1. INTRODUCTION

In [1], robust detection of a known signal in unknown noise was considered, and the approach was to assume a nominal noise density with contamination. Perhaps conversely, in [2] the noise was assumed perfectly modeled, but the signal to be detected was both random and of imperfectly known spectrum. In each case the least favorable class member (noise-density/signal-spectrum) was discovered, and the optimal detector for this constructed. Although this is by no means always the case, the detectors developed were shown to satisfy a saddle-point condition, meaning that they are as good as they can be in the “worst case” situation, and better still (although no longer optimal) in any other.

Marriage of these two types of robustness is of significant practical interest. For example, in passive systems, signals are usually modeled as random; for active systems, the received target echo may also bear some randomness due to unknown target characteristics (e.g., Doppler shift) and channel deterioration (e.g., multipath, clutter, active interference). In either case specification of the signal to be detected is impossible, and even specification of its spectrum is likely to be imprecise. Further, there will be noise, and an unwary assumption of well-behaved Gaussianity can be hazardous if there are “outlying” samples.

In random signal detection, the signal strength is usually unknown, and as such the implied hypothesis test is composite even when the signal spectrum is assumed known. With non-Gaussian noise, a uniformly most powerful (UMP) test usually does not exist, therefore alternative approaches must be adopted. We consider here the locally optimal (LO) approach which is suitable for detecting weak signals. In this context, a natural measure of detector performance is the efficacy of the test statistic. A well-known property of the LO detector is that it maximizes the efficacy [3, pp. 42–46].

The LO detector structure is of the generalized-correlator type [4], meaning that data undergoes a memoryless and probably nonlinear transformation (whose form is closely tied to the noise density) prior to filtering (the form of the filter is likewise closely tied to the spectrum of the signal to be detected). There are thus two inseparable aspects to the robustness we seek: that to noise and that to signal uncertainty. We develop a detector robust to both.

The manuscript is organized as follows. The LO detector for stochastic signals is presented in the next section, and its efficacy is derived explicitly. In Section III we give further background on robustness issues, and subsequently deal with the robust selection of both noise density and signal spectrum. The saddle-point condition is shown to support the robustness argument. In Section IV we give an
example to illustrate the usefulness of this approach. Section V contains some concluding remarks.

II. LOCALLY OPTIMAL DETECTOR

A. Background

The problem of deciding on the presence or absence of a signal embedded in additive noise is usually treated as a statistical hypothesis-testing operation; that is, a receiver must choose between the noise-alone hypothesis $H$ and the signal-plus-noise alternative $K$. Assuming discrete-time processing is to be used, these can be written as

$$ H : X_i = N_i $$

$$ K : X_i = N_i + \theta S_i \quad i = 1, 2, \ldots, n $$

(1)

where $\{X_i\}_{i=1}^n$ is the observation process, $\{N_i\}_{i=1}^n$ is a white noise process, $\{S_i\}_{i=1}^n$ is the signal to be detected. The parameter $\theta$ is a fixed constant; in what follows we must be able to manipulate the strength of the signal explicitly while retaining all other aspects of its stochastic structure, and $\theta$ provides a convenient means to do this. Although extension to the complex case is straightforward, we assume hereafter that all quantities are real. The decision optimal in the Neyman–Pearson sense (maximized probability of detection for a given false-alarm rate) is of course derived from the comparison of the likelihood ratio of the observations to a threshold [5, pp. 33–34]. In the case that the time series is vector-valued, the detection problem can be reposed as

$$ H : X_{ij} = N_{ij} $$

$$ K : X_{ij} = N_{ij} + \theta S_i \quad i = 1, 2, \ldots, n $$

(2)

where $X_{ij}$ is the $i$th observation from the $j$th sensor. If underlying our vector notation is an array of sensors, then it must be assumed that this array is “steered” such that a given (stochastic) noise sample is present at each element at the same time.

In both passive and active systems, due to the unknown channel characteristics, the signal strength $\theta$ is usually not known. As such, the test is composite in nature and with non-Gaussian noise, no UMP test exists. One approach is to use a generalized likelihood ratio test (GLRT), which involves estimating whatever parameters are unknown and hence treating that estimate as perfect and the hypotheses again noncomposite. The GLRT often works well; however, it is optimal only in an asymptotic sense under some restrictive conditions, and in any case is unsuitable for our purpose due to its lack of a manipulable performance measure. Hence here we adopt an alternative approach which is optimal for weak signal detection (strictly, for signals with vanishing strength)—the LO detector—with the motivation that such a detector performs well even if the signal strength is large. For one-sided testing, the LO test statistic is shown, using the generalized Neyman–Pearson Lemma, to be

$$ T(x) = \frac{\partial}{\partial \theta} \left[ \log(p(x | \theta)) \right]_{\theta=0} $$

(3)

where $p(x | \theta)$ is the probability density of the observation sequence $x = \{X_{ij}\}_{j=1}^n$ given $K$ is true and $\theta$ is specified [3, p. 7].

In our case, $p(x | \theta)$ can be rewritten as

$$ p(x | \theta) = \int f(x - \theta s) f_\theta(s) ds $$

(4)

where $f(\cdot)$ and $f_\theta(\cdot)$ are, respectively, the noise and signal densities. However, if we assume that the signal has a density which is symmetric about the origin (i.e., $f_\theta(s) = f_\theta(-s)$) then we can write

$$ p(x | \theta) = p(x | (-\theta)) $$

(5)

which implies that if (3) is to be used,

$$ T(x) = 0. $$

(6)

This, in fact, corresponds to two-sided testing (i.e., $\theta = 0$ versus $\theta \neq 0$). Consequently the LO detector must be modified. Again, the generalized Neyman–Pearson Lemma can be applied to derive the LO test statistic which turns out to be the second derivative of log likelihood,

$$ T(x) = \frac{\partial^2}{\partial \theta^2} \left[ \log(p(x | \theta)) \right]_{\theta=0} $$

(7)

A LO test is most powerful for detecting signals with vanishing strength (i.e., $\theta \to 0$) regardless of record length $n$. However, the performance of a LO test is most easily computed in the asymptotic case $n \to \infty$. Given that the extended regularity conditions given in [3, p. 15] are satisfied, then the performance measure is the efficacy, defined in this case as

$$ \xi = \lim_{n \to \infty} \frac{1}{4n} \left[ \frac{\hat{\mu}_n(0)}{\sigma_n(0)} \right]^2 $$

(8)

where

$$ \hat{\mu}_n(\theta) = \frac{\partial^2}{\partial \theta^2} [E_{\theta}(T_n(x))] $$

$$ \sigma_n^2(\theta) = \mathcal{V}_{\theta}(T_n(x)) $$

where $E$ and $\mathcal{V}$ denote expectation and variance, and where $T_n(x)$ is the LO test statistic for $n$ samples. It is readily seen that the efficacy is the appropriate per-sample signal-to-noise ratio of the test statistic; but in any case the Pitman–Noether Theorem states that the asymptotic relative efficiency (ARE) of two detectors (that is, the asymptotic ratio of the number of samples needed to achieve the same performance) can, under fairly loose regularity conditions, be
expressed as the quotient of the detectors’ efficacies [3]. More precisely, we have
\[
\text{ARE}_{A:B} = \lim_{\theta \to 0} n_A(\alpha, \beta, \theta) = \frac{\xi_A}{\xi_B}
\]
(9)
in which \(n_A(\alpha, \beta, \theta)\) denotes the number of samples needed by detector \(A\) to achieve false-alarm rate \(\alpha\) and probability of detection \(\beta\) given signal strength \(\theta\), and \(n_B(\cdot)\) is defined similarly. In Section IV we show the relationship between efficacy and receiver operating characteristic.

