TIME-FREQUENCY DETECTORS

Akbar M. Sayeed and Douglas L. Jones

Coordinated Science Laboratory
University of Illinois
1308 W. Main Street
Urbana, IL 61801*

ABSTRACT

Time-frequency representations (TFRs) provide a powerful and flexible structure for designing optimal detectors in a variety of nonstationary scenarios. In this paper, we describe a TFR-based framework for optimal detection of arbitrary second-order stochastic signals, with certain unknown or random nuisance parameters, in the presence of Gaussian noise. The framework provides a useful model for many important applications including machine fault diagnostics and radar/sonar. We emphasize a subspace-based formulation of such TFR detectors which can be exploited in a variety of ways to design new techniques. In particular, we explore an extension based on *multi-channel/sensor* measurements that are often available in practice to facilitate improved signal processing. In addition to potentially improved performance, the subspace-based interpretation of such multi-channel detectors provides useful information about the physical mechanisms underlying the signals of interest.

1. INTRODUCTION

Detection and classification of signals in the presence of noise and interference is an old and important problem in communications and signal processing. A wide variety of applications involve signals with nonstationary or time-varying characteristics; examples include radar, sonar, communications, machine fault diagnostics and biomedical and geophysical signal processing. Moreover, in many such scenarios, due to the underlying physical mechanisms, the signals of interest possess certain nuisance parameters which have to be incorporated in the design of detectors. For example, in radar the targets have unknown delay-Doppler, and in machine health monitoring, the fault signals occur at unknown time-offsets and often exhibit unknown or random frequency shifts [1, 2, 3]. The need for detection in such nonstationary scenarios has spurred a great deal of interest in time-frequency-based detection schemes [1, 2, 4, 3, 5].

Time-frequency representations (TFRs) are a versatile set of tools for the analysis and processing of nonstationary signals [6]. They are overparameterized signal representations in terms of time and frequency, and describe the nonstationary signal characteristics via their time-varying spectral content. Prominent examples of such multidimensional representations include the classical short-time Fourier transform (STFT) and the Wigner distribution, and more recent ones such as the continuous wavelet transform (CWT). Linear TFRs such as the STFT (narrowband ambiguity function) have long been used in matched-filter radar detection. However, quadratic TFRs, by virtue of their richer structure, hold promise for a wider variety of detection scenarios. Two important classes of such quadratic TFRs are **Cohen's class** [6] which generalizes the concept of the spectrogram ($|\mathrm{STFT}|^2$), and the **affine class** [7] which generalizes the concept of the scalogram ($|\mathrm{CWT}|^2$).

Recently, a comprehensive theory for optimum TFR-based detection has been developed that has put time-frequency detection on a firm footing [5]. It also overcomes the limitations of previously proposed methods that were mostly *ad hoc* and did not exploit the structure of TFRs [5]. The theory characterizes the detection scenarios in which TFR detectors (from both Cohen's and affine classes) are "natural"; that is, they are optimal from a detection theoretic viewpoint *and* exploit the degrees of freedom available in a TFR. It also provides an explicit characterization of the corresponding TFR detectors.

The TFR detection framework of [5] is optimal for detecting arbitrary *second-order* nonstationary stochastic signals, with certain unknown or random nuisance parameters, in the presence of arbitrary Gaussian noise [5]. For TFR detectors from Cohen's class, the appropriate nuisance signal parameters are time and frequency shifts, and for the affine class, time-shifts and scalings. As mentioned before, such signal parameters are commonly encountered in practice. In particular, our experience with applications in machine fault diagnostics and biomedical signal processing has shown that signals of interest often exhibit different time-frequency shifts and scalings in different observations (see, for example, [3]).

In this paper, we provide an exposition of the optimal TFR detectors based on an equivalent subspace-based formulation in terms of a bank of spectrograms or scalograms. This structural description lends considerable insight into the mechanism of TFR detectors, and can be exploited in a variety of ways to design new techniques. In particular, we discuss an extension of time-frequency detectors based on *multi-channel/sensor* measurements that are collected in many applications in order to facilitate the extraction or classification of relevant signal characteristics. For example, in machine health monitoring, signals from a number of sen-

*This work was supported in part by the Office of Naval Research under Grant No. N00014-95-1-0674, and the Schlumberger Foundation.

sors, mounted at various critical machine components, are often recorded. As we will see, in addition to potentially improved performance, the subspace-based structure of such multi-channel detectors yields useful information about the physical mechanisms underlying the signals of interest.

