Correspondence

Noise Enhanced M-ary Composite Hypothesis-Testing in the Presence of Partial Prior Information

Suat Bayram and Sinan Gezici

Abstract—In this correspondence, noise enhanced detection is studied for M-ary composite hypothesis-testing problems in the presence of partial prior information. Optimal additive noise is obtained according to two criteria, which assume a uniform distribution (Criterion 1) or the least-favorable distribution (Criterion 2) for the unknown priors. The statistical characterization of the optimal noise is obtained for each criterion. Specifically, it is shown that the optimal noise can be represented by a constant signal level or by a randomization of a finite number of signal levels according to Criterion 1 and Criterion 2, respectively. In addition, the cases of unknown parameter distributions under some composite hypotheses are considered, and upper bounds on the risks are obtained. Finally, a detection example is provided in order to investigate the theoretical results.

Index Terms—Bayes risk, composite hypothesis-testing, detection, noise enhanced detection.

I. INTRODUCTION

Although noise commonly degrades performance of a system, outputs of some nonlinear systems can be enhanced by injecting additive noise to their inputs, or by increasing the average power of the noise [1]–[10]. These situations can be considered in the framework of stochastic resonance (SR), which can be regarded as the observation of noise benefits related to signal transmission in nonlinear systems [10]–[13]. Benefits that can be obtained via SR can be in various forms, such as an increase in output signal-to-noise ratio (SNR) [1], [3], [4] or mutual information [5]–[8].

In detection problems, performance of some suboptimal detectors can be enhanced by adding independent noise to their observations [9], [10], [14]–[20]. Such *noise enhanced detection* phenomena have been investigated according to the Bayesian [16]–[18], minimax [19], [20] and Neyman–Pearson [9], [10], [14] criteria. In [16], it is shown that the optimal noise that minimizes the probability of decision error has a constant value, and a Gaussian mixture example is used to illustrate the improvability of a detector. In [17], noise benefits are investigated for threshold neural signal detection in terms of reducing the probability of detection error, and various necessary and sufficient conditions are presented to determine noise enhanced detection for a wide range of signals and symmetric scale-family noise when the detection threshold is suboptimal. In addition, an example is studied in [14] to illustrate that detection performance of a suboptimal detector can be improved by adding white Gaussian noise for the problem of detecting a constant

signal in Gaussian mixture noise. In [9] and [10], the effects of additive noise on detection performance are studied in the Neyman–Pearson framework, and it is shown that the optimal additive noise can be represented by a randomization of at most two different signal values. On the other hand, the studies in [19] and [20] consider the minimax criterion and investigate the effects of additive noise on suboptimal detectors. Finally, [18] considers a nonlinear signal-noise mixture, where non-Gaussian noise acts on the phase of a periodic signal, and illustrates that the performance of an optimal detector can be improved (locally) by increasing the noise level for optimal detection strategies according to the Bayesian, Neyman–Pearson, and minimax criteria.

The Bayesian and minimax frameworks can be considered as two extreme cases of prior information. In the former, perfect (exact) prior information is available whereas no prior information exists in the latter. In practice, having perfect prior information is a very exceptional case [21]. In most cases, prior information is incomplete and only partial prior information is available [21], [22]. Since the Bayesian approach is ineffective in the absence of exact prior information, and since the minimax approach, which ignores the partial prior information, can result in poor performance due to its conservative approach, there have been various studies that take partial prior information into account [21]–[28]. The restricted Bayes, Γ-minimax, empirical Bayes, robust Bayes and mean-max criteria are the main approaches considering partial prior information [21]–[25].

In this correspondence, noise enhanced detection is studied in the presence of partial prior information. Optimal additive noise is formulated according to two different criteria. In the first one, a uniform distribution is assumed for the unknown priors, whereas in the second one the worst-case distributions are considered for the unknown priors by taking a conservative approach, which can be regarded as a Γ-minimax approach. In both cases, the statistics of the optimal additive noise are characterized. Specifically, it is shown that the optimal additive noise can be represented by a constant signal level according to the first criterion, whereas it can be represented by a discrete random variable with a finite number of mass points according to the second criterion (see Proposition 2 for the exact number of mass points). Two other contributions of the study are to investigate noise enhanced detection with partial prior information in the most generic hypotheses formulation; that is, M-ary composite hypotheses, and to employ a very generic cost function in the definition of the conditional risks (see (7)). Therefore, it covers some of the previous studies on noise enhanced detection as special cases. For example, if simple[1] binary hypotheses, uniform cost assignment (UCA), and perfect prior information are assumed, the results reduces to those in [16]. As another example, if simple M-ary hypotheses and no prior information are assumed, the results reduces to those in [20]. Furthermore, for composite hypothesis-testing problems, the cases of unknown parameter distributions under some hypotheses are also considered, and upper bounds on the risks are obtained. Finally, a detection example is presented to investigate the theoretical results.

II. PROBLEM FORMULATION

Consider the following M-ary composite hypothesis-testing problem:

$$\mathcal{H}_i : p_\theta^X(\mathbf{x}), \ \theta \in \Lambda_i, \quad i = 0, 1, \dots, M-1 \quad (1)$$

[1]A *simple* hypothesis means that there is only one possible probability distribution under the hypothesis, whereas a *composite* hypothesis corresponds to multiple possible probability distributions.
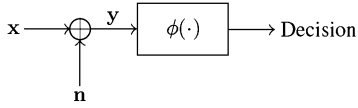
Fig. 1. Independent noise $\mathbf{n}$ is added to observation $\mathbf{x}$ in order to improve the performance of the detector, represented by $\phi(\cdot)$.

where $\mathcal{H}_i$ denotes the ith hypothesis and $p_\theta^X(\mathbf{x})$ represents the probability density function (PDF) of observation $\mathbf{X}$ for a given value of $\Theta = \theta$. Each observation (measurement) $\mathbf{x}$ is a vector with K components; i.e., $\mathbf{x} \in \mathbb{R}^K$, and $\Lambda_0, \Lambda_1, \ldots, \Lambda_{M-1}$ form a partition of the parameter space Λ. The distribution of the unknown parameter Θ for hypothesis i is represented by $w_i(\theta)$ for $i = 0, 1, \ldots, M - 1$. In addition, the prior probability of hypothesis $\mathcal{H}_i$ is denoted by π_i for $i = 0, 1, \ldots, M - 1$. Composite hypothesis-testing problems as in (1) are encountered in various problems, such as in non-coherent communications receivers, pattern recognition, and time series analysis [29], [30]. Note that when Λ_i's consist of single elements, the problem reduces to a *simple* hypothesis-testing problem.