B. Locally-Optimal Detector for Stochastic Signals

Using (7) it is possible (although not straightforward) to express the LO test statistic in a form suitable for computation. However, due to the derivative-taking operations involved, it is necessary to make some specifications on the differentiability of the noise density \(f(\cdot)\).

**Assumption 1.** The density \(f(\cdot)\) and its derivative \(f'(\cdot)\) are continuous and differentiable except at a finite number of points.

Discrete time detection of an independent identically distributed (IID) stochastic signal in white noise in the single-sensor (i.e., \(L = 1\)) case is relatively straightforward. We write
\[
p(x | \theta) = \int_{R^L} \prod_{i=1}^{n} f(X_i - \theta S_i) f_S(S_i) dS_i
\]
(10)
and it is easy to show, noting \(E(S_i) = 0\), that we have
\[
T(x) = \sum_{i=1}^{n} \left[ f(X_i) \int f'(X_j) \right] \left[ \frac{f'(X_i)}{f(X_i)} \right]
\]
(11)
where \(\tilde{f}\) refers to the second derivative of the noise density \(f\). We can write
\[
\xi = \frac{1}{4} \left[ E(S_i^2) \right] \left[ \int \frac{f''(x)}{f(x)^2} dx \right] \]
(12)
for the efficacy in this case.

It is often more realistic to assume that the noise is not white. Kassam in [4] has used a model built on the following.

**Assumption 2.** The noise \(\{N_i\}_{i=1}^{n}\) is IID with symmetric univariate density \(f(\cdot)\). The signal is wide-sense stationary with \(\{S_i\}_{i=1}^{n}\) having density \(f_S(s)\), with \(E(S_i) = 0\), \(E(S_i^2) = 1\), and \(E(S_i S_{i+k}) = r(k)\).

Note that due to the multiplication by \(\theta\), the unit-variance condition of Assumption 2 is not restrictive.

For the general (vector) case we write
\[
p(x | \theta) = \int \prod_{i=1}^{n} \prod_{j=1}^{L} f(X_{ij} - \theta S_i) f_S(S_i) dS_i.
\]
(13)

The LO test statistic for such a model has been shown [4] to be
\[
T(x) = \sum_{i=1}^{n} \left[ \sum_{j=1}^{L} h_{lo}(X_{ij}) \right] + \sum_{i=1}^{n} \sum_{m=1}^{n} r(i-m) \left[ \sum_{j=1}^{L} g_{lo}(X_{ij}) \right] \times \left[ \sum_{j=1}^{L} g_{lo}(X_{pm}) \right]
\]
(14)
where
\[
h_{lo}(x) = \frac{\tilde{f}(x)}{f(x)} \left( \frac{f'(x)}{f(x)} \right)^2
\]
(15)
\[
g_{lo}(x) = \frac{\tilde{f}(x)}{f(x)}
\]
(16)
and \(f(\cdot)\) and \(\tilde{f}(\cdot)\) are first two derivatives of \(f(\cdot)\). Note that in the case \(L = 1\) (a single sensor) this reduces to
\[
T(x) = \sum_{i=1}^{n} h_{lo}(X_i) + \sum_{i=1}^{n} \sum_{m=1}^{n} r(i-m) g_{lo}(X_i) g_{lo}(X_m)
\]
(17)
and that in the white noise case this reduces to (11). It is interesting that (14) and (17) depend only on the correlation structure of the signal to be detected, and not on any more detailed statistical description. Of further note is the appearance of an “energy” sum \(e_{lo} = h_{lo} + g_{lo}^2\) (for this purpose the \(i = m\) terms in the second summations are properly associated with the first summation) and a quadratic form comparing measured autocorrelation (the nonlinear distortion \(g_{lo}\) can be considered as improving this measurement) to expected autocorrelation.

C. Efficacy of Locally-Optimal Detector

Computing the performance of this test is not trivial, and previous treatments have dealt only with the white noise case. It is helpful to derive the efficacy of a generalized correlator
\[
T(x) = \sum_{i=1}^{n} \left[ \sum_{j=1}^{L} h(X_{ij}) \right] + \sum_{i=1}^{n} \sum_{m=1}^{n} \rho(i-m) \left[ \sum_{j=1}^{L} g(X_{ij}) \right] \times \left[ \sum_{j=1}^{L} g(X_{pm}) \right]
\]
(18)
which has the same structure as the LO detector. For best performance \(h\) and \(g\) ought to take on their LO values of (15) and (16) according to the actual noise density \(f\); and \(\rho\) should be the actual signal autocorrelation \(r\). However, these may be unknown, and in order that a robust detector be identified it is important to develop an expression for
the performance of an arbitrary statistic having the structure of (18). We rely on the following.

**Assumption 3** Let the nonlinearity \( h \) have even symmetry (i.e., \( h(x) = h(-x) \)), and likewise \( g \) be odd-symmetric (i.e., \( g(x) = -g(-x) \)). Let both
\[
\int g^2(x)f(x)dx \quad \text{and} \quad \int [h(x) + g^2(x)]^2 f(x)dx
\]
each exist and be finite. Further, let the signal process be ergodic, so that \( \lim_{n \to \infty} \sum_{k=1}^{n} r^2(k)(1 - 1/n) \) converges and is equal to \( \sum_{k=1}^{\infty} r^2(k) \).

**Theorem 1** The efficacy of a statistic of the form (18) is given by
\[
\xi = \frac{L}{4} \left( \frac{\int [h + g^2]f}{f} + 2 \left[-1 + \frac{L}{2\pi} \int_{-\pi}^{\pi} \phi_1(\omega)\phi_2(\omega) d\omega \right] \left[ \int g f \right]^2 \right)^2
\]
(19)
when Assumptions 2 and 3 are satisfied.

**Proof** This was stated without proof in [6]; see Appendix A.

We can use Parseval’s relation to write
\[
\sum_{k=-\infty}^{\infty} r_1(k) r_2(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_1(\omega) \phi_2(\omega) d\omega
\]
(20)
in which
\[
\phi(\omega) = \sum_{k=-\infty}^{\infty} r(k) e^{-j\omega k}
\]
is the spectrum corresponding to an autocorrelation sequence \( r \). We can identify \( r_1 \) and \( r_2 \) variously as \( r \) (the actual correlation) and as \( \rho \) (the assumed correlation), and can hence rewrite (19) as
\[
\xi = \frac{L}{4} \left( \frac{\int [h + g^2]f}{f} + 2 \left[-1 + \frac{L}{2\pi} \int_{-\pi}^{\pi} \phi_1(\omega)\phi_2(\omega) d\omega \right] \left[ \int g f \right]^2 \right)^2
\]
(21)
It is instructive to note the following.

