

ALTERNATIVE APPROACHES AND NOISE
BENEFITS IN HYPOTHESIS-TESTING PROBLEMS IN
THE PRESENCE OF PARTIAL INFORMATION

A DISSERTATION SUBMITTED TO
THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS
ENGINEERING
AND GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By

Suat Bayram

July 2011

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Asst. Prof. Dr. Sinan Gezici (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Prof. Dr. Orhan Arıkan

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Asst. Prof. Dr. Selim Aksoy

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Asst. Prof. Dr. Defne Aktaş

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Asst. Prof. Dr. Ali Cafer Gürbüz

Approved for the Graduate School of Engineering and Science :

Prof. Dr. Levent Onural
Director of Graduate School of Engineering and Science

ABSTRACT

ALTERNATIVE APPROACHES AND NOISE BENEFITS IN HYPOTHESIS-TESTING PROBLEMS IN THE PRESENCE OF PARTIAL INFORMATION

Suat Bayram

Ph.D. in Electrical and Electronics Engineering

Supervisor: Asst. Prof. Dr. Sinan Gezici

July 2011

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

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

In the third part of the dissertation, the effects of additive noise are studied for binary composite hypothesis-testing problems. A Neyman-Pearson (NP) framework is considered, and the maximization of detection performance under a constraint on the maximum probability of false-alarm is studied. The detection performance is quantified in terms of the sum, the minimum and the maximum of the detection probabilities corresponding to possible parameter values under the alternative hypothesis. Sufficient conditions under which detection performance can or cannot be improved are derived for each case. Also, statistical characterization of optimal additive noise is provided, and the resulting false-alarm probabilities and bounds on detection performance are investigated. In addition, optimization theoretic approaches for obtaining the probability distribution of optimal additive noise are discussed. Finally, a detection example is presented to investigate the theoretical results.

Finally, the restricted 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. Optimal decision rules according to the restricted NP criterion are investigated, and an algorithm is provided to calculate the optimal restricted NP decision rule. In addition, it is observed 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, and extensions to more generic scenarios are discussed.

Keywords: Hypothesis-testing, noise enhanced detection, restricted Bayes, stochastic resonance, composite hypotheses, Bayes risk, Neyman-Pearson, max-min, least-favorable prior.

ÖZET

KISMİ BİLGİ BULUNAN HİPOTEZ SINAMA PROBLEMLERİNDE ALTERNATİF YAKLAŞIMLAR VE GÜRÜLTÜ KAZANIMLARI

Suat Bayram

Elektrik ve Elektronik Mühendisliği, Doktora

Tez Yöneticisi: Yrd. Doç. Dr. Sinan Gezici

Temmuz 2011

Optimal olmayan bazı sezicilerin performansı, gözlemlerine bağımsız gürültü eklenerek artırılabilir. Tezin ilk kısmında ek gürültünün etkileri, Bayes ve minimaks kriterlerinin genelleştirilmesini sağlayan kısıtlı Bayes kriterine göre çalışılmaktadır. Genel M 'li bileşik hipotez sınamaları baz alınarak, ek gürültünün optimal olasılık dağılım fonksiyonu incelenmektedir. Aynı zamanda, sezicinin performansının gürültü eklenerek geliştirilip geliştirilemeyeceğiyle ilgili yeter koşullar türetilmektedir. Bunlara ek olarak, basit hipotez sınama problemleri daha ayrıntılı olarak çalışılmakta ve basit hipotezlere özel ek yeter koşullar elde edilmektedir. Ayrıca, belli koşullar altında, bir basit M 'li hipotez sınama problemindeki optimal ek gürültünün olasılık yoğunluk fonksiyonunun, en fazla M farklı değer arasında rasgeleleştirme içerdiği gösterilmektedir. Daha sonra, optimal gürültü dağılımını elde etmek için global optimizasyon, analitik ve dışbükey gevşetme yaklaşımları ele alınmaktadır. Son olarak, kuramsal sonuçları incelemek için sezim örnekleri sunulmaktadır.

Tezin ikinci kısmında, kısmi önsel bilgi bulunan bileşik M 'li hipotez sına ma problemleri için ek gürültünün etkileri çalışılmaktadır. Optimal ek gürültü, bilinmeyen önsel olasılıklar için birbiçimli dağılım (kriter 1) veya en az uygun dağılım (kriter 2) varsayan iki kritere göre elde edilmektedir. Her bir kriter için optimal gürültünün istatistiksel özellikleri elde edilmektedir. Özel olarak, optimal gürültünün kriter 1'e göre sabit bir sinyal seviyesiyle ya da kriter 2'ye göre sonlu sayıdaki sinyal seviyesinin rasgeleleştirilmesiyle ifade edilebileceği gösterilmektedir. Bunlara ek olarak, bazı bileşik hipotezler altındaki parametre dağılımlarının bilinmediği durumlar ele alınmakta ve risklerin üzerine üst sınırlar elde edilmektedir. Son olarak, kuramsal sonuçları göstermek için bir sezim örneği sunulmaktadır.

Tezin üçüncü kısmında, ek gürültünün ikili bileşik hipotez sına ma problemleri üzerindeki etkileri çalışılmaktadır. Bir Neyman-Pearson (NP) çerçevesi ele alınmakta ve en yüksek yanlış alarm olasılığı üzerindeki sınırlama altında sezim performansının en yüksek seviyeye çıkarılmasına çalışılmaktadır. Sezim performansı, alternatif hipotez altındaki muhtemel parametre değerlerine karşılık gelen sezim olasılıklarının toplamı, minimumu ve maksimumu cinsinden hesaplanmaktadır. Her bir durum için sezim performansının geliştirilip geliştirilemeyeceğiyle ilgili yeter koşullar türetilmektedir. Aynı zamanda, optimal ek gürültünün istatistiksel özellikleri sunulmakta ve ortaya çıkan yanlış alarm olasılıkları ve sezim performansı üzerindeki sınırlar incelenmektedir. Bunlara ilave olarak, optimal ek gürültünün olasılık dağılımını elde etmek için optimizasyon kuramı tabanlı yaklaşımlar tartışılmaktadır. Son olarak, kuramsal sonuçları incelemek için bir sezim örneği sunulmaktadır.

Son olarak, alternatif hipotez altındaki önsel olasılık dağılımında belirsizlik bulunan bileşik hipotez sına ma problemleri için kısıtlı NP yaklaşımı çalışılmaktadır. Kısıtlı NP karar kuralı, en kötü durumdaki sezim ve yanlış alarm olasılıkları üzerindeki kısıtlamalar altında, ortalama sezim olasılığını

en yüksek seviyeye çıkarmayı hedefler ve en kötü durumdaki sezim olasılığı üzerindeki kısıtlama seviyesini, önsel olasılık dağılımındaki belirsizliğin miktarına göre ayarlar. Kısıtlı NP kriterine göre optimal karar kuralları incelenmekte ve optimal kısıtlı NP karar kuralının hesaplanması için bir algoritma sağlanmaktadır. Bunlara ek olarak, ortalama sezim olasılığının, minimum sezim olasılığı üzerindeki kısıtlama seviyesinin kesin azalan ve içbükey bir fonksiyonu olduğu gözlenmektedir. Son olarak, bir sezim örneği sunulmakta ve daha genel senaryolara genişletimler tartışılmaktadır.

Anahtar Kelimeler: Hipotez sınama, gürültüyle geliştirilmiş sezim, kısıtlı Bayes, stokastik rezonans, bileşik hipotezler, Bayes riski, Neyman-Pearson, maks-min, en az uygun önsel.

ACKNOWLEDGMENTS

I was so lucky to have Asst. Prof. Dr. Sinan Gezici as my advisor. He has been one of the few people who had vital influence on my life. His patience, perfectionist personality, generosity and inspirational nature have been a great admiration for me. He has always supported me through hard times. It was a real privilege and honor for me to work with such a visionary advisor. I would like to, especially, thank him for providing me great research opportunities and environment. Also I would like to thank Prof. Dr. Orhan Arıkan, Asst. Prof. Dr. Selim Aksoy, Asst. Prof. Dr. Defne Aktaş and Asst. Prof. Dr. Ali Cafer Gürbüz for agreeing to serve on my thesis committee. I would also like to thank TÜBİTAK for its financial support which was vital for me.

I also extend my special thanks to Sara Bergene, Zakir Sözüduyar, Mehmet Barış Tabakcıođlu, Kadir Üstün, Mustafa Ürel, Tolga Özaslan, Abdülkadir Eryıldırım, Saba Öz, Mahmut Yavuzer, Aykut Yıldız, Burak Şekerlisoy, Vahdetin Taş, Hamza Sođancı, M. Emin Tutay, Gökçe Osman Balkan, Osman Gürlevik, Yaşar Kemal Alp and Semih Çaycı for being wonderful friends and sharing unforgettable moments together.

Finally, I would like to give a special thank to my mother Zehra, my brothers Yakup, Cavit (İmdat), Acar Alp and my true friend Elif Eda Demir for their unconditional love and support throughout my studies. They mean everything to me. I have much gratitude towards my mother for helping me believe that there is nothing one cannot accomplish.

Contents

1	Introduction	1
1.1	Objectives and Contributions of the Dissertation	1
1.2	Organization of the Dissertation	10
2	Noise Enhanced Hypothesis-Testing in the Restricted Bayesian Framework	11
2.1	Noise Enhanced M -ary Composite Hypothesis-Testing	12
2.1.1	Problem Formulation and Motivation	12
2.1.2	Improvability and Nonimprovability Conditions	15
2.1.3	On the Optimal Additive Noise	20
2.2	Noise Enhanced Simple Hypothesis-Testing	21
2.2.1	Problem Formulation	22
2.2.2	Optimal Additive Noise	24
2.2.3	Improvability and Nonimprovability Conditions	31
2.3	Numerical Results	34

2.4	Concluding Remarks	49
2.5	Appendices	51
2.5.1	Proof of Theorem 2	51
2.5.2	Proof of Theorem 3	53
2.5.3	Maximum Conditional Risk Achieved by Optimal Noise	53
2.5.4	Proof of Theorem 5	54
2.5.5	Proof of Corollary 1	56
3	Noise Enhanced M-ary Composite Hypothesis-Testing in the Presence of Partial Prior Information	57
3.1	Problem Formulation	58
3.2	Optimal Additive Noise According to Criterion 1	61
3.3	Optimal Additive Noise According to Criterion 2	63
3.4	Unknown Parameter Distributions for Some Hypotheses	65
3.5	A Detection Example and Conclusions	69
3.6	Appendices	73
3.6.1	Proof of Proposition 1	73
4	Noise Enhanced Binary Composite Hypothesis-Testing in the Neyman-Pearson Framework	75
4.1	Problem Formulation and Motivation	76
4.2	Max-Sum Criterion	80

4.2.1	Improvability and Non-improvability Conditions	81
4.2.2	Characterization of Optimal Solution	84
4.2.3	Calculation of Optimal Solution and Convex Relaxation	87
4.3	Max-Min Criterion	91
4.3.1	Improvability and Non-improvability Conditions	91
4.3.2	Characterization of Optimal Solution	94
4.3.3	Calculation of Optimal Solution and Convex Relaxation	97
4.4	Max-Max Criterion	100
4.5	Numerical Results	102
4.5.1	Scenario-1: Λ_0 and Λ_1 with finite number of elements	104
4.5.2	Scenario-2: Λ_0 and Λ_1 are continuous intervals	109
4.6	Concluding Remarks and Extensions	117
5	On the Restricted Neyman-Pearson Approach for Composite Hypothesis-Testing in the Presence of Prior Distribution Uncertainty	118
5.1	Problem Formulation and Motivation	119
5.2	Analysis of Restricted Neyman-Pearson Approach	122
5.2.1	Characterization of Optimal Decision Rule	123
5.2.2	Calculation of Optimal Decision Rule	127

5.2.3	Properties of Average Detection Probability in Restricted NP Solutions	131
5.3	Numerical Results	133
5.4	Alternative Formulation	141
5.5	Concluding Remarks and Extensions	145
5.6	Appendices	147
5.6.1	Proof of Theorem 1	147
5.6.2	Proof of Theorem 2	148
5.6.3	Proof of Theorem 3	149
6	Conclusions and Future Work	151

List of Figures

- 2.1 Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.08$ and $A = 1$. . . 36
- 2.2 Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.12$ and $A = 1$. . . 37
- 2.3 Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.4$ and $A = 1$. . . 39
- 2.4 Bayes risks of original and noise modified detectors versus A in cases of equal priors and unequal priors for $\alpha = 0.08$ and $\sigma = 0.05$. 41
- 2.5 Improvement ratio versus α in the cases of equal priors and unequal priors for $\sigma = 0.01$, $\sigma = 0.05$ and $\sigma = 0.1$, where $A = 1$. . . 42
- 2.6 The second order derivative of $F_0(x)$ at $x = 0$ versus σ for various values of A . Both Theorem 5 and Theorem 3 imply for the detection example in this section that the detector is improvable whenever $F_0''(0)$ is negative. The limit on the conditional risks, α , is set to the original conditional risks for each value of σ . The graph for $A = 1$ is scaled by 0.1 to make view of the figure more convenient (since only the signs of the graphs are important). . . 46

2.7	Bayes risks of original and noise modified detectors versus σ for $\alpha = 0.4$ and $A = 1$	48
2.8	Bayes risks of original and noise modified detectors versus A for $\alpha = 0.4$ and $\sigma = 0.05$	50
3.1	Independent noise \mathbf{n} is added to observation \mathbf{x} in order to improve the performance of the detector, represented by $\phi(\cdot)$	58
3.2	Bayes risks of the original and noise modified detectors versus σ for $A = 1$ according to both criteria.	72
4.1	Independent noise \mathbf{n} is added to data vector \mathbf{x} in order to improve the performance of the detector, $\phi(\cdot)$	76
4.2	Probability mass functions of the optimal additive noise based on the PSO and the convex relaxation techniques for the max-sum case when $A = 1$ and $\sigma = 1$	105
4.3	Probability mass functions of the optimal additive noise based on the PSO and the convex relaxation techniques for the max-min case when $A = 1$ and $\sigma = 1$	106
4.4	Comparison of detection probabilities (normalized) in the absence (“original”) and presence (“SR”) of additive noise according to the max-sum criterion for various values of σ	107
4.5	Comparison of detection probabilities (normalized) in the absence (“original”) and presence (“SR”) of additive noise according to the max-min criterion for various values of σ	108

4.6	The second-order derivatives of $H(t)$ in (4.17) and $H_{\min}(t)$ (4.36) at $t = \alpha$ for various values of σ . Theorem 1 and Theorem 5 imply that the detector is improvable whenever the second-order derivative at $t = \alpha$ is positive.	110
4.7	The optimal additive noise p.d.f. in (4.69) for $A = 1$ and $\sigma = 1$ according to the max-sum criterion. The optimal parameters in (4.69) obtained via the PSO algorithm are $\boldsymbol{\mu} = [0.0969 \ 0 \ 0.0019 \ 0.1401 \ 0.1377 \ 0.0143 \ 0.1470 \ 0.4621]$, $\boldsymbol{\eta} = [25.4039 \ -20.1423 \ 13.7543 \ 17.0891 \ 29.7452 \ -25.0785 \ 17.6887 \ -2.2085]$, and $\boldsymbol{\sigma} = [1.3358 \ 26.2930 \ 11.3368 \ 0 \ 19.5556 \ 11.5953 \ 17.9838 \ 0.0001]$. The mass centers with very small variances ($\eta_i = 17.0891$ and $\eta_i = -2.2085$) are marked by arrows for convenience.	112
4.8	The optimal additive noise p.d.f. in (4.69) for the max-min criterion when $A = 1$ and $\sigma = 1$. The optimal parameters in (4.69) obtained via the PSO algorithm are $\boldsymbol{\mu} = [0.0067 \ 0.1797 \ 0.0411 \ 0.2262 \ 0.0064 \ 0.0498 \ 0 \ 0.4902]$, $\boldsymbol{\eta} = [20.1017 \ 15.0319 \ 0.1815 \ 29.9668 \ 17.2657 \ 22.8092 \ -0.7561 \ -1.4484]$, and $\boldsymbol{\sigma} = [16.5204 \ 15.1445 \ 0.8805 \ 10.1573 \ 12.9094 \ 17.4184 \ 19.0959 \ 0.0102]$. The mass center $\eta_i = -1.4484$ is marked by an arrow for convenience as it has a very small variance.	113
4.9	Comparison of detection probabilities (normalized) in the absence (“original”) and presence (“SR”) of additive noise according to the max-sum criterion for various values of σ	114
4.10	Comparison of detection probabilities (normalized) in the absence (“original”) and presence (“SR”) of additive noise according to the max-min criterion for various values of σ	115

4.11	The second-order derivatives of $H(t)$ in (4.17) and $H_{\min}(t)$ (4.36) at $t = \alpha$ for various values of σ . Theorem 1 and Theorem 5 imply that the detector is improvable whenever the second-order derivative at $t = \alpha$ is positive.	116
5.1	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 $\alpha = 0.2$	135
5.2	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 $\sigma = 0.2$	138
5.3	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 $\rho = 0.9$	140
5.4	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 $\rho = 0.9$	142

In memory of my father . . .

Chapter 1

Introduction

1.1 Objectives and Contributions of the Dissertation

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

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

In the Neyman-Pearson framework, the aim is to increase the probability of detection under a constraint on the probability of false alarm [12], [13], [24], [26]. In [24], an example is presented to illustrate the effects of additive noise on the detection performance for the problem of detecting a constant signal in Gaussian mixture noise. In [12], a theoretical framework for investigating the effects of additive noise on suboptimal detectors is established according to the Neyman-Pearson criterion. Sufficient conditions under which performance of a detector can or cannot be improved via additive noise are derived, and it is proven that optimal additive noise can be generated by a randomization of at most two different signal levels, which is an important result since it greatly simplifies the calculation of the optimal noise probability density function (p.d.f.).

An optimization theoretic framework is provided in [13] for the same problem, which also proves the two mass point structure of the optimal additive noise p.d.f., and, in addition, shows that an optimal noise distribution may not exist in certain scenarios.

The study in [12] is extended to variable detectors in [25], and similar observations as in the case of fixed detectors are made. Also, the theoretical framework in [12] is applied to sequential detection and parameter estimation problems in [38] and [39], respectively. In [38], a binary sequential detection problem is considered, and additive noise that reduces at least one of the expected sample sizes for the sequential detection system is obtained. In [39], improbability of estimation performance via additive noise is illustrated under certain conditions for various estimation criteria, and the form of the optimal noise p.d.f. is obtained for each criterion. The effects of noise are investigated also for detection of weak sinusoidal signals and for locally optimal detectors. In [33] and [34], detection of a weak sinusoidal signal is considered, and improvements on detection performance are investigated. In addition, [35] studies the optimization of noise and detector parameters of locally optimal detectors for the detection of a small amplitude sinusoid in non-Gaussian noise.

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

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

Although it is challenging to obtain the optimal detector according to the restricted Bayes criterion, this criterion can be quite advantageous in practical applications compared to the Bayes and minimax criteria, as studied in [42]. Therefore, in Chapter 2 of the dissertation, the aim is to consider suboptimal detectors and to investigate how their performance can be improved via additive independent noise in the restricted Bayesian framework. In other words, one motivation is to improve performance of suboptimal detectors via additive noise and to provide reasonable performance with low computational complexity. Another motivation is the theoretical interest to investigate the effects of noise on suboptimal detectors and to obtain sufficient conditions under which performance of detectors can or cannot be improved via additive noise in the restricted Bayesian framework.

In Chapter 2 of the dissertation, the effects of additive independent noise on the performance of suboptimal detectors are investigated according to the restricted Bayes criterion [48]. A generic M -ary composite hypothesis-testing problem is considered, and sufficient conditions under which a suboptimal detector can or cannot be improved are derived. In addition, various approaches to obtaining the optimal solution are presented. For simple hypothesis-testing problems, additional improvability conditions that are simple to evaluate are proposed, and it is shown that optimal additive noise can be represented by a p.d.f. with at most M mass points. Furthermore, optimization theoretic approaches to obtaining the optimal noise p.d.f. are discussed; both global optimization techniques and approximate solutions based on convex relaxation are considered. Also, an analytical approach is proposed to obtain the optimal noise p.d.f. under certain conditions. Finally, detection examples are provided to investigate the theoretical results and to illustrate the practical importance of noise enhancement.

In Chapter 3 of the dissertation, noise enhanced detection is studied in the presence of partial prior information [49]. Optimal additive noise is formulated according to two different criteria. In the first one, a uniform distribution is assumed for the unknown priors, whereas in the second one the worst-case distributions are considered for the unknown priors by taking a conservative approach, which can be regarded as a Γ -minimax approach. In both cases, the statistics of the optimal additive noise are characterized. Specifically, it is shown that the optimal additive noise can be represented by a constant signal level according to the first criterion, whereas it can be represented by a discrete random variable with a finite number of mass points according to the second criterion. Two other contributions of the study in Chapter 3 are to investigate noise enhanced detection with partial prior information in the most generic hypotheses formulation; that is, M -ary composite hypotheses, and to employ a very generic cost function in the definition of the conditional risks. Therefore, it covers some of

the previous studies on noise enhanced detection as special cases. For example, if simple¹ binary hypotheses, uniform cost assignment (UCA), and perfect prior information are assumed, the results reduce to those in [23]. As another example, if simple M -ary hypotheses and no prior information are assumed, the results reduce to those in [29]. Furthermore, for composite hypothesis-testing problems, the cases of unknown parameter distributions under some hypotheses are also considered, and upper bounds on the risks are obtained. Finally, a detection example is presented to investigate the theoretical results.

The theoretical studies in [12] and [13] on the effects of additive noise on signal detection in the Neyman-Pearson framework consider *simple* binary hypothesis-testing problems in the sense that there exists a single probability distribution (equivalently, one possible value of the unknown parameter) under each hypothesis. The main purpose of Chapter 4 is to study *composite* binary hypothesis-testing problems, in which there can be multiple possible distributions, hence, multiple parameter values, under each hypothesis [40], [50]. The Neyman-Pearson framework is considered by imposing a constraint on the *maximum* probability of false-alarm, and three detection criteria are studied [41]. In the first one, the aim is to maximize the sum of the detection probabilities for all possible parameter values under the first (alternative) hypothesis \mathcal{H}_1 (*max-sum* criterion), whereas the second one focuses on the maximization of the minimum detection probability among all parameter values under \mathcal{H}_1 (*max-min* criterion). Although it is not commonly used in practice, the maximization of the maximum detection probability among all parameter values under \mathcal{H}_1 is also studied briefly for theoretical completeness (*max-max* criterion). For all detection criteria, sufficient conditions under which performance of a suboptimal detector can or cannot be improved via additive noise are derived. Also, statistical characterization of optimal additive noise is provided in terms its p.d.f. structure in each case. In

¹A *simple* hypothesis means that there is only one possible probability distribution under the hypothesis, whereas a *composite* hypothesis corresponds to multiple possible probability distributions.

addition, the probability of false-alarm in the presence of optimal additive noise is investigated for the max-sum criterion, and upper and lower bounds on the detection performance are obtained for the max-min criterion. Furthermore, optimization theoretic approaches to obtaining the optimal additive noise p.d.f. are discussed for each detection criterion. Both particle swarm optimization (PSO) [51]-[54] and approximate solutions based on convex relaxation [55] are proposed. Finally, a detection example is provided to investigate the theoretical results.

The main contributions in Chapter 4 can be summarized as follows: 1) Theoretical investigation of the effects of additive noise in binary *composite* hypothesis-testing problem in the Neyman-Pearson framework. 2) Extension of the improvability and non-improvability results in [12] for simple hypothesis-testing problems to composite hypothesis-testing problems. 3) Statistical characterization of optimal additive noise according to various detection criteria. 4) Derivation of upper and lower bounds on the detection performance of suboptimal detectors according to the max-min criterion.

Bayesian and minimax hypothesis-testings are two common approaches for the formulation of testing [40], [56], [57]. 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 [40], [58]. 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 [59]. In most cases, prior information is incomplete and only partial prior information is available [42], [59]. 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 [42], [45], [59]-[63], which can be considered as a mixture of Bayesian and frequentist approaches [64]. The most prominent of these approaches are the empirical Bayes, Γ -minimax, restricted Bayes and mean-max approaches [42], [49], [59], [60], [63]. 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 [46], [64]. 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 [46], [64].

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 [41]. 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 [42]. An important characteristic of the restricted Bayes approach is that it combines probabilistic and non-probabilistic descriptions of uncertainty, which are also called measurable and unmeasurable uncertainty [65], [66], because the calculation of the Bayes (average) risk requires uncertainty to be measured and imposing a constraint on the conditional risks is a non-probabilistic description of uncertainty. In Chapter 5, the focus is on the application of the notion of the restricted Bayes approach to the Neyman-Pearson (NP) framework, in which probabilistic and non-probabilistic descriptions of uncertainty are combined [42].

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 [40], [67]. When the null hypothesis is composite, it is common to apply the false-alarm constraint for all possible distributions under that hypothesis [68], [69]. 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 [40], [67]. However, such a decision rule exists only under special circumstances [40]. 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 [70]-[73]. Another approach is to maximize the average detection probability under the false-alarm constraint [64], [74]-[76]. 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 [68], [69]. 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 [77], [78].²

²The 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 [40], [67].

In Chapter 5, 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 [41], [42], it is called the *restricted NP* approach in order to emphasize the considered NP framework [79]. The investigation of the restricted NP criterion provides 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.

1.2 Organization of the Dissertation

The organization of the dissertation is as follows. In Chapter 2, the effects of additive noise are investigated according to the restricted Bayes criterion, which provides a generalization of the Bayes and minimax criteria.

In Chapter 3, noise enhanced detection is studied for M -ary composite hypothesis-testing problems in the presence of partial prior information.

In Chapter 4, the effects of additive noise are investigated for binary composite hypothesis-testing problems in the NP framework.

In Chapter 5, The restricted NP approach is studied for composite hypothesis-testing problems in the presence of uncertainty in the prior probability distribution under the alternative hypothesis.

Chapter 2

Noise Enhanced

Hypothesis-Testing in the

Restricted Bayesian Framework

This chapter is organized as follows. Section 2.1 studies composite hypothesis-testing problems, and provides a generic formulation of the problem. In addition, improvability and nonimprovability conditions are presented and an approximate solution of the optimal noise problem is discussed. Then, Section 2.2 considers simple hypothesis-testing problems and provides additional improvability conditions. Also, the discrete structure of the optimal noise probability distribution is specified. Then, detection examples are presented to illustrate the theoretical results in Section 2.3. Finally, concluding remarks are made in Section 2.4.

2.1 Noise Enhanced M -ary Composite Hypothesis-Testing

2.1.1 Problem Formulation and Motivation

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

$$\mathcal{H}_i : p_{\theta}^{\mathbf{X}}(\mathbf{x}) , \theta \in \Lambda_i , \quad i = 0, 1, \dots, M - 1 , \quad (2.1)$$

where $p_{\theta}^{\mathbf{X}}(\cdot)$ represents the p.d.f. of observation \mathbf{X} for a given value of parameter, $\Theta = \theta$, and θ belongs to parameter set Λ_i under hypotheses \mathcal{H}_i . The observation (measurement), \mathbf{x} , is a vector with K components; i.e., $\mathbf{x} \in \mathbb{R}^K$, and $\Lambda_0, \Lambda_1, \dots, \Lambda_{M-1}$ form a partition of the parameter space Λ . The prior distribution of Θ is denoted by $w(\theta)$, and it is assumed that $w(\theta)$ is known with some uncertainty [41], [42]. For example, it can be a p.d.f. estimate based on previous decisions.

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

$$\phi(\mathbf{x}) = i , \quad \text{if } \mathbf{x} \in \Gamma_i , \quad (2.2)$$

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

In some cases, addition of noise to observations can improve the performance of a suboptimal detector. By adding noise \mathbf{n} to the original observation \mathbf{x} , the noise modified observation is formed as $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where \mathbf{n} has a p.d.f. denoted by $p_{\mathbf{N}}(\cdot)$, and is independent of \mathbf{x} . As in [12] and in Section II of [13], it is assumed that the detector in (2.2) is fixed, and that the only means for improving the performance of the detector is to optimize the additive noise \mathbf{n} . In other words, the aim is to find the best $p_{\mathbf{N}}(\cdot)$ according to the restricted Bayes criterion [41]; namely, to minimize the Bayes risk under certain constraints on the conditional

risks, as specified below.

$$\begin{aligned} \min_{p_{\mathbf{N}}(\cdot)} \int_{\Lambda} R_{\theta}^{\mathbf{Y}}(\phi) w(\theta) d\theta, \\ \text{subject to } \max_{\theta \in \Lambda} R_{\theta}^{\mathbf{Y}}(\phi) \leq \alpha, \end{aligned} \quad (2.3)$$

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

$$R_{\theta}^{\mathbf{Y}}(\phi) = E\{C[\phi(\mathbf{Y}), \Theta] \mid \Theta = \theta\} = \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \quad (2.4)$$

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

In the restricted Bayes formulation in (2.3), any undesired effects due to the uncertainty in the prior distribution can be controlled via parameter α , which can be considered as an upper bound on the Bayes risk [42]. Specifically, as the amount of uncertainty in the prior information increases, a smaller (more restrictive) value of α is employed. In that way, the restricted Bayes formulation provides a generalization of the Bayesian and the minimax approaches [41]. In the Bayesian framework, the prior distribution of the parameter is perfectly known, whereas it is completely unknown in the minimax framework. On the other hand, the restricted Bayesian framework considers some amount of uncertainty in the prior distribution and converges to the Bayesian and minimax formulations as special cases depending on the value of α in (2.3) [41], [42]. Therefore, the study of noise enhanced hypothesis-testing in this chapter covers the previous works on noise enhanced hypothesis-testing according to the Bayesian and minimax criteria as special cases [23], [25], [29].

Two main motivations for studying the effects of additive noise on the detector performance are as follows. First, optimal detectors according to the restricted Bayes criterion are difficult to obtain, or require intense computations

[42]. Therefore, in some cases, a suboptimal detector with additive noise can provide acceptable performance with low computational complexity. Second, it is of theoretical interest to investigate the improvements that can be achieved via additive noise [29].

In order to provide an explicit formulation of the optimization problem in (2.3), which indicates the dependence of $R_{\theta}^{\mathbf{y}}(\phi)$ on the p.d.f. of the additive noise explicitly, $R_{\theta}^{\mathbf{y}}(\phi)$ in (2.4) is manipulated as follows:¹

$$R_{\theta}^{\mathbf{y}}(\phi) = \int_{\Gamma} \int_{\mathbb{R}^K} C[\phi(\mathbf{y}), \theta] p_{\theta}^{\mathbf{x}}(\mathbf{y} - \mathbf{n}) p_{\mathbf{N}}(\mathbf{n}) d\mathbf{n} d\mathbf{y} \quad (2.5)$$

$$= \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) \left[\int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^{\mathbf{x}}(\mathbf{y} - \mathbf{n}) d\mathbf{y} \right] d\mathbf{n} \quad (2.6)$$

$$= \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) F_{\theta}(\mathbf{n}) d\mathbf{n} \quad (2.7)$$

$$= E\{F_{\theta}(\mathbf{N})\} \quad (2.8)$$

where

$$F_{\theta}(\mathbf{n}) \triangleq \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^{\mathbf{x}}(\mathbf{y} - \mathbf{n}) d\mathbf{y} . \quad (2.9)$$

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

From (2.8), the optimization problem in (2.3) can be formulated as follows:

$$\begin{aligned} & \min_{p_{\mathbf{N}}(\cdot)} \int_{\Lambda} E\{F_{\theta}(\mathbf{N})\} w(\theta) d\theta , \\ & \text{subject to } \max_{\theta \in \Lambda} E\{F_{\theta}(\mathbf{N})\} \leq \alpha . \end{aligned} \quad (2.10)$$

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

$$F(\mathbf{n}) \triangleq \int_{\Lambda} F_{\theta}(\mathbf{n}) w(\theta) d\theta , \quad (2.11)$$

¹Note that the independence of \mathbf{X} and \mathbf{N} are used to obtain (2.5) from (2.4).

the optimization problem in (2.10) can be reformulated in the following simple form:

$$\begin{aligned} & \min_{p_{\mathbf{N}}(\cdot)} \mathbb{E}\{F(\mathbf{N})\} , \\ & \text{subject to } \max_{\theta \in \Lambda} \mathbb{E}\{F_{\theta}(\mathbf{N})\} \leq \alpha . \end{aligned} \quad (2.12)$$

From (2.9) and (2.11), it is noted that $F(\mathbf{0}) = r^{\mathbf{x}}(\phi)$. Namely, $F(\mathbf{0})$ is equal to the Bayes risk for the original observation \mathbf{x} ; that is, the Bayes risk in the absence of additive noise.

2.1.2 Improvability and Nonimprovability Conditions

In general, it is quite complex to obtain a solution of the optimization problem in (2.12) as it requires a search over all possible noise p.d.f.s. Therefore, it is useful to determine, without solving the optimization problem, whether additive noise can improve the performance of the original system. In the restricted Bayesian framework, a detector is called *improvable*, if there exists a noise p.d.f. such that $\mathbb{E}\{F(\mathbf{N})\} < r^{\mathbf{x}}(\phi) = F(\mathbf{0})$ and $\max_{\theta \in \Lambda} R_{\theta}^y(\phi) = \max_{\theta \in \Lambda} \mathbb{E}\{F_{\theta}(\mathbf{N})\} \leq \alpha$ (cf. (2.12)). Otherwise, the detector is called *nonimprovable*.

First, the following nonimprovability condition is obtained based on the properties of F_{θ} in (2.9) and F in (2.11).

Theorem 1: *Assume that there exists $\theta^* \in \Lambda$ such that $F_{\theta^*}(\mathbf{n}) \leq \alpha$ implies $F(\mathbf{n}) \geq F(\mathbf{0})$ for all $\mathbf{n} \in \mathcal{S}_{\mathbf{n}}$, where $\mathcal{S}_{\mathbf{n}}$ is a convex set² consisting of all possible values of additive noise \mathbf{n} . If $F_{\theta^*}(\mathbf{n})$ and $F(\mathbf{n})$ are convex functions over $\mathcal{S}_{\mathbf{n}}$, then the detector is nonimprovable.*

Proof: The proof employs an approach that is similar to the proof of Proposition 1 in [26]. Due to the convexity of $F_{\theta^*}(\cdot)$, the conditional risk in (2.8) can

² $\mathcal{S}_{\mathbf{n}}$ can be modeled as convex because convex combination of individual noise components can be obtained via randomization [80].

be bounded, via Jensen's inequality, as

$$R_{\theta^*}^{\mathbf{y}}(\phi) = \mathbb{E}\{F_{\theta^*}(\mathbf{N})\} \geq F_{\theta^*}(\mathbb{E}\{\mathbf{N}\}) . \quad (2.13)$$

As $R_{\theta^*}^{\mathbf{y}}(\phi) \leq \alpha$ is a necessary condition for improvability, (2.13) implies that $F_{\theta^*}(\mathbb{E}\{\mathbf{N}\}) \leq \alpha$ must be satisfied. Since $\mathbb{E}\{\mathbf{N}\} \in \mathcal{S}_n$, $F_{\theta^*}(\mathbb{E}\{\mathbf{N}\}) \leq \alpha$ means $F(\mathbb{E}\{\mathbf{N}\}) \geq F(\mathbf{0})$ due to the assumption in the proposition. Hence,

$$r^{\mathbf{y}}(\phi) = \mathbb{E}\{F(\mathbf{N})\} \geq F(\mathbb{E}\{\mathbf{N}\}) \geq F(\mathbf{0}) , \quad (2.14)$$

where the first inequality results from the convexity of F . Then, from (2.13) and (2.14), it is concluded that $R_{\theta^*}^{\mathbf{y}}(\phi) \leq \alpha$ implies $r^{\mathbf{y}}(\phi) \geq F(\mathbf{0}) = r^{\mathbf{x}}(\phi)$. Therefore, the detector is nonimprovable. \square

The conditions in Theorem 1 can be used to determine when the detector performance cannot be improved via additive noise, which prevents unnecessary efforts for trying to solve the optimization problem in (2.12). However, it should also be noted that Theorem 1 provides only sufficient conditions; hence, the detector can still be nonimprovable although the conditions in the theorem are not satisfied.

In order to provide an example application of Theorem 1, consider a Gaussian location testing problem [40], in which the observation has a Gaussian p.d.f. with mean $\theta\mu$ and variance σ^2 , denoted by $\mathcal{N}(\theta\mu, \sigma^2)$, where μ and σ are known values. Hypotheses \mathcal{H}_0 and \mathcal{H}_1 correspond to $\theta = 0$ and $\theta = 1$, respectively (that is, $\Lambda_0 = \{0\}$ and $\Lambda_1 = \{1\}$). In addition, consider a decision rule that selects \mathcal{H}_1 if $y \geq 0.5\mu$ and \mathcal{H}_0 otherwise. Let $\mathcal{S}_n = (-0.5\mu, 0.5\mu)$ represent the set of additive noise values for possible performance improvement. For uniform cost assignment (UCA) [40], (2.9) can be used to obtain $F_0(n)$ as follows:

$$F_0(n) = \int_{-\infty}^{\infty} C[\phi(y), 0] p_0^X(y - n) dy \quad (2.15)$$

$$= \int_{-\infty}^{\infty} \phi(y) p_0^X(y - n) dy \quad (2.16)$$

$$= \int_{0.5\mu}^{\infty} \frac{e^{-\frac{(y-n)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dy = Q\left(\frac{0.5\mu - n}{\sigma}\right) , \quad (2.17)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$ denotes the Q -function, and $C[i, j] = 1$ for $i \neq j$ and $C[i, j] = 0$ for $i = j$ are used in (2.15) due to the UCA. Similarly, $F_1(n)$ can be obtained as $F_1(n) = Q\left(\frac{0.5\mu+n}{\sigma}\right)$. For equal priors, $F(n)$ in (2.11) is obtained as $F(n) = 0.5(F_0(n) + F_1(n))$; that is,

$$F(n) = 0.5 Q\left(\frac{0.5\mu - n}{\sigma}\right) + 0.5 Q\left(\frac{0.5\mu + n}{\sigma}\right). \quad (2.18)$$

Let α be set to $Q(0.5\mu/\sigma)$, which determines the upper bound on the conditional risks. Regarding the assumption in Theorem 1, it can be shown for $\theta^* = 0$ that $F_{\theta^*}(n) \leq \alpha$ implies $F(n) \geq F(0) = Q(0.5\mu/\sigma)$ for all $n \in \mathcal{S}_n$. This follows from the facts that $F_0(n) \leq \alpha = Q(0.5\mu/\sigma)$ requires that $n \in (-0.5\mu, 0]$ and that $F(n)$ in (2.18) satisfies $F(n) \geq Q(0.5\mu/\sigma) = \alpha$ for $n \in (-0.5\mu, 0]$ due to the convexity of $Q(x/\sigma)$ for $x > 0$. In addition, it can be shown that both $F_0(n)$ and $F_1(n)$ are convex functions over \mathcal{S}_n , which implies that $F(n)$ is also convex over \mathcal{S}_n . Then, Theorem 1 implies that the detector is nonimprovable for this example. Therefore, there is no need to tackle the optimization problem in (2.12) in this case, since $p_N^{\text{opt}}(n) = \delta(n)$ is concluded directly from the theorem.

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

$$f_\theta^{(1)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{i=1}^K z_i \frac{\partial F_\theta(\mathbf{x})}{\partial x_i}, \quad (2.19)$$

$$f^{(1)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{i=1}^K z_i \frac{\partial F(\mathbf{x})}{\partial x_i}, \quad (2.20)$$

$$f_\theta^{(2)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^K \sum_{i=1}^K z_l z_i \frac{\partial^2 F_\theta(\mathbf{x})}{\partial x_l \partial x_i}, \quad (2.21)$$

$$f^{(2)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^K \sum_{i=1}^K z_l z_i \frac{\partial^2 F(\mathbf{x})}{\partial x_l \partial x_i}, \quad (2.22)$$

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

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

Then, the detector is improvable

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

Proof: Please see Appendix 2.5.1.

In order to comprehend the conditions in Theorem 2, it is first noted from (2.9) that $F_\theta(\mathbf{0})$ represents the conditional risk given θ in the absence of additive noise, $R_\theta^{\mathbf{x}}(\phi)$. Therefore, θ^* in the theorem corresponds to the value of θ for which the original conditional risk $R_\theta^{\mathbf{x}}(\phi)$ is maximum and that maximum value is assumed to be equal to the upper limit α . In other words, it is assumed that, in the absence of additive noise, the original detector already achieves the upper limit on the conditional risks for the modified observations specified in (2.3). Then, the results in the theorem imply that, under the stated conditions, it is possible to obtain a noise p.d.f. with multiple mass points around $\mathbf{n} = \mathbf{0}$, which can reduce the Bayes risk under the constraint on the conditional risks.