We start with a brief review of TFRs in the next section, followed by a description of the time-frequency detection framework in Section 3. This includes a characterization of the detection scenarios for which TFR detectors are canonical, and a subspace-based characterization of the corresponding detectors. In Section 4, we provide an insightful interpretation of the subspace-based formulation. An extension of the TFR detectors to multi-sensor scenarios is explored in Section 5. Finally, in Section 6, we present some conclusions regarding the applications and scope of time-frequency detectors.

2. TIME-FREQUENCY REPRESENTATIONS

Two well-known linear TFRs will be instrumental in the subspace-based formulation: the short-time Fourier transform (STFT) and the continuous wavelet transform (CWT). The STFT is defined as

$$\mathrm{STFT}_s(t,f;g) \equiv \int s(\tau) g^*(\tau - t) e^{-j2\pi f\tau} d\tau \ , \qquad (1)$$

where g is the *analysis window* which is usually lowpass. The STFT represents signal characteristics jointly in terms of **time** (t) and **frequency** (f). The continuous wavelet transform (CWT), on the other hand, is defined as

$$\mathrm{CWT}_s(t,a;g) \equiv \frac{1}{\sqrt{a}} \int s(\tau) g^* \left(\frac{\tau - t}{a} \right) d\tau \qquad (2)$$

and is a joint representation in terms of **time** (t) and **scale** ($a > 0$). The window g is called the *mother wavelet* which is usually bandpass. The squared magnitude of the STFT is known as the spectrogram, and that of the CWT is known as the scalogram.

Bilinear TFRs provide a richer structure than linear ones, and two important classes are Cohen's and affine classes [6, 7]. Both classes can be defined as *smoothed* versions of the Wigner distribution (WD) which is defined as

$$W_s(t,f) \equiv \int s\left(t + \frac{\tau}{2}\right) s^* \left(t - \frac{\tau}{2}\right) e^{-j2\pi f \tau} d\tau \ . \qquad (3)$$

Cohen's class, which is a generalization of the spectrogram, can be expressed as a *convolutional* smoothing of the WD [6]

$$P_s(t,f;\Phi) \equiv \int\int W_s(u,v) \Phi(u-t, v-f) du dv \qquad (4)$$

where the two-dimensional (2D) kernel Φ completely characterizes the TFR $P_s(\Phi)$. The affine class, on the other hand, is a generalization of the scalogram, and is characterized by an *affine* smoothing of the WD [7]

$$C_s(t,a;\Pi) \equiv \int\int W_s(u,v) \Pi \left(\frac{u-t}{a}, av \right) du dv \qquad (5)$$

where the kernel Π completely characterizes the TFR $C_s(\Pi)$.

Two fundamental properties of TFRs underlie the theory of time-frequency detection: Cohen's class is *covariant* to time and frequency shifts; that is, $s(t) \mapsto s(t-\tau)e^{j2\pi\nu t} \Rightarrow P_s(t,f;\Phi) \mapsto P_s(t-\tau, f-\nu;\Phi)$, and the affine class is covariant to time-shifts and scalings; that is, $s(t) \mapsto \frac{1}{\sqrt{c}} s\left(\frac{t-\tau}{c}\right) \Rightarrow C_s(t,a;\Pi) \mapsto C_s\left(\frac{t-\tau}{c}, \frac{a}{c}; \Pi\right)$.

3. TIME-FREQUENCY DETECTION FRAMEWORK

Signal detection is a binary hypothesis testing problem, and we consider hypotheses of the form

$$\begin{aligned} H_1 &: \ x(t) = s(t) + n(t) \\ H_0 &: \ x(t) = n(t) \end{aligned} \qquad (6)$$

where $t \in T$, the observation interval, x is the observed signal, s is the underlying signal to be detected, and n is additive noise. Based on the observation x, it has to be decided whether the signal s is present (H_1) or not (H_0). For a variety of performance criteria, such as Neyman-Pearson, Bayesian or minimax, the optimal decision is made be comparing a real-valued function $L(x)$, the *test statistic*, to a threshold.