A generic decision rule (detector) can be defined as

$$\phi(\mathbf{x}) = i, \ \text{if } \mathbf{x} \in \Gamma_i \tag{2}$$

for $i = 0, 1, \ldots, M - 1$, where $\Gamma_0, \Gamma_1, \ldots, \Gamma_{M-1}$ form a partition of the observation space Γ. As shown in Fig. 1, the aim is to add noise to the original observation $\mathbf{x}$ (which commonly consists of a signal component and measurement noise) in order to improve the performance of the detector according to certain criteria [31]. By adding noise $\mathbf{n}$ to the original observation $\mathbf{x}$, the modified observation is formed as $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ has a PDF denoted by $p_\mathbf{N}(\cdot)$, and is independent of $\mathbf{x}$. It should be noted that the additive noise can cause both positive and negative shifts in the observations [16], [20]. As in [9] and [16], it is assumed that the detector ϕ, described by (2), is fixed, and the only means for improving the performance of the detector is to optimize the additive noise $\mathbf{n}$ (please see [20] for motivations).

When all the prior probabilities $\pi_0, \pi_1, \ldots, \pi_{M-1}$ of the hypotheses in (1) are known, the Bayesian approach can be taken, and the optimal additive noise that minimizes the Bayes risk can be sought for. This problem is studied in [16] for simple hypothesis-testing problems under UCA. On the other hand, when none of the prior probabilities are known, the minimax approach can be taken to obtain the optimal additive noise that minimizes the maximum conditional risk, which is investigated in [20] for simple hypothesis-testing problems. In this study, we focus on a more generic scenario by considering both *composite hypotheses* and *partial prior information*, meaning that the prior probabilities of some hypotheses and the probability distributions of the unknown parameters under some hypotheses may be unknown. Such a generalization can be important in practice since composite hypothesis-testing problems are encountered in many applications, and the prior information may not be available for all hypotheses (see Section VI for an example).

In order to introduce a generic problem formulation, define sets $\mathcal{S}_1, \ldots, \mathcal{S}_G$ that form a partition of set $\{0, 1, \ldots, M - 1\}$. Suppose that the prior probability π_i of $\mathcal{H}_i$ is known if $i \in \mathcal{S}_1$ and it is unknown otherwise, and assume that the size of set $\mathcal{S}_1$ is $M - N_\mathrm{u}$. In other words, $\mathcal{S}_1$ corresponds to $M - N_\mathrm{u}$ hypotheses with known prior probabilities. In addition, assume that the hypotheses with unknown prior probabilities are grouped into sets $\mathcal{S}_2, \ldots, \mathcal{S}_G$ in such a way that the sum of the prior probabilities of the hypotheses in set $\mathcal{S}_j$ is known for $j = 2, \ldots, G$. If no such information is available, then $G = 2$ can be employed; that is, all the hypotheses with unknown probabilities can be grouped together into $\mathcal{S}_2$.

In order to define the *optimal* additive noise, we consider the following two criteria:

Criterion 1: For all the hypotheses with unknown prior probabilities, assume uniform distribution of the prior probability in each group $\mathcal{S}_j$ for $j = 2, \ldots, G$, and the define the corresponding Bayes risk as

$$r_1(\phi) = \sum_{i \in \mathcal{S}_1} \pi_i \mathrm{R}_i(\phi) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} \mathrm{R}_i(\phi) \tag{3}$$

where $\mathrm{R}_i(\phi)$ is the conditional risk of decision rule ϕ when hypothesis i is true [29], $|\mathcal{S}_j|$ denotes the number of elements in set $\mathcal{S}_j$, and $\tilde{\pi}_j \triangleq \sum_{i \in \mathcal{S}_j} \pi_i$ defines the sum of the prior probabilities of the hypotheses in $\mathcal{S}_j$ for $j = 2, \ldots, G$. According to Criterion 1, the optimal additive noise is defined as $p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \arg \min_{p_\mathbf{N}(\mathbf{n})} r_1(\phi)$, where $r_1(\phi)$ is given by (3). It should be noted that assuming uniform distribution for the unknown priors is a very popular classical approach [32].

Criterion 2: For the hypotheses with unknown prior probabilities, the least-favorable distribution of the priors is considered in each group, and the corresponding risk is defined as

$$r_2(\phi) = \sum_{i \in \mathcal{S}_1} \pi_i \mathrm{R}_i(\phi) + \sum_{j=2}^G \tilde{\pi}_j \max_{i \in \mathcal{S}_j} \mathrm{R}_i(\phi). \tag{4}$$

In other words, a conservative approach is taken in Criterion 2, and the worst-case Bayes risk is considered as the performance metric. Such an approach can be considered in the framework of Γ-minimax decision rules [21]. According to Criterion 2, the optimal additive noise is calculated from $p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \arg \min_{p_\mathbf{N}(\mathbf{n})} r_2(\phi)$.

In Sections III and IV, the optimal additive noise will be investigated when the probability distributions of the unknown parameters are known under all hypotheses (the prior probabilities can still be unknown). Then, in Section V, the results will be extended to the cases in which the probability distributions of the unknown parameters are unknown under some hypotheses.

III. OPTIMAL ADDITIVE NOISE ACCORDING TO CRITERION 1

According to Criterion 1, the optimal additive noise is calculated from

$$p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \arg \min_{p_\mathbf{N}(\mathbf{n})} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i \mathrm{R}_i(\phi) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} \mathrm{R}_i(\phi) \right\}. \tag{5}$$

Since $\mathrm{R}_i(\phi)$ is the conditional risk for hypotheses i, it can be expressed as

$$\mathrm{R}_i(\phi) = \int_\Lambda \mathrm{R}_\theta(\phi) w_i(\theta) d\theta \tag{6}$$

where $\mathrm{R}_\theta(\phi)$ denotes the conditional risk that is defined as the average cost of decision rule ϕ for a given $\theta \in \Lambda$ [29]. The conditional risk can be calculated from

$$\mathrm{R}_\theta(\phi) = \mathrm{E} \left\{ \mathrm{C} \left[\phi(\mathbf{Y}), \Theta \right] | \Theta = \theta \right\} = \int_\Gamma \mathrm{C} \left[\phi(\mathbf{y}), \theta \right] p_\theta^Y(\mathbf{y}) d\mathbf{y} \tag{7}$$

where $p_\theta^Y(\mathbf{y})$ is the PDF of the noise modified observation for a given value of $\Theta = \theta$, and $\mathrm{C}[j, \theta] \geq 0$ is the cost of deciding $\mathcal{H}_j$ when $\Theta = \theta$, for $\theta \in \Lambda$ [29].