In order to present alternative improvability conditions to those in Theorem 2, we extend the conditions that are developed for simple binary hypothesis-testing problems in the Neyman-Pearson framework in [12] to our problem in (2.12). To that aim, we first define a new function $H(t)$ as

$$H(t) \triangleq \inf \left\{ F(\mathbf{n}) \mid \max_{\theta \in \Lambda} F_\theta(\mathbf{n}) = t, \mathbf{n} \in \mathbb{R}^K \right\}, \quad (2.23)$$

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

From (2.23), it is observed that if there exists $t_0 \leq \alpha$ such that $H(t_0) < F(\mathbf{0})$, then the system is improvable, because under such a condition there exists a noise component \mathbf{n}_0 such that $F(\mathbf{n}_0) < F(\mathbf{0})$ and $\max_{\theta \in \Lambda} F_\theta(\mathbf{n}_0) \leq \alpha$, meaning that the detector performance can be improved by adding a constant \mathbf{n}_0 to the observation. However, improvability of a detector via constant noise is not very common in practice. Therefore, the following improvability condition is obtained for more practical scenarios.

Theorem 3: *Let the maximum value of the conditional risks in the absence of additive noise be defined as $\tilde{\alpha} \triangleq \max_{\theta \in \Lambda} R_\theta^x(\phi)$ and $\tilde{\alpha} \leq \alpha$. If $H(t)$ in (2.23) is second-order continuously differentiable around $t = \tilde{\alpha}$ and satisfies $H''(\tilde{\alpha}) < 0$, then the detector is improvable.*

Proof: Please see Appendix 2.5.2.

Similar to Theorem 2, Theorem 3 provides sufficient conditions that guarantee the improvability of a detector according to the restricted Bayes criterion. Note that $H(t)$ in Theorem 3 is always a single-variable function irrespective of the dimension of the observation vector, which facilitates simple evaluation of the conditions in the theorem. However, the main challenge can be to obtain an expression for $H(t)$ in (2.23) in certain scenarios. On the other hand, Theorem 2 deals with $F_\theta(\cdot)$ and $F(\cdot)$ directly, without defining an auxiliary function like $H(t)$. Therefore, implementation of Theorem 2 can be more efficient in some cases. However, the functions in Theorem 2 are always K -dimensional, which can make the evaluation of its conditions more complicated than that in Theorem 3 in some other cases. In Section 2.3, comparisons of the improvability results based on direct evaluations of $F_\theta(\cdot)$ and $F(\cdot)$, and those based on $H(t)$ are provided.

2.1.3 On the Optimal Additive Noise

In general, the optimization problem in (2.12) is a non-convex problem and has very high computational complexity since the optimization needs to be performed over functions. In Section 2.2, it is shown that (2.12) simplifies significantly in the case of simple hypothesis-testing problems. However, in the composite case, the solution is quite difficult to obtain in general. Therefore, a p.d.f. approximation technique [50] can be employed in this section in order to obtain an approximate solution of the problem.

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

$$p_{\mathbf{N}}(\mathbf{n}) = \sum_{i=1}^L \nu_i \psi_i(\mathbf{n} - \mathbf{n}_i) , \quad (2.24)$$

where $\nu_i \geq 0$, $\sum_{i=1}^L \nu_i = 1$, and $\psi_i(\cdot)$ is a window function with $\psi_i(\mathbf{x}) \geq 0 \forall \mathbf{x}$ and $\int \psi_i(\mathbf{x}) d\mathbf{x} = 1$, for $i = 1, \dots, L$. In addition, let ς_i denote a scaling parameter for the i th window function $\psi_i(\cdot)$, which controls the “width” of the window function. The p.d.f. approximation technique in (2.24) is referred to as *Parzen window density estimation*, which has the property of mean-square convergence to the true p.d.f. under certain conditions [81]. From (2.24), the optimization problem in (2.12) can be expressed as³

$$\begin{aligned} & \min_{\{\nu_i, \mathbf{n}_i, \varsigma_i\}_{i=1}^L} \sum_{i=1}^L \nu_i f_{\mathbf{n}_i}(\varsigma_i) , \\ & \text{subject to } \max_{\theta \in \Lambda} \sum_{i=1}^L \nu_i f_{\theta, \mathbf{n}_i}(\varsigma_i) \leq \alpha , \end{aligned} \quad (2.25)$$

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

In (2.25), the optimization is performed over all the parameters of the window functions in (2.24). Therefore, the performance of the approximation technique is determined mainly by the the number of window functions, L . As L increases,

³As in [12], it is possible to perform the optimization over single-variable functions by considering mapping of the noise \mathbf{n} via $F(\mathbf{n})$ or $F_{\theta}(\mathbf{n})$.

the approximate solution can get closer to the optimal solution for the additive noise p.d.f. Therefore, in general, an improved detector performance can be expected for larger values of L .

Although (2.25) is significantly simpler than (2.12), it is still not a convex optimization problem in general. Therefore, global optimization techniques, such as particle-swarm optimization (PSO) [51], [53], [54], genetic algorithms and differential evolution [82], can be used to calculate the optimal solution [29], [50]. In Section 2.3, the PSO algorithm is used to obtain the optimal noise p.d.f.s for the numerical examples.

Although the calculation of the optimal noise p.d.f. requires significant effort as discussed above, some of its properties can be obtained without solving the optimization problem in (2.12). To that aim, let F_{\min} represent the minimum value of $H(t)$ in (2.23); that is, $F_{\min} = \min_t H(t)$. In addition, suppose that this minimum is attained at $t = t_m$.⁴ Then, one immediate observation is that if t_m is less than or equal to the conditional risk limit α , then the noise component \mathbf{n}_m that results in $\max_{\theta \in \Lambda} F_{\theta}(\mathbf{n}_m) = t_m$ is the optimal noise component; that is, the optimal noise is a constant in that scenario, $p_{\mathbf{N}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{n}_m)$. On the other hand, if $t_m > \alpha$, then it can be shown that the optimal solution of (2.12) satisfies $\max_{\theta \in \Lambda} R_{\theta}^y(\phi) = \alpha$ (Appendix 2.5.3).

2.2 Noise Enhanced Simple Hypothesis-Testing

In this section, noise enhanced detection is studied in the restricted Bayesian framework for *simple* hypothesis-testing problems. In simple hypothesis-testing problems, each hypothesis corresponds to a single probability distribution [40].

⁴If there are multiple t values that result in the minimum value F_{\min} , then the minimum of those values can be considered.

In other words, the generic composite hypothesis-testing problem in (2.1) reduces to a simple hypothesis-testing problem if each Λ_i consists of a single element.

Since the simple hypothesis-testing problem is a special case of the composite one, the results in Section 2.1 are also valid for this section. However, by using the special structure of simple hypotheses, we obtain additional results in this section that are not valid for composite hypothesis-testing problems. It should be noted that both composite and simple hypothesis-testing problems are used to model various practical detection examples [40], [83]; hence, specific results can be useful in different applications.

2.2.1 Problem Formulation

The problem can be formulated as in Section 2.1.1 by defining $\Lambda_i = \{\theta_i\}$ for $i = 0, 1, \dots, M - 1$ in (2.1). In addition, instead of the prior p.d.f. $w(\theta)$, the prior probabilities of the hypotheses can be defined by $\pi_0, \pi_1, \dots, \pi_{M-1}$ with $\sum_{i=0}^{M-1} \pi_i = 1$. Then, the optimal additive noise problem in (2.3) becomes

$$\begin{aligned} \min_{p_{\mathbf{N}}(\cdot)} \quad & \sum_{i=0}^{M-1} \pi_i R_i^{\mathbf{y}}(\phi) , \\ \text{subject to} \quad & \max_{i \in \{0, 1, \dots, M-1\}} R_i^{\mathbf{y}}(\phi) \leq \alpha , \end{aligned} \quad (2.26)$$

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

$$R_i^{\mathbf{y}}(\phi) = \sum_{j=0}^{M-1} C_{ji} P_i^{\mathbf{y}}(\Gamma_j) , \quad (2.27)$$

with $P_i^{\mathbf{y}}(\Gamma_j)$ denoting the probability that $\mathbf{y} \in \Gamma_j$ when \mathcal{H}_i is the true hypothesis, and C_{ji} defining the cost of deciding \mathcal{H}_j when \mathcal{H}_i is true. As in Section 2.1.1, the constraint α sets an upper limit on the conditional risks, and its value is determined depending on the amount of uncertainty in the prior probabilities.

In order to investigate the optimal solution of (2.26), an alternative expression for $R_i^Y(\phi)$ is obtained first. Since the additive noise \mathbf{n} is independent of the observation \mathbf{x} , $P_i^Y(\Gamma_j)$ becomes

$$P_i^Y(\Gamma_j) = \int_{\Gamma_j} p_i^Y(\mathbf{y}) d\mathbf{y} = \int_{\Gamma_j} \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) p_i^{\mathbf{X}}(\mathbf{y} - \mathbf{n}) d\mathbf{n} d\mathbf{y} , \quad (2.28)$$

where $p_i^{\mathbf{X}}(\cdot)$ and $p_i^Y(\cdot)$ represent the p.d.f.s of the original observation and the noise modified observation, respectively, when hypothesis \mathcal{H}_i is true. Then, (2.27) can be expressed, from (2.28), as

$$\begin{aligned} R_i^Y(\phi) &= \sum_{j=0}^{M-1} C_{ji} \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) \int_{\Gamma_j} p_i^{\mathbf{X}}(\mathbf{y} - \mathbf{n}) d\mathbf{y} d\mathbf{n} \\ &= \sum_{j=0}^{M-1} C_{ji} E\{F_{ij}(\mathbf{N})\} = E\{F_i(\mathbf{N})\} , \end{aligned} \quad (2.29)$$

with

$$F_{ij}(\mathbf{n}) \triangleq \int_{\Gamma_j} p_i^{\mathbf{X}}(\mathbf{y} - \mathbf{n}) d\mathbf{y} , \quad (2.30)$$

$$F_i(\mathbf{n}) \triangleq \sum_{j=0}^{M-1} C_{ji} F_{ij}(\mathbf{n}) . \quad (2.31)$$

Based on the relation in (2.29), the optimization problem in (2.26) can be reformulated as

$$\begin{aligned} \min_{P_{\mathbf{N}}(\cdot)} \sum_{i=0}^{M-1} \pi_i E\{F_i(\mathbf{N})\} , \\ \text{subject to } \max_{i \in \{0,1,\dots,M-1\}} E\{F_i(\mathbf{N})\} \leq \alpha . \end{aligned} \quad (2.32)$$

If a new auxiliary function is defined as $F(\mathbf{n}) \triangleq \sum_{i=0}^{M-1} \pi_i F_i(\mathbf{n})$, (2.32) becomes

$$\begin{aligned} \min_{P_{\mathbf{N}}(\cdot)} E\{F(\mathbf{N})\} , \\ \text{subject to } \max_{i \in \{0,1,\dots,M-1\}} E\{F_i(\mathbf{N})\} \leq \alpha . \end{aligned} \quad (2.33)$$

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

It should be noted from the definitions in (2.30) and (2.31) that $F_i(\mathbf{0})$ corresponds to the conditional risk given \mathcal{H}_i for the original observation \mathbf{x} , $R_i^{\mathbf{x}}(\phi)$. Therefore, $F(\mathbf{0})$ defines the original Bayes risk, $r^{\mathbf{x}}(\phi)$.

2.2.2 Optimal Additive Noise

The optimization problem in (2.33) seems quite difficult to solve in general as it requires a search over all possible noise p.d.f.s. However, in the following, it is shown that an optimal additive noise p.d.f. can be represented by a discrete probability distribution with at most M mass points in most practical cases. To that aim, suppose that all possible additive noise values satisfy $\mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b}$ for any finite \mathbf{a} and \mathbf{b} ; that is, $n_j \in [a_j, b_j]$ for $j = 1, \dots, K$, which is a reasonable assumption since additive noise cannot have infinitely large amplitudes in practice. Then, the following theorem states the discrete nature of the optimal additive noise.

Theorem 4: *If $F_i(\cdot)$ in (2.32) are continuous functions, then the p.d.f. of an optimal additive noise can be expressed as $p_{\mathbf{N}}(\mathbf{n}) = \sum_{l=1}^M \lambda_l \delta(\mathbf{n} - \mathbf{n}_l)$, where $\sum_{l=1}^M \lambda_l = 1$ and $\lambda_l \geq 0$ for $l = 1, 2, \dots, M$.*

Proof: The proof employs a similar approach to those used for the related results in [12], [29] and [50]. First, the following set is defined:

$$U = \{(u_0, u_1, \dots, u_{M-1}) : u_i = F_i(\mathbf{n}), i = 0, 1, \dots, M-1, \text{ for } \mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b}\} . \quad (2.34)$$

In addition, V is defined as the convex hull of U [84]. Since $F_i(\cdot)$ are continuous functions, U is a bounded and closed subset of \mathbb{R}^M . Hence, U is a compact set. Therefore, its convex hull V is a closed subset of \mathbb{R}^M [29]. Next, set W is defined

as

$$W = \left\{ (w_0, w_1, \dots, w_{M-1}) : w_i = \mathbb{E}\{F_i(\mathbf{n})\}, i = 0, 1, \dots, M-1, \right. \\ \left. \forall p_{\mathbf{N}}(\mathbf{n}), \mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b} \right\}, \quad (2.35)$$

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

As V is the convex hull of U , each element of V can be expressed as $\mathbf{v} = \sum_{l=1}^{N_L} \lambda_l (F_0(\mathbf{n}_l), F_1(\mathbf{n}_l), \dots, F_{M-1}(\mathbf{n}_l))$, where $\sum_{l=1}^{N_L} \lambda_l = 1$, and $\lambda_l \geq 0 \forall l$. On the other hand, each \mathbf{v} is also an element of W as it can be obtained for $p_{\mathbf{N}}(\mathbf{n}) = \sum_{l=1}^{N_L} \lambda_l \delta(\mathbf{n} - \mathbf{n}_l)$. Hence, $V \subseteq W$ [29]. In addition, since for any vector random variable Θ taking values in set Ω , its expected value, $\mathbb{E}\{\Theta\}$, is in the convex hull of Ω [85], (2.34) and (2.35) implies that W is in the convex hull V of U ; that is, $V \supseteq W$. Since $V \subseteq W$ and $V \supseteq W$, it means that $W = V$ [29]. Therefore, according to Carathéodory's theorem [86], [87], any point in V (or, W) can be expressed as the convex combination of at most $(M + 1)$ points in U as the dimension of U is smaller than or equal to M . Since the aim is to minimize the average of the conditional risks, the optimal solution corresponds to the boundary of W . As W (or, V) is a closed set as mentioned at the beginning of the proof, it contains its own boundary [29]. Since any point at the boundary of W can be expressed as the convex combination of at most M elements in U [86], an optimal noise p.d.f. can be represented by a discrete random variable with M mass points as stated in the theorem. \square

From Theorem 4, the optimization problem in (2.33) can be simplified as

$$\begin{aligned} & \min_{\{\lambda_l, \mathbf{n}_l\}_{l=1}^M} \sum_{l=1}^M \lambda_l F(\mathbf{n}_l), \\ & \text{subject to } \max_{i \in \{0, 1, \dots, M-1\}} \sum_{l=1}^M \lambda_l F_i(\mathbf{n}_l) \leq \alpha, \\ & \sum_{l=1}^M \lambda_l = 1, \quad \lambda_l \geq 0, \quad l = 1, \dots, M. \end{aligned} \quad (2.36)$$

The optimization in (2.36) is considerably simpler than that in (2.33) since the former is over a set of variables instead of functions. However, (2.36) can still be

a nonconvex optimization problem in general; hence, global optimization techniques, such as PSO [51] and differential evolution [82] may be needed.

In order to provide a convex relaxation [55] of the optimization problem in (2.36) and to obtain an approximate solution in polynomial time, one can assume that additive noise \mathbf{n} can take only finitely many known values specified by $\tilde{\mathbf{n}}_1, \dots, \tilde{\mathbf{n}}_L$ [29]. This scenario, for example, corresponds to digital systems in which the signals can take only finitely many different levels. Then, the aim becomes the determination of the weights $\tilde{\lambda}_1, \dots, \tilde{\lambda}_L$ of those possible noise values. In that case, (2.33) can be formulated as

$$\begin{aligned} & \min_{\{\tilde{\lambda}_l\}_{l=1}^L} \sum_{l=1}^L \tilde{\lambda}_l F(\tilde{\mathbf{n}}_l) , \\ & \text{subject to } \max_{i \in \{0,1,\dots,M-1\}} \sum_{l=1}^L \tilde{\lambda}_l F_i(\tilde{\mathbf{n}}_l) \leq \alpha , \\ & \sum_{l=1}^L \tilde{\lambda}_l = 1 , \quad \tilde{\lambda}_l \geq 0 , \quad l = 1, \dots, L , \end{aligned} \quad (2.37)$$

which is a linearly constrained linear programming (LCLP) problem; hence, can be solved in polynomial time [55]. It should be noted that as the optimization is performed over more noise values (as L increases), the solution gets closer to the optimal solution of (2.33).

As an alternative approach, an analytical solution similar to that in [12] can also be proposed for obtaining the optimal additive noise. First, consider the optimization problem in (2.32) for $M = 2$; i.e., the binary case. If functions $F_0(\mathbf{n})$ and $F_1(\mathbf{n})$ are monotone, then t_0 and t_1 can be defined as $t_0 = F_0(\mathbf{n})$ and $t_1 = F_1(\mathbf{n})$. Otherwise, let t_0 and t_1 be defined as follows:

$$\begin{aligned} t_0(t) &\triangleq \inf \{ F_0(\mathbf{n}) \mid F_1(\mathbf{n}) = t , \mathbf{n} \in \mathbb{R}^K \} , \\ t_1(t) &\triangleq \inf \{ F_1(\mathbf{n}) \mid F_0(\mathbf{n}) = t , \mathbf{n} \in \mathbb{R}^K \} . \end{aligned} \quad (2.38)$$

In general, there can exist multiple values of $F_1(\mathbf{n})$ corresponding to a given value of $F_0(\mathbf{n})$. However, the definitions of t_0 and t_1 in (2.38) make sure that only the

best (minimum) value of $F_1(\mathbf{n})$ corresponding to a given $F_0(\mathbf{n})$ is considered, and vice versa. Therefore, t_1 can be expressed as $t_1 = g(t_0)$, where $g(t_0)$ is a monotone function of t_0 and is defined on the range of t_0 , which is denoted by $[t_{0,\min}, t_{0,\max}]$ with $t_{0,\min} = \min t_0$ and $t_{0,\max} = \max t_0$. We call the set of t_0 for which $g(t_0)$ and t_0 satisfy the constraints (cf. (2.32)) as the *feasible domain*. Then, let a new function B be defined as follows:

$$B(t_0) \triangleq \pi_0 t_0 + \pi_1 g(t_0) . \quad (2.39)$$

If $B(t_0)$ takes its global minimum value in the feasible domain, then the optimal Bayes risk is equal to that minimum value and the optimal additive noise can be represented by a constant value. For example, if $t_0^* = \arg \min_{t_0} B(t_0)$, then the optimal additive noise p.d.f. can be expressed as $p_{\mathbf{N}}(\mathbf{n}) = \delta(\mathbf{n} - \mathbf{n}_0)$, where \mathbf{n}_0 satisfies $F_0(\mathbf{n}_0) = t_0^*$.⁵ On the other hand, if $B(t_0)$ achieves its global minimum value outside the feasible domain, then an analytic solution for the optimal additive noise p.d.f. can be obtained as explained in the following. At the end of Section 2.1.3, it was stated that the maximum value of the optimal conditional risks must be equal to the constraint level α for the case considered here. This implies that the optimal (t_0, t_1) pair is equal to one of the following: (α, β) or (γ, α) , where β and γ are such that $g(\alpha) = \beta$ and $g(\gamma) = \alpha$. It should be noted that if $g(t_0)$ is a decreasing function and γ is larger than α , then the feasible domain is an empty set implying that there is no solution satisfying the constraint.

Since $g(t_0)$ is a monotone function and the maximum of the optimal conditional risks must be equal to α , the feasible domain must be in the form of an interval, say $[a, b]$, and the value of t_0 corresponding to the optimal solution must be equal to either a or b . In the following derivations, it is assumed that the value of t_0 corresponding to the optimal solution is b , and $B(t_0)$ takes its global minimum value for $t_0 > b$. However, it should be noted that these assumptions

⁵If there are multiple such \mathbf{n}_0 's, then the one that minimizes $F_1(\mathbf{n}_0)$ should be chosen.

do not reduce the generality of the results. In other words, the derivations based on the other possible assumptions yield the same result.

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

$$Z(t_0, k) \triangleq B(t_0) + kt_0, \quad (2.40)$$

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

$$\begin{aligned} v_1(k) &\triangleq \min_{t_0 \in \mathcal{I}_1} Z(t_0, k) = Z(t_{01}(k), k), \\ v_2(k) &\triangleq \min_{t_0 \in \mathcal{I}_2} Z(t_0, k) = Z(t_{02}(k), k), \end{aligned} \quad (2.41)$$

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

From (2.40) and (2.41), it is obtained for $k = 0$ that $v_2(0) = \min B(t_0) < v_1(0) = B(t_{01}(k))$. On the other hand, as $k \rightarrow \infty$, $v_1(k) = B(t_{0,\min}) + kt_{0,\min} < v_2(k) = B(b) + kb$. Therefore, there must exist a $k = k_0$, where $0 < k_0 < \infty$, such that

$$v = v_1(k_0) = Z(t_{01}(k_0), k_0) = v_2(k_0) = Z(t_{02}(k_0), k_0). \quad (2.42)$$

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

$$p_{\mathbf{n},t_0}(t_0) = \lambda_1 \delta(t_0 - t_{01}(k_0)) + \lambda_2 \delta(t_0 - t_{02}(k_0)) + \mathcal{I}_A(t_0) p_{\mathbf{n},t_0}(t_0), \quad (2.43)$$

where $\mathcal{I}_A(t_0)$ is an indicator function such that $\mathcal{I}_A(t_0) = 1$ if $t_0 \in A$, $\mathcal{I}_A(t_0) = 0$ otherwise [12]. By definition, $\lambda_1 + \lambda_2 + \int_A p_{\mathbf{n},t_0}(t_0) dt_0 = 1$ should be satisfied. In

addition, the expectation of Z in (2.40) over t_0 can be bounded as follows:

$$\begin{aligned} \mathbb{E}\{Z(t_0, k_0)\} &= \lambda_1 v + \lambda_2 v + \int_A Z(t_0, k_0) p_{\mathbf{n}, t_0}(t_0) dt_0, \\ &= v + \int_A [Z(t_0, k_0) - v] p_{\mathbf{n}, t_0}(t_0) dt_0, \\ &\geq v, \end{aligned} \tag{2.44}$$

where the first expression is obtained from (2.42) and (2.43), and the final inequality is obtained from the fact that $Z(t_0, k_0) \geq v$ for $t_0 \in A$ (cf. (2.41) and (2.42)). This lower bound is achieved for $p_{\mathbf{n}, t_0}(t_0) = \lambda_1 \delta(t_0 - t_{01}(k_0)) + \lambda_2 \delta(t_0 - t_{02}(k_0))$, with $\lambda_1 + \lambda_2 = 1$. Hence, $p_{\mathbf{n}, t_0}(t_0) = 0$ for $t_0 \in A$.

From (2.39) and (2.40), the Bayes risk $r^{\mathcal{Y}}(\phi)$ can be expressed as $r^{\mathcal{Y}}(\phi) = \mathbb{E}\{B(t_0)\} = \mathbb{E}\{Z(t_0, k_0)\} - k_0 \mathbb{E}\{t_0\}$. Since $t_{01}(k_0) < b$ and $t_{02}(k_0) \geq b$, one can achieve $\mathbb{E}\{t_0\} = b$ by using a noise component with p.d.f. $p_{\mathbf{n}, t_0}(t_0) = \lambda_1 \delta(t_0 - t_{01}(k_0)) + \lambda_2 \delta(t_0 - t_{02}(k_0))$, where $\lambda_1 + \lambda_2 = 1$ with appropriate values for λ_1 and λ_2 . Thus, the optimal additive noise p.d.f. is $p_{\mathbf{n}, t_0}(t_0) = \lambda_1 \delta(t_0 - t_{01}(k_0)) + \lambda_2 \delta(t_0 - t_{02}(k_0))$, where $\lambda_1 + \lambda_2 = 1$ and $\lambda_1 t_{01}(k_0) + \lambda_2 t_{02}(k_0) = b$, and the optimal Bayes risk is given by $r_{\text{opt}}^{\mathcal{Y}}(\phi) = \mathbb{E}\{B(t_0)\} = v - k_0 b$.

Since $Z(t_0, k_0)$ has (local) minimum values at $t_0 = t_{01}(k_0)$ and $t_0 = t_{02}(k_0)$, if $B(t_0)$ is continuously differentiable, then $\partial Z(t_{01}(k_0), k_0) / \partial t_0 = \partial Z(t_{02}(k_0), k_0) / \partial t_0 = 0$. Then, (2.40) implies the following equalities:

$$\frac{dB(t_{01}(k_0))}{dt_0} = \frac{dB(t_{02}(k_0))}{dt_0} = -k_0. \tag{2.45}$$

From (2.42), we also have the following relation:

$$\frac{B(t_{01}(k_0)) - B(t_{02}(k_0))}{t_{01}(k_0) - t_{02}(k_0)} = -k_0. \tag{2.46}$$

Therefore, (2.45) and (2.46) can be used to obtain the following result:

$$\frac{B(t_{01}(k_0)) - B(t_{02}(k_0))}{t_{01}(k_0) - t_{02}(k_0)} = \frac{dB(t_{01}(k_0))}{dt_0} = \frac{dB(t_{02}(k_0))}{dt_0}. \tag{2.47}$$

From the equalities in (2.47), one can find $t_{01}(k_0)$ and $t_{02}(k_0)$, and the corresponding mass points \mathbf{n}_1 and \mathbf{n}_2 that satisfy $t_{01}(k_0) = F_0(\mathbf{n}_1)$ and $t_{02}(k_0) = F_0(\mathbf{n}_2)$.⁶

⁶If there are multiple such \mathbf{n}_1 's (\mathbf{n}_2 's), then the one that minimizes $F_1(\mathbf{n}_1)$ ($F_1(\mathbf{n}_2)$) should be chosen.

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

The analytic approach described above for the binary case can also be extended to the M -ary case for $M > 2$. However, in that case, only the mass points, $\mathbf{n}_1, \dots, \mathbf{n}_M$, can be found analytically. The weights, $\lambda_1, \dots, \lambda_M$, should be found via a numerical approach. Such a semi-analytical solution can still provide significant computational complexity reduction in some cases since the weights, which are not determined analytically, are easier to search for than the mass points, as the weights are always scalar whereas the mass points can also be multidimensional. The analytical approach to obtaining the mass points in the M -ary case is a simple extension of that in the binary case. Mainly, a function t_{M-1} should be defined as $t_{M-1} \triangleq g(t_0, \dots, t_{M-2}) \triangleq \inf \{F_{M-1}(\mathbf{n}) \mid F_0(\mathbf{n}) = t_0, \dots, F_{M-2}(\mathbf{n}) = t_{M-2}, \mathbf{n} \in \mathbb{R}^K\}$, function B in (2.39) should be generalized as

$B(t_0, \dots, t_{M-2}) \triangleq \pi_0 t_0 + \dots + \pi_{M-1} g(t_0, \dots, t_{M-2})$, and Z should be modified as $Z(t_0, \dots, t_{M-2}, k_1, \dots, k_{M-1}) \triangleq B(t_0, \dots, t_{M-2}) + k_1 t_0 + \dots + k_{M-1} t_{M-2}$. The resulting equations provide a generalization of those in (2.47), the details of which are not presented here due to the space limitations.

2.2.3 Improvability and Nonimprovability Conditions

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

For the nonimprovability conditions, Theorem 1 in Section 2.1.2 already provides a quite explicit statement to evaluate the nonimprovability. Therefore, it is also practical for simple hypothesis-testing problems, as observed in the example after Theorem 1. In accordance with the notation in this section, Theorem 1 can be restated for simple hypothesis-testing problems as follows. *Assume that there exists $i \in \{0, 1, \dots, M - 1\}$ such that $F_i(\mathbf{n}) \leq \alpha$ implies $F(\mathbf{n}) \geq F(\mathbf{0})$ for all $\mathbf{n} \in \mathcal{S}_n$, where \mathcal{S}_n is a convex set consisting of all possible values of additive noise \mathbf{n} . If $F_i(\mathbf{n})$ and $F(\mathbf{n})$ are convex functions over \mathcal{S}_n , then the detector is nonimprovable.*

Regarding the improvability conditions, in addition to Theorem 2 and Theorem 3 in Section 2.1.2, new sufficient conditions that are specific to simple hypothesis-testing problems are provided in the following. To that aim, it is first assumed that $F_i(\mathbf{x})$ for $i = 0, 1, \dots, M - 1$ and $F(\mathbf{x})$, defined in Section 2.2.1, are second-order continuously differentiable around $\mathbf{x} = \mathbf{0}$. In addition, similar to (2.19)-(2.22), the following functions are defined.

$$f_j^{(1)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{i=1}^K z_i \frac{\partial F_j(\mathbf{x})}{\partial x_i}, \quad (2.48)$$

$$f^{(1)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{i=1}^K z_i \frac{\partial F(\mathbf{x})}{\partial x_i}, \quad (2.49)$$

$$f_j^{(2)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^K \sum_{i=1}^K z_l z_i \frac{\partial^2 F_j(\mathbf{x})}{\partial x_l \partial x_i}, \quad (2.50)$$

$$f^{(2)}(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^K \sum_{i=1}^K z_l z_i \frac{\partial^2 F(\mathbf{x})}{\partial x_l \partial x_i}, \quad (2.51)$$

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

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

$$\mathcal{S}_\alpha = \{i \in \{0, 1, \dots, M-1\} \mid F_i(\mathbf{0}) = \alpha\} , \quad (2.52)$$

$$\bar{\mathcal{S}}_\alpha = \{i \in \{0, 1, \dots, M-1\} \mid F_i(\mathbf{0}) < \alpha\} . \quad (2.53)$$

In addition, let $\mathcal{S}_\alpha \cup \bar{\mathcal{S}}_\alpha = \{0, 1, \dots, M-1\}$, meaning that $F_i(\mathbf{0}) = R_i^x(\phi) \leq \alpha$ for $i = 0, 1, \dots, M-1$. Consider the functions in (2.48)-(2.51), and define set \mathcal{F}_n ($n = 1, 2$) as the set that consists of $f^{(n)}(\mathbf{x}, \mathbf{z})$ and $f_i^{(n)}(\mathbf{x}, \mathbf{z})$ for $i \in \mathcal{S}_\alpha$; that is,

$$\mathcal{F}_n = \left\{ f^{(n)}(\mathbf{x}, \mathbf{z}), f_i^{(n)}(\mathbf{x}, \mathbf{z}) \text{ for } i \in \mathcal{S}_\alpha \right\} \quad (2.54)$$

for $n = 1, 2$. Note that \mathcal{F}_n has $|\mathcal{S}_\alpha| + 1$ elements, where $|\mathcal{S}_\alpha|$ represents the number of elements in \mathcal{S}_α . In addition, $\mathcal{F}_n(j)$ will be used to refer to the j th element of \mathcal{F}_n . It should be noted that $\mathcal{F}_n(1) = f^{(n)}(\mathbf{x}, \mathbf{z})$ and $\mathcal{F}_n(j) = f_{\mathcal{S}_\alpha(j-1)}^{(n)}(\mathbf{x}, \mathbf{z})$ for $j = 2, \dots, |\mathcal{S}_\alpha| + 1$, where $\mathcal{S}_\alpha(j-1)$ is the $(j-1)$ th element of \mathcal{S}_α . Finally, the following sets are introduced to define the set of indices j for which $\mathcal{F}_1(j)$ is zero, negative or positive:

$$\mathcal{S}_z = \{j \in \{1, 2, \dots, |\mathcal{S}_\alpha| + 1\} \mid \mathcal{F}_1(j) = 0\} , \quad (2.55)$$

$$\mathcal{S}_n = \{j \in \{1, 2, \dots, |\mathcal{S}_\alpha| + 1\} \mid \mathcal{F}_1(j) < 0\} , \quad (2.56)$$

$$\mathcal{S}_p = \{j \in \{1, 2, \dots, |\mathcal{S}_\alpha| + 1\} \mid \mathcal{F}_1(j) > 0\} . \quad (2.57)$$

Based on the definitions in (2.48)-(2.57), the following theorem provides sufficient conditions for improvability.

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

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

2. One of the following is satisfied:

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

$$\min_{j \in \mathcal{S}_n} \mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l) > \max_{j \in \mathcal{S}_p} \mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l). \quad (2.58)$$

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

$$\min_{j \in \mathcal{S}_p} \mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l) > \max_{j \in \mathcal{S}_n} \mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l). \quad (2.59)$$

Proof: Please see Appendix 2.5.4.

Theorem 5 states that whenever the two conditions in the theorem are satisfied, it can be concluded that the detection performance can be improved via additive independent noise. It should be noted that after defining the sets in (2.52)-(2.57), it is straightforward to check the conditions stated in the theorem. An example application of Theorem 5 is provided in Section 2.3, where its practicality and effectiveness are observed.

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

Corollary 1: Assume that $F(\mathbf{x})$ and $F_i(\mathbf{x}), i = 0, 1, \dots, M - 1$, are second-order continuously differentiable around $\mathbf{x} = \mathbf{0}$ and that $\max_{i \in \{0, 1, \dots, M-1\}} F_i(\mathbf{0}) < \alpha$.

Let \mathbf{f} denote the gradient of $F(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$. Then, the detector is improvable

- if $\mathbf{f} \neq \mathbf{0}$; or,

- if $F(\mathbf{x})$ is not convex around $\mathbf{x} = \mathbf{0}$.

Proof: Please see Appendix 2.5.5.

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

2.3 Numerical Results

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

$$\mathcal{H}_0 : x = v , \quad \text{versus} \quad \mathcal{H}_1 : x = A + v , \quad (2.60)$$

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

$$p_V(x) = \sum_{i=1}^{N_m} w_i \psi_i(x - \mu_i) , \quad (2.61)$$

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

$$\psi_i(x) = \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left(\frac{-x^2}{2\sigma_i^2}\right) , \quad (2.62)$$

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

described by

$$\phi(y) = \begin{cases} 1, & y \geq A/2 \\ 0, & y < A/2 \end{cases}, \quad (2.63)$$

where $y = x + n$, with n representing the additive independent noise term. The aim is to obtain the optimal p.d.f. for the additive noise based on the optimization problem in (2.26).

Under the assumption of UCA, (2.60)-(2.63) can be used to calculate $F_0(x)$ and $F_1(x)$ from (2.30) and (2.31) as

$$\begin{aligned} F_0(x) &= \sum_{i=1}^{N_m} w_i Q\left(\frac{A/2 - x - \mu_i}{\sigma_i}\right), \\ F_1(x) &= \sum_{i=1}^{N_m} w_i Q\left(\frac{A/2 + x + \mu_i}{\sigma_i}\right), \end{aligned} \quad (2.64)$$

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

The symmetric Gaussian mixture noise specified above is observed in many practical scenarios [88]-[90]. One important scenario is multiuser wireless communications, in which the desired signal is corrupted by interference from other users as well as by zero-mean Gaussian background noise [91]. In other words, the signal detection example in (2.60) with symmetric Gaussian mixture noise finds various practical applications.

Since the problem in (2.60) models a signal detection problem in the presence of noise, we consider two common scenarios in the following simulations. In the first one, it is assumed that the noise-only hypothesis \mathcal{H}_0 has a higher prior probability than the signal-plus-noise hypothesis \mathcal{H}_1 . An example of this scenario is the signal acquisition problem, in which a number of correlation outputs are compared against a threshold to determine the timing/phase of the signal [92]. In the second scenario, equal prior probabilities are assumed for the hypotheses, which can be well-suited for binary communications systems that transmit no

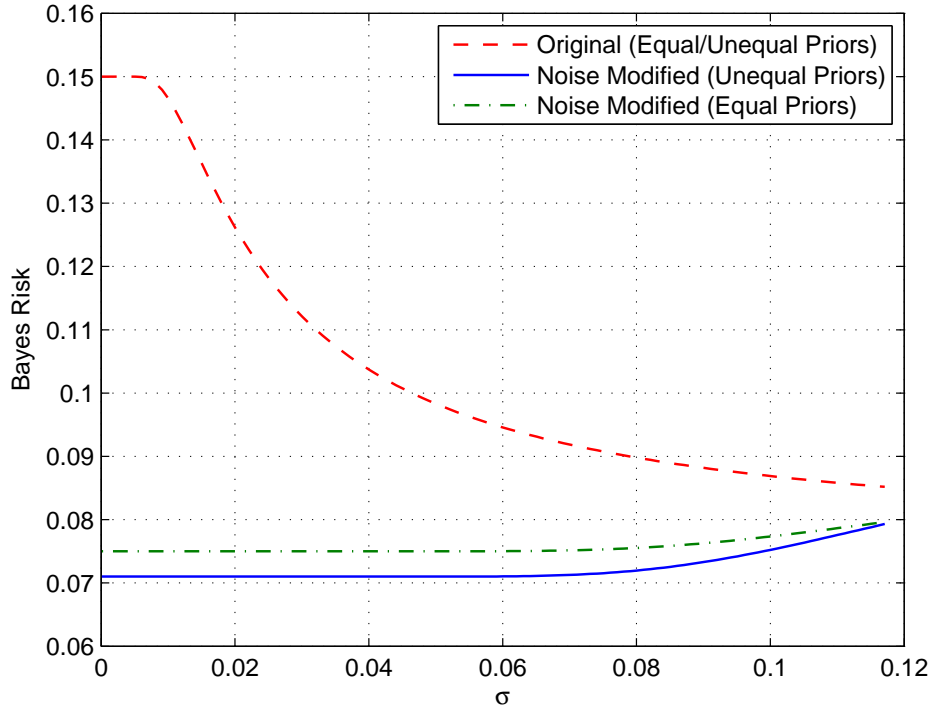


Figure 2.1: Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.08$ and $A = 1$.

signal for bit 0 and a signal for bit 1 (i.e., on-off keying) [93]. For the first scenario, it is assumed that the prior probabilities are known, with some uncertainty, to be equal to $\pi_0 = 0.9$ and $\pi_1 = 0.1$, which is called the *unequal priors* case in the following. On the other hand, $\pi_0 = \pi_1 = 0.5$ is considered for the *equal priors* case. As mentioned in Section 2.1.1, the restricted Bayes criterion mitigates the undesired effects due to the uncertainty in prior probabilities via parameter α , which sets an upper limit on the conditional risks. In the numerical results, symmetric Gaussian mixture noise with $N_m = 4$ is considered, where the mean values of the Gaussian components in the mixture noise in (2.61) are specified as $[0.033 \ 0.52 \ -0.52 \ -0.033]$ with corresponding weights of $[0.35 \ 0.15 \ 0.15 \ 0.35]$. In addition, for all the cases, the variances of the Gaussian components in the mixture noise are assumed to be the same; i.e., $\sigma_i = \sigma$ for $i = 1, \dots, N_m$ in (2.62).

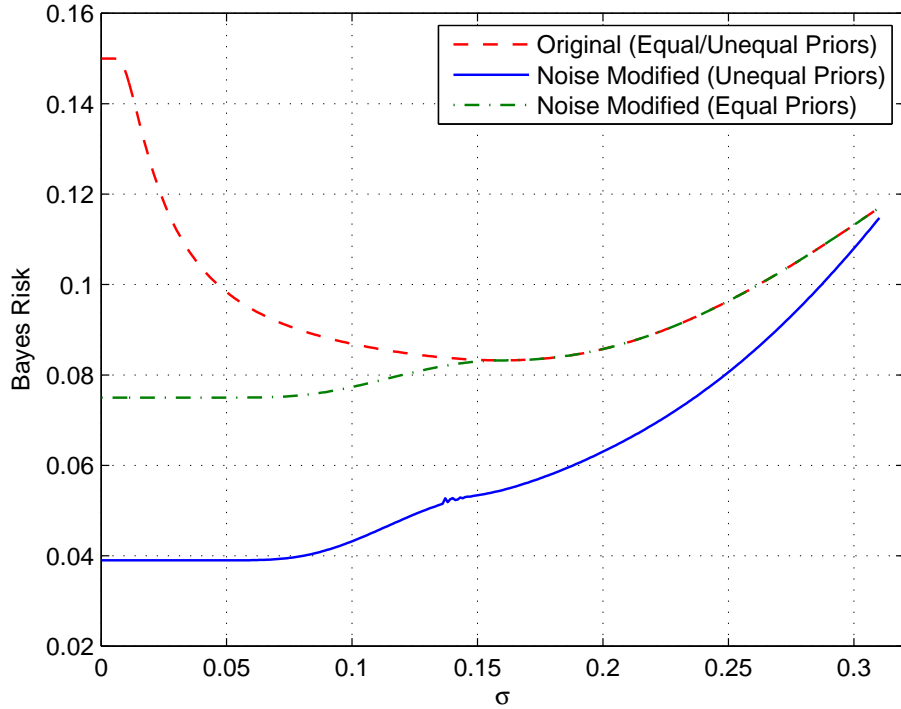


Figure 2.2: Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.12$ and $A = 1$.