From the viewpoint of time-frequency detection, there are two key observations:

1. Bilinear TFRs are **quadratic** in the observations, and
2. TFRs possess **additional degrees of freedom** provided by the TFR parameters; **time** and **frequency** for Cohen's class, and **time** and **scale** for the affine class.

The first observation leads us to consider scenarios in which quadratic detectors are optimal, namely the detection of Gaussian or arbitrary *second-order* signals in the presence of Gaussian noise [8, 5]. The second observation leads to *composite hypothesis testing* in which the signal to be detected has a couple of nuisance parameters that are unknown or random [8, 5]. That is, the hypothesis testing problem becomes [5]

$$\begin{aligned} H_1 &: \ x(t) = s(t;\alpha,\beta) + n(t) \\ H_0 &: \ x(t) = n(t) \end{aligned} \qquad (7)$$

where

$$\begin{aligned} (\alpha,\beta) &: \ \text{nuisance signal parameters} \\ s(t;\alpha,\beta) &: \ \text{Gaussian or arbitrary } \textit{second-order} \text{ signal} \\ n(t) &: \ \text{arbitrary Gaussian noise .} \end{aligned}$$

We focus our attention on zero-mean signals and noise,[1] and assume that the signal $s(t;\alpha,\beta)$ is characterized (up to second-order) by the correlation function $R_s^{(\alpha,\beta)}(t_1,t_2) \equiv E[s(t_1;\alpha,\beta)s^*(t_2;\alpha,\beta)]$, and the noise is characterized by the correlation function R_n.

[1] Nonzero-mean situations can be handled by including linear TFRs in the detectors.

In order to characterize the situations for which TFR detectors are "naturally" suited, we need to identify the nature of the nuisance signal parameters (α, β) in (7). In other words, we need to characterize the dependence of the signal correlation function $R_s^{(\alpha,\beta)}$ on (α, β), which is done next.

3.1. Signal models for time-frequency detectors

It is shown in [5] that the following signal models characterize the scenarios for which TFR-based detectors are canonical.

3.1.1. Cohen's class

Not surprisingly, for TFRs from Cohen's class, the parameters (α, β) must correspond to **time-frequency shifts**; that is, $(\alpha, \beta) \equiv (\tau, \nu) \in T \times \mathbb{R}$ and for $(t_1, t_2) \in T \times T$

$$R_s^{(\tau,\nu)}(t_1, t_2) = R_{TF}(t_1 - \tau, t_2 - \tau) e^{j2\pi\nu t_1} e^{-j2\pi\nu t_2} \qquad (8)$$

for some correlation function R_{TF}.[2] Note that (8) is equivalent to $s(t; \tau, \nu) \equiv s_{(\tau,\nu)}(t - \tau) e^{j2\pi\nu t}$ in (7), where, for each (τ, ν), $s_{(\tau,\nu)}$ is any Gaussian or second-order signal with correlation function R_{TF}; that is, for each (τ, ν), $s(t; \tau, \nu)$ is a time-frequency shifted version of some random signal with correlation function R_{TF}.

3.1.2. Affine class

For TFRs from the affine class, the parameters must correspond to **time-shifts and scalings**; that is, $(\alpha, \beta) = (\tau, c) \in T \times (0, \infty)$ and for $(t_1, t_2) \in T \times T$

$$R_s^{(\tau,c)}(t_1, t_2) = c R_{TS}(c(t_1 - \tau), c(t_2 - \tau)) , \qquad (9)$$

for some correlation function R_{TS}. Again, (9) is equivalent to $s(t; \tau, c) \equiv \sqrt{c}\, s_{(\tau,c)}(c(t - \tau))$ for any second-order signal $s_{(\tau,c)}$ with correlation function R_{TS}. That is, for each (τ, c), $s(t; \tau, c)$ is a time-shifted and scaled version of some random signal with correlation R_{TS}.

