Noise Enhancement for Weighted Sum of Type I and II Error Probabilities with Constraints

Shujun Liu, Ting Yang and Kui Zhang *

College of Communication Engineering, Chongqing University, Chongqing 400044, China;
liusj@cqu.edu.cn (S.L.); yandtt2008@126.com (T.Y.)
* Correspondence: zk@cqu.edu.cn; Tel.: +86-236-5103-544

Abstract: In this paper, the noise-enhanced detection problem is investigated for the binary hypothesis-testing. The optimal additive noise is determined according to a criterion proposed by DeGroot and Schervish (2011), which aims to minimize the weighted sum of type I and II error probabilities under constraints on type I and II error probabilities. Based on a generic composite hypothesis-testing formulation, the optimal additive noise is obtained. The sufficient conditions are also deduced to verify whether the usage of the additive noise can or cannot improve the detectability of a given detector. In addition, some additional results are obtained according to the specificity of the binary hypothesis-testing, and an algorithm is developed for finding the corresponding optimal noise. Finally, numerical examples are given to verify the theoretical results and proofs of the main theorems are presented in the Appendix.

Keywords: noise enhancement; hypothesis testing; weighted sum; error probability

1. Introduction

In the binary hypothesis testing problem, there are usually a null hypothesis H_0 and an alternative hypothesis H_1, and the objective of testing is to be determine truthfulness of them based on the observation data and a decision rule. Due to the presence of noise, the decision result obviously cannot be absolutely correct. Generally, two erroneous decisions may occur in the signal detection: type I error that rejects a true null hypothesis and type II error that accepts a false null hypothesis [1].

In the classical statistical theory, the Neyman–Pearson criterion is usually applied to obtain a decision rule that minimizes the type II error probability β with a constraint on the type I error probability α. However, the minimum β may not always correspond to the optimal decision result. For instance, in the example 1 of [2], a binary hypothesis testing is designed to determine the mean of a normal data: the mean equals to -1 under H_0 and equals to 1 under H_1. Under the constraint that the type I error probability is fixed to 0.05, the type II error probability is decreased from 0.0091 to 0.00000026 when the data size increases from 20 to 100, whereas the rejection region of the null hypothesis is changed from $(0.1, +\infty)$ to $(-0.51, +\infty)$. In such case, more information brings a worse decision result even a smaller type II error probability is achieved with the fixed type I error probability. Similarly, the decision rule that minimizes the type I error probability for a fixed type II error probability may not perform well. Therefore, it could not be appropriate to simply minimize one of the two error probabilities in practice. The ideal case is to develop a decision criterion that minimizes the two types of error probabilities simultaneously, but it is almost impossible in practical applications.

In order to obtain a better decision result to balance the type I and II error probabilities, DeGroot and Schervish [1] proposed a criterion to minimize a weighted sum of type I and II error probabilities, i.e., $\min_{\phi}[c_1 \cdot \alpha(\phi) + c_2 \cdot \beta(\phi)]$, where ϕ represents the decision rule, c_1 and c_2 are the weight coefficients corresponding to α and β, respectively, and $c_1, c_2 > 0$. Furthermore, DeGroot

also provided the optimal decision procedure to minimize the weighted sum. The decision rule is given as follows. If $c_1 f(x|H_0) < c_2 f(x|H_1)$, the null hypothesis H_0 is rejected, where $f(x|H_0)$ and $f(x|H_1)$ are the respective probability density functions (pdfs) of the observation x under H_0 and H_1. If $c_1 f(x|H_0) > c_2 f(x|H_1)$, the alternative hypothesis H_1 is rejected. In addition, if $c_1 f(x|H_0) = c_2 f(x|H_1)$, the hypothesis H_0 can be either rejected or not. The optimal detector in this case is closely related to the distribution of the observation. This implies that once the distribution changes, the detector should be adjusted accordingly. But in the cases where the detector is fixed, this weighted sum rule cannot be directly applied. In such a case, finding an alternative method to minimize the weighted sum of type I and II error probabilities instead of changing the detector is important. Fortunately, the stochastic resonance (SR) theory provides a means to solve this problem.

The SR, first discovered by Benzi et al. [3] in 1981, is a phenomenon where noise plays a positive role in enhancing signal and system through a nonlinear system under certain conditions. The phenomenon of SR in the signal detection is also called noise-enhanced detection. Recent studies indicate that the system output performance can be improved significantly by adding noise to the system input or increasing the background noise level [4–22]. The improvements achieved via noise can be measured in the forms of increased signal-to-noise ratio (SNR) [7–10], mutual information (MI) [11,12] or detection probability [13–16], or in the form of decreased Bayes risk [17,18]. For example, the SNR gain of a parallel uncoupled array of bistable oscillators, operating in a mixture of sinusoidal signal and Gaussian white noise, is maximized via extra array noise [8]. In addition, due to the added array noise, the performance of a finite array closely approaches to an infinite array. In [11], the throughput MI of threshold neurons is increased by increasing the intensity of faint input noise. The optimal additive noise to maximize the detection probability with a constraint on false-alarm probability is studied in [13], and the sufficient conditions for improvability and non-improvability are deduced. In [17], the effects of additive independent noise on the performance of suboptimal detectors are investigated according to the restricted Bayes criterion, where the minimum noise modified Bayes risk is explored with certain constraints on the conditional risks. Inspired by this concept, it is reasonable to conjecture that a proper noise can decrease the weighted sum of type I and II error probabilities for a fixed detector.

In the absence of constraints, it is obvious that the additive noise that minimizes the weighted sum is a constant vector, whereas the corresponding type I or II error probability may exceed a certain value to cause a bad decision result. To avoid this problem, two constraints are enforced on type I and II error probabilities, respectively, to keep a balance. The aim of this work is to find the optimal additive noise that minimizes the weighted sum of type I and II error probabilities with the constraints on type I and II error probabilities for a fixed detector. Furthermore, the work can also be extended to some applications, such as the energy detection in sensor networks [23,24] and the independent Bernoulli trials [25]. The main contributions of this paper are summarized as follows:

- Formulation of the optimization problem for minimizing the noise modified weighted sum of type I and II error probabilities under the constraints on the two error probabilities is presented.
- Derivations of the optimal noise that minimizes the weighted sum and sufficient conditions for improvability and nonimprovability for a general composite hypothesis testing problem are provided.
- Analysis of the characteristics of the optimal additive noise that minimizes the weighted sum for a simple hypothesis testing problem is studied and the corresponding algorithm to solve the optimization problem is developed.
- Numerical results are presented to verify the theoretical results and to demonstrate the superior performance of the proposed detector.

The remainder of this paper is organized as follows: in Section 2, a noise modified composite hypothesis testing problem is formulated first for minimizing the weighted sum of type I and II error probabilities under different constraints. Then the sufficient conditions for improvability and

nonimprovability are given and the optimal additive noise is derived. In Section 3, additional theoretical results are analyzed for a simple hypothesis testing problem. Finally, simulation results are shown in Section 4 and conclusions are made in Section 5.

Notation: Lower-case bold letters denote vectors, with v_i denoting the i-th element of $\boldsymbol{v}$; θ denotes the value of parameter Θ; $f_{\boldsymbol{v}}(\boldsymbol{v}|\theta)$ denotes the pdf of $\boldsymbol{v}$ for a given parameter value $\Theta = \theta$; Ω_i denotes the set of all possible parameter values of Θ under H_i; $\delta(\cdot)$ denotes the Dirac function; $\cap$, $\cup$ and $\varnothing$ denote intersection, union and null set, respectively; $*$, $(\cdot)^T$, $\int$, $E\{\cdot\}$, min, max and arg denote convolution, transpose, integral, expectation, minimum, maximum and argument operators, respectively; $\inf\{\cdot\}$ and $\sup\{\cdot\}$ denote the infimum and supremum operators, respectively; $\sum$ means summation; ∇ and H denote the respective gradient and Hessian operators.

2. Noise Enhanced Composite Hypothesis Testing

2.1. Problem Formulation

Consider the following binary composite hypothesis testing problem:

$$\left\{ \begin{array}{l} H_0 : f_{\mathbf{x}}(\mathbf{x}|\theta), \theta \in \Omega_0 \\ H_1 : f_{\mathbf{x}}(\mathbf{x}|\theta), \theta \in \Omega_1 \end{array} \right. \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^N$ is the observation vector, H_0 and H_1 are the null and the alternative hypothesizes, respectively, θ denotes the value of parameter Θ, $f_{\mathbf{x}}(\mathbf{x}|\theta)$ represents the pdf of $\mathbf{x}$ for a given parameter value $\Theta = \theta$. The parameter Θ has multiple possible values under each hypothesis and denote the pdf of any parameter value $\Theta = \theta$ under H_0 and H_1 by $\varpi_0(\theta)$ and $\varpi_1(\theta)$. In addition, Ω_0 and Ω_1 denote the respective sets of all possible values of Θ under H_0 and H_1. It is true that $\Omega_0 \cap \Omega_1 = \varnothing$ and the union of them forms the parameter space Ω, i.e., $\Omega = \Omega_0 \cup \Omega_1$.

Without loss of generality, a decision rule (detector) is considered as:

$$\phi(\mathbf{x}) = \left\{ \begin{array}{l} 1, \mathbf{x} \in \Gamma_1 \\ 0, \mathbf{x} \in \Gamma_0 \end{array} \right. \tag{2}$$

where Γ_0 and Γ_1 form the observation space Γ. Actually, the detector chooses H_1 if $\mathbf{x} \in \Gamma_1$, otherwise chooses H_0 if $\mathbf{x} \in \Gamma_0$.

In order to investigate the performance of the detector achieved via an additive noise, a noise modified observation $\mathbf{y}$ is obtained by adding an independent additive noise $\mathbf{n}$ to the original observation $\mathbf{x}$, i.e., $\mathbf{y} = \mathbf{x} + \mathbf{n}$. For a given parameter value $\Theta = \theta$, the pdf of $\mathbf{y}$ is calculated by the convolution of the pdfs of $\mathbf{x}$ and $\mathbf{n}$, given by:

$$f_{\mathbf{y}}(\mathbf{y}|\theta) = f_{\mathbf{x}}(\mathbf{x}|\theta) * p_{\mathbf{n}}(\mathbf{n}) = \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) f_{\mathbf{x}}(\mathbf{y} - \mathbf{n}|\theta) d\mathbf{n} \tag{3}$$

where $p_{\mathbf{n}}(\mathbf{n})$ denotes the pdf of $\mathbf{n}$. For a fixed detector, the noise modified type I and II error probabilities of the detector for given parameter values now is expressed as:

$$\alpha^{\mathbf{y}}(\phi; \theta) = \int_{\Gamma} \phi(\mathbf{y}) f_{\mathbf{y}}(\mathbf{y}|\theta) d\mathbf{y} = \int_{\Gamma_1} f_{\mathbf{y}}(\mathbf{y}|\theta) d\mathbf{y}, \theta \in \Omega_0 \tag{4}$$

$$\beta^{\mathbf{y}}(\phi; \theta) = \int_{\Gamma} (1 - \phi(\mathbf{y})) f_{\mathbf{y}}(\mathbf{y}|\theta) d\mathbf{y} = \int_{\Gamma_0} f_{\mathbf{y}}(\mathbf{y}|\theta) d\mathbf{y}, \theta \in \Omega_1 \tag{5}$$

Correspondingly, the average noise modified type I and II error probabilities are calculated by:

$$\alpha^{\mathbf{y}}(\phi) = \int_{\Omega_0} \alpha^{\mathbf{y}}(\phi; \theta) \varpi_0(\theta) d\theta \tag{6}$$

$$\beta^{\mathbf{y}}(\phi) = \int_{\Omega_1} \beta^{\mathbf{y}}(\phi;\theta)\omega_1(\theta)d\theta \tag{7}$$

From (6) and (7), the weighted sum of the two types of average error probabilities is obtained as:

$$
\begin{aligned}
Er^{\mathbf{y}} &= c_1 \cdot \alpha^{\mathbf{y}}(\phi) + c_2 \cdot \beta^{\mathbf{y}}(\phi) \\
&= c_1 \cdot \int_{\Omega_0} \alpha^{\mathbf{y}}(\phi;\theta)\omega_0(\theta)d\theta + c_2 \cdot \int_{\Omega_1} \beta^{\mathbf{y}}(\phi;\theta)\omega_1(\theta)d\theta
\end{aligned}
\tag{8}
$$

where c_1 and c_2 are the weights assigned for the type I and II error probabilities, which can be predefined according to the actual situations. For example, if the prior probabilities are known, the value of c_1 and c_2 equal the prior probabilities corresponding to H_0 and H_1, respectively. Besides, the values of c_1 and c_2 can also be determined based on the expected decision results.

