OPTIMAL STOCHASTIC APPROACHES FOR SIGNAL DETECTION AND ESTIMATION UNDER INEQUALITY CONSTRAINTS

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

By

Berkan Dülek

June 2012

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. Ahmet Enis Çetin (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.

Assoc. Prof. Dr. Uğur Güdükbay

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. Behçet Uğur Töreyin

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 the Graduate School of Engineering and Science

ABSTRACT

OPTIMAL STOCHASTIC APPROACHES FOR SIGNAL DETECTION AND ESTIMATION UNDER INEQUALITY CONSTRAINTS

Berkan Dülek
Ph.D. in Electrical and Electronics Engineering
Supervisors: Asst. Prof. Dr. Sinan Gezici
and Prof. Dr. Ahmet Enis Çetin
June 2012

Fundamental to the study of signal detection and estimation is the design of optimal procedures that operate on the noisy observations of some random phenomenon. For detection problems, the aim is to decide among a number of statistical hypotheses, whereas estimating certain parameters of the statistical model is required in estimation problems. In both cases, the solution depends on some goodness criterion by which detection (or estimation) performance is measured. Despite being a well-established field, the advances over the last several decades in hardware and digital signal processing have fostered a renewed interest in designing optimal procedures that take more into account the practical considerations. For example, in the detection of binary-valued scalar signals corrupted with additive noise, an analysis on the convexity properties of the error probability with respect to the transmit signal power has suggested that the error performance cannot be improved via signal power randomization/sharing under an average transmit power constraint when the noise has a unimodal distribution

(such as the Gaussian distribution). On the contrary, it is demonstrated that performance enhancement is possible in the case of multimodal noise distributions and even under Gaussian noise for three or higher dimensional signal constellations. Motivated by these results, in this dissertation we adopt a structured approach built on concepts called stochastic signaling and detector randomization, and devise optimal detection procedures for power constrained communications systems operating over channels with arbitrary noise distributions.

First, we study the problem of jointly designing the transmitted signals, decision rules, and detector randomization factors for an M-ary communications system with multiple detectors at the receiver. For each detector employed at the receiver, it is assumed that the transmitter can randomize its signal constellation (i.e., transmitter can employ stochastic signaling) according to some probability density function (PDF) under an average transmit power constraint. We show that stochastic signaling without detector randomization cannot achieve a smaller average probability of error than detector randomization with deterministic signaling for the same average power constraint and noise statistics when optimal maximum a-posteriori probability (MAP) detectors are employed in both cases. Next, we prove that a randomization between at most two MAP detectors corresponding to two deterministic signal vectors results in the optimal performance. Sufficient conditions are also provided to conclude ahead of time whether the correct decision performance can or cannot be improved by detector randomization.

In the literature, the discussions on the benefits of stochastic signaling and detector randomization are severely limited to the Bayesian criterion. Therefore, we study the convexity/concavity properties for the problem of detecting the presence of a signal emitted from a power constrained transmitter in the presence of additive Gaussian noise under the Neyman-Pearson (NP) framework. First, it is proved that the detection probability corresponding to the $\alpha-$level likelihood ratio test (LRT) is either concave or has two inflection points such that the function is concave, convex and finally concave with respect to increasing values of the signal power. Based on this result, optimal and near-optimal power sharing/randomization strategies are proposed for average and/or peak power constrained transmitters. Using a similar approach, the convexity/concavity properties of the detection probability are also investigated with respect to the jammer power. The results indicate that a weak Gaussian jammer should employ on-off time sharing to degrade the detection performance.

Next, the previous analysis for the NP criterion is generalized to channels with arbitrary noise PDFs. Specifically, we address the problem of jointly designing the signaling scheme and the decision rule so that the detection probability is maximized under constraints on the average false alarm probability and average transmit power. In the case of a single detector at the receiver, it is shown that the optimal solution can be obtained by employing randomization between at most two signal values for the on-signal and using the corresponding NP-type LRT at the receiver. When multiple detectors are available at the receiver, the optimal solution involves a randomization among no more than three NP decision rules corresponding to three deterministic signal vectors.

