Noise Enhanced Hypothesis-Testing in the Restricted Bayesian Framework

Suat Bayram, *Student Member, IEEE*, Sinan Gezici, *Member, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

***Abstract*— Performance of some suboptimal detectors can be enhanced by adding independent noise to their observations. In this paper, the effects of additive noise are investigated according to the restricted Bayes criterion, which provides a generalization of the Bayes and minimax criteria. Based on a generic M-ary composite hypothesis-testing formulation, the optimal probability distribution of additive noise is investigated. Also, sufficient conditions under which the performance of a detector can or cannot be improved via additive noise are derived. In addition, simple hypothesis-testing problems are studied in more detail, and additional improvability conditions that are specific to simple hypotheses are obtained. Furthermore, the optimal probability distribution of the additive noise is shown to include at most M mass points in a simple M-ary hypothesis-testing problem under certain conditions. Then, global optimization, analytical and convex relaxation approaches are considered to obtain the optimal noise distribution. Finally, detection examples are presented to investigate the theoretical results.**

***Index Terms*— Noise enhanced detection, restricted Bayes, stochastic resonance, M-ary hypothesis-testing, composite hypotheses.**

I. Introduction

Although noise commonly degrades performance of a system, outputs of some nonlinear systems can be improved by adding noise to their inputs or by increasing the noise level in the system via a mechanism called stochastic resonance (SR) [1]-[14]. SR is said to be observed when increases in noise levels cause an increase in a metric of the quality of signal transmission or detection performance. This counterintuitive effect is mainly due to system nonlinearities and/or some parameters being suboptimal [14]. Improvements that can be obtained via SR can be in various forms, such as an increase in output signal-to-noise ratio (SNR) [1], [4], [5] or mutual information [6]-[11], [15], [16]. The first study of SR was performed in [1] to investigate the periodic recurrence of ice gases. In that work, the presence of noise was taken into account in order to explain a natural phenomenon. Since then, SR has been investigated for numerous nonlinear systems, such as optical, electronic, magnetic, and neuronal systems [3]. Also, it has been extensively studied for biological systems [17], [18].

From a signal processing perspective, SR can be viewed as noise benefits in a signal processing system, or, alternatively, noise enhanced signal processing [13], [14]. Specifically, in detection theory, SR can be considered for performance improvements of some suboptimal detectors by adding independent noise to their observations, or by increasing the noise level in the observations. One of the first studies of SR for signal detection is reported in [19], which deals with signal extraction from background noise. After that study, some works in the physics literature also investigate SR for detection purposes [15], [16], [20]-[22]. In the signal processing community, SR is regarded as a mechanism that can be used to improve the performance of a suboptimal detector according to the Bayes, minimax, or Neyman-Pearson criteria [12], [13], [23]-[37]. In fact, noise enhancements can also be observed in optimal detectors, as studied in [13] and [37]. Various scenarios are investigated in [37] for optimal Bayes, minimax and Neyman-Pearson detectors, which shows that performance of optimal detectors can be improved (locally) by raising the noise level in some cases. In addition, randomization between two anti-podal signal pairs and the corresponding maximum *a posteriori* probability (MAP) decision rules is studied in [13], and it is shown that power randomization can result in significant performance improvement.

In the Neyman-Pearson framework, the aim is to increase the probability of detection under a constraint on the probability of false alarm [12], [13], [24], [26]. In [24], an example is presented to illustrate the effects of additive noise on the detection performance for the problem of detecting a constant signal in Gaussian mixture noise. In [12], a theoretical framework for investigating the effects of additive noise on suboptimal detectors is established according to the Neyman-Pearson criterion. Sufficient conditions under which performance of a detector can or cannot be improved via additive noise are derived, and it is proven that optimal additive noise can be generated by a randomization of at most two discrete signals, which is an important result since it greatly simplifies the calculation of the optimal noise probability density function (p.d.f.). An optimization theoretic framework is provided in [13] for the same problem, which also proves the two mass point structure of the optimal additive noise p.d.f., and, in addition, shows that an optimal noise distribution may not exist in certain scenarios.

The study in [12] is extended to variable detectors in [25], and similar observations as in the case of fixed detectors are made. Also, the theoretical framework in [12] is applied to sequential detection and parameter estimation problems in [38] and [39], respectively. In [38], a binary sequential detection problem is considered, and additive noise that reduces at least one of the expected sample sizes for the sequential detection system is obtained. In [39], improvability of estimation performance via additive noise is illustrated under certain conditions for various estimation criteria, and the form of the optimal noise p.d.f. is obtained for each criterion. The effects of noise are investigated also for detection of weak sinusoidal signals

S. Bayram and S. Gezici are with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara 06800, Turkey, Tel: +90 (312) 290-3139, Fax: +90 (312) 266-4192, e-mails: {sbayram,gezici}@ee.bilkent.edu.tr

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA, e-mail: poor@princeton.edu

This research was supported in part by the U.S. Office of Naval Research under Grant N00014-09-1-0342. Part of this work was presented at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010).

and for locally optimal detectors. In [33] and [34], detection of a weak sinusoidal signal is considered, and improvements on detection performance are investigated. In addition, [35] studies the optimization of noise and detector parameters of locally optimal detectors for the detection of a small amplitude sinusoid in non-Gaussian noise.

In [23], the effects of additive noise are investigated according to the Bayes criterion under uniform cost assignment. It is shown that the optimal noise that minimizes the probability of decision error has a constant value, and a Gaussian mixture example is presented to illustrate the improvability of a suboptimal detector via adding constant "noise". On the other hand, [25] and [29] consider the minimax criterion, which aims to minimize the maximum of the conditional risks [40], and they investigate the effects of additive noise on suboptimal detectors. It is shown in [29] that the optimal additive noise can be represented, under mild conditions, by a randomization of at most M signal levels for an M-ary hypothesis testing problem in the minimax framework.

Although both the Bayes and minimax criteria have been considered for noise enhanced hypothesis-testing [23], [25], [29], no studies have considered the *restricted Bayes* criterion [41]. In the Bayesian framework, the prior information is precisely known, whereas it is not available in the minimax framework [40]. However, having prior information with some uncertainty is the most common situation, and the restricted Bayes criterion is well-suited in that case [41], [42]. In the restricted Bayesian framework, the aim is to minimize the Bayes risk under a constraint on the individual conditional risks [41]. Depending on the value of the constraint, the restricted Bayes criterion covers the Bayes and minimax criteria as special cases [42]. In general, it is challenging to obtain the optimal decision rule under the restricted Bayes criterion [42]-[46]. In [42], a number of theorems are presented to obtain the optimal decision rule by modifying Wald's minimax theory [47]. However, the application of those theorems requires certain conditions to hold and commonly intensive computations. Therefore, [42] states that the widespread application of the optimal detectors according to the restricted Bayes criterion would require numerical methods in combination with theoretical results derived in [42].

Although it is challenging to obtain the optimal detector according to the restricted Bayes criterion, this criterion can be quite advantageous in practical applications compared to the Bayes and minimax criteria, as studied in [42]. Therefore, in this paper, the aim is to consider suboptimal detectors and to investigate how their performance can be improved via additive independent noise in the restricted Bayesian framework. In other words, one motivation is to improve performance of suboptimal detectors via additive noise and to provide reasonable performance with low computational complexity. Another motivation is the theoretical interest to investigate the effects of noise on suboptimal detectors and to obtain sufficient conditions under which performance of detectors can or cannot be improved via additive noise in the restricted Bayesian framework.

In this paper, the effects of additive independent noise on the performance of suboptimal detectors are investigated according to the restricted Bayes criterion. A generic M-ary composite hypothesis-testing problem is considered, and sufficient conditions under which a suboptimal detector can or cannot be improved are derived. In addition, various approaches to obtaining the optimal solution are presented. For simple hypothesis-testing problems, additional improvability conditions that are simple to evaluate are proposed, and it is shown that optimal additive noise can be represented by a p.d.f. with at most M mass points. Furthermore, optimization theoretic approaches to obtaining the optimal noise p.d.f. are discussed; both global optimization techniques and approximate solutions based on convex relaxation are considered. Also, an analytical approach is proposed to obtain the optimal noise p.d.f. under certain conditions. Finally, detection examples are provided to investigate the theoretical results and to illustrate the practical importance of noise enhancement.

The remainder of the paper is organized as follows. Section II studies composite hypothesis-testing problems, and provides a generic formulation of the problem. In addition, improvability and nonimprovability conditions are presented and an approximate solution of the optimal noise problem is discussed. Then, Section III considers simple hypothesis-testing problems and provides additional improvability conditions. Also, the discrete structure of the optimal noise probability distribution is specified. Then, detection examples are presented to illustrate the theoretical results in Section IV. Finally, concluding remarks are made in Section V.

II. Noise Enhanced M-ary Composite Hypothesis-Testing

A. Problem Formulation and Motivation

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

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

where $p_\theta^{\mathbf{X}}(\cdot)$ represents the p.d.f. of observation $\mathbf{X}$ for a given value of parameter, $\Theta = \theta$, and θ belongs to parameter set Λ_i under hypotheses $\mathcal{H}_i$. The 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 prior distribution of Θ is denoted by $w(\theta)$, and it is assumed that $w(\theta)$ is known with some uncertainty [41], [42]. For example, it can be a p.d.f. estimate based on previous decisions.

A generic decision rule (detector) is considered, which can be expressed as

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

for $i = 0, 1, \ldots, M-1$, where $\Gamma_0, \Gamma_1, \ldots, \Gamma_{M-1}$ form a partition of the observation space Γ.

In some cases, addition of noise to observations can improve the performance of a suboptimal detector. By adding noise $\mathbf{n}$ to the original observation $\mathbf{x}$, the noise modified observation is formed as $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ has a p.d.f. denoted by $p_{\mathbf{N}}(\cdot)$, and is independent of $\mathbf{x}$.[1] As in [12] and in Section II of [13], it is assumed that the detector in (2) is fixed, and that the only means for improving the performance of the detector is to optimize the additive noise $\mathbf{n}$. In other words, the aim is to find the best $p_{\mathbf{N}}(\cdot)$ according to the restricted

[1]As discussed in [12] and [24], additional improvements in detector performance can be obtained by adding noise that depends on the original background noise and/or that has a p.d.f. depending on which hypothesis is true. However, adding such a dependent noise is not commonly possible in practice since the related prior information is usually not available [12].

Bayes criterion [41]; namely, to minimize the Bayes risk under certain constraints on the conditional risks, as specified below.

$$\min_{p_{\mathbf{N}}(\cdot)} \int_{\Lambda} \mathrm{R}_{\theta}^{\mathbf{y}}(\phi) w(\theta)\, d\theta\ ,$$
$$\text{subject to } \max_{\theta\in\Lambda} \mathrm{R}_{\theta}^{\mathbf{y}}(\phi) \leq \alpha\ , \tag{3}$$

where α represents the upper limit on the conditional risks, $\int_{\Lambda} \mathrm{R}_{\theta}^{\mathbf{y}}(\phi) w(\theta)\, d\theta = \mathrm{E}\{\mathrm{R}_{\Theta}^{\mathbf{y}}(\phi)\} \triangleq r^{\mathbf{y}}(\phi)$ is the Bayes risk, and $\mathrm{R}_{\theta}^{\mathbf{y}}(\phi)$ denotes the conditional risk of ϕ for a given value of θ for the noise modified observation $\mathbf{y}$. More specifically, $\mathrm{R}_{\theta}^{\mathbf{y}}(\phi)$ is defined as the average cost of decision rule ϕ for a given θ,

$$\mathrm{R}_{\theta}^{\mathbf{y}}(\phi) = \mathrm{E}\left\{\mathrm{C}[\phi(\boldsymbol{Y}), \Theta] \mid \Theta = \theta\right\} = \int_{\Gamma} \mathrm{C}[\phi(\mathbf{y}), \theta]\, p_{\theta}^{\boldsymbol{Y}}(\mathbf{y})\, d\mathbf{y} \tag{4}$$

where $p_{\theta}^{\boldsymbol{Y}}(\cdot)$ is the p.d.f. of the noise modified observation for a given value of $\Theta = \theta$, and $\mathrm{C}[i, \theta]$ is the cost of selecting $\mathcal{H}_i$ when $\Theta = \theta$, for $\theta \in \Lambda$ [40].

In the restricted Bayes formulation in (3), any undesired effects due to the uncertainty in the prior distribution can be controlled via parameter α, which can be considered as an upper bound on the Bayes risk [42]. Specifically, as the amount of uncertainty in the prior information increases, a smaller (more restrictive) value of α is employed. In that way, the restricted Bayes formulation provides a generalization of the Bayesian and the minimax approaches [41]. In the Bayesian framework, the prior distribution of the parameter is perfectly known, whereas it is completely unknown in the minimax framework. On the other hand, the restricted Bayesian framework considers some amount of uncertainty in the prior distribution and converges to the Bayesian and minimax formulations as special cases depending on the value of α in (3) [42], [41]. Therefore, the study of noise enhanced hypothesis-testing in this paper covers the previous works on noise enhanced hypothesis-testing according to the Bayesian and minimax criteria as special cases [23], [25], [29].

Two main motivations for studying the effects of additive noise on the detector performance are as follows. First, optimal detectors according to the restricted Bayes criterion are difficult to obtain, or require intense computations [42]. Therefore, in some cases, a suboptimal detector with additive noise can provide acceptable performance with low computational complexity. Second, it is of theoretical interest to investigate the improvements that can be achieved via additive noise [29].

In order to provide an explicit formulation of the optimization problem in (3), which indicates the dependence of $\mathrm{R}_{\boldsymbol{\theta}}^{\mathbf{y}}(\phi)$ on the p.d.f. of the additive noise explicitly, $\mathrm{R}_{\boldsymbol{\theta}}^{\mathbf{y}}(\phi)$ in (4) is manipulated as follows:[2]

$$\mathrm{R}_{\theta}^{\mathbf{y}}(\phi) = \int_{\Gamma}\int_{\mathbb{R}^K} \mathrm{C}[\phi(\mathbf{y}), \theta]\, p_{\theta}^{\mathbf{X}}(\mathbf{y}-\mathbf{n})\, p_{\mathbf{N}}(\mathbf{n})\, d\mathbf{n}\, d\mathbf{y} \tag{5}$$
$$= \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) \left[\int_{\Gamma} \mathrm{C}[\phi(\mathbf{y}), \theta] p_{\theta}^{\mathbf{X}}(\mathbf{y}-\mathbf{n})\, d\mathbf{y}\right] d\mathbf{n} \tag{6}$$
$$= \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n})\, F_{\theta}(\mathbf{n})\, d\mathbf{n} \tag{7}$$
$$= \mathrm{E}\{F_{\theta}(\mathbf{N})\} \tag{8}$$

where

$$F_{\theta}(\mathbf{n}) \triangleq \int_{\Gamma} \mathrm{C}[\phi(\mathbf{y}), \theta]\, p_{\theta}^{\mathbf{X}}(\mathbf{y}-\mathbf{n})\, d\mathbf{y}\ . \tag{9}$$

Note that $F_{\theta}(\mathbf{n})$ defines the conditional risk given θ for a constant value of additive noise, $\mathbf{N} = \mathbf{n}$. Therefore, for $\mathbf{n} = \mathbf{0}$, $F_{\boldsymbol{\theta}}(\mathbf{0}) = \mathrm{R}_{\theta}^{\mathbf{x}}(\phi)$ is obtained; that is, $F_{\boldsymbol{\theta}}(\mathbf{0})$ is equal to the conditional risk of the decision rule given θ for the original observation $\mathbf{x}$.

From (8), the optimization problem in (3) can be formulated as follows:

$$\min_{p_{\mathbf{N}}(\cdot)} \int_{\Lambda} \mathrm{E}\{F_{\theta}(\mathbf{N})\} w(\theta)\, d\theta\ ,$$
$$\text{subject to } \max_{\theta\in\Lambda} \mathrm{E}\{F_{\theta}(\mathbf{N})\} \leq \alpha\ . \tag{10}$$

If a new function $F(\mathbf{n})$ is defined as in the following expression,

$$F(\mathbf{n}) \triangleq \int_{\Lambda} F_{\theta}(\mathbf{n}) w(\theta)\, d\theta\ , \tag{11}$$

the optimization problem in (10) can be reformulated in the following simple form:

$$\min_{p_{\mathbf{N}}(\cdot)} \mathrm{E}\{F(\mathbf{N})\}\ ,$$
$$\text{subject to } \max_{\theta\in\Lambda} \mathrm{E}\{F_{\theta}(\mathbf{N})\} \leq \alpha\ . \tag{12}$$

From (9) and (11), it is noted that $F(\mathbf{0}) = r^{\mathbf{x}}(\phi)$. Namely, $F(\mathbf{0})$ is equal to the Bayes risk for the original observation $\mathbf{x}$; that is, the Bayes risk in the absence of additive noise.

