REFERENCES

[1] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proc. IEEE*, vol. 95, no. 1, pp. 138–162, 2007.

[2] C. O. Savage and B. F. L. Scala, "Optimal scheduling of scalar gauss-markov systems with a terminal cost function," *IEEE Trans. Autom. Control*, vol. 54, no. 5, pp. 1100–1105, 2009.

[3] M. Shakeri, K. R. Pattipati, and D. L. Kleinman, "Optimal measurement scheduling for state estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 716–729, 1995.

[4] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," presented at the Proc. Amer. Control Conf., Baltimore, MD, 2010.

[5] S. Arai, Y. Iwatani, and K. Hashimoto, "Fast sensor scheduling for estimation of networked sensor systems," presented at the ICROS-SICE Int. Joint Conf., Fukuoka, Japan, 2009.

[6] A. S. Chhetri, D. Morrell, and A. Papandreou-Suppappola, "On the use of binary programming for sensor scheduling," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2826–2839, 2007.

[7] V. Krishnamurthy, "Algorithms for optimal scheduling and management of hidden Markov model sensors," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1382–1397, 2002.

[8] K. Cohen and A. Leshem, "A time-varying opportunistic approach to lifetime maximization of wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5307–5319, 2010.

[9] Y. Chen, Q. Zhao, V. Krishnamurthy, and D. Djonin, "Transmission scheduling for optimizing sensor network lifetime: A stochastic shortest path approach," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2294–2309, May 2007.

[10] M. Dong, L. Tong, and B. M. Sadler, "Information retrieval and processing in sensor networks: Deterministic scheduling versus random access," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5806–5820, 2007.

[11] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME*, vol. 82, no. 1, pp. 35–45, 1960.

[12] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

On the Restricted Neyman–Pearson Approach for Composite Hypothesis-Testing in Presence of Prior Distribution Uncertainty

Suat Bayram and Sinan Gezici

***Abstract*—The restricted Neyman–Pearson (NP) approach is studied for composite hypothesis-testing problems in the presence of uncertainty in the prior probability distribution under the alternative hypothesis. A restricted NP decision rule aims to maximize the average detection probability under the constraints on the worst-case detection and false-alarm probabilities, and adjusts the constraint on the worst-case detection probability according to the amount of uncertainty in the prior probability distribution. In this study, optimal decision rules according to the restricted NP criterion are investigated. Also, an algorithm is provided to calculate the optimal restricted NP decision rule. In addition, it is shown that the average detection probability is a strictly decreasing and concave function of the constraint on the minimum detection probability. Finally, a detection example is presented to investigate the theoretical results, and extensions to more generic scenarios are provided.**

***Index Terms*—Composite hypothesis, hypothesis-testing, max-min, Neyman–Pearson (NP), restricted Bayes.**

I. INTRODUCTION

Bayesian and minimax hypothesis-testings are two common approaches for the formulation of testing [1, pp. 5–22]–[3]. In the Bayesian approach, all forms of uncertainty are represented by a prior probability distribution, and the decision is made based on posterior probabilities. On the other hand, no prior information is assumed in the minimax approach, and a minimax decision rule minimizes the maximum of risk functions defined over the parameter space [1, pp. 13–22], [4]. The Bayesian and minimax frameworks can be considered as two extreme cases of prior information. In the former, perfect (exact) prior information is available whereas no prior information exists in the latter. In practice, having perfect prior information is a very exceptional case [5]. In most cases, prior information is incomplete and only partial prior information is available [5], [6]. Since the Bayesian approach is ineffective in the absence of exact prior information, and since the minimax approach, which ignores the partial prior information, can result in poor performance due to its conservative perspective, there have been various studies that take partial prior information into account [5]–[11], which can be considered as a mixture of Bayesian and frequentist approaches [12]. The most prominent of these approaches are the empirical Bayes, Γ-minimax, restricted Bayes and mean-max approaches [5]–[7], [11], [13]. As a solution to the impossibility of complete subjective specification of the model and the prior distribution in the Bayesian approach, the robust Bayesian analysis has been proposed [12], [14, pp. 195–214]. Although the robust Bayesian analysis is considered purely in the Bayesian framework in general, it also has strong connections with the empirical Bayes, Γ-minimax and restricted Bayes approaches [12], [14, pp. 215–235].

Manuscript received August 19, 2010; revised November 23, 2010, March 21, 2011, and August 06, 2011; accepted August 08, 2011. Date of current version September 14, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Cédric Richard.

The authors are with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara 06800, Turkey (e-mail: sbayram@ee.bilkent.edu.tr; gezici@ee.bilkent.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2153846

1053-587X/$26.00 © 2011 IEEE

Among the decision rules that take partial prior information into account, the restricted Bayes decision rule minimizes the Bayes risk under a constraint on the individual conditional risks [15, p. 15]. Depending on the value of the constraint, which is determined according to the amount of uncertainty in the prior information, the restricted Bayes approach covers the Bayes and minimax approaches as special cases [6]. An important characteristic of the restricted Bayes approach is that it combines probabilistic and nonprobabilistic descriptions of uncertainty, which are also called measurable and unmeasurable uncertainty [16], [17, Pt. III, Ch. VII], because the calculation of the Bayes (average) risk requires uncertainty to be measured and imposing a constraint on the conditional risks is a nonprobabilistic description of uncertainty. In this study, the focus is on the application of the notion of the restricted Bayes approach to the Neyman–Pearson (NP) framework, in which probabilistic and nonprobabilistic descriptions of uncertainty are combined [6].

In the NP approach for deciding between two simple hypotheses, the aim is to maximize the detection probability under a constraint on the false-alarm probability [1, pp. 22–29], [18, pp. 33–24]. When the null hypothesis is composite, it is common to apply the false-alarm constraint for all possible distributions under that hypothesis [19], [20]. On the other hand, various approaches can be taken when the alternative hypothesis is composite. One approach is to search for a uniformly most powerful (UMP) decision rule that maximizes the detection probability under the false-alarm constraint for all possible probability distributions under the alternative hypothesis [1, pp. 34–38], [18, pp. 86–92]. However, such a decision rule exists only under special circumstances [1]. Therefore, a generalized notion of the NP criterion, which aims to maximize the misdetection exponent uniformly over all possible probability distributions under the alternative hypothesis subject to the constraint on the false-alarm exponent, is employed in some studies [21]–[24]. Another approach is to maximize the average detection probability under the false-alarm constraint [12], [25]–[27]. In this case, the problem can be formulated in the same form as an NP problem for a simple alternative hypothesis (by defining the probability distribution under the alternative hypothesis as the expectation of the conditional probability distribution over the prior distribution of the parameter under the alternative hypothesis). Therefore, the classical NP lemma can be employed in this scenario. Hence, this max-mean approach for composite alternative hypotheses can be called as the "classical" NP approach. One important requirement for this approach is that a prior distribution of the parameter under the alternative hypothesis should be known in order to calculate the average detection probability. When such a prior distribution is not available, the max-min approach addresses the problem. In this approach, the aim is to maximize the minimum detection probability (the smallest power) under the false-alarm constraint [19], [20]. The solution to this problem is an NP decision rule corresponding to the least favorable distribution of the unknown parameter under the alternative hypothesis. It should be noted that considering the least favorable distribution is equivalent to considering the worst-case scenario, which can be unlikely to occur. Therefore, the max-min approach is quite conservative in general. Some modifications to this approach are proposed by employing the interval probability concept [28], [29].[1]

In this study, a generic criterion is investigated for composite hypothesis-testing problems in the NP framework, which covers the classical NP (max-mean) and the max-min criteria as special cases. Since this criterion can be regarded as an application of the restricted Bayes approach (Hodges–Lehmann rule) to the NP framework [6], [15], it is called the *restricted NP* approach in this study (in order to emphasize the considered NP framework). The investigation of the restricted NP criterion is intended to provide the signal processing community with an illustration of the Hodges-Lehmann rule in the NP framework. A restricted NP decision rule maximizes the average detection probability (average power) under the constraints that the minimum detection probability (the smallest power) cannot be less than a predefined value and that the false-alarm probability cannot be larger than a significance level. In this way, the uncertainty in the knowledge of the prior distribution under the alternative hypothesis is taken into account, and the constraint on the minimum (worst-case) detection probability is adjusted depending on the amount of uncertainty.

[1]The generalized likelihood ratio test (GLRT) is another approach for composite hypothesis-testing, which can be used to test a null hypothesis against an alternative hypothesis [1, p. 38], [18, pp. 92–96].

II. Problem Formulation and Motivation

Consider a family of probability densities $p_\theta(\mathbf{x})$ indexed by parameter θ that takes values in a parameter set Λ, where $\mathbf{x} \in \mathbb{R}^K$ represents the observation (data). A binary composite hypothesis-testing problem can be stated as

$$\mathcal{H}_0 : \theta \in \Lambda_0, \quad \mathcal{H}_1 : \theta \in \Lambda_1 \tag{1}$$

where $\mathcal{H}_i$ denotes the ith hypothesis and Λ_i is the set of possible parameter values under $\mathcal{H}_i$ for $i = 0, 1$ [1]. Parameter sets Λ_0 and Λ_1 are disjoint, and their union forms the parameter space, $\Lambda = \Lambda_0 \cup \Lambda_1$. It is assumed that the probability distributions of parameter θ under $\mathcal{H}_0$ and $\mathcal{H}_1$, denoted by $w_0(\theta)$ and $w_1(\theta)$, respectively, are known *with some uncertainty* (see [16] and [17, Pt. III, Ch. VII] for discussions on the concept of uncertainty). For example, these distributions can be obtained as probability density function (pdf) estimates based on previous decisions (experience). In that case, uncertainty is related to estimation errors, and higher amount of uncertainty is observed as the estimation errors increase.

