (affine group), $(t,c)^{-1} = (-t/c, 1/c)$, induced by the unitary transformation

$$\left(\mathbf{U}_{(t,c)}s\right)(\tau) = (\mathbf{T}_t\mathbf{D}_c s)(\tau) \equiv \frac{1}{\sqrt{c}} s\left(\frac{\tau - t}{c}\right) \ . \qquad (36)$$

Using the above unitary transform in (13), we generate the affine class of bilinear TSRs [20, 19]

$$(\mathbf{P}s)(t,c;\phi) = \int\!\!\int \Pi((u-t)/c, fc) W_s(u,f) du df = \int\!\!\int \Phi(\theta,\tau) A_s(\theta/c, c\tau) e^{j\frac{2\pi\theta t}{c}} d\theta d\tau \ , \qquad (37)$$

where W_s is the Wigner distribution and A_s the ambiguity function (AF) of s defined as

$$\begin{aligned}
W_s(t,f) &= \int s(t+\tau/2)s^*(t-\tau/2)e^{-j2\pi f\tau} d\tau \ , \qquad (38)\\
A_s(\theta,\tau) &= \int s(u+\tau/2)s^*(u-\tau/2)e^{-j2\pi\theta u} du \ , \qquad (39)
\end{aligned}$$

[12] If the two methods yield the same class of JSRs, as is true in the case of Cohen's class of TFRs, the two estimation procedures also yield identical solutions; in particular, the integral equations in (31) and (34) collapse (after an invertible transformation) into a pointwise characterization as in Cohen's method.

and Π and Φ are different forms of the operator kernel K_ϕ defined as

$$\Pi(u,v) = \int K_\phi(u+\tau/2, u-\tau/2) e^{-j2\pi v\tau} d\tau \ , \tag{40}$$

$$\Phi(\theta,\tau) = \int K_\phi(-(u+\tau/2), -(u-\tau/2)) e^{-j2\pi\theta u} du \ . \tag{41}$$

Using (39) and (32), G_X can be expressed in terms of AFs as

$$G_X(t_1,t_2;u_1,u_2) = \int_0^\infty \int E\left\{A_X(\theta, c(t_1-t_2)) A_X^*(\theta, c(u_2-u_1))\right\} e^{j2\pi\theta c((u_1+u_2)/2-(t_1+t_2)/2)} d\theta dc \ , \tag{42}$$

and G_{YX} can be similarly expressed as

$$G_{YX}(t_1,t_2;u_1,u_2) = \int_0^\infty \int E\{A_Y(\theta, c(t_1-t_2))\}\, E\left\{A_X^*(\theta, c(u_2-u_1))\right\} e^{j2\pi\theta c((u_1+u_2)/2-(t_1+t_2)/2)} d\theta dc \ . \tag{43}$$

(42) and (43) can then be substituted in (31) which, after a few changes of variables, yields the following linear characterization of the optimal operator for estimation of average characteristics:

$$\int \Phi_{opt}(\theta',\tau) H_X(\tau;\theta',\tau') d\tau = \int \Phi_o(\theta',\tau) H_{YX}(\tau;\theta',\tau') d\tau \quad \text{for all } (\theta',\tau') \in \mathbb{R}^2 \ , \tag{44}$$

where

$$H_X(\tau;\theta',\tau') \equiv \int E\left\{A_X(\theta'/c, c\tau) A_X^*(\theta'/c, c\tau')\right\} \frac{dc}{c} \quad \text{and} \tag{45}$$

$$H_{YX}(\tau;\theta',\tau') \equiv \int E\left\{A_Y(\theta'/c, c\tau)\right\} E\left\{A_X^*(\theta'/c, c\tau')\right\} \frac{dc}{c} \ . \tag{46}$$

The optimal kernel for realization-based estimation is also characterized by the linear equation (44) after replacing Φ_o with Φ_r and H_{YX} with $\widehat{H}_{YX}$ defined as

$$\widehat{H}_{YX}(\tau;\theta',\tau') \equiv \int E\left\{A_Y(\theta'/c, c\tau) A_X^*(\theta'/c, c\tau')\right\} \frac{dc}{c} \ . \tag{47}$$

We note that the optimal solution for the affine class is simpler than the completely general solution. The operator kernel is defined via an integral equation only in one variable; it is defined pointwise in the other variable.

3.3 Local estimation

Thus far we have discussed estimation procedures that are global in the sense that the kernels/operators characterizing the JSRs are not a function of the variables a,b (analogous to the time-frequency invariant kernels in [3]). The optimality criterion, being the mse integrated over the entire a-b plane, also reflects the global nature of the problem. However, in some cases it might be more appropriate to allow the kernels to vary with the variables in order to better track the nonstationary structure of the processes. For example, in the case of TSRs, allowing the kernel to vary with time facilitates the TSR to be better matched to a possibly "time-varying Q"[13] of a process. We now briefly discuss such local estimation problems.

We start with local estimation of realization-based characteristics. The optimal a-b varying kernel/operator is one which solves the optimization problem

$$\phi_{opt}^{(a,b)} \equiv \arg\inf_\phi E\left\{\left|(\mathbf{P}Y)(a,b;\phi_r^{(a,b)}) - (\mathbf{P}X)(a,b;\phi)\right|^2\right\} \ , \tag{48}$$

[13]That is, the ratio of bandwidth to center frequency, of dominant signal components, is constant over frequency but changes with time.

where we note that the reference kernel $\phi_r^{(a,b)}$ may itself vary with a and b. In both Cohen's method and the covariance-based method, the JSRs can be expressed as

$$(\mathbf{P}s)(a,b;\phi) = \int\int T_\phi(u,v) Q_s(a,b;u,v)\,dudv \ , \tag{49}$$

where

$$T_\phi(u,v) = \begin{cases} \phi(u,v) & \text{Cohen's method} \\ K_\phi(u,v) & \text{covariance-based method} \end{cases} , \tag{50}$$

and Q_s is a quadratic form in the signal given by

$$Q_s(a,b;u,v) = \begin{cases} (\mathbf{M}_o s)(u,v) e^{j2\pi ua} e^{-j2\pi vb} & \text{Cohen's method} \\ \left(\mathbf{U}_{(a,b)}^{-1} s\right)(v) \left(\mathbf{U}_{(a,b)}^{-1} s\right)^*(u) & \text{covariance-based method} \end{cases} . \tag{51}$$

