Data-Driven Time-Frequency and Time-Scale Detectors

Akbar M. Sayeed

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

Presented at SPIE'97

ABSTRACT

In many practical signal detection problems, the detectors have to designed from training data. Due to limited training data, which is usually the case, it is imperative to exploit some inherent signal structure for reliable detector design. The signals of interest in a variety of applications manifest such structure in the form of nuisance parameters. However, data-driven design of detectors by exploiting nuisance parameters is virtually impossible in general due to two major difficulties: *identifying* the appropriate nuisance parameters, and *estimating* the corresponding detector statistics. We address this problem by using recent results that relate joint signal representations (JSRs), such as time-frequency and time-scale representations, to quadratic detectors for a wide variety of nuisance parameters. We propose a general data-driven framework that: 1) identifies the appropriate nuisance parameters from an arbitrarily chosen finite set, and 2) estimates the second-order statistics that characterize the corresponding JSR-based detectors. Simulation results demonstrate that for limited training data, exploiting the structure of nuisance parameters via our framework can deliver substantial gains in performance over empirical detectors which ignore such structure.

Keywords: Detection, training, nuisance parameters, Bayesian detection, generalized likelihood ratio test

1. INTRODUCTION

Optimal detection of signals in the presence of noise requires the knowledge of certain underlying statistics. In most practical problems, those statistics are not known *a priori* and must be estimated from training data. Moreover, only limited training data is usually available.

Due to limited training data, reliable detector design is possible only if the complexity of the detector is sufficiently low. This may be possible if the signal of interest possesses some inherent structure that can be exploited. In many practical scenarios, such structure is manifested in terms of nuisance parameters. For example, the radar returns from an object can be modeled as a signal exhibiting unknown delay and Doppler parameters. The received signal appears to have a much higher complexity (degrees of freedom) when viewed over an ensemble, without the knowledge of the delay-Doppler nuisance parameters. On the other hand, given the delay-Doppler values, the signal has much fewer intrinsic degrees of freedom. By exploiting the structure of nuisance parameters, the effective degrees of freedom in the data can be reduced, which can then be leveraged into reliable detector design.

In order to exploit nuisance parameters for data-driven detector design, there are two key issues that have to be addressed:

1. **Identification of parameters.** What is the nature of the nuisance parameters that underlie a given data set?
2. **Estimation of detector statistics.** Given the structure defined by the nuisance parameters, how can the corresponding detectors be designed from training data?

Other author information: Email: akbar@rice.edu, Web: http://www-dsp.rice.edu. Supported by the ONR under grant nos. N00014-95-1-0674 and N00014-95-1-0849, and the NSF under grant no. MIP-9457438.

In this paper, we exploit the structure of joint signal representations (JSRs), such as time-frequency representations (TFRs)[1,2] and time-scale representations (TSRs)[3,4] to address the problem of data-driven detector design. JSRs admit a unified formulation in terms of *parameterized unitary operators* which can be used to model a wide variety of nuisance parameters (and corresponding detectors) that are quite relevant from a practical viewpoint (the operator parameters define the nuisance parameters)[5,6] — radar/sonar, machine fault diagnostics, and biomedical signal classification are some typical applications.[7,6,8,9] For example, TFR-based detectors are appropriate for unknown time- and frequency-offset nuisance parameters and have been successfully applied in machine fault diagnostics.[9,8] Time-scale detectors have shown promise in biomedical applications such as detection of emboli in Doppler ultrasound.[10] More generally, arbitrary JSRs impose a well-defined structure on the space of nuisance parameters for addressing the above-mentioned issues encountered in data-driven detector design.

By exploiting the structure and unified formulation of JSR-based detectors, we propose a general data-driven detection framework in this paper. Using *labeled training data*, our framework:

1. *Identifies* the nuisance parameters, from a finite set chosen *a priori*, that "best fit" the data in a precise sense.
2. *Estimates* the statistics that characterize the detectors corresponding to the "best" nuisance parameters.

For example, our framework can determine whether time- and frequency-shifts, or time-shifts and scale-changes are the appropriate nuisance parameters in a given data set, and design the corresponding (TFR or TSR) detectors.

In the next section, we motivate the nuisance parameter based approach to data-driven detector design. Section 2 provides a brief description of JSRs and interprets them in the context of detection. In Section 4, we describe our data-driven JSR-based framework for designing quadratic detectors. Section 5 illustrates the performance of TFR- and TSR-based algorithms on simulated data. Some concluding remarks are presented in Section 6.

2. SIGNAL STRUCTURE: NUISANCE PARAMETERS

Detection of signals exhibiting nuisance parameters can be formally cast as a composite hypothesis testing problem[11]

$$\begin{aligned} H_1 &: \quad x(t) = s^\theta(t) + n(t) \\ H_0 &: \quad x(t) = n(t) \end{aligned} \tag{1}$$

where $t \in T$, the observation interval, $x(t)$ is the observed waveform, $s^\theta(t)$ is a (nonstationary) stochastic signal with nuisance parameters $\theta \stackrel{def}{=} (\theta_1, \theta_2, \cdots, \theta_M) \in S \subset \mathbb{R}^M$, and $n(t)$ is the additive noise. The optimal detector is a real-valued test statistic $L(x)$, derived from the likelihood ratio, that is compared to an appropriately chosen threshold γ to determine whether the signal is present (H_1) or not (H_0). In data-driven applications, the objective is to design the test statistic $L(x)$ from labeled training data.

In view of limited training data, we focus on scenarios in which $L(x)$ is characterized by second-order statistics. In particular, we assume that the noise is zero-mean, Gaussian and white with variance σ^2, and the signal is zero-mean, Gaussian and with an arbitrary correlation function $R_s^\theta(t_1, t_2) \stackrel{def}{=} \mathrm{E}[s^\theta(t_1) s^{\theta *}(t_2)]$. We further assume the low SNR (signal-to-noise ratio) regime in which the *locally optimal* test statistic is given by[11,7]

$$L^\theta(x) = \left\langle \mathbf{R}_s^\theta x, x \right\rangle \ , \tag{2}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, and $\mathbf{R}_s^\theta$ denotes the operator defined by $R_s^\theta(t_1, t_2)$ as

$$\left(\mathbf{R}_s^\theta x\right)(t) \stackrel{def}{=} \int R_s^\theta(t, \tau) x(\tau) d\tau \ . \tag{3}$$