In the NP framework, the aim is to maximize (a function of) the detection probability under a constraint on the false-alarm probabilities [1]. For composite hypothesis-testing problems in the NP framework, it is common to consider the conservative approach in which the false-alarm probability should be below a certain constraint for all possible values of parameter θ in set Λ_0 [19], [20]. In this case, whether the probability distribution of the parameter under $\mathcal{H}_0$, $w_0(\theta)$, is known completely or with uncertainty does not change the problem formulation (see Section V for extensions). On the other hand, the problem formulation depends heavily on the amount of knowledge about the probability distribution of the parameter under $\mathcal{H}_1$, $w_1(\theta)$.[2] In that respect, two extreme cases can be considered. In the first case, there is no uncertainty in $w_1(\theta)$. Then, the average detection probability can be considered, and the classical NP approach can be employed to obtain the detector that maximizes the average detection probability under the given false-alarm constraint [12], [25]–[27]. In the second case, there is full uncertainty in $w_1(\theta)$, meaning that the prior distribution under $\mathcal{H}_1$ is completely unknown. Then, maximizing the worst-case (minimum) detection probability can be considered under the false-alarm constraint, which is called as the max-min criterion or the "generalized" NP criterion [19], [20]. In fact, these two extreme cases, complete knowledge and full uncertainty of the prior distribution, are rarely encountered in practice. In most practical cases, there exists some uncertainty in $w_1(\theta)$, and the classical NP and the max-min approaches do not address those cases. The main motivation behind this study is to investigate a criterion that takes various amounts of uncertainty into account, and covers the approaches designed for the complete knowledge and the full uncertainty scenarios as special cases [6].

In practice, the prior distribution $w_1(\theta)$ is commonly estimated based on previous observations, and there exists some uncertainty

[2]In accordance with these observations, the term uncertainty will be used to refer to uncertainties in $w_1(\theta)$ unless stated otherwise.

in the knowledge of $w_1(\theta)$ due to estimation errors. Therefore, the amount of uncertainty depends on the amount of estimation errors. If the average detection probability is calculated based on the estimated prior distribution and the maximization of that average detection probability is performed based on the classical NP approach, it means that the estimation errors (hence, the uncertainty related to the prior distribution) are ignored. In such cases, very poor detection performance can be observed when the estimated distribution differs significantly from the correct one. On the other hand, if the max-min approach is used and the worst-case detection probability is maximized, it means that the prior information (contained in the prior distribution estimate) about the parameter is completely ignored, and the decision rule is designed as if there existed no prior information. Therefore, this approach does not utilize the available prior information at all and employs a very conservative perspective. In this study, we focus on a criterion that aims to maximize the average detection probability, calculated based on the estimated prior distribution, under the constraint that the minimum (worst-case) detection probability stays above a certain threshold, which can be adjusted depending on the amount of uncertainty in the prior distribution. In this way, both the prior information in the distribution estimate is utilized and the uncertainty in this estimate is considered. This criterion is referred to as the restricted NP criterion in this study, since it can be considered as an application of the restricted Bayes criterion (Hodges–Lehmann rule) to the NP framework [6]. The restricted NP criterion generalizes the classical NP and max-min approaches and covers them as special cases.

In order to provide a mathematical formulation of the restricted NP criterion, we first define the detection and false-alarm probabilities of a decision rule for given parameter values as follows:

$$P_D(\phi;\theta) \triangleq \int_\Gamma \phi(\mathbf{x})\; p_\theta(\mathbf{x})\; d\mathbf{x}, \quad \text{for } \theta \in \Lambda_1 \tag{2}$$

$$P_F(\phi;\theta) \triangleq \int_\Gamma \phi(\mathbf{x})\; p_\theta(\mathbf{x})\; d\mathbf{x}, \quad \text{for } \theta \in \Lambda_0 \tag{3}$$

where Γ represents the observation space, and $\phi(\mathbf{x})$ denotes a generic decision rule (detector) that maps the data vector into a real number in $[0,1]$, which represents the probability of selecting $\mathcal{H}_1$ [1]. Then, the restricted NP problem can be formulated as the following optimization problem:

$$\max_{\phi} \quad \int_{\Lambda_1} P_D(\phi;\theta)\; w_1(\theta)\; d\theta \tag{4}$$

$$\text{subject to} \quad P_D(\phi;\theta) \geq \beta, \quad \forall\, \theta \in \Lambda_1 \tag{5}$$

$$P_F(\phi;\theta) \leq \alpha, \quad \forall\, \theta \in \Lambda_0 \tag{6}$$

where α is false-alarm constraint, and β is the design parameter to compensate for the uncertainty in $w_1(\theta)$. In other words, a restricted NP decision rule maximizes the average detection probability, where the average is performed based on the prior distribution estimate $w_1(\theta)$, under the constraints on the worst-case detection and false-alarm probabilities. Parameter β in (5) is defined as $\beta \triangleq (1-\epsilon)\zeta$ for $0 \leq \epsilon \leq 1$, with ζ denoting the max-min detection probability. Namely, ζ is the maximum worst-case detection probability that can be obtained as follows:

$$\zeta = \max_{\phi} \min_{\theta\in\Lambda_1} P_D(\phi;\theta) \quad \text{subject to } P_F(\phi;\theta) \leq \alpha, \quad \forall\, \theta \in \Lambda_0. \tag{7}$$

From the definition of β, it is observed that β ranges from zero to ζ. In the case of full uncertainty in $w_1(\theta)$, ϵ is set to zero (i.e., $\beta = \zeta$), which reduces the restricted NP problem in (4)–(6) to the max-min problem in(7). On the other hand, in the case of complete knowledge of $w_1(\theta)$, ϵ can be set to 1, and the restricted NP problem reduces to the classical NP problem, specified by (4) and (6), which can be expressed as

$$\max_{\phi} \quad P_D^{\text{avg}}(\phi) \quad \text{subject to} \quad P_F(\phi;\theta) \leq \alpha, \quad \forall\, \theta \in \Lambda_0 \tag{8}$$

where $P_D^{\text{avg}}(\phi) \triangleq \int_{\Lambda_1} P_D(\phi;\theta)\; w_1(\theta)\; d\theta$ is the average detection probability. Therefore, the max-min and the classical NP approaches are two special cases of the restricted NP approach.

III. Analysis of Restricted Neyman–Pearson Approach

In this section, the aim is to investigate the optimal solution of the restricted NP problem in (4)–(6). For this purpose, the definitions in (2) and (3) can be used to reformulate the problem in (4)–(6) as follows:

$$\max_{\phi} \quad \int_\Gamma \phi(\mathbf{x})\; p_1(\mathbf{x})\; d\mathbf{x} \tag{9}$$

$$\text{subject to} \quad \min_{\theta\in\Lambda_1} \int_\Gamma \phi(\mathbf{x})\; p_\theta(\mathbf{x})\; d\mathbf{x} \geq \beta \tag{10}$$

$$\max_{\theta\in\Lambda_0} \int_\Gamma \phi(\mathbf{x})\; p_\theta(\mathbf{x})\; d\mathbf{x} \leq \alpha \tag{11}$$

where $p_1(\mathbf{x}) \triangleq \int_{\Lambda_1} p_\theta(\mathbf{x}) w_1(\theta)\; d\theta$ defines the pdf of the observation under $\mathcal{H}_1$, which is obtained based on the prior distribution estimate $w_1(\theta)$. In addition, an alternative representation of the problem in (9)–(11) can be expressed as

$$\max_{\phi} \quad \lambda \int_\Gamma \phi(\mathbf{x})\; p_1(\mathbf{x})\; d\mathbf{x} + (1-\lambda) \min_{\theta\in\Lambda_1} \int_\Gamma \phi(\mathbf{x})\; p_\theta(\mathbf{x})\; d\mathbf{x} \tag{12}$$

$$\text{subject to} \max_{\theta\in\Lambda_0} \int_\Gamma \phi(\mathbf{x})\; p_\theta(\mathbf{x})\; d\mathbf{x} \leq \alpha \tag{13}$$

where $0 \leq \lambda \leq 1$ is a design parameter that is selected according to β.

A. Characterization of Optimal Decision Rule

Based on the formulation in (12) and (13), the following theorem provides a method to characterize the optimal solution of the restricted NP problem under certain conditions.

Theorem 1: Define a pdf $v(\theta)$ as $v(\theta) \triangleq \lambda\; w_1(\theta) + (1-\lambda)\; \mu(\theta)$, where $\mu(\theta)$ is any valid pdf. If ϕ^* is the NP solution for $v(\theta)$ under the false-alarm constraint and satisfies

$$\int_\Gamma \phi^*(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x})\; \mu(\theta)\; d\theta\; d\mathbf{x} = \min_{\theta\in\Lambda_1} \int_\Gamma \phi^*(\mathbf{x})\; p_\theta(\mathbf{x})\; d\mathbf{x}, \tag{14}$$

then it is a solution of the problem in (12) and (13).

Proof: The proof is similar to the proof of Theorem 1 in [6]. Please see [30] for the details.

Theorem 1 states that if one can find a pdf $\mu(\theta)$ that satisfies the condition in (14), then the NP solution corresponding to $\lambda\; w_1(\theta) + (1-\lambda)\; \mu(\theta)$ is a solution of the restricted NP problem in (12) and (13). Also it should be noted that Theorem 1 is an optimality result; it does not guarantee existence or uniqueness. However, in most cases, the optimal solution proposed by Theorem 1 exists, which can be proven as in [6] based on some assumptions on the interchangeability of supremum and infimum operators, and on the existence of a probability distribution (a decision rule) that minimizes (maximizes) the maximum (minimum) average detection probability (see Assumptions 1–3 in [6]). In fact, those assumptions hold when a set of conditions specified in [31, pp. 191–205] are satisfied. From a practical perspective, the assumptions hold, for example, when the probability distributions are discrete or absolutely continuous (i.e., have cumulative distributions function that are absolutely continuous with respect to the Lebesgue measure), the parameter space is compact, and the problem is nonsequential [6].