B. Improvability and Nonimprovability Conditions

In general, it is quite complex to obtain a solution of the optimization problem in (12) as it requires a search over all possible noise p.d.f.s. Therefore, it is useful to determine, without solving the optimization problem, whether additive noise can improve the performance of the original system. In the restricted Bayesian framework, a detector is called *improvable*, if there exists a noise p.d.f. such that $\mathrm{E}\{F(\mathbf{N})\} < r^{\mathbf{x}}(\phi) = F(\mathbf{0})$ and $\max_{\theta\in\Lambda} \mathrm{R}_{\theta}^{\mathbf{y}}(\phi) = \max_{\theta\in\Lambda} \mathrm{E}\{F_{\theta}(\mathbf{N})\} \leq \alpha$ (cf. (12)). Otherwise, the detector is called *nonimprovable*.

First, the following nonimprovability condition is obtained based on the properties of F_{θ} in (9) and F in (11).

Theorem 1: *Assume that there exits $\theta^* \in \Lambda$ such that $F_{\theta^*}(\mathbf{n}) \leq \alpha$ implies $F(\mathbf{n}) \geq F(\mathbf{0})$ for all $\mathbf{n} \in \mathcal{S}_{\mathrm{n}}$, where $\mathcal{S}_{\mathrm{n}}$ is a convex set[3] consisting of all possible values of additive noise $\mathbf{n}$. If $F_{\theta^*}(\mathbf{n})$ and $F(\mathbf{n})$ are convex functions over $\mathcal{S}_{\mathrm{n}}$, then the detector is nonimprovable.*

Proof: The proof employs an approach that is similar to the proof of Proposition 1 in [26]. Due to the convexity of $F_{\theta^*}(\cdot)$, the conditional risk in (8) can be bounded, via Jensen's inequality, as

$$\mathrm{R}_{\theta^*}^{\mathbf{y}}(\phi) = \mathrm{E}\{F_{\theta^*}(\mathbf{N})\} \geq F_{\theta^*}\left(\mathrm{E}\{\mathbf{N}\}\right)\ . \tag{13}$$

As $\mathrm{R}_{\theta^*}^{\mathbf{y}}(\phi) \leq \alpha$ is a necessary condition for improvability, (13) implies that $F_{\theta^*}\left(\mathrm{E}\{\mathbf{N}\}\right) \leq \alpha$ must be satisfied. Since $\mathrm{E}\{\mathbf{N}\} \in \mathcal{S}_{\mathrm{n}}$, $F_{\theta^*}\left(\mathrm{E}\{\mathbf{N}\}\right) \leq \alpha$ means $F\left(\mathrm{E}\{\mathbf{N}\}\right) \geq F(\mathbf{0})$ due to the assumption in the proposition. Hence,

$$r^{\mathbf{y}}(\phi) = \mathrm{E}\{F(\mathbf{N})\} \geq F\left(\mathrm{E}\{\mathbf{N}\}\right) \geq F(\mathbf{0})\ , \tag{14}$$

[2]Note that the independence of $\mathbf{X}$ and $\mathbf{N}$ are used to obtain (5) from (4).

[3]$\mathcal{S}_{\mathrm{n}}$ can be modeled as convex because convex combination of individual noise components can be obtained via randomization [48].

where the first inequality results from the convexity of F. Then, from (13) and (14), it is concluded that $\mathrm{R}^{\mathbf{y}}_{\theta^*}(\phi) \leq \alpha$ implies $r^{\mathbf{y}}(\phi) \geq F(\mathbf{0}) = r^{\mathbf{x}}(\phi)$. Therefore, the detector is nonimprovable. $\square$

The conditions in Theorem 1 can be used to determine when the detector performance cannot be improved via additive noise, which prevents unnecessary efforts for trying to solve the optimization problem in (12). However, it should also be noted that Theorem 1 provides only sufficient conditions; hence, the detector can still be nonimprovable although the conditions in the theorem are not satisfied.

In order to provide an example application of Theorem 1, consider a Gaussian location testing problem [40], in which the observation has a Gaussian p.d.f. with mean $\theta\mu$ and variance σ^2, denoted by $\mathcal{N}(\theta\mu, \sigma^2)$, where μ and σ are known values. Hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ correspond to $\theta = 0$ and $\theta = 1$, respectively (that is, $\Lambda_0 = \{0\}$ and $\Lambda_1 = \{1\}$). In addition, consider a decision rule that selects $\mathcal{H}_1$ if $y \geq 0.5\mu$ and $\mathcal{H}_0$ otherwise. Let $\mathcal{S}_{\mathrm{n}} = (-0.5\mu, 0.5\mu)$ represent the set of additive noise values for possible performance improvement. For uniform cost assignment (UCA) [40], (9) can be used to obtain $F_0(n)$ as follows:

$$F_0(n) = \int_{-\infty}^{\infty} \mathrm{C}[\phi(y), 0] p_0^X(y - n) dy \tag{15}$$

$$= \int_{-\infty}^{\infty} \phi(y) p_0^X(y - n) dy \tag{16}$$

$$= \int_{0.5\mu}^{\infty} \frac{\mathrm{e}^{-\frac{(y-n)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} dy = Q\left(\frac{0.5\mu - n}{\sigma}\right) , \tag{17}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \mathrm{e}^{-t^2/2} dt$ denotes the Q-function, and $\mathrm{C}[i, j] = 1$ for $i \neq j$ and $\mathrm{C}[i, j] = 0$ for $i = j$ are used in (15) due to the UCA. Similarly, $F_1(n)$ can be obtained as $F_1(n) = Q\left(\frac{0.5\mu+n}{\sigma}\right)$. For equal priors, $F(n)$ in (11) is obtained as $F(n) = 0.5(F_0(n) + F_1(n))$; that is,

$$F(n) = 0.5Q\left(\frac{0.5\mu - n}{\sigma}\right) + 0.5Q\left(\frac{0.5\mu + n}{\sigma}\right) . \tag{18}$$

Let α be set to $Q\left(0.5\mu/\sigma\right)$, which determines the upper bound on the conditional risks. Regarding the assumption in Theorem 1, it can be shown for $\theta^* = 0$ that $F_{\theta^*}(n) \leq \alpha$ implies $F(n) \geq F(0) = Q(0.5\mu/\sigma)$ for all $n \in \mathcal{S}_{\mathrm{n}}$. This follows from the facts that $F_0(n) \leq \alpha = Q\left(0.5\mu/\sigma\right)$ requires that $n \in (-0.5\mu, 0]$ and that $F(n)$ in (18) satisfies $F(n) \geq Q(0.5\mu/\sigma) = \alpha$ for $n \in (-0.5\mu, 0]$ due to the convexity of $Q(x/\sigma)$ for $x > 0$. In addition, it can be shown that both $F_0(n)$ and $F_1(n)$ are convex functions over $\mathcal{S}_{\mathrm{n}}$, which implies that $F(n)$ is also convex over $\mathcal{S}_{\mathrm{n}}$. Then, Theorem 1 implies that the detector is nonimprovable for this example. Therefore, there is no need to tackle the optimization problem in (12) in this case, since $p_N^{\mathrm{opt}}(n) = \delta(n)$ is concluded directly from the theorem.

Next, sufficient conditions under which the detector performance can be improved via additive noise are obtained. To that aim, it is first assumed that $F(\mathbf{x})$ and $F_\theta(\mathbf{x})$ $\forall \theta \in \Lambda$ are second-order continuously differentiable around $\mathbf{x} = \mathbf{0}$. In addition, the following functions are defined for notational convenience:

$$f_\theta^{(1)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{i=1}^{K} z_i \frac{\partial F_\theta(\mathbf{x})}{\partial x_i} , \tag{19}$$

$$f^{(1)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{i=1}^{K} z_i \frac{\partial F(\mathbf{x})}{\partial x_i} , \tag{20}$$

$$f_\theta^{(2)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^{K} \sum_{i=1}^{K} z_l z_i \frac{\partial^2 F_\theta(\mathbf{x})}{\partial x_l \partial x_i} , \tag{21}$$

$$f^{(2)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^{K} \sum_{i=1}^{K} z_l z_i \frac{\partial^2 F(\mathbf{x})}{\partial x_l \partial x_i} , \tag{22}$$

where x_i and z_i represent the ith components of $\mathbf{x}$ and $\mathbf{z}$, respectively. Then, the following theorem provides sufficient conditions for improvability based on the function definitions above.

Theorem 2: *Let $\theta = \theta^*$ be the unique maximizer of $F_\theta(\mathbf{0})$ and $\alpha = F_{\theta^*}(\mathbf{0})$. Then, the detector is improvable*

- *if there exists a K-dimensional vector $\mathbf{z}$ such that $f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) f^{(1)}(\mathbf{x}, \mathbf{z}) > 0$ is satisfied at $\mathbf{x} = \mathbf{0}$; or,*
- *if there exists a K-dimensional vector $\mathbf{z}$ such that $f^{(1)}(\mathbf{x}, \mathbf{z}) > 0$, $f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) < 0$, and $f^{(2)}(\mathbf{x}, \mathbf{z}) f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) > f_{\theta^*}^{(2)}(\mathbf{x}, \mathbf{z}) f^{(1)}(\mathbf{x}, \mathbf{z})$ are satisfied at $\mathbf{x} = \mathbf{0}$.*

Proof: Please see Appendix A.

In order to better understand the conditions in Theorem 2, it is first noted from (9) that $F_\theta(\mathbf{0})$ represents the conditional risk given θ in the absence of additive noise, $\mathrm{R}^{\mathbf{x}}_\theta(\phi)$. Therefore, θ^* in the theorem corresponds to the value of θ for which the original conditional risk $\mathrm{R}^{\mathbf{x}}_\theta(\phi)$ is maximum and that maximum value is assumed to be equal to the upper limit α. In other words, it is assumed that, in the absence of additive noise, the original detector already achieves the upper limit on the conditional risks for the modified observations specified in (3). Then, the results in the theorem imply that, under the stated conditions, it is possible to obtain a noise p.d.f. with multiple mass points around $\mathbf{n} = \mathbf{0}$, which can reduce the Bayes risk under the constraint on the conditional risks.

In order to present alternative improvability conditions to those in Theorem 2, we extend the conditions that are developed for simple binary hypothesis-testing problems in the Neyman-Pearson framework in [12] to our problem in (12). To that aim, we first define a new function $H(t)$ as

$$H(t) \triangleq \inf\left\{F(\mathbf{n}) \;\middle|\; \max_{\theta \in \Lambda} F_\theta(\mathbf{n}) = t \ , \ \mathbf{n} \in \mathbb{R}^K \right\} , \tag{23}$$

which specifies the minimum Bayes risk for a given value of the maximum conditional risk considering constant values of additive noise.

From (23), it is observed that if there exists $t_0 \leq \alpha$ such that $H(t_0) < F(\mathbf{0})$, then the system is improvable, because under such a condition there exists a noise component $\mathbf{n}_0$ such that $F(\mathbf{n}_0) < F(\mathbf{0})$ and $\max_{\theta \in \Lambda} F_\theta(\mathbf{n}_0) \leq \alpha$, meaning that the detector performance can be improved by adding a constant $\mathbf{n}_0$ to the observation. However, improvability of a detector via constant noise is not very common in practice. Therefore, the following improvability condition is obtained for more practical scenarios.

Theorem 3: *Let the maximum value of the conditional risks in the absence of additive noise be defined as $\tilde{\alpha} \triangleq \max_{\theta \in \Lambda} \mathrm{R}^{\mathbf{x}}_\theta(\phi)$*

and $\tilde{\alpha} \leq \alpha$. If $H(t)$ in (23) is second-order continuously differentiable around $t = \tilde{\alpha}$ and satisfies $H''(\tilde{\alpha}) < 0$, then the detector is improvable.

Proof: Please see Appendix B.

Similar to Theorem 2, Theorem 3 provides sufficient conditions that guarantee the improvability of a detector according to the restricted Bayes criterion. Note that $H(t)$ in Theorem 3 is always a single-variable function irrespective of the dimension of the observation vector, which facilitates simple evaluation of the conditions in the theorem. However, the main challenge can be to obtain an expression for $H(t)$ in (23) in certain scenarios. On the other hand, Theorem 2 deals with $F_\theta(\cdot)$ and $F(\cdot)$ directly, without defining an auxiliary function like $H(t)$. Therefore, implementation of Theorem 2 can be more efficient in some cases. However, the functions in Theorem 2 are always K-dimensional, which can make the evaluation of its conditions more complicated than that in Theorem 3 in some other cases. In Section IV, comparisons of the improvability results based on direct evaluations of $F_\theta(\cdot)$ and $F(\cdot)$, and those based on $H(t)$ are provided.

C. On the Optimal Additive Noise

In general, the optimization problem in (12) is a non-convex problem and has very high computational complexity since the optimization needs to be performed over functions. In Section III, it is shown that (12) simplifies significantly in the case of simple hypothesis-testing problems. However, in the composite case, the solution is quite difficult to obtain in general. Therefore, a p.d.f. approximation technique [49] can be employed in this section in order to obtain an approximate solution of the problem.

Let the optimal noise p.d.f. be approximated by

$$p_{\mathbf{N}}(\mathbf{n}) = \sum_{i=1}^{L} \nu_i \, \psi_i(\mathbf{n} - \mathbf{n}_i) \ , \tag{24}$$

where $\nu_i \geq 0$, $\sum_{i=1}^{L} \nu_i = 1$, and $\psi_i(\cdot)$ is a window function with $\psi_i(\mathbf{x}) \geq 0 \ \forall \mathbf{x}$ and $\int \psi_i(\mathbf{x}) d\mathbf{x} = 1$, for $i = 1, \ldots, L$. In addition, let ς_i denote a scaling parameter for the ith window function. The p.d.f. approximation technique in (24) is referred to as *Parzen window density estimation*, which has the property of mean-square convergence to the true p.d.f. under certain conditions [50]. From (24), the optimization problem in (12) can be expressed as[4]

$$\begin{aligned} &\min_{\{\nu_i, \mathbf{n}_i, \varsigma_i\}_{i=1}^{L}} \ \sum_{i=1}^{L} \nu_i f_{\mathbf{n}_i}(\varsigma_i) \ , \\ &\text{subject to} \ \max_{\theta \in \Lambda} \sum_{i=1}^{L} \nu_i f_{\theta, \mathbf{n}_i}(\varsigma_i) \leq \alpha \ , \end{aligned} \tag{25}$$

where $f_{\mathbf{n}_i}(\varsigma_i) \triangleq \int F(\mathbf{n}) \psi_i(\mathbf{n} - \mathbf{n}_i) d\mathbf{n}$ and $f_{\theta, \mathbf{n}_i}(\varsigma_i) \triangleq \int F_\theta(\mathbf{n}) \psi_i(\mathbf{n} - \mathbf{n}_i) d\mathbf{n}$.

In (25), the optimization is performed over all the parameters of the window functions in (24). Therefore, the performance of the approximation technique is determined mainly by the the number of window functions, L. As L increases, the approximate solution can get closer to the optimal solution for the additive noise p.d.f. Therefore, in general, an improved detector performance can be expected for larger values of L.

Although (25) is significantly simpler than (12), it is still not a convex optimization problem in general. Therefore, global optimization techniques, such as particle-swarm optimization (PSO) [51]-[53], genetic algorithms and differential evolution [54], can be used to calculate the optimal solution [29], [49]. In Section IV, the PSO algorithm is used to obtain the optimal noise p.d.f.s for the numerical examples.

Although the calculation of the optimal noise p.d.f. requires significant effort as discussed above, some of its properties can be obtained without solving the optimization problem in (12). To that aim, let $F_{\min}$ represent the minimum value of $H(t)$ in (23); that is, $F_{\min} = \min_t H(t)$. In addition, suppose that this minimum is attained at $t = t_{\mathrm{m}}$.[5] Then, one immediate observation is that if t_{m} is less than or equal to the conditional risk limit α, then the noise component $\mathbf{n}_{\mathrm{m}}$ that results in $\max_{\theta \in \Lambda} F_\theta(\mathbf{n}_{\mathrm{m}}) = t_{\mathrm{m}}$ is the optimal noise component; that is, the optimal noise is a constant in that scenario, $p_{\mathbf{N}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{n}_{\mathrm{m}})$. On the other hand, if $t_{\mathrm{m}} > \alpha$, then it can be shown that the optimal solution of (12) satisfies $\max_{\theta \in \Lambda} \mathrm{R}_\theta^{\mathbf{y}}(\phi) = \alpha$ (Appendix C).