3.2. Characterization of TFR detectors

The correlation functions R_{TF} and R_{TS} in (8) and (9) completely characterize the second-order statistics of the underlying signal, and play a fundamental role in the characterization of TFR detectors. In [5], it is shown that the optimal detectors for various composite hypothesis testing problems of the form (7), coupled with the signal models in the previous section, can be naturally realized using the following TFR-based structures:

$$\begin{aligned} L_{TF}(x) &= \max_{(\tau,\nu)} [P_y(\tau, \nu; \Phi) + F_{TF}(\tau, \nu)] && (10) \\ L_{TS}(x) &= \max_{(\tau,c)} [C_y(\tau, 1/c; \Pi) + F_{TS}(\tau, c)] , && (11) \end{aligned}$$

where the kernels Φ and Π characterizing the TFRs can be expressed explicitly in terms of the underlying correlation functions R_{TF} and R_{TS}, respectively [5]. In the above expressions

$$y = \mathbf{R}_n^{-1} x \qquad (12)$$

where $\mathbf{R}_n^{-1}$ is the inverse[3] of the operator $\mathbf{R}_n$ defined by the noise correlation function as[4]

$$(\mathbf{R}_n s)(t) \equiv \int R_n(t, u) s(u) du . \qquad (13)$$

The deterministic functions F_{TF} and F_{TS} depend on the joint probability density function of the parameters in the case of random parameters [5].

A note on optimality. The above TFR structures can realize the optimal detector based on the likelihood-ratio (LR) for Gaussian signals in white Gaussian noise (WGN), and the locally optimal[5] or deflection-optimal detectors for arbitrary second-order signals in arbitrary Gaussian noise [5]. Moreover, due to the presence of nuisance parameters, a generalized-likelihood-ratio-test (GLRT) or its locally optimal analogue is used: maximum likelihood (ML) estimates of the parameters are used in the case of unknown parameters, and the maximum a posteriori probability (MAP) estimates are used in the case of random parameters [5, 8].

3.3. Subspace-based formulation

For all practical purposes,[6] the underlying correlation functions admit eigenexpansions:

$$\begin{aligned} R_{TF}(t_1, t_2) &= \sum_k \lambda_k u_k(t_1) u_k^*(t_2) , && (14) \\ R_{TS}(t_1, t_2) &= \sum_k \mu_k v_k(t_1) v_k^*(t_2) , && (15) \end{aligned}$$

where $\lambda_k > 0$, $\mu_k > 0$, and $\{u_k\}$ and $\{v_k\}$ are orthonormal. Using these eigenexpansions in conjunction with some fundamental properties of TFRs, the TFR detectors can be expressed in terms of weighted sums of a bank of spectrograms/scalograms [5]. More specifically, the TFRs $P_y(\Phi)$ and $C_y(\Pi)$, in the detector structures (10) and (11), can be expressed in terms of spectrograms and scalograms (using the eigenfunctions as windows) as

$$\begin{aligned} P_y(t, f; \Phi) &= \sum_k \gamma_k \, |\mathrm{STFT}_y(t, f; u_k)|^2 && (16) \\ C_y(t, 1/c; \Pi) &= \sum_k \delta_k \, |\mathrm{CWT}_y(t, 1/c; v_k)|^2 && (17) \end{aligned}$$

where

$$\gamma_k = \begin{cases} \frac{\lambda_k}{\lambda_k + N_o} & \text{LR detector for WGN} \\ \lambda_k & \text{locally optimal detector for arbitrary GN} \end{cases}$$

$$\delta_k = \begin{cases} \frac{\mu_k}{\mu_k + N_o} & \text{LR detector for WGN} \\ \mu_k & \text{locally optimal detector for arbitrary GN} \end{cases}$$

and N_o denotes the power spectral density of WGN. Equivalently, the kernels Φ and Π can be expressed in terms of

[2] We assume that the support of R_{TF} is small compared to $T \times T$. A similar assumption applies to R_{TS} in (9).

[3] Which we assume to exist. In particular, the presence of a white noise component guarantees its existence.

[4] Note that for white noise, $\mathbf{R}_n = \mathbf{I}$, the identity operator.

[5] Which is optimal under a weak signal (low SNR) assumption [8, 5].