Since the additive noise is independent of the original observation, $p_\theta^Y(\mathbf{y}) = \int_{\mathbb{R}^K} p_\theta^X(\mathbf{y} - \mathbf{n}) p_\mathbf{N}(\mathbf{n}) d\mathbf{n}$. Then, the expression in (6) for

the conditional risk of hypotheses i can be manipulated from (7) as follows:

$$
\begin{aligned}
R_i(\phi) &= \int_\Lambda \int_\Gamma \int_{\mathbb{R}^K} C\left[\phi(\mathbf{y}), \theta\right] p_\theta^X(\mathbf{y}-\mathbf{n}) p_\mathbf{N}(\mathbf{n}) w_i(\theta) d\mathbf{n} d\mathbf{y} d\theta \\
&= \int_{\mathbb{R}^K} p_\mathbf{N}(\mathbf{n}) \left[\int_\Lambda \int_\Gamma C\left[\phi(\mathbf{y}), \theta\right] p_\theta^X(\mathbf{y}-\mathbf{n}) w_i(\theta) d\mathbf{y} d\theta \right] d\mathbf{n} \\
&\triangleq \int_{\mathbb{R}^K} p_\mathbf{N}(\mathbf{n}) f_i(\mathbf{n}) d\mathbf{n} = E\{f_i(\mathbf{N})\}
\end{aligned} \tag{8}
$$

where

$$
f_i(\mathbf{n}) \triangleq \int_\Lambda \int_\Gamma C\left[\phi(\mathbf{y}), \theta\right] p_\theta^X(\mathbf{y}-\mathbf{n}) w_i(\theta) d\mathbf{y} d\theta. \tag{9}
$$

Note that $f_i(\mathbf{n}) \geq 0\ \forall \mathbf{n}$ since the cost function is non-negative by definition; that is, $C[j, \theta] \geq 0$.

Based on (8), the optimization problem in (5) can be expressed as

$$
\begin{aligned}
p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) &= \arg\min_{p_\mathbf{N}(\mathbf{n})} E\left\{ \sum_{i \in \mathcal{S}_1} \pi_i f_i(\mathbf{N}) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} f_i(\mathbf{N}) \right\} \\
&\triangleq \arg\min_{p_\mathbf{N}(\mathbf{n})} E\{f(\mathbf{N})\}
\end{aligned} \tag{10}
$$

where $f(\mathbf{n})$ is defined as $f(\mathbf{n}) \triangleq \sum_{i \in \mathcal{S}_1} \pi_i f_i(\mathbf{n}) + \sum_{j=2}^G (\tilde{\pi}_j/|\mathcal{S}_j|) \sum_{i \in \mathcal{S}_j} f_i(\mathbf{n})$. From (10), the optimal noise PDF can be obtained by assigning all the probability to the minimizer of $f(\mathbf{n})$; i.e.,

$$
p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \delta(\mathbf{n} - \mathbf{n}_0), \quad \mathbf{n}_0 = \arg\min_\mathbf{n} f(\mathbf{n}). \tag{11}
$$

In other words, the optimal additive noise according to Criterion 1 can be expressed as a constant corresponding to the minimum value of $f(\mathbf{n})$. Of course, when $f(\mathbf{n})$ has multiple minima, then the optimal noise PDF can be represented as $p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \sum_{i=1}^{\tilde{L}} \lambda_i \delta(\mathbf{n} - \mathbf{n}_{0i})$, for any $\lambda_i \geq 0$ such that $\sum_{i=1}^{\tilde{L}} \lambda_i = 1$, where $\mathbf{n}_{01}, \ldots, \mathbf{n}_{0\tilde{L}}$ represent the values corresponding to the minimum values of $f(\mathbf{n})$.

The main implication of the result in (11) is that among all PDFs for the additive independent noise $\mathbf{N}$, the ones that assign all the probability to a single noise value can be used as the optimal additive signal components in Fig. 1. In other words, in this scenario, addition of independent noise to observations corresponds to shifting the decision region of the detector.

Based on the expressions in (10), a detector is called *improvable* according to Criterion 1 if there exists noise $\mathbf{N}$ that satisfies $E\{f(\mathbf{N})\} < f(\mathbf{0})$, where $f(\mathbf{0})$ represents the Bayes risk in (3) in the absence of additive noise. For example, if there exists a noise component $\mathbf{n}_*$ that satisfies $f(\mathbf{n}_*) < f(\mathbf{0})$, the detector can be classified as an improvable one according to Criterion 1. In the following, sufficient conditions are provided to determine the improvability of a detector without actually solving the optimization problem in (11).

Proposition 1: Assume that $f(\mathbf{x})$ in (10) is second-order continuously differentiable around $\mathbf{x} = \mathbf{0}$. Let $\mathbf{f}$ denote the gradient of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$. Then, the detector is improvable:

- if $\mathbf{f} \neq \mathbf{0}$; or
- if $f(\mathbf{x})$ is strictly concave at $\mathbf{x} = \mathbf{0}$.

Proof: Please see the Appendix.

Although Proposition 1 may not be very crucial for scalar observations (since it can be easy to find the optimal solution from (11) di-

rectly), it can be useful for vector observations by providing simple sufficient conditions to check if the detector can be improved via additive noise.

IV. OPTIMAL ADDITIVE NOISE ACCORDING TO CRITERION 2

According to Criterion 2, the optimal additive noise is calculated from

$$
p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \arg\min_{p_\mathbf{N}(\mathbf{n})} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i R_i(\phi) + \sum_{j=2}^G \tilde{\pi}_j \max_{l \in \mathcal{S}_j} R_l(\phi) \right\} \tag{12}
$$

which can also be expressed as

$$
p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \arg\min_{p_\mathbf{N}(\mathbf{n})} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i R_i(\phi) + \max_{\boldsymbol{l} \in \tilde{\mathcal{S}}} \sum_{j=2}^G \tilde{\pi}_j R_{l_j}(\phi) \right\} \tag{13}
$$

where $\boldsymbol{l} \triangleq [l_2 \cdots l_G]$, and $\tilde{\mathcal{S}} \triangleq \mathcal{S}_2 \times \cdots \times \mathcal{S}_G$ is the Cartesian product of sets $\mathcal{S}_2, \ldots, \mathcal{S}_G$.