From (49), we note that for each value of (a,b), $(\mathbf{P}X)(a,b;\phi)$ is a second-order[14] random variable that lies in the linear span of the family of second-order random variables $\{Q_X(a,b;u,v) : (u,v) \in \mathbb{R}^2\}$. Thus, (48) is again a linear mmse estimation problem and, by the orthogonality principle, for each fixed value of (a,b), the optimal solution, $\phi_{opt}^{(a,b)}$, is characterized by the linear equation

$$\begin{aligned} & \int\int T_{\phi_{opt}^{(a,b)}}(u,v) E\left\{Q_X(a,b;u,v) Q_X^*(a,b;u',v')\right\} dudv \\ = & \int\int T_{\phi_r^{(a,b)}}(u,v) E\left\{Q_Y(a,b;u,v) Q_X^*(a,b;u',v')\right\} dudv \quad \text{for all } (u',v') \in \mathbb{R}^2 \ . \end{aligned} \tag{52}$$

The optimal solution for average characteristics is also characterized by the above equation by replacing $T_{\phi_r^{(a,b)}}$ with $T_{\phi_o^{(a,b)}}$ and $E\{Q_Y(a,b;u,v)Q_X^*(a,b;u',v')\}$ with $E\{Q_Y(a,b;u,v)\}E\{Q_X^*(a,b;u',v')\}$ on the right-hand side.

3.4 Discussion

Sections 3.2 and 3.3 developed a mmse estimation theory based on generalized JSRs. The developed results allow optimal estimation of any JSR of a desired random process from realizations of a correlated random process. Procedures were developed for estimating both the expected value of the JSR or the value of the JSR for individual realizations. Since the theory applies to any class of bilinear generalized JSRs, the results allow optimal estimation of a wide variety of *average* and *realization-based* nonstationary characteristics of random signals form observations that are corrupted by noise and interference. Moreover, if two classes of JSRs have common members, the estimation of a JSR from the intersection of the two classes can be effected by JSRs from either of the two classes. For example, the Wigner-Ville spectrum [27, 28] of a random process can be estimated using TFRs from Cohen's class or TSRs from the affine class, since the Wigner distribution is a member of both classes.[15]

The estimation theory encompasses generalized JSRs which analyze signals in terms of arbitrary variables of interest. Although the characteristics of such generalized JSRs are radically different from joint TFRs, in many cases, because of the fundamental importance of time and frequency, the final objective is to obtain an estimate of signal energy as a function of time and frequency. This requires a remapping of the coordinates of the generalized JSR into time and frequency and is usually accomplished with a unitary axis

[14] Under appropriate assumptions on ϕ.

[15] Our framework also facilitates the choice of the best estimator in such cases; namely the one for which the integrated mse (the quantity minimized in (15) or (16)) is lower.

transformation [12, 17]. Also, given a class of JSRs, a new class of JSRs having radically different properties can be defined by unitarily preprocessing the signals to be analyzed [12, 15]. Although such input and output unitary transforms change the properties of a given class of JSRs, they are structurally equivalent to the original class of JSR via the principle of unitary equivalence [12, 16]. We now describe the effect of such input and output unitary transforms on the optimal kernel/operator solutions.

Unitary equivalence. Suppose that we are given a class, C, of JSRs and consider a new class C' of JSRs that is unitarily equivalent to C via the unitary transformations $\mathbf{U}$ and $\mathbf{V}$:

$$C' = \{\mathbf{VPU} : \mathbf{P} \in C\} \ . \tag{53}$$

How are the optimal estimators for C' related to those for C?

We first examine the effect of $\mathbf{V}$. In global estimation, the postprocessing unitary transformation $\mathbf{V}$ has no effect on the design of optimal estimators. The reason is that, by definition of unitarity,

$$\int\int |(\mathbf{VP}s)(\alpha,\beta;\phi)|^2 \, d\mu'(\alpha,\beta) = \int\int |(\mathbf{P}s)(a,b;\phi)|^2 \, d\mu(a,b) \ , \tag{54}$$

where μ and μ' are the appropriate measures for the spaces to which the JSRs belong, and thus the optimality criteria (15) and (16) are unaffected by the transformation $\mathbf{V}$. In local estimation, however, the optimal kernel solution is affected if $\mathbf{V}$ is a completely general unitary transformation. However, for the special and important case in which $\mathbf{V}$ constitutes an axis warping, that is,

$$(\mathbf{VP}s)(\alpha,\beta;\phi) \equiv (\mathbf{P}s)(a(\alpha,\beta), b(\alpha,\beta);\phi) \ , \tag{55}$$

the optimal kernels for C' are related to those for C by

$$\tilde{\phi}_{opt}^{(\alpha,\beta)} = \phi_{opt}^{(a(\alpha,\beta),b(\alpha,\beta))} \ , \tag{56}$$

where $\tilde{\phi}_{opt}^{(\alpha,\beta)}$ is the optimal kernel for the post-warped class $C' = \mathbf{V}C$. Thus, the performance of the local estimators is not affected by the warping $\mathbf{V}$ either; only the variation of the optimal estimator as a function of the variables is altered in a one-to-one fashion. Note that remapping of arbitrary variables into time and frequency is accomplished precisely through such an axis-warping transformation $\mathbf{V}$ [12, 17].

The effect of the preprocessing transformation $\mathbf{U}$ is simply determined by the altered form of the various statistics that are required for computing the optimal estimators. More precisely, the various statistics B_{YX}, $\widehat{B}_{YX}$, B_X, G_{YX}, $\widehat{G}_{YX}$ and G_X defined in (20), (23), (21), (33), (35) and (32), respectively, and those used in the local estimation equation (52), are computed by replacing X and Y with $\mathbf{U}X$ and $\mathbf{U}Y$.

4 Optimum Detection Framework

Linear TFRs and TSRs have been used extensively for detecting transient, nonstationary signals in the presence of additive noise; the short time Fourier transform (STFT) and the wavelet transform (WT) are widely used in narrowband and wideband radar, respectively, for detecting deterministic targets with unknown range and velocity. The key observation in such TFR/TSR-based detectors is that the signal parameters (range/velocity) naturally correspond to the variables of the JSR (time-frequency shifts and time-scale shifts). However, a major limitation of linear TFRs/TSRs is that they constitute optimal detectors only for deterministic signals with possibly unknown phase or random amplitude. Quadratic TFRs/TSRs, on the other hand, facilitate optimal detection of arbitrary nonstationary random signals, with certain unknown parameters, in Gaussian noise [6, 8].