In this work, the aim is to find the optimal independent additive noise, which minimizes the weighted sum of the average error probabilities under the constraints on the maximum type I and II error probabilities for different parameter values. The optimization problem can be formulated as below:

$$p_{\mathbf{n}}^{opt}(\mathbf{n}) = \arg\min_{p_{\mathbf{n}}(\mathbf{n})} Er^{\mathbf{y}} \tag{9}$$

subject to

$$
\begin{cases}
\max\limits_{\theta \in \Omega_0} \alpha^{\mathbf{y}}(\phi;\theta) \leq \alpha_o \\
\max\limits_{\theta \in \Omega_1} \beta^{\mathbf{y}}(\phi;\theta) \leq \beta_o
\end{cases}
\tag{10}
$$

where α_o and β_o are the upper limits for the type I and II error probabilities, respectively.

In order to explicitly express the optimization problem described in (9) and (10), substituting (3) into (4) produces:

$$
\begin{aligned}
\alpha^{\mathbf{y}}(\phi;\theta) &= \int_{\Gamma_1} \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) f_{\mathbf{x}}(\mathbf{y}-\mathbf{n}|\theta)d\mathbf{n}\,d\mathbf{y} \\
&= \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) \int_{\Gamma_1} f_{\mathbf{x}}(\mathbf{y}-\mathbf{n}|\theta)d\mathbf{y}d\mathbf{n} \quad ,\theta \in \Omega_0 \\
&= \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) A_{\theta}(\mathbf{n})d\mathbf{n} \\
&= E\{A_{\theta}(\mathbf{n})\}
\end{aligned}
\tag{11}
$$

where

$$A_{\theta}(\mathbf{n}) = \int_{\Gamma_1} f_{\mathbf{x}}(\mathbf{y}-\mathbf{n}|\theta)d\mathbf{y} = \int_{\Gamma} \phi(\mathbf{y})f_{\mathbf{x}}(\mathbf{y}-\mathbf{n}|\theta)d\mathbf{y}, \theta \in \Omega_0 \tag{12}$$

It should be noted that $A_{\theta}(\mathbf{n})$ can be viewed as the type I error probability obtained by adding a constant vector $\mathbf{n}$ to $\mathbf{x}$ for $\theta \in \Omega_0$. Therefore, $\alpha^{\mathbf{x}}(\phi;\theta) = A_{\theta}(0) = \int_{\Gamma} \phi(\mathbf{x})f_{\mathbf{x}}(\mathbf{x}|\theta)d\mathbf{x}$ denotes the type I error probability for the original observation $\mathbf{x}$.

Similarly, $\beta^{\mathbf{y}}(\phi;\theta)$ in (5) can be expressed as:

$$\beta^{\mathbf{y}}(\phi;\theta) = E\{B_{\theta}(\mathbf{n})\}, \theta \in \Omega_1 \tag{13}$$

where

$$B_{\theta}(\mathbf{n}) = \int_{\Gamma_0} f_{\mathbf{x}}(\mathbf{y}-\mathbf{n}|\theta)d\mathbf{y} = \int_{\Gamma} (1-\phi(\mathbf{y}))f_{\mathbf{x}}(\mathbf{y}-\mathbf{n}|\theta)d\mathbf{y}, \theta \in \Omega_1 \tag{14}$$

The $B_{\theta}(\mathbf{n})$ can be treated as the type II error probability obtained by adding a constant vector $\mathbf{n}$ to $\mathbf{x}$ for $\theta \in \Omega_1$ and $\beta^{\mathbf{x}}(\phi;\theta) = B_{\theta}(0) = \int_{\Gamma}(1-\phi(\mathbf{x}))f_{\mathbf{x}}(\mathbf{x}|\theta)d\mathbf{x}$ is the original type II error probability without adding noise for $\theta \in \Omega_1$.

With (11) and (13), (8) becomes:

$$Er^{\mathbf{y}} = \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) \left[c_1 \cdot \int_{\Omega_0} A_{\theta}(\mathbf{n})\omega_0(\theta)d\theta + c_2 \cdot \int_{\Omega_1} B_{\theta}(\mathbf{n})\omega_1(\theta)d\theta \right] d\mathbf{n} = E\{Er(\mathbf{n})\} \tag{15}$$

where

$$Er(\mathbf{n}) = c_1 \cdot \int_{\Omega_0} A_{\theta}(\mathbf{n})\omega_0(\theta)d\theta + c_2 \cdot \int_{\Omega_1} B_{\theta}(\mathbf{n})\omega_1(\theta)d\theta \tag{16}$$

Accordingly, $Er(\mathbf{n})$ is the weighted sum of two types of average error probabilities achieved by adding a constant vector $\mathbf{n}$ to the original observation $\mathbf{x}$. Naturally, $Er^{\mathbf{x}} = Er(0)$ denotes the weighted sum of type I and II average error probabilities for the original observation $\mathbf{x}$.

Combined (11), (13) and (15), the optimization problem in (9) and (10) now is:

$$p_{\mathbf{n}}^{opt}(\mathbf{n}) = \arg\min_{p_{\mathbf{n}}(\mathbf{n})} E\{Er(\mathbf{n})\} \tag{17}$$

subject to

$$\begin{cases} \max_{\theta \in \Omega_0} E\{A_\theta(\mathbf{n})\} \leq \alpha_o \\ \max_{\theta \in \Omega_1} E\{B_\theta(\mathbf{n})\} \leq \beta_o \end{cases} \tag{18}$$

2.2. Sufficient Conditions for Improvability and Non-improvability

In practice, the solution of the optimization problem in (17) and (18) requires a research over all possible noises and this procedure is complicated. Therefore, it is worthwhile to determine whether the detector can or cannot be improved by adding additive noise in advance. From (17) and (18), a detector is considered to be improvable if there exists one noise $\mathbf{n}$ that satisfies $E\{Er(\mathbf{n})\} < Er^{\mathbf{x}} = Er(0)$, $\max_{\theta \in \Omega_0} E\{A_\theta(\mathbf{n})\} \leq \alpha_o$ and $\max_{\theta \in \Omega_0} E\{B_\theta(\mathbf{n})\} \leq \beta_o$ simultaneously; otherwise, the detector is considered to be non-improvable.

The sufficient conditions for non-improvability can be obtained according to the characteristics of $A_\theta(\mathbf{n})$, $B_\theta(\mathbf{n})$ and $Er(\mathbf{n})$, which are provided in Theorem 1.

Theorem 1. *If there exists $\theta^* \in \Omega_0$ ($\theta^* \in \Omega_1$) such that $A_{\theta^*}(\mathbf{n}) \leq \alpha_o$ ($B_{\theta^*}(\mathbf{n}) \leq \beta_o$) implies $Er(\mathbf{n}) \geq Er(0)$ for any $\mathbf{n} \in P_{\mathbf{n}}$, where $P_{\mathbf{n}}$ represents the convex set of all possible additive noises, and if $A_{\theta^*}(\mathbf{n})$ ($B_{\theta^*}(\mathbf{n})$) and $Er(\mathbf{n})$ are convex functions over $P_{\mathbf{n}}$, then the detector is non-improvable.*

The proof is provided in Appendix A.

Under the conditions in Theorem 1, the detector cannot be improved and it is unnecessary to solve the optimization problem in (17) and (18). In other words, if the conditions in Theorem 1 are satisfied, the three inequities $Er^{\mathbf{y}} \geq Er^{\mathbf{x}}$, $\max_{\theta \in \Omega_0} E\{A_\theta(\mathbf{n})\} \leq \alpha_o$ and $\max_{\theta \in \Omega_1} E\{B_\theta(\mathbf{n})\} \leq \beta_o$ cannot be achieved simultaneously by adding any additive noise. In addition, even if the conditions in Theorem 1 are not satisfied, the detector can also be non-improvable. This implies the sufficient conditions for improvability need to be addressed.

The sufficient conditions for improvability are discussed now. Suppose that $A_\theta(\mathbf{x})$ ($\forall \theta \in \Omega_0$), $B_\theta(\mathbf{x})$ ($\forall \theta \in \Omega_1$) and $Er(\mathbf{x})$ are second-order continuously differentiable around $\mathbf{x} = 0$. In order to facilitate the subsequent analysis, six auxiliary functions are predefined as follows based on the first and second partial derivatives of $A_\theta(\mathbf{x})$, $B_\theta(\mathbf{x})$ and $Er(\mathbf{x})$ with respect to the elements of $\mathbf{x}$. The first three auxiliary functions $a_\theta^{(1)}(\mathbf{x}, \mathbf{g})$, $b_\theta^{(1)}(\mathbf{x}, \mathbf{g})$ and $er^{(1)}(\mathbf{x}, \mathbf{g})$ are defined as the weight sums of the first partial derivatives of $A_\theta(\mathbf{x})$, $B_\theta(\mathbf{x})$ and $Er(\mathbf{x})$, respectively, based on the coefficient vector $\mathbf{g}$. Specifically:

$$a_\theta^{(1)}(\mathbf{x}, \mathbf{g}) \triangleq \sum_{i=1}^{N} g_i \frac{\partial A_\theta(\mathbf{x})}{\partial \mathbf{x}_i} = \mathbf{g}^T \nabla A_\theta(\mathbf{x}), \theta \in \Omega_0 \tag{19}$$

$$b_\theta^{(1)}(\mathbf{x}, \mathbf{g}) \triangleq \sum_{i=1}^{N} g_i \frac{\partial B_\theta(\mathbf{x})}{\partial \mathbf{x}_i} = \mathbf{g}^T \nabla B_\theta(\mathbf{x}), \theta \in \Omega_1 \tag{20}$$

$$er^{(1)}(\mathbf{x}, \mathbf{g}) \triangleq \sum_{i=1}^{N} g_i \frac{\partial Er(\mathbf{x})}{\partial \mathbf{x}_i} = \mathbf{g}^T \nabla Er(\mathbf{x}) \tag{21}$$

where $\mathbf{g}$ is a N-dimensional column vector, $\mathbf{g}^T$ is the transposition of $\mathbf{g}$, $\mathbf{x}_i$ and $\mathbf{g}_i$ are the i-th elements of $\mathbf{x}$ and $\mathbf{g}$, respectively. In addition, ∇ denotes the gradient operator, thereby $\nabla A_\theta(\mathbf{x})$ $(\nabla B_\theta(\mathbf{x}), \nabla Er(\mathbf{x}))$ is a N-dimensional column vector with i-th element $\partial A_\theta(\mathbf{x})/\partial \mathbf{x}_i$ $(\partial B_\theta(\mathbf{x})/\partial \mathbf{x}_i, \partial E_r(\mathbf{x})/\partial \mathbf{x}_i)$, $i = 1, \ldots, N$. The last three auxiliary functions $a_\theta^{(2)}(\mathbf{x}, \mathbf{g})$, $b_\theta^{(2)}(\mathbf{x}, \mathbf{g})$ and $er^{(2)}(\mathbf{x}, \mathbf{g})$ are defined as the weight sums of the second partial derivatives of $A_\theta(\mathbf{x})$, $B_\theta(\mathbf{x})$ and $Er(\mathbf{x})$ based on the $N \times N$ coefficient matrix $\mathbf{g}\mathbf{g}^T$, i.e.,

$$a_\theta^{(2)}(\mathbf{x}, \mathbf{g}) \triangleq \sum_{j=1}^{N} \sum_{i=1}^{N} g_j g_i \frac{\partial^2 A_\theta(\mathbf{x})}{(\partial \mathbf{x}_j \partial \mathbf{x}_i)} = \mathbf{g}^T H(A_\theta(\mathbf{x}))\mathbf{g}, \theta \in \Omega_0 \tag{22}$$

$$b_\theta^{(2)}(\mathbf{x}, \mathbf{g}) \triangleq \sum_{j=1}^{N} \sum_{i=1}^{N} g_j g_i \frac{\partial^2 B_\theta(\mathbf{x})}{(\partial \mathbf{x}_j \partial \mathbf{x}_i)} = \mathbf{g}^T H(B_\theta(\mathbf{x}))\mathbf{g}, \theta \in \Omega_1 \tag{23}$$

$$er^{(2)}(\mathbf{x}, \mathbf{g}) \triangleq \sum_{j=1}^{N} \sum_{i=1}^{N} g_j g_i \frac{\partial^2 Er(\mathbf{x})}{(\partial \mathbf{x}_j \partial \mathbf{x}_i)} = \mathbf{g}^T H(Er(\mathbf{x}))\mathbf{g} \tag{24}$$

where H denote the Hessian operator, $H(A_\theta(\mathbf{x}))$ $(H(B_\theta(\mathbf{x})), H(Er(\mathbf{x})))$ is a $N \times N$ matrix with its (j,i)-th element denoted by $\partial^2 A_\theta(\mathbf{x})/(\partial \mathbf{x}_j \partial \mathbf{x}_i)$ $(\partial^2 B_\theta(\mathbf{x})/(\partial \mathbf{x}_j \partial \mathbf{x}_i), \partial^2 Er(\mathbf{x})/(\partial \mathbf{x}_j \partial \mathbf{x}_i))$, where $i, j = 1, \ldots, N$.

Based on the definitions in (19)–(24), Theorem 2 presents the sufficient conditions for improvability.