Up to this point, we have focused on signal detection problems. In the following, the trade-offs between parameter estimation accuracy and measurement device cost are investigateed under the influence of noise. First, we seek to determine the most favorable allocation of the total cost to measurement devices so that the average Fisher information of the resulting measurements is maximized for arbitrary observation and measurement statistics. Based on a recently proposed measurement device cost model, we present a generic optimization problem without assuming any specific estimator structure. Closed form expressions are obtained in the case of Gaussian observations and measurement noise.

Finally, a more elaborate analysis of the relationship between parameter estimation accuracy and measurement device cost is presented. More specifically, novel convex measurement cost minimization problems are proposed based on various estimation accuracy constraints assuming a linear system subject to additive Gaussian noise for the deterministic parameter estimation problem. Robust allocation of the total cost to measurement devices is also considered by assuming a specific uncertainty model on the system matrix. Closed form solutions are obtained in the case of an invertible system matrix for two estimation accuracy criteria. Through numerical examples, various aspects of the proposed optimization problems are compared. Lastly, the discussion is extended to the Bayesian framework assuming that the estimated parameter is Gaussian distributed.

Keywords: Detection, Stochastic Signaling, Detector Randomization, Probability of Error, Neyman-Pearson (NP), Convexity, Gaussian Noise, Multimodal Noise, Power Constraint, Jamming, Parameter Estimation, Measurement Cost, Cramer-Rao Bound (CRB), Wireless Sensor Networks (WSN).

ÖZET

EŞİTSİZLİK KISITLARI ALTINDA İŞARET SEZİMİ VE KESTİRİMİ İÇİN OPTİMAL STOKASTİK YAKLAŞIMLAR

Berkan Dülek
Elektrik ve Elektronik Mühendisliği, Doktora
Tez Yöneticileri: Yrd. Doç. Dr. Sinan Gezici
ve Prof. Dr. Ahmet Enis Çetin
Haziran 2012

İşaret sezimi ve kestirimi çalışmalarının temelinde, rasgele bir olaya ait gürültülü gözlemler üzerinde işlem gören optimal yöntemlerin tasarlanması yer almaktadır. Sezim problemlerinde amaç, bir takım istatistiksel hipotezler arasında karar verilmesi iken kestirim problemlerinde istatistiksel modele ait belirli parametrelerin kestirimi gerekmektedir. Her iki durumda da çözüm, sezim (veya kestirim) başarımının ölçüleceği bazı kriterlere dayanır. Sezim ve kestirim kuramının köklü bir alan olmasına karşın, son yıllarda sayısal işaret işleme ve donanım alanlarındaki gelişmeler, pratik hususların daha fazla dikkate alındığı optimal yöntemlerin tasarlanmasına olan ilgiyi artırmıştır. Örneğin, toplanır gürültü altında ikili sayıl işaretlerin sezimi konusunda, hata olasılığının verici işaret gücüne bağlı dışbükeylik özelliklerinin analizi sonucunda gürültünün tek doruklu dağılıma (Gauss dağılımı gibi) sahip olduğu durumlarda, hata başarımının işaret gücü rasgeleleştirme/paylaşımı yöntemiyle artırılamayacağı saptanmıştır. Öte yandan, gürültünün çok doruklu olduğu ya da Gauss gürültüsü altında üç veya yüksek boyutlu işaret yıldız kümelerinin kullanıldığı durumlarda başarımda artışların mümkün olduğu gösterilmiştir. Bu sonuçlardan hareketle tezde,

stokastik işaretleme ve sezici rasgeleleştirme kavramlarına dayanan yapısal bir yöntem izlenmekte ve genel gürültü dağılımına sahip kanallar üzerinde çalışan güç kısıtlı iletişim sistemleri için optimal sezim yöntemleri geliştirilmektedir.