Given $L^\theta(x)$, two detectors can be defined depending on the nature of the nuisance parameters θ. If θ is modeled as random with probability density function $p_\Theta(\theta)$, the (locally optimal) Bayesian detector is given by

$$L_{\mathrm{B}}(x) \stackrel{def}{=} \mathrm{E}\left[L^\theta(x)\right] = \int p_\Theta(\theta) L^\theta(x) d\theta = \left\langle \mathbf{R}_s x, x \right\rangle \ , \tag{4}$$

where

$$\mathbf{R}_s \stackrel{def}{=} \mathrm{E}\left[\mathbf{R}_s^\theta\right] = \int \mathbf{R}_s^\theta p_\Theta(\theta) d\theta \ . \tag{5}$$

If θ is modeled as deterministic but unknown, the test statistic based on a generalized likelihood ratio test (GLRT) may be used in which an estimate of θ is first formed and then the $L^\theta(x)$ corresponding to the estimate is used as the detector. The maximum likelihood (ML) estimate is often used in practice which results in the GLRT detector

$$L_{\mathrm{G}}(x) \stackrel{def}{=} L^{\hat{\theta}_{ML}}(x) = \max_{\theta \in S} L^\theta(x) \ . \tag{6}$$

Note that for sufficiently smooth $L^\theta(x)$ (as a function of θ), $L_{\mathrm{B}}(x) = L^{\hat{\theta}_B}(x)$ for some $\hat{\theta}_B$, but $\hat{\theta}_B \neq \hat{\theta}_{ML}$, in general.

The detector $L_{\mathrm{B}}(x)$ is characterized by $\mathbf{R}_s$, or, equivalently, by $\{\mathbf{R}_s^\theta, p_\Theta(\theta)\}$, whereas the detector $L_{\mathrm{G}}(x)$ is characterized by $\mathbf{R}_s^\theta$. Thus, data-driven detector design corresponds to estimating the required signal correlation function from training data. In this context, the power of the structure of nuisance parameters comes from the fact that the complexity of $\mathbf{R}_s^\theta$ (for a fixed θ) is typically much lower than the complexity of $\mathbf{R}_s$. For example, for any given value of θ, $\mathbf{R}_s^\theta$ may be rank-1 (corresponding to a deterministic signal with random amplitude), whereas $\mathbf{R}_s$ as given by (5) would be of a much higher rank, in general. The key point to note is that without exploiting the structure of nuisance parameters θ, only $\mathbf{R}_s$ can be estimated (leading to $L_{\mathrm{B}}(x)$), whereas by incorporating the information about the nuisance parameters, we may be able to estimate $\mathbf{R}_s^\theta$ (leading to $L_{\mathrm{G}}(x)$) or $\{\mathbf{R}_s^\theta, p_\Theta(\theta)\}$ (leading to $L_{\mathrm{B}}(x)$). Clearly, for reliable detector design based on limited training data, it is advantageous to estimate the lower complexity $\mathbf{R}_s^\theta$, as opposed to the higher complexity $\mathbf{R}_s$.

The essential idea behind our data-driven framework is to exploit the structure of the nuisance parameters to estimate $\{\mathbf{R}_s^\theta, p_\Theta(\theta)\}$ from which both $L_{\mathrm{B}}(x)$ and $L_{\mathrm{G}}(x)$ can be designed. However, in order to exploit the structure of nuisance parameters, we need to address the *identification* and *estimation* issues mentioned in the previous section. The next section presents a description of JSR-based detection which provides a natural framework for addressing them.

3. MODELING SIGNAL STRUCTURE

JSRs represent signal characteristics jointly in terms of two or more variables or physical quantities, which define the nuisance parameters — for example, time and frequency in TFRs, and time and scale in TSRs. From a detection viewpoint, JSRs possessing certain *covariance* properties are the appropriate tools,[6] and we now provide a brief description of such JSRs.[5,12,13]

Let $G \subset \mathbb{R}^M$ be an M-parameter Lie group[14] and $\{\mathbf{U}_\theta : \theta \in G\}$ be a family of unitary operators that is a unitary representation of G on $L^2(\mathbb{R})$; that is $\mathbf{U}_\theta : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $\langle \mathbf{U}_\theta s_1, \mathbf{U}_\theta s_2 \rangle = \langle s_1, s_2 \rangle$ for all $s_1,\ s_2 \in L^2(\mathbb{R})$, and $\mathbf{U}_\theta \mathbf{U}_{\theta'} = \mathbf{U}_{\theta \bullet \theta'}$ for all $\theta, \theta' \in G$, where $\bullet$ denotes the group operation.[5,15] The "coordinates" of $\theta = (\theta_1, \theta_2 \cdots \theta_M)$ represent the variables of interest, and the unitary operators $\mathbf{U}_\theta$ represent the signal transformations of interest that produce the nuisance parameters, such as time and frequency shifts, scale changes etc.

Consider a given family $\{\mathbf{U}_\theta : \theta \in G\}$. On one hand, $\{\mathbf{U}_\theta\}$ defines a class C of JSRs with respect to the variables $\theta = (\theta_1, \theta_2 \cdots \theta_M)$ via[5,15]

$$(\mathbf{P}s)(\theta; \mathbf{K}) = \langle \mathbf{K}\mathbf{U}_\theta^{-1} s, \mathbf{U}_\theta^{-1} s \rangle \ , \tag{7}$$

where the JSR is denoted by the operator $\mathbf{P}$ that maps $s \in L^2(\mathbb{R})$ into the space of functions defined on G. Different choices of the operator $\mathbf{K} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ yield different JSRs from C. On the other hand, $\{\mathbf{U}_\theta\}$ defines a class of composite hypothesis testing situations of the form (1) characterized by[6]

$$\mathbf{R}_s^\theta = \mathbf{U}_\theta \mathbf{R}_r \mathbf{U}_\theta^{-1} \ , \tag{8}$$

for some "reference" correlation function $\mathbf{R}_r$, where the nuisance parameters correspond to the JSR variables $\theta = (\theta_1, \theta_2, \cdots \theta_M)$. The correlation function model in (8) is equivalent to the signal model

$$s^\theta(t) = (\mathbf{U}_\theta s_r)(t) \ , \tag{9}$$

for some reference signal s_r with correlation function $\mathbf{R}_r$. The key point is that the JSRs from C constitute the test statistics $L^\theta(x)$ for the hypothesis testing problem (1) associated with the signal correlation function (8):[6]

$$L^\theta(x) = (\mathbf{P}x)(\theta; \mathbf{R}_r) \ , \tag{10}$$

where note that the operator characterizing the JSR is $\mathbf{K} = \mathbf{R}_r$. The Bayesian and GLRT detectors can then be realized via (4) and (6).