III. Noise Enhanced Simple Hypothesis-Testing

In this section, noise enhanced detection is studied in the restricted Bayesian framework for *simple* hypothesis-testing problems. In simple hypothesis-testing problems, each hypothesis corresponds to a single probability distribution [40]. In other words, the generic composite hypothesis-testing problem in (1) reduces to a simple hypothesis-testing problem if each Λ_i consists of a single element.

Since the simple hypothesis-testing problem is a special case of the composite one, the results in Section II are also valid for this section. However, by using the special structure of simple hypotheses, we obtain additional results in this section that are not valid for composite hypothesis-testing problems. It should be noted that both composite and simple hypothesis-testing problems are used to model various practical detection examples [40], [55]; hence, specific results can be useful in different applications.

A. Problem Formulation

The problem can be formulated as in Section II-A by defining $\Lambda_i = \{\theta_i\}$ for $i = 0, 1, \ldots, M-1$ in (1). In addition, instead of the prior p.d.f. $w(\theta)$, the prior probabilities of the hypotheses can be defined by $\pi_0, \pi_1, \ldots, \pi_{M-1}$ with $\sum_{i=0}^{M-1} \pi_i = 1$. Then, the optimal additive noise problem in (3) becomes

$$\begin{aligned} &\min_{p_{\mathbf{N}}(\cdot)} \ \sum_{i=0}^{M-1} \pi_i \mathrm{R}_i^{\mathbf{y}}(\phi) \ , \\ &\text{subject to} \ \max_{i \in \{0,1,\ldots,M-1\}} \mathrm{R}_i^{\mathbf{y}}(\phi) \leq \alpha \ , \end{aligned} \tag{26}$$

where $\sum_{i=0}^{M-1} \pi_i \mathrm{R}_i^{\mathbf{y}}(\phi) \triangleq r^{\mathbf{y}}(\phi)$ is the Bayes risk and $\mathrm{R}_i^{\mathbf{y}}(\phi)$ is the conditional risk of ϕ given $\mathcal{H}_i$ for the noise modified observation $\mathbf{y}$, which is given by

$$\mathrm{R}_i^{\mathbf{y}}(\phi) = \sum_{j=0}^{M-1} \mathrm{C}_{ji} \mathrm{P}_i^{\mathbf{y}}(\Gamma_j) \ , \tag{27}$$

[4] As in [12], it is possible to perform the optimization over single-variable functions by considering mapping of the noise $\mathbf{n}$ via $F(\mathbf{n})$ or $F_\theta(\mathbf{n})$.

[5] If there are multiple t values that result in the minimum value $F_{\min}$, then the minimum of those values can be considered.

with $\mathrm{P}_i^{\mathbf{y}}(\Gamma_j)$ denoting the probability that $\mathbf{y} \in \Gamma_j$ when $\mathcal{H}_i$ is the true hypothesis, and C_{ji} defining the cost of deciding $\mathcal{H}_j$ when $\mathcal{H}_i$ is true. As in Section II-A, the constraint α sets an upper limit on the conditional risks, and its value is determined depending on the amount of uncertainty in the prior probabilities.

In order to investigate the optimal solution of (26), an alternative expression for $\mathrm{R}_i^{\mathbf{y}}(\phi)$ is obtained first. Since the additive noise $\mathbf{n}$ is independent of the observation $\mathbf{x}$, $\mathrm{P}_i^{\mathbf{y}}(\Gamma_j)$ becomes

$$\mathrm{P}_i^{\mathbf{y}}(\Gamma_j) = \int_{\Gamma_j} p_i^{\boldsymbol{Y}}(\mathbf{y}) d\mathbf{y} = \int_{\Gamma_j} \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) p_i^{\mathbf{X}}(\mathbf{y}-\mathbf{n})\, d\mathbf{n}\, d\mathbf{y} \ , \tag{28}$$

where $p_i^{\mathbf{X}}(\cdot)$ and $p_i^{\boldsymbol{Y}}(\cdot)$ represent the p.d.f.s of the original observation and the noise modified observation, respectively, when hypothesis $\mathcal{H}_i$ is true. Then, (27) can be expressed, from (28), as

$$\begin{aligned} \mathrm{R}_i^{\mathbf{y}}(\phi) &= \sum_{j=0}^{M-1} \mathrm{C}_{ji} \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) \int_{\Gamma_j} p_i^{\mathbf{X}}(\mathbf{y}-\mathbf{n})\, d\mathbf{y}\, d\mathbf{n} \\ &= \sum_{j=0}^{M-1} \mathrm{C}_{ji}\, \mathrm{E}\{F_{ij}(\mathbf{N})\} = \mathrm{E}\{F_i(\mathbf{N})\} \ , \end{aligned} \tag{29}$$

with

$$F_{ij}(\mathbf{n}) \triangleq \int_{\Gamma_j} p_i^{\mathbf{X}}(\mathbf{y}-\mathbf{n}) d\mathbf{y} \ , \tag{30}$$

$$F_i(\mathbf{n}) \triangleq \sum_{j=0}^{M-1} \mathrm{C}_{ji} F_{ij}(\mathbf{n}) \ . \tag{31}$$

Based on the relation in (29), the optimization problem in (26) can be reformulated as

$$\begin{aligned} &\min_{p_{\mathbf{N}}(\cdot)} \sum_{i=0}^{M-1} \pi_i \mathrm{E}\{F_i(\mathbf{N})\} \ , \\ &\text{subject to} \max_{i\in\{0,1,\dots,M-1\}} \mathrm{E}\{F_i(\mathbf{N})\} \le \alpha \ . \end{aligned} \tag{32}$$

If a new auxiliary function is defined as $F(\mathbf{n}) \triangleq \sum_{i=0}^{M-1} \pi_i F_i(\mathbf{n})$, (32) becomes

$$\begin{aligned} &\min_{p_{\mathbf{N}}(\cdot)} \ \mathrm{E}\{F(\mathbf{N})\} \ , \\ &\text{subject to} \max_{i\in\{0,1,\dots,M-1\}} \mathrm{E}\{F_i(\mathbf{N})\} \le \alpha \ . \end{aligned} \tag{33}$$

Note that under UCA; that is, when $\mathrm{C}_{ji} = 1$ for $j \neq i$, and $\mathrm{C}_{ji} = 0$ for $j = i$, $F_i(\mathbf{N})$ becomes equal to $1 - F_{ii}(\mathbf{N})$.

It should be noted from the definitions in (30) and (31) that $F_i(\mathbf{0})$ corresponds to the conditional risk given $\mathcal{H}_i$ for the original observation $\mathbf{x}$, $\mathrm{R}_i^{\mathbf{x}}(\phi)$. Therefore, $F(\mathbf{0})$ defines the original Bayes risk, $r^{\mathbf{x}}(\phi)$.

B. Optimal Additive Noise

The optimization problem in (33) seems quite difficult to solve in general as it requires a search over all possible noise p.d.f.s. However, in the following, it is shown that an optimal additive noise p.d.f. can be represented by a discrete probability distribution with at most M mass points in most practical cases. To that aim, suppose 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 $j = 1,\dots,K$, which is a reasonable assumption since additive noise cannot have infinitely large amplitudes in practice. Then, the following theorem states the discrete nature of the optimal additive noise.

Theorem 4: *If $F_i(\cdot)$ in (32) are continuous functions, then the p.d.f. of an optimal additive noise can be expressed as $p_{\mathbf{N}}(\mathbf{n}) = \sum_{l=1}^{M} \lambda_l\, \delta(\mathbf{n}-\mathbf{n}_l)$, where $\sum_{l=1}^{M} \lambda_l = 1$ and $\lambda_l \ge 0$ for $l = 1, 2, \dots, M$.*

Proof: The proof employs a similar approach to those used for the related results in [12], [29] and [49]. First, the following set is defined:

$$\begin{aligned} U = \{(u_0, u_1, \dots, u_{M-1}) \ : \ u_i = F_i(\mathbf{n}) \ , \\ i = 0, 1, \dots, M-1 \ , \ \text{for} \ \boldsymbol{a} \preceq \mathbf{n} \preceq \boldsymbol{b}\} \ . \end{aligned} \tag{34}$$

In addition, V is defined as the convex hull of U [56]. Since $F_i(\cdot)$ are continuous functions, U is a bounded and closed subset of $\mathbb{R}^M$. Hence, U is a compact set. Therefore, its convex hull V is a closed subset of $\mathbb{R}^M$ [29]. Next, set W is defined as

$$\begin{aligned} W = \Big\{(w_0, w_1, \dots, w_{M-1}) \ : \ w_i = \mathrm{E}\{F_i(\mathbf{n})\} \ , \\ i = 0, 1, \dots, M-1, \ \forall p_{\mathbf{N}}(\mathbf{n}), \ \boldsymbol{a} \preceq \mathbf{n} \preceq \boldsymbol{b}\Big\} \ , \end{aligned} \tag{35}$$

where $p_{\mathbf{N}}(\mathbf{n})$ is the p.d.f. of the additive noise.

As V is the convex hull of U, each element of V can be expressed as $\mathbf{v} = \sum_{l=1}^{N_L} \lambda_l \left(F_0(\mathbf{n}_l), F_1(\mathbf{n}_l), \dots, F_{M-1}(\mathbf{n}_l)\right)$, where $\sum_{l=1}^{N_L} \lambda_l = 1$, and $\lambda_l \ge 0 \ \forall l$. On the other hand, each $\mathbf{v}$ is also an element of W as it can be obtained for $p_{\mathbf{N}}(\mathbf{n}) = \sum_{l=1}^{N_L} \lambda_l\, \delta(\mathbf{n}-\mathbf{n}_l)$. Hence, $V \subseteq W$ [29]. In addition, since for any vector random variable $\boldsymbol{\Theta}$ taking values in set Ω, its expected value, $\mathrm{E}\{\boldsymbol{\Theta}\}$, is in the convex hull of Ω [57], (34) and (35) implies that W is in the convex hull V of U; that is, $V \supseteq W$. Since $V \subseteq W$ and $V \supseteq W$, it means that $W = V$ [29]. Therefore, according to Carathéodory's theorem [58], [59], any point in V (or, W) can be expressed as the convex combination of at most $(M+1)$ points in U as the dimension of U is smaller than or equal to M. Since the aim is to minimize the average of the conditional risks, the optimal solution corresponds to the boundary of W. As W (or, V) is a closed set as mentioned at the beginning of the proof, it contains its own boundary [29]. Since any point at the boundary of W can be expressed as the convex combination of at most M elements in U [58], an optimal noise p.d.f. can be represented by a discrete random variable with M mass points as stated in the theorem. □

From Theorem 4, the optimization problem in (33) can be simplified as

$$\begin{aligned} &\min_{\{\lambda_l, \mathbf{n}_l\}_{l=1}^M} \ \sum_{l=1}^{M} \lambda_l F(\mathbf{n}_l) \ , \\ &\text{subject to} \max_{i\in\{0,1,\dots,M-1\}} \sum_{l=1}^{M} \lambda_l F_i(\mathbf{n}_l) \le \alpha \ , \\ &\sum_{l=1}^{M} \lambda_l = 1 \ , \quad \lambda_l \ge 0 \ , \quad l = 1, \dots, M \ . \end{aligned} \tag{36}$$

The optimization in (36) is considerably simpler than that in (33) since the former is over a set of variables instead of functions. However, (36) can still be a nonconvex optimization problem in general; hence, global optimization techniques, such as PSO [51] and differential evolution [54] may be needed.

In order to provide a convex relaxation [60] of the optimization problem in (36) and to obtain an approximate solution in polynomial time, one can assume that additive noise $\mathbf{n}$ can take only finitely many known values specified by $\tilde{\mathbf{n}}_1,\ldots,\tilde{\mathbf{n}}_L$ [29]. This scenario, for example, corresponds to digital systems in which the signals can take only finitely many different levels. Then, the aim becomes the determination of the weights $\tilde{\lambda}_1,\ldots,\tilde{\lambda}_L$ of those possible noise values. In that case, (33) can be formulated as

$$\begin{aligned}&\min_{\{\tilde{\lambda}_l\}_{l=1}^L} \sum_{l=1}^{L}\tilde{\lambda}_l F(\tilde{\mathbf{n}}_l) \ ,\\ &\text{subject to} \max_{i\in\{0,1,\ldots,M-1\}} \sum_{l=1}^{L}\tilde{\lambda}_l F_i(\tilde{\mathbf{n}}_l) \leq \alpha \ ,\\ &\sum_{l=1}^{L}\tilde{\lambda}_l = 1 \ , \quad \tilde{\lambda}_l \geq 0 \ , \quad l=1,\ldots,L \ ,\end{aligned} \tag{37}$$

which is a linearly constrained linear programming (LCLP) problem; hence, can be solved in polynomial time [60]. It should be noted that as the optimization is performed over more noise values (as L increases), the solution gets closer to the optimal solution of (33).

As an alternative approach, an analytical solution similar to that in [12] can also be proposed for obtaining the optimal additive noise. First, consider the optimization problem in (32) for $M=2$; i.e., the binary case. If functions $F_0(\mathbf{n})$ and $F_1(\mathbf{n})$ are monotone, then t_0 and t_1 can be defined as $t_0 = F_0(\mathbf{n})$ and $t_1 = F_1(\mathbf{n})$. Otherwise, let t_0 and t_1 be defined as follows:

$$\begin{aligned}t_0(t) &\triangleq \inf\left\{F_0(\mathbf{n}) \ \middle| \ F_1(\mathbf{n}) = t \ , \ \mathbf{n}\in\mathbb{R}^K\right\} \ ,\\ t_1(t) &\triangleq \inf\left\{F_1(\mathbf{n}) \ \middle| \ F_0(\mathbf{n}) = t \ , \ \mathbf{n}\in\mathbb{R}^K\right\} \ .\end{aligned} \tag{38}$$

In general, there can exist multiple values of $F_1(\mathbf{n})$ corresponding to a given value of $F_0(\mathbf{n})$. However, the definitions of t_0 and t_1 in (38) make sure that only the best (minimum) value of $F_1(\mathbf{n})$ corresponding to a given $F_0(\mathbf{n})$ is considered, and vice versa. Therefore, t_1 can be expressed as $t_1 = g(t_0)$, where $g(t_0)$ is a monotone function of t_0 and is defined on the range of t_0, which is denoted by $[t_{0,\min}, t_{0,\max}]$ with $t_{0,\min} = \min t_0$ and $t_{0,\max} = \max t_0$. We call the set of t_0 for which $g(t_0)$ and t_0 satisfy the constraints (cf. (32)) as the *feasible domain*. Then, let a new function B be defined as follows:

$$B(t_0) \triangleq \pi_0 t_0 + \pi_1 g(t_0) \ . \tag{39}$$

If $B(t_0)$ takes its global minimum value in the feasible domain, then the optimal Bayes risk is equal to that minimum value and the optimal additive noise can be represented by a constant value. For example, if $t_0^* = \arg\min_{t_0} B(t_0)$, then the optimal additive noise p.d.f. can be expressed as $p_{\mathbf{N}}(\mathbf{n}) = \delta(\mathbf{n}-\mathbf{n}_0)$, where $\mathbf{n}_0$ satisfies $F_0(\mathbf{n}_0) = t_0^*$.[6] On the other hand, if $B(t_0)$ achieves its global minimum value outside the feasible domain, then an analytic solution for the optimal additive noise p.d.f. can be obtained as explained in the following. At the end of Section II-C, it was stated that the maximum value of the optimal conditional risks must be equal to the constraint level α for the case considered here. This implies that the optimal (t_0,t_1) pair is equal to one of the following: (α,β) or (γ,α), where β and γ are such that $g(\alpha)=\beta$ and $g(\gamma)=\alpha$. It should be noted that if $g(t_0)$ is a decreasing function and γ is larger than α, then the feasible domain is an empty set implying that there is no solution satisfying the constraint.

[6]If there are multiple such $\mathbf{n}_0$'s, then the one that minimizes $F_1(\mathbf{n}_0)$ should be chosen.

Since $g(t_0)$ is a monotone function and the maximum of the optimal conditional risks must be equal to α, the feasible domain must be in the form of an interval, say $[a,b]$, and the value of t_0 corresponding to the optimal solution must be equal to either a or b. In the following derivations, it is assumed that the value of t_0 corresponding to the optimal solution is b, and $B(t_0)$ takes its global minimum value for $t_0 > b$. However, it should be noted that these assumptions do not reduce the generality of the results. In other words, the derivations based on the other possible assumptions yield the same result.