From (8), the optimization problem in (13) can be stated as

$$
\begin{aligned}
p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) &= \arg\min_{p_\mathbf{N}(\mathbf{n})} \max_{\boldsymbol{l} \in \tilde{\mathcal{S}}} E\left\{ \sum_{i \in \mathcal{S}_1} \pi_i f_i(\mathbf{N}) + \sum_{j=2}^G \tilde{\pi}_j f_{l_j}(\mathbf{N}) \right\} \\
&\triangleq \arg\min_{p_\mathbf{N}(\mathbf{n})} \max_{\boldsymbol{l} \in \tilde{\mathcal{S}}} E\{f_{\boldsymbol{l}}(\mathbf{N})\}
\end{aligned} \tag{14}
$$

where $f_i(\cdot)$ and $f_{l_j}(\cdot)$ are as defined in (9), and $f_{\boldsymbol{l}}(\mathbf{N}) \triangleq \sum_{i \in \mathcal{S}_1} \pi_i f_i(\mathbf{N}) + \sum_{j=2}^G \tilde{\pi}_j f_{l_j}(\mathbf{N})$.

Although the optimization problem in (14) seems quite difficult to solve in general, the following proposition states that the optimization can be performed over a significantly reduced space as the optimal solution can be characterized by a discrete probability distribution under certain conditions. To that aim, assume that all possible additive noise values satisfy $\boldsymbol{a} \preceq \mathbf{n} \preceq \boldsymbol{b}$ for any finite $\boldsymbol{a}$ and $\boldsymbol{b}$; that is, $n_j \in [a_j, b_j]$ for $j = 1, \ldots, K$, which is a reasonable assumption since additive noise cannot have infinitely large amplitudes in practice. Then, the following proposition states the discrete nature of the optimal additive noise.

Proposition 2: If $f_{\boldsymbol{l}}(\cdot)$ in (14) are continuous functions, the PDF of optimal additive noise can be expressed as

$$
p_\mathbf{N}(\mathbf{n}) = \sum_{j=1}^{|\tilde{\mathcal{S}}|} \lambda_j \delta(\mathbf{n} - \mathbf{n}_j) \tag{15}
$$

where $|\tilde{\mathcal{S}}|$ denotes the number of elements in set $\tilde{\mathcal{S}}$ (equivalently, $|\tilde{\mathcal{S}}| = |\mathcal{S}_2| \cdots |\mathcal{S}_G|$), with $\sum_{j=1}^{|\tilde{\mathcal{S}}|} \lambda_j = 1$ and $\lambda_j \geq 0$ for $j = 1, 2, \ldots, |\tilde{\mathcal{S}}|$.

Proof: The proof is omitted since the result can be proven similarly to [9], [20]. The assumption $\boldsymbol{a} \preceq \mathbf{n} \preceq \boldsymbol{b}$ is used to guarantee the existence of the optimal solution [20]. □

Proposition 2 implies that optimal additive noise can be represented by a randomization of no more than $|\tilde{\mathcal{S}}|$ different signal levels. Therefore, the solution of the optimization problem in (14) can be obtained from the following:

$$
\min_{\{\mathbf{n}_j, \lambda_j\}_{j=1}^{|\tilde{\mathcal{S}}|}} \max_{\boldsymbol{l} \in \tilde{\mathcal{S}}} \sum_{j=1}^{|\tilde{\mathcal{S}}|} \lambda_j f_{\boldsymbol{l}}(\mathbf{n}_j)
$$

$$
\text{subject to } \sum_{j=1}^{|\tilde{\mathcal{S}}|} \lambda_j = 1, \ \lambda_j \geq 0, \ j = 1, \ldots, |\tilde{\mathcal{S}}|. \tag{16}
$$

Although (16) is significantly simpler than (14), it can still be a nonconvex optimization problem. Therefore, global optimization techniques, such as particle-swarm optimization (PSO) [33], genetic algorithms, and differential evolution [34] can be employed to obtain the op-

timal additive noise PDF. Alternatively, a convex relaxation approach can be taken as in [20] in order to obtain an approximate solution.

V. UNKNOWN PARAMETER DISTRIBUTIONS FOR SOME HYPOTHESES

In the previous formulations, it is assumed that the distribution of the unknown parameter for hypothesis i, denoted by $w_i(\theta)$, is known for $i = 0, 1, \ldots, M - 1$ (see (6)).[2] If this information is not available for certain hypotheses, an approach similar to that in [25] can be taken, and the conditional risks for those hypotheses can be defined as the worst-case conditional risks; that is, $\mathrm{R}_i(\phi) = \sup_{\theta \in \Lambda_i} \mathrm{R}_\theta(\phi)$, where $\mathrm{R}_\theta(\phi)$ is as in (7). In other words, for hypotheses with unknown parameter distributions, the maximum conditional risk is set by taking a conservative approach. On the other hand, for hypotheses with known parameter distributions, the average conditional risk in (6) can still be obtained. Therefore, the definition of $\mathrm{R}_i(\phi)$ can be extended as

$$\mathrm{R}_i(\phi) = \begin{cases} \int_\Lambda \mathrm{R}_\theta(\phi) w_i(\theta) d\theta, & \text{if } w_i(\theta) \text{ is known} \\ \sup_{\theta \in \Lambda_i} \mathrm{R}_\theta(\phi), & \text{if } w_i(\theta) \text{ is unknown,} \end{cases} \quad (17)$$

for $i = 0, 1, \ldots, M - 1$. Then, Criterion 1 in (3) and Criterion 2 in (4) can still be used in evaluating the performance of detectors.

Remark: Instead of considering the worst-case conditional risks as in (17), another approach is to assume a uniform distribution of parameter θ over Λ_i when $w_i(\theta)$ is unknown. In that case, all the results in Sections III and IV are still valid. Hence, we focus on the approach in (17) in this section.