The JSR-based formulation of $L^\theta(x)$ in (10) admits a useful subspace-based interpretation in terms of the eigenexpansion of $\mathbf{R}_r$

$$R_r(t_1, t_2) = \sum_{k=1}^{N_r} \lambda_k u_k(t_1) u_k^*(t_2) \ , \tag{11}$$

where the $u_k(t)$'s are the orthonormal eigenfunctions and the λ_k's are the corresponding positive eigenvalues of $\mathbf{R}_r$. N_r denotes the rank of $\mathbf{R}_r$ (number of nonzero eigenvalues). Using (11), the JSR-based test statistic (10) becomes[16,5]

$$L^\theta(x) = (\mathbf{P}x)(\theta; \mathbf{R}_r) = \sum_{k=1}^{N_r} \lambda_k \left| \langle \mathbf{U}_\theta x, u_k \rangle \right|^2 \ . \tag{12}$$

The transform

$$(\mathbf{T}x)(\theta; g) \stackrel{def}{=} \langle \mathbf{U}_\theta x, g \rangle \ , \tag{13}$$

denotes a linear JSR with respect to the variables θ,[5,15] and is a generalization of the short-time Fourier transform (STFT)[17,2] and the wavelet transform (WT).[4] The function $g(t)$ serves as a "window" in computing the linear JSR $\mathbf{T}$. Thus, the quadratic JSR-based test statistic admits a series expansion in terms of a bank of linear JSRs corresponding to the eigenfunctions $u_k(t)$'s, as in (12) and illustrated in Figure 1. The eigenfunctions $u_k(t)$'s are the natural modes of the stochastic signal $s^\theta(t)$, and the different linear transforms are matched for processing the natural modes. The effect of the nuisance parameters is accounted for by the variables of the transform.

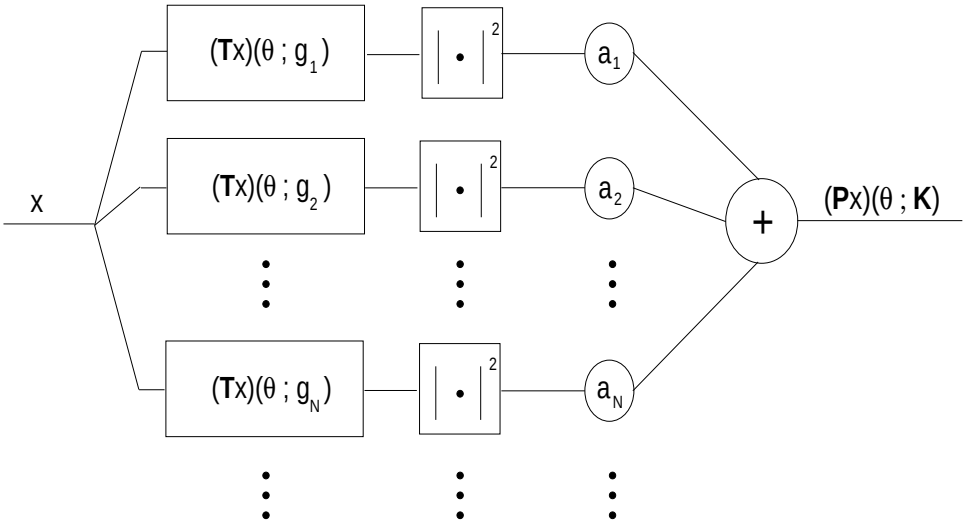

Figure 1. Subspace-based formulation of quadratic JSRs in terms of linear JSRs. In the detection context, $g_k = u_k$ and $a_k = \lambda_k$, in terms of the eigenexpansion of $\mathbf{R}_r$.

3.1. Time-frequency and time-scale representations

We now provide a brief description of two important classes of JSRs: TFRs and TSRs. For Cohen's class of TFRs, the variables (nuisance parameters) are time and frequency shifts; that is, $\theta = (t, f) \in \mathbb{R}^2$ and $(\mathbf{U}_{(t,f)}s)(\tau) =$

$s(\tau - t)e^{j2\pi f\tau}$.[7] TFR-based test statistics admit a subspace-based implementation in terms of a bank of STFTs[18,5]

$$L^{(t,f)}(x) = (\mathbf{P}x)(t, f; \mathbf{R}_r) = \sum_{k=1}^{N_r} \lambda_k \left|(\mathbf{STFT}x)(t, f; u_k)\right|^2 \ , \tag{14}$$

where the STFT is defined as

$$(\mathbf{STFT}x)(t, f; g) \stackrel{def}{=} \int x(\tau)g^*(\tau - t)e^{-j2\pi f\tau}d\tau \ . \tag{15}$$

In Figure 1, the subspace-based formulation of TFRs corresponds to $\theta = (t, f)$ and $\mathbf{T} = \mathbf{STFT}$.

For the affine class of TSRs, the nuisance parameters are time-shifts and scale changes; that is, $\theta = (t, c) \in \mathbb{R} \times (0, \infty)$, and $(\mathbf{U}_{(t,c)}s)(\tau) = \frac{1}{\sqrt{c}}s\left(\frac{\tau - t}{c}\right)$.[7] TSR-based test statistics admit a subspace-based implementation in terms of a bank of WTs[5]

$$L^{(t,c)}(x) = (\mathbf{P}x)(t, c; \mathbf{R}_r) = \sum_{k} \lambda_k \left|(\mathbf{WT}x)(t, c; u_k)\right|^2 \ , \tag{16}$$

where the WT is defined as

$$(\mathbf{WT}x)(t, c; g) \stackrel{def}{=} \frac{1}{\sqrt{c}} \int x(\tau)g^*\left(\frac{\tau - t}{c}\right)d\tau \ . \tag{17}$$

In Figure 1, the subspace-based formulation of TSRs corresponds to $\theta = (t, c)$ and $\mathbf{T} = \mathbf{WT}$.

3.2. Intuition behind the JSR-based approach

The variables of JSRs, corresponding to different choices of $\{G, \mathbf{U}_\theta\}$, can model a wide variety of nuisance parameters that may be encountered in practice. Moreover, JSRs have an inherent structure (determined by $\mathbf{U}_\theta$) that makes them natural for addressing the identification and estimation issues, as elaborated below.