Similar to [12], the following auxiliary function is defined:

$$Z(t_0,k) \triangleq B(t_0) + k t_0 \ , \tag{40}$$

where $k\in\mathbb{R}$. It is observed that Z is an increasing function of k. Let the range of t_0 be partitioned into $\mathcal{I}_1 = [t_{0,\min}, b)$ and $\mathcal{I}_2 = [b, t_{0,\max}]$. In addition, two new functions are defined as follows:

$$\begin{aligned}v_1(k) &\triangleq \min_{t_0\in\mathcal{I}_1} Z(t_0,k) = Z\left(t_{01}(k),k\right) \ ,\\ v_2(k) &\triangleq \min_{t_0\in\mathcal{I}_2} Z(t_0,k) = Z\left(t_{02}(k),k\right) \ ,\end{aligned} \tag{41}$$

where $t_{01}(k)$ is the value of $t_0\in\mathcal{I}_1$ that minimizes Z for a given k, and similarly, $t_{02}(k)$ is value of $t_0\in\mathcal{I}_2$ that minimizes Z for a given k.

From (40) and (41), it is obtained for $k=0$ that $v_2(0) = \min B(t_0) < v_1(0) = B(t_{01}(k))$. On the other hand, as $k\to\infty$, $v_1(k) = B(t_{0,\min}) + k t_{0,\min} < v_2(k) = B(b) + kb$. Therefore, there must exist a $k=k_0$, where $0<k_0<\infty$, such that

$$v = v_1(k_0) = Z(t_{01}(k_0),k_0) = v_2(k_0) = Z(t_{02}(k_0),k_0) \ . \tag{42}$$

Consider the division of the range of t_0 into two disjoint sets A and $\{t_{01}(k_0), t_{02}(k_0)\}$ such that $\{t_{01}(k_0), t_{02}(k_0)\}\cup A = [t_{0,\min}, t_{0,\max}]$. Then, any additive noise p.d.f. can be expressed in the following form:

$$\begin{aligned}p_{\mathbf{n},t_0}(t_0) = &\lambda_1\delta(t_0 - t_{01}(k_0)) + \lambda_2\delta(t_0 - t_{02}(k_0))\\ &+ \mathcal{I}_A(t_0) p_{\mathbf{n},t_0}(t_0) \ ,\end{aligned} \tag{43}$$

where $\mathcal{I}_A(t_0)$ is an indicator function such that $\mathcal{I}_A(t_0)=1$ if $t_0\in A$, $\mathcal{I}_A(t_0)=0$ otherwise [12]. By definition, $\lambda_1+\lambda_2+\int_A p_{\mathbf{n},t_0}(t_0)\,dt_0 = 1$ should be satisfied. In addition, the expectation of Z in (40) over t_0 can be bounded as follows:

$$\begin{aligned}\mathrm{E}\{Z(t_0,k_0)\} &= \lambda_1 v + \lambda_2 v + \int_A Z(t_0,k_0) p_{\mathbf{n},t_0}(t_0)\,dt_0 \ ,\\ &= v + \int_A [Z(t_0,k_0) - v] p_{\mathbf{n},t_0}(t_0)\,dt_0 \ ,\\ &\geq \ v \ ,\end{aligned} \tag{44}$$

where the first expression is obtained from (42) and (43), and the final inequality is obtained from the fact that $Z(t_0,k_0)\geq v$ for $t_0\in A$ (cf. (41) and (42)). This lower bound is achieved for $p_{\mathbf{n},t_0}(t_0) = \lambda_1\delta(t_0 - t_{01}(k_0)) + \lambda_2\delta(t_0 - t_{02}(k_0))$, with $\lambda_1+\lambda_2 = 1$. Hence, $p_{\mathbf{n},t_0}(t_0) = 0$ for $t_0\in A$.

From (39) and (40), the Bayes risk $r^{\mathbf{y}}(\phi)$ can be expressed as $r^{\mathbf{y}}(\phi) = \mathrm{E}\{B(t_0)\} = \mathrm{E}\{Z(t_0,k_0)\} - k_0\mathrm{E}\{t_0\}$. Since $t_{01}(k_0) < b$ and $t_{02}(k_0)\geq b$, one can achieve $\mathrm{E}\{t_0\} = b$ by using a noise component with p.d.f. $p_{\mathbf{n},t_0}(t_0) = \lambda_1\delta(t_0 -$

$t_{01}(k_0)) + \lambda_2\delta(t_0 - t_{02}(k_0))$, where $\lambda_1 + \lambda_2 = 1$ with appropriate values for λ_1 and λ_2. Thus, the optimal additive noise p.d.f. is $p_{\mathbf{n},t_0}(t_0) = \lambda_1\delta(t_0-t_{01}(k_0))+\lambda_2\delta(t_0-t_{02}(k_0))$, where $\lambda_1+\lambda_2 = 1$ and $\lambda_1 t_{01}(k_0) + \lambda_2 t_{02}(k_0) = b$, and the optimal Bayes risk is given by $r^{\mathbf{y}}_{\mathrm{opt}}(\phi) = \mathrm{E}\{B(t_0)\} = v-k_0 b$.

Since $Z(t_0, k_0)$ has (local) minimum values at $t_0 = t_{01}(k_0)$ and $t_0 = t_{02}(k_0)$, if $B(t_0)$ is continuously differentiable, then $\partial Z(t_{01}(k_0), k_0)/\partial t_0 = \partial Z(t_{02}(k_0), k_0)/\partial t_0 = 0$. Then, (40) implies the following equalities:

$$\frac{dB(t_{01}(k_0))}{dt_0} = \frac{dB(t_{02}(k_0))}{dt_0} = -k_0 \ . \tag{45}$$

From (42), we also have the following relation:

$$\frac{B(t_{01}(k_0)) - B(t_{02}(k_0))}{t_{01}(k_0) - t_{02}(k_0)} = -k_0 \ . \tag{46}$$

Therefore, (45) and (46) can be used to obtain the following result:

$$\frac{B(t_{01}(k_0)) - B(t_{02}(k_0))}{t_{01}(k_0) - t_{02}(k_0)} = \frac{dB(t_{01}(k_0))}{dt_0} = \frac{dB(t_{02}(k_0))}{dt_0} \ . \tag{47}$$

From the equalities in (47), one can find $t_{01}(k_0)$ and $t_{02}(k_0)$, and the corresponding mass points $\mathbf{n}_1$ and $\mathbf{n}_2$ that satisfy $t_{01}(k_0) = F_0(\mathbf{n}_1)$ and $t_{02}(k_0) = F_0(\mathbf{n}_2)$.[7]

After obtaining $\mathbf{n}_1$ and $\mathbf{n}_2$ as described above, the corresponding weights λ_1 and λ_2 calculated from the following equations: $\lambda_1 + \lambda_2 = 1$ and $\lambda_1 t_{01}(k_0) + \lambda_2 t_{02}(k_0) = b$. Due to the fact that the maximum of the optimal conditional risks must be α, b must be equal to the constraint level α or must satisfy $g(b) = \alpha$. These two cases should be checked separately and then the one corresponding to the optimal solution should be determined. In other words, the weight pairs corresponding to $t_0 = \alpha$ and $t_1 = g(t_0) = \alpha$ should be calculated separately, and then the one that results in better performance should be selected. An alternative approach to determine b is to find where $B(t_0)$ takes its global minimum value. If $B(t_0)$ takes its global minimum value for $t_0 > \alpha$, then b must be equal to α; otherwise, b must be found from $g(b) = \alpha$. After finding b, the optimal weight pair can easily be obtained from $\lambda_1+\lambda_2 = 1$ and $\lambda_1 t_{01}(k_0)+\lambda_2 t_{02}(k_0) = b$.

The analytic approach described above for the binary case can also be extended to the M-ary case for $M > 2$. However, in that case, only the mass points, $\mathbf{n}_1, \ldots, \mathbf{n}_M$, can be found analytically. The weights, $\lambda_1, \ldots, \lambda_M$, should be found via a numerical approach. Such a semi-analytical solution can still provide significant computational complexity reduction in some cases since the weights, which are not determined analytically, are easier to search for than the mass points, as the weights are always scalar whereas the mass points can also be multidimensional. The analytical approach to obtaining the mass points in the M-ary case is a simple extension of that in the binary case. Mainly, a function t_{M-1} should be defined as $t_{M-1} \triangleq g(t_0, \ldots, t_{M-2}) \triangleq \inf\{F_{M-1}(\mathbf{n}) \,\big|\, F_0(\mathbf{n}) = t_0, \ldots, F_{M-2}(\mathbf{n}) = t_{M-2}\,, \mathbf{n} \in \mathbb{R}^K\}$, function B in (39) should be generalized as $B(t_0, \ldots, t_{M-2}) \triangleq \pi_0 t_0 + \cdots + \pi_{M-1} g(t_0, \ldots, t_{M-2})$, and Z should be modified as $Z(t_0, \ldots, t_{M-2}, k_1, \ldots, k_{M-1}) \triangleq B(t_0, \ldots, t_{M-2}) + k_1 t_0 + \cdots + k_{M-1} t_{M-2}$. The resulting equations provide a generalization of those in (47), the details of which are not presented here due to the space limitations.

[7] If there are multiple such $\mathbf{n}_1$'s ($\mathbf{n}_2$'s), then the one that minimizes $F_1(\mathbf{n}_1)$ ($F_1(\mathbf{n}_2)$) should be chosen.

C. Improvability and Nonimprovability Conditions

In this section, various sufficient conditions are derived in order to determine when the performance of a detector can or cannot be improved via additive independent noise according to the restricted Bayes criterion.

For the nonimprovability conditions, Theorem 1 in Section II-B already provides a quite explicit statement to evaluate the nonimprovability. Therefore, it is also practical for simple hypothesis-testing problems, as observed in the example after Theorem 1. In accordance with the notation in this section, Theorem 1 can be restated for simple hypothesis-testing problems as follows. *Assume that there exits $i \in \{0, 1, \ldots, M-1\}$ such that $F_i(\mathbf{n}) \le \alpha$ implies $F(\mathbf{n}) \ge F(\mathbf{0})$ for all $\mathbf{n} \in \mathcal{S}_{\mathrm{n}}$, where $\mathcal{S}_{\mathrm{n}}$ is a convex set consisting of all possible values of additive noise $\mathbf{n}$. If $F_i(\mathbf{n})$ and $F(\mathbf{n})$ are convex functions over $\mathcal{S}_{\mathrm{n}}$, then the detector is nonimprovable.*

Regarding the improvability conditions, in addition to Theorem 2 and Theorem 3 in Section II-B, new sufficient conditions that are specific to simple hypothesis-testing problems are provided in the following. To that aim, it is first assumed that $F_i(\mathbf{x})$ for $i = 0, 1, \ldots, M-1$ and $F(\mathbf{x})$, defined in Section III-A, are second-order continuously differentiable around $\mathbf{x} = \mathbf{0}$. In addition, similar to (19)-(22), the following functions are defined.

$$f^{(1)}_j(\mathbf{x}, \mathbf{z}) \triangleq \sum_{i=1}^{K} z_i \frac{\partial F_j(\mathbf{x})}{\partial x_i} \ , \tag{48}$$

$$f^{(1)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{i=1}^{K} z_i \frac{\partial F(\mathbf{x})}{\partial x_i} \ , \tag{49}$$

$$f^{(2)}_j(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^{K}\sum_{i=1}^{K} z_l z_i \frac{\partial^2 F_j(\mathbf{x})}{\partial x_l \partial x_i} \ , \tag{50}$$

$$f^{(2)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^{K}\sum_{i=1}^{K} z_l z_i \frac{\partial^2 F(\mathbf{x})}{\partial x_l \partial x_i} \ , \tag{51}$$

for $j = 0, 1, \ldots, M-1$, where x_i and z_i represent the ith components of $\mathbf{x}$ and $\mathbf{z}$, respectively.

Note that the result in Theorem 2 can also be used for simple hypothesis-testing problems when there exists a unique maximizer $i = i^*$ of the original conditional risks, $F_i(\mathbf{0}) = \mathrm{R}^{\mathbf{x}}_i(\phi)$. In the following, more generic improvability conditions, which cover the cases with multiple maximizers of $F_i(\mathbf{0})$ as well, are obtained for simple hypothesis-testing problems. Let $\mathcal{S}_\alpha$ denote the set of indices for which $F_i(\mathbf{0})$ achieves the maximum value of α, and let $\bar{\mathcal{S}}_\alpha$ represent the set of indices with $F_i(\mathbf{0}) < \alpha$; that is,

$$\mathcal{S}_\alpha = \{i \in \{0, 1, \ldots, M-1\} \mid F_i(\mathbf{0}) = \alpha\} \ , \tag{52}$$

$$\bar{\mathcal{S}}_\alpha = \{i \in \{0, 1, \ldots, M-1\} \mid F_i(\mathbf{0}) < \alpha\} \ . \tag{53}$$

In addition, let $\mathcal{S}_\alpha \cup \bar{\mathcal{S}}_\alpha = \{0, 1, \ldots, M-1\}$, meaning that $F_i(\mathbf{0}) = \mathrm{R}^{\mathbf{x}}_i(\phi) \le \alpha$ for $i = 0, 1, \ldots, M-1$. Consider the functions in (48)-(51), and define set $\mathcal{F}_n$ ($n = 1, 2$) as the set that consists of $f^{(n)}(\mathbf{x}, \mathbf{z})$ and $f^{(n)}_i(\mathbf{x}, \mathbf{z})$ for $i \in \mathcal{S}_\alpha$; that is,

$$\mathcal{F}_n = \left\{ f^{(n)}(\mathbf{x}, \mathbf{z}), f^{(n)}_i(\mathbf{x}, \mathbf{z}) \text{ for } i \in \mathcal{S}_\alpha \right\} \tag{54}$$

for $n = 1, 2$. Note that $\mathcal{F}_n$ has $|\mathcal{S}_\alpha| + 1$ elements, where $|\mathcal{S}_\alpha|$ represents the number of elements in $\mathcal{S}_\alpha$. In addition, $\mathcal{F}_n(j)$ will be used to refer to the jth element of $\mathcal{F}_n$. It should be noted that $\mathcal{F}_n(1) = f^{(n)}(\mathbf{x}, \mathbf{z})$ and $\mathcal{F}_n(j) = f^{(n)}_{\mathcal{S}_\alpha(j-1)}(\mathbf{x}, \mathbf{z})$

for $j = 2, \ldots, |\mathcal{S}_\alpha| + 1$, where $\mathcal{S}_\alpha(j-1)$ is the $(j-1)$th element of $\mathcal{S}_\alpha$. Finally, the following sets are introduced to define the set of indices j for which $\mathcal{F}_1(j)$ is zero, negative or positive:

$$\mathcal{S}_z = \{j \in \{1, 2, \ldots, |\mathcal{S}_\alpha| + 1\} \mid \mathcal{F}_1(j) = 0\} \ , \quad (55)$$
$$\mathcal{S}_n = \{j \in \{1, 2, \ldots, |\mathcal{S}_\alpha| + 1\} \mid \mathcal{F}_1(j) < 0\} \ , \quad (56)$$
$$\mathcal{S}_p = \{j \in \{1, 2, \ldots, |\mathcal{S}_\alpha| + 1\} \mid \mathcal{F}_1(j) > 0\} \ . \quad (57)$$

Based on the definitions in (48)-(57), the following theorem provides sufficient conditions for improvability.

Theorem 5: *For simple hypothesis-testing problems, a detector is improvable according to the restricted Bayes criterion if there exists a K-dimensional vector $\mathbf{z}$ such that the following two conditions are satisfied at $\mathbf{x} = \mathbf{0}$:*

1) $\mathcal{F}_2(j) < 0 \ , \forall j \in \mathcal{S}_z$.

2) *One of the following is satisfied:*

- $|\mathcal{S}_n| = 0$ *or* $|\mathcal{S}_p| = 0$.
- $|\mathcal{S}_n|$ *is a positive even number,* $|\mathcal{S}_p| > 0$*, and*

$$\min_{j \in \mathcal{S}_n} \mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l) \ > \ \max_{j \in \mathcal{S}_p} \mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l). \quad (58)$$

- $|\mathcal{S}_n|$ *is an odd number,* $|\mathcal{S}_p| > 0$*, and*

$$\min_{j \in \mathcal{S}_p} \mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l) \ > \ \max_{j \in \mathcal{S}_n} \mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l). \quad (59)$$

Proof: Please see Appendix D.