Theorem 2. *Suppose that Λ_0 and Λ_1 are the sets of all possible values of θ that maximize $A_\theta(0)$ and $B_\theta(0)$, respectively, $\alpha_o = \max_{\theta \in \Omega_0} A_\theta(0)$ and $\beta_o = \max_{\theta \in \Omega_1} B_\theta(0)$. The detector is improvable, if there exists a N-dimensional column vector $\mathbf{g}$ that satisfies one of the following conditions for all $\theta_0 \in \Lambda_0$ and $\theta_1 \in \Lambda_1$:*

(1) $er^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$, $a_{\theta_0}^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$, $b_{\theta_1}^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$;

(2) $er^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} > 0$, $a_{\theta_0}^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} > 0$, $b_{\theta_1}^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} > 0$;

(3) $er^{(2)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$, $a_{\theta_0}^{(2)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$, $b_{\theta_1}^{(2)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$.

The proof is presented in Appendix B.

Theorem 2 indicates that under the condition (1), (2) or (3), there always exist noises that decrease the weighted sum of average error probabilities under the constraints on the type I and II error probabilities. In addition, alternative sufficient conditions for improvability can be obtained by defining the following two functions, and they are:

$$I(t) = \inf\left\{ Er(\mathbf{n}) \Big| \max_{\theta \in \Omega_0} A_\theta(\mathbf{n}) = t, \mathbf{n} \in \mathbb{R}^N \right\} \tag{25}$$

$$S(t) = \sup\left\{ \max_{\theta \in \Omega_1} B_\theta(\mathbf{n}) \Big| \max_{\theta \in \Omega_0} A_\theta(\mathbf{n}) = t, \mathbf{n} \in \mathbb{R}^N \right\} \tag{26}$$

where $I(t)$ and $S(t)$ are the minimum weighted sum of two types of average error probabilities and the maximum type II error probability for a given maximum type I error probability obtained via adding a constant vector, respectively. If there is a $t_0 \leq \alpha_o$ such that $I(t_0) \leq Er(0)$ and $S(0) \leq \beta_o$, the detector is improvable. More specifically, there exists a constant vector $\mathbf{n}_0$ that satisfies $\max_{\theta \in \Omega_0} A_\theta(\mathbf{n}_0) = t_0 \leq \alpha_o$, $Er(\mathbf{n}_0) \leq Er(0)$ and $\max_{\theta \in \Omega_0} B_\theta(\mathbf{n}_0) \leq \beta_o$ simultaneously. However, in most cases, the solution of the optimization problem in (17) and (18) is not a constant vector. A more practical sufficient condition for improvability is shown in Theorem 3.

Theorem 3. *Let $\widetilde{\alpha} = \max_{\theta \in \Omega_0} \alpha^{\mathbf{x}}(\phi; \theta)$ and $\widetilde{\beta} = \max_{\theta \in \Omega_1} \beta^{\mathbf{x}}(\phi; \theta)$ be the respective maximum type I and II error probabilities without adding any noise, and suppose that $\widetilde{\alpha} \leq \alpha_o$, $\widetilde{\beta} \leq \beta_o$ and $S(\widetilde{\alpha}) = \widetilde{\beta}$. If $I(t)$ and $S(t)$*

are second-order continuously differentiable around $t = \widetilde{\alpha}$, and $I''(\widetilde{\alpha}) < 0$ and $S''(\widetilde{\alpha}) < 0$ hold at the same time, then the detector is improvable.

The proof is given in Appendix C.

Additionally, the following functions $J(\varepsilon)$ and $G(\varepsilon)$ are defined:

$$J(\varepsilon) = \inf\left\{ Er(\mathbf{n}) \middle| \max_{\theta \in \Omega_1} B_\theta(\mathbf{n}) = \varepsilon, \mathbf{n} \in \mathbb{R}^N \right\} \tag{27}$$

$$G(\varepsilon) = \sup\left\{ \max_{\theta \in \Omega_0} A_\theta(\mathbf{n}) \middle| \max_{\theta \in \Omega_1} B_\theta(\mathbf{n}) = \varepsilon, \mathbf{n} \in \mathbb{R}^N \right\} \tag{28}$$

A similar conclusion to the Theorem 3 can be made as well, provided in Corollary 1.

Corollary 1. *The detector is improvable, if $J''(\widetilde{\beta}) \leq 0$ and $G''(\widetilde{\beta}) \leq 0$ hold, where $J(\varepsilon)$ and $G(\varepsilon)$ are second-order continuously differentiable around $\varepsilon = \widetilde{\beta}$, and $G(\widetilde{\beta}) = \widetilde{\alpha}$.*

The proof is similar to that of Theorem 3 and it is omitted here.

2.3. Optimal Additive Noise

In general, it is difficult to solve the optimization problem in (17) and (18) directly, because the solution is obtained based on the search over all possible additive noises. Hence, in order to reduce the computational complexity, one can utilize Parzen window density estimation to obtain an approximate solution. Actually, the pdf of the optimal additive noise can be approximated by:

$$p_{\mathbf{n}}(\mathbf{n}) = \sum_{l=1}^{L} \eta_l \vartheta_l(\mathbf{n}) \tag{29}$$

where $\eta_l \geq 0$ and $\sum_{l=1}^{L} \eta_l = 1$, while $\vartheta_l(\cdot)$ represents the window function that satisfies $\vartheta_l(\mathbf{x}) \geq 0$ for any $\mathbf{x}$ and $\int \vartheta_l(\mathbf{x})d\mathbf{x} = 0$ for $l = 1, \ldots, L$. The window function can be a cosine window, rectangular window, or Gauss window function. With (29), the optimization problem is simplified to obtain the parameter values corresponding to each window function. In such cases, global optimization algorithms can be applied such as Particle swarm optimization (PSO), Ant colony algorithm (ACA), and Genetic algorithm (GA) [26–28].

If the numbers of parameter values in Ω_0 and Ω_1 are finite, the optimal additive noise for (17) and (18) is a randomization of no more than $M + K$ constant vectors. In this case, Ω_0 and Ω_1 can be expressed by $\Omega_0 = \{\theta_{01}, \theta_{02}, \ldots, \theta_{0M}\}$ and $\Omega_1 = \{\theta_{11}, \theta_{12}, \ldots, \theta_{1K}\}$, where M and K are finite positive integers. The Theorem 4 states this claim.

Theorem 4. *Suppose that each component in the optimal additive noise is finite, namely $n_i \in [a_i, b_i]$ for $i = 1, \ldots, N$, where a_i and b_i are two finite values. If $A_{\theta_{0i}}(\cdot)$ and $B_{\theta_{1i}}(\cdot)$ are continuous functions, the pdf of the optimal additive noise for the optimization problem in (17) and (18) can be expressed as:*

$$p_{\mathbf{n}}(\mathbf{n}) = \sum_{l=1}^{M+K} \eta_l \delta(\mathbf{n} - \mathbf{n}_l) \tag{30}$$

where $\eta_l \geq 0$ and $\sum_{l=1}^{M+K} \eta_l = 1$.

The proof is similar to that of Theorem 4 in [17] and Theorem 3 in [13], and omitted here. In some special cases, the optimal additive can be solved directly based on the characteristics of $I(t)$ ($H(\varepsilon)$). For example, let $Er_{\min} = \min_t I(t) = I(t_m)$ ($Er_{\min} = \min_\varepsilon H(\varepsilon) = I(\varepsilon_m)$) and $\max_{\theta \in \Omega_0} A_\theta(\mathbf{n}_m) = t_m$ ($\max_{\theta \in \Omega_1} B_\theta(\mathbf{n}_m) = \varepsilon_m$). If $t_m \leq \alpha_o$ ($\varepsilon_m \leq \beta_o$) and $\max_{\theta \in \Omega_1} B_\theta(\mathbf{n}_m) \leq \beta_o$ ($\max_{\theta \in \Omega_0} A_\theta(\mathbf{n}_m) \leq \alpha_o$), the optimal additive

noise is a constant vector with pdf of $p_{\mathbf{n}}(\mathbf{n}) = \delta(\mathbf{n} - \mathbf{n}_m)$. In addition, equality of $\max\limits_{\theta \in \Omega_0} E\{A_\theta(\mathbf{n})\} = \alpha_o$ $(\max\limits_{\theta \in \Omega_1} E\{B_\theta(\mathbf{n})\} = \beta_o)$ holds if $t_m > \alpha_o$ $(\varepsilon_m > \beta_o)$.

3. Noise Enhanced Simple Hypothesis Testing

In this section, the noise enhanced binary simple hypothesis testing problem is considered, which is a special case of the optimization problem in (9) and (10). Therefore, the conclusions obtained in Section 2 are also applicable in this section. Furthermore, due to the specificity of simple binary hypothesis testing problem, some additional results are also obtained.

3.1. Problem Formulation

When $\Omega_i = \{\theta_i\}$, $i = 0, 1$, the composite binary hypothesis testing problem described in (1) is simplified to a simple binary hypothesis testing problem. In this case, the probability of θ_i under H_i equals to 1, i.e., $\varpi_i(\theta) = 1$ for $i = 0, 1$. Therefore, the corresponding noise modified type I and II error probabilities is rewritten as:

$$\alpha^{\mathbf{y}}(\phi) = \alpha^{\mathbf{y}}(\phi; \theta_0) = \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) \int_{\Gamma_1} f_0(\mathbf{y} - \mathbf{n}) d\mathbf{y} d\mathbf{n} = E\{A_0(\mathbf{n})\} \tag{31}$$

$$\beta^{\mathbf{y}}(\phi) = \beta^{\mathbf{y}}(\phi; \theta_1) = \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) \int_{\Gamma_0} f_1(\mathbf{y} - \mathbf{n}) d\mathbf{y} d\mathbf{n} = E\{B_1(\mathbf{n})\} \tag{32}$$

where $f_0(\cdot)$ and $f_1(\cdot)$ represent the pdfs of $\mathbf{x}$ under H_0 and H_1, respectively, and $A_0(\mathbf{n})$ and $B_1(\mathbf{n})$ are:

$$A_0(\mathbf{n}) = \int_{\Gamma_1} f_0(\mathbf{y} - \mathbf{n}) d\mathbf{y} \tag{33}$$

$$B_1(\mathbf{n}) = \int_{\Gamma_0} f_1(\mathbf{y} - \mathbf{n}) d\mathbf{y} \tag{34}$$

Correspondingly, the weighted sum of noise modified type I and II error probabilities is calculated by:

$$\begin{aligned} Er^{\mathbf{y}} &= c_1 \alpha^{\mathbf{y}}(\phi) + c_2 \beta^{\mathbf{y}}(\phi) = c_1 E\{A_0(\mathbf{n})\} + c_2 E\{B_1(\mathbf{n})\} \\ &= \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n})(c_1 A_0(\mathbf{n}) + c_2 B_1(\mathbf{n})) d\mathbf{n} \\ &= E\{Er(\mathbf{n})\} \end{aligned} \tag{35}$$

where

$$Er(\mathbf{n}) = c_1 A_0(\mathbf{n}) + c_2 B_1(\mathbf{n}) \tag{36}$$

As a result, the optimization problem in (9) and (10) becomes:

$$p_{\mathbf{n}}^{opt}(\mathbf{n}) = \arg\min_{p_{\mathbf{n}}(\mathbf{n})} E\{Er(\mathbf{n})\} \tag{37}$$

subject to

$$\begin{cases} E\{A_0(\mathbf{n})\} \leq \alpha_o \\ E\{B_1(\mathbf{n})\} \leq \beta_o \end{cases} \tag{38}$$

Based on the definitions in (33) and (34), $A_0(\mathbf{n})$ and $B_1(\mathbf{n})$ are viewed as the noise modified type I and II error probabilities obtained by adding a constant vector noise. Furthermore, $A_0(0)$ and $B_1(0)$ are the original type I and II error probabilities, respectively.

3.2. Algorithm for the Optimal Additive Noise

According to the Theorem 4 in Section 2.3, the optimal additive noise for the optimization problem in (37) and (38) is a randomization of most two constant vectors with the pdf $p_{\mathbf{n}}^{opt}(\mathbf{n}) = \eta \delta(\mathbf{n} - \mathbf{n}_1) +$

$(1 - \eta)\delta(\mathbf{n} - \mathbf{n}_2)$. In order to find the values of η, $\mathbf{n}_1$ and $\mathbf{n}_2$, we first divide each constant vector $\mathbf{n}$ into four disjoint sets according to the relationships of $A_0(\mathbf{n})$ and α_o, $B_1(\mathbf{n})$ and β_o. To be specific, the four disjoint sets are $Q_1 = \{\mathbf{n}|A_0(\mathbf{n}) \leq \alpha_o, B_1(\mathbf{n}) \leq \beta_o\}$, $Q_2 = \{\mathbf{n}|A_0(\mathbf{n}) \leq \alpha_o, B_1(\mathbf{n}) > \beta_o\}$, $Q_3 = \{\mathbf{n}|A_0(\mathbf{n}) > \alpha_o, B_1(\mathbf{n}) \leq \beta_o\}$, and $Q_4 = \{\mathbf{n}|A_0(\mathbf{n}) > \alpha_o, B_1(\mathbf{n}) > \beta_o\}$. Then, we calculate the minimum $Er(\mathbf{n})$, the corresponding set of all possible values of $\mathbf{n}$ is denoted by $Q_e = \left\{\mathbf{n}\middle|\mathbf{n} = \underset{\mathbf{n}}{\arg\min} Er(\mathbf{n})\right\}$. It should be noted that $\mathbf{n} \in Q_e$ is the optimal additive noise that minimizes the weighted sum without constraints.