Intuition for estimation. Note from (8) that if $\mathbf{U}_\theta$ is known, we need to estimate $\mathbf{R}_s^\theta$ for only *one* value of θ because of the group structure. The reason is that for any $\theta, \theta' \in G$, $\mathbf{R}_s^\theta$ and $\mathbf{R}_s^{\theta'}$ can be transformed into each other as

$$\mathbf{R}_s^{\theta'} = \mathbf{U}_{\theta' \bullet \theta^{-1}} \mathbf{R}_s^\theta \mathbf{U}_{\theta' \bullet \theta^{-1}}^{-1} \ . \tag{18}$$

Similarly,

$$s^{\theta'} = \mathbf{U}_{\theta' \bullet \theta^{-1}} s^\theta \ , \tag{19}$$

for the same underlying s_r in (9). This suggests a mechanism for estimating $\mathbf{R}_s^{\theta_o}$ at a particular value θ_o by "aligning" different realizations of $s^\theta(t)$, corresponding to different values of θ, to θ_o. From a purely detection viewpoint, the actual value of θ_o does not matter because of the group structure: $\mathbf{R}_s^{\theta_o}$ for any θ_o can serve as the reference $\mathbf{R}_r$ in (8) and (10). Figure 2 illustrates the process of alignment to estimate $\mathbf{R}_r$.

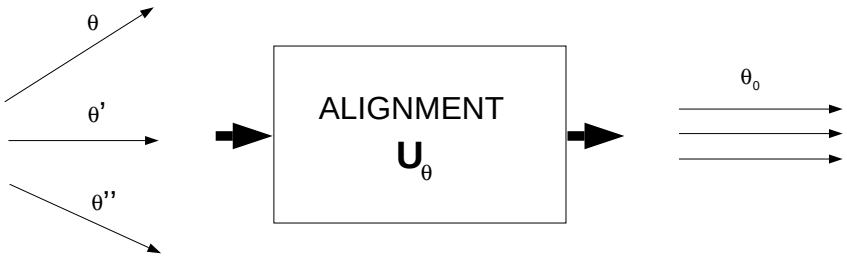

Figure 2. Estimation: reducing the spread of the nuisance parameters in the training data via "alignment" to estimate the underlying correlation function $\mathbf{R}_r$.

Intuition for identification. First of all, from a practical perspective, the choice of nuisance parameters that best describe a data set can only be made from a finite set. Different classes of JSRs, characterized by

different unitary representations $\{\mathbf{U}_\theta\}$ of different groups, can model a wide range of practically relevant nuisance parameters. Depending on the particular application at hand, a finite candidate set of such nuisance parameters can be chosen *a priori*. The "best" nuisance parameters from the set (with a corresponding class of JSRs) are those which result in an estimate of $\mathbf{R}_r$ after "alignment" that has the smallest effective rank N_r. In terms of the subspace-based interpretation, the best nuisance parameters are those which assign the smallest dimension to the underlying stochastic signal $s_r(t)$. The intuition behind the criterion is that, in general, the "unaligned" signal correlation function is of a higher rank than $\mathbf{R}_r$ due to nuisance parameters. Moreover, a mismatch of parameters in "alignment" will also in general lead to a higher rank estimate of $\mathbf{R}_r$. Figure 3 illustrates the idea behind the identification procedure. For quadratic detection, *deflection*[11,7] captures the notion of smallest rank, and is also a measure of SNR. Thus, the class of JSRs that corresponds to the "best" nuisance parameters can be defined to be the one for which the corresponding estimate of $\mathbf{R}_r$ yields the largest deflection[11]

$$H\left(\widehat{\mathbf{R}}_r\right) \stackrel{def}{=} \mathrm{Tr}\left[\left(\widehat{\mathbf{R}}_r\right)^2\right] \ , \tag{20}$$

where Tr[·] denotes the trace of an operator (sum of eigenvalues). This captures the notion that for detection in Gaussian noise, for given signal energy, the smaller the effective rank of $\widehat{\mathbf{R}}_r$ the better the SNR[19] — other criteria, entropy-based, for example, can capture this notion as well.

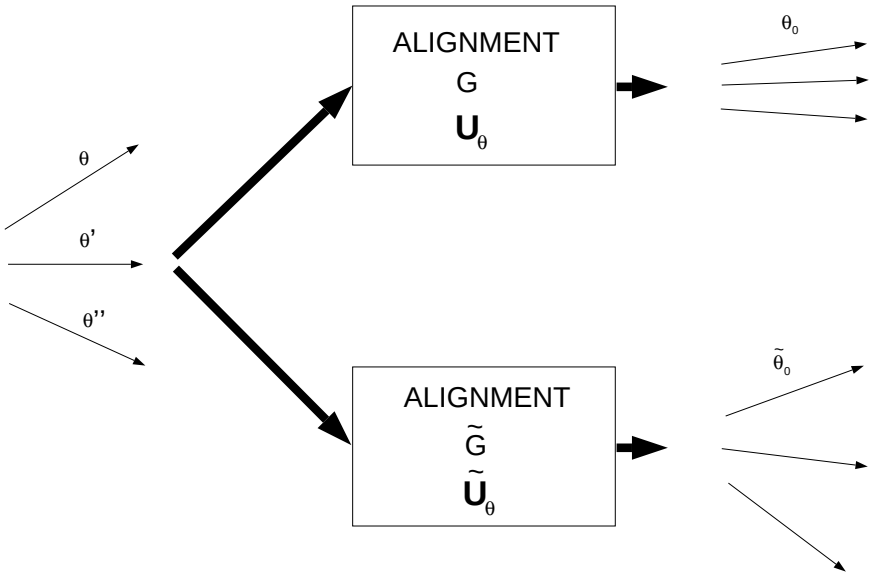

Figure 3. Identification: the "best" nuisance parameters result in the least spread in the aligned data. In the illustration, the nuisance parameters corresponding to G better fit the data.

4. DATA-DRIVEN FRAMEWORK

Consider the hypothesis testing problem (1) characterized by R_s^θ and σ^2 assumed to be unknown. Suppose that we have N_0 training realizations, x_i^0, $i = 1, 2, \cdots N_0$, under H_0, and N_1 realizations, x_i^1, $i = 1, 2, \cdots N_1$, under H_1, available to us. Further suppose that that we have Q classes, C_i, $i = 1, 2, \cdots Q$, of JSRs (defined by corresponding G_i and $\mathbf{U}_{\theta^i}^i$), that have been chosen *a priori* to reflect the types of likely nuisance parameters, such as time-frequency shifts, scale changes, chirp rate changes etc. The structure of the JSR-based detectors suggests the following two-part data-driven algorithm to estimate the "best" $\mathbf{R}_r$ to design the detectors.