İlk olarak, alıcıda birden çok sezicinin bulunduğu M-li iletişim sistemleri için gönderilen işaretlerin, karar kurallarının ve sezici rasgeleleştirme oranlarının ortak olarak tasarlanması problemine çalışılmaktadır. Alıcıdaki her bir sezici için vericinin işaret yıldız kümesini belirli bir olasılık yoğunluk fonksiyonuna (OYF) göre ortalama bir güç kısıtı altında rasgeleleştirebildiği varsayılmaktadır. (Yani verici, stokastik işaretleme uygulayabilmektedir.) Öncelikle, sezicilerde MAP kuralı kullanıldığında, aynı ortamala güç kısıtı ve gürültü istatistikleri altında stokastik işaretleme ile ulaşılan ortalama hata olasılığının, sezici rasgeleleştirme ile ulaşılan ortalama hata olasılığından daha düşük olamayacağı gösterilmektedir. Devamında, optimal başarıma deterministik işaretlemeyle çalışan en fazla iki MAP sezicisi arasındaki rasgeleleştirme ile ulaşılacağı kanıtlanmaktadır. Doğru karar başarımının sezici rasgeleleştirme yöntemi ile artırılıp artırılamayacağına önceden karar verebilmek maksadıyla yeterli koşullar belirtilmektedir.

Literatürde, stokastik işaretleme ve sezici rasgeleleştirme konusundaki çalışmalar Bayes kriteriyle sınırlı kalmıştır. Dolayısıyla bu bölümde Neyman-Pearson (NP) kriteri çerçevesinde, güç kısıtlı bir vericiden gönderilen işaretin sezimlenmesi probleminin dışbükeylik/içbükeylik özellikleri toplanır Gauss gürültüsü altında incelenmektedir. İlk olarak, α düzeyli olabilirlik oran sınamasına (OOS) ait sezim olasılığının ya işaret gücünün içbükey bir fonksiyonu ya da işaret gücünün artan değerleri için sırasıyla içbükey, dışbükey ve son olarak içbükey bir fonksiyonu olduğu kanıtlanmaktadır. Bu sonuç temelinde, ortalama ve tepe güç kısıtlı vericiler için optimal ve optimale yakın güç rasgeleleştirme/paylaşım stratejileri önerilmektedir. Benzer bir yöntemle, sezim olasılığının karıştırıcı gücü cinsinden dışbükeylik/içbükeylik özellikleri incelenmektedir. Sonuçlar, sezim olasılığını düşürmek için düşük güçlü Gauss karıştırıcıların aç-kapa zaman paylaşım yöntemini kullanması gerektiğini ortaya koymaktadır.

Ek olarak, bir önceki kısımda değinilen analiz herhangi bir gürültü OYF'sine sahip kanallar için genelleştirilmektedir. Daha açık bir deyişle, ortalama yanlış alarm olasılığı ve ortalama verici gücü kısıtları altında sezim olasılığının enbüyütülmesi amacıyla, işaretleme yöntemi ve sezici kuralının ortak tasarımı problemi ele alınmaktadır. Alıcıda tek bir sezicinin olduğu durumda, optimal çözüme var simgesi için en fazla iki işaret değeri arasında rasgeleleştirme yapılarak ve alıcıda buna karşılık gelen NP-türü OOS'nin kullanılmasıyla ulaşılacağı bildirilmektedir. Alıcıda birden çok sezicinin olduğu durumda ise optimal çözüm, deterministik işaretlere karşılık gelen en fazla üç NP-türü OOS arasında rasgeleleştirme uygulanarak elde edilmektedir.

Bu aşamaya kadar, işaret sezimi problemlerine odaklanılmaktadır. Tezin devamında ise, gürültü altında parametre kestiriminin doğruluğu ve ölçüm aygıtlarının maliyeti arasındaki ilişkiye değinmektedir. Ilk olarak genel gözlem ve ölçüm istatistikleri altında, ölçümlere ait ortalama Fisher bilgisinin enbüyütülmesi maksadıyla, toplam bütçenin ölçüm aygıtlarına en iyi dağıtımının belirlenmesi amaçlanmaktadır. Yakın zamanda önerilmiş bir ölçüm aygıtı maliyet modeline dayanılarak, belirli bir sezici yapısı varsayılmadan genel bir eniyileme problemi sunulmaktadır. Gauss dağılımlı gözlem ve ölçümlerin olduğu durumlarda çözüm için kapalı formda ifadeler elde edilmektedir.