More specifically, for the problem formulation in this study, all the assumptions are satisfied when $p_\theta(\mathbf{x}), \forall\, \theta \in \Lambda$, is discrete, or cumulative distributions corresponding to $p_\theta(\mathbf{x}), \forall\, \theta \in \Lambda$, are absolutely continuous (with respect to the Lebesgue measure), and the parameter space Λ is compact.

Remark 1: In Theorem 1, the meaning of ϕ^* being the NP solution for $v(\theta)$ under the false-alarm constraint is that ϕ^* solves the following optimization problem:

$$\begin{aligned} \max_{\phi} \quad & \int_\Gamma \phi(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta\, d\mathbf{x} \\ \text{subject to} \quad & \max_{\theta\in\Lambda_0} \int_\Gamma \phi(\mathbf{x})\, p_\theta(\mathbf{x})\, d\mathbf{x} \le \alpha \end{aligned} \tag{15}$$

where $v(\theta) = \lambda\, w_1(\theta) + (1-\lambda)\, \mu(\theta)$. Based on the NP lemma [1, pp. 22–29], it can be shown that the solution of (15) is in the form of a likelihood ratio test (LRT); that is,[3]

$$\phi^*(\mathbf{x}) = \begin{cases} 1, & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta > \eta\, p_{\tilde\theta_0}(\mathbf{x}) \\ \kappa(\mathbf{x}), & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta = \eta\, p_{\tilde\theta_0}(\mathbf{x}) \\ 0, & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta < \eta\, p_{\tilde\theta_0}(\mathbf{x}) \end{cases} \tag{16}$$

where $\eta \ge 0$ and $0 \le \kappa(\mathbf{x}) \le 1$ are such that $\max_{\theta\in\Lambda_0} P_F(\phi^*;\theta) = \alpha$, and $\tilde\theta_0$ is defined as

$$\tilde\theta_0 = \arg\max_{\theta\in\Lambda_0} P_F(\phi^*;\theta). \tag{17}$$

Therefore, the solution of the restricted NP problem in (12) and (13) can be expressed by the LRT specified in (16) and (17), once a pdf $\mu(\theta)$ and the corresponding decision rule ϕ^* that satisfy the constraint in (14) are obtained (see Section III-B). It should also be noted that having multiple solutions for $\tilde\theta_0$ does not present a problem since it can be shown that the same average detection probability is achieved for all the solutions.

The following corollary is presented in order to show the equivalence between the formulation in (12) and (13) and that in (4)–(6) [30].

Corollary 1: Under the conditions in Theorem 1, ϕ^* solves the optimization problem in (4)–(6) when $\min_{\theta\in\Lambda_1} \int_\Gamma \phi^*(\mathbf{x})\, p_\theta(\mathbf{x})\, d\mathbf{x} = \beta$.

Corollary 1 states that when the decision rule ϕ^* specified in Theorem 1 satisfies the constraint in (10) with equality, it also provides a solution of the restricted NP problem specified in (9)–(11); equivalently, in (4)–(6). In other words, the average detection probability can be maximized when the minimum of the detection probabilities for all possible parameter values $\theta \in \Lambda_1$ is equal to the lower limit β. It should also be noted that Corollary 1 establishes a formal link between parameters λ and β. For any λ, β can be calculated through the equation in the corollary.

Another property of the optimal decision rule ϕ^* described in Theorem 1 is that it can be defined as an NP solution corresponding to the *least favorable* distribution $v(\theta)$ specified in Theorem 1. In other words, among a family of pdf's, $v(\theta)$ is the least favorable one since it minimizes the average detection probability. This observation is similar, for example, to the fact that the minimax decision rule is the Bayes rule corresponding to the least favorable priors [1, pp. 15–16]. For the following theorem, an approach similar to that in [6] is taken in order to show that $v(\theta)$ in Theorem 1 corresponds to a least favorable distribution (please see [30] for the proof).

[3]The proof follows from the observation that $(\phi^*(\mathbf{x}) - \phi(\mathbf{x}))\left(\int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta - \eta\, p_{\tilde\theta_0}(\mathbf{x})\right) \ge 0, \forall\, \mathbf{x}$, for any decision rule ϕ due to the definition of ϕ^* in (16). Then, the approach of [1, p. 24] can be used to prove that $\int_\Gamma \phi^*(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta\, d\mathbf{x} \ge \int_\Gamma \phi(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta\, d\mathbf{x}$ for any decision rule ϕ that satisfies $P_F(\phi;\theta) \le \alpha, \forall\, \theta \in \Lambda_0$.

Theorem 2: Under the conditions in Theorem 1, $v(\theta) = \lambda\, w_1(\theta) + (1-\lambda)\, \mu(\theta)$ minimizes the average detection probability among all prior distributions in the form of

$$\tilde v(\theta) = \tilde\lambda\, w_1(\theta) + (1-\tilde\lambda)\, \tilde\mu(\theta) \tag{18}$$

for $\tilde\lambda \ge \lambda$, where $\theta \in \Lambda_1$ and $\tilde\mu(\theta)$ is any probability distribution. Equivalently,

$$\int_\Gamma \phi^*(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta\, d\mathbf{x} \le \int_\Gamma \phi^\star(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x})\, \tilde v(\theta)\, d\theta\, d\mathbf{x}$$

for any $\tilde v(\theta)$ described above, where ϕ^* and $\phi^\star$ are the α-level NP decision rules corresponding to $v(\theta)$ and $\tilde v(\theta)$, respectively.

Although Theorem 2 provides a definition of the least favorable distribution in a family of prior distributions in the form of $\tilde v(\theta) = \tilde\lambda\, w_1(\theta) + (1-\tilde\lambda)\, \tilde\mu(\theta)$ for $\tilde\lambda \ge \lambda$, only the case $\tilde\lambda = \lambda$ is of interest in practice since λ in (12) is commonly set as a design parameter depending on the amount of uncertainty in the prior distribution. Therefore, in calculating the optimal decision rule according to the restricted NP criterion, the special case of Theorem 2 for $\tilde\lambda = \lambda$ will be employed in Section IV.

B. Calculation of Optimal Decision Rule

The analysis in Section III-A reveals that a density $\mu(\theta)$ and a corresponding NP rule (as specified in Remark 1) that satisfy the constraint in Theorem 1 need to be obtained for the solution of the restricted NP problem. To this aim, the condition in Theorem 1 can be expressed based on (2) as

$$\int_{\Lambda_1} \mu(\theta)\, P_D(\phi^*;\theta)\, d\theta = \min_{\theta\in\Lambda_1} P_D(\phi^*;\theta). \tag{19}$$

This condition requires that $\mu(\theta)$ assigns nonzero probabilities only to the values of θ that result in the global minimum of $P_D(\phi^*;\theta)$. First, assume that $P_D(\phi^*;\theta)$ has a unique minimizer that achieves the global minimum (the extensions in the absence of this assumption will be discussed as well). Then, $\mu(\theta)$ can be expressed as

$$\mu(\theta) = \delta(\theta - \theta_1) \tag{20}$$

which means that $\theta = \theta_1$ with probability one under this distribution. Based on this observation, the following algorithm can be proposed to obtain the optimal restricted NP decision rule.

Algorithm

1) Obtain $P_D(\phi^*_{\theta_1};\theta)$ for all $\theta_1 \in \Lambda_1$, where $\phi^*_{\theta_1}$ denotes the α-level NP decision rule corresponding to $v(\theta) = \lambda\, w_1(\theta) + (1-\lambda)\delta(\theta - \theta_1)$ as described in (16) and (17).
2) Calculate

$$\theta_1^* = \arg\min_{\theta_1\in\Lambda_1} f(\theta_1) \tag{21}$$

where

$$f(\theta_1) \triangleq \lambda \int_{\Lambda_1} w_1(\theta) P_D\left(\phi^*_{\theta_1};\theta\right)\, d\theta + (1-\lambda)\, P_D\left(\phi^*_{\theta_1};\theta_1\right). \tag{22}$$

3) If $P_D(\phi^*_{\theta_1^*};\theta_1^*) = \min_{\theta\in\Lambda_1} P_D(\phi^*_{\theta_1^*};\theta)$, output $\phi^*_{\theta_1^*}$ as the solution of the restricted NP problem; otherwise, the solution does not exist.

It should be noted that $f(\theta_1)$ in (22) is the average detection probability corresponding to $v(\theta) = \lambda\, w_1(\theta) + (1-\lambda)\, \delta(\theta - \theta_1)$.[4] Since

[4]It should be noted that λ is related to the design parameter β in (5) through Corollary 1. In addition, the fact that as λ increases (decreases), β decreases (increases) can be used to adjust the corresponding parameter value.

Theorem 2 (for $\tilde{\lambda} = \lambda$) states that the optimal restricted NP solution corresponds to the least favorable prior distribution, which results in the minimum average detection probability, the only possible solution is the NP decision rule corresponding to θ_1^* in (21), $\phi_{\theta_1^*}^*$. Therefore, only that rule is considered in the last step of the algorithm, and the optimality condition is checked. If the condition is satisfied, the optimal restricted NP solution is obtained. Although not common in practice, the optimal solution may not exist in some cases since Theorem 1 does not guarantee existence. Also, it should be noted that there may be multiple solutions of (21), and in that case any solution of (21) satisfying the third condition in the algorithm is an optimal solution according to Theorem 1. Therefore, one such solution can be selected for the optimal restricted NP solution.