It is obvious that Q_1, Q_2 and Q_3 do not exist if all the elements of Q_e belong to Q_4. In other words, if $Q_e \subset Q_4$, there is no additive noise that satisfies $E\{Er(\mathbf{n})\} < Er(0)$ under the constraints of $E\{A_0(\mathbf{n})\} \leq \alpha_o$ and $E\{B_1(\mathbf{n})\} \leq \beta_o$. Therefore, if the detector is improvable, the elements of Q_e must come from Q_1, Q_2 and/or Q_3. Theorem 5 is now provided to find the values of η, $\mathbf{n}_1$ and $\mathbf{n}_2$.

Theorem 5. *Let* $\eta_1 = \frac{\alpha_o - A_0(\mathbf{n}_2)}{A_0(\mathbf{n}_1) - A_0(\mathbf{n}_2)}$ *and* $\eta_2 = \frac{\beta_o - B_1(\mathbf{n}_2)}{B_1(\mathbf{n}_1) - B_1(\mathbf{n}_2)}$.

(1) *If* $Q_e \cap Q_1 \neq \varnothing$, *then* $\eta = 1$ *and* $\mathbf{n}_1 \in Q_e \cap Q_1$ *such that* $Er_{opt}^{\mathbf{y}} = Er(\mathbf{n}_1) = \underset{\mathbf{n}}{\min} Er(\mathbf{n})$.

(2) *If* $Q_e \cap Q_2 \neq \varnothing$ *and* $Q_e \cap Q_3 \neq \varnothing$ *are true, then we have* $\mathbf{n}_1 \in Q_e \cap Q_2$, $\mathbf{n}_2 \in Q_e \cap Q_3$, $\eta_1 \leq \eta \leq \eta_2$, *and* $Er_{opt}^{\mathbf{y}} = \underset{\mathbf{n}}{\min} Er(\mathbf{n})$.

(3) *If* $Q_e \subset Q_2$, *then* $Er_{opt}^{\mathbf{y}}$ *is obtained when* $\eta = \eta_2$, *and the corresponding* $E\{A_0(\mathbf{n})\}$ *achieves the minimum and* $E\{B_1(\mathbf{n})\} = \beta_o$.

(4) *If* $Q_e \subset Q_3$, *then* $Er_{opt}^{\mathbf{y}}$ *is achieved when* $\eta = \eta_1$, *and the corresponding* $E\{A_0(\mathbf{n})\} = \alpha_o$ *and* $E\{B_1(\mathbf{n})\}$ *reaches the minimum.*

The corresponding proofs are provided in Appendix D.

From (3) and (4) in Theorem 5, under the constraints on $E\{A_0(\mathbf{n})\} \leq \alpha_o$ and $E\{B_1(\mathbf{n})\} \leq \beta_o$, the solution of the optimization problem in (37) and (38) is identical with the additive noise that minimizes $E\{A_0(\mathbf{n})\}$ ($E\{B_1(\mathbf{n})\}$) when $Q_e \subset Q_2$ ($Q_e \subset Q_3$). In such cases, the optimal solution can be obtained easily by referring the algorithm provided in [14].

4. Numerical Results

In this section, a binary hypothesis testing problem is studied to verify the theoretical analysis, and it is:

$$\begin{cases} H_0 : x = v \\ H_1 : x = \Theta + v \end{cases} \tag{39}$$

where $x \in \mathbb{R}$ is an observation, Θ is a constant or random variable, and v is the background noise with pdf $p_v(\cdot)$. From (39), the pdf of x under H_0 is $f_0(x) = p_v(x)$, and the pdf of x under H_1 for a given parameter value $\Theta = \theta$ is denoted by $f_\theta(x) = p_v(\cdot) * p_\theta(\cdot)$, where $p_\theta(\cdot)$ represents the pdf of $\Theta = \theta$. A noise modified observation y is obtained via adding an additive independent noise n to the observation x, i.e., $y = x + n$. If the additive noise n is a constant vector, the pdf of y under H_0 is calculated as $f_0(y) = f_0(x - n)$, and the pdf of y under H_1 for $\Theta = \theta$ is $f_\theta(y) = f_\theta(x - n)$. In addition, a linear- quadratic detector is utilized here, given by:

$$T(y) = d_0 y^2 + d_1 y + d_2 \underset{\underset{H_0}{<}}{\overset{\overset{H_1}{>}}{}} \gamma \tag{40}$$

where d_0, d_1 and d_2 are detector parameters, and γ denotes the detection threshold. In the numerical examples, $\alpha_o = \alpha^x$ and $\beta_o = \beta^x$, where α^x and β^x are the original type I and II error probabilities, respectively.

4.1. Rayleigh Distribution Background Noise

Suppose that $\Theta = s$ is a constant, the problem shown in (39) represents a simple binary hypothesis testing problem. Here, we set $d_0 = d_2 = 0$ and $d_1 = 1$, then the detector becomes

$$T(y) = y \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{41}$$

It is assumed that the background noise v obeys the mixture of Rayleigh distributions with zero-means such that $p_v(v) = \sum_{i=1}^{M} m_i \varphi_i(v - \mu_i)$, where $m_i \geq 0$ for $i = 1, \ldots, M$, $\sum_{i=1}^{M} m_i = 1$, and

$$\varphi_i(x) = \begin{cases} \frac{x}{\sigma_i^2} \exp(-\frac{x^2}{2\sigma_i^2}), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{42}$$

In the simulations, the variances of all the Rayleigh components are assumed to be the same, i.e., $\sigma_i = \sigma$ for $i = 1, \ldots, M$. In addition, the parameters are specified as $M = 4$, $u_1 = 0.2$, $u_2 = 0.4$, $u_3 = -2\sigma\sqrt{\frac{\pi}{2}} - 0.2$, $u_4 = -2\sigma\sqrt{\frac{\pi}{2}} - 0.4$ and $m_i = 0.25$ for $i = 1, \ldots, 4$. From (33) and (34), the noise modified type I error probability $A_0(n)$ and type II error probability $B_1(n)$ obtained by adding a constant vector n is calculated as:

$$A_0(n) = \int_{\gamma}^{+\infty} f_0(y)dy = \sum_{i=1}^{4} m_i \Phi(\gamma - \mu_i - n) \tag{43}$$

$$B_1(n) = \int_{-\infty}^{\gamma} f_1(y)dy = 1 - \sum_{i=1}^{4} m_i \Phi(\gamma - \mu_i - n - s) \tag{44}$$

where $\Phi(x) = \int_x^{\infty} \frac{x}{\sigma^2} \exp(-\frac{x^2}{2\sigma^2})dt$, when $x > 0$; $\Phi(x) = 1$, when $x \leq 0$. Accordingly, $\alpha^x = A_0(0) = \sum_{i=1}^{4} m_i \Phi(\gamma - \mu_i)$ and $\beta^x = B_1(0) = 1 - \sum_{i=1}^{4} m_i \Phi(\gamma - \mu_i - s)$. Let $c_1 = \beta^x / (\alpha^x + \beta^x)$ and $c_2 = \alpha^x / (\alpha^\mathbf{x} + \beta^x)$, the noise modified weighted sum of the two types of error probabilities obtained via adding a constant vector is $Er(n) = c_1 \sum_{i=1}^{4} m_i \Phi(\gamma - \mu_i - n) + c_2 \sum_{i=1}^{4} m_i \Phi(\gamma - \mu_i - n - s)$.

From Section 3.2, the pdf of the optimal additive noise that minimizes weighted sum of type I and II error probabilities is denoted by $p_\mathbf{n}^{opt}(\mathbf{n}) = \eta \delta(\mathbf{n} - \mathbf{n}_1) + (1 - \eta)\delta(\mathbf{n} - \mathbf{n}_2)$, under the two constraints that $\alpha^y \leq \alpha^x$ and $\beta^y \leq \beta^x$. Moreover, the optimal additive noise for the case without any constraints is a constant vector.

Figure 1 plots the minimum noise modified weighted sums of type I and II error probabilities obtained under no constraint and two constraints that $\alpha^y \leq \alpha^x$ and $\beta^y \leq \beta^x$, and the original weighted sum without adding any noise for different values of σ when $s = 3$ and $\gamma = s/2$. When $\sigma \leq \sigma_1$, there is no noise that decreases the weighted sum. With the increase of σ, noise exhibits a positive effect on the detection performance. To be specific, when $\sigma_1 < \sigma < \sigma_2$, the weighted sum can be decreased by adding a constant vector for the no constraint case. When $\sigma > \sigma_2$, the weighted sum can be decreased adding the noise under two constraints. The noise modified weighted sum obtained without any constraints is less than or equal to that obtained under the two constraints, and the difference between them first decreases to zero for $\sigma_3 < \sigma < \sigma_4$ and then gradually increases when $\sigma > \sigma_4$. In addition, once σ exceeds a certain value, no noise exists that can decrease the weighted sum for any cases.
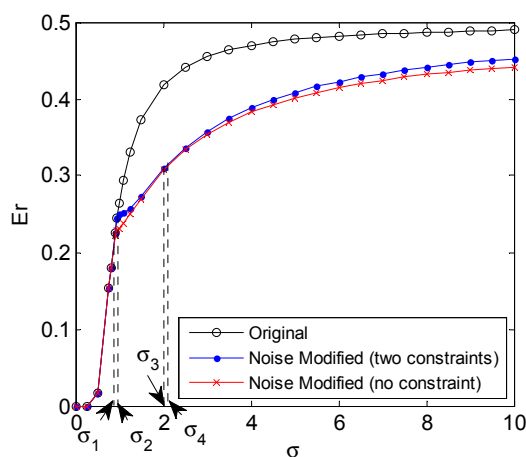
Figure 1. The minimum noise modified weighted sums of the type I and II error probabilities obtained under no constraint and two constraints, and the original weighted sum for different σ when $s = 3$ and $\gamma = s/2$.

Figure 2 shows the type I and II error probabilities corresponding to the weighted sum in Figure 1. From both Figures 1 and 2, it is observed that one of the noise modified Type I and II error probabilities performs worse than the original one for the no constraint case. Therefore, though the noise modified weighted sum obtained with no constraint is less than that obtained under the two constraints, the corresponding noise is actually not suitable to add to the observation. Furthermore, when the minimum value of the noise modified weighted sum is obtained under the two constraints, the corresponding type II probability equals to the original one and the type I probability achieves the minimum for $\sigma_2 < \sigma < \sigma_3$. Conversely, when $\sigma > \sigma_4$, the corresponding type I probability equals to the original one and the type II probability achieves the minimum. The results are consistent with part (3) and part (4) in Theorem 5. Especially, for $\sigma_3 < \sigma < \sigma_4$, the minimum values of the noise modified weighted sum obtained under no constraint is equal to that obtained under two constraints, and the corresponding type I and II error probabilities are the same, which also agrees with part (2) in Theorem 5. In order to further illustrate the results in Figures 1 and 2, Table 1 provides the optimal additive noises added for the two different cases.
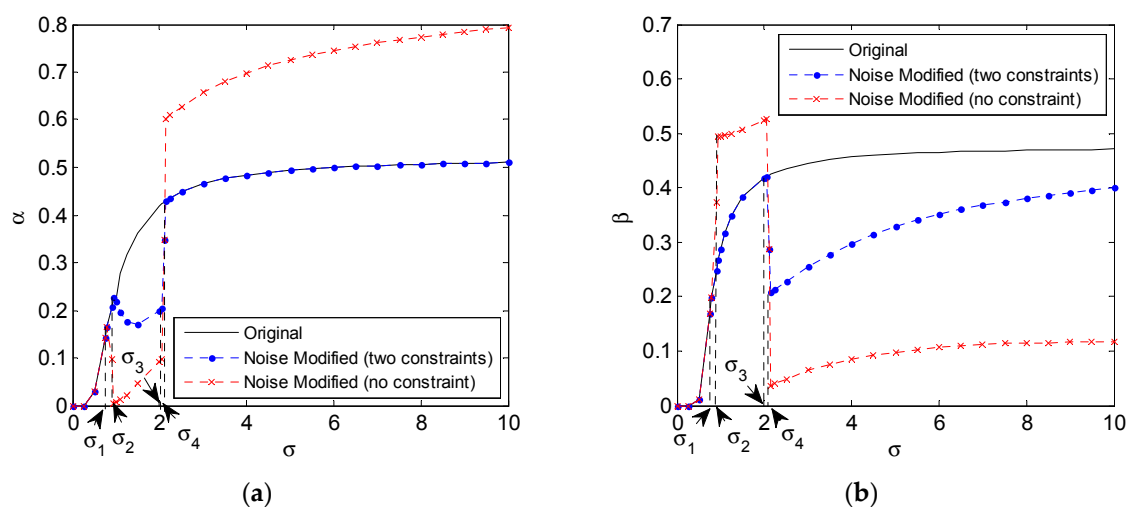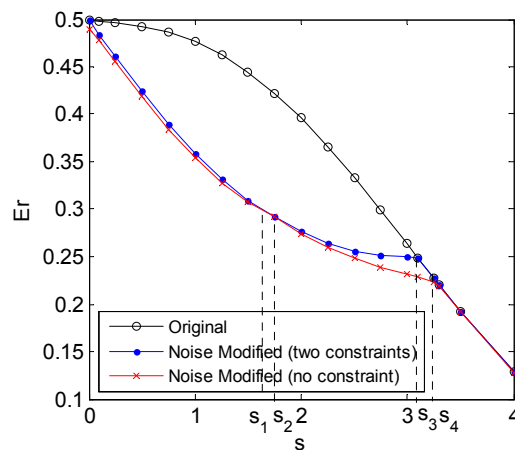

Figure 2. The type I (**a**) and II (**b**) error probabilities corresponding to the weighted sum in Figure 1.