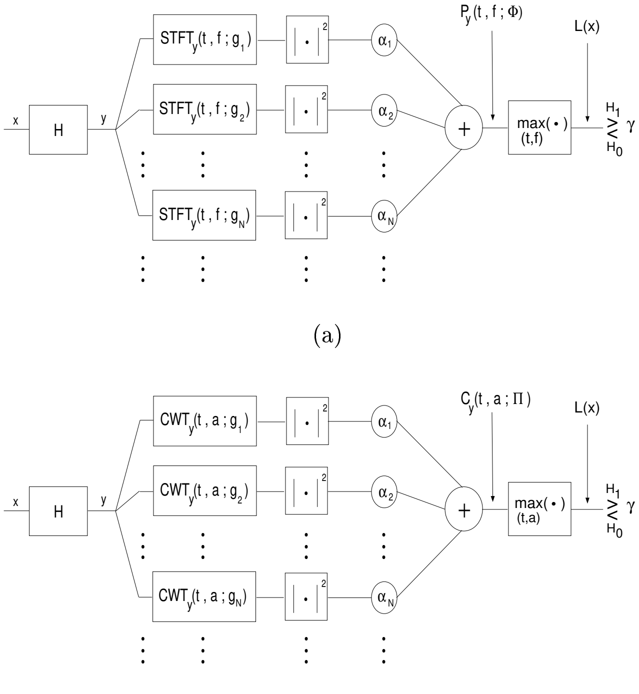
[6] If, for example, the observation interval T is compact.

the WDs of the windows and the combining weights as

$$\Phi(t,f) = \sum_k \gamma_k W_{u_k}(t,f) \ , \tag{18}$$

$$\Pi(t,f) = \sum_k \delta_k W_{v_k}(t,f) \ . \tag{19}$$

Thus, the optimal detectors project a time-frequency shifted (STFT), or time-shifted and scaled (CWT), version of the *preprocessed* observed signal (y) onto the eigenfunction of the underlying signal correlation function, and take a weighted sum of the magnitude-squared output of such matched-filter processors to produce the test statistic. This yields a subspace-based formulation of the TFR detectors: a nonlinear function of the projection of the signal onto a subspace, spanned by the signal eigenfunctions, realizes the optimal detectors. This subspace-based formulation, in terms of the eigen-modes of the signal, is depicted in Figure 1.



(a)

(b)

Figure 1. Subspace-based structure of TFR detectors; $\mathbf{H} = \mathbf{R}_n^{-1}$. (a) Cohen's class; $g_k = u_k$ and $\alpha_k = \gamma_k$. (b) Affine class; $g_k = v_k$ and $\alpha_k = \delta_k$.

4. INTERPRETATION OF THE SUBSPACE-BASED FORMULATION

As evident from Figure 1, general TFR detectors are composed of simpler elements consisting of matched-filter detectors realized by the different spectrograms and scalograms. This interpretation is intimately related to the structure and complexity of the underlying stochastic signal to be detected. From the viewpoint of second-order statistics, the rank[7] of the signal correlation function (R_{TS} or R_{TS}) is a measure of complexity. The number of spectrograms/scalograms in the TFR detector structure is exactly equal to the rank of R_{TF}/R_{TS}. Moreover, the windows for the different spectrograms/scalograms are precisely the eigenfunctions of the correlation function.

Each realization of a second-order signal admits the (second-order) Karhunen-Loève expansion

$$s(t) = \sum_k s_k w_k(t) \tag{20}$$

where the w_k's are the eigenfunctions, and the s_k's are uncorrelated random variables with variance, $E|s_k|^2$, equal to the corresponding eigenvalues. The eigenfunctions are the *natural modes* of the signal, and the number of such modes is equal to the rank. Each spectrogram/scalogram in the TFR detector, with an eigenfunction as the window, is optimally matched for processing the corresponding natural mode.

Thus, the subspace-based structure of TFR detectors provides a whole range of possibilities from simple rank-1 processors (only one element in the bank) for detecting essentially deterministic signals, to the more general higher-rank detectors for more complex stochastic signals. The subspace-based formulation can be exploited to optimize this *complexity* versus *performance* tradeoff [9].