Such unknown or random nuisance signal parameters are encountered commonly in practice. For example, in machine fault diagnostics, the faults occur at unknown time-offsets and often exhibit unknown modulations [9]. In other applications, such as sleep data classification, our experience has shown that scaling parameters also seem relevant. In general, however, due to the variety of nonstationary signals encountered in practice, the variables of time, frequency and scale are not adequate for accurately representing all such nuisance signal parameters.

Generalized JSRs provide the ability of detecting signals with parameters that are more general than time-frequency or time-scale shifts. Essentially, time-frequency or time-scale shifts are unitary operations, and generalized JSRs facilitate the detection of signals that have undergone more general parameterized unitary transformations; the parameters of the unitary transformation constitute the unknown (or random) signal parameters. This observation, and we will elaborate upon it later, makes the covariance-based approach the appropriate vehicle for the JSR-based detection framework. The underlying unitary transformations may model the effect of channels or systems that produce nondissipative (energy preserving) signal distortion.

In this section, we develop an optimal detection framework based on JSRs that is a generalization of the theory developed in [6, 8] for TFRs and TSRs. For each class of covariance-based JSRs, we characterize the corresponding class of detection problems for which such JSRs constitute canonical detectors. We also explicitly characterize the form of the corresponding JSR-based detectors. The structure of the JSR detectors yields a useful subspace-based formulation in terms of linear JSRs which we discuss as well. We start with a brief review of relevant detection theory.

4.1 Classical detection theory

We consider the following binary hypothesis testing problem:

$$\begin{aligned} H_0 &: \quad x(t) = n(t) \\ H_1 &: \quad x(t) = as(t) + n(t) \end{aligned} \tag{57}$$

where $t \in T$, the observation interval, x is the observed signal, s is the signal to be detected, n is additive noise and a is a positive parameter. Based on the observed signal x, it has to be decided whether the signal is present (H_1) or not (H_0). The optimal decision is made by comparing a particular real-valued function of data, $L(x)$, called the test statistic, to a threshold.

We are interested in JSR-based detectors and will be restricting our attention primarily to quadratic JSRs (as opposed to linear ones) because they can realize a richer class of detectors (we will briefly discuss linear JSR-based detectors as well). In this context, a key observation is that a quadratic JSR, say $(\mathbf{P}x)(a, b)$, realizes a quadratic function of the observed signal, x, at each (a, b) location. Thus, for optimal JSR-based detection, we focus on scenarios in which the optimal detector is a quadratic function of the observations. And, in order to exploit the degrees of freedom in a JSR (the ability to realize a different detector at each (a, b) location) we resort to composite hypothesis testing scenarios which involve certain nuisance signal parameters.[16]

It is well-known that for detecting a Gaussian signal in Gaussian noise, the optimal test statistic is a quadratic function of the observations [25, 32]. The *locally optimal* test statistic for detecting a weak arbitrary second-order signal in Gaussian noise is also quadratic [25, 32]. We thus focus attention on these

[16] One of the first time-frequency formulations of optimum detection were proposed in [30, 31]. However, these equivalent formulations of classical quadratic detectors do not exploit the degrees of freedom available in TFRs, and can be realized more efficiently in other domains [8, 6].

two detection scenarios[17] and assume that both s and n are zero-mean and independent of each other, n is a complex Gaussian processes characterized by the correlation function $R_n(t_1, t_2) = E[n(t_1)n^*(t_2)]$, and s has finite second-order statistics characterized by $R_s(t_1, t_2)$; that is, $E[n(t_1)n(t_2)] = E[s(t_1)s(t_2)] = 0$ for all $t_1, t_2 \in T$.[18]

Case I. Arbitrary Gaussian signal in white Gaussian noise. $a = 1$, s is Gaussian, and n is white with its real and imaginary parts independent and of equal power spectral densities so that $R_n(t_1, t_2) = N_0\delta(t_1 - t_2)$, where $\delta(t)$ is the Dirac delta function. The optimum test statistic is given by [25]

$$L_O(x) = \frac{1}{N_0} \left\langle \mathbf{R}_s \left(\mathbf{R}_s + N_0\mathbf{I}\right)^{-1} x, x \right\rangle \ , \tag{58}$$

$\mathbf{I}$ is the identity operator, and $\mathbf{R}_s$ denotes the linear operator defined by the corresponding correlation function R_s as

$$(\mathbf{R}_s x)(t) \equiv \int R_s(t, \tau) x(\tau) d\tau \ . \tag{59}$$

Case II. Weak second-order signal in arbitrary Gaussain noise (locally optimal). In the case of arbitrary Gaussian noise, for reasons that will become clear later[19], we do not use the optimal detector based on the likelihood ratio. Instead, we consider the weak signal scenario in which $a > 0$ is sufficiently small so that the SNR[20] is low. In this case, we use the locally optimal test statistic [25] which has the following form even for arbitrary second-order stochastic signals (not necessarily Gaussian)

$$L_{LO}(x) = \left\langle \mathbf{R}_s \mathbf{R}_n^{-1} x, \mathbf{R}_n^{-1} x \right\rangle - \mathrm{Trace}\left(\mathbf{R}_n^{-1} \mathbf{R}_s\right) \tag{60}$$

where $\mathrm{Trace}(\cdot)$ denotes the trace of an operator (sum of the eigenvalues).

We note that the first term of the above test statistic, $\left\langle \mathbf{R}_s \mathbf{R}_n^{-1} x, \mathbf{R}_n^{-1} x \right\rangle$, can also be interpreted as the quadratic test statistic that maximizes deflection[21] H [25, 8]

$$H(L) = \frac{\left(E_1[L] - E_0[L]\right)^2}{E_0[L^2] - E_0^2[L]} \tag{61}$$

in the case of detecting an arbitrary second-order signal in Gaussian noise, where E_i denotes the expectation given that the i-th hypothesis is true.