Table 1. The optimal additive noises that minimize the weighted sum under two constraints and no constraint for various σ where $s = 3$ and $\gamma = s/2$.

σ	Two Constraints			No Constraints
	n_1	n_2	η	n_o
0.950	-	-	-	-1.7089
1.250	-1.9082	1.7963	0.6950	-1.9218
2.125	-2.5136	3.1896	0.7862	$-2.5136/3.1896$
3.000	-3.3771	4.6942	0.3770	4.7449

Figure 3 depicts the minimum noise modified weighted sums of the type I and II error probabilities versus s for the cases of no constraint and two constraints, and the original weighted sum, when $\sigma = 1$ and $\gamma = s/2$. The corresponding type I and II error probabilities are depicted in Figure 4a,b, respectively. It is seen in Figure 3, the improvement of the weighted sum obtained by adding noise first increases and then decreases with the increase of s, and finally they all converge to the same value. The differences for the cases with and without constraints are very small in most cases. In the small interval of s, i.e., $s \in (s_1, s_2)$, the difference even decreases to zero. On the other hand, the noise modified type I error probability obtained under no constraint is significantly greater than the original one for $s < s_1$, while the corresponding type II error probability is less than that obtained under the two constraints. The situation, however, is reversed for $s_2 < s < s_3$. When $s > s_3$, there is no noise that decreases the weighted sum under the two constraints, while the weighted sum is still decreased by adding a constant vector for no constraint case. When $s > s_4$, the weighted sum cannot be decreased by adding any noise for all the cases. Furthermore, Table 2 shows the optimal additive noises that minimize the weighted sum under the cases of no and two constraints.

Figure 3. The minimum noise modified weighted sums of the type I and II error probabilities obtained under no constraint and two constraints, and the original weighted sum for different s when $\sigma = 1$ and $\gamma = s/2$.

Table 2. The optimal additive noises that minimize the weighted sum under two constraints and no constraint for various s where $\sigma = 1$ and $\gamma = s/2$.

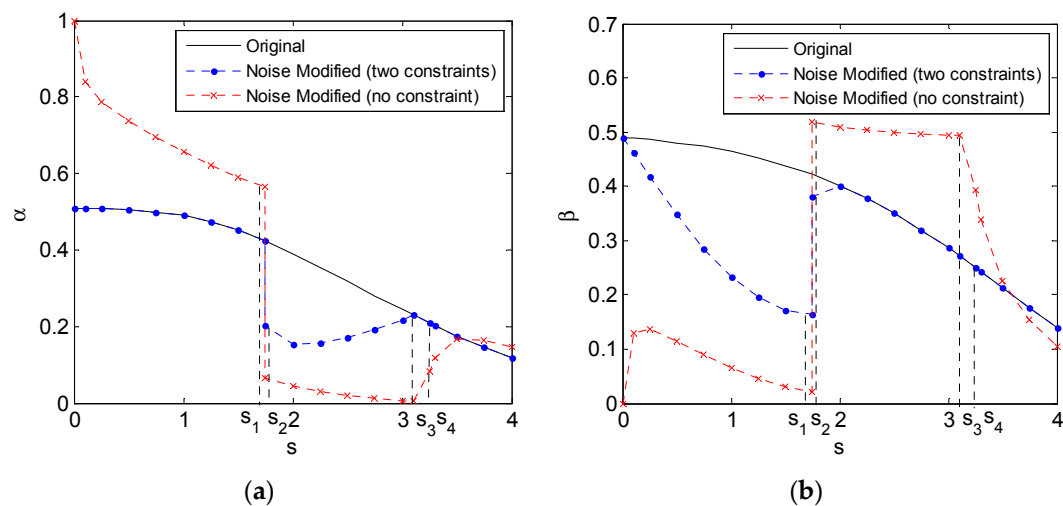| s	Two Constraints			No Constraints
	n_1	n_2	η	n_o
1.25	-1.3682	1.7327	0.2918	1.7474
1.75	-1.4408	1.6563	0.7265	$-1.4408/1.6563$
2.5	-1.6052	1.4690	0.6983	-1.6201
3.25	-	-	-	-0.5866

(**a**) (**b**)

Figure 4. The type I and II error probabilities corresponding to the weighted sum in Figure 3 are shown in (**a**) and (**b**), respectively.

Figure 5 shows the minimum noise modified weighted sums of type I and II error probabilities versus γ for the cases of no constraint and two constraints, and the original weighted sum, when $\sigma = 1$ and $s = 3$. The corresponding type I and II error probabilities are depicted in Figure 6a,b, respectively. As illustrated in Figure 5, when γ is close to zero, the original weighted sum Er^x approaches to zero. In such case, no additive noise exists to decrease the weighted sum. For the case of two constraints, the improvement of the weighted sum first increases for $\gamma_1 < \gamma < \gamma_2$ and then decreases for $\gamma_2 < \gamma < \gamma_3$, and no improvement can be obtained when $\gamma > \gamma_2$. On the other hand, the minimum noise modified weighted sum obtained under no constraint is smaller than that obtained under the two constraints for $\gamma_1 < \gamma < \gamma_3$, and the difference between them first increases and then decreases for both $\gamma_1 < \gamma < \gamma_2$ and $\gamma_2 < \gamma < \gamma_3$. When $\gamma > \gamma_3$, there still exists a constant vector that decreases the weighted sum, but it may be not a suitable noise in the practical application according to the type II probability depicted in Figure 6b. Furthermore, in order to study the results illustrated in Figures 5 and 6, Table 3 shows the optimal additive noises that minimize the weighted sum for the cases of no and two constraints.
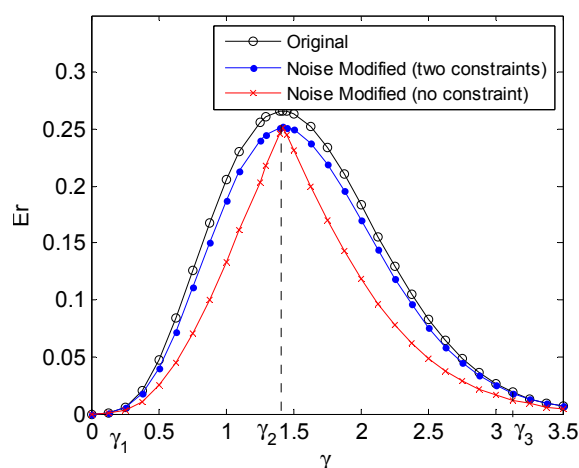

Figure 5. The minimum noise modified weighted sums of the type I and II error probabilities obtained under no constraint and two constraints, and the original weighted sum for different γ when $\sigma = 1$ and $s = 3$.
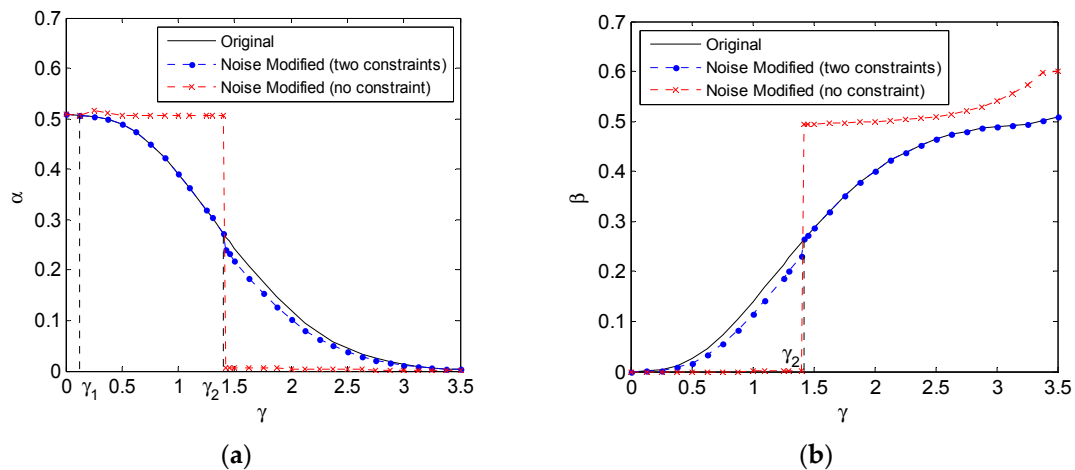
Figure 6. The type I and II error probabilities corresponding to the weighted sum in Figure 5 are shown in (**a**) and (**b**), respectively.

Table 3. The optimal additive noises that minimize the weighted sum under two constraints and no constraint for various γ where $\sigma = 1$ and $s = 3$.

γ	Two Constraints			No Constraints
	n_1	n_2	η	n_o
0.050	-	-	-	-
1.100	-2.1213	0.9341	0.2878	0.9691
1.425	-1.7947	1.2585	0.5355	-1.7957
2.250	-0.9693	2.0836	0.8867	-1.1763
3.375	-	-	-	-0.5775

4.2. Gaussian Mixture Background Noise

Suppose that Θ is a random variable with following pdf:

$$\omega_1(\theta) = \rho\delta(\theta - s) + (1 - \rho)\delta(\theta + s) \tag{45}$$

Therefore, we have $\Omega_0 = \{0\}$ and $\Omega_1 = \{s, -s\}$. In the simulations, we set $d_0 = 1$, $d_1 = 0$, $d_2 = -s^2/4$ and $\gamma = 0$, the detector is expressed as:

$$T(y) = y^2 - \frac{s^2}{4} \underset{H_0}{\overset{H_1}{\underset{<}{\gtrless}}} 0 \tag{46}$$

Moreover, we assume that v is a zero-mean symmetric Gaussian mixture noise with pdf of $p_v(v) = \sum_{i=1}^{M} m_i\psi_i(v - \mu_i)$, where $m_i \geq 0$, $\sum_{i=1}^{M} m_i = 1$ and:

$$\psi_i(v) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(-\frac{v^2}{2\sigma_i^2}) \tag{47}$$

Let $M = 4$ and the mean values of the symmetric Gaussian components are set as $[0.05\ 0.52\ -0.52\ -0.05]$ with corresponding weights $[0.35\ 0.15\ 0.15\ 0.35]$. In addition, the variances of Gaussian components are the same, i.e., $\sigma_i = \sigma$ for $i = 0, \ldots, 4$. According to (12) and (14), the noise modified type I error probability obtained by adding a constant vector n to x is calculated by:

$$A_0(n) = \sum_{i=1}^{4} m_i (\Psi(\frac{s/2 + \mu_i + n}{\sigma}) + \Psi(\frac{s/2 - \mu_i - n}{\sigma})) \tag{48}$$

and the corresponding type II error probabilities for $\Theta = s$ and $-s$ are respectively calculated as:

$$B_s(n) = 1 - \sum_{i=1}^{4} m_i (\Psi(\frac{3s/2 + \mu_i + n}{\sigma}) + \Psi(\frac{-s/2 - \mu_i - n}{\sigma})) \tag{49}$$

$$B_{-s}(n) = 1 - \sum_{i=1}^{4} m_i (\Psi(\frac{-s/2 + \mu_i + n}{\sigma}) + \Psi(\frac{3s/2 - \mu_i - n}{\sigma})) \tag{50}$$

where $\Psi(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2\sigma^2}) dt$. Accordingly:

$$B_1(n) = \rho B_s(n) + (1 - \rho) B_{-s}(n) \tag{51}$$

Therefore, the original type I and type II error probabilities for $\Theta = s$ and $-s$ are $\alpha^x(\phi; 0) = A_0(0)$, $\beta^x(\phi; s) = B_s(0)$ and $\beta^x(\phi; -s) = B_{-s}(0)$, respectively.