Theorem 5 states that whenever the two conditions in the theorem are satisfied, it can be concluded that the detection performance can be improved via additive independent noise. It should be noted that after defining the sets in (52)-(57), it is straightforward to check the conditions stated in the theorem. An example application of Theorem 5 is provided in Section IV, where its practicality and effectiveness are observed.

Finally, another improvability condition is derived as a corollary of Theorem 5.

Corollary 1: *Assume that $F(\mathbf{x})$ and $F_i(\mathbf{x})$, $i = 0, 1, \ldots, M-1$, are second-order continuously differentiable around $\mathbf{x} = \mathbf{0}$ and that $\max_{i \in \{0,1,\ldots,M-1\}} F_i(\mathbf{0}) < \alpha$. 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 not convex around $\mathbf{x} = \mathbf{0}$.*

Proof: Please see Appendix E.

Although Corollary 1 provides simpler improvability conditions than those in Theorem 5, the assumption of $\max_{i \in \{0,1,\ldots,M-1\}} F_i(\mathbf{0}) < \alpha$ makes it less practical. In other words, Corollary 1 assumes that, in the absence of additive noise, the maximum of the original conditional risks is strictly smaller than the upper limit, α. Since it is usually possible to increase the maximum of the conditional risks to reduce the Bayes risk, the scenario in Corollary 1 considers a more trivial case than that in Theorem 5.

IV. Numerical Results

In this section, a binary hypothesis-testing problem is studied first in order to provide a practical example of the results presented in the previous sections. The hypotheses are defined as

$$\mathcal{H}_0 \ : \ x = v \ , \quad \text{versus} \quad \mathcal{H}_1 \ : \ x = A + v \ , \quad (60)$$

where $x \in \mathbb{R}$, $A > 0$ is a known scalar value, and v is symmetric Gaussian mixture noise with the following p.d.f.

$$p_V(x) = \sum_{i=1}^{N_m} w_i \, \psi_i(x - \mu_i) \ , \quad (61)$$

where $w_i \geq 0$ for $i = 1, \ldots, N_m$, $\sum_{i=1}^{N_m} w_i = 1$, and

$$\psi_i(x) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left(\frac{-x^2}{2\,\sigma_i^2}\right) \ , \quad (62)$$

for $i = 1, \ldots, N_m$. Due to the symmetry assumption, $\mu_i = -\mu_{N_m - i + 1}$, $w_i = w_{N_m - i + 1}$ and $\sigma_i = \sigma_{N_m - i + 1}$ for $i = 1, \ldots, \lfloor N_m/2 \rfloor$. In addition, the detector is described by

$$\phi(y) = \begin{cases} 1 \ , & y \geq A/2 \\ 0 \ , & y < A/2 \end{cases} \ , \quad (63)$$

where $y = x + n$, with n representing the additive independent noise term. The aim is to obtain the optimal p.d.f. for the additive noise based on the optimization problem in (26).

Under the assumption of UCA, (60)-(63) can be used to calculate $F_0(x)$ and $F_1(x)$ from (30) and (31) as

$$F_0(x) = \sum_{i=1}^{N_m} w_i \, Q\left(\frac{A/2 - x - \mu_i}{\sigma_i}\right) \ ,$$
$$F_1(x) = \sum_{i=1}^{N_m} w_i \, Q\left(\frac{A/2 + x + \mu_i}{\sigma_i}\right) \ , \quad (64)$$

where $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty \mathrm{e}^{-t^2/2} dt$ denotes the Q-function.

The symmetric Gaussian mixture noise specified above is observed in many practical scenarios [61]-[63]. One important scenario is multiuser wireless communications, in which the desired signal is corrupted by interference from other users as well as by zero-mean Gaussian background noise [64]. In other words, the signal detection example in (60) with symmetric Gaussian mixture noise finds various practical applications.

Since the problem in (60) models a signal detection problem in the presence of noise, we consider two common scenarios in the following simulations. In the first one, it is assumed that the noise-only hypothesis $\mathcal{H}_0$ has a higher prior probability than the signal-plus-noise hypothesis $\mathcal{H}_1$. An example of this scenario is the signal acquisition problem, in which a number of correlation outputs are compared against a threshold to determine the timing/phase of the signal [65]. In the second scenario, equal prior probabilities are assumed for the hypotheses, which can be well-suited for binary communications systems that transmit no signal for bit 0 and a signal for bit 1 (i.e., on-off keying) [66]. For the first scenario, it is assumed that the prior probabilities are known, with some uncertainty, to be equal to $\pi_0 = 0.9$ and $\pi_1 = 0.1$, which is called the *unequal priors* case in the following. On the other hand, $\pi_0 = \pi_1 = 0.5$ is considered for the *equal priors* case. As mentioned in Section II-A, the restricted Bayes criterion mitigates the undesired effects due to the uncertainty in prior probabilities via parameter α, which sets an upper limit on the conditional risks. In the numerical results, symmetric Gaussian mixture noise with $N_m = 4$ is considered, where the mean values of the Gaussian components in the mixture noise in (61) are specified as $[0.033\ 0.52\ {-0.52}\ {-0.033}]$ with corresponding weights of $[0.35\ 0.15\ 0.15\ 0.35]$. In addition, for all the cases, the variances of the Gaussian components in the mixture noise

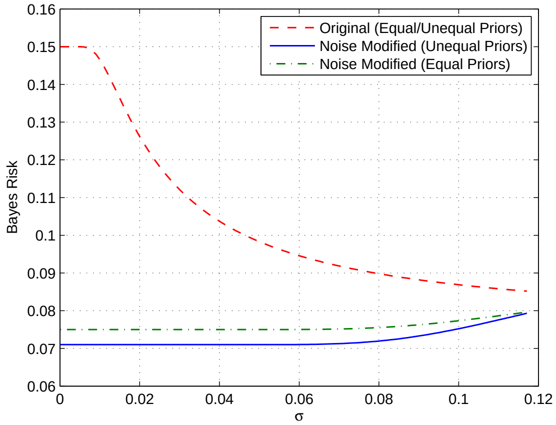


Fig. 1. Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.08$ and $A = 1$.

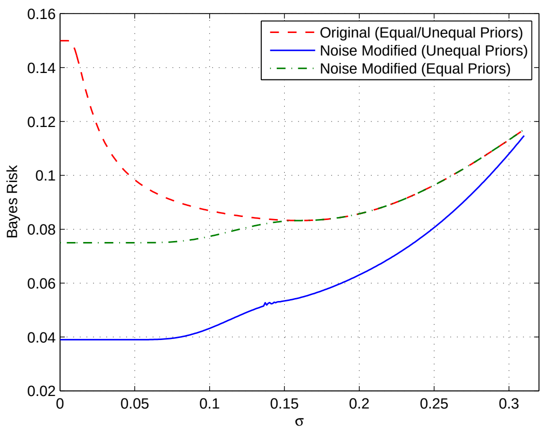


Fig. 2. Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.12$ and $A = 1$.

are assumed to be the same; i.e., $\sigma_i = \sigma$ for $i = 1, \ldots, N_m$ in (62).

For the detection problem described above, the optimal additive noise can be represented by a probability distribution with at most two mass points according to Theorem 4. Therefore, the optimal additive noise p.d.f. can be calculated as the solution of the optimization problem in (36) for $M = 2$. In this section, the PSO algorithm is employed to obtain the optimal solution, since it is based on simple iterations with low computational complexity and has been successfully applied to numerous problems in various fields [67]-[70] (please refer to [51]-[53] for detailed descriptions of the PSO algorithm).[8]

Figs. 1, 2 and 3 illustrate the Bayes risks for the noise modified and the original (i.e., in the absence of additive noise) detectors for various values of σ in the cases of equal and unequal priors for $\alpha = 0.08$, $\alpha = 0.12$, $\alpha = 0.4$, respectively, where $A = 1$ is used.[9] From the figures, it is observed that as σ decreases, the improvement obtained via additive noise increases. This is mainly due to the fact that

[8]In the implementation of the PSO algorithm, we employ 50 particles and 1000 iterations. Also, the other parameters are set to $c_1 = c_2 = 2.05$ and $\chi = 0.72984$, and the inertia weight ω is changed from 1.2 to 0.1 linearly with the iteration number. Please refer to [51] for the details of the PSO algorithm and the definitions of the parameters.

[9]Due to the symmetry of the Gaussian mixture noise, the conditional risks in the absence of noise, $F_0(0)$ and $F_1(0)$, are equal. Therefore, the original Bayes risks are the same for both the equal and the unequal priors cases.

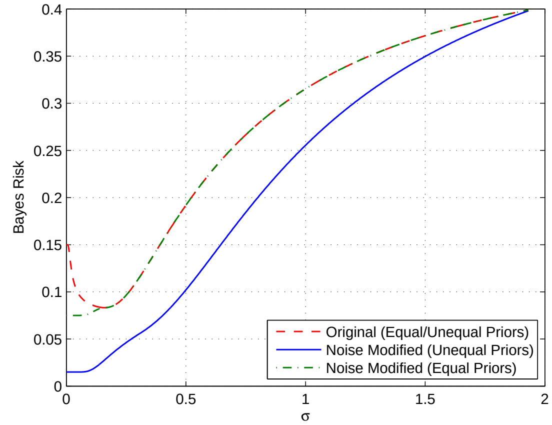


Fig. 3. Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.4$ and $A = 1$.

TABLE I

OPTIMAL ADDITIVE NOISE P.D.F.S FOR VARIOUS VALUES OF σ FOR $\alpha = 0.08$ AND $A = 1$.

	$\pi_0 = 0.5$ / $\pi_0 = 0.9$		
σ	λ	n_1	n_2
0	0.4719 / 0.5333	-0.1057 / -0.2492	0.0901 / 0.0352
0.03	0.4881 / 0.5333	-0.2420 / -0.1995	0.2416 / 0.2982
0.06	0.4858 / 0.5332	-0.2360 / -0.2351	0.2360 / 0.2370
0.09	0.4997 / 0.5251	-0.2189 / -0.2189	0.2189 / 0.2189
0.117	0.5011 / 0.5029	-0.1847 / -0.1847	0.1847 / 0.1847

noise enhancements commonly occur when observations have multimodal p.d.f.s [12], and the multimodal structure is more pronounced for small σ's. In addition, the figures indicate that there is always more improvement in the unequal priors case than that in the equal priors case, which is expected since there is more room for noise enhancement in the unequal priors case due to the asymmetry between the weights of the conditional risks in determining the Bayes risk. Another important point to note from the figures is that the feasible ranges of σ values are different for different values of α. In other words, for each α, the constraint on the maximum conditional risks (cf. (26)) cannot be satisfied after a specific value of σ. This is expected since as σ (which determines the average noise power) exceeds a certain value, it becomes impossible to keep the conditional risks below the given limit α. Therefore, Figs. 1, 2 and 3 are plotted only up to those specific σ values. From the figures, it is observed that those maximum σ values are 0.117, 0.31 and 1.93 for $\alpha = 0.08$, $\alpha = 0.12$ and $\alpha = 0.4$, respectively.

In order to investigate the results in Figs. 1, 2 and 3 further, Tables I, II and III show the optimal additive noise p.d.f.s for various values of σ in the cases of equal and unequal priors for $\alpha = 0.08$, $\alpha = 0.12$ and $\alpha = 0.4$ respectively, where $A = 1$. From Theorem 4, it is known that the optimal additive noise in this example can be represented by a discrete probability distribution with at most two mass points, which can be described as $p_N(x) = \lambda\,\delta(x - n_1) + (1 - \lambda)\,\delta(x - n_2)$. It is observed from the tables that the optimal additive noise p.d.f. has two mass points for certain values of σ, whereas it has a single mass point for other σ's. Also, in the case of equal priors for $\alpha = 0.12$ and $\alpha = 0.4$, the optimal noise p.d.f.s contain only one mass point at the origin for some values of σ, which implies that the detector is nonimprovable in those scenarios. However, there is always improvement for the unequal priors case, which can be also verified from Figs.

TABLE II

OPTIMAL ADDITIVE NOISE P.D.F.S FOR VARIOUS VALUES OF σ FOR $\alpha = 0.12$ AND $A = 1$.

	$\pi_0 = 0.5 / \pi_0 = 0.9$		
σ	λ	n_1	n_2
0	0.2553 / 0.8	-0.2849 / -0.4063	0.0421 / 0.0598
0.08	0.4436 / 0.2028	-0.2266 / 0.2266	0.2266 / -0.2266
0.15	0.7492 / 1	0.0944 / -0.0959	-0.0944 / —
0.23	1 / 1	0 / -0.0693	— / —
0.31	1 / 1	0 / -0.0067	— / —

TABLE III

OPTIMAL ADDITIVE NOISE P.D.F.S FOR VARIOUS VALUES OF σ FOR $\alpha = 0.4$ AND $A = 1$.

	$\pi_0 = 0.5 / \pi_0 = 0.9$		
σ	λ	n_1	n_2
0	0.6518 / 0.1170	-0.3578 / -0.0283	-0.2941 / -0.3879
0.5	1 / 1	0 / -0.3549	— / —
1	1 / 1	0 / -0.2366	— / —
1.5	1 / 1	0 / -0.1131	— / —
1.93	1 / 1	0 / -0.0057	— / —

1, 2 and 3.

Fig. 4 illustrates the Bayes risks for the original and the noise modified detectors for various values of A in the cases of equal and unequal priors for $\alpha = 0.08$ and $\sigma = 0.05$. It is noted that the original conditional risks are above the specified limit $\alpha = 0.08$ for $A < 1.03$.[10] However, after the addition of optimal noise, the noise modified detectors result in conditional risks that are below the limit, which is expected since the optimal noise p.d.f.s are obtained from the solution of the constrained optimization problem in (26). Another observation from Fig. 4 is that, in the equal priors case, the improvement decreases as A increases, and there is no improvement after a certain value of A. However, for the unequal priors case, improvement can be observed over a wider range of A values, which is expected due to the the same reasons argued for Figs. 1-3.

Fig. 5 illustrates the improvement ratio, defined as the ratio of the Bayes risks in the absence and presence of additive noise, versus α for the cases of equal and unequal priors for $\sigma = 0.01$, $\sigma = 0.05$ and $\sigma = 0.1$, where $A = 1$ is used. In the unequal priors case, as α increases, an increase is observed

[10]For the original detector, the conditional risks are equal; hence, $\mathrm{R}_0^x(\phi) = \mathrm{R}_1^x(\phi) = r^x(\phi)$.

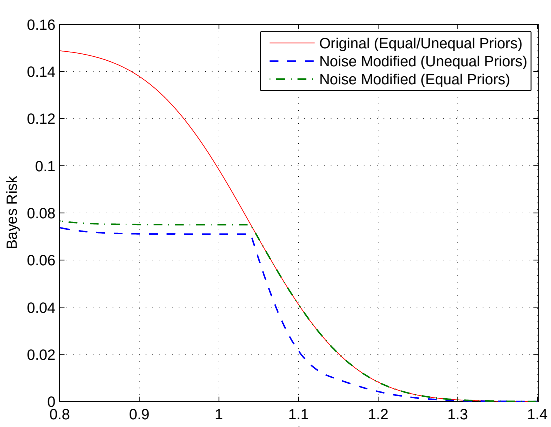


Fig. 4. Bayes risks of original and noise modified detectors versus A in cases of equal priors and unequal priors for $\alpha = 0.08$ and $\sigma = 0.05$.