4.2 Composite hypotheses

Recall that a JSR, $(\mathbf{P}x)(a, b)$, can realize a quadratic function of the observation, x, at each (a, b) location. To exploit these degrees of freedom, we consider composite hypothesis testing scenarios in which, under H_1, the signal, s, has two parameters that may be associated with the variables of the JSR; that is, the two hypotheses are

$$\begin{aligned} H_0 &: \ x(t) = n(t) \\ H_1 &: \ x(t) = as(t; \alpha, \beta) + n(t) \end{aligned} \tag{62}$$

[17] Our development will show that certain composite hypothesis generalizations of these scenarios are naturally suited to JSR-based detectors.

[18] Such processes exist [33] and are sometimes referred to as "circular." We also assume that perfect detection is not possible which, in particular, implies that the test statistics (58) and (60) exist [34].

[19] See footnote 24.

[20] For example, the ratio of the largest eigenvalue of aR_s to that of R_n

[21] Deflection is a measure of SNR, and hence a good indicator of detector performance [32, 25].

where (α,β) are the nuisance signal parameters that may be random or unknown. Under our assumptions, the dependence of $s(t;\alpha,\beta)$ on (α,β) is completely characterized by the signal correlation function which we denote by $R_s^{(\alpha,\beta)}$. Optimal test statistics for such composite hypotheses are described next.

Unknown parameters. If the parameters (α,β) are modeled as deterministic but unknown, we use a generalized likelihood ratio test (GLRT) [25] corresponding to the maximum likelihood (ML) estimate of (α,β).[22] The optimal ML detectors in the two cases are given by

$$L_{ML}(x) = \begin{cases} \max_{(\alpha,\beta)} L_O^{(\alpha,\beta)}(x) & \text{Case I} \\ \\ \max_{(\alpha,\beta)} L_{LO}^{(\alpha,\beta)}(x) & \text{Case II (locally optimal)} \end{cases}, \tag{63}$$

where the superscript (α,β) denotes the test statistic corresponding to the signal correlation function $R_s^{(\alpha,\beta)}$.

Random parameters. For situations in which the parameters (α,β) are random with known joint pdf, $p(\alpha,\beta)$, the optimal test statistic is difficult to compute analytically. Thus, in this case we propose the following suboptimal "MAP GLRT detectors" [8] in which the maximum *a posteriori* probability (MAP) estimate of (α,β) is used in the GLRT:

$$L_{MAP}(x) = \begin{cases} \max_{(\alpha,\beta)} \left[L_O^{(\alpha,\beta)}(x) + \log p(\alpha,\beta) \right] & \text{Case I} \\ \\ \max_{(\alpha,\beta)} \left[L_{LO}^{(\alpha,\beta)}(x) + \log p(\alpha,\beta) \right] & \text{Case II (locally optimal)} \end{cases}. \tag{64}$$

4.3 Parameter dependence for JSR-based detectors

How do JSR-based detectors exactly relate to composite quadratic detection? From (63) and (64), and from the expressions for the test statistics in (58) and (60), we see that in all cases the component of the test statistic that is a function of the observed signal is of the form

$$L^{(\alpha,\beta)}(x) = \left\langle \mathbf{Q}^{(\alpha,\beta)} x, x \right\rangle \tag{65}$$

where $\mathbf{Q}^{(\alpha,\beta)}$ is a nonnegative definite operator (a function of $R_s^{(\alpha,\beta)}$ and R_n). Since (65) is a quadratic form, if the parameters (α,β) could be identified as the variables (a,b) of a class of JSRs, the various detectors could be easily and efficiently realized using those JSRs. So, the question is how should the signal s depend on the nuisance parameters (α,β) so that they naturally correspond to the JSR variables (a,b)?

The covariance-based formulation is canonical for characterizing the appropriate signal dependence on the parameters. By equating the quadratic form (65) for the test statistics with the characterizing quadratic form (13) for covariance-based JSRs, we find that for a given class of JSRs, characterized by the unitary operator $\mathbf{U}_{(a,b)}$ in (13), the quadratic test statistics can be naturally realized by the JSRs if and only if the operator $\mathbf{Q}^{(\alpha,\beta)}$ in (65) is of the form

$$\mathbf{Q}^{(a,b)} = \mathbf{U}_{(a,b)} \phi \mathbf{U}_{(a,b)}^{-1} \tag{66}$$

for some nonnegative definite linear operator ϕ.[23]

We note that since all bilinear JSRs can be expressed as $(\mathbf{P}s)(a,b) = \langle \mathbf{K}^{(a,b)} s, s \rangle$ for some parameterized linear operator $\mathbf{K}^{(a,b)}$, it might appear at first glance that by equating $\mathbf{K}^{(a,b)}$ with $\mathbf{Q}^{(a,b)}$, we could characterize the required signal dependence on the composite parameters for any class of JSR-based detectors.

[22] In a GLRT, an estimate of the parameters (a,b) is first formed, and then used in the likelihood ratio for the corresponding value of parameters.

[23] It follows from (13) that $(\mathbf{P}x)(a,b;\phi) \geq 0$ for all $x \Longleftrightarrow \phi$ nonnegative definite.

However, this is not true in general. The reason is that equating $\mathbf{K}^{(a,b)}$ with $\mathbf{Q}^{(a,b)}$ results in the requisite parameter dependence for the signal correlation function, but translating it into the dependence of the signal itself is fairly complicated, in general, and does not yield useful, meaningful answers. The covariance-based approach, however, naturally yields a meaningful signal model as evident from our discussion in the next section.