1) With an accurate estimate of the signal correlation (i.e., \( \rho(k) = r(k) \)), (19) reduces to
\[
\xi = \frac{L}{4} \left( \frac{\int [h + g^2]f}{f} + 2 \left[-1 + \frac{L}{2\pi} \int_{-\pi}^{\pi} \phi_1(\omega)\phi_2(\omega) d\omega \right] \left[ \int g f \right]^2 \right)^2
\]
(22)

2) With an accurate estimate of the noise density \( f \), we can write \( g = g_{10} \) and \( h = h_{10} \); the efficacy expression (19) thus reduces to
\[
\xi = \frac{L}{4} \left( \frac{\int [h + g^2]f}{f} + 2 \left[-1 + \frac{L}{2\pi} \int_{-\pi}^{\pi} \phi_1(\omega)\phi_2(\omega) d\omega \right] \left[ \int f \right]^2 \right)^2
\]
(23)
where \( I(f) = \int (f)^2 f \) is Fisher’s information.
3) The efficacy of a LO detector can be written as
\[
\xi = \frac{L}{4} \left( \frac{\int [h + g^2]f}{f} + 2 \left[-1 + \frac{L}{2\pi} \int_{-\pi}^{\pi} \phi_1(\omega)\phi_2(\omega) d\omega \right] \left[ \int f \right]^2 \right)^2
\]
(24)
In the latter two cases we have used the fact that \( E_0 \{ h_{10} + g_{10} \} = 0 \). In general \( E_0 \{ h + g^2 \} \neq 0 \). As expected, the performance improves with increased correlation between signal samples. It is instructive to observe (for the LO case) that with \( r(k) \equiv 1 \), the efficacy is infinite; this is tantamount to the known signal problem, for which the efficacy of (8) is inappropriate and must be modified [3].

The equations above represent different levels of knowledge about the statistical test. In (24) all parameters are known and the (LO) test is used. In (23) the noise density \( f \) is known, but the precise form of the signal correlation structure is not. This is a situation in which it is reasonable to seek a robust \( \rho \); the performance of whose corresponding detector is not degraded as the actual correlation \( r \) varies over its uncertainty class. In (22) the correct signal correlation is known but the noise density is not, and it is reasonable to seek a detector whose nonlinearities \( g \) and \( h \) engender a similar lack of degradation as the actual noise density \( f \) varies. The situation of interest in this work is (21): neither the signal correlation structure nor the noise density is known, and a detector with joint robustness is sought.

### III. ROBUSTNESS

#### A. Review on Robustness Issue

Statistical inference not only refer to the final decision procedure, but also includes the a priori assumption of the distributional models, either explicit or implicit, of the observations being studied. Problems arise when the true underlying statistical models deviate from the nominal assumptions about noise and/or signal.

Robustness signifies insensitivity against small deviations from our assumptions regarding the noise and/or the signal. The ideas were first propounded by Huber [7, 8] in a general context of statistical inference, and later the subject of a wide literature in signal detection (e.g., [1, 4]). For example, suppose we are to estimate the location parameter from some
IID observations. With the noise Gaussian, the sample mean is the best unbiased estimate; but if the true noise density deviates only a little from normal, use of the sample mean can be catastrophic.

Minimax is not the only kind of robustness, but it seems to have drawn the most attention. There are three steps to finding a minimax robust approach.

1) Identify the Uncertainty Class: This may seem straightforward, but as will be seen, is often not, since great care must be taken at this stage that a solution exist. For example, in testing for a known signal with efficacy as a metric, the usual uncertainty class is all additive noise densities expressible as mixtures of a known “nominal” and unknown “contaminant”

\[ f(x) = (1 - \epsilon)f_{\text{norn}}(x) + \epsilon f_{\text{con}}(x). \]

There are some technical constraints; but there is also the condition that the nominal density be “strongly unimodal”, meaning that \( d^2/dx^2 \log(f_{\text{norn}}(x)) < 0 \).

Without this condition there is no robust solution. In our problem we have to impose this and another two conditions.

2) Find the Worst Element of this Class: With the best detector for each class element derived, we select the pair (class-element/detector) which has the worst performance; we have thus solved the minimax problem. Returning to the known-signal example, it can be shown via variational calculus that the worst-case class-element has an exponential density “tail.”

3) Prove Robustness: Finding the minimax solution may not be meaningful if the class has been chosen carelessly, since its performance may be poor at its design point and even worse elsewhere. It is necessary to show that the minimax solution is also maximin, meaning that deviation from the design point can only improve performance. In the known-signal example this can be done directly and in a similar manner to what follows.

Geometrically and intuitively speaking, the minimax solution is a saddle-point, with behavior, respectively, convex and concave along the “dimensions” of uncertainty class and detector structure.

There has been considerable study of minimax robustness in signal detection and estimation, and an excellent literature survey is available in [9]. For example, in [10] the spectral uncertainty class that we adopt is explored in the context of estimation. As mentioned, there has been considerable examination of robustness in the detection of known signals in both independent and dependent noise (e.g. [1, 8, 11]). There has been extensive work in the robust detection of stochastic signals (e.g. [4, 12, 13]) which has generally dealt with the white-signal case or with constrained detector structures. Most works on robustness use efficacy as a performance measure; but treatments using asymptotic error exponents [14], elegant alternative measures (e.g. [15]), and robustness with respect to signal parameters (e.g. [16]) all make excellent contributions.

In this work we deal with signal correlation explicitly, both in the form of the detector and in the sense of an associated uncertainty class. The work in [4] is relevant, but relies on greater constraint in the noise uncertainty class, and does not treat the case of signal uncertainty. In [18] the problem presented in this work is broached and discussed, but no explicit solution is given.

B. Robust Choice of Signal Spectrum

Now let us consider the practical application of this result. In the succeeding section we deal with the case that \( f(\cdot) \) is unknown; for now let us assume that it is known. We assume further that the signal correlation is not known accurately, and pose the problem of finding a minimax detector, one which is optimal for the least favorable signal correlation possible. We call such a detector robust and show that it will perform reasonably well (is maximin) over a wide class of stochastic signals.

Clearly, if nothing whatever is known about \( r(k) \), the robust detector is that corresponding to \( r(k) = 0, \ k \neq 0 \). This is a somewhat trivial result; the worst case signal is white. It is reasonable to assume, however, that some knowledge is available about signal correlation, and proceeding in a manner similar to [2, 10, 19] we define a class \( \Phi(\phi_U, \phi_L) \) of allowable signal spectra \( \phi(\omega) \) as

\[ \Phi(\phi_U, \phi_L) = \left\{ \phi(\omega) : [\phi_U(\omega) \leq \phi(\omega) \leq \phi_L(\omega)] \right\} \]

\[ \Phi(\phi_U, \phi_L) = \left\{ \phi(\omega) : \int_{-\pi}^{\pi} \phi(\omega) d\omega = 1 \right\} \] (25),

\[ \Phi(\phi_U, \phi_L) \] defines the set of possible signal spectra with unity variance and is upper- and lower-bounded by \( \phi_U \) and \( \phi_L \), respectively. Note that the latter condition corresponds to Assumption 2 of the model, and that we must specify

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_U(\omega) d\omega \geq 1 \] (26)

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_L(\omega) d\omega \leq 1 \] (27)

for the set \( \Phi \) to be nonempty. From the discussion in the previous section, the immediate task is seen to be selection of the least favorable signal spectrum (that which minimizes its LO efficacy); and from (24), it is clear that this is the \( \phi(\omega) \) that minimizes \( \int_{-\pi}^{\pi} \phi^2(\omega) d\omega \), i.e.,

\[ \phi^*(\omega) = \arg \min_{\phi \in \Phi(\phi_U, \phi_L)} \left\{ \int_{-\pi}^{\pi} \phi^2(\omega) d\omega \right\} \] (28)
As mentioned in [6], the solution to this minimization can be obtained from variational principles as

$$\phi^*(\omega) = \begin{cases} 
\phi_U(\omega) & \phi_U(\omega) \leq \lambda \\
\lambda & \phi_U(\omega) < \lambda < \phi_U(\omega) \\
\phi_L(\omega) & \phi_U(\omega) \geq \lambda 
\end{cases}$$

(29)

where \(\lambda\) is chosen to satisfy

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \phi^*(\omega) d\omega = 1.$$  

(30)

We note that this solution is identical to that derived in [10] for the robust Wiener filter in Gaussian noise; it is also similar to that in [2] for the Eckart filter.