In order to extend the algorithm to the cases in which $P_D(\phi^*;\theta)$ has multiple values of θ that achieve the global minimum, express $\mu(\theta)$ as

$$\mu(\theta) = \sum_{l=1}^{N} \nu_l \; \delta(\theta - \theta_l) \tag{23}$$

where $\nu_l \geq 0$, $\sum_{l=1}^{N} \nu_l = 1$, and N is the number of θ values that minimize $P_D(\phi^*;\theta)$. For simplicity of notation, let $\boldsymbol{\vartheta}$ denote the vector of unknown parameters of $\mu(\theta)$; that is, $\boldsymbol{\vartheta} = [\theta_1 \cdots \theta_N \; \nu_1 \cdots \nu_N]$. Based on (23), the calculations in the algorithm should be updated as follows:

$$\boldsymbol{\vartheta}^* = \underset{\boldsymbol{\vartheta}}{\operatorname{argmin}} \, f(\boldsymbol{\vartheta}) \tag{24}$$

where

$$f(\boldsymbol{\vartheta}) \triangleq \lambda \int_{\Lambda_1} w_1(\theta) P_D\left(\phi_{\boldsymbol{\vartheta}}^*;\theta\right) \, d\theta + (1-\lambda) \sum_{l=1}^{N} \nu_l \, P_D\left(\phi_{\boldsymbol{\vartheta}}^*;\theta_l\right) \tag{25}$$

with $\phi_{\boldsymbol{\vartheta}}^*$ denoting the NP solution corresponding to $v(\theta) = \lambda \; w_1(\theta) + (1-\lambda) \sum_{l=1}^{N} \nu_l \; \delta(\theta - \theta_l)$. Then, the condition $P_D(\phi_{\boldsymbol{\vartheta}^*}^*;\boldsymbol{\vartheta}^*) = \min_{\boldsymbol{\vartheta}} P_D(\phi_{\boldsymbol{\vartheta}^*}^*;\boldsymbol{\vartheta})$ is checked to verify the optimal solution as $\phi_{\boldsymbol{\vartheta}^*}^*$. It is noted from (24) that the computational complexity can increase significantly when the detection probability is minimized by multiple θ values. In such cases, global optimization algorithms, such as particle-swarm optimization (PSO) [32], [33], genetic algorithms and differential evolution [34], can be used to calculate $\boldsymbol{\vartheta}^*$.

Finally, if the global minimum of $P_D(\phi^*;\theta)$ is achieved by infinitely many θ values, then all possible $\mu(\theta)$ need to be considered, which can have prohibitive complexity in general. In order to obtain an approximate solution in such cases, *Parzen window density estimation* [35, pp. 161–168] can be employed as in [36]. Specifically, $\mu(\theta)$ is expressed approximately by a linear combination of a number of window functions as

$$\mu(\theta) \approx \sum_{l=1}^{N_w} \xi_l \; \varphi_l(\theta - \theta_l), \tag{26}$$

and the unknown parameters of $\mu(\theta)$ such as θ_l and ξ_l can be collected into $\boldsymbol{\vartheta}$ as for the discrete case above. Then, (24) and (25) can be employed in the algorithm by replacing ν_l and N with ξ_l and N_w, respectively, and by defining $\phi_{\boldsymbol{\vartheta}}^*$ as the NP solution corresponding to $v(\theta) = \lambda \; w_1(\theta) + (1-\lambda) \sum_{l=1}^{N_w} \xi_l \; \varphi_l(\theta - \theta_l)$.

In Section IV, an example is provided to illustrate how to calculate the optimal restricted NP solution based on the techniques discussed in this section. Since the number of minimizers of $P_D(\phi^*;\theta)$ may not be known in advance, a practical approach can be taken as follows. First, it is assumed that there is only one value of θ that achieves the global minimum, and the algorithm is applied based on this assumption [see (21) and (22)]. If the condition in Step 3) is satisfied, then the optimal solution is obtained. Otherwise, it is assumed that there are two (or, more) θ values that achieve the global minimum, and the algorithm is run based on (24) and (25). In this way, the complexity of the solution can be increased gradually until a solution is obtained. Please see [30] for a discussion on the computational complexity of the three-step algorithm proposed in this section.

C. Properties of Average Detection Probability in Restricted NP Solutions

In the restricted NP approach, the average detection probability is maximized under some constraints on the worst-case detection and false-alarm probabilities [see (4)–(6)]. On the other hand, the classical NP approach in (8) does not consider the constraint on the worst-case detection probability, and maximizes the average detection probability under the constraint on the worst-case false-alarm probability only. Therefore, the average detection probability achieved by the classical NP approach is larger than or equal to that of the restricted NP approach; however, its worst-case detection probability is smaller than or equal to that of the restricted NP solution. Considering the max-min approach in (7), the aim is to maximize the worst-case detection probability under the constraint on the worst-case false-alarm probability. Therefore, the worst-case detection probability achieved by the max-min decision rule is larger than or equal to that of the restricted NP decision rule, whereas the average detection probability of the max-min approach is smaller than or equal to that of the restricted NP solution.

In order to express the relations above in mathematical terms, let ϕ_r^β, ϕ_m, and ϕ_c denote the solutions of the restricted NP, max-min and classical NP problems in (4)–(6), (7), and (8), respectively. In addition, let $L \triangleq \min_{\theta \in \Lambda_1} P_D(\phi_c;\theta)$ and $U \triangleq \min_{\theta \in \Lambda_1} P_D(\phi_m;\theta)$ define the worst-case detection probabilities of the classical NP and max-min solutions, respectively. It should be noted that, in the restricted NP approach, the constraint β on the worst-case detection probability [see (5)] cannot be larger than U, since the max-min solution provides the maximum value of the worst-case detection probability as discussed before. On the other hand, when β is selected to be smaller than L in the restricted NP formulation, the worst-case detection probability constraint becomes ineffective; hence, the restricted NP and the classical NP approaches become identical. Therefore, β in the restricted NP formulation is defined over the interval $[L, U]$ in practice. As a special case, when $L = U = \beta$, the restricted NP, the max-min and the classical NP solutions all become equal.

For the restricted NP solution ϕ_r^β, the average detection probability can be calculated as

$$P_D^{\text{avg}}\left(\phi_r^\beta\right) = \int_{\Lambda_1} P_D(\phi_r^\beta;\theta) \; w_1(\theta) \; d\theta. \tag{27}$$

The discussions above imply that $P_D^{\text{avg}}\left(\phi_r^\beta\right)$ is constant and equal to the average detection probability of the classical NP solution for $\beta \leq L$. In order to characterize the behavior of $P_D^{\text{avg}}\left(\phi_r^\beta\right)$ for $\beta \in [L, U]$, the following theorem is presented.

Theorem 3: The average detection probability of the restricted NP decision rule, $P_D^{\text{avg}}\left(\phi_r^\beta\right)$, is a strictly decreasing and concave function of β for $\beta \in [L, U]$.

Proof: Please see the Appendix.

Theorem 3 implies that the average detection probability can be improved monotonically as β decreases towards L. In other words, by considering a less strict constraint (i.e., smaller β) on the worst-case detection probability, it is possible to increase the average detection probability. However, it should be noted that β should be selected depending on the amount of uncertainty in the prior distribution; namely, smaller β values are selected as the uncertainty decreases. Therefore, Theorem 3 implies that the reduction in the uncertainty can always be used to improve the average detection probability. Another important conclusion from Theorem 3 is that there is a diminishing return in

improving the average detection probability by reducing β due to the concavity of $P_D^{\text{avg}}\left(\phi_r^{\beta}\right)$. In other words, a unit decrease of β results in a smaller increase in the average detection probability for smaller values of β. Fig. 1 in Section IV provides an illustration of the results of Theorem 3.

IV. Numerical Results

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

$$\mathcal{H}_0 : X = V, \quad \mathcal{H}_1 : X = \Theta + V \tag{28}$$

where $X \in \mathbb{R}$, Θ is an unknown parameter, and V is symmetric Gaussian mixture noise with the following pdf $p_V(v) = \sum_{i=1}^{N_m} \omega_i \ \psi_i(v - m_i)$, where $\omega_i \geq 0$ for $i = 1, \ldots, N_m$, $\sum_{i=1}^{N_m} \omega_i = 1$, and $\psi_i(x) = 1/(\sqrt{2\pi}\,\sigma_i)\ \exp\left(-x^2/(2\,\sigma_i^2)\right)$ for $i = 1, \ldots, N_m$. Due to the symmetry assumption, $m_l = -m_{N_m-l+1}$, $\omega_l = \omega_{N_m-l+1}$ and $\sigma_l = \sigma_{N_m-l+1}$ for $l = 1, \ldots, \lfloor N_m/2 \rfloor$, where $\lfloor y \rfloor$ denotes the largest integer smaller than or equal to y. Note that if N_m is an odd number, $m_{(N_m+1)/2}$ should be zero for symmetry.

Parameter Θ in (28) is modeled as a random variable with a pdf in the form of

$$w_1(\theta) = \rho\ \delta(\theta - A) + (1 - \rho)\ \delta(\theta + A) \tag{29}$$

where A is exactly known, but ρ is known with some uncertainty. With this model, the detection problem in (28) corresponds to the detection of a signal that employs binary modulation, namely, binary phase shift keying (BPSK). It should be noted that prior probabilities of symbols are not necessarily equal (i.e., ρ may not be equal to 0.5) in all communications systems [37]; hence, ρ should be estimated based on (previous) measurements in practice. In the numerical examples, the possible errors in the estimation of ρ are taken into account in the restricted NP framework.