Estimation Algorithm

Step 1. Estimate $\mathbf{R}_n = \mathbf{R}_0 = \sigma^2\mathbf{I}$ and $\mathbf{R}_1$:

$$\hat{\sigma}^2 = \frac{1}{N_0}\sum_{i=1}^{N_0} \|x_i^0\|^2 \ , \tag{21}$$

$$\widehat{R}_1(t_1,t_2) = \frac{1}{N_1}\sum_{i=1}^{N_1} x_i^1(t_1)x_i^{1*}(t_2) \ . \tag{22}$$

Define the operator

$$\widehat{\mathbf{D}} = (\widehat{\mathbf{R}}_1 - \hat{\sigma}^2\mathbf{I})\widehat{\mathbf{R}}_1^{-1} \ , \tag{23}$$

which is the (estimated) Wiener filter for estimating the signal component from an H_1 realization. We assume enough data so that $\widehat{\mathbf{R}}_1$ is invertible; a pseudo-inverse can be used otherwise.

Step 2. Choose the "reference" correlation function template

$$\widehat{\mathbf{R}}_a = \widehat{\mathbf{R}}_1 - \hat{\sigma}^2\mathbf{I} \ , \tag{24}$$

which serves as an estimate of $\mathbf{R}_r$ for the purpose of alignment.

Step 3. For each JSR class C_j, $j = 1, 2, \cdots Q$, do:

Initialize $\widehat{R}_r(t_1,t_2) = 0$, $(t_1,t_2) \in T \times T$.

For $i = 1$ to N_1

begin

$$\hat{\theta} = \arg\max_{\theta\in S}\left[(\mathbf{P}\,x_i)\,(\theta;\widehat{\mathbf{R}}_a)\right] \ , \tag{25}$$

$$\hat{s}_i(t) = \left(\widehat{\mathbf{D}}x_i^1\right)(t) \quad \text{("unaligned" signal estimate)} \tag{26}$$

$$\hat{s}_{r,i}(t) = \left(\mathbf{U}_{\hat{\theta}}^{-1}\hat{s}_i\right)(t) \quad \text{("aligned" signal estimate)} \tag{27}$$

$$\widehat{R}_r(t_1,t_2) = \widehat{R}_r(t_1,t_2) + \hat{s}_{r,i}(t_1)\hat{s}_{r,i}^*(t_2) \tag{28}$$

end

$\widehat{\mathbf{R}}_r^j = \frac{1}{N_1}\widehat{\mathbf{R}}_r$ (corresponding to the JSR class C_j)

Identification Algorithm

Compute the deflections

$$H\left(\widehat{\mathbf{R}}_r^j\right) = \mathrm{Tr}\left[\left(\widehat{\mathbf{R}}_r^j\right)^2\right] \ , \quad j = 1, 2\cdots Q \ . \tag{29}$$

The JSR class C_j corresponding to the largest deflection defines the nuisance parameters that best fit the training data.

4.1. Remarks

The objective of the estimation algorithm is to estimate the underlying $\mathbf{R}_r$ for each JSR class. It does so by aligning (Step 3) all the H_1 realizations to the value of θ implicitly defined by the reference template $\widehat{\mathbf{R}}_a$ (Step 2). Note that the estimate of θ in (25) is the ML estimate if $\widehat{\mathbf{R}}_a = \mathbf{R}_r$. Step 3 may be repeated to yield better $\widehat{\mathbf{R}}_r$ estimates by replacing $\widehat{\mathbf{R}}_a$ with $\widehat{\mathbf{R}}_r$. Note that the estimation algorithm implicity yields an estimate of $p_\Theta(\theta)$ as well (via (25)).

The identification step essentially chooses the nuisance parameters which pack the signal energy in the smallest subspace (corresponding to the $\widehat{\mathbf{R}}_r$ with the smallest rank). Taken together, $\widehat{\mathbf{R}}_r$ and $\mathbf{U}_\theta$ corresponding to best nuisance parameters define $\widehat{\mathbf{R}}_s^\theta$ which then yields the corresponding GLRT and Bayesian detectors.

We note that several choices for the operators $\widehat{\mathbf{D}}$ and $\widehat{\mathbf{R}}_a$ are possible. The purpose of the operator $\widehat{\mathbf{D}}$ in (23) is to mitigate the noise component of H_1 realizations, and here we have proposed a minimum-mean-squared-error approach. Other denoising approaches may be used as well.[20,21]

The choice of the alignment operator $\widehat{\mathbf{R}}_a$ in (24) assumes that the spread of the nuisance parameters is sufficiently small in the H_1 training realizations. In case of a large spread, a "pre-alignment" procedure may be used by aligning the H_1 training realizations to a particular "good" H_1 realization x_{io}^1 by using the alignment operator $\widehat{R}_a(t_1,t_2) \stackrel{def}{=} x_{io}^1(t_1)x_{io}^{1*}(t_2)$.

5. EXAMPLES: TIME-FREQUENCY AND TIME-SCALE DETECTORS

In this section, we demonstrate the utility of the data-driven framework by designing time-frequency and time-scale detectors from simulated data. Recall that the estimation part of our JSR-based data-driven algorithm yields estimates of $\mathbf{R}_1$, $\mathbf{R}_r$ (and hence $\mathbf{R}_s^\theta$, given $\mathbf{U}_\theta$), σ^2, and (implicity) $p_\Theta(\theta)$.

The (locally optimal) Bayesian detector which ignores the nuisance parameters can be directly realized as

$$L_{\mathrm{B1}}(x) = \langle \widehat{\mathbf{R}}_s x, x\rangle \quad , \ \widehat{\mathbf{R}}_s = \widehat{\mathbf{R}}_1 - \hat{\sigma}^2\mathbf{I} \ . \tag{30}$$

Recall that $\widehat{\mathbf{R}}_s$ is an estimate of

$$\mathbf{R}_s = \int \mathbf{R}_s^\theta p_\Theta(\theta) d\theta \tag{31}$$

if the nuisance parameters are distributed according to $p_\Theta(\theta)$. By exploiting the structure of nuisance parameters, the JSR-based GLRT detector can be realized using (6) and (10) as

$$L_{\mathrm{G}}(x) = \max_{\theta\in S} (\mathbf{P}x)\left(\theta; \widehat{\mathbf{R}}_r\right) \ . \tag{32}$$

Moreover, the Bayes detector also admits an alternative JSR-based realization via

$$L_{\mathrm{B2}}(x) = \int (\mathbf{P}x)\left(\theta; \widehat{\mathbf{R}}_r\right) \hat{p}_\Theta(\theta) d\theta \ . \tag{33}$$

Although the two Bayesian detectors, (30) and (33), are equivalent under exact knowledge of statistics, their designs based on training data can have substantially different performance as we will see. We will compare the performance of $L_{\mathrm{G}}(x)$, $L_{\mathrm{B1}}(x)$ and $L_{\mathrm{B2}}(x)$ in the following simulation results.