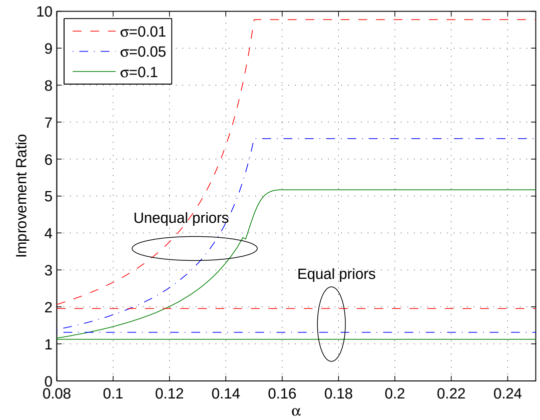


Fig. 5. Improvement ratio versus α in the cases of equal priors and unequal priors for $\sigma = 0.01$, $\sigma = 0.05$ and $\sigma = 0.1$, where $A = 1$.

in the improvement ratio up to a certain value of α, and then the improvement ratio becomes constant. Those critical α values specify the boundaries between the restricted Bayes and the (unrestricted) Bayes criteria. When α gets larger than those values, the constraint in (26) is no longer active; hence, the problem reduces to the Bayesian framework. Therefore, further increases in α do not cause any additional performance improvements. Similarly, as the value of α decreases, the restricted Bayes criterion converges to the minimax criterion [29]. The restricted Bayes criterion achieves its minimum improvement ratio when it becomes equivalent to the minimax criterion, and achieves its maximum improvement ratio when it is equal to the Bayes criterion. In the case of equal priors, the improvement ratio is constant with respect to α, meaning that the improvement for the minimax criterion equals to that for the Bayes criterion. Another observation from the figure is that an increase in σ reduces the improvement ratio, and for the same values of σ, there is more improvement in the unequal priors case. Finally, it should be noted that various values of α in Fig. 5 correspond to different amounts of uncertainty in the prior information [42]. As the prior information gets more accurate, a larger value of α is selected; hence, the constraint on the conditional risks becomes less strict, meaning that the restricted Bayes criterion converges to the Bayes criterion after a certain value of α. On the other hand, as the amount of uncertainty increases, a smaller value of α is selected, and the restricted Bayes criterion converges to the minimax criterion when α becomes equal to the minimax risk [40], [42].

Next, the improvability conditions in Theorem 5 are investigated for the detection example. To that aim, it is first observed that the original conditional risks $F_0(0)$ and $F_1(0)$ are equal to each other for any value of σ due to the symmetry of the Gaussian mixture noise (cf. (64)). Therefore, $F(0) = \pi_0 F_0(0) + \pi_1 F_1(0) = F_0(0) = F_1(0)$. In addition, suppose that the limit on the conditional risks, α, is set to the original conditional risks for each value of σ, which implies that $\mathcal{S}_\alpha = \{0, 1\}$ in (52). Also, the first order derivatives of $F_0(x)$ and $F_1(x)$ at $x = 0$ can be calculated from (64) as

$$F_0^{'}(0) = -F_1^{'}(0) = \sum_{i=1}^{N_m} \frac{w_i}{\sqrt{2\pi}\,\sigma_i} \exp\left(-\frac{(A/2-\mu_i)^2}{2\sigma_i^2}\right). \tag{65}$$

Similarly, the second order derivatives of $F_0(x)$ and $F_1(x)$ at

$x = 0$ are obtained as

$$F_0^{''}(0) = F_1^{''}(0) = \sum_{i=1}^{N_m} \frac{w_i(A/2 - \mu_i)}{\sqrt{2\pi}\,\sigma_i^3} \exp\left(-\frac{(A/2-\mu_i)^2}{2\sigma_i^2}\right). \quad (66)$$

For the unequal priors case, the first and second order derivatives of $F(x) = \pi_0 F_0(x) + \pi_1 F_1(x)$ at $x = 0$ can be expressed as $F^{'}(0) = 0.8F_0^{'}(0)$ and $F^{''}(0) = F_0^{''}(0)$. From (65), it is noted that $F_0^{'}(0) > 0$ and $F_1^{'}(0) < 0$; hence, $F^{'}(0) > 0$ as well. Then, from (48)-(51), set $\mathcal{F}_n$ in (54) can be expressed, at $x = 0$, as

$$\begin{aligned} \mathcal{F}_1 &= \{0.8zF_0^{'}(0), zF_0^{'}(0), -zF_0^{'}(0)\}\ , \\ \mathcal{F}_2 &= \{z^2F_0^{''}(0), z^2F_0^{''}(0), z^2F_0^{''}(0)\}\ . \end{aligned} \quad (67)$$

Therefore, (55)-(57) imply that, at $x = 0$, $\mathcal{S}_z = \emptyset$, $\mathcal{S}_n = \{3\}$ and $\mathcal{S}_p = \{1, 2\}$ for $z > 0$ and $\mathcal{S}_z = \emptyset$, $\mathcal{S}_n = \{1, 2\}$, and $\mathcal{S}_p = \{3\}$ for $z < 0$.[11] Since $\mathcal{S}_z = \emptyset$, the first condition in Theorem 5 is automatically satisfied. For $z > 0$, $|\mathcal{S}_n| = 1$ and $|\mathcal{S}_p| = 2$; hence, the third bullet of the second condition implies that

$$\begin{aligned} \min\{\mathcal{F}_2(1)\mathcal{F}_1(2)\mathcal{F}_1(3)\,, \mathcal{F}_2(2)\mathcal{F}_1(1)\mathcal{F}_1(3)\} \\ > \mathcal{F}_2(3)\mathcal{F}_1(1)\mathcal{F}_1(2) \end{aligned} \quad (68)$$

is required for improvability. For $z < 0$, $|\mathcal{S}_n| = 2$ and $|\mathcal{S}_p| = 1$; hence, the second bullet of the second condition becomes active, which can be shown to yield the same condition as in (68). From (67), the improvability condition in (68) can be expressed more explicitly as

$$\begin{aligned} \min\left\{-z^4F_0^{''}(0)\left(F_0^{'}(0)\right)^2\,,\,-0.8z^4F_0^{''}(0)\left(F_0^{'}(0)\right)^2\right\} \\ > 0.8z^4F_0^{''}(0)\left(F_0^{'}(0)\right)^2\ , \end{aligned} \quad (69)$$

which is satisfied when $F_0^{''}(0) < 0$. Therefore, the detector is improvable whenever the expression in (66) is negative. For the equal priors case, $\mathcal{F}_1$ and $\mathcal{F}_2$ in (67) become $\mathcal{F}_1 = \{0, zF_0^{'}(0), -zF_0^{'}(0)\}$ and $\mathcal{F}_2 = \{z^2F_0^{''}(0), z^2F_0^{''}(0), z^2F_0^{''}(0)\}$, respectively. Therefore, the first improvability condition in Theorem 5 requires that $F_0^{''}(0) < 0$, whereas the third bullet of the second condition requires that $\mathcal{F}_2(2)\mathcal{F}_1(3) > \mathcal{F}_2(3)\mathcal{F}_1(2)$ for $z > 0$ and $\mathcal{F}_2(3)\mathcal{F}_1(2) > \mathcal{F}_2(2)\mathcal{F}_1(3)$ for $z < 0$. However, it can be shown that the conditions in the third bullet are always satisfied when $F_0^{''}(0) < 0$. Therefore, the same improvability condition is obtained for the equal priors case, as well. Fig. 6 illustrates $F_0^{''}(0)$ versus σ for various values of A, where σ represents the standard deviation of the Gaussian mixture noise components ($\sigma_i = \sigma$, $\forall i$ in (62)). It is observed that the detector performance can be improved for $A = 1$ if $\sigma \in [0.005, 0.1597]$, for $A = 0.9$ if $\sigma \in [0.01, 0.1686]$, and for $A = 0.8$ if $\sigma \in [0.02, 0.161]$. On the other hand, the calculations show that the detector is actually improvable for $A = 1$ if $\sigma \leq 0.16$, for $A = 0.9$ if $\sigma \leq 0.17$, and for $A = 0.8$ if $\sigma \leq 0.161$. Hence, the results reveal that the proposed improvability conditions are sufficient but not necessary, and that they are quite effective in determining the range of parameters for which the detector performance can be improved.[12]

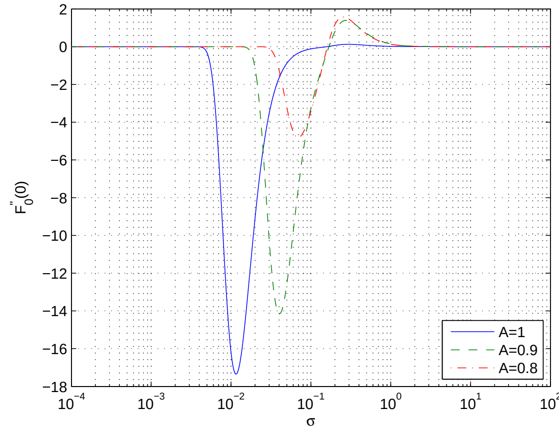


Fig. 6. The second order derivative of $F_0(x)$ at $x = 0$ versus σ for various values of A. Both Theorem 5 and Theorem 3 imply for the detection example in this section that the detector is improvable whenever $F_0^{''}(0)$ is negative. The limit on the conditional risks, α, is set to the original conditional risks for each value of σ. The graph for $A = 1$ is scaled by 0.1 to make view of the figure more convenient (since only the signs of the graphs are important).

Next, the improvability conditions based on Theorem 3 are considered. For the binary hypothesis-testing example in this section, $H(t)$ in (23) becomes $H(t) = \inf\{\pi_0F_0(n) + \pi_1F_1(n) \mid \max\{F_0(n), F_1(n)\} = t\ ,\ n \in \mathbb{R}\}$. From (64), it can be shown that $F_0(n)$ and $F_1(n)$ are monotone increasing and decreasing functions, respectively. In addition, due to the symmetry of the Gaussian mixture noise, $F_1(n) = F_0(-n)$, $\forall n$. Therefore, without loss of generality, $H(t)$ can be expressed as $H(t) = \pi_0 t + \pi_1 F_1\left(F_0^{-1}(t)\right)$. Then, the second derivative of $H(t)$ can be obtained as

$$H^{''}(t) = \pi_1 \frac{F_0^{''}(F_0^{-1}(t)) - F_1^{'}(F_0^{-1}(t))F_0^{''}(F_0^{-1}(t))/F_0^{'}(F_0^{-1}(t))}{\left(F_0^{'}(F_0^{-1}(t))\right)^2}\ . \quad (70)$$

In order to evaluate the condition in Theorem 5, it is first observed that $t = \tilde{\alpha} = \max\{F_0(0), F_1(0)\} = F_0(0)$, since $F_0(0) = F_1(0)$ (cf. (64)). Then, $H^{''}(\tilde{\alpha}) < 0$ implies that $F_1^{''}(0) - F_1^{'}(0)F_0^{''}(0)/F_0^{'}(0) < 0$ for any π_1. Since $F_0^{''}(0) = F_1^{''}(0)$ from (66), and $F_0^{'}(0) > 0$ and $F_1^{'}(0) < 0$ from (65), that improvability condition reduces to $F_0^{''}(0) < 0$, which is the same condition obtained from Theorem 5. Therefore, for this specific example, the improvability conditions in Theorem 3 and Theorem 5 are equivalent (cf. Fig. 6). However, it should be noted that the two conditions are not equivalent in general, and the calculation of $H(t)$ can be difficult in the absence of monotonicity properties related to F_0 and F_1.

Finally, another example is studied in order to investigate the theoretical results on a 4-ary hypothesis-testing problem in the presence of observation noise that is a mixture of non-Gaussian components. The hypotheses $\mathcal{H}_0$, $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$

[11] Note that $\mathcal{S}_z = \{1, 2, 3\}$ for $z = 0$, in which case the first condition in Theorem 5 cannot satisfied since $\mathcal{F}_2 = \{0, 0, 0\}$. Therefore, $z = 0$ is not considered in obtaining improvability conditions.

[12] In fact, $F_0^{''}(0)$ can be shown to be negative even for smaller σ values than specified above; however, very small negative values are computed as zero due to the accuracy limitations.

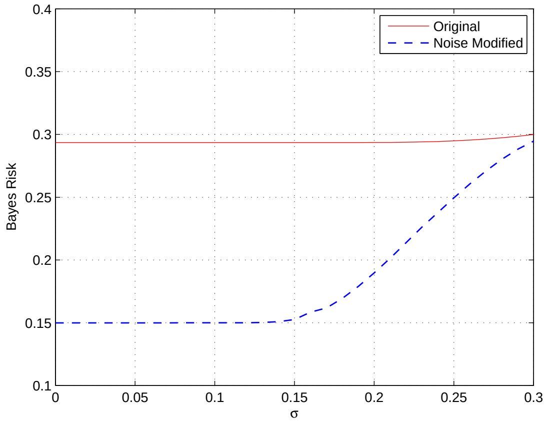


Fig. 7. Bayes risks of original and noise modified detectors versus σ for $\alpha = 0.4$ and $A = 1$.

are defined as

$$\mathcal{H}_0 \;:\; x = -3\sqrt{A} + v \;, \quad \mathcal{H}_1 \;:\; x = -\sqrt{A} + v \;,$$
$$\mathcal{H}_2 \;:\; x = \sqrt{A} + v \;, \qquad \mathcal{H}_3 \;:\; x = 3\sqrt{A} + v \;, \qquad (71)$$

where $x \in \mathbb{R}$, $A > 0$ is a known scalar value, and v is zero-mean observation noise that is a mixture of Rayleigh distributed components; that is, $p_V(x) = \sum_{i=1}^{N_m} w_i\, \psi_i(x - \mu_i)$, where $w_i \geq 0$ for $i = 1, \ldots, N_m$, $\sum_{i=1}^{N_m} w_i = 1$, and

$$\psi_i(x) = \begin{cases} \frac{x}{\sigma_i^2} \exp\left(\frac{-x^2}{2\sigma_i^2}\right) \;, & x \geq 0 \\ 0 \;, & x < 0 \end{cases} \;, \qquad (72)$$

for $i = 1, \ldots, N_m$. In the numerical results, the same variance is considered for all the Rayleigh components, meaning that $\sigma_i = \sigma$, $\forall i$. In addition, the parameters are selected as $N_m = 4$, $\mu_1 = 0.2$, $\mu_2 = 1$, $\mu_3 = -2\sigma\sqrt{\frac{\pi}{2}} - 0.2$, $\mu_4 = -2\sigma\sqrt{\frac{\pi}{2}} - 1$, $w_1 = w_3 = 0.3$ and $w_2 = w_4 = 0.2$.[13] In addition, the detector is described by

$$\phi(y) = \begin{cases} 0 \;, & y \leq -2\sqrt{A} \\ 1 \;, & -2\sqrt{A} < y \leq 0 \\ 2 \;, & 0 < y \leq 2\sqrt{A} \\ 3 \;, & 2\sqrt{A} < y \end{cases} \;, \qquad (73)$$

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

For equal prior probabilities and UCA, Fig. 7 illustrates the Bayes risk versus σ when $A = 1$ and $\alpha = 0.4$. It is observed that the additive noise can significantly improve the detector performance (equivalently, it reduces the average probability of error of a communications system) for small values of σ. In addition, for the scenario in Fig. 7, Table IV illustrates the optimal additive noise p.d.f.s for various values of σ. In accordance with Theorem 4, the optimal noise can have up to four non-zero mass points in this problem. Furthermore, for $\sigma = 0.05$, Fig. 8 plots the Bayes risk versus A for the original and noise modified detectors. A significant improvement is observed for $A \in [0.5, 1]$.

[13]It should be noted that the dependence of the means on σ is necessary in order to keep the noise zero-mean, since the Rayleigh distribution is specified by a single parameter, σ.

TABLE IV

OPTIMAL ADDITIVE NOISE P.D.F.S FOR VARIOUS VALUES OF σ FOR $\alpha = 0.4$ AND $A = 1$.

σ	λ_1	λ_2	λ_3	λ_4
0.05	0.1654	0.1218	0.3552	0.3576
0.15	0.2232	0.7768	0	0
0.25	1	0	0	0
σ	n_1	n_2	n_3	n_4
0.05	-0.4916	0.2175	0.2652	-0.5331
0.15	-0.4288	0.3661	—	—
0.25	-0.2819	—	—	—

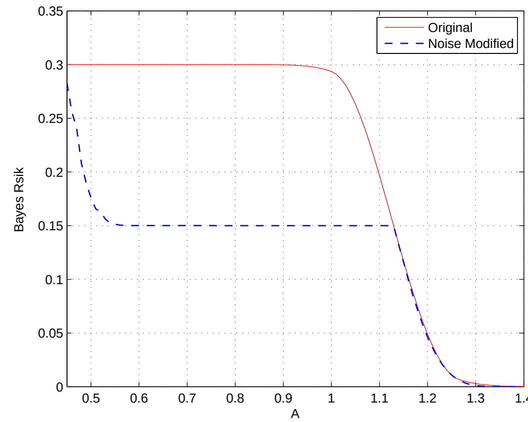


Fig. 8. Bayes risks of original and noise modified detectors versus A for $\alpha = 0.4$ and $\sigma = 0.05$.