4.4 Signal model for JSR-based detectors

Recall from the form of the test statistics (58) and (60), that the operator $\mathbf{Q}^{(a,b)}$ is a function of the signal correlation function, $R_s^{(a,b)}$, and the noise correlation function R_n. It can be verified that the required form for $\mathbf{Q}^{(a,b)}$ in (66) translates into an exactly similar form for $\mathbf{R}_s^{(a,b)}$[24]; that is,

$$\mathbf{R}_s^{(a,b)} \equiv \mathbf{U}_{(a,b)} \mathbf{R}_0 \mathbf{U}_{(a,b)}^{-1} \tag{67}$$

for some correlation function operator $\mathbf{R}_0$, or, equivalently,

$$R_s^{(a,b)}(t_1, t_2) = \sum_k \lambda_k \, \left(\mathbf{U}_{(a,b)} u_k\right)(t_1) \left(\mathbf{U}_{(a,b)} u_k\right)^* (t_2) \ , \tag{68}$$

in terms of the eigenfunctions u_k's and the corresponding nonnegative eigenvalues λ_k's of the correlation function R_0. The above expression for $R_s^{(a,b)}$ implies that the signal in (62) is of the form

$$s(t; a, b) \equiv \left(\mathbf{U}_{(a,b)} s^{(a,b)}\right)(t) \ , \tag{69}$$

where, for each (a, b), $s^{(a,b)}$ is any zero-mean Gaussian (or arbitrary second-order) signal with correlation function R_0. In particular, $s(t; a, b)$ could be one fixed signal (with correlation function R_0), say s_o, unitarily transformed by $\mathbf{U}_{(a,b)}$; that is, $s(t; a, b) = (\mathbf{U}_{(a,b)} s_o)(t)$.

4.5 JSR-based realization of the optimal detectors

Parameterized signal correlation functions of the form (68) characterize the composite hypothesis testing scenarios for which $\mathbf{U}_{(a,b)}$-covariant JSRs provide a canonical realization of the optimal quadratic detectors. We now explicitly derive the forms of the JSR-based detectors for the various cases discussed in Section 4.2. Substituting $\mathbf{R}_s^{(a,b)}$ for $\mathbf{R}_s$ in (58) and (60), and using the form for $\mathbf{R}_s^{(a,b)}$ in (67), we find that the two parameterized test statistics can be realized using covariance-based JSRs as

$$\begin{aligned}
L_O^{(a,b)}(x) &= \frac{1}{N_0} \left\langle \mathbf{R}_0 \left(\mathbf{R}_0 + N_0 \mathbf{I}\right)^{-1} \mathbf{U}_{(a,b)}^{-1} x, \mathbf{U}_{(a,b)}^{-1} x \right\rangle = (\mathbf{P}x) \left(a, b; \phi = \frac{1}{N_0} \mathbf{R}_0 \left(\mathbf{R}_0 + N_0 \mathbf{I}\right)^{-1}\right) \ , &(70) \\
L_{LO}^{(a,b)}(x) &= \left\langle \mathbf{R}_0 \mathbf{U}_{(a,b)}^{-1} \mathbf{R}_n^{-1} x, \mathbf{U}_{(a,b)}^{-1} \mathbf{R}_n^{-1} x \right\rangle - \text{Trace}\left(\mathbf{R}_n^{-1} \mathbf{R}_s^{(a,b)}\right) \\
&= \left(\mathbf{P} \mathbf{R}_n^{-1} x\right)(a, b; \phi = \mathbf{R}_0) - \sum_k \lambda_k \left\langle \mathbf{R}_n^{-1} \mathbf{U}_{(a,b)} u_k, \mathbf{U}_{(a,b)} u_k \right\rangle \ . &(71)
\end{aligned}$$

The above realizations of the test statistics provide the complete JSR-based description of all the ML and MAP GLRT detectors discussed in Section 4.2. The results for the detection framework are summarized in the following proposition.

[24] In conjunction with a preprocessing by $\mathbf{R}_n^{-1}$ in Case II. We also note that the optimal test statistic for Gaussian signal in arbitrary Gaussian noise does not yield the simple signal model (67) that is amenable to JSR-based realization. This is the reason for making the weak signal assumption and using the locally optimal test statistic in Case II, which, incidentally, does not require the signal to be Gaussian either.

Proposition. JSR-based detection framework. *Consider the class C of JSRs covariant to the family of unitary operators $\{\mathbf{U}_{(a,b)} : (a,b) \in G\}$ in the sense of (12). Then, the class of composite hypothesis testing problems of the form (62), for which JSRs from C constitute canonical detectors, are characterized by the family of signal correlation functions (with $(\alpha, \beta) = (a,b)$ in (62))*

$$\mathbf{R}_s^{(a,b)} \equiv \mathbf{U}_{(a,b)} \mathbf{R}_0 \mathbf{U}_{(a,b)}^{-1} \ , (a,b) \in G \ , \tag{72}$$

for some fixed correlation function R_0. The correlation model (72) is equivalent to the signal model

$$s(t; a, b) = \left(\mathbf{U}_{(a,b)} s^{(a,b)} \right) (t) \ , \tag{73}$$

where, for each (a,b), $s^{(a,b)}$ is any zero-mean Gaussian (Case I), or arbitrary second-order (Case II), signal with correlation function R_0. Moreover, the corresponding ML and MAP GLRT detectors discussed in Section 4.2 admit the JSR-based realization

$$L(x) = \max_{(a,b)} \left[(\mathbf{P}y)(a,b;\phi) + F(a,b) \right] \ , \tag{74}$$

where

$$y = \begin{cases} x & \text{Case I} \\ \mathbf{R}_n^{-1} x & \text{Case II} \end{cases} \ , \tag{75}$$

$$F(a,b) = \begin{cases} 0 & \text{for ML detectors : Case I} \\ \mathrm{Trace}\left(\mathbf{R}_n^{-1} \mathbf{R}_s^{(a,b)} \right) & \text{for ML detectors : Case II} \\ \log p(a,b) & \text{for MAP GLRT detectors : Case I} \\ \log p(a,b) + \mathrm{Trace}\left(\mathbf{R}_n^{-1} \mathbf{R}_s^{(a,b)} \right) & \text{for MAP GLRT detectors : Case II} \end{cases} \ , \tag{76}$$

and the operator ϕ characterizing the JSR, $(\mathbf{P}y)(a,b;\phi)$, is given by

$$\phi = \begin{cases} \frac{1}{N_0} \mathbf{R}_0 \left(\mathbf{R}_0 + N_0 \mathbf{I} \right)^{-1} & \text{Case I} \\ \\ \mathbf{R}_0 & \text{Case II} \end{cases} \ . \tag{77}$$

We note that $\mathrm{Trace}\left(\mathbf{R}_n^{-1} \mathbf{R}_s^{(a,b)} \right)$ *in (76) can be alternatively expressed as*

$$\mathrm{Trace}\left(\mathbf{R}_n^{-1} \mathbf{R}_s^{(a,b)} \right) = \sum_k \lambda_k \left\langle \mathbf{R}_n^{-1} \mathbf{U}_{(a,b)} u_k, \mathbf{U}_{(a,b)} u_k \right\rangle = \sum_k \lambda_k \left(\mathbf{P} u_k \right) \left((a,b)^{-1}; \phi = \mathbf{R}_n^{-1} \right) \ , \tag{78}$$

where the u_k's and the λ_k's are the eigenfunctions and eigenvalues of R_0.