It remains to be shown that this choice of signal spectrum is truly robust, as would be indicated by the saddle-point condition:

$$\xi_{f\phi^*}(T_{f\phi}) \leq \xi_{f\phi}(T_{f\phi}) \leq \xi_{f\phi^*}(T_{f\phi^*})$$

(31)

for all \(\phi \in \mathcal{F}(\phi_U, \phi_L)\). Here \(\xi_{f\phi}(T_{f\phi})\) denotes the efficacy of a statistical T based on an assumption of signal spectrum \(\phi_2\) and noise density \(f_2\) when the actual signal spectrum and noise density are \(\phi_1\) and \(f_1\), respectively; more precisely, \(\xi_{f\phi}(T_{f\phi})\) is given by (21) with the substitutions \(g = -f_2/f_2\), \(h = f_2/f_2 - (f_2/f_2)^2\), \(f = f_1\), \(\phi = \phi_1\), and \(\phi = \phi_2\). Equation (31) means that the robust detector (that corresponding to \(\phi^*\)) has higher efficacy than any other detector when the actual spectrum is indeed \(\phi^*\), and whose efficacy is at least this value, although perhaps no longer optimal, when the actual spectrum is \(\phi \neq \phi^*\).

The first inequality follows from the fact that \(T_{f\phi^*}\) is LO. We prove the second inequality in the following.

**Theorem 2** For \(\phi \in \mathcal{F}(\phi_U, \phi_L)\) and \(\phi^*\) as given in (29), and any \(f\),

$$\xi_{f\phi^*}(T_{f\phi^*}) \leq \xi_{f\phi}(T_{f\phi})$$

(32)

with equality if and only if

$$[\phi(\omega) - \phi^*(\omega)][\phi^*(\omega) - \lambda] \equiv 0$$

(33)

for all \(\omega\).

**Proof** For any \(\phi \in \mathcal{F}(\phi_U, \phi_L)\), write

$$\phi(\omega) = \phi^*(\omega) + \tilde{\phi}(\omega)$$

where we know

$$\int_{-\pi}^{\pi} \tilde{\phi}(\omega) d\omega = 0.$$

from the unit-power condition on the signal. Now form the sets:

\(A = \{\omega : [-\pi \leq \omega \leq \pi] \cap [\phi_U(\omega) < \lambda]\}\)
\(B = \{\omega : [-\pi \leq \omega \leq \pi] \cap [\phi_U(\omega) \leq \lambda \leq \phi_U(\omega)]\}\)
\(C = \{\omega : [-\pi \leq \omega \leq \pi] \cap [\phi_U(\omega) > \lambda]\}\)

From (23) it is clear that \(\xi_{f\phi}(T_{f\phi})\) is increasing in

$$\int_{-\pi}^{\pi} \phi^*(\omega) d\omega.$$  

We can write

$$\int_{-\pi}^{\pi} \phi(\omega)\phi^*(\omega) d\omega - \int_{-\pi}^{\pi} \phi^*(\omega)^2 d\omega$$

$$= \int_A \tilde{\phi}(\omega)\phi_U(\omega) d\omega + \int_B \tilde{\phi}(\omega)\lambda d\omega$$

$$+ \int_C \tilde{\phi}(\omega)\phi_U(\omega) d\omega$$

$$\geq \int_A \tilde{\phi}(\omega)\phi_U(\omega) d\omega + \int_B \tilde{\phi}(\omega)\lambda d\omega + \int_C \tilde{\phi}(\omega)\lambda d\omega$$

$$= \int_A \tilde{\phi}(\omega)\phi_U(\omega) d\omega - \int_A \tilde{\phi}(\omega)\lambda d\omega$$

$$= \int_A \tilde{\phi}(\omega)[\phi_U(\omega) - \lambda] d\omega$$

$$\geq 0.$$  

This last inequality results from the fact that on \(\omega \in A\), both \(\phi(\omega) \leq 0\) and \(\phi_U(\omega) \leq \lambda\). There is equality if and only if \(\phi(\omega) \equiv 0\) on \(\omega \in A \cup C\).

C. Robust Selection of Noise Density

In the previous section we discussed the problem of stochastic signal detection in the case that the noise density \(f\) was known, but there was uncertainty as to the actual signal spectrum \(\phi(\omega)\). Here we begin by exploring the converse problem, that in which the signal spectrum is known but the actual noise density is uncertain. A popular approach in modeling noise uncertainty is to use an \(\epsilon\)-contamination class.

As such, we restrict attention to the class of noise densities given by the following.

**Assumption 4** The noise density \(f\) satisfies

**Assumption 1** and is a member of a class

$$\mathcal{F}_\epsilon = \{f : f(x) = (1 - \epsilon)\alpha(x) + \epsilon\beta(x)\}$$

where \(\alpha(x)\) is symmetric and strongly unimodal density function (\(d^2/dx^2\log(\alpha(x)) \leq 0\)). We also assume that \(\alpha(x)/\alpha(x)\) is convex. Further, with a determined by

$$\int_{-\alpha}^{\alpha} \alpha(x) dx = 2\alpha(a)^2/\alpha(a) = (1 - \epsilon)^{-1}$$

(34)

we have

$$\left(\frac{\alpha(x)}{\alpha(a)}\right)^2 \geq \frac{\alpha(x)}{\alpha(x)} \quad \text{for} \quad |x| \leq a.$$  

(35)

The contaminant density \(\beta(x)\) is symmetric but otherwise arbitrary, and the “mixing proportion” \(\epsilon \in \mathbb{R}\) is bounded as \(0 \leq \epsilon \leq 1\).

Both the “nominal” density \(\alpha(x)\) and the proportion of contaminating samples \(\epsilon\) are known. This assumption allows us to derive the following result.
THEOREM 3 Suppose the robust spectrum $\phi^*$ is used in the correlator. Define

$$f^*(x) = \begin{cases} 
(1 - \varepsilon)\alpha(x) & |x| \leq a \\
(1 - \varepsilon)\alpha(a) \exp \left( \frac{\alpha(a)(|x| - a)}{\alpha(a)} \right) & |x| > a 
\end{cases}$$

Then the correlator with nonlinearities being optimal for $f^*(x)$ is minimax robust.

PROOF See Appendix B.