For the detection problem described above, the optimal additive noise can be represented by a probability distribution with at most two mass points according to Theorem 4. Therefore, the optimal additive noise p.d.f. can be calculated as the solution of the optimization problem in (2.36) for $M = 2$. In this section, the PSO algorithm is employed to obtain the optimal solution, since it is based on simple iterations with low computational complexity and has been successfully applied to numerous problems in various fields [94]-[97] (please refer to [51]-[54] for detailed descriptions of the PSO algorithm).⁷

Figures 2.1, 2.2 and 2.3 illustrate the Bayes risks for the noise modified and the original (i.e., in the absence of additive noise) detectors for various values

⁷In the implementation of the PSO algorithm, we employ 50 particles and 1000 iterations. Also, the other parameters are set to $c_1 = c_2 = 2.05$ and $\chi = 0.72984$, and the inertia weight ω is changed from 1.2 to 0.1 linearly with the iteration number. Please refer to [51] for the details of the PSO algorithm and the definitions of the parameters.

Table 2.1: Optimal additive noise p.d.f.s for various values of σ for $\alpha = 0.08$ and $A = 1$.

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

of σ in the cases of equal and unequal priors for $\alpha = 0.08$, $\alpha = 0.12$, $\alpha = 0.4$, respectively, where $A = 1$ is used.⁸ From the figures, it is observed that as σ decreases, the improvement obtained via additive noise increases. This is mainly due to the fact that noise enhancements commonly occur when observations have multimodal p.d.f.s [12], and the multimodal structure is more pronounced for small σ 's. In addition, the figures indicate that there is always more improvement in the unequal priors case than that in the equal priors case, which is expected since there is more room for noise enhancement in the unequal priors case due to the asymmetry between the weights of the conditional risks in determining the Bayes risk. Another important point to note from the figures is that the feasible ranges of σ values are different for different values of α . In other words, for each α , the constraint on the maximum conditional risks (cf. (2.26)) cannot be satisfied after a specific value of σ . This is expected since as σ (which determines the average noise power) exceeds a certain value, it becomes impossible to keep the conditional risks below the given limit α . Therefore, Figures 2.1, 2.2 and 2.3 are plotted only up to those specific σ values. From the figures, it is observed that those maximum σ values are 0.117, 0.31 and 1.93 for $\alpha = 0.08$, $\alpha = 0.12$ and $\alpha = 0.4$, respectively.

⁸Due to the symmetry of the Gaussian mixture noise, the conditional risks in the absence of noise, $F_0(0)$ and $F_1(0)$, are equal. Therefore, the original Bayes risks are the same for both the equal and the unequal priors cases.

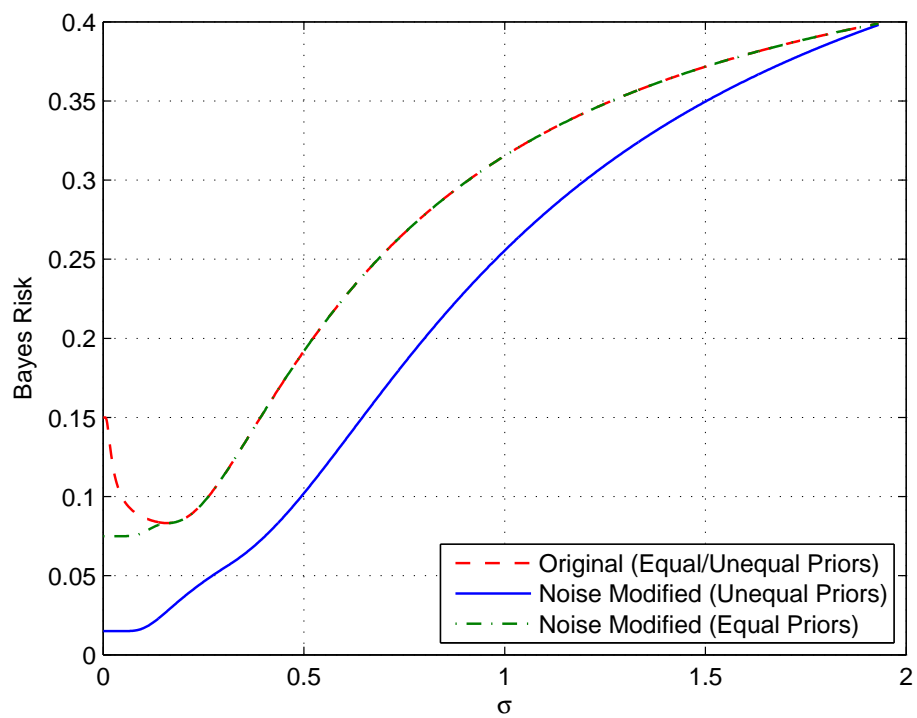


Figure 2.3: Bayes risks of original and noise modified detectors versus σ in cases of equal priors and unequal priors for $\alpha = 0.4$ and $A = 1$.

Table 2.2: Optimal additive noise p.d.f.s for various values of σ for $\alpha = 0.12$ and $A = 1$.

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

Table 2.3: Optimal additive noise p.d.f.s for various values of σ for $\alpha = 0.4$ and $A = 1$.

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

In order to investigate the results in Figures 2.1, 2.2 and 2.3 further, Tables 2.1, 2.2 and 2.3 show the optimal additive noise p.d.f.s for various values of σ in the cases of equal and unequal priors for $\alpha = 0.08$, $\alpha = 0.12$ and $\alpha = 0.4$ respectively, where $A = 1$. From Theorem 4, it is known that the optimal additive noise in this example can be represented by a discrete probability distribution with at most two mass points, which can be described as $p_N(x) = \lambda \delta(x - n_1) + (1 - \lambda) \delta(x - n_2)$. It is observed from the tables that the optimal additive noise p.d.f. has two mass points for certain values of σ , whereas it has a single mass point for other σ 's. Also, in the case of equal priors for $\alpha = 0.12$ and $\alpha = 0.4$, the optimal noise p.d.f.s contain only one mass point at the origin for some values of σ , which implies that the detector is nonimprovable in those scenarios. However, there is always improvement for the unequal priors case, which can be also verified from Figures 2.1, 2.2 and 2.3.

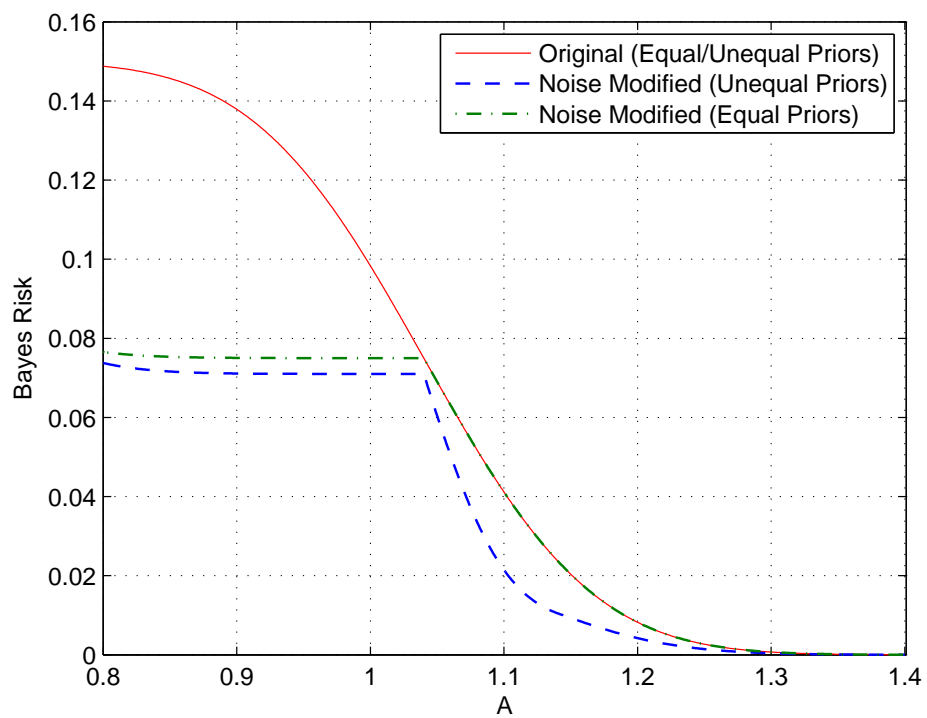


Figure 2.4: Bayes risks of original and noise modified detectors versus A in cases of equal priors and unequal priors for $\alpha = 0.08$ and $\sigma = 0.05$.

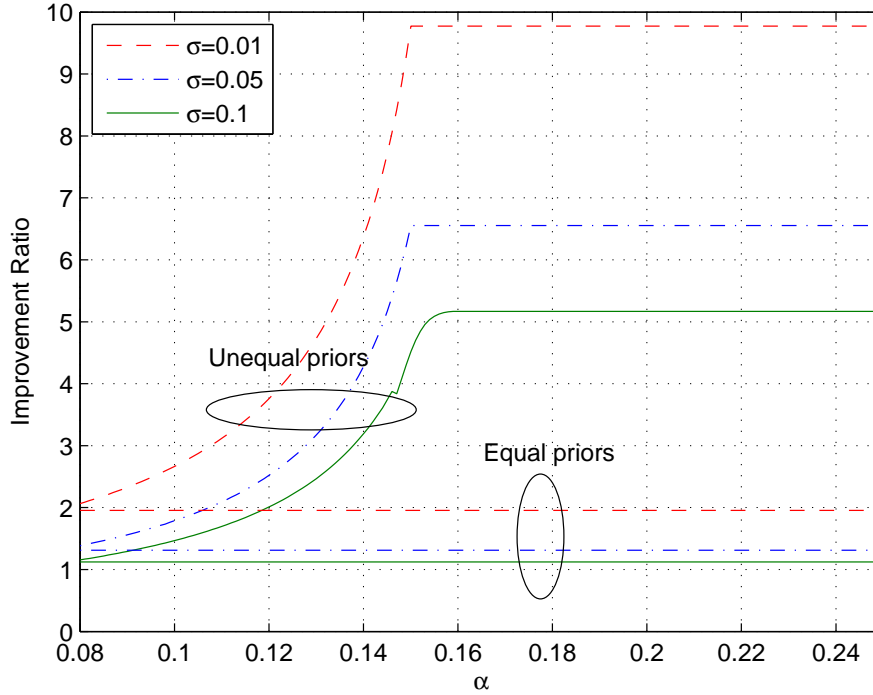


Figure 2.5: Improvement ratio versus α in the cases of equal priors and unequal priors for $\sigma = 0.01$, $\sigma = 0.05$ and $\sigma = 0.1$, where $A = 1$.

Figure 2.4 illustrates the Bayes risks for the original and the noise modified detectors for various values of A in the cases of equal and unequal priors for $\alpha = 0.08$ and $\sigma = 0.05$. It is noted that the original conditional risks are above the specified limit $\alpha = 0.08$ for $A < 1.03$.⁹ However, after the addition of optimal noise, the noise modified detectors result in conditional risks that are below the limit, which is expected since the optimal noise p.d.f.s are obtained from the solution of the constrained optimization problem in (2.26). Another observation from Figure 2.4 is that, in the equal priors case, the improvement decreases as A increases, and there is no improvement after a certain value of A . However, for the unequal priors case, improvement can be observed over a wider range of A values, which is expected due to the the same reasons argued for Figures 2.1-2.3.

⁹For the original detector, the conditional risks are equal; hence, $R_0^x(\phi) = R_1^x(\phi) = r^x(\phi)$.

Figure 2.5 illustrates the improvement ratio, defined as the ratio of the Bayes risks in the absence and presence of additive noise, versus α for the cases of equal and unequal priors for $\sigma = 0.01$, $\sigma = 0.05$ and $\sigma = 0.1$, where $A = 1$ is used. In the unequal priors case, as α increases, an increase is observed in the improvement ratio up to a certain value of α , and then the improvement ratio becomes constant. Those critical α values specify the boundaries between the restricted Bayes and the (unrestricted) Bayes criteria. When α gets larger than those values, the constraint in (2.26) is no longer active; hence, the problem reduces to the Bayesian framework. Therefore, further increases in α do not cause any additional performance improvements. Similarly, as the value of α decreases, the restricted Bayes criterion converges to the minimax criterion [29]. The restricted Bayes criterion achieves its minimum improvement ratio when it becomes equivalent to the minimax criterion, and achieves its maximum improvement ratio when it is equal to the Bayes criterion. In the case of equal priors, the improvement ratio is constant with respect to α , meaning that the improvement for the minimax criterion equals to that for the Bayes criterion. Another observation from the figure is that an increase in σ reduces the improvement ratio, and for the same values of σ , there is more improvement in the unequal priors case. Finally, it should be noted that various values of α in Figure 2.5 correspond to different amounts of uncertainty in the prior information [42]. As the prior information gets more accurate, a larger value of α is selected; hence, the constraint on the conditional risks becomes less strict, meaning that the restricted Bayes criterion converges to the Bayes criterion after a certain value of α . On the other hand, as the amount of uncertainty increases, a smaller value of α is selected, and the restricted Bayes criterion converges to the minimax criterion when α becomes equal to the minimax risk [40], [42].

Next, the improvability conditions in Theorem 5 are investigated for the detection example. To that aim, it is first observed that the original conditional risks $F_0(0)$ and $F_1(0)$ are equal to each other for any value of σ

due to the symmetry of the Gaussian mixture noise (cf. (2.64)). Therefore, $F(0) = \pi_0 F_0(0) + \pi_1 F_1(0) = F_0(0) = F_1(0)$. In addition, suppose that the limit on the conditional risks, α , is set to the original conditional risks for each value of σ , which implies that $\mathcal{S}_\alpha = \{0, 1\}$ in (2.52). Also, the first order derivatives of $F_0(x)$ and $F_1(x)$ at $x = 0$ can be calculated from (2.64) as

$$F_0'(0) = -F_1'(0) = \sum_{i=1}^{N_m} \frac{w_i}{\sqrt{2\pi} \sigma_i} \exp\left(-\frac{(A/2 - \mu_i)^2}{2\sigma_i^2}\right). \quad (2.65)$$

Similarly, the second order derivatives of $F_0(x)$ and $F_1(x)$ at $x = 0$ are obtained as

$$F_0''(0) = F_1''(0) = \sum_{i=1}^{N_m} \frac{w_i(A/2 - \mu_i)}{\sqrt{2\pi} \sigma_i^3} \exp\left(-\frac{(A/2 - \mu_i)^2}{2\sigma_i^2}\right). \quad (2.66)$$

For the unequal priors case, the first and second order derivatives of $F(x) = \pi_0 F_0(x) + \pi_1 F_1(x)$ at $x = 0$ can be expressed as $F'(0) = 0.8F_0'(0)$ and $F''(0) = F_0''(0)$. From (2.65), it is noted that $F_0'(0) > 0$ and $F_1'(0) < 0$; hence, $F'(0) > 0$ as well. Then, from (2.48)-(2.51), set \mathcal{F}_n in (2.54) can be expressed, at $x = 0$, as

$$\begin{aligned} \mathcal{F}_1 &= \{0.8zF_0'(0), zF_0'(0), -zF_0'(0)\}, \\ \mathcal{F}_2 &= \{z^2F_0''(0), z^2F_0''(0), z^2F_0''(0)\}. \end{aligned} \quad (2.67)$$

Therefore, (2.55)-(2.57) imply that, at $x = 0$, $\mathcal{S}_z = \emptyset$, $\mathcal{S}_n = \{3\}$ and $\mathcal{S}_p = \{1, 2\}$ for $z > 0$ and $\mathcal{S}_z = \emptyset$, $\mathcal{S}_n = \{1, 2\}$, and $\mathcal{S}_p = \{3\}$ for $z < 0$.¹⁰ Since $\mathcal{S}_z = \emptyset$, the first condition in Theorem 5 is automatically satisfied. For $z > 0$, $|\mathcal{S}_n| = 1$ and $|\mathcal{S}_p| = 2$; hence, the third bullet of the second condition implies that

$$\min\{\mathcal{F}_2(1)\mathcal{F}_1(2)\mathcal{F}_1(3), \mathcal{F}_2(2)\mathcal{F}_1(1)\mathcal{F}_1(3)\} > \mathcal{F}_2(3)\mathcal{F}_1(1)\mathcal{F}_1(2) \quad (2.68)$$

is required for improvability. For $z < 0$, $|\mathcal{S}_n| = 2$ and $|\mathcal{S}_p| = 1$; hence, the second bullet of the second condition becomes active, which can be shown to yield the same condition as in (2.68). From (2.67), the improvability condition in (2.68)

¹⁰Note that $\mathcal{S}_z = \{1, 2, 3\}$ for $z = 0$, in which case the first condition in Theorem 5 cannot be satisfied since $\mathcal{F}_2 = \{0, 0, 0\}$. Therefore, $z = 0$ is not considered in obtaining improvability conditions.

can be expressed more explicitly as

$$\min \left\{ -z^4 F_0''(0) \left(F_0'(0) \right)^2, -0.8z^4 F_0''(0) \left(F_0'(0) \right)^2 \right\} > 0.8z^4 F_0''(0) \left(F_0'(0) \right)^2, \quad (2.69)$$

which is satisfied when $F_0''(0) < 0$. Therefore, the detector is improvable whenever the expression in (2.66) is negative. For the equal priors case, \mathcal{F}_1 and \mathcal{F}_2 in (2.67) become $\mathcal{F}_1 = \{0, zF_0'(0), -zF_0'(0)\}$ and $\mathcal{F}_2 = \{z^2 F_0''(0), z^2 F_0''(0), z^2 F_0''(0)\}$, respectively. Therefore, the first improvability condition in Theorem 5 requires that $F_0''(0) < 0$, whereas the third bullet of the second condition requires that $\mathcal{F}_2(2)\mathcal{F}_1(3) > \mathcal{F}_2(3)\mathcal{F}_1(2)$ for $z > 0$ and $\mathcal{F}_2(3)\mathcal{F}_1(2) > \mathcal{F}_2(2)\mathcal{F}_1(3)$ for $z < 0$. However, it can be shown that the conditions in the third bullet are always satisfied when $F_0''(0) < 0$. Therefore, the same improvability condition is obtained for the equal priors case, as well. Figure 2.6 illustrates $F_0''(0)$ versus σ for various values of A , where σ represents the standard deviation of the Gaussian mixture noise components ($\sigma_i = \sigma, \forall i$ in (2.62)). It is observed that the detector performance can be improved for $A = 1$ if $\sigma \in [0.005, 0.1597]$, for $A = 0.9$ if $\sigma \in [0.01, 0.1686]$, and for $A = 0.8$ if $\sigma \in [0.02, 0.161]$. On the other hand, the calculations show that the detector is actually improvable for $A = 1$ if $\sigma \leq 0.16$, for $A = 0.9$ if $\sigma \leq 0.17$, and for $A = 0.8$ if $\sigma \leq 0.161$. Hence, the results reveal that the proposed improvability conditions are sufficient but not necessary, and that they are quite effective in determining the range of parameters for which the detector performance can be improved.¹¹

Next, the improvability conditions based on Theorem 3 are considered. For the binary hypothesis-testing example in this section, $H(t)$ in (2.23) becomes $H(t) = \inf \{ \pi_0 F_0(n) + \pi_1 F_1(n) \mid \max\{F_0(n), F_1(n)\} = t, n \in \mathbb{R} \}$. From (2.64), it can be shown that $F_0(n)$ and $F_1(n)$ are monotone increasing and decreasing functions, respectively. In addition, due to the symmetry of the Gaussian mixture

¹¹In fact, $F_0''(0)$ can be shown to be negative even for smaller σ values than specified above; however, very small negative values are computed as zero due to the accuracy limitations.

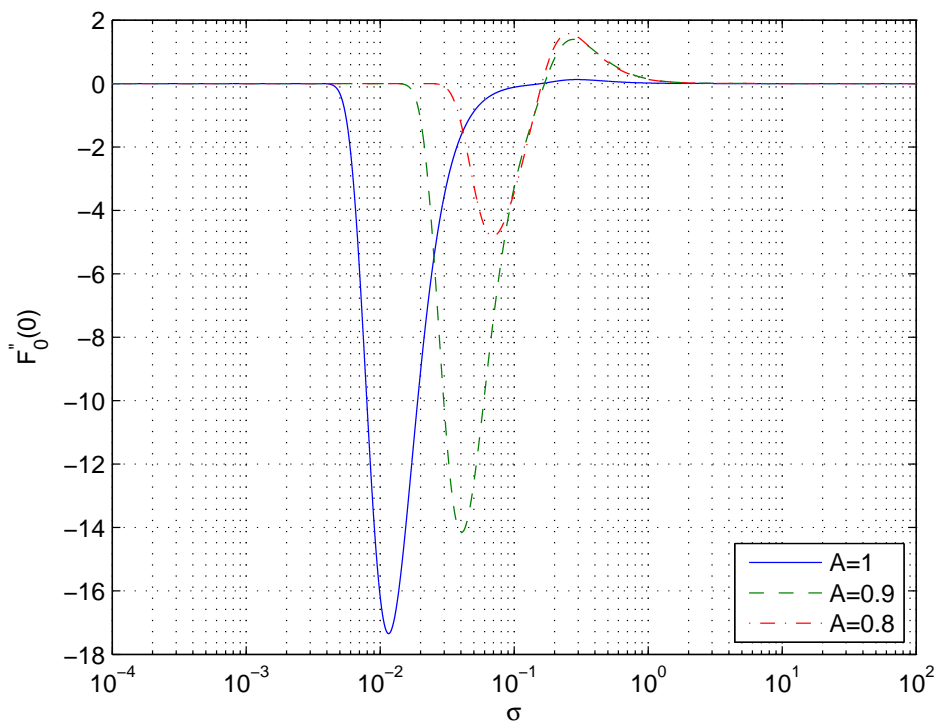


Figure 2.6: The second order derivative of $F_0(x)$ at $x = 0$ versus σ for various values of A . Both Theorem 5 and Theorem 3 imply for the detection example in this section that the detector is improvable whenever $F_0''(0)$ is negative. The limit on the conditional risks, α , is set to the original conditional risks for each value of σ . The graph for $A = 1$ is scaled by 0.1 to make view of the figure more convenient (since only the signs of the graphs are important).

noise, $F_1(n) = F_0(-n)$, $\forall n$. Therefore, without loss of generality, $H(t)$ can be expressed as $H(t) = \pi_0 t + \pi_1 F_1(F_0^{-1}(t))$. Then, the second derivative of $H(t)$ can be obtained as

$$H''(t) = \pi_1 \frac{F_0''(F_0^{-1}(t)) - F_1'(F_0^{-1}(t))F_0''(F_0^{-1}(t))/F_0'(F_0^{-1}(t))}{(F_0'(F_0^{-1}(t)))^2}. \quad (2.70)$$

In order to evaluate the condition in Theorem 5, it is first observed that $t = \tilde{\alpha} = \max\{F_0(0), F_1(0)\} = F_0(0)$, since $F_0(0) = F_1(0)$ (cf. (2.64)). Then, $H''(\tilde{\alpha}) < 0$ implies that $F_1''(0) - F_1'(0)F_0''(0)/F_0'(0) < 0$ for any π_1 . Since $F_0''(0) = F_1''(0)$ from (2.66), and $F_0'(0) > 0$ and $F_1'(0) < 0$ from (2.65), that improvability condition reduces to $F_0''(0) < 0$, which is the same condition obtained from Theorem 5. Therefore, for this specific example, the improvability conditions in Theorem 3 and Theorem 5 are equivalent (cf. Figure 2.6). However, it should be noted that the two conditions are not equivalent in general, and the calculation of $H(t)$ can be difficult in the absence of monotonicity properties related to F_0 and F_1 .

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

$$\begin{aligned} \mathcal{H}_0 &: x = -3\sqrt{A} + v, \\ \mathcal{H}_1 &: x = -\sqrt{A} + v, \\ \mathcal{H}_2 &: x = \sqrt{A} + v, \\ \mathcal{H}_3 &: x = 3\sqrt{A} + v, \end{aligned} \quad (2.71)$$

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

$$\psi_i(x) = \begin{cases} \frac{x}{\sigma_i^2} \exp\left(\frac{-x^2}{2\sigma_i^2}\right), & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (2.72)$$

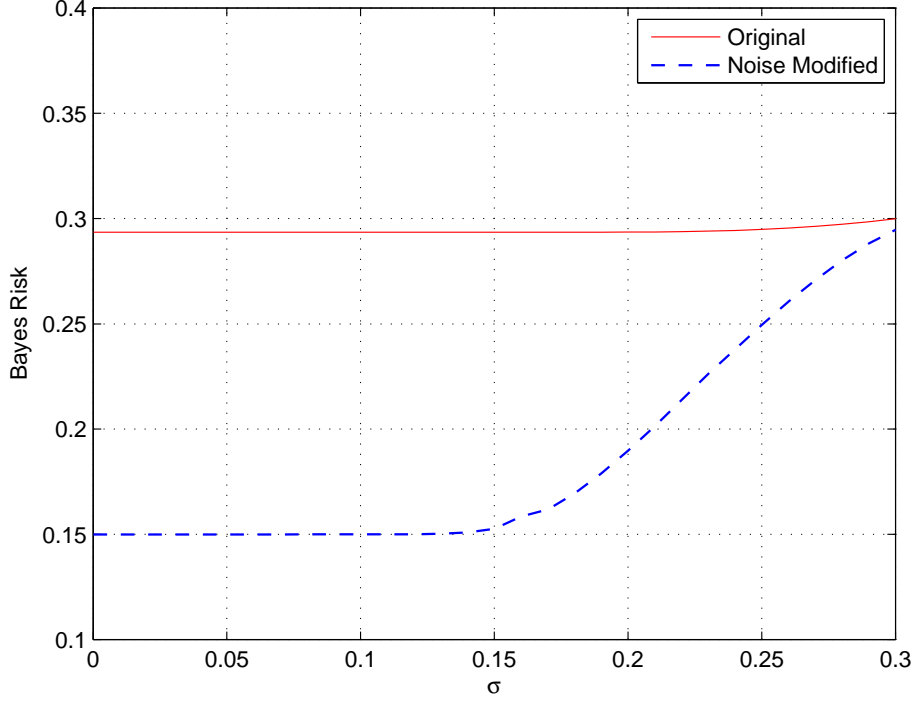


Figure 2.7: Bayes risks of original and noise modified detectors versus σ for $\alpha = 0.4$ and $A = 1$.

for $i = 1, \dots, N_m$. In the numerical results, the same variance is considered for all the Rayleigh components, meaning that $\sigma_i = \sigma, \forall i$. In addition, the parameters are selected as $N_m = 4, \mu_1 = 0.2, \mu_2 = 1, \mu_3 = -2\sigma\sqrt{\frac{\pi}{2}} - 0.2, \mu_4 = -2\sigma\sqrt{\frac{\pi}{2}} - 1, w_1 = w_3 = 0.3$ and $w_2 = w_4 = 0.2$.¹² In addition, the detector is described by

$$\phi(y) = \begin{cases} 0, & y \leq -2\sqrt{A} \\ 1, & -2\sqrt{A} < y \leq 0 \\ 2, & 0 < y \leq 2\sqrt{A} \\ 3, & 2\sqrt{A} < y \end{cases}, \quad (2.73)$$

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

¹²It should be noted that the dependence of the means on σ is necessary in order to keep the noise zero-mean, since the Rayleigh distribution is specified by a single parameter, σ .

Table 2.4: Optimal additive noise p.d.f.s for various values of σ for $\alpha = 0.4$ and $A = 1$.

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

For equal prior probabilities and UCA, Figure 2.7 illustrates the Bayes risk versus σ when $A = 1$ and $\alpha = 0.4$. It is observed that the additive noise can significantly improve the detector performance (equivalently, it reduces the average probability of error of a communications system) for small values of σ . In addition, for the scenario in Figure 2.7, Table 2.4 illustrates the optimal additive noise p.d.f.s for various values of σ . In accordance with Theorem 4, the optimal noise can have up to four non-zero mass points in this problem. Furthermore, for $\sigma = 0.05$, Figure 2.8 plots the Bayes risk versus A for the original and noise modified detectors. A significant improvement is observed for $A \in [0.5, 1]$.

2.4 Concluding Remarks

In this chapter, noise enhanced hypothesis-testing has been studied in the restricted Bayesian framework. First, the most generic formulation of the problem has been considered based on M -ary composite hypothesis-testing, and sufficient conditions for improvability and nonimprovability of detection via additive independent noise have been derived. In addition, an approximate formulation of the optimal noise p.d.f. has been presented. Then, simple hypothesis-testing problems have been studied and additional improvability conditions that are specific to simple hypotheses have been obtained. Also, the optimal noise p.d.f. has

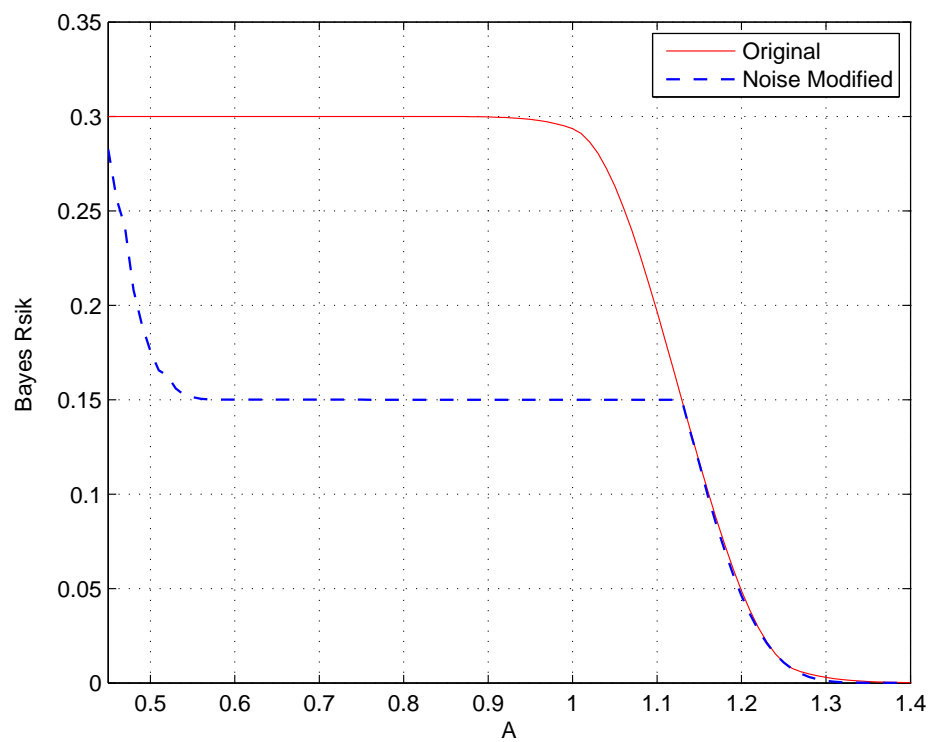


Figure 2.8: Bayes risks of original and noise modified detectors versus A for $\alpha = 0.4$ and $\sigma = 0.05$.

been shown to include at most M mass points for M -ary simple hypothesis-testing problems under certain conditions. Then, various approaches to solving for the optimal noise p.d.f. have been considered, including global optimization techniques, such as the PSO, and a convex relaxation technique. Finally, two detection examples have been studied to illustrate the practicality of the theoretical results.

2.5 Appendices

2.5.1 Proof of Theorem 2

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

$$\sum_{j=1}^L \lambda_j F(\boldsymbol{\epsilon}_j) < F(\mathbf{0}), \quad \sum_{j=1}^L \lambda_j F_{\theta}(\boldsymbol{\epsilon}_j) \leq \alpha, \quad \forall \theta \in \Lambda. \quad (2.74)$$

Since the $\boldsymbol{\epsilon}_j$'s are infinitesimally small, $F(\boldsymbol{\epsilon}_j)$ and $F_{\theta}(\boldsymbol{\epsilon}_j)$ can be approximated by using the Taylor series expansion as $F(\mathbf{0}) + \boldsymbol{\epsilon}_j^T \mathbf{f} + 0.5 \boldsymbol{\epsilon}_j^T \mathbf{H} \boldsymbol{\epsilon}_j$ and $F_{\theta}(\mathbf{0}) + \boldsymbol{\epsilon}_j^T \mathbf{f}_{\theta} + 0.5 \boldsymbol{\epsilon}_j^T \mathbf{H}_{\theta} \boldsymbol{\epsilon}_j$ respectively, where \mathbf{H} and \mathbf{f} (\mathbf{H}_{θ} and \mathbf{f}_{θ}) are the Hessian and the gradient of $F(\mathbf{x})$ ($F_{\theta}(\mathbf{x})$) at $\mathbf{x} = \mathbf{0}$, respectively. Therefore, (2.74) requires that

$$\begin{aligned} \sum_{j=1}^L \lambda_j \boldsymbol{\epsilon}_j^T \mathbf{H} \boldsymbol{\epsilon}_j + 2 \sum_{j=1}^L \lambda_j \boldsymbol{\epsilon}_j^T \mathbf{f} &< 0, \\ \sum_{j=1}^L \lambda_j \boldsymbol{\epsilon}_j^T \mathbf{H}_{\theta} \boldsymbol{\epsilon}_j + 2 \sum_{j=1}^L \lambda_j \boldsymbol{\epsilon}_j^T \mathbf{f}_{\theta} &\leq 2(\alpha - F_{\theta}(\mathbf{0})), \quad \forall \theta \in \Lambda. \end{aligned} \quad (2.75)$$

Let $\boldsymbol{\epsilon}_j = \rho_j \mathbf{z}$ for $j = 1, 2, \dots, L$, where ρ_j for $j = 1, 2, \dots, L$ are infinitesimally small real numbers, and \mathbf{z} is a K -dimensional real vector. Then, based on the

function definitions in (2.19)-(2.22), the conditions in (2.75) can be simplified, after some manipulation, as

$$(f^{(2)}(\mathbf{x}, \mathbf{z}) + c f^{(1)}(\mathbf{x}, \mathbf{z})) \Big|_{\mathbf{x}=\mathbf{0}} < 0, \quad (2.76)$$

$$\left(f_{\theta}^{(2)}(\mathbf{x}, \mathbf{z}) + c f_{\theta}^{(1)}(\mathbf{x}, \mathbf{z}) \right) \Big|_{\mathbf{x}=\mathbf{0}} < \frac{2(\alpha - F_{\theta}(\mathbf{0}))}{\sum_{j=1}^L \lambda_j \rho_j^2}, \quad \forall \theta \in \Lambda, \quad (2.77)$$

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

Since $\alpha = F_{\theta^*}(\mathbf{0})$ and $\alpha > \max_{\theta \in \Lambda \setminus \theta^*} F_{\theta}(\mathbf{0})$, the right-hand-side of (2.77) goes to infinity for $\theta \neq \theta^*$. Hence, we should consider only the $\theta = \theta^*$ case. Thus, (2.76) and (2.77) can be expressed as

$$(f^{(2)}(\mathbf{x}, \mathbf{z}) + c f^{(1)}(\mathbf{x}, \mathbf{z})) \Big|_{\mathbf{x}=\mathbf{0}} < 0, \quad (2.78)$$

$$\left(f_{\theta^*}^{(2)}(\mathbf{x}, \mathbf{z}) + c f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) \right) \Big|_{\mathbf{x}=\mathbf{0}} < 0. \quad (2.79)$$

It is noted that c can take any real value by definition via selection of appropriate λ_i and infinitesimally small ρ_i values for $i = 1, 2, \dots, L$. Therefore, for the first part of the theorem, under the condition of $f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) f^{(1)}(\mathbf{x}, \mathbf{z}) > 0$ at $\mathbf{x} = \mathbf{0}$, which states that the second term in (2.78) has the same sign as the second term in (2.79), there always exists c that satisfies the improvability conditions in (2.78) and (2.79). For the second part of the theorem, since $f^{(1)}(\mathbf{x}, \mathbf{z}) > 0$ and $f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) < 0$ at $\mathbf{x} = \mathbf{0}$, (2.78) and (2.79) can also be expressed as

$$\left(f^{(2)}(\mathbf{x}, \mathbf{z}) f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) + c f^{(1)}(\mathbf{x}, \mathbf{z}) f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) \right) \Big|_{\mathbf{x}=\mathbf{0}} > 0, \quad (2.80)$$

$$\left(f_{\theta^*}^{(2)}(\mathbf{x}, \mathbf{z}) f^{(1)}(\mathbf{x}, \mathbf{z}) + c f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) f^{(1)}(\mathbf{x}, \mathbf{z}) \right) \Big|_{\mathbf{x}=\mathbf{0}} < 0. \quad (2.81)$$

Under the condition of $f^{(2)}(\mathbf{x}, \mathbf{z}) f_{\theta^*}^{(1)}(\mathbf{x}, \mathbf{z}) > f_{\theta^*}^{(2)}(\mathbf{x}, \mathbf{z}) f^{(1)}(\mathbf{x}, \mathbf{z})$ at $\mathbf{x} = \mathbf{0}$, which states that the first term in (2.80) is larger than the first term in (2.81), there always exists c that satisfies the improvability conditions in (2.80) and (2.81).

2.5.2 Proof of Theorem 3

Since $H''(\tilde{\alpha}) < 0$ and $H(t)$ in (2.23) is second-order continuously differentiable around $t = \tilde{\alpha}$, there exist $\epsilon > 0$, \mathbf{n}_1 and \mathbf{n}_2 such that $\max_{\theta \in \Lambda} F_\theta(\mathbf{n}_1) = \tilde{\alpha} + \epsilon$ and $\max_{\theta \in \Lambda} F_\theta(\mathbf{n}_2) = \tilde{\alpha} - \epsilon$. Then, it is proven in the following that an additive noise component with $p_{\mathbf{N}}(\mathbf{n}) = 0.5 \delta(\mathbf{x} - \mathbf{n}_1) + 0.5 \delta(\mathbf{x} - \mathbf{n}_2)$ improves the detector performance under the conditional risk constraint. First, the maximum value of the conditional risks in the presence of additive noise is shown not to exceed α :

$$\max_{\theta \in \Lambda} E\{F_\theta(\mathbf{N})\} \leq E\left\{\max_{\theta \in \Lambda} F_\theta(\mathbf{N})\right\} = 0.5(\tilde{\alpha} + \epsilon) + 0.5(\tilde{\alpha} - \epsilon) = \tilde{\alpha} \leq \alpha. \quad (2.82)$$

Then, the decrease in the Bayes risk is proven as follows. Due to the assumptions in the theorem, $H(t)$ is concave in an interval around $t = \tilde{\alpha}$. Since $E\{F(\mathbf{N})\}$ can attain the value of $0.5 H(\tilde{\alpha} + \epsilon) + 0.5 H(\tilde{\alpha} - \epsilon)$, which is always smaller than $H(\tilde{\alpha})$ due to concavity, it is concluded that $E\{F(\mathbf{N})\} < H(\tilde{\alpha})$. As $H(\tilde{\alpha}) \leq F(\mathbf{0})$ by definition of $H(t)$ in (2.23), $E\{F(\mathbf{N})\} < F(\mathbf{0})$ is satisfied; hence, the detector is improvable.

2.5.3 Maximum Conditional Risk Achieved by Optimal Noise

Consider the case in which $t_m = \arg \min_t H(t) > \alpha$. In order to prove that “ $\max_{\theta \in \Lambda} R_\theta^{\mathbf{y}}(\phi) = \alpha$ for the optimal noise” by contradiction, first assume that the optimal solution of (2.12) is given by $p_{\tilde{\mathbf{N}}}(\mathbf{x})$ with $\beta \triangleq \max_{\theta \in \Lambda} R_\theta^{\mathbf{y}}(\phi) < \alpha$. As in the proof of Theorem 4 in [12], we define another noise \mathbf{N} with the following p.d.f.:

$$p_{\mathbf{N}}(\mathbf{n}) = \frac{\alpha - \beta}{t_m - \beta} \delta(\mathbf{n} - \mathbf{n}_m) + \frac{t_m - \alpha}{t_m - \beta} p_{\tilde{\mathbf{N}}}(\mathbf{n}), \quad (2.83)$$

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

For the noise p.d.f. in (2.83), the Bayes risk and conditional risks can be calculated as

$$r^{\mathbf{y}}(\phi) = \mathbb{E}\{F(\mathbf{N})\} = \frac{\alpha - \beta}{t_m - \beta} F(\mathbf{n}_m) + \frac{t_m - \alpha}{t_m - \beta} r^{\tilde{\mathbf{y}}}(\phi) , \quad (2.84)$$

$$R_\theta^{\mathbf{y}}(\phi) = \mathbb{E}\{F_\theta(\mathbf{N})\} = \frac{\alpha - \beta}{t_m - \beta} F_\theta(\mathbf{n}_m) + \frac{t_m - \alpha}{t_m - \beta} R_\theta^{\tilde{\mathbf{y}}}(\phi) , \quad (2.85)$$

for all $\theta \in \Lambda$. Since $F(\mathbf{n}_m) < r^{\tilde{\mathbf{y}}}(\phi)$, (2.84) implies $r^{\mathbf{y}}(\phi) < r^{\tilde{\mathbf{y}}}(\phi)$. On the other hand, as $F_\theta(\mathbf{n}_m) \leq t_m$ and $R_\theta^{\tilde{\mathbf{y}}}(\phi) \leq \beta$, $R_\theta^{\mathbf{y}}(\phi) \leq \alpha$ is obtained. Therefore, $\tilde{\mathbf{N}}$ cannot be an optimal solution, which implies a contradiction. In other words, any noise p.d.f. that satisfies $\max_{\theta \in \Lambda} R_\theta^{\mathbf{y}}(\phi) < \alpha$ cannot be optimal.