When the parameter distributions for some hypotheses are unknown and the extended definition of $\mathrm{R}_i(\phi)$ in (17) is used, the discrete structures of the probability distributions of optimal additive noise (see (11) and Proposition 2) may not be guaranteed anymore. In other words, the optimal additive noise may also have continuous probability distributions in that scenario. Therefore, in order to obtain the (approximate) PDF of the optimal additive noise, the approach in [35] can be taken in order to search over possible PDFs in the form of $p_\mathbf{N}(\mathbf{n}) = \sum_l \zeta_l \psi(\mathbf{n} - \mathbf{n}_l)$, where $\zeta_l \geq 0$, $\sum_l \zeta_l = 1$, and $\psi_l(\cdot)$ is a window function that satisfies $\psi_l(\mathbf{x}) \geq 0, \forall \mathbf{x}$ and $\int \psi_l(\mathbf{x}) d\mathbf{x} = 1, \forall l$.

Since the computational complexity of searching over possible additive noise PDFs in the form of $p_\mathbf{N}(\mathbf{n}) = \sum_l \zeta_l \psi(\mathbf{n} - \mathbf{n}_l)$ can be high in some cases, it becomes important to specify theoretical upper bounds on $r_1(\phi)$ in (3) and $r_2(\phi)$ in (4) (with $\mathrm{R}_i(\phi)$ being given by (17)), which can be achieved under certain scenarios. The following lemma presents such upper bounds.

Lemma 1: When the conditional risk $\mathrm{R}_i(\phi)$ is defined as in (17), $r_1(\phi)$ in (3) and $r_2(\phi)$ in (4) are upper bounded as follows:

$$r_1(\phi) \leq \mathrm{E} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i \tilde{f}_i(\mathbf{N}) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} \tilde{f}_i(\mathbf{N}) \right\} \quad (18)$$

$$r_2(\phi) \leq \max_{l \in \tilde{\mathcal{S}}} \mathrm{E} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i \tilde{f}_i(\mathbf{N}) + \sum_{j=2}^G \tilde{\pi}_j \tilde{f}_{l_j}(\mathbf{N}) \right\} \quad (19)$$

[2]Note that this assumption is not needed for simple hypotheses since there is only one possible parameter value.

for any additive noise PDF $p_\mathbf{N}(\cdot)$, where

$$\tilde{f}_i(\mathbf{n}) \triangleq \begin{cases} f_i(\mathbf{n}), & \text{if } w_i(\theta) \text{ is known} \\ \sup_{\theta \in \Lambda_i} \int_\Gamma \mathrm{C}\,[\phi(\mathbf{y}), \theta]\, p_\theta^X(\mathbf{y} - \mathbf{n}) d\mathbf{y}, & \text{if } w_i(\theta) \text{ is unknown.} \end{cases} \quad (20)$$

Proof: The conditional risk in (7) can be expressed as $\mathrm{R}_\theta(\phi) = \int_\Gamma \int_{\mathbb{R}^K} \mathrm{C}[\phi(\mathbf{y}), \theta] p_\theta^X(\mathbf{y} - \mathbf{n}) p_\mathbf{N}(\mathbf{n}) d\mathbf{n} d\mathbf{y}$, which is equal to $\mathrm{R}_\theta(\phi) = \mathrm{E}\{\int_\Gamma \mathrm{C}[\phi(\mathbf{y}), \theta] p_\theta^X(\mathbf{y} - \mathbf{N}) d\mathbf{y}\}$. Based on this expression, $\mathrm{R}_i(\phi)$ in (17) becomes equal to (21), shown at the bottom of the page, where $f_i(\mathbf{N})$ is as in (9). When the expression in (21) is inserted into (3), and the fact that

$$\sup_{\theta \in \Lambda_i} \mathrm{E} \left\{ \int_\Gamma \mathrm{C}\,[\phi(\mathbf{y}), \theta]\, p_\theta^X(\mathbf{y} - \mathbf{N}) d\mathbf{y} \right\}$$

$$\leq \mathrm{E} \left\{ \sup_{\theta \in \Lambda_i} \int_\Gamma \mathrm{C}\,[\phi(\mathbf{y}), \theta]\, p_\theta^X(\mathbf{y} - \mathbf{N}) d\mathbf{y} \right\} \quad (22)$$

is employed, it can be shown that $r_1(\phi)$ is upper bounded as in (18) and (20). Similarly, the expression in (13) can be manipulated to obtain the upper bound specified by (19) and (20). □

Note that when all the $w_i(\theta)$'s are known, the terms on the right-hand-sides of (18) and (19) reduce to the objective functions in the minimization problems in (10) and (14), respectively. Therefore, they become equal to $r_1(\phi)$ and $r_2(\phi)$, respectively (since $p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \arg\min_{p_\mathbf{N}(\mathbf{n})} r_1(\phi)$ in (10) and $p_\mathbf{N}^{\mathrm{opt}}(\mathbf{n}) = \arg\min_{p_\mathbf{N}(\mathbf{n})} r_2(\phi)$ in (14) by definition); hence the upper bounds in Lemma 1 are achieved. Also, in the absence of additive noise (that is, $p_\mathbf{N}(\mathbf{n}) = \delta(\mathbf{n})$ and $\mathbf{Y} = \mathbf{X}$), (3), (4), (20) and (21) can be used to show that the upper bounds in (18) and (19) are achieved again. Specifically, in the absence of noise, the expectation operators are removed and $\tilde{f}_i(\mathbf{N})$ terms are replaced by $\tilde{f}_i(\mathbf{0})$ terms for the upper bounds in (18) and (19). Also, $\mathrm{R}_i(\phi)$ in (21) becomes equal to $\tilde{f}_i(\mathbf{0})$ in the absence of noise (see (20)). Therefore, the definitions of $r_1(\phi)$ in (3) and $r_2(\phi)$ in (4) can be used to show that the upper bounds are achieved in this scenario. In addition, it can be shown that *any additive noise component that improves (i.e., reduces) the upper bounds on $r_1(\phi)$ or $r_2(\phi)$ with respect to the case without additive noise also improves the detector performance over the noiseless case according to Criterion 1 or Criterion 2, respectively.* In order to verify this claim, let $r_1^X(\phi)$ and $r_2^X(\phi)$ denote, respectively, the performance metrics $r_1(\phi)$ and $r_2(\phi)$ when no additive noise is employed. As stated before, the upper bounds are achieved in the absence of additive noise (that is, $r_1^X(\phi)$ and $r_2^X(\phi)$ are equal to the corresponding upper bounds in the absence of additive noise). Next, suppose that noise with PDF $p_\mathbf{N}^{(1)}(\mathbf{n})$ or $p_\mathbf{N}^{(2)}(\mathbf{n})$ is added to the original observation $\mathbf{x}$, which results in a reduction of the corresponding upper bound; that is, the upper bounds become strictly less than $r_1^X(\phi)$ and $r_2^X(\phi)$, respectively. On the other hand, since $r_1(\phi)$ and $r_2(\phi)$ are always smaller than or equal to the specified upper bounds due to Lemma 1, they also become strictly less than $r_1^X(\phi)$ and $r_2^X(\phi)$, respectively. Hence, the detector performance is improved via additive noise specified by $p_\mathbf{N}^{(1)}(\mathbf{n})$ and $p_\mathbf{N}^{(2)}(\mathbf{n})$ according to Criterion 1 and Criterion 2, respectively, relative to the case without additive noise. Therefore, if an additive noise component