Son olarak, parametre kestiriminin doğruluğu ve ölçüm aygıtlarının maliyeti arasındaki ilişkiye yönelik daha detaylı bir analiz sunulmaktadır. Daha açık bir deyişle, deterministik parametre kestirimi için Gauss gürültüsü altında çalışan doğrusal bir sistem varsayılarak, çeşitli kestirim doğruluğu kısıtlarına dayanan yeni dışbükey ölçüm aygıtı maliyeti enküçültme problemleri önerilmektedir. Toplam maliyetin ölçüm aygıtlarına gürbüz dağıtımı konusu da sistem matrisi

için belirli bir hata modeli ele alınarak incelenmektedir. Tersi alınabilen sistem matrisleri için iki kestirim doğruluğu kriteri altında kapalı formda çözümler elde edilmektedir. Sayısal örnekler üzerinden, önerilen eniyileme problemleri karşılaştırılmaktadır. Ek olarak, kestirilen parametrenin Gauss dağılımlı olduğu varsayılarak analizler Bayes çerçevesine taşınmaktadır.

Anahtar Kelimeler: Sezim, Stokastik İşaretleme, Sezici Rasgeleleştirme, Hata Olasılığı, Neyman-Pearson (NP), Dışbükeylik, Gauss Gürültüsü, Çok Doruklu Gürültü, Güç Kısıtı, Karıştırma, Parametre Kestirimi, Ölçüm Maliyeti, Cramer-Rao Sınırı (CRS), Telsiz Algılayıcı Ağlar (TAA).

Contents

List of Figures

List of Tables

To my parents

Ahmet and Fikriye,

and to my family

Bengi and Berin

1

Introduction

Nature uses only the longest threads to weave her patterns, so that each small piece of her fabric reveals the organization of the entire tapestry.

– Richard P. Feynman, *The Character of Physical Law*

The main work in this dissertation has originated from the recent application of the stochastic resonance (SR) theory to signal detection and estimation problems [1–18]. The idea is that the detection performance of suboptimal detectors can be enhanced by intentionally injecting randomized noise samples at the input of the receiver of a binary communications system. In this dissertation, we have been able to reflect these ideas into concepts called *stochastic signaling* and *detector randomization*, in which we show that correct decision performance over channels corrupted by multimodal noise distributions can be improved by jointly randomizing the transmitted signals and decision rules employed at the receiver. The main building block in our analysis is Carathéodory's theorem from convex analysis [19], which lets us notice that the optimal signal distributions are discrete with a certain maximum number of mass points. More interestingly, we have discovered that it is possible to increase the detection probability by a similar randomized signaling mechanism for the classical textbook example [20, Example II.D.1] of detecting the presence of a signal immersed in Gaussian noise under the Neyman-Pearson framework whenever the false alarm requirement is smaller than $\mathcal{Q}(2) \approx 0.02275$ as is the case in most practical applications ($\mathcal{Q}(\cdot)$ denotes the tail probability of the standard Gaussian random variable). The second part of the dissertation provides several results on the trade-offs between measurement device cost and parameter estimation accuracy in the presence of noise. We consider estimation scenarios based on noise corrupted observations with arbitrary distribution functions as well as a linear system model with Gaussian observations. The following sections introduce the context, describe the previous work in the literature, and summarize our contributions for both parts.

1.1 Stochastic Signaling and Detector Randomization in Power Constrained Communications Systems

In coherent detection applications, despite the ubiquitous restrictions on the transmission power, there is often some flexibility in the choice of signals transmitted over the communications medium [20]. Due to crosstalk limitation between adjacent wires and frequency blocks, wired systems require that the signal power should be carefully controlled [21]. A more pronounced example from wireless systems dictates the signal power to be limited both to conserve battery power and to meet restrictions by regulatory bodies. Optimal signaling and detector design in the presence of Gaussian noise has been studied extensively in the literature [20, 22, 23]. For a binary communications system corrupted by additive white Gaussian noise (AWGN) as shown in Figure 1.1 with equally likely priors and subject to individual average power constraints in the form of $\mathbb{E}\left\{|S_i|^2\right\} \leq \mathrm{A}$ for $i \in \{0, 1\}$, it is well-known that the average probability of error is minimized when deterministic antipodal signals (i.e., $S_1 = -S_0$) are utilized