Each rank-1 matched-filter component in Figure 1 naturally incorporates nuisance parameters (time-frequency shifts in Cohen's class, and time-shifts and scalings in the affine class) due to the implementation via spectrograms and scalograms. In fact, the exact time-frequency or time-scale location at which the maximum occurs (see (10) and (11)) in the TFR detector is precisely the ML or MAP estimate of the nuisance parameters. As mentioned earlier, the TFR detectors effectively realize a GLRT in which an estimate of the parameters is first formed and then used in the optimal detector structure corresponding to those values of the parameters [5].

5. SOME EXTENSIONS

Measurements from multiple sensors or channels are often collected to facilitate improved extraction and classification of signal characteristics. For example, for enhancement and classification of electrocardiogram signals, signals from multiple probes are often recorded. Similarly, multi-sensor measurements are fairly common in machine monitoring.

In this section, we explore a straightforward extension of the TFR detectors to such multi-sensor measurements. For simplicity, we will discuss the simpler case of quadratic detectors as opposed to TFR detectors; the essential ideas remain the same. But first, we briefly describe an extension of the binary hypothesis testing scenario (6) which is more realistic and would be helpful in interpreting the multi-sensor detectors.

[7] The number of nonzero terms in the eigenexpansion.

5.1. A more general binary hypothesis testing problem

The *signal + noise* model under H_1 adopted in (6) is not always appropriate in practice. A more general and more realistic setting, which we will adopt in the multi-sensor discussion, is one in which the signals under the two hypotheses are zero-mean Gaussian with different correlation functions R_0 and R_1. In this case, the optimal detector is given by

$$L(x) \equiv \langle \mathbf{Q}_o x, x\rangle \equiv \int (\mathbf{Q}_o x)(t) x^*(t) dt \tag{21}$$

where the linear operator $\mathbf{Q}_o$ is defined as[8] [8, 5]

$$\mathbf{Q}_o \equiv \mathbf{R}_0^{-1} - \mathbf{R}_1^{-1} \ . \tag{22}$$

We note that although $\mathbf{Q}_o$ is nonnegative definite in the *signal + noise* scenario, it is not necessarily so in the general case.

5.2. An extension to multi-sensor measurements

Suppose that measurements from N sensors, $x_1, x_2, \cdots x_N$, are available over a finite observation interval $[0, T)$. If we define a "concatenated" observation z over the interval $[0, NT)$ as

$$z(t) \equiv x_i(t) \quad , \quad \text{for} \quad \mathrm{t} \in [(\mathrm{i}-1)\mathrm{T}, \mathrm{iT}) \quad , \ \mathrm{i} = 1, 2 \cdots \mathrm{N} \tag{23}$$

then we can directly use the optimal detector (22) if we know the correlation functions of z under the two hypotheses.[9]

The eigenexpansion (subspace-based formulation) of the multi-sensor optimal detector

$$Q_o^N(t_1, t_2) = \sum_k \lambda_k w_k(t_1) w_k^*(t_2) \ , \tag{24}$$

$(t_1, t_2) \in [0, NT) \times [0, NT)$, yields some very useful information. The operator $\mathbf{Q}_o^N$ can be decomposed into two definite operators $\mathbf{Q}_{o+}^N$ (positive definite) and $\mathbf{Q}_{o-}^N$ (negative definite) as

$$\begin{aligned} Q_o^N(t_1, t_2) &\equiv Q_{o+}^N(t_1, t_2) \ + \ Q_{o-}^N(t_1, t_2) \\ &\equiv \sum_{k \in \mathcal{I}_+} \lambda_k w_k(t_1) w_k^*(t_2) \ + \ \sum_{k \in \mathcal{I}_-} \lambda_k w_k(t_1) w_k^*(t_2) \end{aligned} \tag{25}$$

where $\mathcal{I}_+ \equiv \{k : \lambda_k > 0\}$ and $\mathcal{I}_- \equiv \{k : \lambda_k < 0\}$. The eigenvectors w_k's corresponding to $\mathcal{I}_+$ roughly correspond to the "natural" modes dominant under H_1, whereas those corresponding to $\mathcal{I}_-$ roughly reflect the modes dominant under H_0. Moreover, the eigenvectors can be partitioned into components, corresponding to different channels, exactly similar to the partitioning of z in (23). Thus, in the case of machine fault diagnostics, for example, if H_1 represents "fault present", then the *presence* of $\mathcal{I}_+$ modes or the *absence* of $\mathcal{I}_-$ modes would roughly reflect the presence of the fault. Similarly, in terms of different channels, the presence ($\mathcal{I}_+$) or absence ($\mathcal{I}_-$) of certain signal characteristics in the different sensor measurements would reflect the occurrence of the fault. Moreover, the relative energy in the different partitions of an eigenfunction reflect the relative importance of the corresponding sensors in that mode. These concepts are illustrated in Figure 2, and explained in the next example.