2.5.4 Proof of Theorem 5

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

$$\sum_{l=1}^M \lambda_l F(\mathbf{n}_l) < F(\mathbf{0}) , \quad \max_{i \in \{0,1,\dots,M-1\}} \sum_{l=1}^M \lambda_l F_i(\mathbf{n}_l) \leq \alpha . \quad (2.86)$$

As in the proof of Theorem 2 in Appendix 2.5.1, consider the improvability conditions in (2.86) with infinitesimally small noise components, $\mathbf{n}_l = \boldsymbol{\epsilon}_l = \rho_l \mathbf{z}$ for $l = 1, 2, \dots, M$, where ρ_l 's are infinitesimally small real numbers, and \mathbf{z} is a K -dimensional real vector. Then, similar manipulations to those in Appendix A (cf. (2.75)-(2.77)) can be performed to obtain

$$(f^{(2)}(\mathbf{x}, \mathbf{z}) + c f^{(1)}(\mathbf{x}, \mathbf{z})) \Big|_{\mathbf{x}=\mathbf{0}} < 0 , \quad (2.87)$$

$$(f_i^{(2)}(\mathbf{x}, \mathbf{z}) + c f_i^{(1)}(\mathbf{x}, \mathbf{z})) \Big|_{\mathbf{x}=\mathbf{0}} < \frac{2(\alpha - F_i(\mathbf{0}))}{\sum_{j=1}^M \lambda_j \rho_j^2} , \quad (2.88)$$

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

Since $F_i(0) < \alpha$, $\forall i \in \bar{\mathcal{S}}_\alpha$, the right-hand-side of (2.88) goes to infinity for $i \in \bar{\mathcal{S}}_\alpha$. Hence, one can consider $i \in \mathcal{S}_\alpha$ only. Thus, (2.87) and (2.88) can be expressed as

$$\left(f^{(2)}(\mathbf{x}, \mathbf{z}) + c f^{(1)}(\mathbf{x}, \mathbf{z}) \right) \Big|_{\mathbf{x}=\mathbf{0}} < 0, \quad (2.89)$$

$$\left(f_i^{(2)}(\mathbf{x}, \mathbf{z}) + c f_i^{(1)}(\mathbf{x}, \mathbf{z}) \right) \Big|_{\mathbf{x}=\mathbf{0}} < 0, \quad \forall i \in \mathcal{S}_\alpha. \quad (2.90)$$

Based on the definition in (2.54), (2.89) and (2.90) can be restated as

$$\left(\mathcal{F}_2(j) + c \mathcal{F}_1(j) \right) \Big|_{\mathbf{x}=\mathbf{0}} < 0 \quad \text{for } j = 1, 2, \dots, |\mathcal{S}_\alpha| + 1. \quad (2.91)$$

It is noted that c can take any real value by selecting appropriate λ_i and infinitesimally small ρ_i values for $i = 0, 1, \dots, M - 1$. From (2.55), it is concluded that in order for the conditions in (2.91) to hold,

$$\mathcal{F}_2(j) \Big|_{\mathbf{x}=\mathbf{0}} < 0 \quad (2.92)$$

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

In addition to (2.92), one of the following conditions should be satisfied for the improvability conditions in (2.91) to hold:

- When $|\mathcal{S}_n| = 0$ or $|\mathcal{S}_p| = 0$, as stated in the first part of the second condition in Theorem 5, all the second terms in (2.91) (namely, $\mathcal{F}_1(1), \dots, \mathcal{F}_1(|\mathcal{S}_\alpha| + 1)$) are either all non-negative or all non-positive. Therefore, there always exists a c that satisfies the improvability conditions in (2.91) when the first condition in Theorem 5 (cf. (2.92)) is satisfied.
- When $|\mathcal{S}_n|$ is a positive even number and $|\mathcal{S}_p| > 0$, (2.91) can be expressed, after some manipulation, as

$$\mathcal{F}_2(j) \Big|_{\mathbf{x}=\mathbf{0}} < 0, \quad \forall j \in \mathcal{S}_z, \quad (2.93)$$

$$\left(\mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l) + c \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n} \mathcal{F}_1(l) \right) \Big|_{\mathbf{x}=\mathbf{0}} < 0, \quad \forall j \in \mathcal{S}_p, \quad (2.94)$$

$$\left(\mathcal{F}_2(j) \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l) + c \prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n} \mathcal{F}_1(l) \right) \Big|_{\mathbf{x}=\mathbf{0}} > 0, \quad \forall j \in \mathcal{S}_n. \quad (2.95)$$

Note that (2.94) and (2.95) are obtained by multiplying (2.91) by $\prod_{l \in \mathcal{S}_p \cup \mathcal{S}_n \setminus \{j\}} \mathcal{F}_1(l)$, which is a positive (negative) quantity when $j \in \mathcal{S}_p$ ($j \in \mathcal{S}_n$) since $|\mathcal{S}_n|$ is even. The condition in (2.93) is satisfied due to the first condition in Theorem 5. In addition, under the condition in (2.58), there always exists a c that satisfies the improvability conditions in (2.94) and (2.95).

- When $|\mathcal{S}_n|$ is an odd number and $|\mathcal{S}_p| > 0$, (2.91) can be expressed by three conditions as in (2.93)-(2.95) with the only difference being that the signs of the inequalities in (2.94) and (2.95) are switched. In that case, the first condition (cf. (2.93)) is satisfied due to the first condition in Theorem 5. Also, under the condition in (2.59), there always exists a c that satisfies the second and third conditions.

2.5.5 Proof of Corollary 1

Consider the proof of Theorem 5 above. Since $\alpha > \max_{i \in \{0,1,\dots,M-1\}} F_i(\mathbf{0})$, the right-hand-side of (2.88) becomes infinity for any i . Therefore, we can consider the condition in (2.87) only; that is,

$$(f^{(2)}(\mathbf{x}, \mathbf{z}) + c f^{(1)}(\mathbf{x}, \mathbf{z})) \Big|_{\mathbf{x}=\mathbf{0}} < 0. \quad (2.96)$$

In terms of the gradient \mathbf{f} and the Hessian \mathbf{H} of $F(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$, (2.96) becomes $\mathbf{z}^T \mathbf{H} \mathbf{z} + c \mathbf{z}^T \mathbf{f} < 0$. Since c can take any real value by definition (cf. Appendix 2.5.4) and \mathbf{z} can be chosen arbitrarily small, the improvability condition can always be satisfied if $\mathbf{f} \neq \mathbf{0}$. On the other hand, if $\mathbf{f} = \mathbf{0}$, then the improvability condition becomes $\mathbf{z}^T \mathbf{H} \mathbf{z} < 0$. If $F(\mathbf{x})$ is not convex around $\mathbf{x} = \mathbf{0}$, \mathbf{H} is not positive semidefinite. Therefore, there exists \mathbf{z} such that $\mathbf{z}^T \mathbf{H} \mathbf{z} < 0$ is satisfied; hence, the detector is improvable.

Chapter 3

Noise Enhanced M -ary

Composite Hypothesis-Testing in the Presence of Partial Prior

Information

This chapter is organized as follows. Section 3.1 introduces M -ary composite hypothesis-testing problems under partial prior information, and defines two criteria for the calculation of optimal additive noise. Investigations of optimal additive noise and improvability conditions for those criteria are provided in Sections 3.2 and 3.3. In Section 3.4, the cases of unknown parameter distributions for some composite hypotheses are studied, and upper bounds on the risks are provided. Finally, a detection example is studied in Section 3.5 in order to investigate the theoretical results.

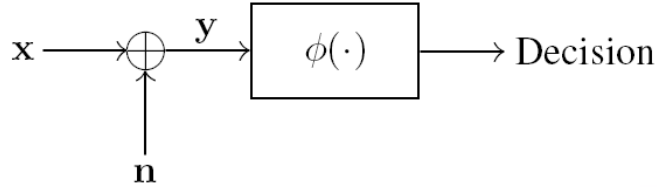


Figure 3.1: Independent noise \mathbf{n} is added to observation \mathbf{x} in order to improve the performance of the detector, represented by $\phi(\cdot)$.

3.1 Problem Formulation

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

$$\mathcal{H}_i : p_{\theta}^{\mathbf{X}}(\mathbf{x}) , \theta \in \Lambda_i , \quad i = 0, 1, \dots, M - 1 , \quad (3.1)$$

where \mathcal{H}_i denotes the i th hypothesis and $p_{\theta}^{\mathbf{X}}(\mathbf{x})$ represents the probability density function (p.d.f.) of observation \mathbf{X} for a given value of $\Theta = \theta$. Each observation (measurement) \mathbf{x} is a vector with K components; i.e., $\mathbf{x} \in \mathbb{R}^K$, and $\Lambda_0, \Lambda_1, \dots, \Lambda_{M-1}$ form a partition of the parameter space Λ . The distribution of the unknown parameter Θ for hypothesis i is represented by $w_i(\theta)$ for $i = 0, 1, \dots, M - 1$. In addition, the prior probability of hypothesis \mathcal{H}_i is denoted by π_i for $i = 0, 1, \dots, M - 1$. Composite hypothesis-testing problems as in (3.1) are encountered in various problems, such as in non-coherent communications receivers, pattern recognition, and time series analysis [40], [98]. Note that when Λ_i 's consist of single elements, the problem reduces to a *simple* hypothesis-testing problem.

A generic decision rule (detector) can be defined as

$$\phi(\mathbf{x}) = i , \quad \text{if } \mathbf{x} \in \Gamma_i , \quad (3.2)$$

for $i = 0, 1, \dots, M - 1$, where $\Gamma_0, \Gamma_1, \dots, \Gamma_{M-1}$ form a partition of the observation space Γ . As shown in Figure 3.1, the aim is to add noise to the original observation \mathbf{x} (which commonly consists of a signal component and measurement noise) in order to improve the performance of the detector according to certain

criteria [80]. By adding noise \mathbf{n} to the original observation \mathbf{x} , the modified observation is formed as $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where \mathbf{n} has a p.d.f. denoted by $p_{\mathbf{N}}(\cdot)$, and is independent of \mathbf{x} . It should be noted that the additive noise can cause both positive and negative shifts in the observations [23], [29]. As in [12] and [23], it is assumed that the detector ϕ , described by (3.2), is fixed, and the only means for improving the performance of the detector is to optimize the additive noise \mathbf{n} .

When all the prior probabilities $\pi_0, \pi_1, \dots, \pi_{M-1}$ of the hypotheses in (3.1) are known, the Bayesian approach can be taken, and the optimal additive noise that minimizes the Bayes risk can be sought for. This problem is studied in [23] for simple hypothesis-testing problems under uniform cost assignment (UCA). On the other hand, when none of the prior probabilities are known, the minimax approach can be taken to obtain the optimal additive noise that minimizes the maximum conditional risk, which is investigated in [29] for simple hypothesis-testing problems. In this chapter, we focus on a more generic scenario by considering both *composite hypotheses* and *partial prior information*, meaning that the prior probabilities of some hypotheses and the probability distributions of the unknown parameters under some hypotheses may be unknown. Such a generalization can be important in practice since composite hypothesis-testing problems are encountered in many applications, and the prior information may not be available for all hypotheses (see Section 3.5 for an example).

In order to introduce a generic problem formulation, define sets $\mathcal{S}_1, \dots, \mathcal{S}_G$ that form a partition of set $\{0, 1, \dots, M-1\}$. Suppose that the prior probability π_i of \mathcal{H}_i is known if $i \in \mathcal{S}_1$ and it is unknown otherwise, and assume that the size of set \mathcal{S}_1 is $M - N_u$. In other words, \mathcal{S}_1 corresponds to $M - N_u$ hypotheses with known prior probabilities. In addition, assume that the hypotheses with unknown prior probabilities are grouped into sets $\mathcal{S}_2, \dots, \mathcal{S}_G$ in such a way that the sum of the prior probabilities of the hypotheses in set \mathcal{S}_j is known for $j = 2, \dots, G$.

If no such information is available, then $G = 2$ can be employed; that is, all the hypotheses with unknown probabilities can be grouped together into \mathcal{S}_2 .

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

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

$$r_1(\phi) = \sum_{i \in \mathcal{S}_1} \pi_i R_i(\phi) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} R_i(\phi), \quad (3.3)$$

where $R_i(\phi)$ is the conditional risk of decision rule ϕ when hypothesis i is true [40], $|\mathcal{S}_j|$ denotes the number of elements in set \mathcal{S}_j , and $\tilde{\pi}_j \triangleq \sum_{i \in \mathcal{S}_j} \pi_i$ defines the sum of the prior probabilities of the hypotheses in \mathcal{S}_j for $j = 2, \dots, G$. According to Criterion 1, the optimal additive noise is defined as $p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg \min_{p_{\mathbf{N}}(\mathbf{n})} r_1(\phi)$, where $r_1(\phi)$ is given by (3.3). It should be noted that assuming uniform distribution for the unknown priors is a very popular classical approach [99].

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

$$r_2(\phi) = \sum_{i \in \mathcal{S}_1} \pi_i R_i(\phi) + \sum_{j=2}^G \tilde{\pi}_j \max_{i \in \mathcal{S}_j} R_i(\phi). \quad (3.4)$$

In other words, a conservative approach is taken in Criterion 2, and the worst-case Bayes risk is considered as the performance metric. Such an approach can be considered in the framework of Γ -minimax decision rules [59]. According to Criterion 2, the optimal additive noise is calculated from $p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg \min_{p_{\mathbf{N}}(\mathbf{n})} r_2(\phi)$.

In Section 3.2 and Section 3.3, the optimal additive noise will be investigated when the probability distributions of the unknown parameters are known under all hypotheses (the prior probabilities can still be unknown). Then, in Section

3.4, the results will be extended to the cases in which the probability distributions of the unknown parameters are unknown under some hypotheses.

3.2 Optimal Additive Noise According to Criterion 1

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

$$p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg \min_{p_{\mathbf{N}}(\mathbf{n})} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i R_i(\phi) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} R_i(\phi) \right\}. \quad (3.5)$$

Since $R_i(\phi)$ is the conditional risk for hypotheses i , it can be expressed as

$$R_i(\phi) = \int_{\Lambda} R_{\theta}(\phi) w_i(\theta) d\theta, \quad (3.6)$$

where $R_{\theta}(\phi)$ denotes the conditional risk that is defined as the average cost of decision rule ϕ for a given $\theta \in \Lambda$ [40]. The conditional risk can be calculated from

$$R_{\theta}(\phi) = E\{C[\phi(\mathbf{Y}), \Theta] | \Theta = \theta\} = \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^Y(\mathbf{y}) d\mathbf{y}, \quad (3.7)$$

where $p_{\theta}^Y(\mathbf{y})$ is the p.d.f. of the noise modified observation for a given value of $\Theta = \theta$, and $C[j, \theta] \geq 0$ is the cost of deciding \mathcal{H}_j when $\Theta = \theta$, for $\theta \in \Lambda$ [40].

Since the additive noise is independent of the original observation, $p_{\theta}^Y(\mathbf{y}) = \int_{\mathbb{R}^K} p_{\theta}^X(\mathbf{y} - \mathbf{n}) p_{\mathbf{N}}(\mathbf{n}) d\mathbf{n}$. Then, the expression in (3.6) for the conditional risk of hypotheses i can be manipulated from (3.7) as follows:

$$\begin{aligned} R_i(\phi) &= \int_{\Lambda} \int_{\Gamma} \int_{\mathbb{R}^K} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{n}) p_{\mathbf{N}}(\mathbf{n}) w_i(\theta) d\mathbf{n} d\mathbf{y} d\theta \\ &= \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) \left[\int_{\Lambda} \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{n}) w_i(\theta) d\mathbf{y} d\theta \right] d\mathbf{n} \\ &\triangleq \int_{\mathbb{R}^K} p_{\mathbf{N}}(\mathbf{n}) f_i(\mathbf{n}) d\mathbf{n} = E\{f_i(\mathbf{N})\} \end{aligned} \quad (3.8)$$

where

$$f_i(\mathbf{n}) \triangleq \int_{\Lambda} \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{n}) w_i(\theta) d\mathbf{y} d\theta . \quad (3.9)$$

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

Based on (3.8), the optimization problem in (3.5) can be expressed as

$$\begin{aligned} p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) &= \arg \min_{p_{\mathbf{N}}(\mathbf{n})} \mathbb{E} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i f_i(\mathbf{N}) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} f_i(\mathbf{N}) \right\} \\ &\triangleq \arg \min_{p_{\mathbf{N}}(\mathbf{n})} \mathbb{E} \{ f(\mathbf{N}) \} , \end{aligned} \quad (3.10)$$

where $f(\mathbf{n})$ is defined as $f(\mathbf{n}) \triangleq \sum_{i \in \mathcal{S}_1} \pi_i f_i(\mathbf{n}) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} f_i(\mathbf{n})$. From (3.10), the optimal noise p.d.f. can be obtained by assigning all the probability to the minimizer of $f(\mathbf{n})$; i.e.,

$$p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \delta(\mathbf{n} - \mathbf{n}_0) , \quad \mathbf{n}_0 = \arg \min_{\mathbf{n}} f(\mathbf{n}) . \quad (3.11)$$

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

The main implication of the result in (3.11) is that among all p.d.f.s for the additive independent noise \mathbf{N} , the ones that assign all the probability to a single noise value can be used as the optimal additive signal components in Figure 3.1. In other words, in this scenario, addition of independent noise to observations corresponds to shifting the decision region of the detector.

Based on the expressions in (3.10), a detector is called *improvable* according to Criterion 1 if there exists noise \mathbf{N} that satisfies $\mathbb{E}\{f(\mathbf{N})\} < f(\mathbf{0})$, where $f(\mathbf{0})$ represents the Bayes risk in (3.3) in the absence of additive noise. For example,

if there exists a noise component \mathbf{n}_* that satisfies $f(\mathbf{n}_*) < f(\mathbf{0})$, the detector can be classified as an improvable one according to Criterion 1. In the following, sufficient conditions are provided to determine the improvability of a detector without actually solving the optimization problem in (3.11).

Proposition 1: *Assume that $f(\mathbf{x})$ in (3.10) is second-order continuously differentiable around $\mathbf{x} = \mathbf{0}$. Let \mathbf{f} denote the gradient of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$. Then, the detector is improvable*

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

Proof: Please see Appendix 3.6.1.

Although Proposition 1 may not be very crucial for scalar observations (since it can be easy to find the optimal solution from (3.11) directly), it can be useful for vector observations by providing simple sufficient conditions to check if the detector can be improved via additive noise.

3.3 Optimal Additive Noise According to Criterion 2

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

$$p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg \min_{p_{\mathbf{N}}(\mathbf{n})} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i R_i(\phi) + \sum_{j=2}^G \tilde{\pi}_j \max_{l \in \mathcal{S}_j} R_l(\phi) \right\}, \quad (3.12)$$

which can also be expressed as

$$p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg \min_{p_{\mathbf{N}}(\mathbf{n})} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i R_i(\phi) + \max_{l \in \tilde{\mathcal{S}}} \sum_{j=2}^G \tilde{\pi}_j R_{l_j}(\phi) \right\}, \quad (3.13)$$

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

From (3.8), the optimization problem in (3.13) can be stated as

$$\begin{aligned} p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) &= \arg \min_{p_{\mathbf{N}}(\mathbf{n})} \max_{\mathbf{l} \in \tilde{\mathcal{S}}} \mathbb{E} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i f_i(\mathbf{N}) + \sum_{j=2}^G \tilde{\pi}_j f_{l_j}(\mathbf{N}) \right\} \\ &\triangleq \arg \min_{p_{\mathbf{N}}(\mathbf{n})} \max_{\mathbf{l} \in \tilde{\mathcal{S}}} \mathbb{E} \{ f_{\mathbf{l}}(\mathbf{N}) \} , \end{aligned} \quad (3.14)$$

where $f_i(\cdot)$ and $f_{l_j}(\cdot)$ are as defined in (3.9), and $f_{\mathbf{l}}(\mathbf{N}) \triangleq \sum_{i \in \mathcal{S}_1} \pi_i f_i(\mathbf{N}) + \sum_{j=2}^G \tilde{\pi}_j f_{l_j}(\mathbf{N})$.

Although the optimization problem in (3.14) seems quite difficult to solve in general, the following proposition states that the optimization can be performed over a significantly reduced space as the optimal solution can be characterized by a discrete probability distribution under certain conditions. To that aim, assume that all possible additive noise values satisfy $\mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b}$ for any finite \mathbf{a} and \mathbf{b} ; that is, $n_j \in [a_j, b_j]$ for $j = 1, \dots, K$, which is a reasonable assumption since additive noise cannot have infinitely large amplitudes in practice. Then, the following proposition states the discrete nature of the optimal additive noise.

Proposition 2: *If $f_{\mathbf{l}}(\cdot)$ in (3.14) are continuous functions, the p.d.f. of optimal additive noise can be expressed as*

$$p_{\mathbf{N}}(\mathbf{n}) = \sum_{j=1}^{|\tilde{\mathcal{S}}|} \lambda_j \delta(\mathbf{n} - \mathbf{n}_j) , \quad (3.15)$$

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

Proof: The proof is omitted since the result can be proven similarly to [12], [29]. The assumption $\mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b}$ is used to guarantee the existence of the optimal solution [29]. \square

Proposition 2 implies that optimal additive noise can be represented by a randomization of no more than $|\tilde{\mathcal{S}}|$ different signal levels. Therefore, the solution

of the optimization problem in (3.14) can be obtained from the following:

$$\begin{aligned} & \min_{\{\mathbf{n}_j, \lambda_j\}_{j=1}^{|\tilde{\mathcal{S}}|}} \max_{l \in \tilde{\mathcal{S}}} \sum_{j=1}^{|\tilde{\mathcal{S}}|} \lambda_j f_l(\mathbf{n}_j) \\ & \text{subject to } \sum_{j=1}^{|\tilde{\mathcal{S}}|} \lambda_j = 1, \quad \lambda_j \geq 0, \quad j = 1, \dots, |\tilde{\mathcal{S}}|. \end{aligned} \quad (3.16)$$

Although (3.16) is significantly simpler than (3.14), it can still be a non-convex optimization problem. Therefore, global optimization techniques, such as particle-swarm optimization (PSO) [51], genetic algorithms, and differential evolution [82] can be employed to obtain the optimal additive noise p.d.f.. Alternatively, a convex relaxation approach can be taken as in [29] in order to obtain an approximate solution.

3.4 Unknown Parameter Distributions for Some Hypotheses

In the previous formulations, it is assumed that the distribution of the unknown parameter for hypothesis i , denoted by $w_i(\theta)$, is known for $i = 0, 1, \dots, M-1$ (see (3.6)).¹ If this information is not available for certain hypotheses, an approach similar to that in [63] can be taken, and the conditional risks for those hypotheses can be defined as the worst-case conditional risks; that is, $R_i(\phi) = \sup_{\theta \in \Lambda_i} R_\theta(\phi)$, where $R_\theta(\phi)$ is as in (3.7). In other words, for hypotheses with unknown parameter distributions, the maximum conditional risk is set by taking a conservative approach. On the other hand, for hypotheses with known parameter distributions, the average conditional risk in (3.6) can still be obtained. Therefore, the

¹Note that this assumption is not needed for simple hypotheses since there is only one possible parameter value.

definition of $R_i(\phi)$ can be extended as

$$R_i(\phi) = \begin{cases} \int_{\Lambda} R_{\theta}(\phi) w_i(\theta) d\theta, & \text{if } w_i(\theta) \text{ is known} \\ \sup_{\theta \in \Lambda_i} R_{\theta}(\phi), & \text{if } w_i(\theta) \text{ is unknown} \end{cases}, \quad (3.17)$$

for $i = 0, 1, \dots, M - 1$. Then, Criterion 1 in (3.3) and Criterion 2 in (3.4) can still be used in evaluating the performance of detectors.

Remark: Instead of considering the worst-case conditional risks as in (3.17), another approach is to assume a uniform distribution of parameter θ over Λ_i when $w_i(\theta)$ is unknown. In that case, all the results in Section 3.2 and Section 3.3 are still valid. Hence, we focus on the approach in (3.17) in this section.

When the parameter distributions for some hypotheses are unknown and the extended definition of $R_i(\phi)$ in (3.17) is used, the discrete structures of the probability distributions of optimal additive noise (see (3.11) and Proposition 2) may not be guaranteed anymore. In other words, the optimal additive noise may also have continuous probability distributions in that scenario. Therefore, in order to obtain the (approximate) p.d.f. of the optimal additive noise, the approach in [50] can be taken in order to search over possible p.d.f.s in the form of $p_{\mathbf{N}}(\mathbf{n}) = \sum_l \zeta_l \psi(\mathbf{n} - \mathbf{n}_l)$, where $\zeta_l \geq 0$, $\sum_l \zeta_l = 1$, and $\psi_l(\cdot)$ is a window function that satisfies $\psi_l(\mathbf{x}) \geq 0, \forall \mathbf{x}$ and $\int \psi_l(\mathbf{x}) d\mathbf{x} = 1, \forall l$.

Since the computational complexity of searching over possible additive noise p.d.f.s in the form of $p_{\mathbf{N}}(\mathbf{n}) = \sum_l \zeta_l \psi(\mathbf{n} - \mathbf{n}_l)$ can be high in some cases, it becomes important to specify theoretical upper bounds on $r_1(\phi)$ in (3.3) and $r_2(\phi)$ in (3.4) (with $R_i(\phi)$ being given by (3.17)), which can be achieved under certain scenarios. The following lemma presents such upper bounds.

Lemma 1: When the conditional risk $R_i(\phi)$ is defined as in (3.17), $r_1(\phi)$ in (3.3) and $r_2(\phi)$ in (3.4) are upper bounded as follows:

$$r_1(\phi) \leq \mathbb{E} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i \tilde{f}_i(\mathbf{N}) + \sum_{j=2}^G \frac{\tilde{\pi}_j}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} \tilde{f}_i(\mathbf{N}) \right\} \quad (3.18)$$

$$r_2(\phi) \leq \max_{\mathbf{l} \in \tilde{\mathcal{S}}} \mathbb{E} \left\{ \sum_{i \in \mathcal{S}_1} \pi_i \tilde{f}_i(\mathbf{N}) + \sum_{j=2}^G \tilde{\pi}_j \tilde{f}_{l_j}(\mathbf{N}) \right\} \quad (3.19)$$

for any additive noise p.d.f. $p_{\mathbf{N}}(\cdot)$, where

$$\tilde{f}_i(\mathbf{n}) \triangleq \begin{cases} f_i(\mathbf{n}) , & \text{if } w_i(\theta) \text{ is known} \\ \sup_{\theta \in \Lambda_i} \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{n}) d\mathbf{y} , & \text{if } w_i(\theta) \text{ is unknown} \end{cases} . \quad (3.20)$$

Proof: The conditional risk in (3.7) can be expressed as

$$R_{\theta}(\phi) = \int_{\Gamma} \int_{\mathbb{R}^K} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{n}) p_{\mathbf{N}}(\mathbf{n}) d\mathbf{n} d\mathbf{y} ,$$

which is equal to

$$R_{\theta}(\phi) = \mathbb{E} \left\{ \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{N}) d\mathbf{y} \right\} .$$

Based on this expression, $R_i(\phi)$ in (3.17) becomes equal to

$$R_i(\phi) = \begin{cases} \mathbb{E}\{f_i(\mathbf{N})\} , & \text{if } w_i(\theta) \text{ is known} \\ \sup_{\theta \in \Lambda_i} \mathbb{E} \left\{ \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{N}) d\mathbf{y} \right\} , & \text{if } w_i(\theta) \text{ is unknown} \end{cases} , \quad (3.21)$$

where $f_i(\mathbf{N})$ is as in (3.9). When the expression in (3.21) is inserted into (3.3), and the fact that

$$\sup_{\theta \in \Lambda_i} \mathbb{E} \left\{ \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{N}) d\mathbf{y} \right\} \leq \mathbb{E} \left\{ \sup_{\theta \in \Lambda_i} \int_{\Gamma} C[\phi(\mathbf{y}), \theta] p_{\theta}^X(\mathbf{y} - \mathbf{N}) d\mathbf{y} \right\} \quad (3.22)$$

is employed, it can be shown that $r_1(\phi)$ is upper bounded as in (3.18) and (3.20). Similarly, the expression in (3.13) can be manipulated to obtain the upper bound specified by (3.19) and (3.20). \square

Note that when all the $w_i(\theta)$'s are known, the terms on the right-hand-sides of (3.18) and (3.19) reduce to the objective functions in the minimization problems in (3.10) and (3.14), respectively. Therefore, they become equal to $r_1(\phi)$ and $r_2(\phi)$, respectively (since $p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg \min_{p_{\mathbf{N}}(\mathbf{n})} r_1(\phi)$ in (3.10) and $p_{\mathbf{N}}^{\text{opt}}(\mathbf{n}) = \arg \min_{p_{\mathbf{N}}(\mathbf{n})} r_2(\phi)$ in (3.14) by definition); hence the upper bounds in Lemma 1 are achieved. Also, in the absence of additive noise (that is, $p_{\mathbf{N}}(\mathbf{n}) = \delta(\mathbf{n})$ and $\mathbf{Y} = \mathbf{X}$), (3.3), (3.4), (3.20) and (3.21) can be used to show that the upper bounds in (3.18) and (3.19) are achieved again. Specifically, in the absence of noise, the expectation operators are removed and $\tilde{f}_i(\mathbf{N})$ terms are replaced by $\tilde{f}_i(\mathbf{0})$ terms for the upper bounds in (3.18) and (3.19). Also, $R_i(\phi)$ in (3.21) becomes equal to $\tilde{f}_i(\mathbf{0})$ in the absence of noise (see (3.20)). Therefore, the definitions of $r_1(\phi)$ in (3.3) and $r_2(\phi)$ in (3.4) can be used to show that the upper bounds are achieved in this scenario. In addition, it can be shown that *any additive noise component that improves (i.e., reduces) the upper bounds on $r_1(\phi)$ or $r_2(\phi)$ with respect to the case without additive noise also improves the detector performance over the noiseless case according to Criterion 1 or Criterion 2, respectively*. In order to verify this claim, let $r_1^X(\phi)$ and $r_2^X(\phi)$ denote, respectively, the performance metrics $r_1(\phi)$ and $r_2(\phi)$ when no additive noise is employed. As stated before, the upper bounds are achieved in the absence of additive noise (that is, $r_1^X(\phi)$ and $r_2^X(\phi)$ are equal to the corresponding upper bounds in the absence of additive noise). Next, suppose that noise with p.d.f. $p_{\mathbf{N}}^{(1)}(\mathbf{n})$ or $p_{\mathbf{N}}^{(2)}(\mathbf{n})$ is added to the original observation \mathbf{x} , which results in a reduction of the corresponding upper bound; that is, the upper bounds become strictly less than $r_1^X(\phi)$ and $r_2^X(\phi)$, respectively. On the other hand, since $r_1(\phi)$ and $r_2(\phi)$ are always smaller than or equal to the specified upper bounds due to Lemma 1, they also become strictly less than $r_1^X(\phi)$ and $r_2^X(\phi)$, respectively. Hence, the detector performance is improved via additive noise specified by $p_{\mathbf{N}}^{(1)}(\mathbf{n})$ and $p_{\mathbf{N}}^{(2)}(\mathbf{n})$ according to Criterion 1 and Criterion 2, respectively, relative to the case without additive noise. Therefore, if an additive noise component reduces the upper

bound in (3.18) (in (3.19)) compared to the case without additive noise, it also improves the detection performance according to Criterion 1 (Criterion 2) over the noiseless case.

The additive noise components that minimize the upper bounds in (3.18) and (3.19) can be represented by discrete probability distributions as specified by (3.11) and Proposition 2 since the upper bounds are in the same form as the objective functions in the minimization problems in (3.10) and (3.14). Specifically, the p.d.f. that minimizes the upper bound on $r_1(\phi)$ can be represented by a constant signal value, and the p.d.f. that minimizes the upper bound on $r_2(\phi)$ can be represented by a randomization of no more than $|\tilde{\mathcal{S}}|$ different signal values. It should also be noted that although these additive noise p.d.f.s minimize the upper bounds in Lemma 1, they may not be the optimal additive noise p.d.f.s for the original problem in general. The optimal solution needs to be calculated based on some p.d.f. approximations as discussed before. However, the approach based on Lemma 1 can still be useful to obtain certain improvability conditions and to achieve performance improvements with low complexity solutions in some cases.

3.5 A Detection Example and Conclusions

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

$$\begin{aligned}
 \mathcal{H}_0 &: x = -3\sqrt{A} + v , \\
 \mathcal{H}_1 &: x = -\sqrt{A} + v , \\
 \mathcal{H}_2 &: x = \sqrt{A} + v , \\
 \mathcal{H}_3 &: x = 3\sqrt{A} + v ,
 \end{aligned} \tag{3.23}$$

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

$$p_V(x) = \sum_{i=1}^M w_i \psi_i(x - \mu_i) , \quad (3.24)$$

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

$$\phi(y) = \begin{cases} 0 , & y \leq -2\sqrt{A} \\ 1 , & -2\sqrt{A} < y \leq 0 \\ 2 , & 0 < y \leq 2\sqrt{A} \\ 3 , & 2\sqrt{A} < y \end{cases} , \quad (3.25)$$

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

The hypothesis-testing problem in (3.23) is the form of pulse amplitude modulation (PAM); that is, the information is carried in the signal amplitude. The Gaussian mixture noise specified above can be encountered in PAM communications systems in the presence of interference or jamming [88]. In the following example, four different amplitudes corresponding four different underlying hypotheses are transmitted using the PAM technique above over such a communication environment. It is assumed that only the prior probability of \mathcal{H}_1 , π_1 , is known. Such a scenario can be encountered in practice when previous measurements can successfully discriminate between the underlying hypotheses for \mathcal{H}_1 and the other hypotheses (\mathcal{H}_0 , \mathcal{H}_2 and \mathcal{H}_3), whereas it is difficult to specify reliably which of the underlying hypotheses for \mathcal{H}_0 , \mathcal{H}_2 and \mathcal{H}_3 is actually true. For instance, if we assume four fish species with three of them (corresponding to \mathcal{H}_0 , \mathcal{H}_2 and \mathcal{H}_3) having similar characteristics, we cannot assume a known prior

for each of those species (as we do not have reliable information from measurements); however, we can regard $\pi_0 + \pi_2 + \pi_3$ (equivalently, π_1) as a known value, since these three fish species can be distinguished easily from the other one.²

Since only the prior probability of \mathcal{H}_1 is known, there are two groups ($G = 2$), $\mathcal{S}_1 = \{1\}$ and $\mathcal{S}_2 = \{0, 2, 3\}$ (see (3.3)-(3.4)). Also, UCA is assumed in the following calculations. Based on the expressions in (3.9), (3.10) and (3.14), $f(\mathbf{n})$ and $f_{\mathbf{l}}(\mathbf{n})$ can be obtained, and the optimization problems in (3.11) and (3.16) can be solved. Specifically, $f(\mathbf{n})$ in (3.10) can be calculated as

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

for $\mathbf{n} = n \in \mathbb{R}$, and similarly $f_{\mathbf{l}}(\mathbf{n})$ in (3.14) becomes

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

for $\mathbf{l} = l_2 \in \mathcal{S}_2$, where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$ denotes the Q -function, $c_2 = c_3 = 1$, $c_0 = -1$, $m_0 = m_3 = 0$, and $m_2 = 1$. For the simulation results, symmetric Gaussian mixture noise with $M = 6$ is considered, where the mean values of the Gaussian components in the mixture noise in (3.24) are specified as $\pm[0.01 \ 0.7 \ 1.1]$ with corresponding weights of $[0.35 \ 0.1 \ 0.05]$. In addition, the variances of the Gaussian components in the mixture noise are assumed to be the same; i.e., $\sigma_i = \sigma$ for $i = 1, \dots, M$.

Figure 3.2 illustrates the Bayes risks for the modified and original detectors for various values of σ when $A = 1$ and $\pi_1 = 0.25$. From the figure, it is observed that the use of additive noise can significantly improve the performance according to both criteria. Also, as σ increases the improvement ratio decreases, and after

²Consider a scenario in which a device measures some parameters of the fish (such as their length or color), and this information is transmitted to a data processing center using PAM.

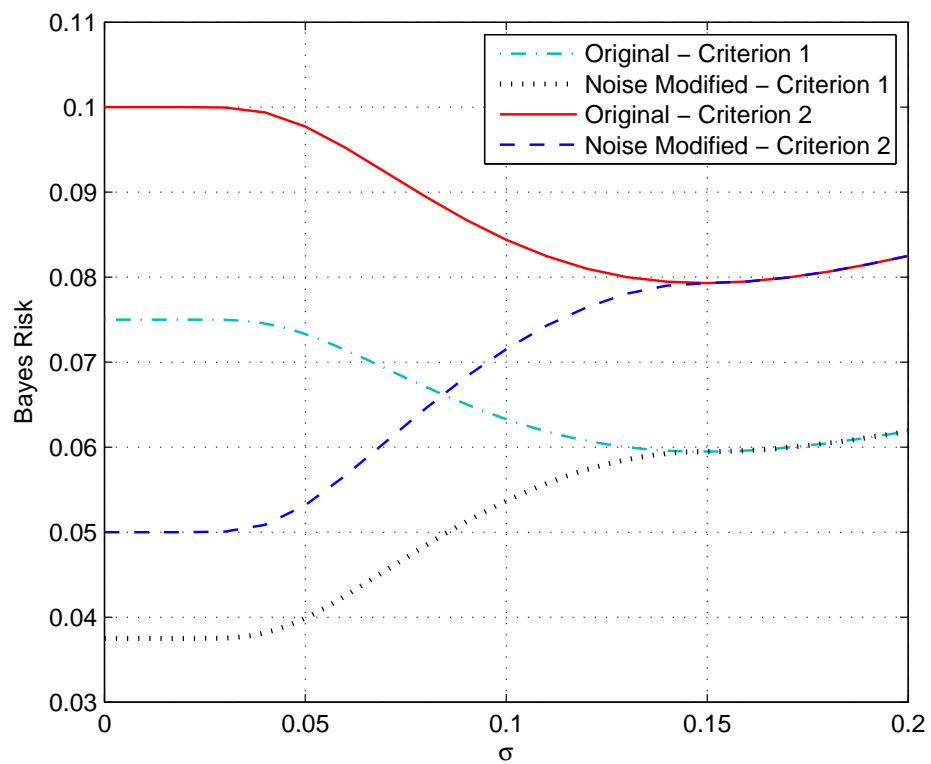


Figure 3.2: Bayes risks of the original and noise modified detectors versus σ for $A = 1$ according to both criteria.

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

In order to investigate the result in Proposition 2, Table 3.1 shows the optimal noise p.d.f.s for various values of σ according to Criterion 2. In accordance with the proposition, the optimal noise p.d.f.s are expressed as randomization of three or fewer mass points.

3.6 Appendices

3.6.1 Proof of Proposition 1

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

$$\boldsymbol{\epsilon}_*^T \mathbf{H} \boldsymbol{\epsilon}_* + 2\boldsymbol{\epsilon}_*^T \mathbf{f} < 0 . \quad (3.26)$$

Let $\boldsymbol{\epsilon}_* = \rho_* \mathbf{z}$, where ρ_* is an infinitesimally small real number, and \mathbf{z} is a K -dimensional real vector. Then, (3.26) can be simplified, after some manipulation,

Table 3.1: Optimal additive noise p.d.f., $p_N(n) = \lambda_1\delta(n - n_1) + \lambda_2\delta(n - n_2) + \lambda_3\delta(n - n_3)$, according to Criterion 2.

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

as

$$\mathbf{z}^T \mathbf{H} \mathbf{z} + \frac{2}{\rho_*} \mathbf{z}^T \mathbf{f} < 0 . \quad (3.27)$$

For the first part of the proposition, if $\mathbf{f} \neq \mathbf{0}$, then ρ_* and \mathbf{z} satisfying (3.27) can always be found. For the second part of the proposition, if $f(\mathbf{x})$ is strictly concave at $\mathbf{x} = \mathbf{0}$, which means that \mathbf{H} is negative definite, then ρ_* and \mathbf{z} satisfying (3.27) always exist. \square

Chapter 4

Noise Enhanced Binary

Composite Hypothesis-Testing in the Neyman-Pearson Framework

This chapter is organized as follows. Section 4.1 describes the composite hypothesis-testing problem, and introduces the detection criteria. Then, Section 4.2 and Section 4.3 study the effects of additive noise according to the max-sum and the max-min criteria, respectively. In Section 4.4, the results in the previous sections are extended to the max-max case, and the main implications are briefly summarized. A detection example is provided in Section 4.5, which is followed by the concluding remarks.

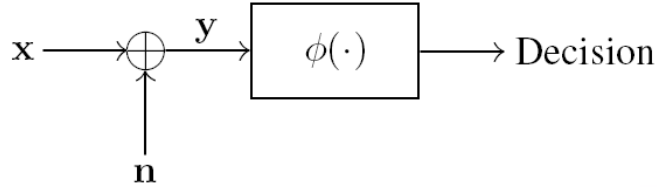


Figure 4.1: Independent noise \mathbf{n} is added to data vector \mathbf{x} in order to improve the performance of the detector, $\phi(\cdot)$.