Due to the symmetry property of v, one obtains $B_s(0) = B_{-s}(0)$. In this case, the original average type II error probability is $\beta^x = B_1(0) = \rho \beta^x(\phi; s) + (1 - \rho)\beta^x(\phi; -s) = B_s(0) = B_{-s}(0)$. The noise modified weighted sum of type I and average type II error probabilities corresponding to the constant vector is expressed by $Er(n) = c_1 A_0(n) + c_2 B_1(n)$. The values of c_1 and c_2 are still specified as $\beta^x / (\alpha^x + \beta^x)$ and $\alpha^x / (\alpha^x + \beta^x)$, respectively. From Theorem 4 in Section 2.3, the optimal additive noise that minimizes the weighted sum is a randomization with a pdf of $p_n^{opt}(n) = \eta_1 \delta(n - n_1) + \eta_2 \delta(n - n_2) + \eta_3 \delta(n - n_3)$, where $\eta_i \geq 0$ for $i = 1, \ldots, 3$, and $\sum_{i=1}^{3} \eta_i = 1$.

Figure 7 shows the detection performance of the original detector and the noise enhanced detector that minimizes the weighted sum of type I and average type II error probabilities under the constraints that $\alpha^y(\phi; 0) \leq \alpha_o$ and $\max_{\theta \in \Omega_1} \beta^y(\phi; \theta) \leq \beta_o$, for different values of σ where $s = 1$ and $\rho = 0.6$. The minimum achievable noise modified weighted sum is plotted in Figure 7a, and the corresponding type I error probability and type II error probabilities for $\Theta = s$ and $-s$ are depicted in Figure 7b–d, respectively.

From Figure 7, the original weighted sums, type I error probabilities, and type II error probabilities for $\Theta = s$ and $-s$ increase as σ decreases towards zero. In Figure 7a, when σ is close to zero, the weighted sum can be decreased significantly. With the increase of σ, the improvement obtained by adding noise is reduced gradually to zero. In other words, the phenomenon of noise-enhanced detection performance cannot occur when σ exceeds a certain value. In Figure 7b, the noise modified type I error probability stays at 0.1500 for $\sigma < 0.07$ and then increases gradually to equal to the original type I error probability. Moreover, the noise modified type II error probabilities for $\Theta = s$ corresponding to the minimum weighted sum increases from zero to that of original detector, shown in Figure 7c, while the type II error probabilities for $\Theta = -s$ of the noise enhanced detector is equal to that of the original detector all the time. In fact, the type II error probability for $\Theta = s$ also reaches the minimums under the constraints that $\alpha^y(\phi; 0) \leq \alpha_o$ and $\max_{\theta \in \Omega_1} \beta^y(\phi; \theta) \leq \beta_o$ in this example. In addition, Table 4 offers the optimal additive noises that minimize the weighted sum for different values of σ to explain the results in Figure 7. It should be noted that the optimal noise is not unique.y.
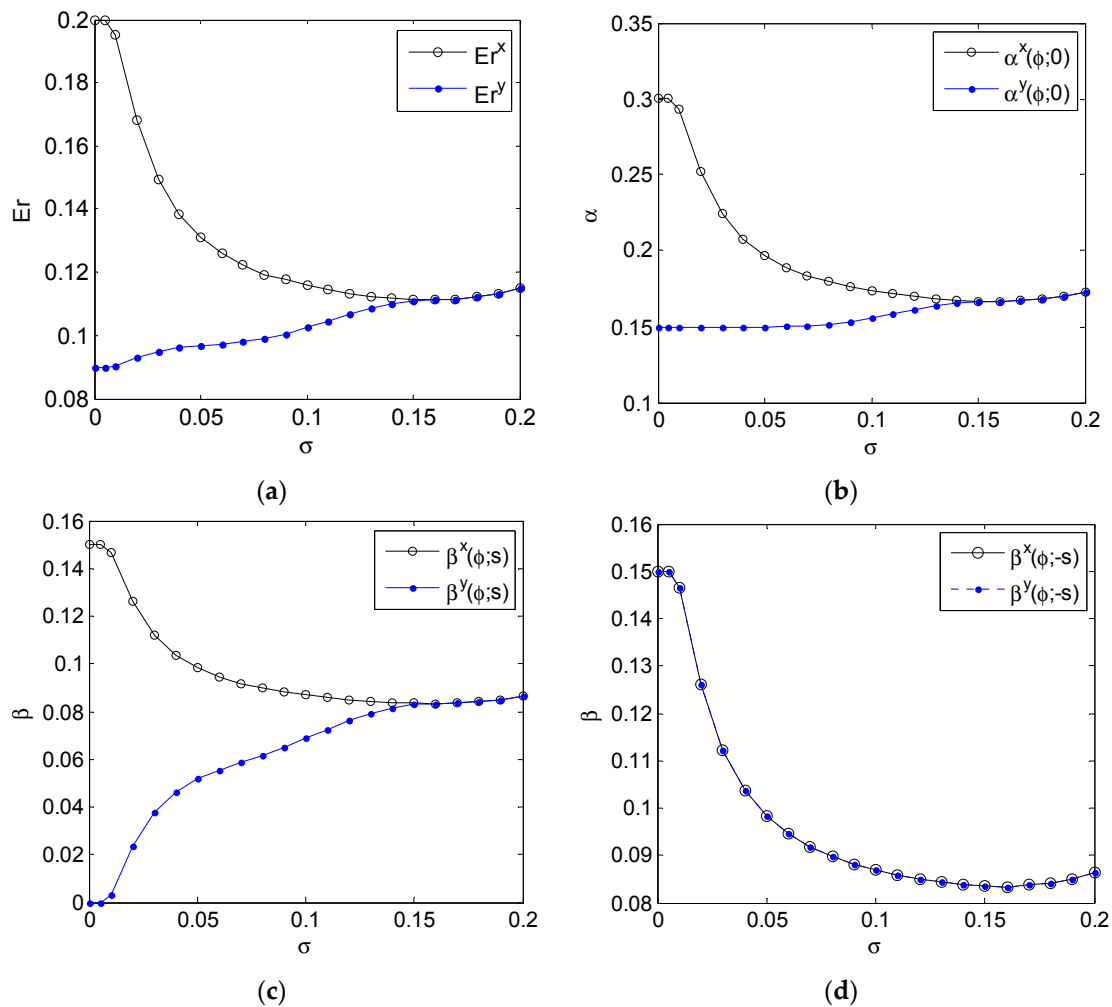
Figure 7. The weighted sums, type I error probabilities, and type II error probabilities for $\Theta = s$ and $-s$ of the original detector and the noise enhanced detector for different σ where $s = 1$ and $\rho = 0.6$ shown in (**a**), (**b**), (**c**) and (**d**), respectively

Table 4. The optimal additive noises that minimize the weighted sum under two constraints for various σ where $s = 1$ and $\rho = 0.6$.

σ	n_1	n_2	n_3	η_1	η_2	η_3
0.0001	0.2286	-	-	1.0000	-	-
0.02	0.2286	-0.2255	-	0.8413	0.1587	-
0.05	0.2287	-0.2208	0.2421	0.5310	0.3446	0.1244
0.08	0.2180	-0.2185	-0.2168	0.5943	0.2449	0.1608

Figure 8a demonstrates the weighted sums of type I and average type II error probabilities of the original detector and the noise enhanced detector versus s, where $\sigma = 0.08$ and $\rho = 0.6$. The corresponding type I error probability and type II error probabilities for $\Theta = s$ and $-s$ are depicted in Figure 8b–d, respectively. From Figure 8a, the weighted sum cannot be decreased under the constraints on different error probabilities for $s < s_1$ and $s < s_2$. Conversely, there exists additive noise under the constraints that reduces the weighted sum for $s_1 < s < s_2$, and the corresponding improvement first increases and then decreases with the increase of s. Comparing Figure 8b with Figure 8a, it is noted that the change of the noise modified type I error probability is similar to that of the noise modified weighted sum. In Figure 8c, the noise modified type II error probability for $\Theta = s$

first decreases to the minimum and then increases as *s* increases, while the type II error probability for $\Theta = -s$ of the noise modified detector is always equal to that of the original detector, shown in Figure 8d. In addition, in order to further illustrate the results in Figure 8, Table 5 shows the optimal noises that minimize the weighted sum under the case of two constraints.
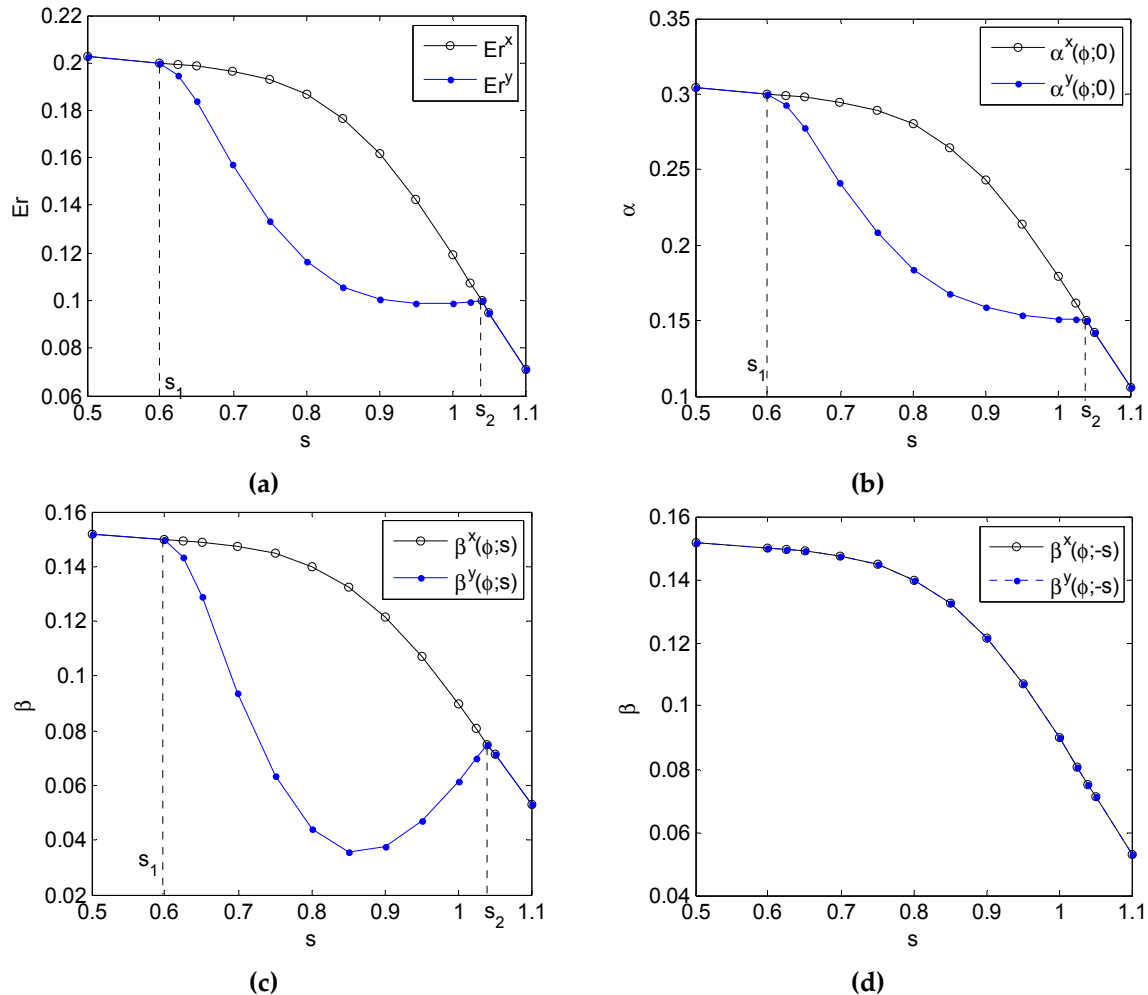

(a)

(b)

(c)

(d)

Figure 8. The weighted sums, type I error probabilities, and type II error probabilities for $\Theta = s$ and $-s$ of the original detector and the noise enhanced detector for different *s* where $\sigma = 0.08$ and $\rho = 0.6$ shown in (a), (b), (c) and (d), respectively.

Table 5. The optimal additive noises that minimize the weighted sum under two constraints for various *s* where $\sigma = 0.08$ and $\rho = 0.6$.

s	n_1	n_2	n_3	η_1	η_2	η_3
0.65	0.1613	−0.1613	-	0.6267	0.3733	-
0.75	0.2026	−0.2026	-	0.7949	0.2051	-
0.85	0.2148	−0.2149	-0.2150	0.8262	0.1300	0.0438
0.95	0.2195	−0.2196	-0.2190	0.7006	0.1916	0.1078

5. Conclusions

In this paper, a noise-enhanced detection problem has been investigated for a general composite hypothesis testing. Under the constraints of type I and II error probabilities, the minimization of the weighted sum of average type I and II error probabilities has been explored by adding an additive

independent noise. The sufficient conditions for improvability of the weighted sum are provided, and a simple algorithm to search the optimal noise is developed. Then some additional theoretical results are made based on the specificity of the binary simple hypothesis testing problem. The studies on different noise distributions confirm the theoretical analysis that the optimal additive noise indeed minimizes the weighted sum under certain conditions. To be noted that, theoretical results can also be extended to a broad class of noise enhanced optimization problems under two inequality constraints such as the minimization of Bayes risk under the different constraints of condition risks for a binary hypothesis testing problem.