The use of robust signal spectrum is not always required. For example, use of the true signal spectrum (provided it is available) also warrants the minimax optimality of the defined noise density. The requirement of the use of robust signal spectrum is more or less of technical requirement for the proof of the theorem.

The robust density is of exactly the same form as that derived in [4] except for some extra constraints on $\alpha(\cdot)$; this is not surprising given the extra level of complication in our problem. These extra constraints (convexity of $\alpha(x)/\alpha(x)$ and the technical condition of (35)) are non-trivial, but are satisfied for a rich class of nominal densities $\alpha(x)$ provided the contaminant proportion $\varepsilon$ is not too large. Constraint (35) in fact defines the range of possible values for $a$, which in turn, from (34), determines the possible range of the contamination constant $\varepsilon$. For example, with a Gaussian nominal, it is straightforward to show that the resulting range for $a$ and $\varepsilon$ are $a > 1$ and $\varepsilon < 0.14$. In most robustness studies, an accepted value of $\varepsilon$ seems to be between 0.01 and 0.1 [7].

D. Joint Robustness

It is straightforward to apply Theorems 2 and 3 into a more general robustness statement, given in the following.

THEOREM 4 With $f^*$ as defined in (36) and $\phi^*$ as defined in (29), the LO detector $T_{f^*}$ is truly robust; that is,

$$\xi_{f^*}(T_{f^*}) \leq \xi_{f^*}(T_{f^*}) \leq \xi_{f^*}(T_{f^*})$$

for all $f \in \Phi_{\Upsilon, \Phi_L}$ and all $f \in F_{\alpha}$.

PROOF The first inequality follows from the fact that $T_{f^*}$ is LO. For the second, write

$$\xi_{f^*}(T_{f^*}) \leq \xi_{f^*}(T_{f^*}) \leq \xi_{f^*}(T_{f^*})$$

where both Theorems 2 and 3 have been used.

IV. APPLICATIONS AND RESULTS

A. Example 1. Detection of Target with Unknown Doppler in Non-Gaussian Noise

As an example of the application of the robust detector just developed, let us consider the following problem. Suppose we are to detect a target with unknown Doppler shift using single pulse radar. For concreteness, we specify: the radar RF $f_z = 4$ GHz, the maximum radial velocity $v_{\text{max}} = 600$ m/s, radar pulsewidth $\tau = 40$ $\mu$s, and the IF $f_d = 50$ MHz. With these specifications, it is straightforward to calculate the maximum Doppler shift is $f_{\text{d max}} = 16$ kHz. Thus the actual IF signal spectrum (parametrized by the Doppler shift $f_d$) is a sinc-squared function with center frequency $f_0 + f_d \in (f_0 - f_{\text{d max}}, f_0 + f_{\text{d max}})$, or more precisely

$$\phi(f; f_d) = \frac{\sin(\pi(f + f_0 + f_d)\tau)}{\pi(f + f_0 + f_d)\tau} + \frac{\sin(\pi(f - (f_0 + f_d)\tau)}{\pi(f - (f_0 + f_d)\tau}$$

as is shown in Fig. 1, and in which $c$ is a constant relating to the pulse energy. Clearly this results in an upper- and lower-bound of possible signal spectrum as shown by the dashed and dotted line in Fig. 2, and as calculated by

$$\phi_L(f) = \min_{f \in (-f_{\text{d max}}, f_{\text{d max}})} \{\phi(f; f_d)\}$$

$$\phi_U(f) = \max_{f \in (-f_{\text{d max}}, f_{\text{d max}})} \{\phi(f; f_d)\}$$

in which $\phi(f; f_d)$ is as given in (37). Also plotted in the same graph is the robust spectrum as obtained via Theorem 2. To further illustrate the idea, suppose now we know, based on some different prior knowledge, that the maximum radial velocity is 300 m/s. The resulting robust spectrum together with the upper and lower bound are plotted in Fig. 3. The reader should note that the spectra $\phi$ discussed above are according to “analog” frequency $f$, as seems most appropriate for intuition, as opposed to “digital” frequency $\omega$ of the theorems. Assuming a sufficiently high sampling frequency $f_s$, the conversion is $\omega = 2\pi f / f_s$, for $-\pi \leq \omega < \pi$. Since these plots are in analog frequency they have not been normalized according to (25)-(27).

For the noise process, we assume a normal mixture density

$$f(x) = (1 - \varepsilon) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + \frac{\varepsilon}{\sqrt{2\pi}} e^{-x^2/2\sigma^2}.$$  

(39)

Gaussian mixtures are widely used in modeling moderately heavy-tailed densities. The variance of the contaminant provides easy control of tail mass, and in our example we use $\sigma^2 = 16$. The variance increases linearly with $\varepsilon$ from unity to 16, while the kurtosis (the ratio of the fourth moment to the square of the second moment) is as shown in Fig. 4.

With respect to the model of (1) we have an observation sequence $\{X_t\}$ composed either of noise or noise plus signal. The actual noise $(\{N_t\})$ distribution is a mixture of Gaussians, as above. However, all that is known by the receiver is that the noise is nominally (i.e., with probability $(1 - \varepsilon)$) unit-Gaussian; the contaminant density happens to

WILLET & CHEN: ROBUST DETECTION OF SMALL STOCHASTIC SIGNALS
Fig. 1. Example 1 return signal spectra assuming minimum (negative) Doppler frequency shift (dashed); no Doppler (dashed); maximum (positive) Doppler (solid). Minimum and maximum of continuum of all such spectra for Doppler shifts between minimum and maximum are shown in Fig. 2 as upper and lower spectrum bounds.

Fig. 2. Robust signal spectrum for Example 1, with maximum radial velocity 600 m/s.

Fig. 3. Robust signal spectrum for Example 1, with maximum radial velocity 300 m/s.

Fig. 4. Kurtosis for Gaussian mixture of Example 1, plotted as function of mixing proportion $c$. 

IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS VOL. 35, NO. 1 JANUARY 1999
be high-variance Gaussian, but the receiver does not know this. The $g$ and $h$ nonlinearities (see (18)) are shown in Figs. 5 and 6.

It is assumed that a radar CF pulse (a constant-frequency and rectangular-envelope waveform) has been transmitted, and that the presence or absence of a reflecting target at a certain range is being tested. It is standard to assume that the range is accounted for coarsely by a bulk delay; but that accounting for fine range, as represented by the carrier phase, is impossible. Consequently, if a target is present at the specified range, the received signal is stochastic in that it is a CF pulse of unknown and uniformly distributed phase. Since there may be relative
target/receiver motion, a Doppler shift $f_d$ can be introduced; this frequency shift is not random, but is instead an unknown parameter of the received signal, with the result that the actual power spectral density (loosely, the "spectrum") is given by (37). It is thus appropriate to introduce an uncertainty class in which the true spectrum is to be found, which we have done by specifying that it lies between upper bound $\phi_U(f)$ and lower bound $\phi_L(f)$, with these calculated according to (38).