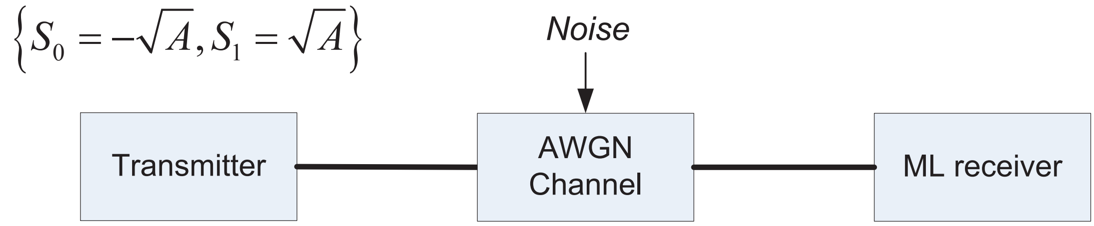

Figure 1.1: Antipodal signaling over AWGN channel for a binary communications system with equal priors and individual average power constraints.

at the power limit ($|S_1|^2 = |S_0|^2 = \mathrm{A}$), and a maximum likelihood (ML) decision rule (detector) is employed at the receiver [20].

In the case of vector observations immersed in additive zero-mean but colored Gaussian noise, it is shown that selecting the deterministic signals along the eigenvector of the covariance matrix of the Gaussian noise corresponding to the minimum eigenvalue maximizes the average correct decision probability of the binary communications system under same individual average power constraint [20]. Optimal deterministic signaling is investigated in [24] for nonequal prior probabilities under an average power constraint in the form of $\sum_{i=0}^{1} \pi_i \mathbb{E}\left\{|S_i|^2\right\} \leq \mathrm{A}$ instead of the individual power constraints for a scalar-valued binary communications system, when the noise is zero-mean Gaussian and the maximum *a posteriori* probability (MAP) decision rule is employed at the receiver. On-off keying is shown to be the optimal signaling strategy for coherent receivers when the signals have nonnegative correlation, and it is also optimal for noncoherent receivers employing envelope detection. As for coherent receivers and allowing for negative correlations, it is proven that the optimal performance is attained by maximizing the Euclidean distance between the signals under the given average power constraint. That is, $S_0 = -\sqrt{A}/\alpha$ and $S_1 = \alpha\sqrt{A}$, where $\alpha \triangleq \sqrt{\pi_0/\pi_1}$, and π_0 and π_1 denoting the prior probabilities. This is depicted in Figure 1.2.

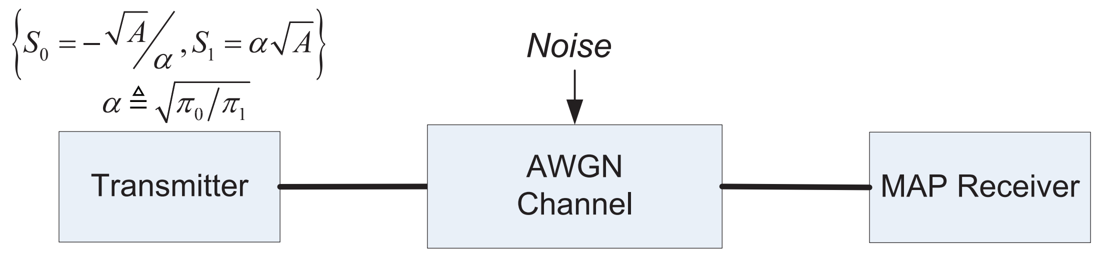

Figure 1.2: Optimal signaling over AWGN channel for a binary communications system with nonequal priors under an average power constraint.