For the problem formulation above, the parameter sets under $\mathcal{H}_0$ and $\mathcal{H}_1$ can be specified as $\Lambda_0 = \{0\}$ and $\Lambda_1 = \{-A, A\}$, respectively. In addition, the conditional pdf of X for a given value of $\Theta = \theta$ is expressed as

$$p_\theta(x) = \sum_{i=1}^{N_m} \frac{\omega_i}{\sqrt{2\pi}\,\sigma_i}\ \exp\left(\frac{-(x - \theta - m_i)^2}{2\,\sigma_i^2}\right). \tag{30}$$

In order to obtain the optimal restricted NP decision rule for this problem, the algorithm in Section III-B is employed. First, it is assumed that $\mu(\theta)$ can be expressed as in (20); namely, $\mu(\theta) = \delta(\theta - \theta_1)$, where $\theta_1 \in \{-A, A\}$, and the algorithm is applied based on (21) and (22). When the condition in the third step of the algorithm is satisfied, then the optimal solution is obtained. Otherwise, $\mu(\theta)$ is represented as $\mu(\theta) = \tilde{\gamma}\ \delta(\theta - A) + (1 - \tilde{\gamma})\ \delta(\theta + A)$ for $\tilde{\gamma} \in [0, 1]$, and the algorithm is run based on this model [consider (23) with $N = 2$, $\nu_1 = 1 - \nu_2 = \tilde{\gamma}$, and $\theta_1 = -\theta_2 = A$]. Note that this model includes all possible pdf's since $\Lambda_1 = \{-A, A\}$. As there is only one unknown variable, $\tilde{\gamma}$, in $\mu(\theta)$, the algorithm can be employed to find the value of $\tilde{\gamma}$ that minimizes the average detection probability [see (24) and (25) with $\boldsymbol{\theta} = \tilde{\gamma}$]. Then, the condition in the third step of the algorithm is checked in order to obtain the optimal decision rule.

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 are specified as $[0.1\ 0.95\ -0.95\ -0.1]$ 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 are assumed to be the same; i.e., $\sigma_i = \sigma$ for $i = 1, \ldots, N_m$.

In Fig. 1, the average detection probabilities of the classical NP, restricted NP, and max-min decision rules are plotted against β, which specifies the lower limit on the minimum (worst-case) detection probability. Various values of ρ in (29) are considered, and $A = 1$, $\sigma = 0.2$, and $\alpha = 0.2$ [see (6)] are used. As discussed in Section III-C, the restricted NP decision rule reduces to the classical NP decision rule when β is smaller than or equal to the worst-case detection probability of the classical NP decision rule.[5] On the other hand, the restricted NP and the max-min decision rules become identical when β is equal to the worst-case detection probability of the max-min decision rule. For the restricted NP decision rule, β is equal to the minimum detection probability; hence, the x axis in Fig. 1 can also be considered as the minimum detection probability except for the constant parts of the lines that correspond to the classical NP. As expected, the highest average detection probabilities are achieved by the classical NP decision rule; however, it also results in the lowest minimum detection probabilities, which are 0.453, 0.431, and 0.389 for $\rho = 0.7$, $\rho = 0.8$, and $\rho = 0.9$, respectively. Conversely, the max-min decision rule achieves the highest minimum detection probabilities, but its average detection probabilities are the worst. On the other hand, the restricted NP decision rules provide tradeoffs between the average and the minimum detection probabilities, and cover the classical NP and the max-min decision rules as the special cases. It is also observed from the figure that as ρ decreases, the difference between the performance of the classical NP and the max-min decision rules reduces. In fact, for $\rho = 0.5$, the restricted NP, the max-min, and the classical NP decision rule all become equal, since it can be shown that $w_1(\theta)$ in (29) becomes the least favorable pdf for $\rho = 0.5$. Fig. 1 can also be used to investigate the results of Theorem 3. It is observed that the average detection probability is a strictly decreasing and concave function of β for the restricted NP decision rule, as claimed in the theorem. Finally, we would like to mention that Fig. 1 can provide guidelines for the designer to choose a β value by observing the corresponding average detection probability for each β. Therefore, in practice, instead of setting a prescribed β directly, Fig. 1 can be used to choose a β value for the problem.

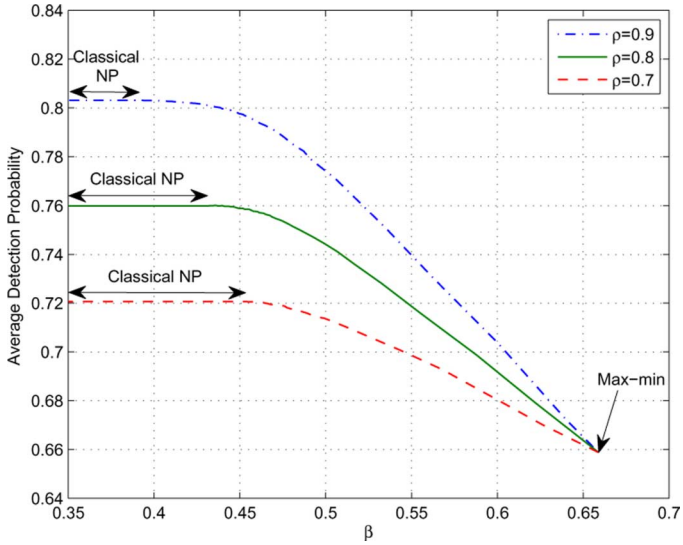

Fig. 1. Average detection probability versus β for the classical NP, restricted NP, and max-min decision rules for $\rho = 0.7$, $\rho = 0.8$ and $\rho = 0.9$, where $A = 1$, $\sigma = 0.2$, and $\alpha = 0.2$.

For the scenario in Fig. 1, the least favorable distributions are investigated for the restricted NP decision rule, and they are compared against the least favorable distribution for the max-min decision rule.

[5]Although the classical NP decision rule can be regarded as a special case of the restricted NP decision rule for $\beta \leq L$, the "restricted NP decision rule" term is used only for $\beta \in [L, U]$ in the following discussions (see Section III-C).

TABLE I
PARAMETER γ FOR LEAST-FAVORABLE DISTRIBUTION $v(\theta) = \gamma\,\delta(\theta-1)+(1-\gamma)\,\delta(\theta+1)$ CORRESPONDING TO RESTRICTED NP DECISION RULES. "NA" MEANS THAT THE GIVEN MINIMUM DETECTION PROBABILITY CANNOT BE ACHIEVED BY A RESTRICTED NP DECISION RULE

Avg. det. prob. $\rho = 0.9/0.8/0.7$	Min. det. prob.	γ
0.7997/0.7597/NA	0.4398	0.765
0.7915/0.7556/0.7197	0.4687	0.63
0.7635/0.7360/0.7086	0.5166	0.54
0.7301/0.7115/0.6930	0.5629	0.522
0.7034/0.6920/0.6806	0.6007	0.513
0.6724/0.6688/0.6652	0.6398	0.504

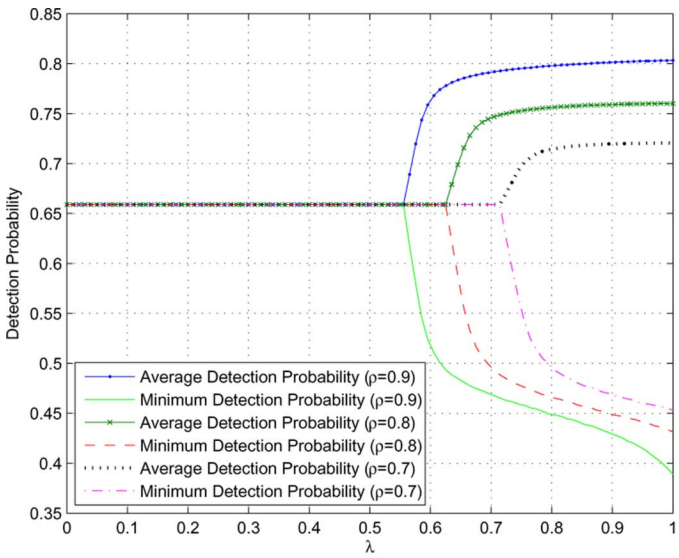

Fig. 2. Average and minimum detection probabilities of the restricted NP decision rules versus λ for $\rho = 0.7$, $\rho = 0.8$ and $\rho = 0.9$, where $A = 1$, $\alpha = 0.2$ and $\sigma = 0.2$.

For the max-min criterion, the least favorable distribution $w_{\rm lf}(\theta)$ in this example can be calculated as $w_{\rm lf}(\theta) = 0.5\,\delta(\theta-1) + 0.5\,\delta(\theta+1)$. Table I shows the least favorable distributions, expressed in the form of $v(\theta) = \gamma\,\delta(\theta-1) + (1-\gamma)\,\delta(\theta+1)$, for the restricted NP solution for various parameters. The corresponding average and minimum detection probabilities are also listed. As the minimum detection probability increases, the least favorable distribution gets closer to that of the max-min decision rule. It is also noted that the least favorable distributions are the same for all the ρ values in this example.