We use an observation interval of $T = 50$ samples in all the simulations. The underlying signal $s_r(t)$ is $N_s = 16$ samples long and of the form $s_r(n) = a\sum_{k=1}^3 Z_k e^{-\beta n^2} e^{j2\pi f_k n}$, $-N_s/2+1 \le n \le N_s/2$, where $Z_k \sim \mathcal{N}(0,1)$, a is positive constant to control the SNR, β is the (fixed) variance of the Gaussian envelope, and $f_k \in [1/16, 1/4]$ (fixed). This corresponds to a rank-3 $N_s \times N_s$ matrix $\mathbf{R}_r$. Without loss of generality, for the "zero" value of the nuisance parameters, the signal $s^{(\theta=0)}(t) = s_r(t)$ is *centered* in the observation interval T. Using the estimation algorithm of Section 4, the detectors are designed using $N_d = 25, 50$ and 100 realizations each under H_0 and H_1, and are tested using 100 new realizations each under H_1 and H_0.

5.1. Estimation

Time-Frequency Detectors. Recall that in the case of TFRs, the nuisance parameters correspond to time and frequency shifts. For realizing the time-frequency detectors, we compute the TFRs via the subspace-based formulation in terms of STFTs given in (14). The STFTs are computed using $N_f = 32$ point FFTs. The time and frequency nuisance parameters take on values uniformly so that the signal $s^{(t,f)}(t)$ can be anywhere in the $T \times N_f$ time-frequency region. In this case, the data are generated at SNR $= 10\log(E[s^H s]/E[n^H n]) = 2.3$dB. ROC curves for the three detectors for different amounts of training data are shown in Figure 4. The time-frequency (TF) GLRT detector $L_{\mathrm{G}}(x)$ performs the best, followed by the TF Bayes detector $L_{\mathrm{B2}}(x)$, which in turn performs better than the unaligned Bayes detector $L_{\mathrm{B1}}(x)$.

Time-Scale Detectors. In the case of time-scale representations, the nuisance parameters correspond to time and scale shifts. We compute the TSRs via the formulation (16) in terms of WTs to realize the TSR detectors.

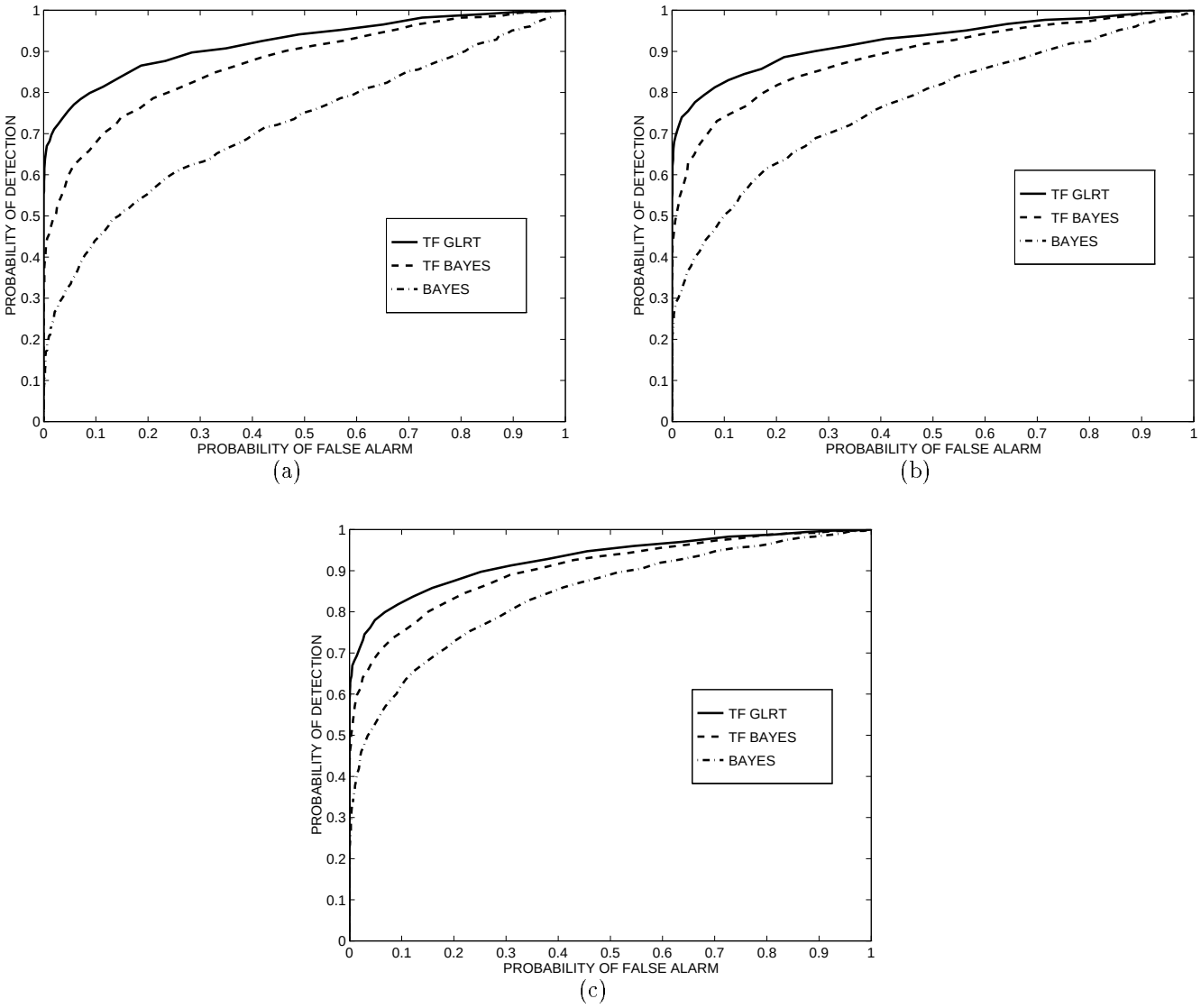

Figure 4. Time-Frequency Detectors: (a) 25 training vectors, (b) 50, and (c) 100.

The WTs are computed between the scales of 1/4 and 2 using $N_c = 65$ samples. The nuisance parameters take on values uniformly so that the entire $T \times N_c$ time-scale region is occupied by the signal $s^{(t,c)}(t)$. The SNR in this case is 1.5dB. Figure 5 shows the comparison of the three detectors for different amounts of training data. Again the time-scale (TS) GLRT detector $L_{\rm G}(x)$ performs the best, followed by the TS Bayes detector $L_{\rm B2}(x)$ which in turn performs better than the unaligned Bayes detector $L_{\rm B1}(x)$.