V. Concluding Remarks

In this paper, noise enhanced hypothesis-testing has been studied in the restricted Bayesian framework. First, the most generic formulation of the problem has been considered based on M-ary composite hypothesis-testing, and sufficient conditions for improvability and nonimprovability of detection via additive independent noise have been derived. In addition, an approximate formulation of the optimal noise p.d.f. has been presented. Then, simple hypothesis-testing problems have been studied and additional improvability conditions that are specific to simple hypotheses have been obtained. Also, the optimal noise p.d.f. has been shown to include at most M mass points for M-ary simple hypothesis-testing problems under certain conditions. Then, various approaches to solving for the optimal noise p.d.f. have been considered, including global optimization techniques, such as the PSO, and a convex relaxation technique. Finally, two detection examples have been studied to illustrate the practicality of the theoretical results.

Appendix

A. Proof of Theorem 2

A detector is improvable if there exists a noise p.d.f. $p_{\mathbf{N}}(\mathbf{n})$ that satisfies $\mathrm{E}\{F(\mathbf{N})\} < F(\mathbf{0})$ and $\max_{\theta\in\Lambda} \mathrm{E}\{F_\theta(\mathbf{N})\} \leq \alpha$, which can be expressed as $\int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n})\, F(\mathbf{n})\, d\mathbf{n} < F(\mathbf{0})$ and $\int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n})\, F_\theta(\mathbf{n})\, d\mathbf{n} \leq \alpha$, $\forall\theta \in \Lambda$. For a noise p.d.f. having L infinitesimally small noise components, $p_{\mathbf{N}}(\mathbf{n}) = \sum_{j=1}^{L} \lambda_j\, \delta(\mathbf{n} - \boldsymbol{\epsilon}_j)$, these conditions become

$$\sum_{j=1}^{L} \lambda_j\, F(\boldsymbol{\epsilon}_j) < F(\mathbf{0}) \;, \quad \sum_{j=1}^{L} \lambda_j\, F_\theta(\boldsymbol{\epsilon}_j) \leq \alpha \;, \; \forall\theta \in \Lambda. \quad (74)$$

Since the $\boldsymbol{\epsilon}_j$'s are infinitesimally small, $F(\boldsymbol{\epsilon}_j)$ and $F_\theta(\boldsymbol{\epsilon}_j)$ can be approximated by using the Taylor series expansion as $F(\mathbf{0}) + \boldsymbol{\epsilon}_j^T\mathbf{f} + 0.5\,\boldsymbol{\epsilon}_j^T\mathbf{H}\boldsymbol{\epsilon}_j$ and $F_\theta(\mathbf{0}) + \boldsymbol{\epsilon}_j^T\mathbf{f}_\theta + 0.5\,\boldsymbol{\epsilon}_j^T\mathbf{H}_\theta\boldsymbol{\epsilon}_j$ respectively, where $\mathbf{H}$ and $\mathbf{f}$ ($\mathbf{H}_\theta$ and $\mathbf{f}_\theta$) are the Hessian and the gradient of $F(\mathbf{x})$ ($F_\theta(\mathbf{x})$) at $\mathbf{x}=\mathbf{0}$, respectively. Therefore, (74) requires that

$$\sum_{j=1}^{L}\lambda_j\,\boldsymbol{\epsilon}_j^T\mathbf{H}\boldsymbol{\epsilon}_j + 2\sum_{j=1}^{L}\lambda_j\,\boldsymbol{\epsilon}_j^T\mathbf{f} < 0\ ,$$
$$\sum_{j=1}^{L}\lambda_j\,\boldsymbol{\epsilon}_j^T\mathbf{H}_\theta\boldsymbol{\epsilon}_j + 2\sum_{j=1}^{L}\lambda_j\,\boldsymbol{\epsilon}_j^T\mathbf{f}_\theta \le 2\,(\alpha - F_\theta(\mathbf{0}))\ \ ,\ \forall\theta\in\Lambda\ . \tag{75}$$

Let $\boldsymbol{\epsilon}_j = \rho_j\,\mathbf{z}$ for $j=1,2,\ldots,L$, where ρ_j for $j=1,2,\ldots,L$ are infinitesimally small real numbers, and $\mathbf{z}$ is a K-dimensional real vector. Then, based on the function definitions in (19)-(22), the conditions in (75) can be simplified, after some manipulation, as

$$\left(f^{(2)}(\mathbf{x},\mathbf{z}) + c\,f^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ , \tag{76}$$
$$\left(f_\theta^{(2)}(\mathbf{x},\mathbf{z}) + c\,f_\theta^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < \frac{2\,(\alpha - F_\theta(\mathbf{0}))}{\sum_{j=1}^{L}\lambda_j\,\rho_j^2}\ ,\forall\theta\in\Lambda, \tag{77}$$

where $c \triangleq 2\sum_{j=1}^{L}\lambda_j\,\rho_j \big/ \sum_{j=1}^{L}\lambda_j\,\rho_j^2$.

Since $\alpha = F_{\theta^*}(\mathbf{0})$ and $\alpha > \max_{\theta\in\Lambda\setminus\theta^*} F_\theta(\mathbf{0})$, the right-hand-side of (77) goes to infinity for $\theta\neq\theta^*$. Hence, we should consider only the $\theta=\theta^*$ case. Thus, (76) and (77) can be expressed as

$$\left(f^{(2)}(\mathbf{x},\mathbf{z}) + c\,f^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ , \tag{78}$$
$$\left(f_{\theta^*}^{(2)}(\mathbf{x},\mathbf{z}) + c\,f_{\theta^*}^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ . \tag{79}$$

It is noted that c can take any real value by definition via selection of appropriate λ_i and infinitesimally small ρ_i values for $i=1,2,\ldots,L$. Therefore, for the first part of the theorem, under the condition of $f_{\theta^*}^{(1)}(\mathbf{x},\mathbf{z})f^{(1)}(\mathbf{x},\mathbf{z}) > 0$ at $\mathbf{x}=\mathbf{0}$, which states that the second term in (78) has the same sign as the second term in (79), there always exists c that satisfies the improvability conditions in (78) and (79). For the second part of the theorem, since $f^{(1)}(\mathbf{x},\mathbf{z}) > 0$ and $f_{\theta^*}^{(1)}(\mathbf{x},\mathbf{z}) < 0$ at $\mathbf{x}=\mathbf{0}$, (78) and (79) can also be expressed as

$$\left(f^{(2)}(\mathbf{x},\mathbf{z})f_{\theta^*}^{(1)}(\mathbf{x},\mathbf{z}) + c\,f^{(1)}(\mathbf{x},\mathbf{z})f_{\theta^*}^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} > 0\ , \tag{80}$$
$$\left(f_{\theta^*}^{(2)}(\mathbf{x},\mathbf{z})f^{(1)}(\mathbf{x},\mathbf{z}) + c\,f_{\theta^*}^{(1)}(\mathbf{x},\mathbf{z})f^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ . \tag{81}$$

Under the condition of $f^{(2)}(\mathbf{x},\mathbf{z})f_{\theta^*}^{(1)}(\mathbf{x},\mathbf{z}) > f_{\theta^*}^{(2)}(\mathbf{x},\mathbf{z})f^{(1)}(\mathbf{x},\mathbf{z})$ at $\mathbf{x}=\mathbf{0}$, which states that the first term in (80) is larger than the first term in (81), there always exists c that satisfies the improvability conditions in (80) and (81).

B. Proof of Theorem 3

Since $H^{''}(\tilde{\alpha}) < 0$ and $H(t)$ in (23) is second-order continuously differentiable around $t=\tilde{\alpha}$, there exist $\epsilon>0$, $\mathbf{n}_1$ and $\mathbf{n}_2$ such that $\max_{\theta\in\Lambda} F_\theta(\mathbf{n}_1) = \tilde{\alpha}+\epsilon$ and $\max_{\theta\in\Lambda} F_\theta(\mathbf{n}_2) = \tilde{\alpha}-\epsilon$. Then, it is proven in the following that an additive noise component with $p_{\mathbf{N}}(\mathbf{n}) = 0.5\,\delta(\mathbf{x}-\mathbf{n}_1) + 0.5\,\delta(\mathbf{x}-\mathbf{n}_2)$ improves the detector performance under the conditional risk constraint. First, the maximum value of the conditional risks in the presence of additive noise is shown not to exceed α:

$$\max_{\theta\in\Lambda}\mathrm{E}\{F_\theta(\mathbf{N})\} \le \mathrm{E}\left\{\max_{\theta\in\Lambda}F_\theta(\mathbf{N})\right\}$$
$$= 0.5(\tilde{\alpha}+\epsilon) + 0.5(\tilde{\alpha}-\epsilon) = \tilde{\alpha} \le \alpha \tag{82}$$

Then, the decrease in the Bayes risk is proven as follows. Due to the assumptions in the theorem, $H(t)$ is concave in an interval around $t=\tilde{\alpha}$. Since $\mathrm{E}\{F(\mathbf{N})\}$ can attain the value of $0.5\,H(\tilde{\alpha}+\epsilon) + 0.5\,H(\tilde{\alpha}-\epsilon)$, which is always smaller than $H(\tilde{\alpha})$ due to concavity, it is concluded that $\mathrm{E}\{F(\mathbf{N})\} < H(\tilde{\alpha})$. As $H(\tilde{\alpha}) \le F(\mathbf{0})$ by definition of $H(t)$ in (23), $\mathrm{E}\{F(\mathbf{N})\} < F(\mathbf{0})$ is satisfied; hence, the detector is improvable.

C. Maximum Conditional Risk Achieved by Optimal Noise

Consider the case in which $t_{\mathrm{m}} = \arg\min_t H(t) > \alpha$. In order to prove that "$\max_{\theta\in\Lambda}\mathrm{R}_\theta^{\mathbf{y}}(\phi) = \alpha$ for the optimal noise" by contradiction, first assume that the optimal solution of (12) is given by $p_{\tilde{\mathbf{N}}}(\mathbf{x})$ with $\beta \triangleq \max_{\theta\in\Lambda}\mathrm{R}_\theta^{\tilde{\mathbf{y}}}(\phi) < \alpha$. As in the proof of Theorem 4 in [12], we define another noise $\mathbf{N}$ with the following p.d.f.:

$$p_{\mathbf{N}}(\mathbf{n}) = \frac{\alpha-\beta}{t_{\mathrm{m}}-\beta}\,\delta(\mathbf{n}-\mathbf{n}_{\mathrm{m}}) + \frac{t_{\mathrm{m}}-\alpha}{t_{\mathrm{m}}-\beta}\,p_{\tilde{\mathbf{N}}}(\mathbf{n})\ , \tag{83}$$

where $\mathbf{n}_{\mathrm{m}}$ is the noise component that results in the minimum Bayes risk; that is, $F(\mathbf{n}_{\mathrm{m}}) = F_{\min}$, and t_{m} is the maximum value of the conditional risks when noise $\mathbf{n}_{\mathrm{m}}$ is employed; that is, $t_{\mathrm{m}} = \max_{\theta\in\Lambda} F_\theta(\mathbf{n}_{\mathrm{m}})$.

For the noise p.d.f. in (83), the Bayes risk and conditional risks can be calculated as

$$r^{\mathbf{y}}(\phi) = \mathrm{E}\{F(\mathbf{N})\} = \frac{\alpha-\beta}{t_{\mathrm{m}}-\beta}\,F(\mathbf{n}_{\mathrm{m}}) + \frac{t_{\mathrm{m}}-\alpha}{t_{\mathrm{m}}-\beta}\,r^{\tilde{\mathbf{y}}}(\phi)\ , \tag{84}$$

$$\mathrm{R}_\theta^{\mathbf{y}}(\phi) = \mathrm{E}\{F_\theta(\mathbf{N})\} = \frac{\alpha-\beta}{t_{\mathrm{m}}-\beta}\,F_\theta(\mathbf{n}_{\mathrm{m}}) + \frac{t_{\mathrm{m}}-\alpha}{t_{\mathrm{m}}-\beta}\,\mathrm{R}_\theta^{\tilde{\mathbf{y}}}(\phi)\ , \tag{85}$$

for all $\theta\in\Lambda$. Since $F(\mathbf{n}_{\mathrm{m}}) < r^{\tilde{\mathbf{y}}}(\phi)$, (84) implies $r^{\mathbf{y}}(\phi) < r^{\tilde{\mathbf{y}}}(\phi)$. On the other hand, as $F_\theta(\mathbf{n}_{\mathrm{m}}) \le t_{\mathrm{m}}$ and $\mathrm{R}_\theta^{\tilde{\mathbf{y}}}(\phi) \le \beta$, $\mathrm{R}_\theta^{\mathbf{y}}(\phi) \le \alpha$ is obtained. Therefore, $\tilde{\mathbf{N}}$ cannot be an optimal solution, which implies a contradiction. In other words, any noise p.d.f. that satisfies $\max_{\theta\in\Lambda}\mathrm{R}_\theta^{\mathbf{y}}(\phi) < \alpha$ cannot be optimal.

D. Proof of Theorem 5

Theorem 4 states that the optimal additive noise can be represented by a discrete probability distribution with at most M mass points. Therefore, a detector is improvable if there exists a noise p.d.f. $p_{\mathbf{N}}(\mathbf{n}) = \sum_{l=1}^{M}\lambda_l\,\delta(\mathbf{n}-\mathbf{n}_l)$ that satisfies $\mathrm{E}\{F(\mathbf{N})\} < F(\mathbf{0})$ and $\max_{i\in\{0,1,\ldots,M-1\}}\mathrm{E}\{F_i(\mathbf{N})\} \le \alpha$, which can be expressed as

$$\sum_{l=1}^{M}\lambda_l\,F(\mathbf{n}_l) < F(\mathbf{0})\ ,\ \max_{i\in\{0,1,\ldots,M-1\}}\sum_{l=1}^{M}\lambda_l\,F_i(\mathbf{n}_l) \le \alpha. \tag{86}$$

As in the proof of Theorem 2 in Appendix A, consider the improvability conditions in (86) with infinitesimally small noise components, $\mathbf{n}_l = \boldsymbol{\epsilon}_l = \rho_l\,\mathbf{z}$ for $l = 1, 2, \ldots, M$, where ρ_l's are infinitesimally small real numbers, and $\mathbf{z}$ is a K-dimensional real vector. Then, similar manipulations to those in Appendix A (cf. (75)-(77)) can be performed to obtain

$$\left(f^{(2)}(\mathbf{x},\mathbf{z}) + c\, f^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ , \tag{87}$$

$$\left(f_i^{(2)}(\mathbf{x},\mathbf{z}) + c\, f_i^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < \frac{2\,(\alpha - F_i(\mathbf{0}))}{\sum_{j=1}^{M}\lambda_j\,\rho_j^2}\ , \tag{88}$$

for $i = 0, 1, \ldots, M-1$, where $c \triangleq 2\sum_{j=1}^{M}\lambda_j\,\rho_j \big/ \sum_{j=1}^{M}\lambda_j\,\rho_j^2$.

Since $F_i(0) < \alpha$, $\forall i \in \bar{\mathcal{S}}_\alpha$, the right-hand-side of (88) goes to infinity for $i \in \bar{\mathcal{S}}_\alpha$. Hence, one can consider $i \in \mathcal{S}_\alpha$ only. Thus, (87) and (88) can be expressed as

$$\left(f^{(2)}(\mathbf{x},\mathbf{z}) + c\, f^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ , \tag{89}$$

$$\left(f_i^{(2)}(\mathbf{x},\mathbf{z}) + c\, f_i^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ ,\ \forall i \in \mathcal{S}_\alpha\ . \tag{90}$$

Based on the definition in (54), (89) and (90) can be restated as

$$\Big(\mathcal{F}_2(j) + c\,\mathcal{F}_1(j)\Big)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ \text{ for }\ j = 1, 2, \ldots, |\mathcal{S}_\alpha| + 1\ . \tag{91}$$

It is noted that c can take any real value by selecting appropriate λ_i and infinitesimally small ρ_i values for $i = 0, 1, \ldots, M-1$. From (55), it is concluded that in order for the conditions in (91) to hold,

$$\mathcal{F}_2(j)\,\big|_{\mathbf{x}=\mathbf{0}} < 0 \tag{92}$$

must be satisfied $\forall j \in \mathcal{S}_z$, which is the first condition in Theorem 5.