Fig. 2 plots the average and minimum detection probabilities of the restricted NP decision rules versus λ in (12) for $\rho = 0.7$, $\rho = 0.8$ and $\rho = 0.9$, where $A = 1$, $\sigma = 0.2$ and $\alpha = 0.2$ are used. It is observed that the average and the minimum detection probabilities are the same when $0 \le \lambda \le 0.555$ for $\rho = 0.9$, when $0 \le \lambda \le 0.625$ for $\rho = 0.8$, and when $0 \le \lambda \le 0.714$ for $\rho = 0.7$. In these cases, the restricted NP decision rule is equivalent to the max-min decision rule. On the other hand, for $\lambda = 1$, the restricted NP decision rule reduces to the classical NP decision rule. These observations can easily be verified from (12) and (13). Another observation from Fig. 2 is that the max-min solution equalizes the detection probabilities for $\theta \in \Lambda_1 = \{-1, 1\}$ values. Therefore, the average and the minimum detection probabilities are equal for the max-min solutions. On the other hand, the classical NP solution maximizes the average detection probability at the expense of reducing the worst-case (minimum) detection probability. For this reason, the difference between the average and the minimum detection probabilities increases with λ. Finally, Fig. 2 shows that the difference between the average and the minimum detection probabilities increases as ρ increases.

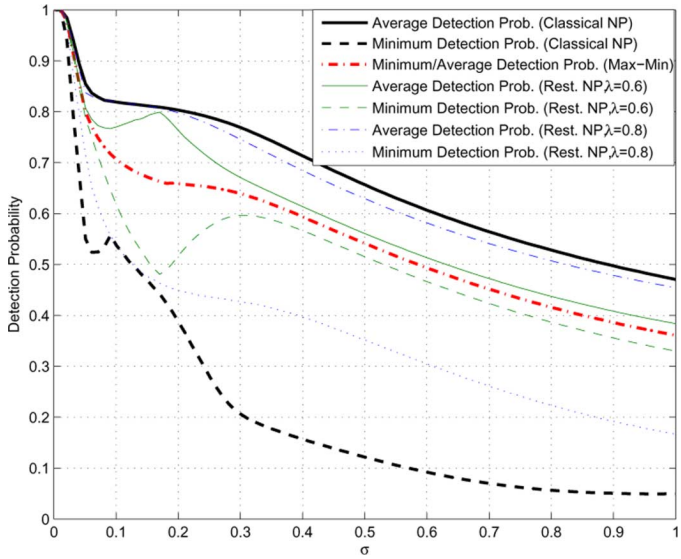

Fig. 3. Average and minimum detection probabilities of the classical NP, max-min, and restricted NP (for $\lambda = 0.6$ and $\lambda = 0.8$) decision rules versus σ for $A = 1$, $\alpha = 0.2$, and $\rho = 0.9$.

Fig. 3 compares the performances of the restricted NP, the max-min, the classical NP decision rules for various standard deviation values σ, where $A = 1$, $\alpha = 0.2$ and $\rho = 0.9$ are used. The restricted NP decision rules are calculated for $\lambda = 0.6$ and $\lambda = 0.8$, where the weight λ is as specified in (12). For each decision rule, both the average detection probability and the minimum (worst-case) detection probability are obtained. As expected, the classical NP decision rule achieves the highest average detection probability and the lowest minimum detection probability for all values of σ. On the other hand, the max-min decision rule achieves the highest minimum detection probability and the lowest average detection probability. It is noted that the max-min decision rule equalizes the detection probabilities for various parameter values, and results in the same average and the minimum detection probabilities. Another observation from Fig. 3 is that the restricted NP decision rule gets closer to the classical NP decision rule as λ increases, and to the max-min decision rule as λ decreases. The restricted NP decision rule provides various advantages over the classical NP and the max-min decision rules when both the average and the minimum detection probabilities are considered. For example, the restricted NP decision rule for $\lambda = 0.8$ has very close average detection probabilities to those of the classical NP decision rule; however, it achieves significantly higher minimum detection probabilities. Therefore, *even if the prior distribution is known perfectly*, it can be advantageous to use the restricted NP decision rule when *both* the average and the minimum detection probabilities are considered as performance metrics.[6] Of course, when there are uncertainties in the knowledge of the prior distribution, the actual average probabilities achieved by the classical NP approach can be significantly lower than those shown in Fig. 3, which can get as low as the lowest curve. In such scenarios, the restricted NP approach has a clear performance advantage. Compared to the max-min decision rule, the advantage of the restricted NP decision is to utilize the prior information, which can include uncertainty, in order to achieve higher average detection probabilities.

[6]In this problem, for $\rho > 0.5$, the minimum detection probability corresponds to $\theta = -1$, which occurs with probability $1-\rho$. Therefore, the minimum detection probability may be considered as an important performance metric along with the average detection probability.

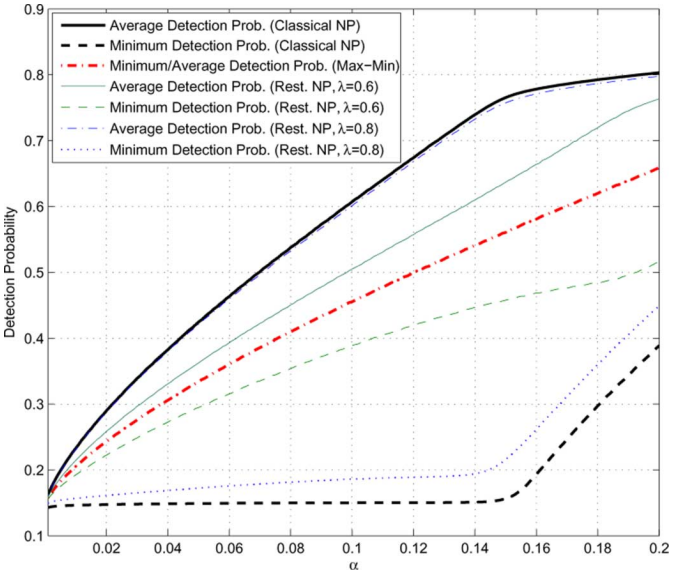

Fig. 4. Average and minimum detection probabilities of the classical NP, max-min, and restricted NP (for $\lambda = 0.6$ and $\lambda = 0.8$) decision rules versus α for $A = 1$, $\sigma = 0.2$, and $\rho = 0.9$.

Finally, in Fig. 4, the average and the minimum detection probabilities of the restricted NP (for $\lambda = 0.6$ and $\lambda = 0.8$), the max-min, and the classical NP decision rules are plotted versus α for $A = 1$, $\sigma = 0.2$, and $\rho = 0.9$. As expected, larger detection probabilities are achieved as α increases. In addition, similar tradeoffs to those in the previous scenario are observed from the figure.

V. Alternative Formulation

Although the formulation in (4)–(6) takes into account uncertainties in $w_1(\theta)$ only, it is possible to extend the results in order to impose a similar constraint also on $w_0(\theta)$. In other words, knowledge on $w_0(\theta)$ can also be incorporated into the problem formulation. Therefore, in this section we provide an alternative formulation that incorporates both the uncertainties in $w_0(\theta)$ and $w_1(\theta)$, and provides an explicit model for the prior uncertainties.

Consider an ε-contaminated model [38] and express the true prior distribution as $w_i^{\rm tr}(\theta) = (1-\varepsilon_i)w_i(\theta) + \varepsilon_i h_i(\theta)$ for $i = 0, 1$, where $w_i(\theta)$ denotes the estimated prior distribution and $h_i(\theta)$ is any unknown probability distribution. In other words, the prior distributions are known as $w_0(\theta)$ and $w_1(\theta)$ with some uncertainty, and the amount of uncertainty is controlled by ε_0 and ε_1. For example, $w_0(\theta)$ and $w_1(\theta)$ can be pdf estimates based on previous decisions (experience), and ε_0 and ε_1 can be determined depending on certain metrics of the estimators, such as the variances of the parameter estimators. Let $\mathcal{W}_i$ denote the set of all possible prior distributions $w_i^{\rm tr}(\theta)$ according to the ε-contaminated model above. Then, the following problem formulation can be considered:

$$\begin{aligned} \max_{\phi} \quad & \min_{w_1^{\rm tr}(\theta)\in\mathcal{W}_1} \int P_D(\phi;\theta)\, w_1^{\rm tr}(\theta)\, d\theta \\ \text{subject to} \quad & \max_{w_0^{\rm tr}(\theta)\in\mathcal{W}_0} \int P_F(\phi;\theta)\, w_0^{\rm tr}(\theta)\, d\theta \le \alpha. \end{aligned} \tag{31}$$

Based on the ε-contaminated model, the problem in (31) can also be expressed from (2) and (3) as

$$\begin{aligned} \max_{\phi} \quad & (1-\varepsilon_1)\int\int \phi(\mathbf{x})p_\theta(\mathbf{x})w_1(\theta)\, d\theta\, d\mathbf{x} \\ & + \varepsilon_1 \min_{h_1(\theta)} \int\int \phi(\mathbf{x})p_\theta(\mathbf{x})h_1(\theta)\, d\theta\, d\mathbf{x} \\ \text{subject to} \quad & \max_{h_0(\theta)} (1-\varepsilon_0)\int\int \phi(\mathbf{x})p_\theta(\mathbf{x})w_0(\theta)\, d\theta\, d\mathbf{x} \\ & + \varepsilon_0 \int\int \phi(\mathbf{x})p_\theta(\mathbf{x})h_0(\theta)\, d\theta\, d\mathbf{x} \le \alpha. \end{aligned} \tag{32}$$

Let $p_i(\mathbf{x}) = \int p_\theta(\mathbf{x})w_i(\theta)\, d\theta$ for $i = 0, 1$. In addition, since

$$\min_{h_1(\theta)} \int\int \phi(\mathbf{x})p_\theta(\mathbf{x})h_1(\theta)\, d\theta\, d\mathbf{x} = \min_{\theta\in\Lambda_1} \int \phi(\mathbf{x})p_\theta(\mathbf{x})\, d\mathbf{x}$$

and

$$\max_{h_0(\theta)} \int\int \phi(\mathbf{x})p_\theta(\mathbf{x})h_0(\theta)\, d\theta\, d\mathbf{x} = \max_{\theta\in\Lambda_0} \int \phi(\mathbf{x})p_\theta(\mathbf{x})\, d\mathbf{x},$$