Example: Detection framework based on the hyperbolic class [21]. The hyperbolic class is covariant to "hyperbolic time-shifts" and scales changes: $\mathbf{U}_{(a,c)} = \mathbf{H}_a \mathbf{D}_c$ where the hyperbolic time-shift operator, $\mathbf{H}_a$, is defined in the frequency domain (for analytic signals) as $(\mathbf{H}_a \mathbb{F} s)(f) \equiv e^{-j2\pi a \ln(f)} (\mathbb{F} s)(f)$. Thus, the TFRs from hyperbolic class form canonical detectors for random signals transformed by $\mathbf{U}_{(a,c)}$; that is, the corresponding family of *spectral* correlation functions is of the form

$$\begin{aligned} \mathbf{S}_s^{(a,c)} &= \mathbf{U}_{(a,c)} \mathbf{S}_0 \mathbf{U}_{(a,c)}^{-1} \iff \\ S_s^{(a,c)}(f_1, f_2) &\equiv E\left\{ (\mathbb{F} s(a,c))(f_1) (\mathbb{F} s(a,c))^*(f_2) \right\} = \sum_k c \mu_k e^{-j2\pi a \ln(f_1/f_2)} V_k(cf_1) V_k^*(cf_2) \end{aligned} \tag{79}$$

where the μ_k's are the eigenvalues and the V_k's are the eigenfunctions (in the spectral domain) of the spectral correlation function S_0. This framework should be contrasted with the results presented in [7] in which the detection of *fixed* Gaussian signals (with no unknown parameters) is equivalently formulated using the hyperbolic class[25], and detection and parameter estimation of hyperbolic chirps is discussed.[26] Our framework, based on the above proposition, provides a natural setting for optimal detection of arbitrary nonstationary random signals with unknown (or random) scale and hyperbolic time-shift parameters.

Similarly, the power class of TFRs [36] can be used for optimal detection of random signals that have undergone scale changes and "power time-shifts". Optimal detection frameworks based on Cohen's class and the affine class, corresponding to the unitary operators $\mathbf{U}_{(t,f)} = \mathbf{T}_t\mathbf{F}_f$ and $\mathbf{U}_{(t,c)} = \mathbf{T}_t\mathbf{D}_c$, respectively, were developed in [6, 8].

4.6 Discussion

Composite quadratic hypothesis testing provides a general framework for optimal detection of random signals with a few uncertain nuisance parameters. Generalized JSR-based detectors impose a structure on the nuisance parameters that is relevant from a practical viewpoint, and provide a natural setting for implementing the optimal detectors based on a GLRT.

The Proposition in the last section encapsulates the main results of the JSR-based detection framework. We showed that in the composite hypothesis testing framework, the covariance-based method is natural for characterizing the JSR-based detectors. Moreover, not surprisingly, the unitary transform $\mathbf{U}_{(a,b)}$, that characterizes a particular class of covariant-based JSRs, also characterizes the class of random signals for which such JSRs form canonical detectors; (a, b) constitute the nuisance signal parameters. Essentially, JSRs characterized by a parameterized unitary transform $\mathbf{U}_{(a,b)}$, as in (13), are naturally suited for detecting random signals that have been transformed by the same unitary transform, with the parameters either unknown or random.

Expressing the JSRs in terms of the eigenexpansion of the linear operator ϕ (see (77)) yields more insight into the structure of the JSR-based detectors:

$$(\mathbf{P}y)(a, b; \phi) = \sum_k \mu_k \left|\left\langle \mathbf{U}_{(a,b)} u_k, y \right\rangle\right|^2 , \tag{80}$$

where the eigenfunctions u_k's of ϕ are the same as those of R_0 and the eigenvalues μ_k's are

$$\mu_k = \begin{cases} \frac{\lambda_k}{N_0(\lambda_k + N_0)} & \text{Case I} \\ \lambda_k & \text{Case II} \end{cases} , \tag{81}$$

where the λ_k's are the eigenvalues of R_0 and N_0 is the noise power in Case I. We first note that the linear transform

$$(\mathbf{T}s)(a, b; g) \equiv \left\langle \mathbf{U}_{(a,b)} g, s \right\rangle \tag{82}$$

defined by the unitary operator $\mathbf{U}_{(a,b)}$ and the "window" function, g, generalizes the concept of the short-time Fourier transform (STFT) [1, 37] and the wavelet transform (WT) [38, 37]. $\mathbf{T}$ in (82) defines a class of linear transforms (characterized by g and $\mathbf{U}_{(a,b)}$) which analyze the joint a-b content in the signal and which are covariant to a-b shifts (covariant to the operator $\mathbf{U}_{(a,b)}$), just as the STFT and WT are covariant to time-frequency and time-scale shifts, respectively. Moreover, the linear transform $\mathbf{T}$ is the optimal matched filter for detecting the unitarily transformed deterministic signal, $g_{(a,b)} \equiv \mathbf{U}_{(a,b)} g$, in additive Gaussian noise. The JSR detector in (80) is then simply a weighted sum of the magnitude-squared outputs of a bank of linear a-b transforms corresponding to the eigenfunctions u_k's.[27] This yields a subspace-based interpretation of

[25] As an extension of [31].
[26] An extension to generalized chirps is presented in [35].
[27] This structure in terms of linear transforms also suggests a natural implementation.

the JSR-based detectors: Essentially, (80) shows that the detection of any signal is accomplished by taking a weighted (nonlinear) projection onto the subspace spanned by the eigenfunctions of the signal correlation function. If the signal correlation function is rank-1, then the quadratic detector is effectively reduced to the magnitude-squared output of a linear detector (matched filter). The effect of the unknown or random parameters is taken into account by using the ML or MAP estimates of the parameters. In fact, the point (a, b) at which the maximum occurs in (74) is actually the ML/MAP estimate of the parameters.