5.3. An example with real data

Figure 2 is based on real data collected from firepumps.[10] It corresponds to a 2-channel quadratic detector (256×256 matrix) that was designed from training data collected from two accelerometers (2 channels), one mounted on the pump side, and the other on the motor side. The training data was used to estimate the correlation matrices R_1 (fault present) and R_0 (fault absent), from which the optimal detector was derived using (22). Each of the correlation matrices can be partitioned into four 128×128 submatrices corresponding to the partitioning of the two channels. We found that indeed such multi-channel detectors improved the performance relative to single-channel processors, albeit at the expense of higher complexity. However, in addition to improved performance, the subspace-based structure of such multi-channel detectors yields useful information about the underlying physical mechanisms which we illustrate next.

Figure 2(a) shows the eigenvalues and certain eigenfunctions of the optimal 2-channel detector. First, note in Figure 2(a) that effectively only about 50 to 60 of the eigenvalues are nonzero; the rank of the detector is substantially smaller (50–60) compared to its size (256). The eigen-modes in Figures 2(b)–2(c) correspond to some dominant *negative* eigenvalues ($\mathcal{I}_-$), whereas those in Figures 2(d)–2(f) correspond to a few dominant positive ones ($\mathcal{I}_+$). Moreover, in the eigen-modes of (c) and (e), channel 2 is dominant, whereas in (d) channel 1 has more energy. In (b) and (f), which correspond to the most dominant eigen-modes (largest absolute eigenvalues), both channels seem to be relevant.

These observations reveal useful information about the nature of the fault(s). For example, since channel 1 corresponds to measurements made on the motor side, the presence of dominant channel 1 components in the eigen-mode (d) ($\mathcal{I}_+$) suggests that the *presence* of such modes on the *motor side* indicates the occurrence of fault. On the other hand, the dominant pump-component (channel 2) in the eigen-mode (c) ($\mathcal{I}_-$) suggests that such modes are normally present on the pump side and their *absence* could possibly indicate the occurrence of fault. Such information can also be used to identify critical modes from the two channels that together provide better classification than individual channels but still result in a detector that is lower in complexity than the full two-channel detector.

[8] The corresponding generalization to the locally optimal or deflection optimal detector is given by $\mathbf{Q}_o = \mathbf{R}_0^{-1}(\mathbf{R}_1 - \mathbf{R}_0)\mathbf{R}_0^{-1}$.

[9] Which require, in addition to the correlation functions for each individual channel, the cross-correlation functions between the different channels under the two hypotheses.

[10] 16 pumps running at 3600 RPM; 5 faulty and 11 OK. Vibration data collected from three (tri-axial) accelerometers each on the motor and pump side (6 channels in total; 50 kHz sampling rate). We thank Dr. Douglas Lake of the Office of Naval Research for providing the firepump data and permitting us to use it in this paper.