(32) becomes

$$\max_{\phi} (1-\varepsilon_1)\int \phi(\mathbf{x})p_1(\mathbf{x})\, d\mathbf{x} + \varepsilon_1 \min_{\theta\in\Lambda_1} \int \phi(\mathbf{x})p_\theta(\mathbf{x})\, d\mathbf{x} \tag{33}$$

$$\text{subject to} \max_{\theta\in\Lambda_0} \int \phi(\mathbf{x})\left[(1-\varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_\theta(\mathbf{x})\right] d\mathbf{x} \le \alpha. \tag{34}$$

It is noted from (12)–(13) and (33)–(34) that the objective functions are in the same form but the constraints are somewhat different in the optimization problems considered in Section III and in this section. Since the proof of Theorem 1 in [30] focuses on the maximization of the objective function considering only the NP decision rules that satisfy the false-alarm constraint, the same proof applies to the problem in (33)–(34) as well if we consider the NP decision rules under the constraint in (34) and define $v(\theta) = (1-\varepsilon_1)w_1(\theta) + \varepsilon_1\mu(\theta)$. Therefore, Theorem 1 is valid in this scenario when the NP solution for $v(\theta)$ under the false-alarm constraint is updated as follows (see Remark 1):

$$\phi^*(\mathbf{x}) = \begin{cases} 1, & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta > \eta\left[(1-\varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\tilde\theta_0}(\mathbf{x})\right] \\ \kappa(\mathbf{x}), & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta = \eta\left[(1-\varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\tilde\theta_0}(\mathbf{x})\right] \\ 0, & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x})\, v(\theta)\, d\theta < \eta\left[(1-\varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\tilde\theta_0}(\mathbf{x})\right] \end{cases} \tag{35}$$

where $\eta \ge 0$ and $0 \le \kappa(\mathbf{x}) \le 1$ are such that $\max_{\theta\in\Lambda_0} \int \phi^*(\mathbf{x})\left[(1-\varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_\theta(\mathbf{x})\right] d\mathbf{x} = \alpha$, and $\tilde\theta_0$ is defined as

$$\tilde\theta_0 = \arg\max_{\theta\in\Lambda_0} \int \phi^*(\mathbf{x})[(1-\varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_\theta(\mathbf{x})]d\mathbf{x}. \tag{36}$$

Hence, the solution of the problem in (33) and (34) can be expressed by the LRT specified in (35) and (36), once a pdf $\mu(\theta)$ and the corresponding decision rule ϕ^* that satisfy the condition in Theorem 1 are obtained.

The problem formulation in (31) can also be regarded as an application of the Γ-minimax approach [12] to the Neyman–Pearson framework, or as Neyman–Pearson testing under interval probability [28], [29]. Although the mathematical approach in obtaining the optimal solution is similar to that of the restricted NP approach investigated in the previous sections, there exist significant differences between these approaches. For the approach in this section, uncertainty needs to be

modeled by a class of possible prior distributions, then the prior distribution that minimizes the detection probability is considered for the alternative hypothesis.[7] On the other hand, the restricted NP approach in (4)–(6) focuses on a scenario in which one has a single prior distribution (e.g., a prior distribution estimate from previous experience) but can only consider decision rules whose detection probability is constrained by a lower limit. In other words, the main idea is that "one can utilize the prior information, but in a way that will be guaranteed to be acceptable to the frequentist who wants to limit frequentist risk" (detection probability in this scenario) [12]. Therefore, there is no model assumption in the restricted NP approach; hence, no efforts are required to find the best model. The two performance metrics, the average and the minimum detection probabilities, can be investigated in order to decide the best value of β. As stated in [39], it can be challenging to represent some uncertainty types via certain mathematical models such as the ε-contaminated class. Therefore, the restricted NP approach can also be useful in such scenarios.

VI. Concluding Remarks and Extensions

In this study, a restricted NP framework has been investigated for composite hypothesis-testing problems in the presence of prior information uncertainty. The optimal decision rule according to the restricted NP criterion has been analyzed and an algorithm has been proposed to calculate it. In addition, it has been observed that the restricted NP decision rule can be specified as a classical NP decision rule corresponding to the least favorable distribution. Furthermore, the average detection probability achieved by the restricted NP approach has been shown to be a strictly decreasing and concave function of the constraint on the worst-case detection probability. Finally, numerical examples have been presented in order to investigate and illustrate the theoretical results.

Similar to the extensions of the restricted Bayesian approach in [6], the notion of a restricted NP decision rule can be extended to cover more generic scenarios, in which there exist sets of distribution families $\Upsilon_0, \Upsilon_1, \ldots, \Upsilon_M$ such that $\Upsilon_0 \subset \Upsilon_1 \cdots \subset \Upsilon_M$. (Suppose we are certain that the prior distribution under the alternative hypothesis lies in Υ_M; that is, $w_1(\theta) \in \Upsilon_M$. However, we get less sure that it lies in Υ_i as i decreases.) Please see [30, p. 146], for the extension of the restricted NP formulation in (9)–(11) and for the generalization of Theorem 1 to this scenario.

Appendix
Proof of Theorem 3

Based on the definition of the restricted NP problem in (4)–(6), $P_D^{\text{avg}}\left(\phi_r^{\beta}\right)$ in (27) is a nonincreasing function of β since larger β values result in a smaller feasible set of decision rules for the optimization problem. In order to use this observation in proving the concavity of $P_D^{\text{avg}}\left(\phi_r^{\beta}\right)$, define a new decision rule as a randomization [1], [6] of two restricted NP decision rules as follows:

$$\phi \triangleq \varsigma \phi_r^{\beta_1} + (1-\varsigma)\phi_r^{\beta_2} \tag{37}$$

where $0 \le \beta_1 < \beta_2 \le U$ and $0 < \varsigma < 1$. From the definition of ϕ, the following equations can be obtained for the detection and false-alarm probabilities of ϕ for specific parameter values:

$$P_D(\phi;\theta) = \varsigma\, P_D\left(\phi_r^{\beta_1};\theta\right) + (1-\varsigma)\, P_D\left(\phi_r^{\beta_2};\theta\right), \quad \theta \in \Lambda_1 \tag{38}$$

$$P_F(\phi;\theta) = \varsigma\, P_F\left(\phi_r^{\beta_1};\theta\right) + (1-\varsigma)\, P_F\left(\phi_r^{\beta_2};\theta\right), \quad \theta \in \Lambda_0. \tag{39}$$

[7]Similarly, the prior distribution that maximizes the false alarm probability is considered for the null hypothesis.

The relation in (39) can be used to show that ϕ is an α-level decision rule. That is,

$$\max_{\theta\in\Lambda_0} P_F(\phi;\theta) \le \varsigma \max_{\theta\in\Lambda_0} P_F\left(\phi_r^{\beta_1};\theta\right) + (1-\varsigma)\max_{\theta\in\Lambda_0} P_F\left(\phi_r^{\beta_2};\theta\right) \le \alpha \tag{40}$$

where (6) is used to obtain the second inequality.

Based on (37) and (38), the average detection probability of ϕ can be calculated as

$$\begin{aligned} P_D^{\text{avg}}(\phi) &= \int_{\Lambda_1} P_D(\phi;\theta)\, w_1(\theta)\, d\theta \\ &= \varsigma\, P_D^{\text{avg}}(\phi_r^{\beta_1}) + (1-\varsigma)\, P_D^{\text{avg}}(\phi_r^{\beta_2}). \end{aligned} \tag{41}$$

Also, from (38), the worst-case detection probability of ϕ can be upper bounded as follows:

$$\begin{aligned} \min_{\theta\in\Lambda_1} P_D(\phi;\theta) &\ge \varsigma \min_{\theta\in\Lambda_1} P_D\left(\phi_r^{\beta_1};\theta\right) + (1-\varsigma)\min_{\theta\in\Lambda_1} P_D\left(\phi_r^{\beta_2};\theta\right) \\ &\ge \varsigma\,\beta_1 + (1-\varsigma)\,\beta_2. \end{aligned} \tag{42}$$

Defining $\beta \triangleq \min_{\theta\in\Lambda_1} P_D(\phi;\theta)$ and $\beta^* \triangleq \varsigma\,\beta_1 + (1-\varsigma)\,\beta_2$, the relations in (41) and (42) can be used to obtain the following inequalities:

$$\begin{aligned} P_D^{\text{avg}}\left(\phi_r^{\beta^*}\right) \ge P_D^{\text{avg}}\left(\phi_r^{\beta}\right) \ge P_D^{\text{avg}}(\phi) &= \varsigma\, P_D^{\text{avg}}\left(\phi_r^{\beta_1}\right) \\ &+ (1-\varsigma)\, P_D^{\text{avg}}\left(\phi_r^{\beta_2}\right) \end{aligned} \tag{43}$$

where the first inequality follows from the nonincreasing property of $P_D^{\text{avg}}(\phi_r^{\beta})$ explained at the beginning of the proof [since $\beta \ge \beta^*$ as shown in (42)], and the second inequality is obtained from the fact that the restricted NP decision rule ϕ_r^{β} maximizes the average detection probability under a given constraint β on the worst case detection probability (among all α-level decision rules). Thus, the concavity of $P_D^{\text{avg}}(\phi_r^{\beta})$ is proven.