4.1 Problem Formulation and Motivation

Consider a composite binary hypothesis-testing problem described as

$$\begin{aligned} \mathcal{H}_0 &: p_{\theta_0}(\mathbf{x}) \quad , \quad \theta_0 \in \Lambda_0 \\ \mathcal{H}_1 &: p_{\theta_1}(\mathbf{x}) \quad , \quad \theta_1 \in \Lambda_1 \end{aligned} \quad (4.1)$$

where \mathcal{H}_i denotes the i th hypothesis for $i = 0, 1$. Under hypothesis \mathcal{H}_i , data (observation) $\mathbf{x} \in \mathbb{R}^K$ has a p.d.f. indexed by $\theta_i \in \Lambda_i$, namely, $p_{\theta_i}(\mathbf{x})$, where Λ_i is the set of possible parameter values under hypothesis \mathcal{H}_i . Parameter sets Λ_0 and Λ_1 are disjoint, and their union forms the parameter space, $\Lambda = \Lambda_0 \cup \Lambda_1$ [40]. In addition, it is assumed that the probability distributions of the parameters are not known *a priori*.

The expressions in (4.1) present a generic formulation of a binary composite hypothesis-testing problem. Such problems are encountered in various scenarios, such as in radar systems and non-coherent communications receivers [40], [100]. In the case that both Λ_0 and Λ_1 consist of single elements, the problem in (4.1) reduces to a *simple* hypothesis-testing problem [40].

A generic detector (decision rule), denoted by $\phi(\mathbf{x})$, is considered, which maps the data vector into a real number in $[0, 1]$ that represents the probability of selecting \mathcal{H}_1 [40]. The aim is to investigate the stochastic resonance (SR) phenomenon by analyzing the effects of additive independent noise to the original data, \mathbf{x} , of a given detector, as shown in Figure 4.1, where \mathbf{y} represents the

modified data vector given by

$$\mathbf{y} = \mathbf{x} + \mathbf{n} , \quad (4.2)$$

with \mathbf{n} denoting the additive noise term that is independent of \mathbf{x} .

The Neyman-Pearson framework is considered in this study, and performance of a detector is specified by its probabilities of detection and false-alarm [40], [41], [68]. Since the additive noise is independent of the data, the probabilities of detection and false-alarm can be expressed, conditioned on θ_1 and θ_0 , respectively, as

$$P_D^{\mathbf{y}}(\theta_1) = \int_{\mathbb{R}^K} \phi(\mathbf{y}) \left[\int_{\mathbb{R}^K} p_{\theta_1}(\mathbf{y} - \mathbf{x}) p_{\mathbf{n}}(\mathbf{x}) d\mathbf{x} \right] d\mathbf{y} , \quad (4.3)$$

$$P_F^{\mathbf{y}}(\theta_0) = \int_{\mathbb{R}^K} \phi(\mathbf{y}) \left[\int_{\mathbb{R}^K} p_{\theta_0}(\mathbf{y} - \mathbf{x}) p_{\mathbf{n}}(\mathbf{x}) d\mathbf{x} \right] d\mathbf{y} , \quad (4.4)$$

where $p_{\mathbf{n}}(\cdot)$ denotes the p.d.f. of the additive noise. After some manipulation, (4.3) and (4.4) can be expressed as [12]

$$P_D^{\mathbf{y}}(\theta_1) = E_{\mathbf{n}}\{F_{\theta_1}(\mathbf{n})\} , \quad (4.5)$$

$$P_F^{\mathbf{y}}(\theta_0) = E_{\mathbf{n}}\{G_{\theta_0}(\mathbf{n})\} , \quad (4.6)$$

for $\theta_1 \in \Lambda_1$ and $\theta_0 \in \Lambda_0$, where

$$F_{\theta_1}(\mathbf{n}) \triangleq \int_{\mathbb{R}^K} \phi(\mathbf{y}) p_{\theta_1}(\mathbf{y} - \mathbf{n}) d\mathbf{y} , \quad (4.7)$$

$$G_{\theta_0}(\mathbf{n}) \triangleq \int_{\mathbb{R}^K} \phi(\mathbf{y}) p_{\theta_0}(\mathbf{y} - \mathbf{n}) d\mathbf{y} . \quad (4.8)$$

Note that $F_{\theta_1}(\mathbf{n})$ and $G_{\theta_0}(\mathbf{n})$ define, respectively, the probability of detection conditioned on θ_1 and the probability of false alarm conditioned on θ_0 when a constant noise \mathbf{n} is added to the data. Also, in the absence of additive noise, i.e., for $\mathbf{n} = \mathbf{0}$, the probabilities of detection and false-alarm are given by $P_D^{\mathbf{x}}(\theta_1) = F_{\theta_1}(\mathbf{0})$ and $P_F^{\mathbf{x}}(\theta_0) = G_{\theta_0}(\mathbf{0})$, respectively, for given values of the parameters.

Various performance metrics can be defined for composite hypothesis-testing problems [40], [41]. In the Neyman-Pearson framework, the main constraint is

to keep the probability of false-alarm below a certain threshold for all possible parameter values θ_0 ; i.e.,

$$\max_{\theta_0 \in \Lambda_0} P_F^y(\theta_0) \leq \tilde{\alpha} . \quad (4.9)$$

In most practical cases, the detectors are designed in such a way that they operate at the maximum allowed false-alarm probability $\tilde{\alpha}$ in order to obtain maximum detection probabilities. Therefore, the constraint on the false-alarm probability can be defined as $\tilde{\alpha} = \max_{\theta_0 \in \Lambda_0} P_F^x(\theta_0) = \max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{0})$ for practical scenarios. In other words, in the absence of additive noise \mathbf{n} , the detectors commonly operate at the false-alarm probability limit.

Under the constraint in (4.9), the aim is to maximize a function of the detection probabilities for possible parameter values $\theta_1 \in \Lambda_1$. In this study, the following performance criteria are considered [41]:

- **Max-sum criterion:** In this case, the aim is to maximize $\int_{\theta_1 \in \Lambda_1} P_D^y(\theta_1) d\theta_1$, which can be regarded as the “sum” of the detection probabilities for different θ_1 values. This is equivalent to assuming uniform distribution for θ_1 and maximizing the average detection probability [41].
- **Max-min criterion:** According to this criterion, the aim is to maximize the worst-case detection probability, defined as $\min_{\theta_1 \in \Lambda_1} P_D^y(\theta_1)$ [41], [68], [69]. The worst-case detection probability corresponds to considering the *least-favorable* distribution for θ_1 [41].
- **Max-max criterion:** This criterion maximizes the best-case detection probability, $\max_{\theta_1 \in \Lambda_1} P_D^y(\theta_1)$. This criterion is not very common in practice, since maximizing the detection probability for a single parameter can result in very low detection probabilities for the other parameters. Therefore, this criterion will only be briefly analyzed in Section 4.4 for completeness of the theoretical results.

There are two main motivations for investigating the effects of additive independent noise in (4.2) for binary composite hypothesis-testing problems. First, it is important to quantify performance improvements that can be achieved via additive noise, and to determine when additive noise can improve detection performance. In other words, theoretical investigation of SR in binary composite hypothesis-testing problems is of interest. Second, in many cases, the optimal detector based on the calculation of likelihood functions is difficult to obtain or requires intense computations [12], [40], [68]. Therefore, a suboptimal detector can be preferable in some practical scenarios. However, the performance of a suboptimal detector may need to be enhanced in order to meet certain system requirements. One way to enhance the performance of a suboptimal detector without changing the detector structure is to modify its original data as in Figure 4.1 [12]. Even though calculation of optimal additive noise causes a complexity increase for the suboptimal detector, the overall computational complexity is still considerably lower than that of an optimal detector based on likelihood function calculations. This is because the optimal detector needs to perform intense calculations for each decision whereas the suboptimal detector with modified data needs to update the optimal additive noise whenever the statistics of the hypotheses change. For instance, in a binary communications system, the optimal detector needs to calculate the likelihood ratio for each symbol, whereas a suboptimal detector as in Figure 4.1 needs to update \mathbf{n} only when the channel statistics change, which can be constant over a large number of symbols for slowly varying channels [101].

4.2 Max-Sum Criterion

In this section, the aim is to determine the optimal additive noise \mathbf{n} in (4.2) that solves the following optimization problem.

$$\max_{p_{\mathbf{n}}(\cdot)} \int_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1) d\theta_1 \quad (4.10)$$

$$\text{subject to } \max_{\theta_0 \in \Lambda_0} P_F^{\mathbf{y}}(\theta_0) \leq \tilde{\alpha} \quad (4.11)$$

where $P_D^{\mathbf{y}}(\theta_1)$ and $P_F^{\mathbf{y}}(\theta_0)$ are as in (4.5)-(4.8). Note that the problem in (4.10) and (4.11) can also be regarded as a *max-mean* problem since the objective function in (4.10) can be normalized appropriately so that it defines the average detection probability assuming that all θ_1 parameters are equally likely [41].¹

From (4.5) and (4.6), the optimization problem in (4.10) and (4.11) can also be expressed as

$$\max_{p_{\mathbf{n}}(\cdot)} E_{\mathbf{n}}\{F(\mathbf{n})\} \quad (4.12)$$

$$\text{subject to } \max_{\theta_0 \in \Lambda_0} E_{\mathbf{n}}\{G_{\theta_0}(\mathbf{n})\} \leq \tilde{\alpha} \quad (4.13)$$

where $F(\mathbf{n})$ is defined by

$$F(\mathbf{n}) \triangleq \int_{\theta_1 \in \Lambda_1} F_{\theta_1}(\mathbf{n}) d\theta_1 . \quad (4.14)$$

Note that $F(\mathbf{n})$ defines the total detection probability for a specific value of additive noise \mathbf{n} .

In the following sections, the effects of additive noise are investigated for this max-sum problem, and various results related to optimal solutions are presented.

¹When Λ_1 does not have a finite volume, the max-mean formulation should be used since $\int_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1) d\theta_1$ may not be finite.

4.2.1 Improvability and Non-improvability Conditions

According to the max-sum criterion, the detector is called *improvable* if there exists additive independent noise \mathbf{n} that satisfies

$$P_{D,\text{sum}}^y \triangleq \int_{\theta_1 \in \Lambda_1} P_D^y(\theta_1) d\theta_1 > \int_{\theta_1 \in \Lambda_1} P_D^x(\theta_1) d\theta_1 \triangleq P_{D,\text{sum}}^x \quad (4.15)$$

under the false-alarm constraint. From (4.5) and (4.14), the condition in (4.15) can also be expressed as

$$P_{D,\text{sum}}^y = E_{\mathbf{n}}\{F(\mathbf{n})\} > F(\mathbf{0}) = P_{D,\text{sum}}^x . \quad (4.16)$$

If the detector cannot be improved, it is called *non-improvable*.

In order to determine the improvability of a detector according to the max-sum criterion without actually solving the optimization problem in (4.12) and (4.13), the approach in [12] for simple hypothesis-testing problems can be extended to composite hypothesis-testing problems in the following manner. First, we introduce the following function

$$H(t) \triangleq \sup \left\{ F(\mathbf{n}) \mid \max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}) = t, \mathbf{n} \in \mathbb{R}^K \right\} , \quad (4.17)$$

which defines the maximum value of the total detection probability for a given value of the maximum false-alarm probability. In other words, among all constant noise components \mathbf{n} that achieve a maximum false-alarm probability of t , $H(t)$ defines the maximum probability of detection.

From (4.17), it is observed that if there exists $t_0 \leq \tilde{\alpha}$ such that $H(t_0) > P_{D,\text{sum}}^x$, then the system is improvable, since under such a condition there exists a noise component \mathbf{n}_0 such that $F(\mathbf{n}_0) > P_{D,\text{sum}}^x$ and $\max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}_0) \leq \tilde{\alpha}$. Hence, the detector performance can be improved by using an additive noise with $p_{\mathbf{n}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{n}_0)$. However, that condition may not hold in many practical scenarios since, for constant additive noise values, larger total detection probabilities than $P_{D,\text{sum}}^x$ are commonly accompanied by false-alarm probabilities that exceed the

false-alarm limit. Therefore, a more generic improvability condition is derived in the following theorem.

Theorem 1: *Define the maximum false-alarm probability in the absence of additive noise as $\alpha \triangleq \max_{\theta_0 \in \Lambda_0} P_F^x(\theta_0)$. If $H(t)$ in (4.17) is second-order continuously differentiable around $t = \alpha$ and satisfies $H''(\alpha) > 0$, then the detector is improvable.*

Proof: Since $H''(\alpha) > 0$ and $H(t)$ in (4.17) is second-order continuously differentiable around $t = \alpha$, there exist $\epsilon > 0$, \mathbf{n}_1 and \mathbf{n}_2 such that $\max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}_1) = \alpha + \epsilon$ and $\max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}_2) = \alpha - \epsilon$. Then, it is proven in the following that an additive noise with $p_{\mathbf{n}}(\mathbf{x}) = 0.5 \delta(\mathbf{x} - \mathbf{n}_1) + 0.5 \delta(\mathbf{x} - \mathbf{n}_2)$ improves the detection performance under the false-alarm constraint. First, the maximum false-alarm probability in the presence of additive noise is shown not to exceed α .

$$\max_{\theta_0 \in \Lambda_0} E_{\mathbf{n}}\{G_{\theta_0}(\mathbf{n})\} \leq E_{\mathbf{n}} \left\{ \max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}) \right\} = 0.5(\alpha + \epsilon) + 0.5(\alpha - \epsilon) = \alpha. \quad (4.18)$$

Then, the increase in the detection probability is proven as follows. Due to the assumptions in the theorem, $H(t)$ is convex in an interval around $t = \alpha$. Since $E_{\mathbf{n}}\{F(\mathbf{n})\}$ can attain the value of $0.5 H(\alpha + \epsilon) + 0.5 H(\alpha - \epsilon)$, which is always larger than $H(\alpha)$ due to convexity, it is concluded that $E_{\mathbf{n}}\{F(\mathbf{n})\} > H(\alpha)$. As $H(\alpha) \geq P_{D,\text{sum}}^x$ by definition of $H(t)$ in (4.17), $E_{\mathbf{n}}\{F(\mathbf{n})\} > P_{D,\text{sum}}^x$ is satisfied; hence, the detector is improvable. \square

Theorem 1 provides a simple condition that guarantees the improvability of a detector according to the max-sum criterion. Note that $H(t)$ is always a single-variable function irrespective of the dimension of the data vector, which facilitates simple evaluations of the conditions in the theorem. However, the main complexity may come into play in obtaining an expression for $H(t)$ in (4.17) in certain scenarios. An example is presented to in Section 4.5 to illustrate the use of Theorem 1.

In addition to the improbability conditions in Theorem 1, sufficient conditions for non-improvability can be obtained by defining the following function.

$$J_{\theta_0}(t) \triangleq \sup \{F(\mathbf{n}) \mid G_{\theta_0}(\mathbf{n}) = t, \mathbf{n} \in \mathbb{R}^K\} . \quad (4.19)$$

This function is similar to that in [12], but it is defined for each $\theta_0 \in \Lambda_0$ here, since a composite hypothesis-testing problem is considered. Therefore, Theorem 2 in [12] can be extended in the following manner.

Theorem 2: *If there exists $\theta_0 \in \Lambda_0$ and a nondecreasing concave function $\Psi(t)$ such that $\Psi(t) \geq J_{\theta_0}(t) \forall t$ and $\Psi(\tilde{\alpha}) = P_{\mathbf{D},\text{sum}}^{\mathbf{x}}$, then the detector is non-improvable.*

Proof: For the θ_0 value in the theorem, the objective function in (4.12) can be expressed as

$$E_{\mathbf{n}}\{F(\mathbf{n})\} = \int p_{\mathbf{n}}(\mathbf{x})F(\mathbf{x}) d\mathbf{x} \leq \int p_{\mathbf{n}}(\mathbf{x})J_{\theta_0}(G_{\theta_0}(\mathbf{x})) d\mathbf{x} , \quad (4.20)$$

where the inequality is obtained by the definition in (4.19).

Since $\Psi(t)$ satisfies $\Psi(t) \geq J_{\theta_0}(t) \forall t$, and is concave, (4.20) becomes

$$E_{\mathbf{n}}\{F(\mathbf{n})\} \leq \int p_{\mathbf{n}}(\mathbf{x})\Psi(G_{\theta_0}(\mathbf{x})) d\mathbf{x} \leq \Psi \left(\int p_{\mathbf{n}}(\mathbf{x})G_{\theta_0}(\mathbf{x}) d\mathbf{x} \right) . \quad (4.21)$$

Finally, the nondecreasing property of $\Psi(t)$ together with $\int p_{\mathbf{n}}(\mathbf{x})G_{\theta_0}(\mathbf{x}) d\mathbf{x} \leq \tilde{\alpha}$ implies that $E_{\mathbf{n}}\{F(\mathbf{n})\} \leq \Psi(\tilde{\alpha})$. Since $\Psi(\tilde{\alpha}) = P_{\mathbf{D},\text{sum}}^{\mathbf{x}}$, $E_{\mathbf{n}}\{F(\mathbf{n})\} \leq P_{\mathbf{D},\text{sum}}^{\mathbf{x}}$ is obtained for any additive noise \mathbf{n} . Hence, the detector is non-improvable. \square

The conditions in Theorem 2 can be used to determine that the detector performance cannot be improved via additive noise, which prevents efforts for solving the optimization problem in (4.10) and (4.11).² However, it should also be noted that the detector can still be non-improvable although the conditions in the theorem are not satisfied; that is, Theorem 2 does not provide necessary conditions for non-improvability.

²The optimization problem yields $p_{\mathbf{n}}(\mathbf{x}) = \delta(\mathbf{x})$ when the detector is non-improvable.

4.2.2 Characterization of Optimal Solution

In this section, the statistical characterization of optimal additive noise components is provided. First, the maximum false-alarm probabilities of optimal solutions are specified. Then, the structures of the optimal noise p.d.f.s are investigated.

In order to investigate the false-alarm probabilities of the optimal solution obtained from (4.10) and (4.11) without actually solving the optimization problem, $H(t)$ in (4.17) can be utilized. Let F_{\max} represent the maximum value of $H(t)$, i.e., $F_{\max} = \max_t H(t)$. Assume that this maximum is attained at $t = t_m$.³ Then, one immediate observation is that if t_m is smaller than or equal to the false-alarm limit, i.e., $t_m \leq \tilde{\alpha}$, then the noise component \mathbf{n}_m that results in $\max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}_m) = t_m$ is the optimal noise component; i.e., $p_{\mathbf{n}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{n}_m)$. However, in many practical scenarios, the maximum of $H(t)$ is attained for $t_m > \tilde{\alpha}$, since larger detection probabilities can be achieved for larger false-alarm probabilities. In such cases, the following theorem specifies the false-alarm probability achieved by the optimal solution.

Theorem 3: *If $t_m > \tilde{\alpha}$, then the optimal solution of (4.10) and (4.11) satisfies $\max_{\theta_0 \in \Lambda_0} P_{\mathbb{F}}^{\mathcal{Y}}(\theta_0) = \tilde{\alpha}$.*

Proof: Assume that the optimal solution to (4.10) and (4.11) is given by $p_{\mathbf{n}}(\mathbf{x})$ with $\beta \triangleq \max_{\theta_0 \in \Lambda_0} P_{\mathbb{F}}^{\mathcal{Y}}(\theta_0) < \tilde{\alpha}$. Define another noise \mathbf{n} with the following p.d.f.:

$$p_{\mathbf{n}}(\mathbf{x}) = \frac{\tilde{\alpha} - \beta}{t_m - \beta} \delta(\mathbf{x} - \mathbf{n}_m) + \frac{t_m - \tilde{\alpha}}{t_m - \beta} p_{\mathbf{n}}(\mathbf{x}), \quad (4.22)$$

where \mathbf{n}_m is the noise component that results in the maximum total detection probability; that is, $F(\mathbf{n}_m) = F_{\max}$, and t_m is the maximum false-alarm probability when noise \mathbf{n}_m is employed; i.e., $t_m = \max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}_m)$.

³If there are multiple t values that result in the maximum value F_{\max} , then the minimum of those values is selected.

For the noise p.d.f. in (4.22), the false-alarm and detection probabilities can be obtained as

$$P_{D,\text{sum}}^{\mathbf{y}} = E_{\mathbf{n}}\{F(\mathbf{n})\} = \frac{\tilde{\alpha} - \beta}{t_m - \beta} F(\mathbf{n}_m) + \frac{t_m - \tilde{\alpha}}{t_m - \beta} P_{D,\text{sum}}^{\tilde{\mathbf{y}}} , \quad (4.23)$$

$$P_{F}^{\tilde{\mathbf{y}}}(\theta_0) = E_{\mathbf{n}}\{G_{\theta_0}(\mathbf{n})\} = \frac{\tilde{\alpha} - \beta}{t_m - \beta} G_{\theta_0}(\mathbf{n}_m) + \frac{t_m - \tilde{\alpha}}{t_m - \beta} P_{F}^{\tilde{\mathbf{y}}}(\theta_0) , \quad (4.24)$$

for all $\theta_0 \in \Lambda_0$. Since $F(\mathbf{n}_m) > P_{D,\text{sum}}^{\tilde{\mathbf{y}}}$, (4.23) implies $P_{D,\text{sum}}^{\mathbf{y}} > P_{D,\text{sum}}^{\tilde{\mathbf{y}}}$. On the other hand, as $G_{\theta_0}(\mathbf{n}_m) \leq t_m$ and $P_{F}^{\tilde{\mathbf{y}}}(\theta_0) \leq \beta$, $P_{F}^{\tilde{\mathbf{y}}}(\theta_0) \leq \tilde{\alpha}$ is obtained. Therefore, $\tilde{\mathbf{n}}$ cannot be an optimal solution, which indicates a contradiction. In other words, any noise p.d.f. that satisfies $\max_{\theta_0 \in \Lambda_0} P_{F}^{\tilde{\mathbf{y}}}(\theta_0) < \tilde{\alpha}$ cannot be optimal. \square

The main implication of Theorem 3 is that, in most practical scenarios, the false-alarm probabilities are set to the maximum false-alarm probability limit; i.e., $\max_{\theta_0 \in \Lambda_0} P_{F}^{\mathbf{y}}(\theta_0) = \tilde{\alpha}$, in order to optimize the detection performance according to the max-sum criterion.

Another important characterization of the optimal noise involves the specification of the optimal noise p.d.f.. In [12] and [13], it is shown for simple hypothesis-testing problems that an optimal noise p.d.f., if exists, can be represented by a randomization of at most 2 discrete signals. In general, the optimal noise specified by (4.10) and (4.11) for the composite hypothesis-testing problem can have more than 2 mass points. The following theorem specifies the structure of the optimal noise p.d.f. under certain conditions.

Theorem 4: *Let $\theta_0 \in \Lambda_0 = \{\theta_{01}, \theta_{02}, \dots, \theta_{0M}\}$. Assume that the additive noise components can take finite values specified by $n_i \in [a_i, b_i]$, $i = 1, \dots, K$, for any finite a_i and b_i . Define set U as*

$$U = \left\{ (u_0, u_1, \dots, u_M) : u_0 = F(\mathbf{n}), u_1 = G_{\theta_{01}}(\mathbf{n}), \dots, u_M = G_{\theta_{0M}}(\mathbf{n}), \right. \\ \left. \text{for } \mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b} \right\} , \quad (4.25)$$

where $\mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b}$ means that $n_i \in [a_i, b_i]$ for $i = 1, \dots, K$. If U is a closed subset of \mathbb{R}^{M+1} , an optimal solution to (4.10) and (4.11) has the following form

$$p_{\mathbf{n}}(\mathbf{x}) = \sum_{i=1}^{M+1} \lambda_i \delta(\mathbf{x} - \mathbf{n}_i), \quad (4.26)$$

where $\sum_{i=1}^{M+1} \lambda_i = 1$ and $\lambda_i \geq 0$ for $i = 1, 2, \dots, M+1$.

Proof: The proof extends the results in [12] and [13] for the two mass point probability distributions to the $(M+1)$ mass point ones. Since the possible additive noise components are specified by $n_i \in [a_i, b_i]$ for $i = 1, \dots, K$, U in (4.25) represents the set of all possible combinations of $F(\mathbf{n})$ and $G_{\theta_{0i}}(\mathbf{n})$ for $i = 1, \dots, M$. Let the convex hull of U be denoted by set V . Since $F(\mathbf{n})$ and $G_{\theta_{0i}}(\mathbf{n})$ are bounded by definition, U is a bounded and closed subset of \mathbb{R}^{M+1} by the assumption in the theorem. Therefore, U is compact, and the convex hull V of U is closed [84]. In addition, since $V \subseteq \mathbb{R}^{M+1}$, the dimension of V is smaller than or equal to $(M+1)$.

Define W as the set of all possible total detection and false-alarm probabilities; that is,

$$W = \{(w_0, w_1, \dots, w_M) : w_0 = \mathbf{E}_{\mathbf{n}}\{F(\mathbf{n})\}, w_1 = \mathbf{E}_{\mathbf{n}}\{G_{\theta_{01}}(\mathbf{n})\}, \dots, \\ w_M = \mathbf{E}_{\mathbf{n}}\{G_{\theta_{0M}}(\mathbf{n})\}, \forall p_{\mathbf{n}}(\mathbf{n}), \mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b}\}. \quad (4.27)$$

Similar to [12] and [85], it can be shown that $W = V$. Therefore, Carathéodory's theorem [86], [87] implies that any point in V (hence, in W) can be expressed as the convex combination of $(M+2)$ points in U . Since an optimal p.d.f. must maximize the total detection probability, it corresponds to the boundary of V [12]. Since V is closed, it always contains its boundary. Therefore, the optimal p.d.f. can be expressed as the convex combination of $(M+1)$ elements in U . \square

In other words, for composite hypothesis-testing problems with a finite number of possible parameter values under hypothesis \mathcal{H}_0 , the optimal p.d.f. can be expressed as a discrete p.d.f. with a finite number of mass points. Therefore,

Theorem 4 generalizes the two mass points result for simple hypothesis-testing problems [12], [13]. It should be noted that the result in Theorem 4 is valid irrespective of the number of parameters under hypothesis \mathcal{H}_1 ; that is, Λ_1 in (4.1) can be discrete or continuous. However, the theorem does not guarantee a discrete p.d.f. if the parameter space for \mathcal{H}_0 includes continuous intervals.

Regarding the first assumption in the proposition, constraining the additive noise values as $\mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b}$ is quite realistic since arbitrarily large/small values cannot be realized in practical systems. In other words, in practice, the minimum and maximum possible values of n_i define a_i and b_i , respectively. In addition, the assumption that U is a closed set guarantees the *existence* of the optimal solution [13], and it holds, for example, when F and $G_{\theta_{0j}}$ are continuous functions.

4.2.3 Calculation of Optimal Solution and Convex Relaxation

After the derivation of the improvability and non-improvability conditions, and the characterization of optimal additive noise in the previous sections, the calculation of optimal noise p.d.f.s is studied in this section.

Let $p_{\mathbf{n},f}(\cdot)$ represent the p.d.f. of $f = F(\mathbf{n})$, where $F(\mathbf{n})$ is given by (4.14). Note that $p_{\mathbf{n},f}(\cdot)$ can be obtained from the noise p.d.f., $p_{\mathbf{n}}(\cdot)$. As studied in [12], working with $p_{\mathbf{n},f}(\cdot)$ is more convenient since it results in an optimization problem in a single-dimensional space. Assume that $F(\mathbf{n})$ is a one-to-one function.⁴ Then, for a given value of noise \mathbf{n} , the false-alarm probabilities in (4.8) can be expressed as $g_{\theta_0} = G_{\theta_0}(F^{-1}(f))$, where $f = F(\mathbf{n})$. Therefore, the optimization problem in

⁴Similar to the approach in [12], the one-to-one assumption can be removed. However, it is employed in this study to obtain convenient expressions.

(4.10) and (4.11) can be stated as

$$\begin{aligned} & \max_{p_{\mathbf{n},f}(\cdot)} \int_0^\infty f p_{\mathbf{n},f}(f) df , \\ & \text{subject to } \max_{\theta_0 \in \Lambda_0} \int_0^\infty g_{\theta_0} p_{\mathbf{n},f}(f) df \leq \tilde{\alpha} . \end{aligned} \quad (4.28)$$

Note that since $p_{\mathbf{n},f}(\cdot)$ specifies a p.d.f., the optimization problem in (4.28) has also implicit constraints that $p_{\mathbf{n},f}(f) \geq 0 \forall f$ and $\int p_{\mathbf{n},f}(f) df = 1$.

In order to solve the optimization problem in (4.28), first consider the case in which the unknown parameter θ_0 under hypothesis \mathcal{H}_0 can take finitely many values specified by $\theta_0 \in \Lambda_0 = \{\theta_{01}, \theta_{02}, \dots, \theta_{0M}\}$. Then, the optimal noise p.d.f. has $(M + 1)$ mass points, under the conditions in Theorem 4. Hence, (4.28) can be expressed as

$$\begin{aligned} & \max_{\{\lambda_i, f_i\}_{i=1}^{M+1}} \sum_{i=1}^{M+1} \lambda_i f_i \\ & \text{subject to } \max_{\theta_0 \in \Lambda_0} \sum_{i=1}^{M+1} \lambda_i g_{\theta_0, i} \leq \tilde{\alpha} \\ & \sum_{i=1}^{M+1} \lambda_i = 1 \\ & \lambda_i \geq 0 , \quad i = 1, \dots, M + 1 \end{aligned} \quad (4.29)$$

where $f_i = F(\mathbf{n}_i)$, $g_{\theta_0, i} = G_{\theta_0}(F^{-1}(f_i))$, and \mathbf{n}_i and λ_i are the optimal mass points and their weights as specified in Theorem 4. Note that the optimization problem in (4.29) may not be formulated as a convex optimization problem in general since $g_{\theta_0, i} = G_{\theta_0}(F^{-1}(f_i))$ may be non-convex. Therefore, global optimization algorithms, such as particle-swarm optimization (PSO) [51]-[54], genetic algorithms and differential evolution [82], can be employed to obtain the optimal solution. In this study, the PSO approach is used since it is based on simple iterations with low computational complexity and has been successfully applied to numerous problems in various fields [94]-[97]. In Section 4.5, the PSO technique is applied to this optimization problem, which results in accurate calculation of the optimal additive noise in the specified scenario (please refer to [51]-[54] for detailed descriptions of the PSO algorithm).

Another approach to solve the optimization problem in (4.29) is to perform *convex relaxation* [55] of the problem. To that end, assume that $f = F(\mathbf{n})$ can take only finitely many known (pre-determined) values $\tilde{f}_1, \dots, \tilde{f}_{\tilde{M}}$. In that case, the optimization can be performed only over the weights $\tilde{\lambda}_1, \dots, \tilde{\lambda}_{\tilde{M}}$ corresponding to those values. Then, (4.29) can be expressed as

$$\begin{aligned}
& \max_{\tilde{\lambda}} \quad \tilde{\mathbf{f}}^T \tilde{\boldsymbol{\lambda}} \\
& \text{subject to} \quad \tilde{\mathbf{g}}_{\theta_0}^T \tilde{\boldsymbol{\lambda}} \leq \tilde{\alpha}, \quad \forall \theta_0 \in \Lambda_0 \\
& \quad \quad \quad \mathbf{1}^T \tilde{\boldsymbol{\lambda}} = 1 \\
& \quad \quad \quad \tilde{\boldsymbol{\lambda}} \succeq \mathbf{0}
\end{aligned} \tag{4.30}$$

where

$$\begin{aligned}
\tilde{\mathbf{f}} &= [\tilde{f}_1 \cdots \tilde{f}_{\tilde{M}}]^T, \\
\tilde{\boldsymbol{\lambda}} &= [\tilde{\lambda}_1 \cdots \tilde{\lambda}_{\tilde{M}}]^T, \\
\tilde{\mathbf{g}}_{\theta_0} &= [G_{\theta_0}(F^{-1}(\tilde{f}_1)) \cdots G_{\theta_0}(F^{-1}(\tilde{f}_{\tilde{M}}))]^T.
\end{aligned}$$

The optimization problem in (4.30) is a linearly constrained linear programming (LCLP) problem. Therefore, it can be solved efficiently in polynomial time [55]. Although (4.30) is an approximation to (4.29), since it assumes that $f = F(\mathbf{n})$ can take only specific values, the solutions can get very close to each other as \tilde{M} is increased; i.e., as more values of $f = F(\mathbf{n})$ are included in the optimization problem in (4.30). Also, it should be noted that the assumption for $F(\mathbf{n})$ to take only finitely many known values can be practical in some cases, since a digital system cannot generate additive noise components with infinite precision due to quantization effects; hence, there can be only finitely many possible values of \mathbf{n} .

For the case in which the unknown parameter θ_0 under hypothesis \mathcal{H}_0 can take infinitely many values, the optimal noise may not be represented by $(M+1)$ mass points as in Theorem 4. In that case, an approximate solution is proposed based on p.d.f. approximation techniques. Let the optimal p.d.f. for the optimization

problem in (4.28) be expressed approximately by

$$p_{\mathbf{n},f}(f) = \sum_{i=1}^L \mu_i \psi_i(f - f_i) , \quad (4.31)$$

where $\mu_i \geq 0$, $\sum_{i=1}^L \mu_i = 1$, and $\psi_i(\cdot)$ is a window function that satisfies $\psi_i(\mathbf{x}) \geq 0 \forall \mathbf{x}$ and $\int \psi_i(\mathbf{x}) d\mathbf{x} = 1$, for $i = 1, \dots, L$. The p.d.f. approximation technique in (4.31) is called *Parzen window density estimation*, which has the property of mean-square convergence to the true p.d.f. under certain conditions [81]. In general, a larger L facilitates better approximation to the true p.d.f.. A common example of a window function is the Gaussian window, which is expressed as

$$\psi_i(f) = \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{f^2}{2\sigma_i^2}} .$$

Based on the approximate p.d.f. in (4.31), the optimization problem in (4.28) can be stated as

$$\begin{aligned} & \max_{\{\mu_i, f_i, \sigma_i\}_{i=1}^L} \sum_{i=1}^L \mu_i \tilde{f}_i \\ & \text{subject to } \max_{\theta_0 \in \Lambda_0} \sum_{i=1}^L \mu_i \tilde{g}_{\theta_0, i} \leq \tilde{\alpha} \\ & \sum_{i=1}^L \mu_i = 1 \\ & \mu_i \geq 0 , \quad i = 1, \dots, L \end{aligned} \quad (4.32)$$

where σ_i represents the parameter⁵ of the i th window function $\psi_i(\cdot)$, $\tilde{f}_i = \int_0^\infty f \psi_i(f - f_i) df$ and $\tilde{g}_{\theta_0, i} = \int_0^\infty g_{\theta_0} \psi_i(f - f_i) df$. Similar to the solution of (4.29), the PSO approach can be applied to obtain the optimal solution. Also, convex relaxation can be employed as in (4.30) when $\sigma_i = \sigma \forall i$ is considered as a pre-determined value, and the optimization problem is considered as determining the weights for a number of pre-determined f_i values.

⁵If there are constraints on this parameter, they should be added to the set of constraints in (4.32).

4.3 Max-Min Criterion

In this section, the aim is to determine the optimal additive noise \mathbf{n} in (4.2) that solves the following optimization problem.

$$\max_{p_{\mathbf{n}}(\cdot)} \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1) \quad (4.33)$$

$$\text{subject to } \max_{\theta_0 \in \Lambda_0} P_F^{\mathbf{y}}(\theta_0) \leq \tilde{\alpha} \quad (4.34)$$

where $P_D^{\mathbf{y}}(\theta_1)$ and $P_F^{\mathbf{y}}(\theta_0)$ are as in (4.5)-(4.8).

4.3.1 Improvability and Non-improvability Conditions

According to this criterion, the detector is called *improvable* if there exists additive noise \mathbf{n} that satisfies

$$\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1) > \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{x}}(\theta_1) = \min_{\theta_1 \in \Lambda_1} F_{\theta_1}(\mathbf{0}) \triangleq P_{D,\min}^{\mathbf{x}} \quad (4.35)$$

under the false-alarm constraint. Otherwise, the detector is *non-improvable*.

A simple sufficient condition for improvability can be obtained from the improvability definition in (4.35). If there exists a noise component $\tilde{\mathbf{n}}$ that satisfies $\min_{\theta_1 \in \Lambda_1} F_{\theta_1}(\tilde{\mathbf{n}}) > \min_{\theta_1 \in \Lambda_1} F_{\theta_1}(\mathbf{0})$ and $G_{\theta_0}(\tilde{\mathbf{n}}) \leq \tilde{\alpha} \forall \theta_0 \in \Lambda_0$, (4.5) and (4.6) implies that addition of noise $\tilde{\mathbf{n}}$ to the data vector increases the probability of detection under the false-alarm constraint for all θ_1 values; hence, $\min_{\theta_1 \in \Lambda_1} P_D^{\tilde{\mathbf{y}}}(\theta_1) > \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{x}}(\theta_1)$ is satisfied, where $\tilde{\mathbf{y}} = \mathbf{x} + \tilde{\mathbf{n}}$. However, such a noise component may not be available in many practical scenarios. Therefore, a more generic improvability condition is obtained in the following.

Similar to the max-sum case, the following function is defined for deriving generic improvability conditions:

$$H_{\min}(t) \triangleq \sup \left\{ \min_{\theta_1 \in \Lambda_1} F_{\theta_1}(\mathbf{n}) \mid t = \max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}), \mathbf{n} \in \mathbb{R}^K \right\}, \quad (4.36)$$

which defines the maximum value of the minimum detection probability for a given value of the maximum false-alarm probability. From (4.36), it is observed that if there exists $t_0 \leq \tilde{\alpha}$ such that $H_{\min}(t_0) > P_{D,\min}^{\mathbf{x}}$, the system is improvable, since under such a condition there exists a noise component \mathbf{n}_0 such that $\min_{\theta_1 \in \Lambda_1} F_{\theta_1}(\mathbf{n}_0) > P_{D,\min}^{\mathbf{x}}$ and $\max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}_0) \leq \tilde{\alpha}$. Hence, the detector performance can be improved by using an additive noise with $p_{\mathbf{n}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{n}_0)$. However, as stated previously, such a condition may not hold in many practical scenarios. Therefore, a more generic improvability condition is derived in the following theorem.

Theorem 5: *Let $\alpha = \max_{\theta_0 \in \Lambda_0} P_{\mathbb{F}}^{\mathbf{x}}(\theta_0)$ denote the maximum false-alarm probability in the absence of additive noise. If $H_{\min}(t)$ in (4.36) is second-order continuously differentiable around $t = \alpha$ and satisfies $H_{\min}''(\alpha) > 0$, then the detector is improvable.*

Proof: Since $H_{\min}''(\alpha) > 0$ and $H_{\min}(t)$ is second-order continuously differentiable around $t = \alpha$, there exist $\epsilon > 0$, \mathbf{n}_1 and \mathbf{n}_2 such that $\max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}_1) = \alpha + \epsilon$ and $\max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}_2) = \alpha - \epsilon$. Then, it is proven in the following that additive noise with $p_{\mathbf{n}}(\mathbf{x}) = 0.5 \delta(\mathbf{x} - \mathbf{n}_1) + 0.5 \delta(\mathbf{x} - \mathbf{n}_2)$ improves the detection performance under the false-alarm constraint. First, the maximum false-alarm probability in the presence of additive noise is shown not to exceed α .

$$\max_{\theta_0 \in \Lambda_0} E_{\mathbf{n}}\{G_{\theta_0}(\mathbf{n})\} \leq E_{\mathbf{n}} \left\{ \max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}) \right\} = 0.5(\alpha + \epsilon) + 0.5(\alpha - \epsilon) = \alpha. \quad (4.37)$$

Then, the increase in the detection probability is proven as follows. Since

$$\min_{\theta_1 \in \Lambda_1} E_{\mathbf{n}}\{F_{\theta_1}(\mathbf{n})\} \geq E_{\mathbf{n}} \left\{ \min_{\theta_1 \in \Lambda_1} F_{\theta_1}(\mathbf{n}) \right\} \quad (4.38)$$

is valid for all noise p.d.f.s,

$$\min_{\theta_1 \in \Lambda_1} E_{\mathbf{n}}\{F_{\theta_1}(\mathbf{n})\} \geq 0.5 H_{\min}(\alpha + \epsilon) + 0.5 H_{\min}(\alpha - \epsilon) \quad (4.39)$$

can be obtained. Due to the assumptions in the theorem, $H_{\min}(t)$ is convex in an interval around $t = \alpha$. Therefore, (4.39) becomes

$$\min_{\theta_1 \in \Lambda_1} \mathbb{E}_{\mathbf{n}} \{F_{\theta_1}(\mathbf{n})\} \geq 0.5 H_{\min}(\alpha + \epsilon) + 0.5 H_{\min}(\alpha - \epsilon) > H_{\min}(\alpha) . \quad (4.40)$$

Since $H_{\min}(\alpha) \geq P_{\text{D},\min}^{\times}$ by definition, (4.40) implies $\min_{\theta_1 \in \Lambda_1} \mathbb{E}_{\mathbf{n}} \{F_{\theta_1}(\mathbf{n})\} > P_{\text{D},\min}^{\times}$. Therefore, the detector is improvable. \square

Similar to Theorem 1 in Section 4.2.1, Theorem 5 provides a convenient sufficient condition that deals with a scalar function $H_{\min}(t)$ irrespective of the dimension of the observation vector.