$$\mathrm{R}_i(\phi) = \begin{cases} \mathrm{E}\left\{\int_\Lambda \int_\Gamma \mathrm{C}\,[\phi(\mathbf{y}), \theta]\, p_\theta^X(\mathbf{y} - \mathbf{N}) w_i(\theta) d\mathbf{y} d\theta\right\} = \mathrm{E}\{f_i(\mathbf{N})\}, & \text{if } w_i(\theta) \text{ is known} \\ \sup_{\theta \in \Lambda_i} \mathrm{E}\left\{\int_\Gamma \mathrm{C}\,[\phi(\mathbf{y}), \theta]\, p_\theta^X(\mathbf{y} - \mathbf{N}) d\mathbf{y}\right\}, & \text{if } w_i(\theta) \text{ is unknown} \end{cases} \quad (21)$$

reduces the upper bound in (18) (in (19)) compared to the case without additive noise, it also improves the detection performance according to Criterion 1 (Criterion 2) over the noiseless case.

The additive noise components that minimize the upper bounds in (18) and (19) can be represented by discrete probability distributions as specified by (11) and Proposition 2 since the upper bounds are in the same form as the objective functions in the minimization problems in (10) and (14). Specifically, the PDF that minimizes the upper bound on $r_1(\phi)$ can be represented by a constant signal value, and the PDF that minimizes the upper bound on $r_2(\phi)$ can be represented by a randomization of no more than $|\tilde{S}|$ different signal values. It should also be noted that although these additive noise PDFs minimize the upper bounds in Lemma 1, they may not be the optimal additive noise PDFs for the original problem in general. The optimal solution needs to be calculated based on some PDF approximations as discussed before. However, the approach based on Lemma 1 can still be useful to obtain certain improvability conditions and to achieve performance improvements with low complexity solutions in some cases.

VI. A DETECTION EXAMPLE AND CONCLUSIONS

In this section, a 4-ary hypothesis-testing problem is studied in order to provide an example of the results presented in the previous sections. The hypotheses $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2$ and $\mathcal{H}_3$ are defined as

$$\mathcal{H}_0 : x = -3\sqrt{A} + v, \quad \mathcal{H}_1 : x = -\sqrt{A} + v,$$
$$\mathcal{H}_2 : x = \sqrt{A} + v, \quad \mathcal{H}_3 : x = 3\sqrt{A} + v \quad (23)$$

where $x \in \mathbb{R}^1$, $A > 0$ is a known scalar value, and v is symmetric Gaussian mixture noise with the following PDF:

$$p_V(x) = \sum_{i=1}^{M} w_i \psi_i(x - \mu_i) \quad (24)$$

where $w_i \geq 0$ for $i = 1, \ldots, M$, $\sum_{i=1}^{M} w_i = 1$, and $\psi_i(x) = 1/(\sqrt{2\pi}\sigma_i) \exp(-x^2/(2\sigma_i^2))$ for $i = 1, \ldots, M$. Due to the symmetry assumption, $\mu_i = -\mu_{M-i+1}$, $w_i = w_{M-i+1}$ and $\sigma_i = \sigma_{M-i+1}$ for $i = 1, \ldots, \lfloor M/2 \rfloor$. In addition, the detector is described by

$$\phi(y) = \begin{cases} 0, & y \leq -2\sqrt{A} \\ 1, & -2\sqrt{A} < y \leq 0 \\ 2, & 0 < y \leq 2\sqrt{A} \\ 3, & 2\sqrt{A} < y \end{cases} \quad (25)$$

where $y = x + n$, with n representing the independent additive noise term.

The hypothesis-testing problem in (23) is the form of pulse amplitude modulation (PAM); that is, the information is carried in the signal amplitude. The Gaussian mixture noise specified above can be encountered in PAM communications systems in the presence of interference or jamming [36]. In the following example, four different amplitudes corresponding to four different underlying hypotheses are transmitted using the PAM technique above over such a communication environment. It is assumed that only the prior probability of $\mathcal{H}_1$, π_1, is known. Such a scenario can be encountered in practice when previous measurements can successfully discriminate between the underlying hypotheses for $\mathcal{H}_1$ and the other hypotheses ($\mathcal{H}_0, \mathcal{H}_2$, and $\mathcal{H}_3$), whereas it is difficult to specify reliably which of the underlying hypotheses for $\mathcal{H}_0, \mathcal{H}_2$, and $\mathcal{H}_3$ is actually true. For instance, if we assume four fish species with three of them (corresponding to $\mathcal{H}_0, \mathcal{H}_2$, and $\mathcal{H}_3$) having similar characteristics, we cannot assume a known prior for each of those species (as we do not have reliable information from measurements); however, we can regard $\pi_0 + \pi_2 + \pi_3$ (equivalently, π_1) as a
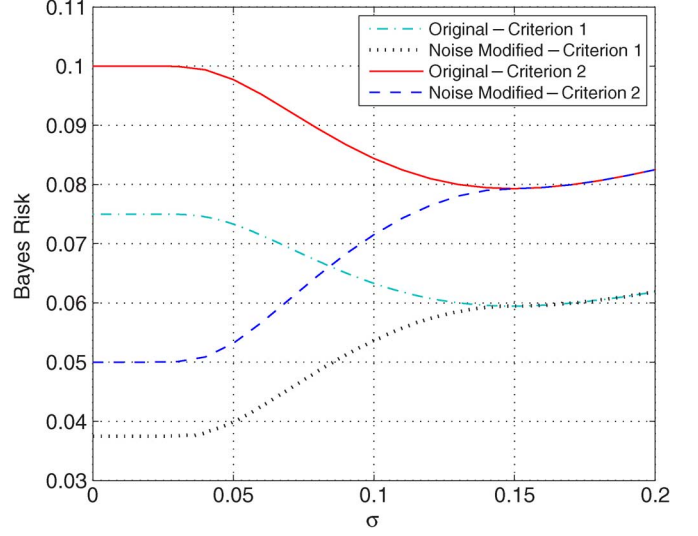