Author Contributions: Shujun Liu raised the idea of the framework to solve noise enhanced detection problems. Ting Yang and Shujun Liu contributed to the drafting of the manuscript, interpretation of the results, some experimental design and checked the manuscript. Kui Zhang contributed to develop the algorithm of finding the corresponding optimal noise and Ting Yang contributed to the proofs of the theories developed in this paper. All authors have read and approved the final manuscript.

Appendix A. Proof of Theorem 1

Proof. Due to the convexity of $A_{\theta^*}(\mathbf{n})$ and according to the Jensen's inequality, the type I error probability in (4) is calculated as:

$$\alpha^{\mathbf{y}}(\phi; \theta^*) = E\{A_{\theta^*}(\mathbf{n})\} \geq A_{\theta^*}(E\{\mathbf{n}\}) \tag{A1}$$

The contradiction method is utilized to prove this theorem. Suppose that the detector can be improved by adding noise. The improvability means that $\alpha^{\mathbf{y}}(\phi; \theta^*) \leq \alpha_o$ for any $\theta^* \in \Omega_0$, and then $A_{\theta^*}(E\{\mathbf{n}\}) \leq \alpha_o$ from (A1). Since $E\{\mathbf{n}\} \in P_{\mathbf{n}}$, $A_{\theta^*}(E\{\mathbf{n}\}) \leq \alpha_o$ implies $Er(E\{\mathbf{n}\}) \geq Er(0)$ based on the assumption in Theorem 1, (15) can be recalculated as:

$$Er^{\mathbf{y}} = E\{Er(\mathbf{n})\} \geq Er(E\{\mathbf{n}\}) \geq Er(0) = Er^{\mathbf{x}} \tag{A2}$$

where the first inequality holds according to the convexity of $Er(\mathbf{n})$. From (A1) and (A2), the inequality $Er^{\mathbf{y}} < Er^{\mathbf{x}}$ cannot be achieved by adding any noise under the conditions presented in Theorem 1. Therefore, the detector is nonimprovable, which contradicts the assumption. Similarly, the alternative conditions for nonimprovability stated in the parentheses can also be proved. □

Appendix B. Proof of Theorem 2

Proof. According to the definitions in (9) and (10), improvability for a detector means that there exists at least one pdf $p_{\mathbf{n}}(\mathbf{n})$ to satisfy three conditions, i.e., $Er^{\mathbf{y}} = \int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) Er(\mathbf{n}) d\mathbf{n} < Er(0)$, $\int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) A_\theta(\mathbf{n}) d\mathbf{n} \leq \alpha_o$ for any $\theta \in \Omega_0$ and $\int_{\mathbb{R}^N} p_{\mathbf{n}}(\mathbf{n}) B_\theta(\mathbf{n}) d\mathbf{n} \leq \beta_o$ for any $\theta \in \Omega_1$. Suppose that the noise pdf $p_{\mathbf{n}}(\mathbf{n})$ consists of L infinitesimal noise components, i.e., $p_{\mathbf{n}}(\mathbf{n}) = \sum_{l=1}^{L} \lambda_l \delta(\mathbf{n} - \boldsymbol{\varepsilon}_l)$. The three conditions can be rewritten as follows:

$$Er^{\mathbf{y}} = \sum_{l=1}^{L} \lambda_l Er(\boldsymbol{\varepsilon}_l) < Er(0) \tag{A3}$$

$$\sum_{l=1}^{L} \lambda_l A_\theta(\boldsymbol{\varepsilon}_l) \leq \alpha_o, \forall \theta \in \Omega_0 \tag{A4}$$

$$\sum_{l=1}^{L} \lambda_l B_\theta(\boldsymbol{\varepsilon}_l) \leq \beta_o, \forall \theta \in \Omega_1 \tag{A5}$$

Since ε_l, $l = 1,\ldots,L$, is infinitesimal, $Er(\varepsilon_l)$, $A_\theta(\varepsilon_l)$ and $B_\theta(\varepsilon_l)$ can be expressed approximately with Taylor series expansion as $Er(0) + \varepsilon_l^T \mathbf{Er} + 0.5\varepsilon_l^T \mathbf{H}\varepsilon_l$, $A_\theta(0) + \varepsilon_l^T \mathbf{A}_\theta + 0.5\varepsilon_l^T \mathbf{H}_\theta^A \varepsilon_l$ and $B_\theta(0) + \varepsilon_l^T \mathbf{B}_\theta + 0.5\varepsilon_l^T \mathbf{H}_\theta^B \varepsilon_l$, where $\mathbf{Er}(\mathbf{A}_\theta,\mathbf{B}_\theta)$ and $\mathbf{H}(\mathbf{H}_\theta^A,\mathbf{H}_\theta^B)$ are the gradient and the Hessian matrix of $Er(\mathbf{x})(A_\theta(\mathbf{x}),B_\theta(\mathbf{x}))$ around $\mathbf{x} = 0$, respectively. Therefore, (A3)–(A5) are rewritten as:

$$\sum_{l=1}^{L} \lambda_l \varepsilon_l^T \mathbf{Er} + 0.5\sum_{l=1}^{L} \lambda_l \varepsilon_l^T \mathbf{H}\varepsilon_l < 0 \tag{A6}$$

$$\sum_{l=1}^{L} \lambda_l \varepsilon_l^T \mathbf{A}_\theta + 0.5\sum_{l=1}^{L} \lambda_l \varepsilon_l^T \mathbf{H}_\theta^A \varepsilon_l \leq \alpha_o - A_\theta(0), \forall\theta \in \Omega_0 \tag{A7}$$

$$\sum_{l=1}^{L} \lambda_l \varepsilon_l^T \mathbf{B}_\theta + 0.5\sum_{l=1}^{L} \lambda_l \varepsilon_l^T \mathbf{H}_\theta^B \varepsilon_l \leq \beta_o - B_\theta(0), \forall\theta \in \Omega_1 \tag{A8}$$

Let ε_l be expressed by $\varepsilon_l = \tau_l \mathbf{g}$, where $\mathbf{g}$ is a N-dimensional real vector and τ_l is an infinitesimal real value, $l = 1,\ldots,L$. Accordingly, one obtains:

$$\sum_{l=1}^{L} \lambda_l \tau_l \mathbf{g}^T \mathbf{Er} + 0.5\sum_{l=1}^{L} \lambda_l \tau_l^2 \mathbf{g}^T \mathbf{H}\mathbf{g} < 0 \tag{A9}$$

$$\sum_{l=1}^{L} \lambda_l \tau_l \mathbf{g}^T \mathbf{A}_\theta + 0.5\sum_{l=1}^{L} \lambda_l \tau_l^2 \mathbf{g}^T \mathbf{H}_\theta^A \mathbf{g} \leq \alpha_o - A_\theta(0), \forall\theta \in \Omega_0 \tag{A10}$$

$$\sum_{l=1}^{L} \lambda_l \tau_l \mathbf{g}^T \mathbf{B}_\theta + 0.5\sum_{l=1}^{L} \lambda_l \tau_l^2 \mathbf{g}^T \mathbf{H}_\theta^B \mathbf{g} \leq \beta_o - B_\theta(0), \forall\theta \in \Omega_1 \tag{A11}$$

Based on the definitions given in (19)–(24), (A9)–(A11) are simplified as:

$$(k \cdot er^{(1)}(\mathbf{x}, \mathbf{g}) + er^{(2)}(\mathbf{x}, \mathbf{g}))\big|_{\mathbf{x}=0} < 0 \tag{A12}$$

$$(k \cdot a_\theta^{(1)}(\mathbf{x}, \mathbf{g}) + a_\theta^{(2)}(\mathbf{x}, \mathbf{g}))\big|_{\mathbf{x}=0} < \frac{2(\alpha_o - A_\theta(0))}{\sum_{l=1}^{L} \lambda_l \tau_l^2}, \forall\theta \in \Omega_0 \tag{A13}$$

$$(k \cdot b_\theta^{(1)}(\mathbf{x}, \mathbf{g}) + b_\theta^{(2)}(\mathbf{x}, \mathbf{g}))\big|_{\mathbf{x}=0} < \frac{2(\beta_o - B_\theta(0))}{\sum_{l=1}^{L} \lambda_l \tau_l^2}, \forall\theta \in \Omega_1 \tag{A14}$$

where $k = 2\sum_{l=1}^{L} \lambda_l \tau_l / \sum_{l=1}^{L} \lambda_l \tau_l^2$. As $\alpha_o = A_\theta(0)$ for $\theta \in \Lambda_0$ and $\alpha_o > A_\theta(0)$ for $\theta \in \Omega_0\Lambda_0$, the right-hand side of (A13) approaches to plus infinity for $\theta \in \Omega_0\Lambda_0$. Similarly, when $\beta_o = B_\theta(0)$ for $\theta \in \Lambda_1$ and $\beta_o > B_\theta(0)$ for $\theta \in \Omega_1\Lambda_1$, the right-hand side of (A14) also goes to plus infinity for $\theta \in \Omega_1\Lambda_1$. Therefore, we only need to consider the cases of $\theta \in \Lambda_0$ and $\theta \in \Lambda_1$. In doing so, (A12)–(A14) are now:

$$(k \cdot er^{(1)}(\mathbf{x}, \mathbf{g}) + er^{(2)}(\mathbf{x}, \mathbf{g}))\big|_{\mathbf{x}=0} < 0 \tag{A15}$$

$$(k \cdot a_\theta^{(1)}(\mathbf{x}, \mathbf{g}) + a_\theta^{(2)}(\mathbf{x}, \mathbf{g}))\big|_{\mathbf{x}=0} < 0, \forall\theta \in \Lambda_0 \tag{A16}$$

$$(k \cdot b_\theta^{(1)}(\mathbf{x}, \mathbf{g}) + b_\theta^{(2)}(\mathbf{x}, \mathbf{g}))\big|_{\mathbf{x}=0} < 0, \forall\theta \in \Lambda_1 \tag{A17}$$

It is obvious that k can be set as any real value by choosing appropriate λ_l and τ_l. As a result, (A15)–(A17) can be satisfied by selecting a suitable value of k under each condition in Theorem 2. That is:

(1) Inequalities (A15)–(A17) can be satisfied by setting k as a sufficiently large positive number, if $er^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$, $a_{\theta_0}^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$, $b_{\theta_1}^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$ hold.

(2) Inequalities (A15)–(A17) can be satisfied by setting k as a sufficiently large negative number, if $er^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} > 0$, $ea_{\theta_0}^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} > 0$, $b_{\theta_1}^{(1)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} > 0$ hold.

(3) Inequalities (A15)–(A17) can be satisfied by setting k as zero, if $er^{(2)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$, $a_{\theta_0}^{(2)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$, $b_{\theta_1}^{(2)}(\mathbf{x}, \mathbf{g})\big|_{\mathbf{x}=0} < 0$ hold. $\square$

Appendix C. Proof of Theorem 3

Proof. Since $I(t)$ and $S(t)$ are second-order continuously differentiable around $t = \widetilde{\alpha}$, there exists a $\xi > 0$ such that $I''(\cdot) < 0$ and $S''(\cdot) < 0$ for $\Delta = (\widetilde{\alpha} - \xi, \widetilde{\alpha} + \xi)$. If one adds a noise with pdf $p_{\hat{\mathbf{n}}}(\mathbf{n}) = 0.5\delta(\mathbf{n} - \mathbf{n}_1) + 0.5\delta(\mathbf{n} - \mathbf{n}_2)$, where $\max_{\theta \in \Omega_0} A_\theta(\mathbf{n}_1) = \widetilde{\alpha} + \xi$ and $\max_{\theta \in \Omega_0} A_\theta(\mathbf{n}_2) = \widetilde{\alpha} - \xi$, to the original observation $\mathbf{x}$, the maximum values of corresponding noise modified type I and II error probabilities are:

$$\max_{\theta \in \Omega_0} E\{A_\theta(\hat{\mathbf{n}})\} \leq E\left\{\max_{\theta \in \Omega_0} A_\theta(\hat{\mathbf{n}})\right\} \leq 0.5(\widetilde{\alpha} + \xi) + 0.5(\widetilde{\alpha} - \xi) = \widetilde{\alpha} \leq \alpha_o \tag{A18}$$

$$\max_{\theta \in \Omega_1} E\{B_\theta(\hat{\mathbf{n}})\} \leq E\left\{\max_{\theta \in \Omega_1} B_\theta(\hat{\mathbf{n}})\right\} \leq 0.5S(\widetilde{\alpha} + \xi) + 0.5S(\widetilde{\alpha} - \xi) \leq S(\widetilde{\alpha}) = \widetilde{\beta} \leq \beta_o \tag{A19}$$

In addition:

$$E\{Er(\hat{\mathbf{n}})\} = E\{I(t)\} = 0.5I(\widetilde{\alpha} + \xi) + 0.5I(\widetilde{\alpha} - \xi) < I(\widetilde{\alpha}) \tag{A20}$$

One obtains $E\{Er(\hat{\mathbf{n}})\} < Er(0)$ because $I(\widetilde{\alpha}) \leq Er(0)$ according to the definition of $I(t)$. As a result, the detector is improvable. $\square$

Appendix D. Proof of Theorem 5

Proof. Part (1): If $Q_e \cap Q_1 \neq \varnothing$, any $\mathbf{n}_1 \in Q_e \cap Q_1$ satisfies the constraints of $A_0(\mathbf{n}_1) \leq \alpha_o$ and $B_1(\mathbf{n}_1) \leq \beta_o$ based on the definition of Q_1 and $Er_{opt}^{\mathbf{y}} = Er(\mathbf{n}_1) = \min_{\mathbf{n}} Er(\mathbf{n}) < Er(0)$ according to the definition of Q_e.