In order to obtain sufficient conditions for non-improvability, the following function is defined as an extension of that in (4.19).

$$J_{\theta_0, \theta_1}(t) \triangleq \sup \{F_{\theta_1}(\mathbf{n}) \mid G_{\theta_0}(\mathbf{n}) = t, \mathbf{n} \in \mathbb{R}^K\} . \quad (4.41)$$

Then, the following theorem can be obtained as an extension of Theorem 2 in Section 4.2.1.

Theorem 6: *Let θ_1^{\min} represent the value of $\theta_1 \in \Lambda_1$ that has the minimum detection probability in the absence of additive noise; that is,*

$$\theta_1^{\min} \triangleq \arg \min_{\theta_1 \in \Lambda_1} P_{\text{D}}^{\times}(\theta_1) . \quad (4.42)$$

If there exists $\theta_0 \in \Lambda_0$ and a nondecreasing concave function $\Psi(t)$ such that $\Psi(t) \geq J_{\theta_0, \theta_1^{\min}}(t) \forall t$ and $\Psi(\tilde{\alpha}) = P_{\text{D}}^{\times}(\theta_1^{\min})$, then the detector is non-improvable.

Proof: If the detector is non-improvable for $\theta_1 = \theta_1^{\min}$, it is non-improvable according to the max-min criterion, since its minimum can never increase by using additive noise components. Therefore, the result in Theorem 6 directly follows from that in Theorem 2 by considering the non-improvability conditions at $\theta_1 = \theta_1^{\min}$. \square

The conditions in Theorem 6 can be used to determine the scenarios in which the detector performance cannot be improved via additive noise. Hence, unnecessary efforts for solving the optimization problem in (4.33) and (4.34) can be prevented.

4.3.2 Characterization of Optimal Solution

In this section, performance bounds for the detector based on $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where the p.d.f. of \mathbf{n} is obtained from (4.33) and (4.34) are derived. In addition, statistical characterization of optimal noise p.d.f.s is provided.

In order to obtain upper and lower bounds on the performance of the detector that employs the noise specified by the optimization problem in (4.33) and (4.34), consider a separate optimization problem for each $\theta_1 \in \Lambda_1$ as follows:

$$\begin{aligned} & \max_{p_{\mathbf{n}}(\cdot)} P_D^{\mathbf{y}}(\theta_1) \\ & \text{subject to } \max_{\theta_0 \in \Lambda_0} P_F^{\mathbf{y}}(\theta_0) \leq \tilde{\alpha} \end{aligned} \quad (4.43)$$

Let $P_{D,\text{opt}}^{\mathbf{y}}(\theta_1)$ represent the solution of (4.43), and $p_{\mathbf{n}_{\theta_1}}(\cdot)$ denote the corresponding optimal p.d.f.. In addition, let $\tilde{\theta}_1$ represent the parameter value with the minimum $P_{D,\text{opt}}^{\mathbf{y}}(\theta_1)$ among all $\theta_1 \in \Lambda_1$. That is,

$$\tilde{\theta}_1 = \arg \min_{\theta_1 \in \Lambda_1} P_{D,\text{opt}}^{\mathbf{y}}(\theta_1) . \quad (4.44)$$

Then, the following theorem provides performance bounds for the noise-modified detector according to the max-min criterion.

Theorem 7: *Let $P_{D,\text{mm}}^{\mathbf{y}}$ represent solution of the optimization problem specified by (4.33) and (4.34). It has the following lower and upper bounds:*

$$\max \left\{ \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{x}}(\theta_1) , \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1) \right\} \leq P_{D,\text{mm}}^{\mathbf{y}} \leq \min_{\theta_1 \in \Lambda_1} P_{D,\text{opt}}^{\mathbf{y}}(\theta_1) , \quad (4.45)$$

where $P_{D,\text{opt}}^{\mathbf{y}}(\theta_1)$ is the solution of the optimization problem in (4.43), $P_D^{\mathbf{x}}(\theta_1)$ is the probability of detection in the absence of additive noise, and $P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1)$ is the

probability of detection in the presence of additive noise $\mathbf{n}_{\tilde{\theta}_1}$, which is specified by the p.d.f. $p_{\mathbf{n}_{\tilde{\theta}_1}}(\cdot)$ that is the optimizer of (4.43) for $\tilde{\theta}_1$ defined by (4.44).

Proof: The upper bound in (4.45) directly follows from (4.33), (4.34) and (4.43), since $\max_{p_{\mathbf{n}}(\cdot)} P_D^{\mathbf{y}}(\theta_1) \geq \max_{p_{\mathbf{n}}(\cdot)} \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1)$ for all $\theta_1 \in \Lambda_1$. For the lower bound, it is first noted that the noise-modified detector can never have lower minimum detection probability than that in the absence of noise, i.e., $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{x}}(\theta_1)$. In addition, using a noise with p.d.f. $p_{\mathbf{n}_{\tilde{\theta}_1}}(\cdot)$, which is the optimal noise for the problem in (4.43) for a specific θ_1 value, can never result in a larger minimum probability $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1)$ than that obtained from the solution of (4.33) and (4.34), since the latter directly maximizes the $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1)$ metric. Therefore, $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1)$ provides another lower bound. \square

The main intuition behind the upper and lower bounds in Theorem 7 can be explained as follows. Note that $P_{D,\text{opt}}^{\mathbf{y}}(\theta_1)$ represents the maximum detection probability when an additive noise component that is optimized for a specific value of θ_1 is used. Therefore, for each $\theta_1 \in \Lambda_1$, $P_{D,\text{opt}}^{\mathbf{y}}(\theta_1)$ is larger than $\max_{p_{\mathbf{n}}(\cdot)} \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1)$, as the latter involves a single additive noise component that is optimized for the minimum detection probability metric and is used for all θ_1 values. In other words, the upper bound is obtained by assuming a more flexible optimization problem in which a different optimal noise component can be used for each θ_1 value. Considering the lower bound, the first lower bound expression is obtained from the fact that the optimal value can never be smaller than $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{x}}(\theta_1)$, which is the minimum detection probability in the absence of additive noise. The second lower bound is obtained from the observation that the optimal noise p.d.f. that maximizes the minimum detection probability, $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1)$, is obtained from the optimization problem in (4.33) and (4.34); hence, the resulting optimal value, $P_{D,\text{mm}}^{\mathbf{y}}$, is larger than or equal to all other $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}}(\theta_1)$ values that are obtained by using a different noise p.d.f..

Both the lower and the upper bounds in Theorem 7 are achievable. For example, when the detector is non-improvable, the lower bound is achieved since $P_{D,\text{mm}}^{\mathbf{y}} = \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{x}}(\theta_1)$ and $P_{D,\text{mm}}^{\mathbf{y}} \geq \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1)$. Note that $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1)$ can be smaller than $P_{D,\text{mm}}^{\mathbf{y}}$ in certain scenarios since the additive noise $p_{\mathbf{n}_{\tilde{\theta}_1}}(\cdot)$ that is optimized for $\theta_1 = \tilde{\theta}_1$ can degrade the detection performance for other θ_1 values. In fact, this is the main reason why a maximum operator is used for the lower bound in Theorem 7. On the other hand, for scenarios in which the detector performance can be improved, $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1)$ can be larger than $\min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{x}}(\theta_1)$. Also, in some cases, $\min_{\theta_1 \in \Lambda_1} P_{D,\text{opt}}^{\mathbf{y}}(\theta_1) = P_{D,\text{mm}}^{\mathbf{y}} = \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1) \geq \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{x}}(\theta_1)$ can be satisfied; that is, the upper and lower bounds in Theorem 7 can be equal. If $P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\tilde{\theta}_1) \leq P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1)$ for all $\theta_1 \in \Lambda_1$, then $p_{\mathbf{n}_{\tilde{\theta}_1}}(\cdot)$ becomes the optimal p.d.f. for the max-min problem as well, since any other noise p.d.f. will have smaller detection probability than $P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\tilde{\theta}_1)$ at $\theta_1 = \tilde{\theta}_1$, and hence will decrease the minimum detection probability. In addition, using a different optimal noise for each θ_1 will not improve the max-min performance since $P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\tilde{\theta}_1)$ will be the limiting factor. Therefore, $\min_{\theta_1 \in \Lambda_1} P_{D,\text{opt}}^{\mathbf{y}}(\theta_1) = \min_{\theta_1 \in \Lambda_1} P_D^{\mathbf{y}_{\tilde{\theta}_1}}(\theta_1)$ is satisfied, and the lower and upper bounds become equal in such a case.

Regarding the statistical characterization of the optimal additive noise according to the max-min criterion, it can be shown that when parameter sets Λ_0 and Λ_1 in (4.1) consist of a finite number of parameters, the optimal additive noise can be represented by a discrete random variable with a finite number of mass points as specified below.

Theorem 8: *Let $\theta_0 \in \Lambda_0 = \{\theta_{01}, \theta_{02}, \dots, \theta_{0M}\}$ and $\theta_1 \in \Lambda_1 = \{\theta_{11}, \theta_{12}, \dots, \theta_{1N}\}$. Assume that the additive noise components can take finite values specified by $n_i \in [a_i, b_i]$, $i = 1, \dots, K$, for any finite a_i and b_i . Define set U as*

$$U = \left\{ (u_1, \dots, u_{N+M}) : u_1 = F_{\theta_{11}}(\mathbf{n}), \dots, u_N = F_{\theta_{1N}}(\mathbf{n}), u_{N+1} = G_{\theta_{01}}(\mathbf{n}), \dots, \right. \\ \left. u_{N+M} = G_{\theta_{0M}}(\mathbf{n}), \text{ for } \mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b} \right\}, \quad (4.46)$$

where $\mathbf{a} \preceq \mathbf{n} \preceq \mathbf{b}$ means that $n_i \in [a_i, b_i]$ for $i = 1, \dots, K$. If U is a closed subset of \mathbb{R}^{N+M} , an optimal solution to (4.33) and (4.34) has the following form

$$p_{\mathbf{n}}(\mathbf{x}) = \sum_{i=1}^{N+M} \lambda_i \delta(\mathbf{x} - \mathbf{n}_i), \quad (4.47)$$

where $\sum_{i=1}^{N+M} \lambda_i = 1$ and $\lambda_i \geq 0$ for $i = 1, 2, \dots, N + M$.

Proof: The proof is omitted since it is a straightforward extension of that of Theorem 4. \square

The main difference of Theorem 8 from Theorem 4 in Section 4.2.2 is that both Λ_0 and Λ_1 should be discrete for the optimal p.d.f. to have a discrete structure in the max-min framework. However, for the max-sum criterion, it is enough to have a discrete Λ_0 in order to have a discrete p.d.f. as stated in Theorem 4. The reason for this is that according to the max-sum criterion, the objective function to maximize becomes $E_{\mathbf{n}}\{F(\mathbf{n})\}$, where $F(\mathbf{n}) = \int_{\theta_1 \in \Lambda_1} F_{\theta_1}(\mathbf{n}) d\theta_1$ is as defined in (4.14). In other words, maximization of a single function is considered in the max-sum problem under the false-alarm constraint.

4.3.3 Calculation of Optimal Solution and Convex Relaxation

In this section, possible approaches to solving the optimization problem in (4.33) and (4.34) are considered. In order to express the optimization problem as optimization over a single-dimensional p.d.f., consider a specific value of $\theta_1 \in \Lambda_1$, for which $F_{\theta_1}(\mathbf{n})$ is one-to-one. Let this value be represented as $\tilde{\theta}_1$. Then, for a given value \mathbf{n} of noise, $f = F_{\tilde{\theta}_1}(\mathbf{n})$ can be used to express $g_{\theta_0} = G_{\theta_0}(\mathbf{n})$ and $f_{\theta_1} = F_{\theta_1}(\mathbf{n})$

as $g_{\theta_0} = G_{\theta_0} \left(F_{\tilde{\theta}_1}^{-1}(f) \right)$ and $f_{\theta_1} = F_{\theta_1} \left(F_{\tilde{\theta}_1}^{-1}(f) \right)$, respectively. Therefore, the optimization problem in (4.33) and (4.34) can be reformulated as

$$\begin{aligned} & \max_{p_{\mathbf{n}, f_{\tilde{\theta}_1}}(\cdot)} \min_{\theta_1 \in \Lambda_1} \int_0^1 f_{\theta_1} p_{\mathbf{n}, f_{\tilde{\theta}_1}}(f) df, \\ & \text{subject to } \max_{\theta_0 \in \Lambda_0} \int_0^1 g_{\theta_0} p_{\mathbf{n}, f_{\tilde{\theta}_1}}(f) df \leq \tilde{\alpha}. \end{aligned} \quad (4.48)$$

First, consider the case in which the parameters can take finitely many values specified by $\theta_0 \in \Lambda_0 = \{\theta_{01}, \theta_{02}, \dots, \theta_{0M}\}$ and $\theta_1 \in \Lambda_1 = \{\theta_{11}, \theta_{12}, \dots, \theta_{1N}\}$. In this case, the optimal noise p.d.f. can be represented by $(N + M)$ mass points under the conditions in Theorem 8. Hence, (4.48) can be expressed as

$$\begin{aligned} & \max_{\{\lambda_i, f_i\}_{i=1}^{N+M}} \min_{\theta_1 \in \Lambda_1} \sum_{i=1}^{N+M} \lambda_i f_{\theta_1, i} \\ & \text{subject to } \max_{\theta_0 \in \Lambda_0} \sum_{i=1}^{N+M} \lambda_i g_{\theta_0, i} \leq \tilde{\alpha} \\ & \sum_{i=1}^{N+M} \lambda_i = 1 \\ & \lambda_i \geq 0, \quad i = 1, \dots, N + M \end{aligned} \quad (4.49)$$

where $f_i = F_{\tilde{\theta}_1}(\mathbf{n}_i)$, $f_{\theta_1, i} = F_{\theta_1}(F_{\tilde{\theta}_1}^{-1}(f_i))$, $g_{\theta_0, i} = G_{\theta_0}(F_{\tilde{\theta}_1}^{-1}(f_i))$, and \mathbf{n}_i and λ_i are, respectively, the optimal mass points and their weights as specified in Theorem 8. Since the optimization problem in (4.49) may not be formulated as a convex optimization problem in general, global optimization techniques, such as PSO [51]-[54] can be employed to obtain the optimal solution, as studied in Section 4.5.

Due to the complexity of the optimization problem in (4.49), an approximate and efficient formulation can be obtained by the convex relaxation approach as in Section 4.2.3. Assume that $f = F_{\tilde{\theta}_1}(\mathbf{n})$ can take known values of $\tilde{f}_1, \dots, \tilde{f}_{\tilde{M}}$ only. In that case, the optimization can be performed only over the weights $\tilde{\lambda}_1, \dots, \tilde{\lambda}_{\tilde{M}}$

corresponding to those values. Hence, (4.49) becomes

$$\begin{aligned}
& \max_{\tilde{\boldsymbol{\lambda}}} \min_{\theta_1 \in \Lambda_1} \tilde{\mathbf{f}}_{\theta_1}^T \tilde{\boldsymbol{\lambda}} \\
& \text{subject to } \tilde{\mathbf{g}}_{\theta_0}^T \tilde{\boldsymbol{\lambda}} \leq \tilde{\alpha}, \quad \forall \theta_0 \in \Lambda_0 \\
& \quad \mathbf{1}^T \tilde{\boldsymbol{\lambda}} = 1 \\
& \quad \tilde{\boldsymbol{\lambda}} \succeq \mathbf{0}
\end{aligned} \tag{4.50}$$

where

$$\begin{aligned}
\tilde{\mathbf{f}}_{\theta_1} &= \left[F_{\theta_1} \left(F_{\tilde{\theta}_1}^{-1}(\tilde{f}_1) \right) \cdots F_{\theta_1} \left(F_{\tilde{\theta}_1}^{-1}(\tilde{f}_{\tilde{M}}) \right) \right]^T \\
\tilde{\mathbf{g}}_{\theta_0} &= \left[G_{\theta_0} \left(F_{\tilde{\theta}_1}^{-1}(\tilde{f}_1) \right) \cdots G_{\theta_0} \left(F_{\tilde{\theta}_1}^{-1}(\tilde{f}_{\tilde{M}}) \right) \right]^T \\
\tilde{\boldsymbol{\lambda}} &= [\tilde{\lambda}_1 \cdots \tilde{\lambda}_{\tilde{M}}]^T .
\end{aligned}$$

The optimization problem (4.50) can be expressed as a convex problem when we define an auxiliary optimization variable t as follows:

$$\begin{aligned}
& \max_{\tilde{\boldsymbol{\lambda}}, t} t \\
& \text{subject to } \tilde{\mathbf{f}}_{\theta_1}^T \tilde{\boldsymbol{\lambda}} \geq t, \quad \forall \theta_1 \in \Lambda_1 \\
& \quad \tilde{\mathbf{g}}_{\theta_0}^T \tilde{\boldsymbol{\lambda}} \leq \tilde{\alpha}, \quad \forall \theta_0 \in \Lambda_0 \\
& \quad \mathbf{1}^T \tilde{\boldsymbol{\lambda}} = 1 \\
& \quad \tilde{\boldsymbol{\lambda}} \succeq \mathbf{0}
\end{aligned} \tag{4.51}$$

In fact, (4.51) can be recognized as an LCLP problem if the new optimization variable is defined as $\mathbf{x} = \left[\tilde{\boldsymbol{\lambda}}^T \ t \right]^T$. Therefore, it can be solved efficiently in polynomial time [55]. Although (4.51) is an approximation to (4.49), the solutions get very close as more values of $f = F_{\tilde{\theta}_1}(\mathbf{n})$ are included in the optimization.

When at least one of θ_0 or θ_1 can take infinitely many values, the optimal noise may not be represented by a finite number of mass points as in Theorem 8. In such cases, the optimization problem in (4.48) can be solved over the set of p.d.f. approximations as in Section 4.2.3. Let the optimal p.d.f. be approximated

by

$$p_{\mathbf{n}, f_{\tilde{\theta}_1}}(f) = \sum_{i=1}^L \mu_i \psi_i(f - f_i), \quad (4.52)$$

where $\mu_i \geq 0$, $\sum_{i=1}^L \mu_i = 1$, and $\psi_i(\cdot)$ is a window function that satisfies $\psi_i(\mathbf{x}) \geq 0 \forall \mathbf{x}$ and $\int \psi_i(\mathbf{x}) d\mathbf{x} = 1$, for $i = 1, \dots, L$. Then, the optimization problem in (4.48) can be stated as

$$\begin{aligned} & \max_{\{\mu_i, f_i, \sigma_i\}_{i=1}^L} \min_{\theta_1 \in \Lambda_1} \sum_{i=1}^L \mu_i \tilde{f}_{\theta_1, i} \\ \text{subject to } & \max_{\theta_0 \in \Lambda_0} \sum_{i=1}^L \mu_i \tilde{g}_{\theta_0, i} \leq \tilde{\alpha} \\ & \sum_{i=1}^L \mu_i = 1 \\ & \mu_i \geq 0, \quad i = 1, \dots, L \end{aligned} \quad (4.53)$$

where σ_i represents the parameter of the i th window function $\psi_i(\cdot)$, $\tilde{f}_{\theta_1, i} = \int f_{\theta_1} \psi_i(f - f_i) df$, and $\tilde{g}_{\theta_0, i} = \int g_{\theta_0} \psi_i(f - f_i) df$. Similar to the solution of (4.49), the PSO approach can be employed, for example, to obtain the optimal solution of (4.53). Also, the convex relaxation technique can be employed as in (4.50) and (4.51) when $\sigma_i = \sigma \forall i$ is considered as a pre-determined value.

4.4 Max-Max Criterion

In this section, the aim is to determine the optimal additive noise \mathbf{n} in (4.2) that solves the following optimization problem.

$$\max_{p_{\mathbf{n}}(\cdot)} \max_{\theta_1 \in \Lambda_1} P_D^{\mathbf{Y}}(\theta_1) \quad (4.54)$$

$$\text{subject to } \max_{\theta_0 \in \Lambda_0} P_F^{\mathbf{Y}}(\theta_0) \leq \tilde{\alpha} \quad (4.55)$$

where $P_D^{\mathbf{Y}}(\theta_1)$ and $P_F^{\mathbf{Y}}(\theta_0)$ are as in (4.5)-(4.8). According to the max-max criterion, the detector is called *improvable* if there exists additive noise \mathbf{n} that

satisfies

$$\max_{\theta_1 \in \Lambda_1} P_D^Y(\theta_1) > \max_{\theta_1 \in \Lambda_1} P_D^X(\theta_1) = \max_{\theta_1 \in \Lambda_1} F_{\theta_1}(\mathbf{0}) \triangleq P_{D,\max}^X \quad (4.56)$$

under the false-alarm constraint. Otherwise, the detector is *non-improvable*.

The results in the previous sections can be extended to cover the max-max case as well. Since the derivations are quite similar, the results for this case are stated without any proofs.

Let θ_1^{\max} represent the value of $\theta_1 \in \Lambda_1$ that has the maximum detection probability in the absence of additive noise; that is, $\theta_1^{\max} \triangleq \arg \max_{\theta_1 \in \Lambda_1} P_D^X(\theta_1)$. In addition, define

$$H_{\theta_1}(t) \triangleq \sup \left\{ F_{\theta_1}(\mathbf{n}) \mid \max_{\theta_0 \in \Lambda_0} G_{\theta_0}(\mathbf{n}) = t, \mathbf{n} \in \mathbb{R}^K \right\}. \quad (4.57)$$

Then, the detector is improvable if $H_{\theta_1^{\max}}(t)$ is second-order continuously differentiable around $t = \alpha$ and satisfies $H_{\theta_1^{\max}}''(\alpha) > 0$, where $\alpha \triangleq \max_{\theta_0 \in \Lambda_0} P_F^X(\theta_0)$. This result can be proven as in Theorem 1. In fact, it directly follows from the observation that if the detector can be improved for $\theta_1 = \theta_1^{\max}$, then the maximum of $\max_{\theta_1 \in \Lambda_1} P_D^Y(\theta_1)$ is always larger than $\max_{\theta_1 \in \Lambda_1} P_D^X(\theta_1)$.

A non-improvability condition can be obtained in a similar way to that in Theorem 6. The detector is non-improvable if there exists $\theta_0 \in \Lambda_0$ and a nondecreasing concave function $\Psi_{\theta_1}(t)$ such that $\Psi_{\theta_1}(t) \geq J_{\theta_0, \theta_1}(t) \forall t$ and $\Psi_{\theta_1}(\tilde{\alpha}) = P_D^X(\theta_1)$ for all $\theta_1 \in \Lambda_1$, where $J_{\theta_0, \theta_1}(t)$ is given by (4.41).

Regarding the structure of the optimal noise p.d.f. for the problem in (4.54) and (4.55), consider a composite hypothesis-testing problem with $\theta_0 \in \Lambda_0 = \{\theta_{01}, \theta_{02}, \dots, \theta_{0M}\}$. Then, it can be concluded that the optimal p.d.f. can be represented by $(M + 1)$ mass points under the conditions in Theorem 4. This follows from the fact that the max-max problem in (4.54) and (4.55) can be solved by choosing the p.d.f. that results in the maximum detection probability

among the p.d.f.s that solve the following optimization problems:

$$\max_{p_{\mathbf{n}}(\cdot)} P_D^Y(\theta_1) \quad (4.58)$$

$$\text{subject to } \max_{\theta_0 \in \Lambda_0} P_F^Y(\theta_0) \leq \tilde{\alpha} \quad (4.59)$$

for $\theta_1 \in \Lambda_1$. In other words, the optimal noise p.d.f. can be calculated for each $\theta_1 \in \Lambda_1$ separately, and the noise p.d.f. that yields the maximum detection probability becomes the solution of the max-max problem. Since the structure of each optimization problem is as in the max-sum formulation, Theorem 4 applies to the max-max case as well.

Finally, for the solution of the max-max problem, the approaches in Section 4.2.3 for the max-sum problem can directly be applied, since the optimization problems in (4.10)-(4.11) and (4.58)-(4.59) have the same structure.

4.5 Numerical Results

In this section, a composite version of the detection example in [12] and [24] is studied in order to illustrate the theoretical results obtained in the previous sections. Namely, the following composite hypothesis-testing problem is considered:

$$\begin{aligned} \mathcal{H}_0 &: x = w \\ \mathcal{H}_1 &: x = A + w \end{aligned} \quad (4.60)$$

where A is a known constant, and w is the noise term that has a Gaussian mixture distribution specified as

$$p_w(w) = \frac{1}{2} \gamma(w; -\theta, \sigma^2) + \frac{1}{2} \gamma(w; \theta, \sigma^2), \quad (4.61)$$

with

$$\gamma(w; \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(w - \theta)^2}{2\sigma^2}\right\}.$$

The p.d.f. of noise w has an unknown parameter θ , which belongs to Λ_0 under hypothesis \mathcal{H}_0 and to Λ_1 under \mathcal{H}_1 with $\Lambda_0 \cap \Lambda_1 = \emptyset$.

From (4.60) and (4.61), the probability distributions of observation x under hypotheses \mathcal{H}_0 and \mathcal{H}_1 are given, respectively, by

$$p_{\theta_0}(x) = \frac{1}{2} \gamma(x; -\theta_0, \sigma^2) + \frac{1}{2} \gamma(x; \theta_0, \sigma^2) , \quad (4.62)$$

$$p_{\theta_1}(x) = \frac{1}{2} \gamma(x; -\theta_1 + A, \sigma^2) + \frac{1}{2} \gamma(x; \theta_1 + A, \sigma^2) . \quad (4.63)$$

Since additive noise can improve the performance of suboptimal detectors only [24], a suboptimal sign detector, as in [12], is considered as the decision rule for the problem in (4.60), which is given by

$$\phi(x) = \begin{cases} 1 , & x > 0 \\ 0 , & x \leq 0 \end{cases} . \quad (4.64)$$

Then, from (4.62)-(4.64), detection and false-alarm probabilities when constant noise is added can be calculated as (c.f. (4.7) and (4.8))

$$\begin{aligned} F_{\theta_1}(x) &= \int_{-\infty}^{\infty} \phi(y) p_{\theta_1}(y-x) dy \\ &= \frac{1}{2} Q\left(\frac{-x + \theta_1 - A}{\sigma}\right) + \frac{1}{2} Q\left(\frac{-x - \theta_1 - A}{\sigma}\right) \end{aligned} \quad (4.65)$$

and

$$\begin{aligned} G_{\theta_0}(x) &= \int_{-\infty}^{\infty} \phi(y) p_{\theta_0}(y-x) dy \\ &= \frac{1}{2} Q\left(\frac{-x + \theta_0}{\sigma}\right) + \frac{1}{2} Q\left(\frac{-x - \theta_0}{\sigma}\right) , \end{aligned} \quad (4.66)$$

respectively, where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt$ is the Q -function. It is noted that both $F_{\theta_1}(x)$ and $G_{\theta_0}(x)$ are monotone increasing functions of x for all parameter values.

The aim is to add noise n to observation x in (4.60), and to improve the detection performance of the sign detector in (4.64) under a false-alarm constraint.

The noise-modified observation is denoted as $y = x + n$, and the probabilities of detection and false-alarm are given by

$$P_D^y(\theta_1) = \int_{-\infty}^{\infty} F_{\theta_1}(x)p_n(x) dx, \quad P_F^y(\theta_0) = \int_{-\infty}^{\infty} G_{\theta_0}(x)p_n(x) dx, \quad (4.67)$$

respectively, where $p_n(\cdot)$ represents the p.d.f. of the additive noise.

4.5.1 Scenario-1: Λ_0 and Λ_1 with finite number of elements

In the first scenario, the parameter sets under \mathcal{H}_0 and \mathcal{H}_1 are specified as $\theta_0 \in \Lambda_0 = \{0.1, 0.4\}$ and $\theta_1 \in \Lambda_1 = \{2, 2.5, 4\}$. According to Theorem 4 and Theorem 8, the optimal additive noise has a p.d.f. of the form $p_n(x) = \sum_{i=1}^{N_m} \lambda_i \delta(x - n_i)$, where $N_m = 3$ for the max-sum case, and $N_m = 5$ for the max-min case. For the noise p.d.f. specified as $p_n(x) = \sum_{i=1}^{N_m} \lambda_i \delta(x - n_i)$, the detection and false-alarm probabilities in (4.67) become

$$\begin{aligned} P_D^y(\theta_1) &= \sum_{i=1}^{N_m} \frac{\lambda_i}{2} \left[Q\left(\frac{-n_i + \theta_1 - A}{\sigma}\right) + Q\left(\frac{-n_i - \theta_1 - A}{\sigma}\right) \right], \\ P_F^y(\theta_0) &= \sum_{i=1}^{N_m} \frac{\lambda_i}{2} \left[Q\left(\frac{-n_i + \theta_0}{\sigma}\right) + Q\left(\frac{-n_i - \theta_0}{\sigma}\right) \right]. \end{aligned} \quad (4.68)$$

For the first simulations, $A = 1$ and $\sigma = 1$ are used. For the max-sum and max-min cases, the original detection probabilities (i.e., in the absence of additive noise) can be calculated from (4.65) and (4.66) as $P_{D,\text{sum}}^x = 1.613$ and $P_{D,\text{min}}^x = 0.5007$, respectively, with $\max_{\theta_0} P_F^x(\theta_0) = \alpha = \tilde{\alpha} = 0.5$. Then, the PSO⁶ and the convex relaxation techniques are applied as described in Sections 4.2.3 and 4.3.3, and the optimal additive noise p.d.f.s are calculated for both the max-sum and max-min cases, which are illustrated in Figure 4.2 and Figure 4.3,

⁶In the PSO algorithm, 50 particles and 1000 iterations are employed. In addition, the other parameters are set to $c_1 = c_2 = 2.05$ and $\chi = 0.72984$, and the inertia weight ω is changed from 1.2 to 0.1 linearly with the iteration number. Please refer to [51] for the details of the PSO algorithm and the definitions of the parameters.

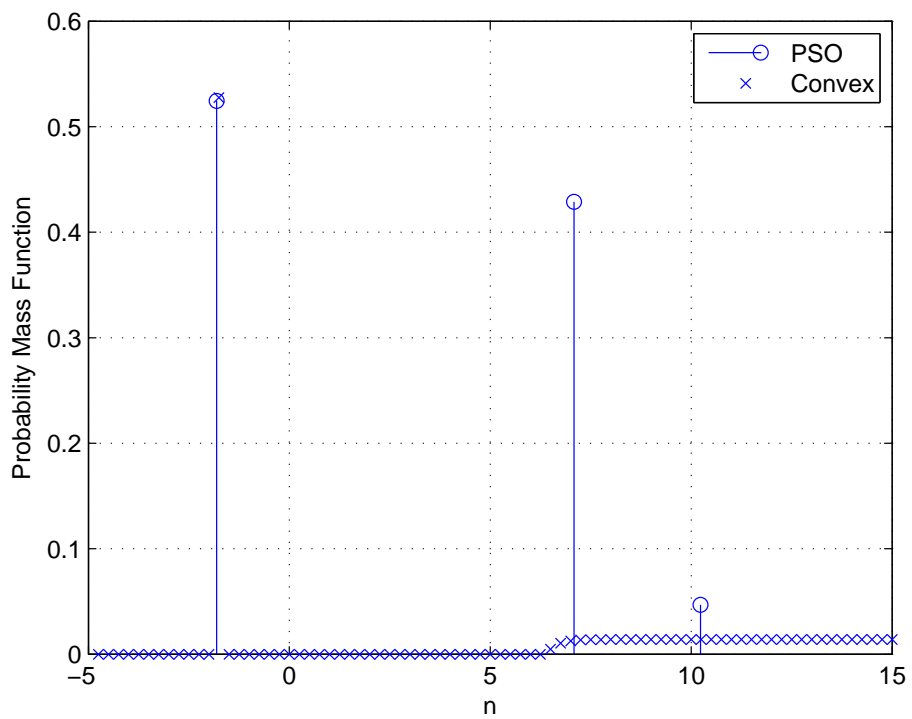


Figure 4.2: Probability mass functions of the optimal additive noise based on the PSO and the convex relaxation techniques for the max-sum case when $A = 1$ and $\sigma = 1$.

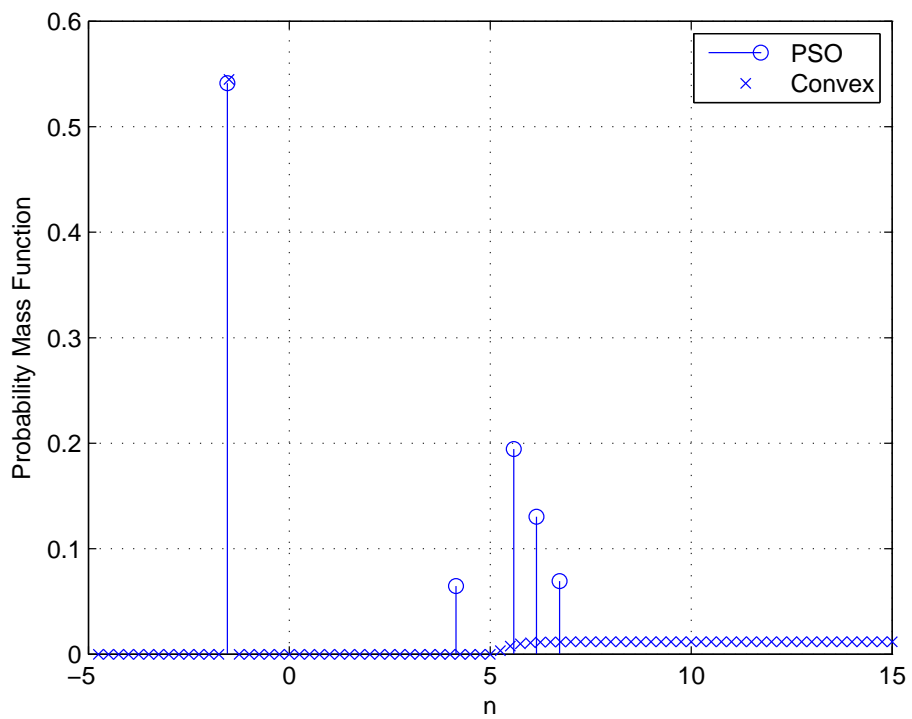


Figure 4.3: Probability mass functions of the optimal additive noise based on the PSO and the convex relaxation techniques for the max-min case when $A = 1$ and $\sigma = 1$.

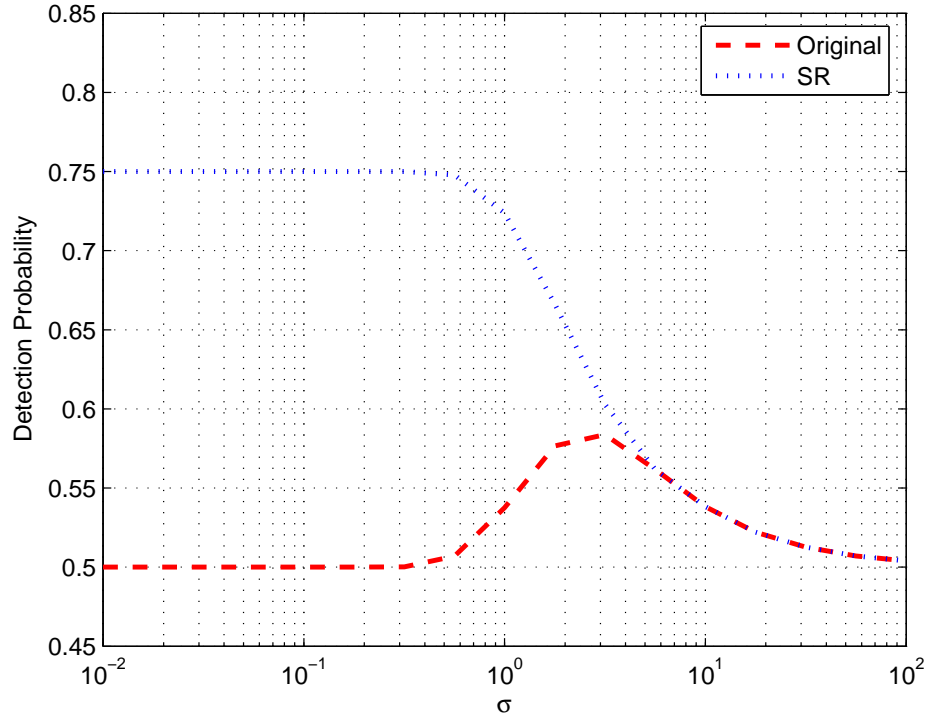


Figure 4.4: Comparison of detection probabilities (normalized) in the absence (“original”) and presence (“SR”) of additive noise according to the max-sum criterion for various values of σ .

respectively. For the convex solutions, the optimizations are performed over the noise values that are specified as $-15 + 0.25i$ for $i = 0, 1, \dots, 120$. The resulting detection probabilities when the PSO algorithm is used are calculated as $P_{D,\text{sum}}^y = 2.172$ and $P_{D,\text{mm}}^y = 0.711$ under the constraint that $\max_{\theta_0} P_F^y(\theta_0) = 0.5$. In other words, improvement ratios of $2.172/1.613 = 1.347$ and $0.711/0.5007 = 1.420$ are obtained according to the max-sum and max-min criteria, respectively. When the convex relaxation approach is employed, the detection probabilities become $P_{D,\text{sum}}^y = 2.171$ and $P_{D,\text{mm}}^y = 0.711$, which are almost the same as those obtained by the PSO technique. It is noted from Figure 4.2 and Figure 4.3 that the convex solutions approximate the optimal PSO solutions with 3 and 5 mass points (for the max-sum and max-min cases, respectively) with a larger number of non-zero mass points.

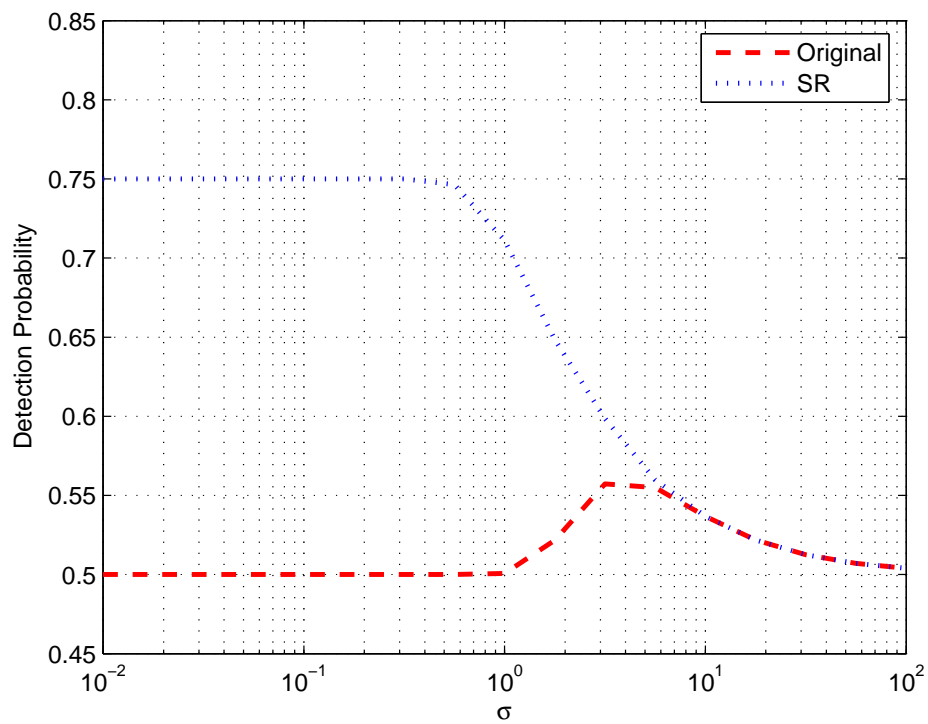


Figure 4.5: Comparison of detection probabilities (normalized) in the absence (“original”) and presence (“SR”) of additive noise according to the max-min criterion for various values of σ .

Next, $A = 1$ is used, and the detection probabilities are plotted for various values of σ in (4.61) in the absence and in the presence of additive noise (referred to as “original” and “SR” detectors, respectively).⁷ Figure 4.4 illustrates the resulting plot for the max-sum criterion. The normalized detection probability is used in the figure, which is defined as $P_{D,\text{sum}}^y/3$ as there are three possible θ_1 values. It is observed from the figure that the improvement via additive noise increases as σ decreases. Figure 4.5 illustrates the case for the max-min criterion. Similar to the max-sum case, the improvement is observed for small σ values. The observation that the detector becomes non-improvable for large σ values is mainly due to the fact that the improvability is commonly caused by the multi-modal nature of the measurement noise p.d.f. in (4.61), which reduces as σ increases.

Figure 4.6 illustrates the sufficient conditions in Theorem 1 and Theorem 5 for the max-sum and max-min cases with respect to σ . It is obtained that the improvement is guaranteed in the interval $\sigma \in [0.1259, 2.639]$ for the max-sum case and in the interval $\sigma \in [0.3981, 3.978]$ for the max-min case. Comparison of Figure 4.6 with Figure 4.4 and Figure 4.5 reveals that whenever the second derivative is positive, the detector is improvable as stated in the related theorems; however, it also indicates that the conditions in Theorem 1 and Theorem 5 are not necessary conditions, as the detector can be improved also for smaller σ values.