Fig. 2. Bayes risks of the original and noise modified detectors versus σ for $A = 1$ according to both criteria.

known value, since these three fish species can be distinguished easily from the other one.[3]

Since only the prior probability of $\mathcal{H}_1$ is known, there are two groups $(G = 2), \mathcal{S}_1 = \{1\}$ and $\mathcal{S}_2 = \{0, 2, 3\}$ (see (3) and (4)). Also, UCA is assumed in the following calculations. Based on the expressions in (9), (10) and (14), $f(\mathbf{n})$ and $f_l(\mathbf{n})$ can be obtained, and the optimization problems in (11) and (16) can be solved. Specifically, $f(\mathbf{n})$ in (10) can be calculated as

$$f(\mathbf{n}) = 1 - \frac{1}{3} \sum_{i=1}^{M} w_i \left[(1 - \pi_1) Q \left(\frac{-\sqrt{A} + n + \mu_i}{\sigma_i} \right) \right.$$
$$\left. + (2 + \pi_1) Q \left(\frac{-\sqrt{A} - n - \mu_i}{\sigma_i} \right) - (1 + 2\pi_1) Q \left(\frac{\sqrt{A} - n - \mu_i}{\sigma_i} \right) \right]$$

for $\mathbf{n} = n \in \mathbb{R}$, and similarly $f_l(\mathbf{n})$ in (14) becomes

$$f_l(\mathbf{n}) = 1 - \sum_{i=1}^{M} w_i \left[\pi_1 Q \left(\frac{-\sqrt{A} - n - \mu_i}{\sigma_i} \right) - \pi_1 Q \left(\frac{\sqrt{A} - n - \mu_i}{\sigma_i} \right) \right.$$
$$+ (1 - \pi_1) Q \left(\frac{-\sqrt{A} - c_{l_2} n - \mu_i}{\sigma_i} \right)$$
$$\left. - m_{l_2}(1 - \pi_1) Q \left(\frac{\sqrt{A} - n - \mu_i}{\sigma_i} \right) \right]$$

for $l = l_2 \in \mathcal{S}_2$, where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$ denotes the Q-function, $c_2 = c_3 = 1$, $c_0 = -1$, $m_0 = m_3 = 0$, and $m_2 = 1$. For the simulation results, symmetric Gaussian mixture noise with $M = 6$ is considered, where the mean values of the Gaussian components in the mixture noise in (24) are specified as $\pm[0.01 \ 0.7 \ 1.1]$ with corresponding weights of $[0.35 \ 0.1 \ 0.05]$. In addition, the variances of the Gaussian components in the mixture noise are assumed to be the same; i.e., $\sigma_i = \sigma$ for $i = 1, \ldots, M$.

Fig. 2 illustrates the Bayes risks for the noise modified and original detectors for various values of σ when $A = 1$ and $\pi_1 = 0.25$. From the figure, it is observed that the use of additive noise can significantly improve the performance according to both criteria. Also, as σ increases

---

[3]Consider a scenario in which a device measures some parameters of the fish (such as their length or color), and this information is transmitted to a data processing center using PAM.

TABLE I
OPTIMAL ADDITIVE NOISE PDF, $p_N(n) = \lambda_1 \delta(n - n_1) + \lambda_2 \delta(n - n_2) + \lambda_3 \delta(n - n_3)$, ACCORDING TO CRITERION 2

	λ_1	λ_2	λ_3	n_1	n_2	n_3
$\sigma = 0$	0.2521	0.2264	0.5215	0.3011	-0.1898	-0.1495
$\sigma = 0.05$	0.1195	0.2715	0.6090	-0.3207	-0.1913	0.1913
$\sigma = 0.1$	0.1549	0.8451	0	0.5208	-0.1634	–

the improvement ratio decreases, and after some value of σ there is no improvement. In addition, as expected, Criterion 1, which considers uniform distribution for the unknown priors, has smaller risks than Criterion 2, which considers the worst case scenario. However, it should be noted that when the priors are actually different from uniform, the additive noise obtained according to Criterion 1 can be quite suboptimal in terms of minimizing the true Bayes risk, $\sum_{i=0}^{3} \pi_i R_i(\phi)$. On the other hand, Criterion 2 considers the worst-case scenario and obtains the additive noise that minimizes the Bayes risk for the least-favorable distribution of the priors.

In order to investigate the result in Proposition 2, Table I shows the optimal noise PDFs for various values of σ according to Criterion 2. In accordance with the proposition, the optimal noise PDFs are expressed as randomization of three or fewer mass points.

APPENDIX
PROOF OF PROPOSITION 1

A sufficient condition for improvability is the existence of $\mathbf{n}_*$ such that $f(\mathbf{n}_*) < f(\mathbf{0})$. Consider an infinitesimally small noise component, $\mathbf{n}_* = \boldsymbol{\epsilon}_*$. Then, $f(\boldsymbol{\epsilon}_*)$ can be approximated by using the Taylor series expansion as $f(\mathbf{0}) + \boldsymbol{\epsilon}_*^T \mathbf{f} + 0.5\boldsymbol{\epsilon}_*^T \mathbf{H}\boldsymbol{\epsilon}_*$, where $\mathbf{H}$ and $\mathbf{f}$ are the Hessian and the gradient of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$. Therefore, $f(\mathbf{n}_*) < f(\mathbf{0})$ requires

$$\boldsymbol{\epsilon}_*^T \mathbf{H}\boldsymbol{\epsilon}_* + 2\boldsymbol{\epsilon}_*^T \mathbf{f} < 0. \tag{26}$$

Let $\boldsymbol{\epsilon}_* = \rho_* \mathbf{z}$, where ρ_* is an infinitesimally small real number, and $\mathbf{z}$ is a K-dimensional real vector. Then, (26) can be simplified, after some manipulation, as

$$\mathbf{z}^T \mathbf{H}\mathbf{z} + \frac{2}{\rho_*}\mathbf{z}^T \mathbf{f} < 0. \tag{27}$$