Part (2): If $Q_e \cap Q_2 \neq \varnothing$ and $Q_e \cap Q_3 \neq \varnothing$ simultaneously, there exists $\mathbf{n}_1 \in Q_e \cap Q_2$ and $\mathbf{n}_2 \in Q_e \cap Q_3$ such that $Er(\mathbf{n}_1) = Er(\mathbf{n}_2) = \min_{\mathbf{n}} Er(\mathbf{n})$ based on the definition of Q_e. In order to meet the constraints that $E\{A_0(\mathbf{n})\} \leq \alpha_o$ and $E\{B_1(\mathbf{n})\} \leq \beta_o$, the noise components η, $\mathbf{n}_1$ and $\mathbf{n}_2$ should satisfy the following two inequalities:

$$\eta A_0(\mathbf{n}_1) + (1 - \eta)A_0(\mathbf{n}_2) \leq \alpha_o \tag{A21}$$

$$\eta B_1(\mathbf{n}_1) + (1 - \eta)B_1(\mathbf{n}_2) \leq \beta_o \tag{A22}$$

Consequently, $\eta \geq \eta_1 = \frac{\alpha_o - A_0(\mathbf{n}_2)}{A_0(\mathbf{n}_1) - A_0(\mathbf{n}_2)}$ and $\eta \leq \eta_2 = \frac{\beta_o - B_1(\mathbf{n}_2)}{B_1(\mathbf{n}_1) - B_1(\mathbf{n}_2)}$ according to the definitions of Q_2 and Q_3. If $\eta_1 \leq \eta \leq \eta_2$, the noise with pdf $p_{\mathbf{n}}^{opt}(\mathbf{n}) = \eta\delta(\mathbf{n} - \mathbf{n}_1) + (1 - \eta)\delta(\mathbf{n} - \mathbf{n}_2)$ can minimize $E\{Er(\mathbf{n})\}$ and satisfy the two inequalities, and $Er_{opt}^{\mathbf{y}} = \eta Er(\mathbf{n}_1) + (1 - \eta)Er(\mathbf{n}_2) = \min_{\mathbf{n}} Er(\mathbf{n})$.

Part (3): If $Q_e \subset Q_2$, the optimal additive noise is not a constant vector, i.e., $\eta \neq 1$. Therefore, one of $\mathbf{n}_1$ and $\mathbf{n}_2$ belongs to Q_2 and the second one comes from Q_1 or Q_3. In addition, η, $\mathbf{n}_1$ and $\mathbf{n}_2$ should also satisfy the two constraints in (A21) and (A22).

First, suppose that $\mathbf{n}_1 \in Q_2$ and $\mathbf{n}_2 \in Q_1$, then (A21) holds based on the definitions of Q_1 and Q_2. We should only consider the constraint in (A22), which implies $\eta \leq \eta_2$. It is true that $A_0(\mathbf{n}_2) \leq \alpha_o$ and $B_1(\mathbf{n}_2) \leq \beta_o$ according to the definition of Q_1. If $Er(\mathbf{n}_1) > Er(\mathbf{n}_2)$, we have $Er(\mathbf{n}_2) < Er_{opt}^{\mathbf{y}} = \eta Er(\mathbf{n}_1) + (1 - \eta)Er(\mathbf{n}_2)$, which contradicts with the definition of $p_{\mathbf{n}}^{opt}(\mathbf{n})$. Hence, $Er(n_1) < Er(n_2)$ and the minimum of $E\{Er(\mathbf{n})\}$ is obtained when $\eta = \eta_2$.

Next, suppose that $\mathbf{n}_1 \in Q_2$ and $\mathbf{n}_2 \in Q_3$. The two inequalities in (A21) and (A22) require that $\eta_1 \leq \eta \leq \eta_2$. If $Er(\mathbf{n}_1) > Er(\mathbf{n}_2)$, the minimum of $E\{Er(\mathbf{n})\}$ is obtained when $\eta = \eta_1$. In such

case, there exists a noise with pdf $p_{\hat{\mathbf{n}}}(\mathbf{n}) = \varsigma p_{\mathbf{n}}^{opt}(\mathbf{n}) + (1 - \varsigma)\delta(\mathbf{n} - \mathbf{n}_e)$ that satisfies $E\{A_0(\hat{\mathbf{n}})\} \leq \alpha_o$ and $E\{B_1(\hat{\mathbf{n}})\} \leq \beta_o$ simultaneously, where $\mathbf{n}_e \in Q_e$ and $0 \leq \varsigma \leq 1$. Therefore, $E\{Er(\hat{\mathbf{n}})\} = \varsigma Er_{opt}^{\mathbf{y}} + (1 - \varsigma)Er(\mathbf{n}_e) < Er_{opt}^{\mathbf{y}}$ since $Er(\mathbf{n}_e) = \min_{\mathbf{n}} Er(\mathbf{n}) < Er_{opt}^{\mathbf{y}}$, which contradicts with the definition of $p_{\mathbf{n}}^{opt}(\mathbf{n})$. As a result, $Er(\mathbf{n}_1) < Er(\mathbf{n}_2)$ and the minimum of $E\{Er(\mathbf{n})\}$ is obtained when $\eta = \eta_2$.

When $\eta = \eta_2$, one obtains $E\{B_1(\mathbf{n})\} = \eta B_1(\mathbf{n}_1) + (1 - \eta)B_1(\mathbf{n}_2) = \beta_o$. In other words, the minimum of $E\{Er(\mathbf{n})\}$ is obtained when $E\{A_0(\mathbf{n})\}$ achieves the minimum and $E\{A_0(\mathbf{n})\}$. Accordingly, one obtains $Er_{opt}^{\mathbf{y}} = c_1\alpha_{opt}^{\mathbf{y}} + c_2\beta_o$.

Part (4): The proof of Part (4) is similar to that of Part (3) and it is omitted here. $\square$

References

1. DeGroot, M.H.; Sxhervish, M.J. *Probability and Statistics*, 4nd ed.; Addison-Wesley: Boston, MA, USA, 2011.
2. Pericchi, L.; Pereira, C. Adaptative significance levels using optimal decision rules: Balancing by weighting the error probabilities. *Braz. J. Probab. Stat.* **2016**, *30*, 70–90. [CrossRef]
3. Benzi, R.; Sutera, A.; Vulpiani, A. The mechanism of stochastic resonance. *J. Phys. A Math.* **1981**, *14*, 453–457. [CrossRef]
4. Patel, A.; Kosko, B. Noise benefits in quantizer-array correlation detection and watermark decoding. *IEEE Trans. Signal Process.* **2011**, *59*, 488–505. [CrossRef]
5. Han, D.; Li, P.; An, S.; Shi, P. Multi-frequency weak signal detection based on wavelet transform and parameter compensation band-pass multi-stable stochastic resonance. *Mech. Syst. Signal Process.* **2016**, *70–71*, 995–1010. [CrossRef]
6. Addesso, P.; Pierro, V.; Filatrella, G. Interplay between detection strategies and stochastic resonance properties. *Commun. Nonlinear Sci. Numer. Simul.* **2016**, *30*, 15–31. [CrossRef]
7. Gingl, Z.; Makra, P.; Vajtai, R. High signal-to-noise ratio gain by stochastic resonance in a double well. *Fluct. Noise Lett.* **2001**, *1*, L181–L188. [CrossRef]
8. Makra, P.; Gingl, Z. Signal-to-noise ratio gain in non-dynamical and dynamical bistable stochastic resonators. *Fluct. Noise Lett.* **2002**, *2*, L147–L155. [CrossRef]
9. Makra, P.; Gingl, Z.; Fulei, T. Signal-to-noise ratio gain in stochastic resonators driven by coloured noises. *Phys. Lett. A* **2003**, *317*, 228–232. [CrossRef]
10. Duan, F.; Chapeau-Blondeau, F.; Abbott, D. Noise-enhanced SNR gain in parallel array of bistable oscillators. *Electron. Lett.* **2006**, *42*, 1008–1009. [CrossRef]
11. Mitaim, S.; Kosko, B. Adaptive stochastic resonance in noisy neurons based on mutual information. *IEEE Trans. Neural Netw.* **2004**, *15*, 1526–1540. [CrossRef] [PubMed]
12. Patel, A.; Kosko, B. Mutual-Information Noise Benefits in Brownian Models of Continuous and Spiking Neurons. In Proceedings of the 2006 International Joint Conference on Neural Network, Vancouver, BC, Canada, 16–21 July 2006; pp. 1368–1375.
13. Chen, H.; Varshney, P.K.; Kay, S.M.; Michels, J.H. Theory of the stochastic resonance effect in signal detection: Part I – fixed detectors. *IEEE Trans. Signal Process.* **2007**, *55*, 3172–3184. [CrossRef]
14. Patel, A.; Kosko, B. Optimal noise benefits in Neyman–Pearson and inequality constrained signal detection. *IEEE Trans. Signal Process.* **2009**, *57*, 1655–1669.
15. Bayram, S.; Gezici, S. Stochastic resonance in binary composite hypothesis-testing problems in the Neyman–Pearson framework. *Digit. Signal Process.* **2012**, *22*, 391–406. [CrossRef]
16. Bayrama, S.; Gultekinb, S.; Gezici, S. Noise enhanced hypothesis-testing according to restricted Neyman–Pearson criterion. *Digit. Signal Process.* **2014**, *25*, 17–27. [CrossRef]
17. Bayram, S.; Gezici, S.; Poor, H.V. Noise enhanced hypothesis-testing in the restricted Bayesian framework. *IEEE Trans. Signal Process.* **2010**, *58*, 3972–3989. [CrossRef]
18. Bayram, S.; Gezici, S. Noise enhanced M-ary composite hypothesis-testing in the presence of partial prior information. *IEEE Trans. Signal Process.* **2011**, *59*, 1292–1297. [CrossRef]
19. Chen, H.; Varshney, L.R.; Varshney, P.K. Noise-enhanced information systems. *Proc. IEEE* **2014**, *102*, 1607–1621. [CrossRef]
20. Weber, J.F.; Waldman, S.D. Stochastic Resonance is a Method to Improve the Biosynthetic Response of Chondrocytes to Mechanical Stimulation. *J. Orthop. Res.* **2015**, *34*, 231–239. [CrossRef] [PubMed]

21. Duan, F.; Chapeau-Blondeau, F.; Abbott, D. Non-Gaussian noise benefits for coherent detection of narrow band weak signal. *Phys. Lett. A* **2014**, *378*, 1820–1824. [CrossRef]

22. Lu, Z.; Chen, L.; Brennan, M.J.; Yang, T.; Ding, H.; Liu, Z. Stochastic resonance in a nonlinear mechanical vibration isolation system. *J. Sound Vib.* **2016**, *370*, 221–229. [CrossRef]

23. Rossi, P.S.; Ciuonzo, D.; Ekman, T.; Dong, H. Energy Detection for MIMO Decision Fusion in Underwater Sensor Networks. *IEEE Sen. J.* **2015**, *15*, 1630–1640. [CrossRef]

24. Rossi, P.S.; Ciuonzo, D.; Kansanen, K.; Ekman, T. Performance Analysis of Energy Detection for MIMO Decision Fusion in Wireless Sensor Networks Over Arbitrary Fading Channels. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 7794–7806. [CrossRef]

25. Ciuonzo, D.; de Maio, A.; Rossi, P.S. A Systematic Framework for Composite Hypothesis Testing of Independent Bernoulli Trials. *IEEE Signal Proc. Lett.* **2015**, *22*, 1249–1253. [CrossRef]

26. Parsopoulos, K.E.; Vrahatis, M.N. *Particle Swarm Optimization Method for Constrained Optimization Problems*; IOS Press: Amsterdam, The Netherlands, 2002; pp. 214–220.

27. Hu, X.; Eberhart, R. Solving constrained nonlinear optimization problems with particle swarm optimization. In Proceedings of the sixth world multiconference on systemics, cybernetics and informatics, Orlando, FL, USA, 14–18 July 2002.

28. Price, K.V.; Storn, R.M.; Lampinen, J.A. *Differential Evolution: A Practical Approach to Global Optimization*; Springer: New York, NY, USA, 2005.