4.5.2 Scenario-2: Λ_0 and Λ_1 are continuous intervals

In the second scenario, $\Lambda_0 = [0.1, 0.4]$ and $\Lambda_1 = [2, 5]$ are used. As discussed in Sections 4.2.3 and 4.3.3, an approximation to the optimal additive noise p.d.f. as in (4.31) can be used to obtain an approximate solution in such a scenario. Considering Gaussian window functions for p.d.f. approximation, the additive

⁷The PSO technique is employed for the SR case.

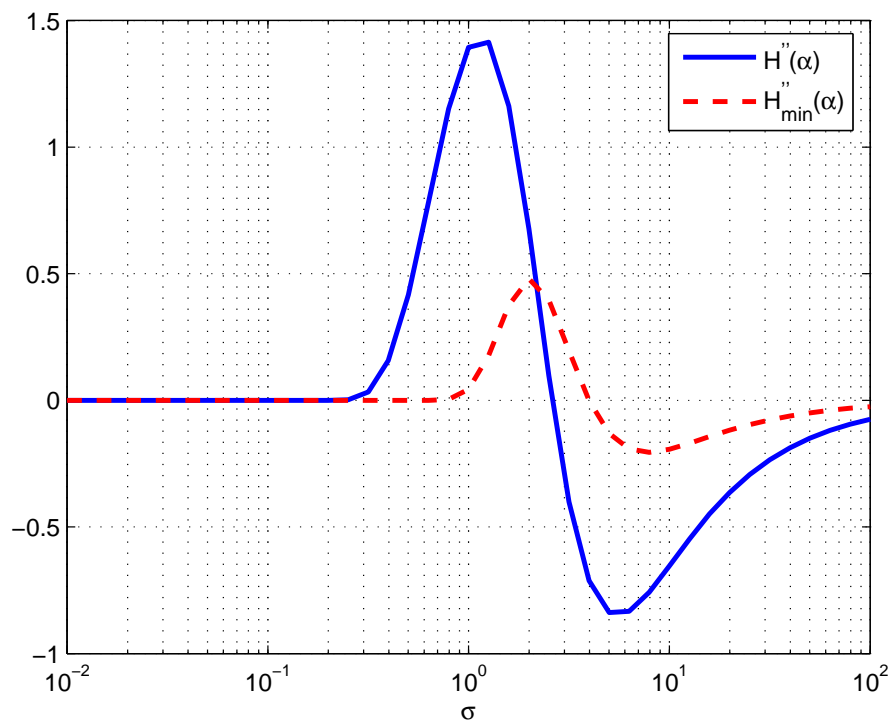


Figure 4.6: The second-order derivatives of $H(t)$ in (4.17) and $H_{\min}(t)$ (4.36) at $t = \alpha$ for various values of σ . Theorem 1 and Theorem 5 imply that the detector is improvable whenever the second-order derivative at $t = \alpha$ is positive.

noise p.d.f. can be expressed as⁸

$$p_n(x) = \sum_{i=1}^L \mu_i \gamma(x; \eta_i, \sigma_i^2). \quad (4.69)$$

Then, the probabilities of detection and false-alarm can be calculated from (4.67), after some manipulation, as

$$P_D^y(\theta_1) = \sum_{i=1}^L \frac{\mu_i}{2} \left[Q \left(\frac{-\theta_1 - \eta_i - A}{\sqrt{\sigma^2 + \sigma_i^2}} \right) + Q \left(\frac{\theta_1 - \eta_i - A}{\sqrt{\sigma^2 + \sigma_i^2}} \right) \right], \quad (4.70)$$

$$P_F^y(\theta_0) = \sum_{i=1}^L \frac{\mu_i}{2} \left[Q \left(\frac{-\theta_0 - \eta_i}{\sqrt{\sigma^2 + \sigma_i^2}} \right) + Q \left(\frac{\theta_0 - \eta_i}{\sqrt{\sigma^2 + \sigma_i^2}} \right) \right]. \quad (4.71)$$

For the following simulations, $L = 8$ is considered, and the parameters $\{\mu_i, \eta_i, \sigma_i\}_{i=1}^8$ are obtained via the PSO algorithm for both the max-sum and max-min cases. First, $A = 1$ and $\sigma = 1$ are used. In the absence of additive noise, the detection probabilities in the max-sum and max-min cases are given, respectively, by $\int_{\theta_1 \in \Lambda_1} P_D^x(\theta_1) d\theta_1 = \int_{\theta_1 \in \Lambda_1} F_{\theta_1}(0) d\theta_1 = 1.5417$ and $\min_{\theta_1 \in \Lambda_1} P_D^x(\theta_1) = \min_{\theta_1 \in \Lambda_1} F_{\theta_1}(0) = 0.5$ with $\max_{\theta_0 \in \Lambda_0} P_F^x(\theta_0) = \max_{\theta_0 \in \Lambda_0} G_{\theta_0}(0) = \alpha = \tilde{\alpha} = 0.5$. When the optimal additive noise p.d.f.s are calculated via the PSO algorithm, the detection probabilities become $\int_{\theta_1 \in \Lambda_1} P_D^y(\theta_1) d\theta_1 = 2.1426$ for the max-sum case, and $\min_{\theta_1 \in \Lambda_1} P_D^y(\theta_1) = 0.6943$ for the max-min case. In other words, improvement ratios of 1.390 and 1.389 are obtained for the max-sum and max-min cases, respectively. The optimal additive noise p.d.f.s for the max-sum and max-min cases are shown in Figure 4.7 and Figure 4.8, respectively.

In Figure 4.9 and Figure 4.10, the detection probabilities according to the max-sum and max-min criteria are plotted, respectively, for both the original detector (i.e., without additive noise) and the noise-modified one when $A = 1$. For the max-sum case, the detection probability is normalized as $\frac{1}{3} \int_2^5 P_D^y(\theta_1) d\theta_1$. Similar to the first scenario, more improvement can be achieved as σ decreases, and no improvement is observed for large values of σ .

⁸Since scalar observations are considered in this example, the optimization problem can also be solved in the original noise domain, instead of the detection probability domain as in (4.28) or (4.48).

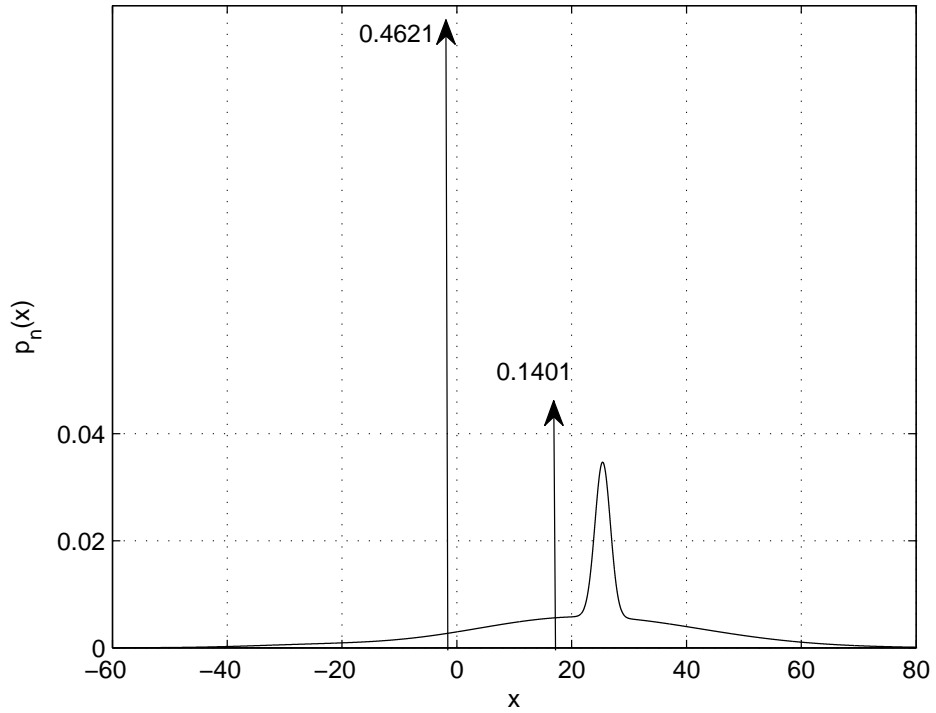


Figure 4.7: The optimal additive noise p.d.f. in (4.69) for $A = 1$ and $\sigma = 1$ according to the max-sum criterion. The optimal parameters in (4.69) obtained via the PSO algorithm are $\boldsymbol{\mu} = [0.0969 \ 0 \ 0.0019 \ 0.1401 \ 0.1377 \ 0.0143 \ 0.1470 \ 0.4621]$, $\boldsymbol{\eta} = [25.4039 \ -20.1423 \ 13.7543 \ 17.0891 \ 29.7452 \ -25.0785 \ 17.6887 \ -2.2085]$, and $\boldsymbol{\sigma} = [1.3358 \ 26.2930 \ 11.3368 \ 0 \ 19.5556 \ 11.5953 \ 17.9838 \ 0.0001]$. The mass centers with very small variances ($\eta_i = 17.0891$ and $\eta_i = -2.2085$) are marked by arrows for convenience.

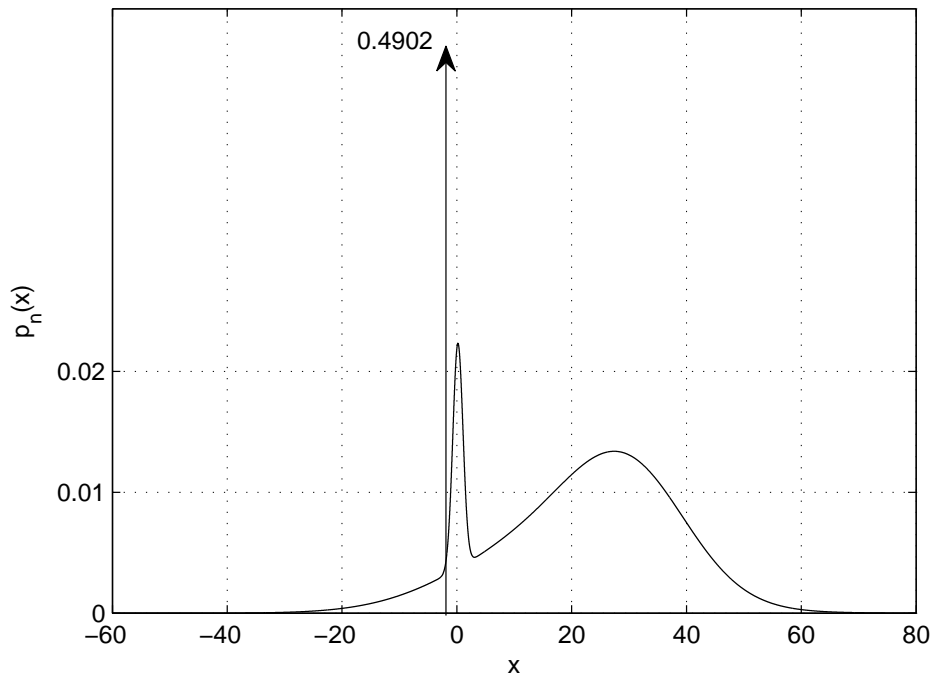


Figure 4.8: The optimal additive noise p.d.f. in (4.69) for the max-min criterion when $A = 1$ and $\sigma = 1$. The optimal parameters in (4.69) obtained via the PSO algorithm are $\boldsymbol{\mu} = [0.0067 \ 0.1797 \ 0.0411 \ 0.2262 \ 0.0064 \ 0.0498 \ 0 \ 0.4902]$, $\boldsymbol{\eta} = [20.1017 \ 15.0319 \ 0.1815 \ 29.9668 \ 17.2657 \ 22.8092 \ -0.7561 \ -1.4484]$, and $\boldsymbol{\sigma} = [16.5204 \ 15.1445 \ 0.8805 \ 10.1573 \ 12.9094 \ 17.4184 \ 19.0959 \ 0.0102]$. The mass center $\eta_i = -1.4484$ is marked by an arrow for convenience as it has a very small variance.

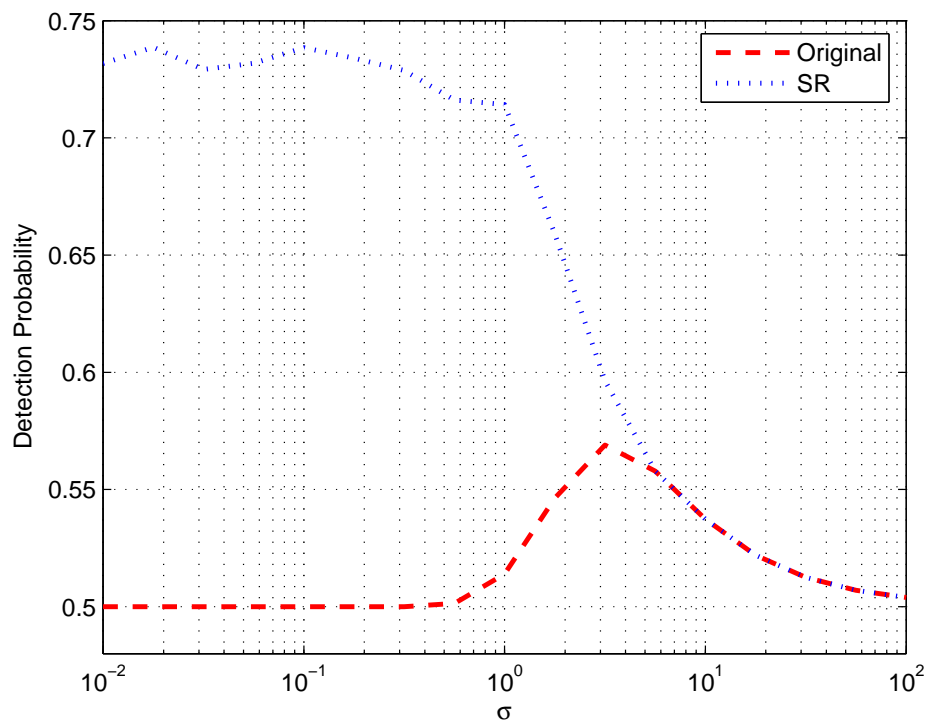


Figure 4.9: Comparison of detection probabilities (normalized) in the absence (“original”) and presence (“SR”) of additive noise according to the max-sum criterion for various values of σ .

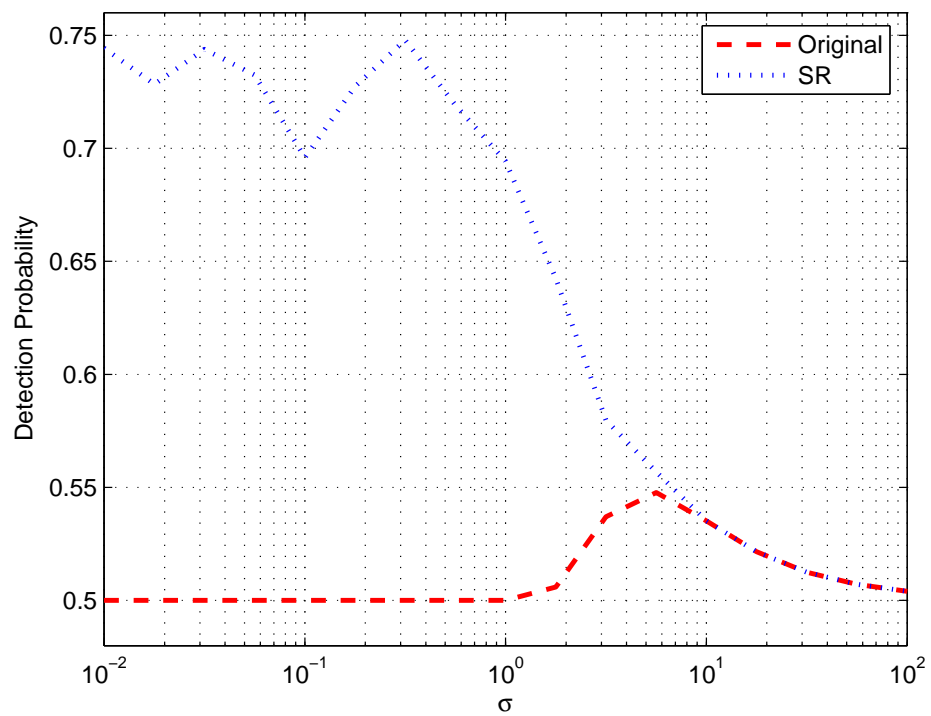


Figure 4.10: Comparison of detection probabilities (normalized) in the absence (“original”) and presence (“SR”) of additive noise according to the max-min criterion for various values of σ .

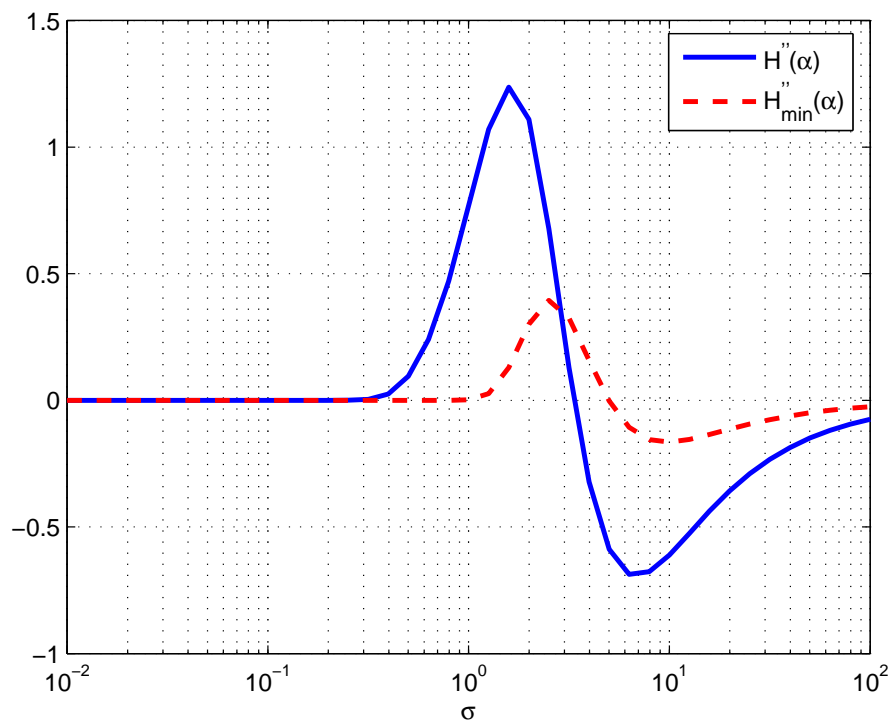


Figure 4.11: The second-order derivatives of $H(t)$ in (4.17) and $H_{\min}(t)$ (4.36) at $t = \alpha$ for various values of σ . Theorem 1 and Theorem 5 imply that the detector is improvable whenever the second-order derivative at $t = \alpha$ is positive.

Finally, the improvability conditions in Theorem 1 and Theorem 5 are investigated in Figure 4.11. It is observed from the figures that the detector is improvable in the interval $\sigma \in [0.1585, 3.398]$ for the max-sum case and in the interval $\sigma \in [0.5012, 4.996]$ for the max-min case, which together with Figure 4.9 and Figure 4.10 imply that the conditions in the theorems are sufficient but not necessary.

4.6 Concluding Remarks and Extensions

In this chapter, the effects of additive independent noise have been investigated for composite hypothesis-testing problems. The Neyman-Pearson framework has been considered, and performance of noise-modified detectors has been analyzed according to the max-sum, max-min and max-max criteria. Improvability and non-improvability conditions have been derived for each case, and the statistical characterization of optimal additive noise p.d.f.s has been provided. A detection example has been presented in order to explain the theoretical results.

Although the additive independent noise as in Figure 4.1 is considered in this study, the results can be extended to other noise injection approaches than the addition operation by considering a nonlinear transformation of the observation, as discussed in [12]. In that case, the nonlinear operator and the original detector can be regarded together as a new detector and the results in this study can directly be applied.

Chapter 5

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

This chapter is organized as follows. In Section 5.1, the formulation of the restricted Neyman-Pearson (NP) criterion and motivations for employing this criterion are presented. Some characteristics of the optimal decision rule and algorithms to obtain the optimal solution are investigated in Section 5.2. An example is provided in Section 5.3 in order to investigate the theoretical results. Section 5.4 presents an alternative formulation to the restricted NP approach. Finally, extensions to more generic scenarios and concluding remarks are presented in Section 5.5.

5.1 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

$$\mathcal{H}_0 : \theta \in \Lambda_0 \quad , \quad \mathcal{H}_1 : \theta \in \Lambda_1 \quad (5.1)$$

where \mathcal{H}_i denotes the i th hypothesis and Λ_i is the set of possible parameter values under \mathcal{H}_i for $i = 0, 1$ [40]. 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 [65] and [66, Part III, Chapter VII] for discussions on the concept of uncertainty). For example, these distributions can be obtained as probability density function (p.d.f.) 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 [40]. 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 [68], [69]. 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 5.4 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 , $w_1(\theta)$.¹ In that respect, two extreme cases can be considered. In the first case, there is no uncertainty in $w_1(\theta)$. Then, the average

¹In accordance with these observations, the term uncertainty will be used to refer to uncertainties in $w_1(\theta)$ unless stated otherwise.

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 [64], [74]-[76]. 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 [68], [69]. 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 [42].

In practice, the prior distribution $w_1(\theta)$ is commonly estimated based on previous observations, and there exists some uncertainty 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 chapter, 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 [42]. 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:

$$P_D(\phi; \theta) \triangleq \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} , \quad \text{for } \theta \in \Lambda_1 \quad (5.2)$$

$$P_F(\phi; \theta) \triangleq \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} , \quad \text{for } \theta \in \Lambda_0 \quad (5.3)$$

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 [40]. Then, the restricted NP problem can be formulated as the following optimization problem:

$$\max_{\phi} \int_{\Lambda_1} P_D(\phi; \theta) w_1(\theta) d\theta \quad (5.4)$$

$$\text{subject to } P_D(\phi; \theta) \geq \beta , \quad \forall \theta \in \Lambda_1 \quad (5.5)$$

$$P_F(\phi; \theta) \leq \alpha , \quad \forall \theta \in \Lambda_0 \quad (5.6)$$

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 (5.5) 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:

$$\begin{aligned} \zeta &= \max_{\phi} \min_{\theta \in \Lambda_1} P_D(\phi; \theta) \\ &\text{subject to } P_F(\phi; \theta) \leq \alpha, \quad \forall \theta \in \Lambda_0. \end{aligned} \quad (5.7)$$

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 (5.4)-(5.6) to the max-min problem in (5.7). 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 (5.4) and (5.6), which can be expressed as

$$\begin{aligned} &\max_{\phi} P_D^{\text{avg}}(\phi) \\ &\text{subject to } P_F(\phi; \theta) \leq \alpha, \quad \forall \theta \in \Lambda_0 \end{aligned} \quad (5.8)$$

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.

5.2 Analysis of Restricted Neyman-Pearson Approach

In this section, the aim is to investigate the optimal solution of the restricted NP problem in (5.4)-(5.6). For this purpose, the definitions in (5.2) and (5.3) can be

used to reformulate the problem in (5.4)-(5.6) as follows:

$$\max_{\phi} \int_{\Gamma} \phi(\mathbf{x}) p_1(\mathbf{x}) d\mathbf{x} \quad (5.9)$$

$$\text{subject to } \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \geq \beta \quad (5.10)$$

$$\max_{\theta \in \Lambda_0} \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \leq \alpha \quad (5.11)$$

where $p_1(\mathbf{x}) \triangleq \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta$ defines the p.d.f. 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 (5.9)-(5.11) can be expressed as

$$\max_{\phi} \lambda \int_{\Gamma} \phi(\mathbf{x}) p_1(\mathbf{x}) d\mathbf{x} + (1 - \lambda) \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \quad (5.12)$$

$$\text{subject to } \max_{\theta \in \Lambda_0} \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \leq \alpha \quad (5.13)$$

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

5.2.1 Characterization of Optimal Decision Rule

Based on the formulation in (5.12) and (5.13), the following theorem provides a method to characterize the optimal solution of the restricted NP problem under certain conditions.

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

$$\int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) \mu(\theta) d\theta d\mathbf{x} = \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} , \quad (5.14)$$

then it is a solution of the problem in (5.12) and (5.13).

Proof: Please see Appendix 5.6.1.

Theorem 1 states that if one can find a p.d.f. $\mu(\theta)$ that satisfies the condition in (5.14), then the NP solution corresponding to $\lambda w_1(\theta) + (1 - \lambda) \mu(\theta)$ is

a solution of the restricted NP problem in (5.12) and (5.13). 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 [42] 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 [42]). In fact, those assumptions hold when a set of conditions specified in [47, 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 non-sequential [42]. 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:

$$\begin{aligned} & \max_{\phi} \int_{\Gamma} \phi(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x}) v(\theta) d\theta d\mathbf{x} \\ & \text{subject to } \max_{\theta \in \Lambda_0} \int_{\Gamma} \phi(\mathbf{x}) p_\theta(\mathbf{x}) d\mathbf{x} \leq \alpha \end{aligned} \quad (5.15)$$

where $v(\theta) = \lambda w_1(\theta) + (1 - \lambda) \mu(\theta)$. Based on the NP lemma [40], it can be shown that the solution of (5.15) is in the form of a likelihood ratio test (LRT);

that is,²

$$\phi^*(\mathbf{x}) = \begin{cases} 1, & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x}) v(\theta) d\theta > \eta p_{\tilde{\theta}_0}(\mathbf{x}) \\ \kappa(\mathbf{x}), & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x}) v(\theta) d\theta = \eta p_{\tilde{\theta}_0}(\mathbf{x}) \\ 0, & \text{if } \int_{\Lambda_1} p_\theta(\mathbf{x}) v(\theta) d\theta < \eta p_{\tilde{\theta}_0}(\mathbf{x}) \end{cases} \quad (5.16)$$

where $\eta \geq 0$ and $0 \leq \kappa(\mathbf{x}) \leq 1$ are such that $\max_{\theta \in \Lambda_0} P_F(\phi^*; \theta) = \alpha$, and $\tilde{\theta}_0$ is defined as

$$\tilde{\theta}_0 = \arg \max_{\theta \in \Lambda_0} P_F(\phi^*; \theta). \quad (5.17)$$

Therefore, the solution of the restricted NP problem in (5.12) and (5.13) can be expressed by the LRT specified in (5.16) and (5.17), once a p.d.f. $\mu(\theta)$ and the corresponding decision rule ϕ^* that satisfy the constraint in (5.14) are obtained (see Section 5.2.2). 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 (5.12) and (5.13) and that in (5.4)-(5.6).

Corollary 1: *Under the conditions in Theorem 1, ϕ^* solves the optimization problem in (5.4)-(5.6) when*

$$\min_{\theta \in \Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_\theta(\mathbf{x}) d\mathbf{x} = \beta.$$

Proof: According to Theorem 1, ϕ^* achieves the maximum value of the objective function in (5.12). That is, for any α -level decision rule ϕ (i.e., for any ϕ that satisfies (5.13)),

$$\begin{aligned} & \lambda \int_{\Gamma} \phi(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \lambda) \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi(\mathbf{x}) p_\theta(\mathbf{x}) d\mathbf{x} \\ & \leq \lambda \int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \lambda) \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_\theta(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (5.18)$$

²The proof follows from the observation that $(\phi^*(\mathbf{x}) - \phi(\mathbf{x})) \left(\int_{\Lambda_1} p_\theta(\mathbf{x}) v(\theta) d\theta - \eta p_{\tilde{\theta}_0}(\mathbf{x}) \right) \geq 0, \forall \mathbf{x}$, for any decision rule ϕ due to the definition of ϕ^* in (5.16). Then, the approach on page 24 of [40] can be used to prove that $\int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x}) v(\theta) d\theta d\mathbf{x} \geq \int_{\Gamma} \phi(\mathbf{x}) \int_{\Lambda_1} p_\theta(\mathbf{x}) v(\theta) d\theta d\mathbf{x}$ for any decision rule ϕ that satisfies $P_F(\phi; \theta) \leq \alpha, \forall \theta \in \Lambda_0$.

is satisfied. Since $\min_{\theta \in \Lambda_1} \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \geq \beta$ due to (5.5) and $\min_{\theta \in \Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} = \beta$ as stated in the corollary, $\int_{\Gamma} \phi(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x}$ should be smaller than or equal to $\int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x}$ in order for the inequality in (5.18) to hold. Equivalently, $\int_{\Lambda_1} P_D(\phi; \theta) w_1(\theta) d\theta \leq \int_{\Lambda_1} P_D(\phi^*; \theta) w_1(\theta) d\theta$ for any α -level decision rule ϕ , which proves that ϕ^* solves the optimization problem in (5.4)-(5.6). \square

Corollary 1 states that when the decision rule ϕ^* specified in Theorem 1 satisfies the constraint in (5.10) with equality, it also provides a solution of the restricted NP problem specified in (5.9)-(5.11); equivalently, in (5.4)-(5.6). 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 p.d.f.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 [40]. In the following theorem, an approach similar to that in [42] is taken in order to show that $v(\theta)$ in Theorem 1 corresponds to a least-favorable distribution.

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*

$$\tilde{v}(\theta) = \tilde{\lambda} w_1(\theta) + (1 - \tilde{\lambda}) \tilde{\mu}(\theta) \quad (5.19)$$

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

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

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

Proof: Please see Appendix 5.6.2.

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 $\tilde{\lambda} \geq \lambda$, only the case $\tilde{\lambda} = \lambda$ is of interest in practice since λ in (5.12) 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 the next section.

5.2.2 Calculation of Optimal Decision Rule

The analysis in Section 5.2.1 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 (5.2) as

$$\int_{\Lambda_1} \mu(\theta) P_D(\phi^*; \theta) d\theta = \min_{\theta \in \Lambda_1} P_D(\phi^*; \theta) . \quad (5.20)$$

This condition requires that $\mu(\theta)$ assigns non-zero probabilities only to the values of θ that result in the 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

$$\mu(\theta) = \delta(\theta - \theta_1) \quad (5.21)$$

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 (5.16) and (5.17).
2. Calculate

$$\theta_1^* = \arg \min_{\theta_1 \in \Lambda_1} f(\theta_1) \quad (5.22)$$

where

$$f(\theta_1) \triangleq \lambda \int_{\Lambda_1} w_1(\theta) P_D(\phi_{\theta_1}^*; \theta) d\theta + (1 - \lambda) P_D(\phi_{\theta_1}^*; \theta_1). \quad (5.23)$$

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 (5.23) is the average detection probability corresponding to $v(\theta) = \lambda w_1(\theta) + (1 - \lambda) \delta(\theta - \theta_1)$.³ Since 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 (5.22), $\phi_{\theta_1^*}^*$. 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

³It should be noted that λ is related to the design parameter β in (5.5) through Corollary 1. In addition, the fact that as λ increases (decreases), β decreases (increases) can be used to adjust the corresponding parameter value.

of (5.22), and in that case any solution of (5.22) 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

$$\mu(\theta) = \sum_{l=1}^N \nu_l \delta(\theta - \theta_l) \quad (5.24)$$

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 (5.24), the calculations in the algorithm should be updated as follows:

$$\boldsymbol{\vartheta}^* = \arg \min_{\boldsymbol{\vartheta}} f(\boldsymbol{\vartheta}) \quad (5.25)$$

where

$$f(\boldsymbol{\vartheta}) \triangleq \lambda \int_{\Lambda_1} w_1(\theta) P_D(\phi_{\boldsymbol{\vartheta}}^*; \theta) d\theta + (1 - \lambda) \sum_{l=1}^N \nu_l P_D(\phi_{\boldsymbol{\vartheta}}^*; \theta_l) \quad (5.26)$$

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 (5.25) 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) [51], [54], genetic algorithms and differential evolution [82], 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* [81] can be employed as in [48]. Specifically, $\mu(\theta)$ is expressed approximately by a linear combination of a number of window

functions as

$$\mu(\theta) \approx \sum_{l=1}^{N_w} \xi_l \varphi_l(\theta - \theta_l) , \quad (5.27)$$

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, (5.25) and (5.26) 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 5.3, 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 (5.22) and (5.23)). 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 (5.25) and (5.26). In this way, the complexity of the solution can be increased gradually until a solution is obtained.

Considering the computational complexity of the three-step algorithm proposed in this section, the first step involves the derivation of a generic NP decision rule as a function of θ_1 . In this derivation, the threshold of the test is obtained based on the likelihood ratio and the false-alarm constraint. Then, the expression for the detection probability can be obtained as a function of θ_1 . The exact number of operations in this step depends on the form of the probability density function of the observation. For example, in the simplest case, the likelihood ratio test can be reduced to a single threshold test. Then, the false-alarm and detection probabilities can be expressed in terms of the cumulative distribution functions (CDFs) of the observation under \mathcal{H}_0 and \mathcal{H}_1 , respectively. In the second step of the algorithm, a minimization problem needs to be solved in order to obtain the parameters of a candidate solution. The complexity of this step depends

on the number of point masses of the optimal solution (i.e., the number of minimizers of the detection probability $P_D(\phi^*; \theta)$ over $\theta \in \Lambda_1$). If a one point mass solution exists, a simple one-dimensional search leads to the candidate parameter for the optimal solution. However, if the solution has multiple, say N , point masses, then a linearly constrained minimization problem over a $2N$ dimensional space needs to be solved (see (5.25)). For convex cost functions, the solution can be obtained by interior-point methods, which are polynomial time in the worst case, and are very fast in practice. However, for nonconvex cost functions, global optimization techniques, such as PSO, need to be employed in order to obtain a solution. In that case, the computational complexity depends on the number of particles and iterations of the algorithm. Finally, the third step of the algorithm involves checking the minimum detection probability for the candidate solution obtained in the second step. This condition can be checked either by calculating the minimum value directly, or by first obtaining the possible minimum points via first order necessary conditions (taking first-order derivatives) and then by evaluating the detection probability at those points.

5.2.3 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 (5.4)-(5.6)). On the other hand, the classical NP approach in (5.8) 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 (5.7),

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 (5.4)-(5.6), (5.7) and (5.8), 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 (5.5)) 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

$$P_D^{\text{avg}}(\phi_r^\beta) = \int_{\Lambda_1} P_D(\phi_r^\beta; \theta) w_1(\theta) d\theta . \quad (5.28)$$

The discussions above imply that $P_D^{\text{avg}}(\phi_r^\beta)$ 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 $P_D^{\text{avg}}(\phi_r^\beta)$ for $\beta \in [L, U]$, the following theorem is presented.

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

Proof: Please see Appendix 5.6.3.

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 $P_D^{\text{avg}}(\phi_r^\beta)$. In other words, a unit decrease of β results in a smaller increase in the average detection probability for smaller values of β . Figure 5.1 in Section 5.3 provides an illustration of the results of Theorem 3.

5.3 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

$$\mathcal{H}_0 : X = V \quad , \quad \mathcal{H}_1 : X = \Theta + V \quad (5.29)$$

where $X \in \mathbb{R}$, Θ is an unknown parameter, and V is symmetric Gaussian mixture noise with the following p.d.f. $p_V(v) = \sum_{i=1}^{N_m} \omega_i \psi_i(v - m_i)$, where $\omega_i \geq 0$ for $i = 1, \dots, N_m$, $\sum_{i=1}^{N_m} \omega_i = 1$, and $\psi_i(x) = 1/(\sqrt{2\pi} \sigma_i) \exp(-x^2/(2\sigma_i^2))$ for $i = 1, \dots, N_m$. 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 $l = 1, \dots, \lfloor N_m/2 \rfloor$, 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 (5.29) is modeled as a random variable with a p.d.f. in the form of

$$w_1(\theta) = \rho \delta(\theta - A) + (1 - \rho) \delta(\theta + A) \quad (5.30)$$

where A is exactly known, but ρ is known with some uncertainty. With this model, the detection problem in (5.29) 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 [103]; 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 p.d.f. of X for a given value of $\Theta = \theta$ is expressed as

$$p_\theta(x) = \sum_{i=1}^{N_m} \frac{\omega_i}{\sqrt{2\pi} \sigma_i} \exp\left(\frac{-(x - \theta - m_i)^2}{2 \sigma_i^2}\right). \quad (5.31)$$

In order to obtain the optimal restricted NP decision rule for this problem, the algorithm in Section 5.2.2 is employed. First, it is assumed that $\mu(\theta)$ can be expressed as in (5.21); namely, $\mu(\theta) = \delta(\theta - \theta_1)$, where $\theta_1 \in \{-A, A\}$, and the algorithm is applied based on (5.22) and (5.23). 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 (5.24) with $N = 2$, $\nu_1 = 1 - \nu_2 = \tilde{\gamma}$, and $\theta_1 = -\theta_2 = A$). Note that this model includes all possible

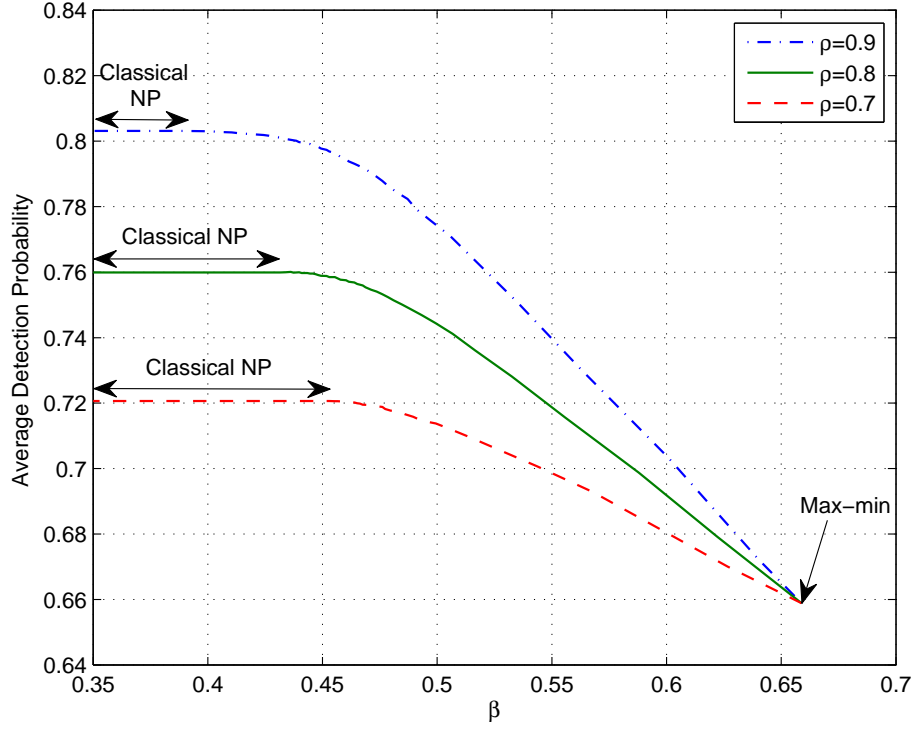


Figure 5.1: Average detection probability versus β for the classical NP, restricted NP, and max-min decision rules for $\rho = 0.7$, $\rho = 0.8$ and $\rho = 0.9$, where $A = 1$, $\sigma = 0.2$, and $\alpha = 0.2$.

p.d.f.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 (5.25) and (5.26) with $\boldsymbol{\vartheta} = \tilde{\gamma}$). 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 $i = 1, \dots, N_m$.

In Figure 5.1, 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 (5.30) are considered, and $A = 1$, $\sigma = 0.2$, and $\alpha = 0.2$ (see (5.6)) are used. As discussed in Section 5.2.3, 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.⁴ 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 (see (5.63)); hence, the x -axis in Figure 5.1 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 (5.30) becomes the least-favorable p.d.f. for $\rho = 0.5$. Figure 5.1 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.

⁴Although 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 5.2.3).

Table 5.1: 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. for $\rho = 0.9 / \rho = 0.8 / \rho = 0.7$	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

Finally, we would like to mention that Figure 5.1 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, Figure 5.1 can be used to choose a β value for the problem.

For the scenario in Figure 5.1, 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. For the max-min criterion, the least-favorable distribution $w_{lf}(\theta)$ in this example can be calculated as $w_{lf}(\theta) = 0.5 \delta(\theta - 1) + 0.5 \delta(\theta + 1)$. Table 5.1 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.