In this example there is uncertainty both as to noise density and to signal spectrum, and hence use of a robust detector structure is worth investigation. Fig. 7 shows the AREs (see (9), and note that the efficacies are calculated according to (8)) of the robust detector versus LO (known noise density and signal spectrum) detector, "linear" detector (that is, the detector which assumes Gaussian noise) with zero Doppler shift spectrum, and linear detector with robust signal spectrum. In these plots the actual noise density is the Gaussian mixture of (39) with $\varepsilon = 0.1$; and the actual Doppler shift is varied. The solid line compares the robust detector (whose structure is fixed) to the LO detector (whose structure varies according to the Doppler shift, and is hence not a realistic detector but instead serves as an upper bound on performance). It appears that a constant 3 dB is lost compared with this clairvoyant optimum, which is not surprising in light of Fig. 2. The dotted line compares the robust detector to the linear detector with the robust signal spectrum. From this we observe a considerable gain in robustness to the heavy-tailed noise. Finally, we compare the robust detector to the most naive detector, that which assumes Gaussian noise and zero Doppler shift (i.e., its assumed signal spectrum is the middle element of Fig. 1), and we not unnaturally note that this latter detector suffers most as the Doppler shift increases. The analysis is repeated in Fig. 8, except we here have $\sigma^2 = 9$. Note the heavier the noise tail, the greater the improvement of the robust versus Gaussian-assumption detector.

Heavy-tailedness can also be controlled using the contamination constant $\epsilon$. For $\sigma^2 = 16$, we plot in Fig. 9 the ARE as a function of $\epsilon$ for the robust versus linear detector. The signal spectrum is assumed known to both, and has zero Doppler shift. It is obvious that the larger $\epsilon$, the greater the preferableity of the robust detector.

B. Example 2. Detection of Random Telegraph Signal with Uncertain Switching Probability

As a further example, consider the detection of a discrete-time random telegraph signal (see, for example [17, p. 290]) in additive noise. Random
telegraph signals are Markov in their dependency structure and have unity magnitude; more specifically, \( \Pr(s_i = -s_{i-1} \mid s_{i-1}) = 1 - \Pr(s_i = s_{i-1} \mid s_{i-1}) = \nu \) for time index \( i \), and from this it can be shown that the correlation sequence is \( r(k) = (1 - 2\nu)^k \). The uncertainty class is parametrized by \( \nu \), and we take \( \nu \in [0.1, 0.3] \) in this example. Thus we have

\[
\phi(\omega; \nu) = \frac{(1 - 2\nu) - 1/(1 - 2\nu)}{2\cos(\omega) - ((1 - 2\nu) + 1/(1 - 2\nu))}. \tag{40}
\]

Following (38), we have

\[
\phi_L(\omega) = \min_{\nu \in [0.1, 0.3]} \{\phi(\omega; \nu)\}
\]

\[
\phi_U(\omega) = \max_{\nu \in [0.1, 0.3]} \{\phi(\omega; \nu)\}. \tag{41}
\]

The minimum, maximum, and robust spectra are shown in Fig. 10.

For variety, we have as our noise a mixture of nominal logistic and high-variance Gaussian contaminant:

\[
f(x) = (1 - \epsilon) \frac{e^{-x}}{(1 + e^{-x})^2} + \epsilon \frac{1}{\sqrt{2\pi} \sigma} e^{-x^2/2\sigma^2}. \tag{42}
\]

The contaminating variance \( \sigma^2 \) is in this example 100. For the logistic nominal, Assumption 4 is satisfied for approximate range \( \epsilon < 0.108 \), meaning that the breakpoint \( \alpha > 1.76 \). In this case we use \( \epsilon = 5\% \). In Fig. 11 we show the robust nonlinearities \( g^* \) and \( h^* \).

The twin purposes of this example are to show some variety in the robustness classes and to demonstrate the applicability of the efficacy measure. Efficacy is an asymptotic signal-to-noise ratio. Since the Central Limit Theorem is implicit to its development, we compare in Fig. 12 the simulated receiver operating characteristic for \( \theta = 0.5, 0.75, \) and \( 1 \) (the total signal energy is \( \theta^2 \)) to the Gaussian approximation

\[
P_d = Q(Q^{-1}(1 - P_f) - \theta^2 \sqrt{n\xi}). \tag{43}
\]

In the above \( P_d \) and \( P_f \) refer to probabilities of detection and false alarm, respectively, and \( Q \) to the unit-normal tail probability. To understand (43), consider that the correct signal-to-noise ratio for detecting a signal of strength \( \theta \) is \( (\mu(\theta) - \mu(0))/\sigma(\theta) \) (see (8)). Use of the efficacy as in (8) requires that for small \( \theta \), both \( \sigma(\theta) \approx \sigma(0) \) and \( \mu(0) = 0 \). We thus have

\[
\frac{\mu(\theta) - \mu(0)}{\sigma(\theta)} \approx \frac{1}{2} \theta^2 \frac{\mu(0)}{\sigma(0)} = \sqrt{n\theta^2} \left[ \frac{\mu(0)}{\sigma(0)} \right]^2 = \theta^2 \sqrt{n\xi}. \tag{44}
\]
and (43) follows. Note from (43) the square-root relationship between efficacy and $\theta^2$; an ARE of 4 indicates that the less efficient detector requires twice the signal-to-noise ratio to achieve the same performance as compared with the more efficient detector.

In Fig. 12 we have used $n = 256$ time samples, $L = 1$ channel, and $10^5$ Monte Carlo runs. Specifically, noise samples were generated according to (42), processed according to (18) using robust spectrum and nonlinearities of Figs. 10, 11, and 12, and a histogram taken of the test statistic under this noise-alone (H) hypothesis. Similar processing was performed for the noise-plus-signal alternative (K), with the difference that a random telegraph signal (see above) having “switching” probability $\nu = 0.25$ (recall that the uncertainty class derives from the assumption $\nu \in (0.1, 0.3)$) is multiplied by the appropriate scalar $\theta$ and added to the noise prior to application of (18). The receiver operating characteristics were obtained in a straightforward manner from the pair of histograms. At any rate, it is clear that the efficacy does indeed provide an accurate approximation for the performance for small signals, and even for relatively large signals it is useful.

C. Application Note

It is clear that a robust detector for a stochastic signal can offer a great improvement in performance. However, the calculation of a statistic of the form (18) might seem unattractive computationally, particularly when compared with a (nonrobust) Gaussian-assumption detector which can be implemented efficiently via a linear-filter/envelope-detector combination [5]. Specifically we note that if (18) is computed directly it would require at least $O(2n^2)$ multiplications. However, if we write

$$\Phi(k) = \sum_{m=-n}^{n-1} \rho(m)e^{-j2\pi mk/2n} \quad k = 0, 1, \ldots, 2n - 1$$

for the $2n$-point discrete Fourier transform of $\rho(m)$, we can write

$$T(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} h(X_{ij}) + \frac{1}{2n} \sum_{k=0}^{2n-1} \Phi(k)|G(k)|^2$$

where

$$G(k) = \sum_{m=0}^{n-1} \left[ \sum_{p=1}^{L} g(X_{i(m+p)}) \right] e^{-j2\pi mk/2n}$$

$$k = 0, 1, \ldots, 2n - 1.$$ 

Thus if $n$ is a power of 2, $T(x)$ can be computed using on the order of $O(2n\log(n))$ multiplications;

this is as compared with the Gaussian-assumption detector discussed above, whose computational load is essentially linear in $n$. It should be noted that $n$, the number of data samples used in each decision, has for much of this work been assumed asymptotically large so that efficacy is a meaningful measure of performance. In the second example, however, we demonstrated that efficacy was a meaningful measure even in nonasymptotic situations. Similarly, in this application note, we address the issue of computational cost in the case of finite and feasible $n$.