Unitary equivalence. We now briefly discuss the relationship between classes of JSR-based detectors that are unitarily equivalent to each other. That is, given a class, C, of covariance-based JSRs, covariant to the family of unitary operators $\mathbf{U}_{(a,b)}$, we are interested in characterizing the detectors corresponding to any unitarily equivalent class, C', as defined in (53), for a given preprocessing transformation $\mathbf{U}$ and postprocessing transformation $\mathbf{V}$. We restrict $\mathbf{V}$ to be an axis-warping transformation because the purpose of a postprocessing transformation is mostly to remap the arbitrary coordinates into time and frequency.

We first analyze the effect of $\mathbf{U}$. If a JSR, $\mathbf{P}$, is covariant to $\mathbf{U}_{(a,b)}$, the JSR $\mathbf{PU}$ is covariant to $\widehat{\mathbf{U}}_{(a,b)} = \mathbf{U}^{-1}\mathbf{U}_{(a,b)}\mathbf{U}$ [12, 16]. Also, the operators $\widehat{\mathbf{U}}_{(a,b)}$ satisfy the same group composition law as $\mathbf{U}_{(a,b)}$:

$$\widehat{\mathbf{U}}_{(a,b)}\widehat{\mathbf{U}}_{(a',b')} = \mathbf{U}^{-1}\mathbf{U}_{(a,b)}\mathbf{U}_{(a',b')}\mathbf{U} = \widehat{\mathbf{U}}_{(a,b)\bullet(a',b')} \ . \tag{83}$$

Thus, the JSRs from the class $C' = C\mathbf{U}$ constitute canonical detectors for random signals with correlation functions of the form

$$\mathbf{R}_s^{(a,b)} = \widehat{\mathbf{U}}_{(a,b)}\mathbf{R}_0\widehat{\mathbf{U}}_{(a,b)}^{-1} \ , \tag{84}$$

or, equivalently,

$$R_s^{(a,b)}(t_1, t_2) = \sum_k \lambda_k \left(\mathbf{U}^{-1}\mathbf{U}_{(a,b)}\mathbf{U}u_k\right)(t_1)\left(\mathbf{U}^{-1}\mathbf{U}_{(a,b)}\mathbf{U}u_k\right)^*(t_2) \ . \tag{85}$$

That is, whereas the JSRs from C are canonical for detecting random signals that have been transformed by $\mathbf{U}_{(a,b)}$, the JSRs from $C' = C\mathbf{U}$ are naturally suited for detecting signals that have been transformed by $\widehat{\mathbf{U}}_{(a,b)} = \mathbf{U}^{-1}\mathbf{U}_{(a,b)}\mathbf{U}$, the unitary transform that characterizes the class C'.

As far as the effect of $\mathbf{V}$ is concerned, the class $C' = \mathbf{V}C$ is still covariant to the operator $\mathbf{U}_{(a,b)}$, albeit with a covariance relation that is induced from (12) via $\mathbf{V}$ [14, 17, 18]. And thus, the underlying class of signals for which the JSRs constitute canonical detectors does not change; only the relationship between the parameters of the signal and the variables of the JSR is modified in an invertible fashion. It follows that the detection characteristics of the C'-based detectors are also the same as those of the original C-based detectors. Hence, the effect of the postwarping transformation, $\mathbf{V}$, is of no consequence in the detection framework.

5 Conclusions

Spurred by the interest in time-frequency-based techniques for nonstationary signal processing, recent research in time-frequency analysis has produced two major advances: a substantial theory for statistically optimal time-frequency techniques and a comprehensive theory for generalized joint signal representations. The theoretical developments in statistical time-frequency analysis provide new time-frequency methods that go beyond merely exploratory data analysis and, for the first time, enable TFRs to be fully exploited in real situations. The theory of generalized JSRs facilitates nonstationary signal processing that, compared to conventional time-frequency analysis, can be well-matched to a broader class of nonstationary signals. This paper fuses the two advances by developing optimal detection and estimation techniques based on generalized JSRs, thereby producing a unified theory applicable to a wide variety of problems in nonstationary statistical signal processing.

By virtue of the estimation theory, any bilinear JSR of a signal of interest can be optimally estimated from an observed snapshot that is corrupted by random noise and/or interference. The signal of interest may itself be random, and the theory facilitates optimal estimation of both the expected value of the JSR (average statistics) and its value for each realization (realization-based statistics). Moreover, since the theory applies to *any* class of bilinear generalized JSRs, it provides optimal estimation of virtually any type of *quadratic* nonstationary statistics. Because of the demonstrated interference-suppression ability of TFR-based estimators [3, 39], we expect the generalized JSR-based estimators to possess that ability too.

The JSR-based detection theory provides optimal detection of noise-corrupted random signals that have undergone certain parameterized unitary transformations; the unitary transformations may model a variety of nondissipative signal distortions that give rise to a few uncertain signal parameters. Each such family of unitary transforms defines, on one hand, a class of JSRs covariant to the unitary transforms, and on the other, a corresponding class of random signals for which those JSRs constitute canonical detectors. Such random signals, with a few random or unknown nuisance parameters (determined by the unitary transform), can serve as useful models in many applications such as machine fault diagnostics and biomedical signal classification. The special cases of TFR- and TSR-based detectors, that are also directly applicable in narrowband and wideband radar, respectively, were discussed in [8, 6].

The form of the quadratic JSR-based detectors yields a very interesting and useful subspace-based interpretation in terms of corresponding linear JSRs. Following the approach in [8, 6], this interpretation can be exploited to design optimal detectors based on partial signal information when only the eigenfunctions of the signal correlation function are known: ML estimates of the eigenvalues are used to determine such optimal detectors.

As is the case in most statistically optimal methods, the techniques developed in this paper require exact knowledge of certain statistics: Certain second- and fourth-order statistics for estimation, and second-order statistics for detection. Due to the nonstationary nature of the signals involved, we cannot resort to asymptotic results as in the stationary case. Thus, an immediate issue of critical importance is the estimation of such statistics when they are not known *a priori*, or the development of techniques that do not assume such *a priori* knowledge.[28] A particular scenario of significant practical importance, in which such issues may be successfully handled, is when multiple realizations are available.