In addition to (92), one of the following conditions should be satisfied for the improvability conditions in (91) to hold:

- When $|\mathcal{S}_n| = 0$ or $|\mathcal{S}_p| = 0$, as stated in the first part of the second condition in Theorem 5, all the second terms in (91) (namely, $\mathcal{F}_1(1), \ldots, \mathcal{F}_1(|\mathcal{S}_\alpha| + 1)$) are either all non-negative or all non-positive. Therefore, there always exists a c that satisfies the improvability conditions in (91) when the first condition in Theorem 5 (cf. (92)) is satisfied.
- When $|\mathcal{S}_n|$ is a positive even number and $|\mathcal{S}_p| > 0$, (91) can be expressed, after some manipulation, as

$$\mathcal{F}_2(j)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ ,\ \text{for all } j \in \mathcal{S}_z\ , \tag{93}$$

$$\Big(\mathcal{F}_2(j)\prod_{l\in\mathcal{S}_p\cup\mathcal{S}_n\setminus\{j\}}\mathcal{F}_1(l) + c\prod_{l\in\mathcal{S}_p\cup\mathcal{S}_n}\mathcal{F}_1(l)\Big)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ , \tag{94}$$

  for all $j \in \mathcal{S}_p$, and

$$\Big(\mathcal{F}_2(j)\prod_{l\in\mathcal{S}_p\cup\mathcal{S}_n\setminus\{j\}}\mathcal{F}_1(l) + c\prod_{l\in\mathcal{S}_p\cup\mathcal{S}_n}\mathcal{F}_1(l)\Big)\Big|_{\mathbf{x}=\mathbf{0}} > 0\ , \tag{95}$$

  for all $j \in \mathcal{S}_n$. Note that (94) and (95) are obtained by multiplying (91) by $\prod_{l\in\mathcal{S}_p\cup\mathcal{S}_n\setminus\{j\}}\mathcal{F}_1(l)$, which is a positive (negative) quantity when $j \in \mathcal{S}_p$ ($j \in \mathcal{S}_n$) since $|\mathcal{S}_n|$ is even. The condition in (93) is satisfied due to the first condition in Theorem 5. In addition, under the condition in (58), there always exists a c that satisfies the improvability conditions in (94) and (95).
- When $|\mathcal{S}_n|$ is an odd number and $|\mathcal{S}_p| > 0$, (91) can be expressed by three conditions as in (93)-(95) with the only difference being that the signs of the inequalities in (94) and (95) are switched. In that case, the first condition (cf. (93)) is satisfied due to the first condition in Theorem 5. Also, under the condition in (59), there always exists a c that satisfies the second and third conditions.

E. Proof of Corollary 1

Consider the proof of Theorem 5 above. Since $\alpha > \max\limits_{i\in\{0,1,\ldots,M-1\}} F_i(\mathbf{0})$, the right-hand-side of (88) becomes infinity for any i. Therefore, we can consider the condition in (87) only; that is,

$$\left(f^{(2)}(\mathbf{x},\mathbf{z}) + c\, f^{(1)}(\mathbf{x},\mathbf{z})\right)\Big|_{\mathbf{x}=\mathbf{0}} < 0\ . \tag{96}$$

In terms of the gradient $\mathbf{f}$ and the Hessian $\mathbf{H}$ of $F(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$, (96) becomes $\mathbf{z}^T\mathbf{H}\mathbf{z} + c\,\mathbf{z}^T\mathbf{f} < 0$. Since c can take any real value by definition (cf. Appendix D) and $\mathbf{z}$ can be chosen arbitrarily small, the improvability condition can always be satisfied if $\mathbf{f} \neq \mathbf{0}$. On the other hand, if $\mathbf{f} = \mathbf{0}$, then the improvability condition becomes $\mathbf{z}^T\mathbf{H}\mathbf{z} < 0$. If $F(\mathbf{x})$ is not convex around $\mathbf{x} = \mathbf{0}$, $\mathbf{H}$ is not positive semidefinite. Therefore, there exists $\mathbf{z}$ such that $\mathbf{z}^T\mathbf{H}\mathbf{z} < 0$ is satisfied; hence, the detector is improvable.

References

[1] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys. A: Math. General*, vol. 14, pp. 453–457, 1981.

[2] P. Makra and Z. Gingl, "Signal-to-noise ratio gain in non-dynamical and dynamical bistable stochastic resonators," *Fluctuat. Noise Lett.*, vol. 2, no. 3, pp. L145–L153, 2002.

[3] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni, "Stochastic resonance," *Rev. Mod. Phys.*, vol. 70, pp. 223–287, Jan. 1998.

[4] G. P. Harmer, B. R. Davis, and D. Abbott, "A review of stochastic resonance: Circuits and measurement," *IEEE Trans. Instrum. Meas*, vol. 51, pp. 299–309, Apr. 2002.

[5] 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.

[6] I. Goychuk and P. Hanggi, "Stochastic resonance in ion channels characterized by information theory," *Phys. Rev. E*, vol. 61, pp. 4272–4280, Apr. 2000.

[7] S. Mitaim and B. Kosko, "Adaptive stochastic resonance in noisy neurons based on mutual information," *IEEE Trans. Neural Netw.*, vol. 15, pp. 1526–1540, Nov. 2004.

[8] N. G. Stocks, "Suprathreshold stochastic resonance in multilevel threshold systems," *Phys. Rev. Lett.*, vol. 84, pp. 2310–2313, Mar. 2000.

[9] 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.

[10] B. Kosko and S. Mitaim, "Stochastic resonance in noisy threshold neurons," *Neural Netw.*, vol. 16, pp. 755–761, 2003.

[11] B. Kosko and S. Mitaim, "Robust stochastic resonance for simple threshold neurons," *Phys. Rev. E*, vol. 70, no. 031911, 2004.

[12] 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. Sig. Processing*, vol. 55, pp. 3172–3184, July 2007.

[13] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection," *IEEE Trans. Sig. Processing*, vol. 57, pp. 1655–1669, May 2009.

[14] M. D. McDonnell and D. Abbott, "What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology," *PLoS Computational Biology*, vol. 5, May 2009.

[15] A. R. Bulsara and A. Zador, "Threshold detection of wideband signals: A noise-induced maximum in the mutual information," *Physical Review E*, vol. 54, pp. R2185–R2188, Sep. 1996.

[16] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, "Information-theoretic stochastic resonance in noise-floor limited systems: The case for adding noise," *Phys. Rev. Lett.*, vol. 85, pp. 3369–3372, Oct. 2000.

[17] P. Hanggi, "Stochastic resonance in biology how noise can enhance detection of weak signals and help improve biological information processing," *ChemPhysChem*, vol. 3, pp. 285–290, Mar. 2002.

[18] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance," *Nature*, vol. 365, pp. 337–340, Sep. 1993.

[19] L. Gammaitoni, E. Menichella-Saetta, S. Santucci, and F. Marchesoni, "Extraction of periodic signals from a noise background," *Physics Letters A*, vol. 142, pp. 59–62, Dec. 1989.

[20] A. R. Bulsara and L. Gammaitoni, "Tuning in to noise," *Physics Today*, vol. 49, no. 3, pp. 39–45, 1996.

[21] M. E. Inchiosa and A. R. Bulsara, "Signal detection statistics of stochastic resonators," *Phys. Rev. E*, vol. 53, pp. R2021–R2024, Mar. 1996.

[22] V. Galdi, V. Pierro, and I. M. Pinto, "Evaluation of stochastic-resonance-based detectors of weak harmonic signals in additive white Gaussian noise," *Physical Review E*, vol. 57, pp. 6470–6479, June 1998.

[23] S. M. Kay, J. H. Michels, H. Chen, and P. K. Varshney, "Reducing probability of decision error using stochastic resonance," *IEEE Sig. Processing Lett.*, vol. 13, pp. 695–698, Nov. 2006.

[24] S. M. Kay, "Can detectability be improved by adding noise?," *IEEE Sig. Processing Lett.*, vol. 7, pp. 8–10, Jan. 2000.

[25] 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. Sig. Processing*, vol. 56, pp. 5031–5041, Oct. 2007.

[26] S. Bayram and S. Gezici, "On the improvability and nonimprovability of detection via additional independent noise," *IEEE Sig. Processing Lett.*, vol. 16, pp. 1001–1004, Nov. 2009.

[27] X. Godivier and F. Chapeau-Blondeau, "Noise-assisted signal transmission in a nonlinear electronic comparator: Experiment and theory," *Signal Processing*, vol. 56, pp. 293–303, Feb. 1997.

[28] S. Zozor and P.-O. Amblard, "Stochastic resonance in discrete time nonlinear AR(1) models," *IEEE Trans. Signal Process.*, vol. 47, pp. 108–122, Jan. 1999.

[29] S. Bayram and S. Gezici, "Noise-enhanced M-ary hypothesis-testing in the minimax framework," in *Proc. International Conference on Signal Processing and Commun. Systems*, (Omaha, Nebraska), Sep. 2009.

[30] F. Chapeau-Blondeau, "Nonlinear test statistic to improve signal detection in non-gaussian noise," *IEEE Sig. Processing Lett.*, vol. 7, pp. 205–207, July 2000.

[31] S. Zozor and P. Amblard, "Can stochastic resonance be used in detection?," in *Proc. of the 10th European Signal Processing Conference (EUSIPCO)*, (Tampere, Finland), Sep. 2000.

[32] H. Chen, P. K. Varshney, J. H. Michels, and S. M. Kay, "Approaching near optimal detection performance via stochastic resonance," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, (Toulouse, France), May 2006.

[33] A. Asdi and A. Tewfik, "Detection of weak signals using adaptive stochastic resonance," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, (Detroit, Michigan), pp. 1332–1335, May 1995.

[34] S. Zozor and P.-O. Amblard, "On the use of stochastic resonance in sine detection," *Signal Process.*, vol. 7, pp. 353–367, Mar. 2002.

[35] S. Zozor and P.-O. Amblard, "Stochastic resonance in locally optimal detectors," *IEEE Trans. Signal Process.*, vol. 51, pp. 3177–3181, Dec. 2003.

[36] A. A. Saha and G. Anand, "Design of detectors based on stochastic resonance," *Signal Process.*, vol. 83, pp. 1193–1212, 2003.

[37] D. Rousseau and F. Chapeau-Blondeau, "Stochastic resonance and improvement by noise in optimal detection strategies," *Digital Signal Processing*, vol. 15, pp. 19–32, 2005.

[38] H. Chen, P. K. Varshney, and J. H. Michels, "Improving sequential detection performance via stochastic resonance," *IEEE Sig. Processing Lett.*, vol. 15, pp. 685–688, Dec. 2008.

[39] H. Chen, P. K. Varshney, and J. H. Michels, "Noise enhanced parameter estimation," *IEEE Trans. Sig. Processing*, vol. 56, pp. 5074–5081, Oct. 2008.

[40] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[41] E. L. Lehmann, *Testing Statistical Hypotheses*. New York: Chapman & Hall, 2 ed., 1986.

[42] J. J. L. Hodges and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *The Annals of Mathematical Statistics*, vol. 23, pp. 396–407, Sep. 1952.

[43] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*. New York: Wiley, 1954.

[44] T. S. Ferguson, *Mathematical Statistics*. New York: Academic, 1967.

[45] L. J. Savage, *The Foundations of Statistics*. New York: Dover, 2nd ed., 1972.

[46] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer, 2nd ed., 1985.

[47] A. Wald, *Statistical Decision Functions*. John Wiley and Sons, 1950.

[48] S. M. Kay, "Noise enhanced detection as a special case of randomization," *IEEE Sig. Processing Lett.*, vol. 15, pp. 709–712, 2008.

[49] S. Bayram and S. Gezici, "Effects of additional independent noise in binary composite hypothesis-testing problems," in *Proc. International Conference on Signal Processing and Commun. Systems*, (Omaha, Nebraska), Sep. 2009.

[50] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2nd ed., 2000.

[51] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization method for constrained optimization problems*, pp. 214–220. IOS Press, 2002. in Intelligent Technologies–Theory and Applications: New Trends in Intelligent Technologies.

[52] S. Koziel and Z. Michalewicz, "Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization," *Evolutionary Computation*, vol. 7, no. 1, pp. 19–44, 1999.

[53] X. Hu and R. Eberhart, "Solving constrained nonlinear optimization problems with particle swarm optimization," in *Proc. Sixth World Multiconference on Systemics, Cybernetics and Informatics 2002 (SCI 2002)*, (Orlando, FL), 2002.

[54] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. New York: Springer, 2005.

[55] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice Hall, Inc., 1998.

[56] C. C. Pugh, *Real Mathematical Analysis*. New York: Springer-Verlag, 2002.

[57] L. Huang and M. J. Neely, "The optimality of two prices: Maximizing revenue in a stochastic network," in *Proc. 45th Annual Allerton Conference on Communication, Control, and Computing*, (Monticello, IL), Sep. 2007.

[58] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1968.

[59] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Boston, MA: Athena Specific, 2003.

[60] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[61] V. Bhatia and B. Mulgrew, "Non-parametric likelihood based channel estimator for Gaussian mixture noise," *Signal Processing*, vol. 87, pp. 2569–2586, Nov. 2007.

[62] T. Erseghe and S. Tomasin, "Optimized demodulation for MAI resilient UWB W-PAN receivers," in *Proc. IEEE Int. Conf. Commun. (ICC)*, (Beijing, China), pp. 4867–4871, May 2008.

[63] T. Erseghe, V. Cellini, and G. Dona, "On UWB impulse radio receivers derived by modeling MAI as a Gaussian mixture process," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2388–2396, June 2008.

[64] S. Verdu, *Multiuser Detection*. 1st ed. Cambridge, UK: Cambridge University Press, 1998.

[65] L. Reggiani and G. M. Maggio, "Rapid search algorithms for code acquisition in UWB impulse radio communications," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 898–908, May 2005.

[66] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 4th ed., 2001.

[67] Y. Chen and V. K. Dubey, "Ultrawideband source localization using a particle-swarm-optimized Capon estimator," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4, (Seoul, Korea), pp. 2825–2829, May 2005.

[68] Y. Rahmat-Samii, D. Gies, and J. Robinson, "Particle swarm optimization (PSO): A novel paradigm for antenna designs," *The Radio Science Bulletin*, vol. 305, pp. 14–22, Sep. 2003.

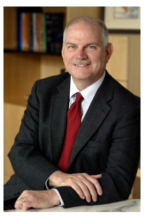
[69] Z. Yangyang, J. Chunlin, Y. Ping, L. Manlin, W. Chaojin, and W. Guangxing, "Particle swarm optimization for base station placement in mobile communication," in *Proc. IEEE International Conf. on Networking, Sensing and Control*, vol. 1, pp. 428–432, May 2004.

[70] J. Pugh, A. Martinoli, and Y. Zhang, "Particle swarm optimization for unsupervised robotic learning," in *Proc. Swarm Intelligence Symposium (SIS)*, (Pasadena, California), pp. 92–99, June 2005.

Suat Bayram (S'09) received the B.S. degree from Middle East Technical University, Ankara, Turkey in 2007 and the M.S. degree from Bilkent University, Ankara, Turkey, in 2009. He is currently working towards the Ph.D. degree in the Department of Electrical and Electronics Engineering, Bilkent University. His research interests are in the statistical signal processing and communications fields. Currently, he has particular interest in alternative detection criteria for hypothesis-testing problems, and in the theory and applications of noise enhanced detection.

Sinan Gezici (S'03, M'06) received the B.S. degree from Bilkent University, Turkey in 2001, and the Ph.D. degree in Electrical Engineering from Princeton University in 2006. From April 2006 to January 2007, he worked as a Visiting Member of Technical Staff at Mitsubishi Electric Research Laboratories, Cambridge, MA. Since February 2007, he has been an Assistant Professor in the Department of Electrical and Electronics Engineering at Bilkent University.

Dr. Gezici's research interests are in the areas of signal detection, estimation and optimization theory, and their applications to wireless communications and localization systems. Among his publications in these areas is the recent book Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols (Cambridge University Press, 2008). Dr. Gezici is a member of the IEEE.



H. Vincent Poor (S'72, M'77, SM'82, F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering and Dean of the School of Engineering and Applied Science. Dr. Poor's research interests are in the areas of stochastic analysis, statistical signal processing and their applications in wireless networks and related fields. Among his publications in these areas are the recent book MIMO Wireless Communications (Cambridge University Press, 2007), and the forthcoming book Quickest Detection (Cambridge University Press, 2009).

Dr. Poor is a member of the National Academy of Engineering, a Fellow of the American Academy of Arts and Sciences, and a former Guggenheim Fellow. He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, and in 2004-07 he served as the Editor-in-Chief of the IEEE Transactions on Information Theory. He is the recipient of the 2005 IEEE Education Medal. Recent recognition of his work includes the 2007 IEEE Marconi Prize Paper Award of the IEEE Communications Society, the 2007 Technical Achievement Award of the IEEE Signal Processing Society, and the 2008 Aaron D. Wyner Award of the IEEE Information Theory Society.