For the first part of the proposition, if $\mathbf{f} \neq \mathbf{0}$, then ρ_* and $\mathbf{z}$ satisfying (27) can always be found. For the second part of the proposition, if $f(\mathbf{x})$ is strictly concave at $\mathbf{x} = \mathbf{0}$, which means that $\mathbf{H}$ is negative definite, then ρ_* and $\mathbf{z}$ satisfying (27) always exist. □

REFERENCES

[1] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys. A, Math. Gen.*, vol. 14, pp. 453–457, 1981.
[2] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni, "Stochastic resonance," *Rev. Mod. Phys.*, vol. 70, no. 1, pp. 223–287, Jan. 1998.
[3] G. P. Harmer, B. R. Davis, and D. Abbott, "A review of stochastic resonance: Circuits and measurement," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 2, pp. 299–309, Apr. 2002.
[4] K. Loerincz, Z. Gingl, and L. Kiss, "A stochastic resonator is able to greatly improve signal-to-noise ratio," *Phys. Lett. A*, vol. 224, pp. 63–67, 1996.
[5] I. Goychuk and P. Hanggi, "Stochastic resonance in ion channels characterized by information theory," *Phys. Rev. E*, vol. 61, no. 4, pp. 4272–4280, 2000.
[6] S. Mitaim and B. Kosko, "Adaptive stochastic resonance in noisy neurons based on mutual information," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1526–1540, Nov. 2004.
[7] N. G. Stocks, "Suprathreshold stochastic resonance in multilevel threshold systems," *Phys. Rev. Lett.*, vol. 84, no. 11, pp. 2310–2313, Mar. 2000.
[8] X. Godivier and F. Chapeau-Blondeau, "Stochastic resonance in the information capacity of a nonlinear dynamic system," *Int. J. Bifurc. Chaos*, vol. 8, no. 3, pp. 581–589, 1998.
[9] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part I—Fixed detectors," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3172–3184, Jul. 2007.
[10] A. Patel and B. Kosko, "Optimal noise benefits in Neyman–Pearson and inequality-constrained signal detection," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1655–1669, May 2009.
[11] P. Hanggi, M. E. Inchiosa, D. Fogliatti, and A. R. Bulsara, "Nonlinear stochastic resonance: The saga of anomalous output-input gain," *Phys. Rev. E*, vol. 62, no. 5, pp. 6155–6163, Nov. 2000.
[12] S. Zozor and P.-O. Amblard, "On the use of stochastic resonance in sine detection," *Signal Process.*, vol. 7, pp. 353–367, Mar. 2002.
[13] V. Galdi, V. Pierro, and I. M. Pinto, "Evaluation of stochastic-resonance-based detectors of weak harmonic signals in additive white Gaussian noise," *Phys. Rev. E*, vol. 57, no. 6, pp. 6470–6479, Jun. 1998.
[14] S. M. Kay, "Can detectability be improved by adding noise?," *IEEE Signal Process. Lett.*, vol. 7, no. 1, pp. 8–10, Jan. 2000.
[15] S. Bayram and S. Gezici, "On the improvability and nonimprovability of detection via additional independent noise," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 1001–1004, Nov. 2009.
[16] S. M. Kay, J. H. Michels, H. Chen, and P. K. Varshney, "Reducing probability of decision error using stochastic resonance," *IEEE Signal Process. Lett.*, vol. 13, no. 11, pp. 695–698, Nov. 2006.
[17] A. Patel and B. Kosko, "Error-probability noise benefits in threshold neural signal detection," *Neural Netw.*, vol. 22, pp. 697–706, 2009.
[18] D. Rousseau and F. Chapeau-Blondeau, "Stochastic resonance and improvement by noise in optimal detection strategies," *Digital Signal Process.*, vol. 15, no. 1, pp. 19–32, Jan. 2005.
[19] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part II—Variable detectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5031–5041, Oct. 2008.
[20] S. Bayram and S. Gezici, "Noise-enhanced M-ary hypothesis-testing in the minimax framework," presented at the Int. Conf. Signal Process. Commun. Syst., Omaha, NE, Sep. 2009.
[21] J. R. Blum and J. Rosenblatt, "On partial a priori information in statistical inference," *Ann. Math. Stat.*, vol. 38, no. 6, pp. 1671–1678, 1967.
[22] J. L. Hodges, Jr. and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Stat.*, vol. 23, no. 3, pp. 396–407, Sep. 1952.
[23] H. Robbins, "The empirical Bayes approach to statistical decision problems," *Ann. Math. Stat.*, vol. 35, no. 1, pp. 1–20, 1964.
[24] J. Berger *et al.*, "An overview of robust Bayesian analysis," *Test*, vol. 3, no. 1, pp. 5–124, 1994.
[25] H. Kudo, "On partial prior information and property of parametric sufficiency," in *Proc. 5th Berkeley Symp. Math. Stat. Prob.*, 1967, vol. 1, pp. 251–265.
[26] L. J. Savage, *The Foundations of Statistics*, 2nd ed. New York: Dover, 1972.
[27] S. R. Watson, "On Bayesian inference with incompletely specified prior distributions," *Biometrika*, vol. 61, pp. 193–196, 1974.
[28] S. Bayram, S. Gezici, and H. V. Poor, "Noise enhanced hypothesis-testing in the restricted Bayesian framework," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3972–3989, Aug. 2010.
[29] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.
[30] D. Luengo, C. Pantaleon, I. Santamara, L. Velva, and J. Ibanez, "Multiple composite hypothesis testing: A competitive approach," *J. VLSI Signal Process.*, vol. 37, pp. 319–331, 2004.
[31] S. M. Kay, "Noise enhanced detection as a special case of randomization," *IEEE Signal Process. Lett.*, vol. 15, pp. 709–712, 2008.
[32] J. Skilling, "Prior probabilities," *Synthese*, vol. 63, no. 1, pp. 1–34, Apr. 1985.
[33] K. E. Parsopoulos and M. N. Vrahatis, "Particle swarm optimization method for constrained optimization problems," in *Intelligent Technologies-Theory and Applications*. Amsterdam, The Netherlands: IOS Press, 2002, pp. 214–220.
[34] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. New York: Springer, 2005.
[35] S. Bayram and S. Gezici, "Effects of additional independent noise in binary composite hypothesis-testing problems," presented at the Int. Conf. Signal Process. Commun. Syst., Omaha, NE, Sep. 2009.
[36] V. Bhatia and B. Mulgrew, "Non-parametric likelihood based channel estimator for Gaussian mixture noise," *Signal Process.*, vol. 87, pp. 2569–2586, Nov. 2007.