Finally, although the detection and estimation theories have been developed independently, it is conceivable that in more complex problems involving detection in the presence of a multitude of interference signals and noise, a fusion of the two techniques might yield promising results by preprocessing the observed data using estimation techniques to suppress interference, and then applying the detection procedures. Such, possibly suboptimal, combination of detection and estimation techniques may be necessary, and possibly fruitful, in complex real-world applications.

References

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

[28] For example, in the estimation framework, information only about the support of the various quantities, characterizing the optimal solution, may be used to design practical estimators. One such technique for TFR estimation in the presence of white noise is proposed in [5].

[2] A. M. Sayeed and D. L. Jones, "Optimal kernels for Wigner-Ville spectral estimation," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '94*, pp. IV–297–300, 1994.

[3] A. M. Sayeed and D. L. Jones, "Optimal kernels for nonstationary spectral estimation," *IEEE Trans. Signal Processing*, vol. 43, pp. 478–491, February 1995.

[4] W. Kozek, "Optimally Karhunen-Loève-like STFT expansion of nonstatonary processes," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '93*, (Minneapolis, MN), pp. IV–428–IV–431, 1993.

[5] W. Kozek and K. Riedel, "Quadratic time-varying spectral estimation for underspread processes," in *Proc. IEEE Int'l Symp. on Time-Frequency and Time-Scale Analysis*, pp. 460–463, 1994.

[6] A. M. Sayeed and D. L. Jones, "Optimal quadratic detection using bilinear time-frequency and time-scale representations," in *Proc. IEEE Int'l Symp. on Time-Frequency and Time-Scale Analysis*, pp. 365–368, 1994.

[7] A. Papandreou, S. M. Kay, and G. F. Boudreaux-Bartels, "The use of hyperbolic time-frequency representations for optimum detection and parameter estimation of hyperbolic chirps," in *Proc. IEEE Int'l Symp. on Time-Frequency and Time-Scale Analysis*, pp. 369–372, 1994.

[8] A. M. Sayeed and D. L. Jones, "Optimal detection using bilinear time-frequency and time-scale representations," *IEEE Trans. Signal Processing*, vol. 43, pp. 2872–2883, December 1995.

[9] B. Samimy, G. Rizzoni, A. M. Sayeed, and D. L. Jones, "Design of training data-based quadratic detectors with application to mechanical systems," *To be presented at the IEEE Int. Conf. on Acoust., Speech and Signal Proc. – ICASSP '96*, 1996.

[10] A. M. Sayeed, P. Lander, and D. L. Jones, "Improved time-frequency filtering of signal-averaged electrocardiograms," *Journal of Electrocardiology*, vol. 28, pp. 53–58, Supplement 1995.

[11] L. Cohen, "The scale representation," *IEEE Trans. Signal Processing*, vol. 41, pp. 3275–3292, December 1993.

[12] R. G. Baraniuk and D. L. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. Signal Processing*, vol. 43, pp. 2269–2282, October 1995.

[13] 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.

[14] R. G. Baraniuk, "Marginals vs. covariance in joint distribution theory," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '95*, vol. 2, pp. 1021–1024, 1995.

[15] 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.

[16] 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*, June 1996.

[17] F. Hlawatsch and H. Bölcskei, "Unified theory of displacement-covariant time-frequency analysis," in *Proc. IEEE Int'l Symp. on Time-Frequency and Time-Scale Analysis*, pp. 524–527, 1994.

[18] 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.

[19] 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*, April 1996.

[20] O. Rioul and P. Flandrin, "Time-scale distributions: A general class extending the wavelet transform," *IEEE Trans. Signal Processing*, vol. 40, pp. 1746–1757, May 1992.

[21] 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.

[22] R. G. Baraniuk, "A limitation of the kernel method for joint distributions of arbitrary variables," *IEEE Signal Processing Letters*, pp. 51–53, February 1996.

[23] A. M. Sayeed, "On the equivalence of the operator and kernel methods for joint distributions of arbitrary variables," *Submitted to IEEE Trans. Signal Processing*, November 1995.

[24] A. Grossman, J. Morlet, and T. Paul, "Transforms associated to square integrable group representations. Part 1. General results," *J. Math. Phys.*, vol. 26, pp. 2473–2479, October 1985.

[25] H. V. Poor, *An Introduction to Signal Detection and Estimation.* Springer-Verlag, 1988.

[26] W. Rudin, *Fourier analysis on groups.* New York: Interscience Publishers, 1962.

[27] W. Martin, "Time-frequency analysis of random signals," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '82*, pp. 1325–1328, 1982.

[28] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of non-stationary processes," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 33, pp. 1461–1470, December 1985.

[29] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications.* San Francisco, CA: Holden-Day, 1968, p. 275.

[30] R. A. Altes, "Detection, estimation and classification with spectrograms," *J. Acoust. Soc. Am.*, vol. 67, pp. 1232–1246, April 1980.

[31] P. Flandrin, "A time-frequency formulation of optimum detection," *IEEE Trans. Signal Processing*, vol. 36, pp. 1377–1384, September 1988.

[32] B. Picinbono and P. Duvaut, "Optimal linear-quadratic systems for detection and estimation," *IEEE Trans. Inform. Theory*, vol. 43, pp. 304–311, March 1988.

[33] K. S. Miller, *Complex Stochastic Processes.* Addison-Wesley, 1974.

[34] E. Parzen, "Probability density functionals and reproducing kernel Hilbert spaces," in *Time Series Analysis*, (John Wiley & Sons, Inc., New York), pp. 155–169, (M. Rosenblatt, Ed.), 1963.

[35] A. Papandreou, G. F. Boudreaux-Bartels, and S. M. Kay, "Detection and estimation of generalized chirps using time-frequency representations," in *Proc. 28th Asilomar Conference on Signals, Systems, and Computers*, pp. 50–54, October 1994.

[36] F. Hlawatsch, A. Papandreou, and G. F. Boudreaux-Bartels, "The power class of quadratic time-frequency representations: A generalization of the affine and hyperbolic classes," in *Proc. 27th Asilomar Conference on Signals, Systems, and Computers*, pp. 1265–1270, 1993.

[37] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, April 1992.

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

[39] A. M. Sayeed and D. L. Jones, "Optimal estimation of time-frequency representations from corrupted observations," in *Proc. IEEE Int'l Symp. on Time-Frequency and Time-Scale Analysis*, pp. 456–459, 1994.