5.2. Identification

Data with time-scale nuisance parameters are generated as before and the estimation algorithms for both TFRs and TSRs are applied to yield corresponding estimates of $\mathbf{R}_r$, whose eigenvalues are shown in Figure 6, along with those for the unaligned estimate $\widehat{\mathbf{R}}_s$. The eigenvalue profile is most concentrated for the time-scale-aligned data as expected. Correspondingly, the deflections of the three correlation function estimates are: 5.7(TS), 4.09(TF) and 1.92(unaligned). The sum of the eigenvalues is the same in the three cases.

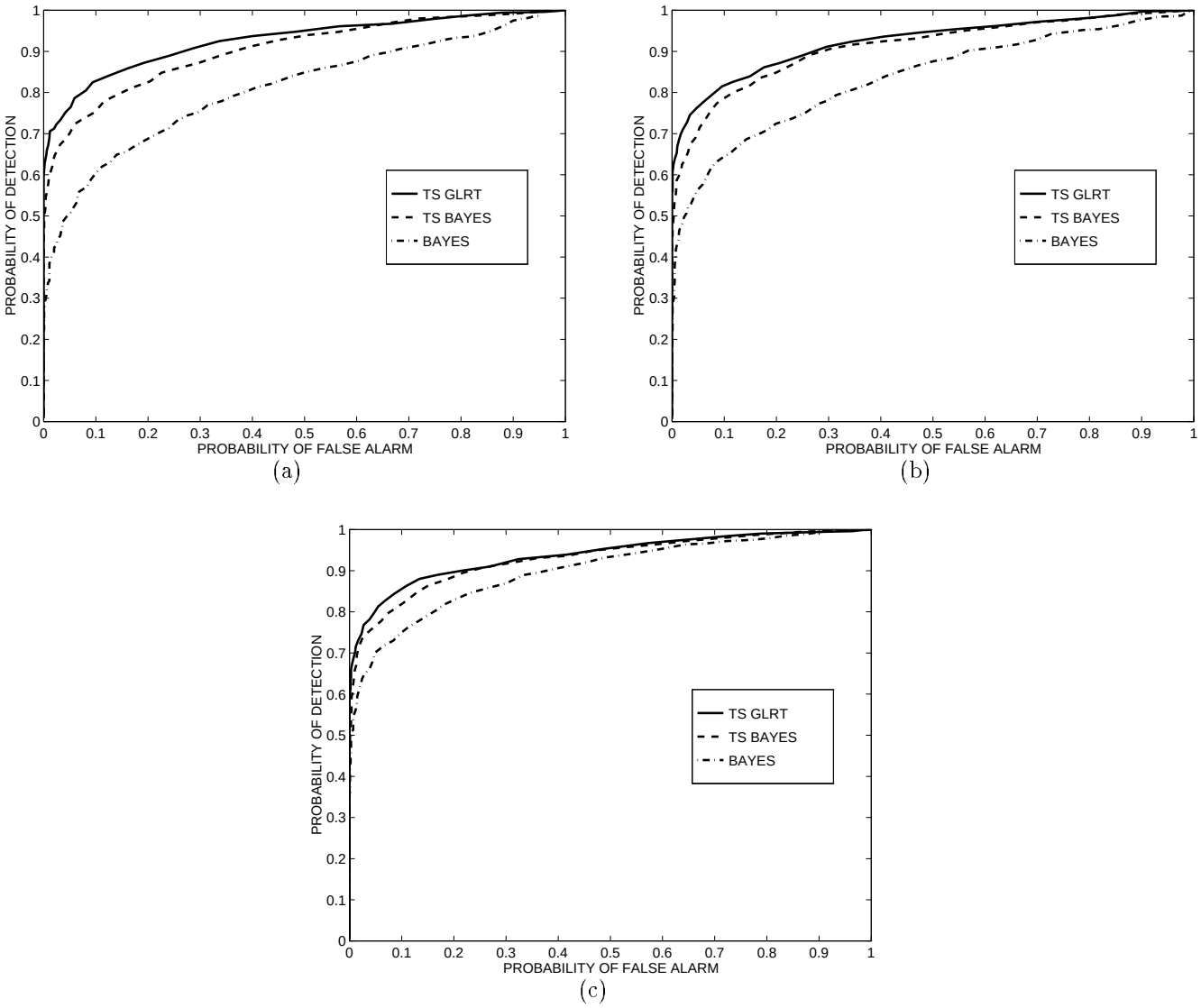

Figure 5. Time-Scale Detectors: (a) 25 training vectors, (b) 50, and (c) 100.

5.3. Discussion

The examples illustrate the ability of the data-driven framework to identify the correct parameters and design the corresponding detectors. Recall that the Bayesian detector $L_{\mathrm{B1}}(x)$ is designed without exploiting the structure of nuisance parameters. On the other hand, the design of GLRT detector $L_{\mathrm{G}}(x)$ and the Bayesian detector $L_{\mathrm{B2}}(x)$ takes the nuisance parameters into account.

Exploitation of the structure of the nuisance parameters reduces the complexity of the detectors in two ways. First, the underlying correlation function $\mathbf{R}_r$ is of a lower rank than $\mathbf{R}_s$. Second, the support of $\mathbf{R}_r$ is smaller than that of $\mathbf{R}_s$, due to spread of the nuisance parameters (time-shifts). Both factors increase the degrees of freedom in $\mathbf{R}_s$ compared to the inherent degrees of freedom in $\mathbf{R}_r$. For example, in the above examples, $N_s = 16$ which implies that $\mathbf{R}_r$ has a 16×16 support, whereas due to the nuisance parameters, $\mathbf{R}_s$ has an effective support of $T \times T$, where $T = 50$. Note that the operators defining the TFR and TSR detectors in the above examples have a 16×16 support corresponding to $\mathbf{R}_r$. The Bayesian detector $L_{\mathrm{B1}}(x)$ can also be designed for a centered support of $N_s \times N_s$, but

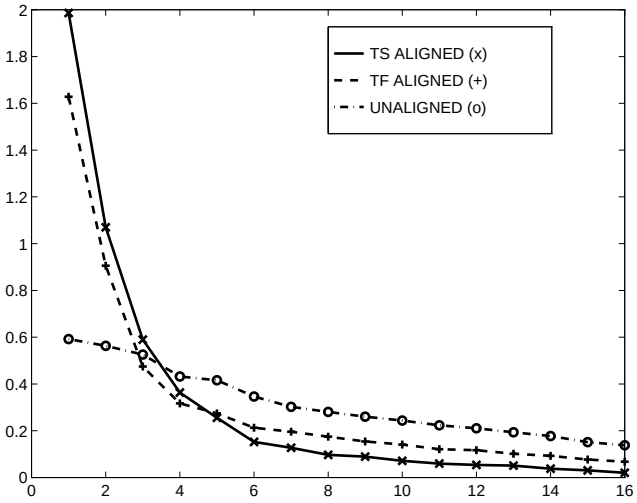

Figure 6. Comparison of time-frequency versus time-scale identification (time-scale data). Eigenvalue profile of the various detectors.