V. SUMMARY

Optimal detection of stochastic and dependent signals is often infeasible due to the complicated nature of likelihood ratio. When the strength of the signal is assumed small, however, the likelihood ratio (the LO detector) is much simpler, and amounts even in the vector time-series case to the sum of an energy term and a term which compares measured-to-expected autocorrelation. While the LO detector is in general not optimal for large signals, its performance is usually good, and in any case its simplicity is appealing.

In many cases the true statistical nature, both of the additive noise and of the dependency structure of the signal to be detected, will be known only partially. We assume here that a proportion $\epsilon$ of noise samples arise from an unknown and relatively unconstrained contaminating source; this models our noise uncertainty class. We further assume that the spectrum of the signal is bounded from above and below by known functions, but is otherwise unconstrained.

The main contribution of this work is to select a detector which is robust to both uncertainties, meaning that there exists a density $f^*$ and a spectrum $\phi^*$ for which it is LO, and that its performance (in the sense of efficacy, the appropriate asymptotic signal-to-noise ratio) improves as $(f, \phi)$ deviate from $(f^*, \phi^*)$. The density $f^*$ is such that its large-argument behavior is exponential; the spectrum $\phi^*$ takes as values either the upper bound, the lower bound, or a constant. Both of these have been seen in previous robustness studies, although their appearance in this joint robustness problem necessitates additional constraints.

The robust detector is applied to an example in which 10% of noise samples are not from the Gaussian nominal density, and in which the spectral uncertainty arises from an unknown Doppler shift. A further example involving a random telegraph signal of uncertain bandwidth and a logistic nominal illustrates the relevance of the efficacy measure.
APPENDIX A

For the proof of Theorem 1, we calculate the efficacy as defined in (8) of a statistic as defined in (18). Here the actual (white) noise has symmetric density \( f(\cdot) \), and the actual signal correlation is \( r(k) \). We rewrite the statistic as

\[
T(x) = T_1(x) + T_2(x) + T_3(x)
\]

where

\[
T_1(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} e(X_{ji})
\]

\[
T_2(x) = \sum_{i=1}^{n} \sum_{m \neq i}^{n} \rho(m - i) \sum_{j=1}^{L} g(X_{ji}) \sum_{p=1}^{L} g(X_{pm})
\]

\[
T_3(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{p \neq j}^{L} g(X_{ji}) g(X_{pm})
\]

and

\[
e(x) = h(x) + g^2(x).
\]

Proceeding with the numerator first, we write

\[
\frac{\partial^2}{\partial \theta^2} E_{\theta} [T(x)]_{\theta=0} = \frac{\partial^2}{\partial \theta^2} E_{\theta} [N_1(\theta)]_{\theta=0} + \frac{\partial^2}{\partial \theta^2} E_{\theta} [N_2(\theta)]_{\theta=0}
\]

\[
+ \frac{\partial^2}{\partial \theta^2} [N_3(\theta)]_{\theta=0}
\]

where the \( N_j \) are expected values of the components of \( T \), as given below. For the first term we write

\[
N_1(\theta) = E_{\theta} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{L} e(X_{ji}) \right\}
\]

or

\[
N_1(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{L} \int e(x_{ji}) f(x_{ji} - \theta s_j) f_j(s) ds dx_{ji}.
\]

Taking the derivative we get

\[
\frac{\partial^2}{\partial \theta^2} [N_1(\theta)]_{\theta=0} = \sum_{i=1}^{n} \sum_{j=1}^{L} \int e(x_{ji}) s_j^2 f_j(s) ds dx_{ji}
\]

\[
\times f(x_{ji} - \theta s_j) f_j(s) ds dx_{ji}
\]

and evaluating at \( \theta = 0 \) we write

\[
\frac{\partial^2}{\partial \theta^2} [N_1(\theta)]_{\theta=0} = nL \int e^2 \hat{f}
\]

where for notational ease we have suppressed the dependence on the variable of integration. Proceeding

with the second term we write

\[
N_2(\theta) = E_{\theta} \left\{ \sum_{i=1}^{n} \sum_{m \neq i}^{n} \sum_{j=1}^{L} \rho(m - i) g(X_{ji}) g(X_{pm}) \right\}
\]

or

\[
N_2(\theta) = \sum_{i=1}^{n} \sum_{m \neq i}^{n} \sum_{j=1}^{L} \rho(m - i) \int \int g(x_{ji}) g(x_{pm})
\]

\[
\times f(x_{ji} - \theta s_j) f(x_{pm} - \theta s_m) f_j(s) f_m(s) ds dx_{ji} dx_{pm}
\]

Taking the derivative and evaluating at \( \theta = 0 \) we get

\[
\frac{\partial^2}{\partial \theta^2} [N_2(\theta)]_{\theta=0} = \sum_{i=1}^{n} \sum_{m \neq i}^{n} \sum_{j=1}^{L} \rho(m - i) 2 \left[ \int g \hat{f} \right]^2 r(m - i)
\]

or

\[
\frac{\partial^2}{\partial \theta^2} [N_2(\theta)]_{\theta=0} = \sum_{i=1}^{n} \sum_{m \neq i}^{n} \sum_{j=1}^{L} \rho(m - i) r(m - i) 2L^2 \left[ \int g \hat{f} \right]^2
\]

or

\[
\frac{\partial^2}{\partial \theta^2} [N_2(\theta)]_{\theta=0} = 4L^2 \sum_{k=1}^{n-1} \rho(k) r(k)(n - k) \left[ \int g \hat{f} \right]^2.
\]

For the third term we write

\[
N_3(\theta) = E_{\theta} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{p \neq j}^{L} g(X_{ji}) g(X_{pj}) \right\}
\]

or

\[
N_3(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{p \neq j}^{L} \int \int g(x_{ji}) g(x_{pj})
\]

\[
\times f(x_{ji} - \theta s_j) f(x_{pj} - \theta s_j) f_j(s) f_p(s) ds dx_{ji} dx_{pj}.
\]

Taking the derivative and evaluating at \( \theta = 0 \) we get

\[
\frac{\partial^2}{\partial \theta^2} [N_3(\theta)]_{\theta=0} = \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{p \neq j}^{L} 2 \left[ \int g \hat{f} \right]^2
\]

or

\[
\frac{\partial^2}{\partial \theta^2} [N_3(\theta)]_{\theta=0} = 2nL(L - 1) \left[ \int g \hat{f} \right]^2.
\]