Figure 5.2 plots the average and minimum detection probabilities of the restricted NP decision rules versus λ in (5.12) 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 $0 \leq \lambda \leq 0.555$ for

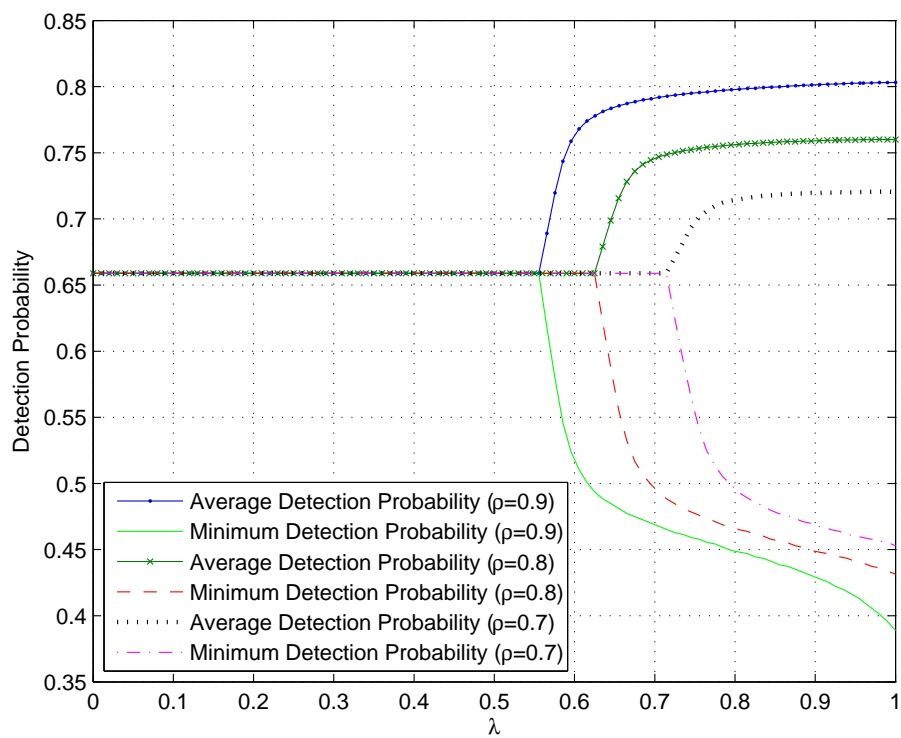


Figure 5.2: Average and minimum detection probabilities of the restricted NP decision rules versus λ for $\rho = 0.7$, $\rho = 0.8$ and $\rho = 0.9$, where $A = 1$, $\alpha = 0.2$ and $\sigma = 0.2$.

$\rho = 0.9$, when $0 \leq \lambda \leq 0.625$ for $\rho = 0.8$, and when $0 \leq \lambda \leq 0.714$ 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 (5.12) and (5.13). Another observation from Figure 5.2 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, Figure 5.2 shows that the difference between the average and the minimum detection probabilities increases as ρ increases.

Figure 5.3 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 (5.12). 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 Figure 5.3 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

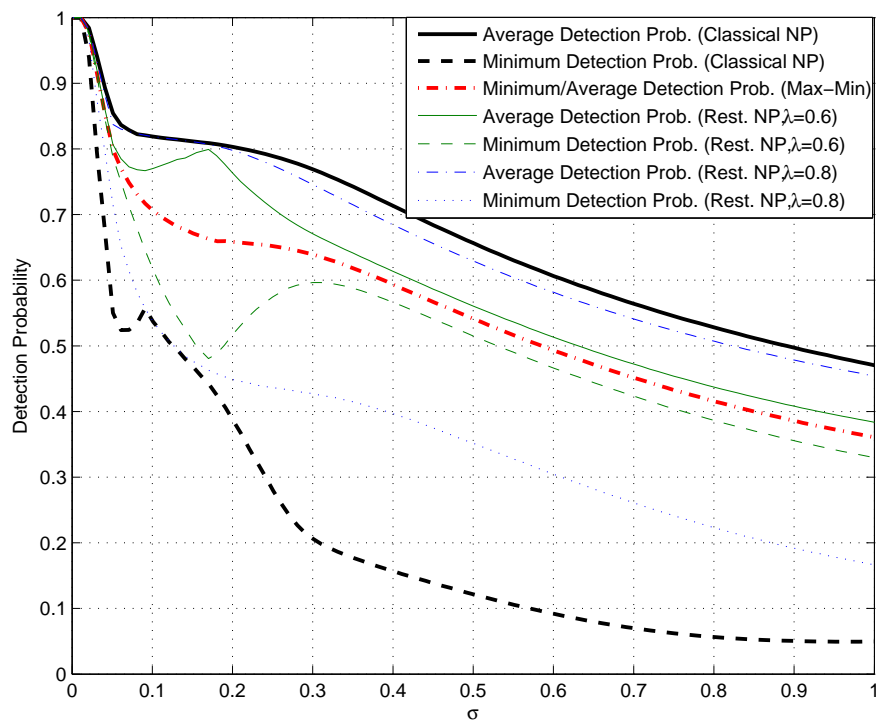


Figure 5.3: Average and minimum detection probabilities of the classical NP, max-min, and restricted NP (for $\lambda = 0.6$ and $\lambda = 0.8$) decision rules versus σ for $A = 1$, $\alpha = 0.2$, and $\rho = 0.9$.

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.⁵ 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 Figure 5.3, 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.

Finally, in Figure 5.4, 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.

5.4 Alternative Formulation

Although the formulation in (5.4)-(5.6) 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

⁵In 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.

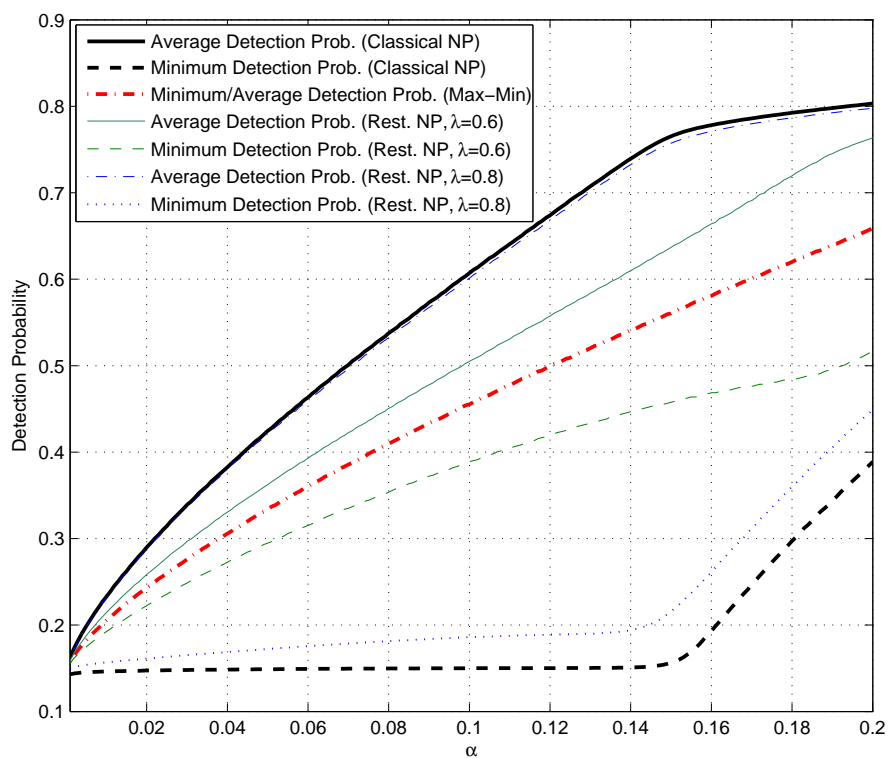


Figure 5.4: Average and minimum detection probabilities of the classical NP, max-min, and restricted NP (for $\lambda = 0.6$ and $\lambda = 0.8$) decision rules versus α for $A = 1$, $\sigma = 0.2$, and $\rho = 0.9$.

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 [104] and express the true prior distribution as $w_i^{\text{tr}}(\theta) = (1 - \varepsilon_i)w_i(\theta) + \varepsilon_i h_i(\theta)$ 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 p.d.f. 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 $w_i^{\text{tr}}(\theta)$ according to the ε -contaminated model above. Then, the following problem formulation can be considered:

$$\begin{aligned} & \max_{\phi} \min_{w_1^{\text{tr}}(\theta) \in \mathcal{W}_1} \int P_D(\phi; \theta) w_1^{\text{tr}}(\theta) d\theta \\ & \text{subject to } \max_{w_0^{\text{tr}}(\theta) \in \mathcal{W}_0} \int P_F(\phi; \theta) w_0^{\text{tr}}(\theta) d\theta \leq \alpha . \end{aligned} \quad (5.32)$$

Based on the ε -contaminated model, the problem in (5.32) can also be expressed from (5.2) and (5.3) as

$$\begin{aligned} & \max_{\phi} (1 - \varepsilon_1) \int \int \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + \varepsilon_1 \min_{h_1(\theta)} \int \int \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) h_1(\theta) d\theta d\mathbf{x} \\ & \text{subject to } \max_{h_0(\theta)} (1 - \varepsilon_0) \int \int \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) w_0(\theta) d\theta d\mathbf{x} \\ & \quad + \varepsilon_0 \int \int \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) h_0(\theta) d\theta d\mathbf{x} \leq \alpha . \end{aligned} \quad (5.33)$$

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

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

(5.33) becomes

$$\max_{\phi} (1 - \varepsilon_1) \int \phi(\mathbf{x}) p_1(\mathbf{x}) d\mathbf{x} + \varepsilon_1 \min_{\theta \in \Lambda_1} \int \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \quad (5.34)$$

$$\text{subject to } \max_{\theta \in \Lambda_0} \int \phi(\mathbf{x}) [(1 - \varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\theta}(\mathbf{x})] d\mathbf{x} \leq \alpha . \quad (5.35)$$

It is noted from (5.12)-(5.13) and (5.34)-(5.35) that the objective functions are in the same form but the constraints are somewhat different in the optimization problems considered in Section 5.2 and in this section. Since the proof of Theorem 1 focuses on the maximization of the objective function considering only the NP decision rules that satisfy the false-alarm constraint (see Appendix 5.6.1), the same proof applies to the problem in (5.34)-(5.35) as well if we consider the NP decision rules under the constraint in (5.35) 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):

$$\phi^*(\mathbf{x}) = \begin{cases} 1, & \text{if } \int_{\Lambda_1} p_{\theta}(\mathbf{x}) v(\theta) d\theta > \eta [(1 - \varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\tilde{\theta}_0}(\mathbf{x})] \\ \kappa(\mathbf{x}), & \text{if } \int_{\Lambda_1} p_{\theta}(\mathbf{x}) v(\theta) d\theta = \eta [(1 - \varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\tilde{\theta}_0}(\mathbf{x})] \\ 0, & \text{if } \int_{\Lambda_1} p_{\theta}(\mathbf{x}) v(\theta) d\theta < \eta [(1 - \varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\tilde{\theta}_0}(\mathbf{x})] \end{cases} \quad (5.36)$$

where $\eta \geq 0$ and $0 \leq \kappa(\mathbf{x}) \leq 1$ are such that

$$\max_{\theta \in \Lambda_0} \int \phi^*(\mathbf{x}) [(1 - \varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\theta}(\mathbf{x})] d\mathbf{x} = \alpha ,$$

and $\tilde{\theta}_0$ is defined as

$$\tilde{\theta}_0 = \arg \max_{\theta \in \Lambda_0} \int \phi^*(\mathbf{x}) [(1 - \varepsilon_0)p_0(\mathbf{x}) + \varepsilon_0 p_{\theta}(\mathbf{x})] d\mathbf{x} . \quad (5.37)$$

Hence, the solution of the problem in (5.34) and (5.35) can be expressed by the LRT specified in (5.36) and (5.37), once a p.d.f. $\mu(\theta)$ and the corresponding decision rule ϕ^* that satisfy the condition in Theorem 1 are obtained.

The problem formulation in (5.32) can also be regarded as an application of the Γ -minimax approach [64] to the NP framework, or as NP testing under interval probability [77], [78]. 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 hypothesis⁶. On the other hand, the restricted NP approach in (5.4)-(5.6) 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) [64]. 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 [105], 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.

5.5 Concluding Remarks and Extensions

In this chapter, 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 characterized theoretically, 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.

⁶Similarly, the prior distribution that maximizes the false alarm probability is considered for the null hypothesis.

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 [42], the notion of a restricted NP decision rule can be extended to cover more generic scenarios. Consider sets of distribution families $\Upsilon_0, \Upsilon_1, \dots, \Upsilon_M$ 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. In this scenario, the restricted NP formulation in (5.9)-(5.11) can be extended as follows:

$$\max_{\phi} \min_{w_1(\theta) \in \Upsilon_0} \int_{\Gamma} \phi(\mathbf{x}) \int p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} \quad (5.38)$$

$$\text{subject to } \min_{w_1(\theta) \in \Upsilon_i} \int_{\Gamma} \phi(\mathbf{x}) \int p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} \geq \beta_i, \quad i = 1, \dots, M \quad (5.39)$$

$$\max_{\theta \in \Lambda_0} \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \leq \alpha \quad (5.40)$$

where $\beta_1 > \dots > \beta_M$ specify the constraints on the worst-case detection probabilities in sets $\Upsilon_1, \dots, \Upsilon_M$, respectively. For this problem, the proof of Theorem 1 can be extended in a straightforward manner in order to obtain the following result:

Theorem 4: *Suppose that there exists a density $v(\theta) = \sum_{i=0}^M \lambda_i \mu_i(\theta)$, with $\lambda_i \geq 0$, $\sum_{i=0}^M \lambda_i = 1$, and $\mu_i(\theta) \in \Upsilon_i$, such that an α -level NP decision rule ϕ^* for $v(\theta)$ satisfies*

$$\int_{\Gamma} \phi^*(\mathbf{x}) \int p_{\theta}(\mathbf{x}) \mu_i(\theta) d\theta d\mathbf{x} = \min_{w_1(\theta) \in \Upsilon_i} \int_{\Gamma} \phi^*(\mathbf{x}) \int p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} = \beta_i \quad (5.41)$$

for $i = 1, 2, \dots, M$, and

$$\int_{\Gamma} \phi^*(\mathbf{x}) \int p_{\theta}(\mathbf{x}) \mu_0(\theta) d\theta d\mathbf{x} = \min_{w_1(\theta) \in \Upsilon_0} \int_{\Gamma} \phi^*(\mathbf{x}) \int p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x}. \quad (5.42)$$

Then ϕ^* solves the optimization problem in (5.38)-(5.40).

5.6 Appendices

5.6.1 Proof of Theorem 1

The proof is similar to the proof of Theorem 1 in [42]. Let ϕ be any α -level decision rule. Then,

$$\lambda \int_{\Gamma} \phi(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \lambda) \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \quad (5.43)$$

$$\leq \lambda \int_{\Gamma} \int_{\Lambda_1} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \lambda) \int_{\Lambda_1} \int_{\Gamma} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) \mu(\theta) d\mathbf{x} d\theta \quad (5.44)$$

since the second term in (5.43) is smaller than or equal to that in (5.44) due to the minimum operator. The expression in (5.44) can also be stated as

$$\int_{\Gamma} \int_{\Lambda_1} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) [\lambda w_1(\theta) + (1 - \lambda) \mu(\theta)] d\theta d\mathbf{x} = \int_{\Gamma} \int_{\Lambda_1} \phi(\mathbf{x}) p_{\theta}(\mathbf{x}) v(\theta) d\theta d\mathbf{x} \quad (5.45)$$

based on the definition of $v(\theta)$ in the theorem. Since ϕ^* is the NP decision rule for $v(\theta)$ under the false-alarm constraint in (5.13), the expression in (5.45) must be smaller than or equal to

$$\int_{\Gamma} \int_{\Lambda_1} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) v(\theta) d\theta d\mathbf{x} = \int_{\Gamma} \int_{\Lambda_1} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) [\lambda w_1(\theta) + (1 - \lambda) \mu(\theta)] d\theta d\mathbf{x} \quad (5.46)$$

(see Remark 1). After some manipulation, (5.46) can be expressed as

$$\lambda \int_{\Gamma} \int_{\Lambda_1} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \lambda) \int_{\Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) \mu(\theta) d\mathbf{x} d\theta \quad (5.47)$$

$$= \lambda \int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \lambda) \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \quad (5.48)$$

where the condition in (5.14) is employed in obtaining (5.48) from (5.47).

The arguments above indicate that the expression in (5.43) is always smaller than or equal to that in (5.48). Therefore, ϕ^* maximizes the objective function in (5.12) among all possible decision rules that satisfy the constraint in (5.13).

□

5.6.2 Proof of Theorem 2

In order to prove that $v(\theta)$ is the least-favorable distribution, we need to show that the average detection probability corresponding to $v(\theta)$ is smaller than or equal to that corresponding to $\tilde{v}(\theta)$ for any $\tilde{v}(\theta)$ specified in the theorem. The average detection probability corresponding to $v(\theta)$ is the average detection probability achieved by decision rule ϕ^* in Theorem 1, which can be expressed as

$$\begin{aligned} & \int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) v(\theta) d\theta d\mathbf{x} \quad (5.49) \\ &= \lambda \int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \lambda) \int_{\Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) \mu(\theta) d\mathbf{x} d\theta \\ &= \lambda \int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \lambda) \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \quad (5.50) \end{aligned}$$

where the condition (5.14) in Theorem 1 is used to obtain (5.50) from (5.49). Since $\int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta \geq \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x}$, the following relations can be obtained for any $\tilde{\lambda} \geq \lambda$:

$$\begin{aligned} & \int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) v(\theta) d\theta d\mathbf{x} \\ & \leq \tilde{\lambda} \int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \tilde{\lambda}) \min_{\theta \in \Lambda_1} \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} \quad (5.51) \\ & \leq \tilde{\lambda} \int_{\Gamma} \phi^*(\mathbf{x}) \int_{\Lambda_1} p_{\theta}(\mathbf{x}) w_1(\theta) d\theta d\mathbf{x} + (1 - \tilde{\lambda}) \int_{\Lambda_1} \tilde{\mu}(\theta) \int_{\Gamma} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} d\theta \quad (5.52) \end{aligned}$$

$$= \int_{\Gamma} \int_{\Lambda_1} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) \left[\tilde{\lambda} w_1(\theta) + (1 - \tilde{\lambda}) \tilde{\mu}(\theta) \right] d\theta d\mathbf{x} \quad (5.53)$$

$$= \int_{\Gamma} \int_{\Lambda_1} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) \tilde{v}(\theta) d\theta d\mathbf{x} \quad (5.54)$$

$$\leq \int_{\Gamma} \int_{\Lambda_1} \phi^*(\mathbf{x}) p_{\theta}(\mathbf{x}) \tilde{v}(\theta) d\theta d\mathbf{x} \quad (5.55)$$

where ϕ^* is the α -level NP solution corresponding to $\tilde{v}(\theta)$. It should be noted that the inequality between (5.51) and (5.52) is valid for any probability distribution $\tilde{\mu}(\theta)$. In addition, (5.55) is larger than or equal to (5.54) since ϕ^* is the α -level NP solution for $\tilde{v}(\theta)$ (see Remark 1).

From (5.51)-(5.55), it is observed that the average detection probability corresponding to $v(\theta)$ is smaller than or equal to that corresponding to $\tilde{v}(\theta) = \tilde{\lambda} w_1(\theta) + (1 - \tilde{\lambda}) \tilde{\mu}(\theta)$ for any $\tilde{\mu}(\theta)$ and $\tilde{\lambda} \geq \lambda$. \square

5.6.3 Proof of Theorem 3

Based on the definition of the restricted NP problem in (5.4)-(5.6), $P_D^{\text{avg}}(\phi_r^\beta)$ in (5.28) is a non-increasing 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 $P_D^{\text{avg}}(\phi_r^\beta)$, define a new decision rule as a randomization [40], [42] of two restricted NP decision rules as follows:

$$\phi \triangleq \varsigma \phi_r^{\beta_1} + (1 - \varsigma) \phi_r^{\beta_2} \quad (5.56)$$

where $0 \leq \beta_1 < \beta_2 \leq U$ 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:

$$P_D(\phi; \theta) = \varsigma P_D(\phi_r^{\beta_1}; \theta) + (1 - \varsigma) P_D(\phi_r^{\beta_2}; \theta), \quad \theta \in \Lambda_1 \quad (5.57)$$

$$P_F(\phi; \theta) = \varsigma P_F(\phi_r^{\beta_1}; \theta) + (1 - \varsigma) P_F(\phi_r^{\beta_2}; \theta), \quad \theta \in \Lambda_0. \quad (5.58)$$

The relation in (5.58) can be used to show that ϕ is an α -level decision rule. That is,

$$\max_{\theta \in \Lambda_0} P_F(\phi; \theta) \leq \varsigma \max_{\theta \in \Lambda_0} P_F(\phi_r^{\beta_1}; \theta) + (1 - \varsigma) \max_{\theta \in \Lambda_0} P_F(\phi_r^{\beta_2}; \theta) \leq \alpha \quad (5.59)$$

where (5.6) is used to obtain the second inequality.

Based on (5.56) and (5.57), the average detection probability of ϕ can be calculated as

$$P_D^{\text{avg}}(\phi) = \int_{\Lambda_1} P_D(\phi; \theta) w_1(\theta) d\theta = \varsigma P_D^{\text{avg}}(\phi_r^{\beta_1}) + (1 - \varsigma) P_D^{\text{avg}}(\phi_r^{\beta_2}). \quad (5.60)$$

Also, from (5.57), the worst-case detection probability of ϕ can be upper bounded as follows:

$$\min_{\theta \in \Lambda_1} P_D(\phi; \theta) \geq \varsigma \min_{\theta \in \Lambda_1} P_D(\phi_r^{\beta_1}; \theta) + (1 - \varsigma) \min_{\theta \in \Lambda_1} P_D(\phi_r^{\beta_2}; \theta) \geq \varsigma \beta_1 + (1 - \varsigma) \beta_2 . \quad (5.61)$$

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 (5.60) and (5.61) can be used to obtain the following inequalities:

$$P_D^{\text{avg}}(\phi_r^{\beta^*}) \geq P_D^{\text{avg}}(\phi_r^\beta) \geq P_D^{\text{avg}}(\phi) = \varsigma P_D^{\text{avg}}(\phi_r^{\beta_1}) + (1 - \varsigma) P_D^{\text{avg}}(\phi_r^{\beta_2}) \quad (5.62)$$

where the first inequality follows from the non-increasing property of $P_D^{\text{avg}}(\phi_r^\beta)$ explained at the beginning of the proof (since $\beta \geq \beta^*$ as shown in (5.61)), 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 $L < \beta < U$

$$\min_{\theta \in \Lambda_1} P_D(\phi_r^\beta; \theta) = \beta . \quad (5.63)$$

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 (5.63). 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 (5.63). 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 . \square

Chapter 6

Conclusions and Future Work

In this thesis, we have first analyzed noise enhanced detection in the restricted Bayesian framework, which covers Bayesian and minimax frameworks as special cases. We have also provided statistical characterization of optimal additive noise, and derived improvability and nonimprovability conditions. Secondly, we have investigated noise enhanced composite hypothesis-testing in the presence of partial prior information. Two criteria for evaluating noise enhancement have been proposed, and the structure of the optimal additive noise p.d.f. has been derived for each criterion. Also, extensions to the cases with unknown parameter distributions for some hypotheses have been discussed. Thirdly, noise enhanced binary composite hypothesis-testing has been studied in the NP framework. The previous studies on noise benefits for simple hypothesis-testing problems in the NP framework have been extended to composite hypothesis-testing problems. Optimal additive noise p.d.f.s have been derived, and improvability and nonimprovability conditions have been obtained. Finally, the restricted NP approach for composite hypothesis-testing in the presence of prior distribution uncertainty has been investigated. The restricted NP criterion is an application of the restricted Bayes approach (Hodges-Lehmann rule) to the NP framework. Algorithms have been proposed for the calculation of the optimal decision rule, and

the characteristics of the optimal decision rule have been investigated. Also, the properties of the average detection probability corresponding to restricted NP decision rules have been studied.

As a future work, noise enhanced detection can be studied in the restricted NP framework, and improvability and nonimprovability conditions can be investigated. Also, the restricted NP approach can be applied to the spectrum sensing problem in cognitive radio systems [106]. In addition, the study of noise enhanced detection in the restricted Bayesian framework can be extended to time varying scenarios, and adaptive noise enhancement algorithms can be obtained.

Bibliography

- [1] R. Benzi, A. Sutera, and A. Vulpiani, “The mechanism of stochastic resonance,” *J. Phys. A: Math. Gen.*, vol. 14, pp. 453–457, 1981.
- [2] P. Makra and Z. Gingl, “Signal-to-noise ratio gain in non-dynamical and dynamical bistable stochastic resonators,” *Fluctuat. Noise Lett.*, vol. 2, no. 3, pp. L145–L153, 2002.
- [3] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni, “Stochastic resonance,” *Rev. Mod. Phys.*, vol. 70, pp. 223–287, Jan. 1998.
- [4] G. P. Harmer, B. R. Davis, and D. Abbott, “A review of stochastic resonance: Circuits and measurement,” *IEEE Trans. Instrum. Meas.*, vol. 51, pp. 299–309, Apr. 2002.
- [5] K. Loerincz, Z. Gingl, and L. Kiss, “A stochastic resonator is able to greatly improve signal-to-noise ratio,” *Phys. Lett. A*, vol. 224, pp. 63–67, 1996.
- [6] I. Goychuk and P. Hanggi, “Stochastic resonance in ion channels characterized by information theory,” *Phys. Rev. E*, vol. 61, no. 4, pp. 4272–4280, 2000.
- [7] S. Mitaim and B. Kosko, “Adaptive stochastic resonance in noisy neurons based on mutual information,” *IEEE Trans. Neural Netw.*, vol. 15, pp. 1526–1540, Nov. 2004.

- [8] N. G. Stocks, “Suprathreshold stochastic resonance in multilevel threshold systems,” *Phys. Rev. Lett.*, vol. 84, pp. 2310–2313, Mar. 2000.
- [9] X. Godivier and F. Chapeau-Blondeau, “Stochastic resonance in the information capacity of a nonlinear dynamic system,” *Int. J. Bifurc. Chaos*, vol. 8, no. 3, pp. 581–589, 1998.
- [10] B. Kosko and S. Mitaim, “Stochastic resonance in noisy threshold neurons,” *Neural Netw.*, vol. 16, pp. 755–761, 2003.
- [11] B. Kosko and S. Mitaim, “Robust stochastic resonance for simple threshold neurons,” *Phys. Rev. E*, vol. 70, no. 031911, 2004.
- [12] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, “Theory of the stochastic resonance effect in signal detection: Part I—Fixed detectors,” *IEEE Trans. Sig. Processing*, vol. 55, pp. 3172–3184, July 2007.
- [13] A. Patel and B. Kosko, “Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection,” *IEEE Trans. Sig. Processing*, vol. 57, pp. 1655–1669, May 2009.
- [14] M. D. McDonnell and D. Abbott, “What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology,” *PLoS Computational Biology*, vol. 5, May 2009.
- [15] A. R. Bulsara and A. Zador, “Threshold detection of wideband signals: A noise-induced maximum in the mutual information,” *Physical Review E*, vol. 54, pp. R2185–R2188, Sep. 1996.
- [16] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, “Information-theoretic stochastic resonance in noise-floor limited systems: The case for adding noise,” *Phys. Rev. Lett.*, vol. 85, pp. 3369–3372, Oct. 2000.

- [17] P. Hanggi, “Stochastic resonance in biology how noise can enhance detection of weak signals and help improve biological information processing,” *ChemPhysChem*, vol. 3, pp. 285–290, Mar. 2002.
- [18] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, “Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance,” *Nature*, vol. 365, pp. 337–340, Sep. 1993.
- [19] L. Gammaitoni, E. Menichella-Saetta, S. Santucci, and F. Marchesoni, “Extraction of periodic signals from a noise background,” *Physics Letters A*, vol. 142, pp. 59–62, Dec. 1989.
- [20] A. R. Bulsara and L. Gammaitoni, “Tuning in to noise,” *Physics Today*, vol. 49, no. 3, pp. 39–45, 1996.
- [21] M. E. Inchiosa and A. R. Bulsara, “Signal detection statistics of stochastic resonators,” *Phys. Rev. E*, vol. 53, pp. R2021–R2024, Mar. 1996.
- [22] V. Galdi, V. Pierro, and I. M. Pinto, “Evaluation of stochastic-resonance-based detectors of weak harmonic signals in additive white Gaussian noise,” *Physical Review E*, vol. 57, pp. 6470–6479, June 1998.
- [23] S. M. Kay, J. H. Michels, H. Chen, and P. K. Varshney, “Reducing probability of decision error using stochastic resonance,” *IEEE Sig. Processing Lett.*, vol. 13, pp. 695–698, Nov. 2006.
- [24] S. M. Kay, “Can detectability be improved by adding noise?,” *IEEE Sig. Processing Lett.*, vol. 7, pp. 8–10, Jan. 2000.
- [25] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, “Theory of the stochastic resonance effect in signal detection: Part II–Variable detectors,” *IEEE Trans. Sig. Processing*, vol. 56, pp. 5031–5041, Oct. 2008.

- [26] S. Bayram and S. Gezici, “On the improvability and nonimprovability of detection via additional independent noise,” *IEEE Sig. Processing Lett.*, vol. 16, pp. 1001–1004, Nov. 2009.
- [27] X. Godivier and F. Chapeau-Blondeau, “Noise-assisted signal transmission in a nonlinear electronic comparator: Experiment and theory,” *Signal Processing*, vol. 56, pp. 293–303, Feb. 1997.
- [28] S. Zozor and P.-O. Amblard, “Stochastic resonance in discrete time nonlinear AR(1) models,” *IEEE Trans. Signal Process.*, vol. 47, pp. 108–122, Jan. 1999.
- [29] S. Bayram and S. Gezici, “Noise-enhanced M -ary hypothesis-testing in the minimax framework,” in *Proc. International Conference on Signal Processing and Commun. Systems*, (Omaha, Nebraska), pp. 31–36, Sep. 2009.
- [30] F. Chapeau-Blondeau, “Nonlinear test statistic to improve signal detection in non-gaussian noise,” *IEEE Sig. Processing Lett.*, vol. 7, pp. 205–207, July 2000.
- [31] S. Zozor and P. Amblard, “Can stochastic resonance be used in detection?,” in *Proc. of the 10th European Signal Processing Conference (EUSIPCO)*, (Tampere, Finland), Sep. 2000.
- [32] H. Chen, P. K. Varshney, J. H. Michels, and S. M. Kay, “Approaching near optimal detection performance via stochastic resonance,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, May 2006.
- [33] A. Asdi and A. Tewfik, “Detection of weak signals using adaptive stochastic resonance,” in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, (Detroit, Michigan), pp. 1332–1335, May 1995.
- [34] S. Zozor and P.-O. Amblard, “On the use of stochastic resonance in sine detection,” *Signal Process.*, vol. 7, pp. 353–367, Mar. 2002.

- [35] S. Zozor and P.-O. Amblard, “Stochastic resonance in locally optimal detectors,” *IEEE Trans. Signal Process.*, vol. 51, pp. 3177–3181, Dec. 2003.
- [36] A. A. Saha and G. Anand, “Design of detectors based on stochastic resonance,” *Signal Process.*, vol. 83, pp. 1193–1212, 2003.
- [37] D. Rousseau and F. Chapeau-Blondeau, “Stochastic resonance and improvement by noise in optimal detection strategies,” *Digital Signal Processing*, vol. 15, pp. 19–32, 2005.
- [38] H. Chen, P. K. Varshney, and J. H. Michels, “Improving sequential detection performance via stochastic resonance,” *IEEE Sig. Processing Lett.*, vol. 15, pp. 685–688, Dec. 2008.
- [39] H. Chen, P. K. Varshney, and J. H. Michels, “Noise enhanced parameter estimation,” *IEEE Trans. Sig. Processing*, vol. 56, pp. 5074–5081, Oct. 2008.
- [40] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.
- [41] E. L. Lehmann, *Testing Statistical Hypotheses*. New York: Chapman & Hall, 2 ed., 1986.
- [42] J. L. Hodges, Jr. and E. L. Lehmann, “The use of previous experience in reaching statistical decisions,” *The Annals of Mathematical Statistics*, vol. 23, pp. 396–407, Sep. 1952.
- [43] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*. New York: Wiley, 1954.
- [44] T. S. Ferguson, *Mathematical Statistics*. New York: Academic, 1967.
- [45] L. J. Savage, *The Foundations of Statistics*. New York: Dover, 2 ed., 1972.

- [46] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer, 2nd ed., 1985.
- [47] A. Wald, *Statistical Decision Functions*. John Wiley and Sons, 1950.
- [48] S. Bayram, S. Gezici, and H. V. Poor, “Noise enhanced hypothesis-testing in the restricted Bayesian framework,” *IEEE Trans. Sig. Processing*, vol. 58, pp. 3972–3989, Aug. 2010.
- [49] S. Bayram and S. Gezici, “Noise enhanced M-ary composite hypothesis-testing in the presence of partial prior information,” *IEEE Trans. Sig. Processing*, vol. 59, pp. 1292–1297, Mar. 2011.
- [50] S. Bayram and S. Gezici, “Effects of additional independent noise in binary composite hypothesis-testing problems,” in *Proc. International Conference on Signal Processing and Commun. Systems*, (Omaha, Nebraska), pp. 22–30, Sep. 2009.
- [51] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization method for constrained optimization problems*, pp. 214–220. IOS Press, 2002. in *Intelligent Technologies—Theory and Applications*.
- [52] A. I. F. Vaz and E. M. G. P. Fernandes, “Optimization of nonlinear constrained particle swarm,” *Baltic Journal on Sustainability*, vol. 12, no. 1, pp. 30–36, 2006.
- [53] S. Koziel and Z. Michalewicz, “Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization,” *Evolutionary Computation*, vol. 7, no. 1, pp. 19–44, 1999.
- [54] X. Hu and R. Eberhart, “Solving constrained nonlinear optimization problems with particle swarm optimization,” in *Proc. Sixth World Multiconference on Systemics, Cybernetics and Informatics 2002 (SCI 2002)*, (Orlando, FL), 2002.

- [55] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [56] E. L. Lehmann, “The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two?,” *Journal of the American Statistical Association*, vol. 88, pp. 1242–1249, Dec. 1993.
- [57] J. O. Berger, “Could Fisher, Jeffreys and Neyman have agreed on testing?,” *Statistical Science*, vol. 18, pp. 1–12, Feb. 2003.
- [58] W. E. Strawderman, “Minimaxity,” *Journal of the American Statistical Association*, vol. 95, pp. 1364–1368, Dec. 2000.
- [59] J. R. Blum and J. Rosenblatt, “On partial a priori information in statistical inference,” *The Annals of Mathematical Statistics*, vol. 38, no. 6, pp. 1671–1678, 1967.
- [60] H. Robbins, “An empirical Bayes approach to statistics.,” in *Proc. 3rd Berkeley Symposium on Math. Statist. and Probability 1*, (Berkeley), pp. 157–164, 1955.
- [61] H. Robbins, “The empirical bayes approach to statistical decision problems,” *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 1–20, 1964.
- [62] S. R. Watson, “On Bayesian inference with incompletely specified prior distributions,” *Biometrika*, vol. 61, pp. 193–196, 1974.
- [63] H. Kudo, “On partial prior information and property of parametric sufficiency,” in *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*, vol. 1, pp. 251–265, 1967.
- [64] M. J. Bayarri and J. O. Berger, “The interplay of Bayesian and frequentist analysis,” *Statistical Science*, vol. 19, pp. 58–80, 2004.

- [65] D. Ellsberg, “Risk, ambiguity, and the savage axioms,” *The Quarterly Journal of Economics*, vol. 75, pp. 643–669, Nov. 1961.
- [66] F. H. Knight, *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin, 1921.
- [67] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Part I*. New York, NY: John Wiley & Sons, Inc., 2 ed., 2001.
- [68] J. Cvitanic and I. Karatzas, “Generalized Neyman-Pearson lemma via convex duality,” *Bernoulli*, vol. 7, no. 1, pp. 79–97, 2001.
- [69] B. Rudloff and I. Karatzas, “Testing composite hypotheses via convex duality,” <http://arxiv.org/abs/0809.4297>, Sep. 2008.
- [70] W. Hoeffding, “Asymptotically optimal tests for multinomial distributions,” *The Annals of Mathematical Statistics*, vol. 36, pp. 369–400, 1965.
- [71] S. Natarajan, “Large deviations, hypotheses testing, and source coding for finite Markov chains,” *IEEE Trans. Inform. Theory*, vol. 31, pp. 360–365, May 1985.
- [72] O. Zeitouni and M. Gutman, “On universal hypotheses testing via large deviations,” *IEEE Trans. Inform. Theory*, vol. 37, pp. 285–290, Mar. 1991.
- [73] E. Levitan and N. Merhav, “A competitive Neyman-Pearson approach to universal hypothesis testing with applications,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 2215–2229, Aug. 2002.
- [74] E. L. Lehmann, “Some history of optimality,” *Lecture Notes-Monograph Series*, vol. 57, pp. 11–17, 2009.
- [75] 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 of the Econometric Society*, 2007.

- [76] J. O. Berger, B. Boukai, and Y. Wang, “Unified frequentist and Bayesian testing of a precise hypothesis,” *Statistical Science*, vol. 12, no. 3, pp. 133–160, 1997.
- [77] P. J. Huber and V. Strassen, “Minimax tests and the Neyman-Pearson lemma for capacities,” *The Annals of Statistics*, vol. 1, pp. 251–263, Mar. 1973.
- [78] 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,” *Journal of Statistical Planning and Inference*, vol. 105, pp. 149–173, June 2002.
- [79] S. Bayram and S. Gezici, “On the restricted Neyman-Pearson approach for composite hypothesis-testing in the presence of prior distribution uncertainty,” *IEEE Trans. Sig. Processing*, vol. 59, Sep. 2011.
- [80] S. M. Kay, “Noise enhanced detection as a special case of randomization,” *IEEE Sig. Processing Lett.*, vol. 15, pp. 709–712, 2008.
- [81] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2nd ed., 2000.
- [82] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. Springer, 2005.
- [83] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice Hall, Inc., 1998.
- [84] C. C. Pugh, *Real Mathematical Analysis*. New York: Springer-Verlag, 2002.
- [85] L. Huang and M. J. Neely, “The optimality of two prices: Maximizing revenue in a stochastic network,” in *Proc. 45th Annual Allerton Conference on Communication, Control, and Computing*, (Monticello, IL), Sep. 2007.

- [86] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1968.
- [87] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Boston, MA: Athena Specific, 2003.
- [88] V. Bhatia and B. Mulgrew, “Non-parametric likelihood based channel estimator for Gaussian mixture noise,” *Signal Processing*, vol. 87, pp. 2569–2586, Nov. 2007.
- [89] T. Erseghe and S. Tomasin, “Optimized demodulation for MAI resilient UWB W-PAN receivers,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, (Beijing, China), pp. 4867–4871, May 2008.
- [90] T. Erseghe, V. Cellini, and G. Dona, “On UWB impulse radio receivers derived by modeling MAI as a Gaussian mixture process,” *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2388–2396, June 2008.
- [91] S. Verdú, *Multiuser Detection*. 1st ed. Cambridge, UK: Cambridge University Press, 1998.
- [92] L. Reggiani and G. M. Maggio, “Rapid search algorithms for code acquisition in UWB impulse radio communications,” *IEEE J. Select. Areas Commun.*, vol. 23, pp. 898–908, May 2005.
- [93] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 4th ed., 2001.
- [94] Y. Chen and V. K. Dubey, “Ultrawideband source localization using a particle-swarm-optimized Capon estimator,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4, (Seoul, Korea), pp. 2825–2829, May 2005.
- [95] Y. Rahmat-Samii, D. Gies, and J. Robinson, “Particle swarm optimization (PSO): A novel paradigm for antenna designs,” *The Radio Science Bulletin*, vol. 305, pp. 14–22, Sep. 2003.

- [96] Z. Yangyang, J. Chunlin, Y. Ping, L. Manlin, W. Chaojin, and W. Guangxing, “Particle swarm optimization for base station placement in mobile communication,” in *Proc. IEEE International Conference on Networking, Sensing and Control*, vol. 1, pp. 428–432, May 2004.
- [97] J. Pugh, A. Martinoli, and Y. Zhang, “Particle swarm optimization for unsupervised robotic learning,” in *Proc. Swarm Intelligence Symposium (SIS)*, (Pasadena, California), pp. 92–99, June 2005.
- [98] D. Luengo, C. Pantaleon, I. Santamaria, L. Vielva, and J. Ibanez, “Multiple composite hypothesis testing: A competitive approach,” *Journal of VLSI Signal Processing*, vol. 37, pp. 319–331, 2004.
- [99] J. Skilling, “Prior probabilities,” *Synthese*, vol. 63, pp. 1–34, April 1985.
- [100] M. A. Richards, *Fundamentals of Radar Signal Processing*. USA: McGraw-Hill, Electronic Engineering Series, 2005.
- [101] A. Goldsmith, *Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005.
- [102] W. Jatmiko, K. Sekiyama, and T. Fukuda, “A PSO-based mobile sensor network for odor source localization in dynamic environment: Theory, simulation and measurement,” in *Proc. IEEE Congress on Evolutionary Computation*, (Vancouver, BC), pp. 1036–1043, July 2006.
- [103] I. Korn, J. P. Fonseka, and S. Xing, “Optimal binary communication with nonequal probabilities,” *IEEE Trans. Commun.*, vol. 51, pp. 1435–1438, Sep. 2003.
- [104] J. Berger and L. M. Berliner, “Robust Bayes and empirical Bayes analysis with ε -contaminated priors,” *The Annals of Statistics*, vol. 14, no. 2, pp. 461–486, 1986.

- [105] J. Berger et. al., “An overview of robust Bayesian analysis,” *Test*, vol. 3, no. 1, pp. 5–124, 1994.
- [106] H. Celebi, I. Guvenc, S. Gezici, and H. Arslan, “Cognitive-radio systems for spectrum, location, and environmental awareness,” *IEEE Antennas and Propagation Magazine*, vol. 52, pp. 41–61, Aug. 2010.