In order to prove the strictly decreasing property, it is first shown that for any $L < \beta < U$

$$\min_{\theta\in\Lambda_1} P_D\left(\phi_r^{\beta};\theta\right) = \beta. \tag{44}$$

Assume that $\min_{\theta\in\Lambda_1} P_D(\phi_r^{\beta};\theta) > \beta$. Then, there exists an α-level classical NP decision rule ϕ_c and $0 < \varsigma < 1$ such that an α-level decision rule ϕ can be defined as $\phi \triangleq \varsigma\,\phi_c + (1-\varsigma)\,\phi_r^{\beta}$, which satisfies $\min_{\theta\in\Lambda_1} P_D(\phi;\theta) = \beta$. It should be noted that ϕ_c achieves a smaller minimum detection probability and a higher average detection probability than ϕ_r^{β} for any $L < \beta < U$ by definition. Therefore, the average detection probability of ϕ satisfies $P_D^{\text{avg}}(\phi) > P_D^{\text{avg}}(\phi_r^{\beta})$, which contradicts with the definition of the restricted NP. Hence, $\min_{\theta\in\Lambda_1} P_D(\phi_r^{\beta};\theta) > \beta$ cannot be true, which proves the result in (44). Next, let $L < \beta_1 < \beta_2 < U$ and suppose that $P_D^{\text{avg}}(\phi_r^{\beta_1}) = P_D^{\text{avg}}(\phi_r^{\beta_2})$. Obviously, this implies that $\phi_r^{\beta_2}$ is also a solution corresponding to β_1, which contradicts with the result in (44). Therefore, $P_D^{\text{avg}}(\phi_r^{\beta_1}) > P_D^{\text{avg}}(\phi_r^{\beta_2})$ must hold. Hence, $P_D^{\text{avg}}(\phi_r^{\beta})$ is a strictly decreasing function of β. □

References

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

[2] E. L. Lehmann, "The Fisher, Neyman–Pearson theories of testing hypotheses: One theory or two?," *J. Amer. Stat. Assoc.*, vol. 88, pp. 1242–1249, Dec. 1993.

[3] J. O. Berger, "Could Fisher, Jeffreys and Neyman have agreed on testing?," *Stat. Sci.*, vol. 18, pp. 1–12, Feb. 2003.

[4] W. E. Strawderman, "Minimaxity," *J. Amer. Stat. Assoc.*, vol. 95, pp. 1364–1368, Dec. 2000.

[5] J. R. Blum and J. Rosenblatt, "On partial *a priori* information in statistical inference," *Ann. Math. Stat.*, vol. 38, no. 6, pp. 1671–1678, 1967.
[6] J. L. Hodges, Jr. and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Stat.*, vol. 23, pp. 396–407, Sep. 1952.
[7] H. Robbins, "An empirical Bayes approach to statistics," in *Proc. 3rd Berkeley Symp. Math. Stat. Probability 1*, Berkeley, CA, 1955, pp. 157–164.
[8] H. Robbins, "The empirical Bayes approach to statistical decision problems," *Ann. Math. Stat.*, vol. 35, no. 1, pp. 1–20, 1964.
[9] L. J. Savage, *The Foundations of Statistics*, 2nd ed. New York: Dover, 1972.
[10] S. R. Watson, "On Bayesian inference with incompletely specified prior distributions," *Biometrika*, vol. 61, pp. 193–196, 1974.
[11] H. Kudo, "On partial prior information and property of parametric sufficiency," in *Proc. 5th Berkeley Symp. Math. Stat. and Prob.*, 1967, vol. 1, pp. 251–265.
[12] M. J. Bayarri and J. O. Berger, "The interplay of Bayesian and frequentist analysis," *Stat. Sci.*, vol. 19, pp. 58–80, 2004.
[13] S. Bayram and S. Gezici, "Noise enhanced M-ary composite hypothesis-testing in the presence of partial prior information," *IEEE Trans. Signal Process.*, vol. 59, pp. 1292–1297, Mar. 2011.
[14] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer, 1985.
[15] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer, 2005.
[16] D. Ellsberg, "Risk, ambiguity, and the savage axioms," *Quart. J. Econom.*, vol. 75, pp. 643–669, Nov. 1961.
[17] F. H. Knight, *Risk, Uncertainty and Profit*. Boston, MA: Houghton Mifflin, 1921.
[18] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Part I*, 2nd ed. New York, NY: Wiley, 2001.
[19] J. Cvitanic and I. Karatzas, "Generalized Neyman-Pearson lemma via convex duality," *Bernoulli*, vol. 7, no. 1, pp. 79–97, 2001.
[20] B. Rudloff and I. Karatzas, Testing Composite Hypotheses Via Convex Duality, 2008 [Online]. Available: http://arxiv.org/abs/0809.4297
[21] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Ann. Math. Stat.*, vol. 36, pp. 369–400, 1965.
[22] S. Natarajan, "Large deviations, hypotheses testing, and source coding for finite Markov chains," *IEEE Trans. Inf. Theory*, vol. 31, pp. 360–365, May 1985.
[23] O. Zeitouni and M. Gutman, "On universal hypotheses testing via large deviations," *IEEE Trans. Inf. Theory*, vol. 37, pp. 285–290, Mar. 1991.
[24] E. Levitan and N. Merhav, "A competitive Neyman-Pearson approach to universal hypothesis testing with applications," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2215–2229, Aug. 2002.
[25] E. L. Lehmann, "Some history of optimality," *Lecture Notes-Monograph Series*, vol. 57, pp. 11–17, 2009.
[26] N. Begum and M. L. King, "A new class of test for testing a composite null against a composite alternative hypothesis," in *Proc. Australasian Meeting Econom. Soc.*, 2007.
[27] J. O. Berger, B. Boukai, and Y. Wang, "Unified frequentist and Bayesian testing of a precise hypothesis," *Stat. Sci.*, vol. 12, no. 3, pp. 133–160, 1997.
[28] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Stat.*, vol. 1, pp. 251–263, Mar. 1973.
[29] T. Augustin, "Neyman–Pearson testing under interval probability by globally least favorable pairs: A survey of Huber–Strassen theory and some results on its extension to general interval probability," *J. Stat. Planning Inference*, vol. 105, pp. 149–173, Jun. 2002.
[30] S. Bayram, "Alternative approaches and noise benefits in hypothesis-testing problems in the presence of partial information," Ph.D. dissertation, Bilkent Univ., Ankara, Turkey, 2011.
[31] A. Wald, *Statistical Decision Functions*. New York: Wiley, 1950.
[32] K. E. Parsopoulos and M. N. Vrahatis, *Particle Swarm Optimization Method for Constrained Optimization Problems*. Amsterdam, The Netherlands: IOS Press, 2002, pp. 214–220.
[33] X. Hu and R. Eberhart, "Solving constrained nonlinear optimization problems with particle swarm optimization," presented at the World Multiconf. Systemics, Cybern., Inf., Orlando, FL, 2002.
[34] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. New York: Springer, 2005.
[35] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2000.
[36] S. Bayram, S. Gezici, and H. V. Poor, "Noise enhanced hypothesis-testing in the restricted Bayesian framework," *IEEE Trans. Signal Process.*, vol. 58, pp. 3972–3989, Aug. 2010.
[37] I. Korn, J. P. Fonseka, and S. Xing, "Optimal binary communication with nonequal probabilities," *IEEE Trans. Commun.*, vol. 51, pp. 1435–1438, Sep. 2003.
[38] J. Berger and L. M. Berliner, "Robust Bayes and empirical Bayes analysis with ε-contaminated priors," *Ann. Stat.*, vol. 14, no. 2, pp. 461–486, 1986.
[39] J. Berger *et al.*, "An overview of robust Bayesian analysis," *Test*, vol. 3, pp. 5–124, 1994.

A Noniterative Frequency Estimator With Rational Combination of Three Spectrum Lines

Cui Yang and Gang Wei

***Abstract*—A noniterative frequency estimator is proposed in this paper. The maximum bin of fast Fourier transform (FFT) is searched as a coarse estimation, then the rational combination of three spectrum lines (RCTSL) is used as the fine estimation. Based on least square approximation in frequency domain, the combinational weights of RCTSL are found to be constants depending only on data length, therefore the RCTSL is very computational efficient. Theoretical bound is also obtained, which shows the performance of RCTSL approaches the Cramér–Rao Bound (CRB). Simulation results demonstrate that RCTSL has a lower signal-to-noise ratio (SNR) threshold compared with other known estimators.**

***Index Terms*—Fourier transforms, frequency estimation, least square approximation.**

I. Introduction

Frequency estimation for exponential signals in additive white Gaussian noise is of great importance in a wide range of applications such as signal interception, detection, wireless communications, etc. The complex exponential sample data is presented as

$$y(n)=x(n)+z(n)=Ae^{j(\omega n+\varphi)}+z(n),\quad n=0,1\ldots,N-1 \quad (1)$$

where A, $\omega \in [-\pi,\pi]$ and $\varphi \in [-\pi,\pi]$ are signal amplitude, frequency and phase, respectively, which are all unknown constants. It is the frequency ω which we are interested to estimate. $z(n)$ is complex white Gaussian noise with zero mean and unknown variance σ^2. Given

Manuscript received May 28, 2010; revised October 13, 2010 and February 13, 2011; accepted June 07, 2011. Date of publication June 23, 2011; date of current version September 14, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sofia C. Olhede. This work was supported by the Guangdong Provincial key lab of Short-Range Wireless Detections and Communications, China, and the National Natural Science Foundation of China and Guangdong province under Grant 60625101, U1035003, 9351064101000003, and the Fundamental Research Funds for the Central Universities (Grants 2011ZM0034 and 2009ZM0005).

C. Yang is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510640, China (e-mail: yangcui26@163.com).

G. Wei is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: ecgwei@scut.edu.cn).

Digital Object Identifier 10.1109/TSP.2011.2160257

1053-587X/$26.00 © 2011 IEEE