Taking all these terms together we can write

\[
\frac{\partial^2}{\partial \theta^2} E_{\theta} [T(x)]_{\theta=0}
\]

\[
= nL \left[ \int e\hat{f} \right]^2 + 2nL \left[ (L - 1) + 2L \sum_{i=1}^{n-1} r(k)(1 - k/n) \right] \left[ \int g \hat{f} \right]^2
\]

for the numerator.
Now, examining the denominator, we note that to calculate the variance we need an expression for the test statistic’s mean value under the noise-alone hypothesis. Recalling that by assumption $f$ is even and $g$ is odd, we can easily obtain

$$E_0(T(x)) = nL \int ef.$$

We split the second moment into six expressions as

$$E_0(T^2(x)) = D_{11}(0) + D_{12}(0) + D_{13}(0) + 2D_{12}(0) + 2D_{13}(0) + 2D_{23}(0).$$

The first three of these terms can be written as

$$D_{11}(0) = E_0 \left[ \sum_{i=1}^{n} \sum_{j=1}^{L} c_i X_j \right]^2$$

$$= nL \int e^2 f + (n^2L^2 - nL) \left[ \int ef \right]^2$$

$$D_{12}(0) = E_0 \left[ \sum_{i=1}^{n} \sum_{m=1}^{n} r(m-i) \sum_{j=1}^{L} g(X_j) \sum_{p=1}^{L} g(X_p) \right]^2$$

$$= 4L^2 \sum_{k=1}^{L} \tilde{c}(k)(a-k) \left[ \int g^2 f \right]^2$$

$$D_{13}(0) = E_0 \left[ \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{p=1}^{L} g(X_i) g(X_p) \right]^2$$

$$= 2nL(L-1) \left[ \int g^2 f \right]^2.$$

The assumption $\int g f = 0$ implies that all three cross-terms are zero, hence we write

$$D_{12}(0) = D_{13}(0) = D_{23}(0).$$

Combining all terms we write

$$V_0(T(x)) = nL \left( \int e^2 f - \left[ \int ef \right]^2 \right)$$

$$+ 2nL \left[ (L-1) + 2L \sum_{k=1}^{n-1} \tilde{c}(k)(1-k/n) \right] \left[ \int g^2 f \right]^2$$

for the denominator.

Combining the numerator and denominator and taking the limit as $n \to \infty$, we write

$$\xi = \frac{L}{4} \left( \frac{\left( \int e^2 f + \left[ \int ef \right]^2 \right)}{\left( \int e^2 f - \left[ \int ef \right]^2 \right)^2} \right)^2$$

for the efficacy, and replacing $e$ by $h + g^2$ we get the result.

**APPENDIX B**

**Proof** We prove that the nonlinearities $g(x)$ and $h(x)$ based on $f^*$ are minimax optimal. With $f^*$ defined as in Theorem 3, we can write

$$g_{lo}^*(x) = \begin{cases} \frac{\tilde{c}(x)}{\alpha(x)} & |x| \leq a \\ \frac{\tilde{c}(x)}{\alpha(x)} & x > a \end{cases}$$

Further we define

$$e_{lo}^*(x) = \frac{\hat{h}_{lo}^*(x) + [g_{lo}^*(x)]^2}{\hat{f}_{lo}^*} = \begin{cases} \frac{\tilde{c}(x)}{\alpha(x)} & |x| \leq a \\ \frac{\tilde{c}(x)}{\alpha(x)} & |x| > a \end{cases}$$

Denote

$$K_N = 2 \left[ -1 + \frac{L}{2\pi} \int_{-\pi}^{\pi} f_1(\omega) f_2(\omega) d\omega \right]$$

$$K_D = 2 \left[ -1 + \frac{L}{2\pi} \int_{-\pi}^{\pi} f_3(\omega) f_4(\omega) d\omega \right]$$

where we have replaced $f_r$ with $f^*$. Then the efficacy becomes

$$\xi_{f_r}(T_{f^*};\phi) = \frac{L}{4} \left( \frac{\left( \int e_{lo}^* f + K_N \left[ \int g_{lo}^* f \right]^2 \right)}{\left( \int e_{lo}^* f - \left( \int e_{lo}^* f \right)^2 + K_D \left[ \int (g_{lo}^*)^2 f \right]^2 \right)} \right)^2$$

$$= \frac{L}{4} \left( \frac{\left( \int e_{lo}^* f + K_N \left[ \int g_{lo}^* f \right]^2 \right)}{\left( \int e_{lo}^* f - \left( \int e_{lo}^* f \right)^2 + K_D \left[ \int (g_{lo}^*)^2 f \right]^2 \right)} \right)^2,$$

where

$$N_{f_r}(T_{f^*};\phi) = \int e_{lo}^* f + K_N \left[ \int g_{lo}^* f \right]^2$$

$$D_{f_r}(T_{f^*};\phi) = \left( \int e_{lo}^* f - \left( \int e_{lo}^* f \right)^2 + K_D \left[ \int (g_{lo}^*)^2 f \right]^2 \right).$$

Before proceeding, we first prove the positivity of $K_N$ and $K_D$. Since $f_r(\omega)$ is a member of the class $\Phi(f_r, f_r)$ (equation (25)) and that $f_r(\omega)$ minimizes $\int_{-\pi}^{\pi} \phi_r^2(\omega) d\omega$, it is shown in Theorem 2, that

$$\int_{-\pi}^{\pi} \phi_r(\omega) f_r^2(\omega) d\omega \geq \int_{-\pi}^{\pi} \phi_r(\omega)^2 d\omega.$$ But

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_r(\omega)^2 d\omega \geq \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} \phi_r(\omega) d\omega \right)^2 = 2\pi.$$ It follows directly that $K_N \geq K_D > 0$. 28
We first consider the denominator of (45). Constraint (35) indicates that \( e_{0}^{1}(x) \leq e_{0}^{1}(a) \) for any \( |x| \leq a \), and is by construction constant for all \( |x| \geq a \). Thus since the contaminant component of \( f^{*} \) places all its mass on values of \( x \) such that \( |x| > a \), the first term \( \int (e_{0}^{1})^2 f \) is clearly maximized by \( f^{*} \). Further, \( f^{*} \) maximizes the second term, \(-\int e_{0}^{1} f^{*} \), as it is easy to show \( \int e_{0}^{1} f^{*} = 0 \). The sum of the first two terms specifies the variance of \( e_{0}^{1} \) hence is always nonnegative. The third term, given that \( K_{0} > 0 \) and that strong unimodality forces \( g_{0}^{*} \) to achieve its maximum at \( |x| > a \), is maximized again for \( f^{*} \).

We next consider the numerator of (45). Due to the symmetry of both \( f \) and \( e_{0} \), it is easy to see that \( \int e_{0} f = \int \bar{e}_{0} f \). As \( e_{0} \) is convex for \( |x| < a \) and constant otherwise, \( \int \bar{e}_{0} f \) is minimized by \( f^{*} \). For the same reason we can write \( \int g_{0} f = \int g_{0}^{*} f \). This again is minimized by \( f^{*} \) as \( g_{0}^{*} \) is increasing for \( |x| < a \) and is constant otherwise.

Consequently \( f^{*} \) maximizes the denominator while minimizing the numerator of the efficacy defined in (45), meaning that under any other density \( f \in \mathcal{F} \), performance will improve. Theorem 3 is thus proven.

ACKNOWLEDGMENT

The authors wish to express their appreciation to associate editor Gary Krumpholz for his helpful comments and suggestions.

REFERENCES


Biao Chen received his B.E. in 1992, M.A. in 1994, both in electrical engineering, from Tsinghua University.

He is a graduate student at the University of Connecticut, Storrs, working towards his Ph.D. His research interest focuses on detection and estimation theory.

Peter Willett (S’83—M’86) received the B.A.Sc. in 1982 from the University of Toronto, Toronto, Canada, and the Ph.D. from Princeton University, Princeton, NJ, in 1986.

He is an Associate Professor at the University of Connecticut, Storrs, where he has worked since 1986. His interests are generally in the areas of detection theory and signal processing, and, lately, particularly in the area of data fusion.