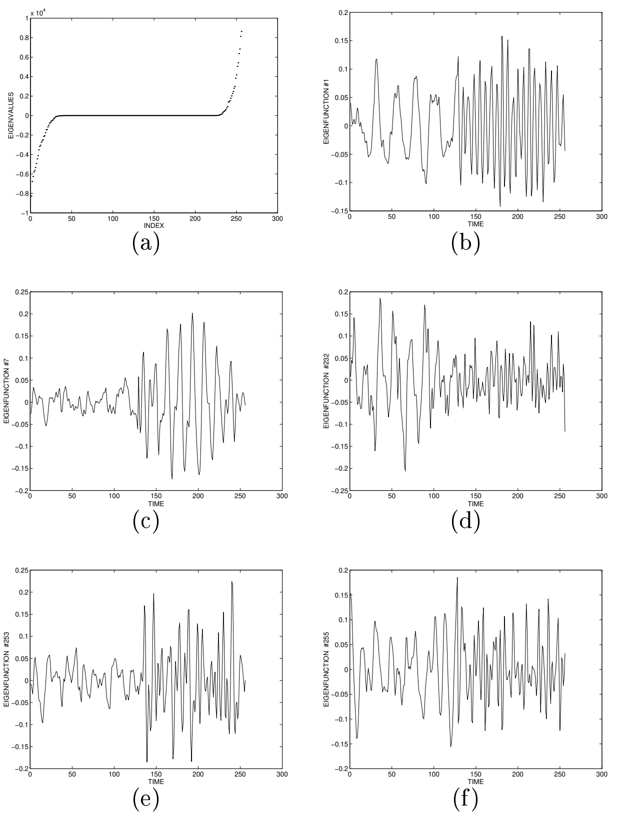


Figure 2. Eigenvalues and eigenfunctions of a 2-channel quadratic detector (256×256 matrix); the first 128 samples correspond to channel 1 and the second half to channel 2. (a) Eigenvalues $\lambda_k : k = 1, 2, \cdots 256$. (b) First eigenfunction: w_1 ($\lambda_1 < 0$). (c) w_7 ($\lambda_7 < 0$). (d) w_{232} ($\lambda_{232} > 0$). (e) w_{253} ($\lambda_{253} > 0$). (f) w_{255} ($\lambda_{255} > 0$).

6. CONCLUSIONS

In many real applications involving detection and classification, the signals of interest are nonstationary and often possess certain nuisance parameters which are either unknown or random. For example, the machine fault signatures occur at unknown times and often exhibit random modulations. The TFR-based optimum detection framework described in this paper provides powerful detector structures that can be successfully applied in such scenarios.

The subspace-based formulation in terms of a bank of spectrograms/scalograms illuminates the flexible structure of such TFR-detectors: they range from the simple rank-1 matched-filter detectors, optimal for detecting essentially deterministic signals, to the more complex higher-rank processors for the detection of arbitrary stochastic signals. Time-frequency shifts and scalings are the appropriate nuisance parameters and are naturally incorporated by such TFR detectors via a generalized likelihood ratio test.

Our experience with real data in machine fault diagnostics has shown significant promise for such TFR detectors; the fault signatures have unknown time-offsets and often exhibit frequency modulations and/or scalings. For an application to engine knock detection see [3]. In cyclostationary applications, such as those involving rotating machinery, we have found that time-alignment is crucial to the performance of fault detectors. In addition, scalings also seem to be relevant in biomedical applications such as sleep data classifications. Moreover, our experience has shown that the effective rank of the underlying signals is usually relatively low, making the subspace-based interpretation particularly useful for designing computationally efficient low-rank detectors which yield most of the performance gain [9].

Such reduced-rank structures could also be potentially useful in the *multi-sensor* detection that we briefly discussed in this paper; the detector size increases by a factor of N for N sensors. The simple multi-sensor structure explored in this paper has shown promising results with real data, suggesting that higher performance gains may be achievable with more sophisticated techniques. In addition to improved performance, the subspace-based interpretation of such multi-sensor detectors yields useful information about the underlying physical mechanisms that may in turn be exploited for improved signal processing.

REFERENCES

[1] R. A. Rohrbaugh, "Application of time-frequency analysis to machinery condition assessment", in *Proc. 27th Asilomar Conference on Signals, Systems, and Computers*, November 1993.

[2] L. P. Heck, "Signal processing research in automatic tool wear monitoring", in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. — ICASSP '93*, 1993, pp. I–55–I–58.

[3] 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.

[4] B. Samimy and G. Rizzoni, "Time-frequency analysis for improved detection of internal combustion engine knock", in *Proc. IEEE Int'l Symp. on Time-Frequency and Time-Scale Analysis*, 1994, pp. 178–181.

[5] 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.

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

[7] 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.

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

[9] A. M. Sayeed and D. L. Jones, "Optimal reduced-rank time-frequency/time-scale quadratic detectors", *To be presented at the IEEE Int'l Symp. on Time-Frequency and Time-Scale Analysis*, 1996.