our simulations showed that the performance of this reduced-support Bayesian detector is inferior to that of the full (T) support one, since it fails to capture to significant signal components for large time-shifts. The JSR-based formulation of the Bayesian detector, $L_{\mathrm{B2}}(x)$, alleviates this problem to some extent by using an underlying detector of support $N_s \times N_s$ and explicitly averaging over the nuisance parameters. However, the performance of $L_{\mathrm{B2}}(x)$ is still inferior to that of $L_{\mathrm{G}}(x)$ due to errors in estimating $\hat{p}_{\Theta}(\theta)$.

Finally, it is worth noting that the performance of the three detectors converges as the amount of training data increases (see Figures 4 and 5). Thus, it appears that even in the limit of perfect training, the GLRT detector performs as well as the Bayesian detector.

6. CONCLUSIONS

Detector design from limited training data is a challenging problem since in the absence of some underlying (low-complexity) signal structure there is usually not enough training data for reliable detector design. However, in many practical scenarios, the inherent signal complexity is much lower than that apparent from the training data due to the presence of nuisance parameters, such as unknown time offsets, frequency shifts, and scale changes. Exploiting the structure of nuisance parameters provides a possibility for reliable detector design from limited training data.

In practice, in order to make such data-driven design viable, some structure needs to be imposed on the class of possible nuisance parameters. Such structure can then facilitate the identification of the appropriate nuisance parameters from the training data, and the estimation of the requisite statistics to design the corresponding detectors.

In this paper, we have exploited the structure of JSR-based detectors to propose a data-driven framework that can model a wide variety of nuisance parameters. We have presented algorithms that automatically identify the nuisance parameters (from a finite set) that best fit the training data, and estimate the corresponding detector statistics. Simulation results based on time-frequency and time-scale representations have demonstrated that if the data at hand do exhibit nuisance parameters, exploiting their structure via JSRs to design the detectors can deliver substantial performance improvements for limited training data (see Figures 4 and 5). In particular, our results indicate that the JSR-based GLRT detector consistently outperforms the Bayesian detector for limited training data. Moreover, as evidenced by Figure 6, our approach has the ability of correctly identifying the underlying nuisance parameters. Thus, the framework proposed in this paper is suitable for the numerous practical detection applications in which only limited training data is available.

ACKNOWLEDGEMENTS

The author would like to thank Dr. B. Scott Krongold of the University of Illinois for graciously providing his code (based on[22]) for computing the continuous wavelet transform.

REFERENCES

1. L. Cohen, *Time-Frequency Analysis*, Prentice Hall, 1995.
2. F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine* , April 1992.
3. O. Rioul and P. Flandrin, "Time-scale distributions: A general class extending the wavelet transform," *IEEE Trans. Signal Processing* **40**, pp. 1746–1757, May 1992.
4. O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine* , October 1991.
5. A. M. Sayeed and D. L. Jones, "A canonical covariance-based method for generalized joint signal representations," *IEEE Signal Processing Letters* **3**, pp. 121–123, April 1996.
6. A. M. Sayeed and D. L. Jones, "Optimum quadratic detection and estimation using generalized joint signal representations," *IEEE Trans. Signal Processing* **44**, pp. 3031–3043, December 1996.
7. A. M. Sayeed and D. L. Jones, "Optimal detection using bilinear time-frequency and time-scale representations," *IEEE Trans. Signal Processing* **43**, pp. 2872–2883, December 1995. See also the corrections in *IEEE Trans. Signal Processing*, vol. 45, pp. 761–762, March 1997.
8. B. Samimy and G. Rizzoni, "Mechanical signal analysis using time-frequency signal processing: Application to internal combustion engine knock detection," *Proc. IEEE (special issue on time-frequency analysis)* **84**, pp. 1330–1343, September 1996.
9. B. Samimy, G. Rizzoni, A. M. Sayeed, and D. L. Jones, "Design of training data-based quadratic detectors with application to mechanical systems," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. – ICASSP '96*, vol. 3, pp. 1767–1770, 1996.
10. B. S. Krongold, A. M. Sayeed, M. Moehring, J. A. Ritcey, M. Spencer, and D. L. Jones, "Quadratic time-scale detection of microemboli in flowing blood with doppler ultrasound," in *Proc. SPIE 3162, Advanced Signal-Processing Algorithms , Architectures and Implementation II*, F. T. Luk, ed., 1997.
11. H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, 1988.
12. 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.
13. R. G. Baraniuk, "Covariant time-frequency representations through unitary equivalence," *IEEE Signal Processing Letters* , pp. 79–81, March 1996.
14. A. A. Kirillov, *Elements of the Theory of Representations*, Springer-Verlag, 1976.
15. A. M. Sayeed and D. L. Jones, "A simple covariance-based characterization of joint signal representations of arbitrary variables," in *Proc. IEEE Int'l Symp. on Time-Frequency and Time-Scale Analysis*, pp. 433–436, 1996.
16. A. M. Sayeed and D. L. Jones, "Generalized joint signal representations and optimum detection," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '96*, vol. 3, pp. 1431–1434, 1996.
17. L. Cohen, "Time-frequency distributions — a review," *Proc. IEEE* **77**, pp. 941–981, July 1989.
18. G. S. Cunningham and W. J. Williams, "Kernel decomposition of time-frequency distributions," *IEEE Trans. Signal Processing* **42**, pp. 1425–1442, June 1994.
19. C. R. Baker, "Optimum quadratic detection of a random vector in Gaussian noise," *IEEE Trans. Comm. Tech.* **COM-14**, pp. 802–805, December 1966.
20. D. Donoho and I. Johnstone, "Adapting to unknown smoothnes via wavelet shrinkage," *J. Amer. Stat. Assoc.* **90**, pp. 1200–1224, December 1995.
21. S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Subspace-based signal detection and estimation using wavelets," in *Proc. SPIE 3169, Advanced Signal-Processing Algorithms, Architectures and Implementation II*, F. T. Luk, ed., 1997.
22. D. L. Jones and R. G. Baraniuk, "Efficient approximation of continuous wavelet transforms," *Electronics Letters* , April 25 1991.