Further insights are obtained by studying the convexity properties of error probability in [2] for the optimal detection of binary-valued scalar signals corrupted by additive noise under an average power constraint. It is shown that the average probability of error is a nonincreasing convex function of the signal power when the channel noise has a continuously differentiable unimodal noise probability density function (PDF) with finite variance. This discussion is extended from binary modulations to arbitrary signal constellations in [18] by concentrating on the maximum likelihood (ML) detection for AWGN channels. The symbol error rate (SER) is shown to be always convex in signal-to-noise ratio (SNR) for 1-D and 2-D constellations, but nonconvexity in higher dimensions at low to intermediate SNRs is possible, while convexity is always guaranteed at high SNRs with an odd number of inflection points in-between. When the transmitter is average power constrained, this result suggests the possibility of improving the error performance in high dimensional constellations through time sharing/randomization of the signal power, as opposed to the case for low dimensions (1-D and 2-D). This conclusion is illustrated in Figure 1.3 for the BPSK communications system given in Figure 1.1.

With the advent of the optimization techniques, there has been a renewed interest in designing randomized signaling schemes that improve/degrade (jamming problem) the error performance of communications systems operating under signal power constraints. Since performance gains in AWGN channels due to such

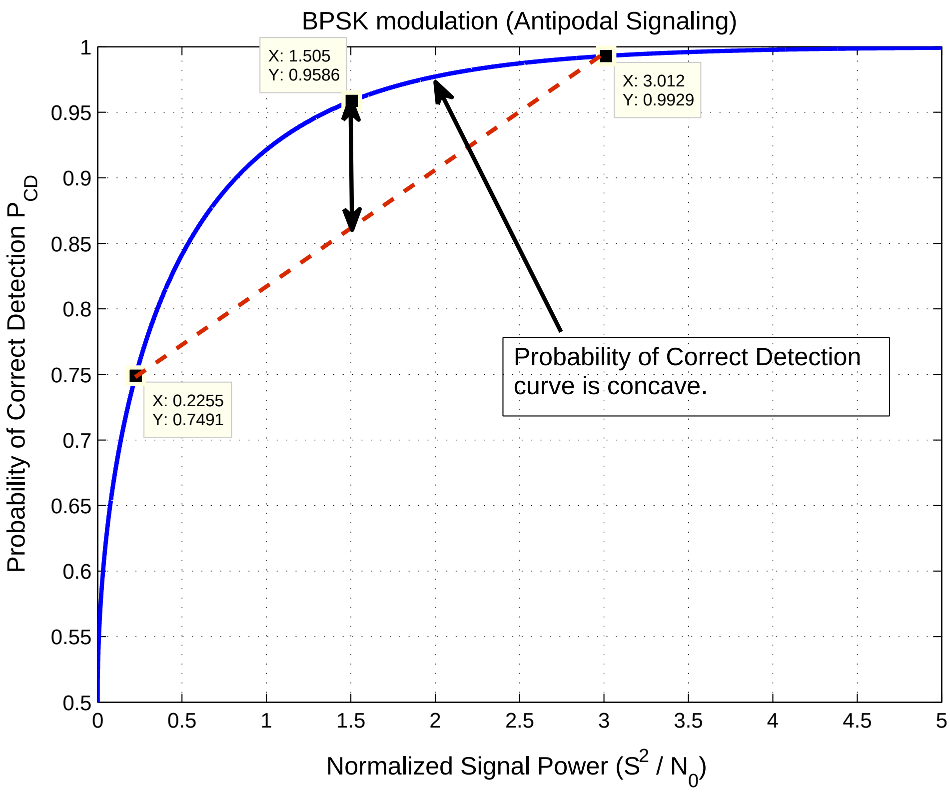

Figure 1.3: Probability of correct detection versus average signal power for the binary communications system given in Figure 1.1 operating over an AWGN channel. More generally, unimodal noise PDFs result in concave probability of correct detection curves.

stochastic approaches are restricted to higher dimensional constellations[1], the attempts to exploit the convexity properties of the error probability have been diverted towards channels with multimodal noise PDFs [25]. In practice, the noise can have significantly different probability distribution than the Gaussian distribution due to effects such as multiuser interference and jamming [26, 27].

In power constrained binary communications systems, stochastic signaling; that is, modeling signals for transmitted symbols as random variables instead of deterministic quantities, can provide performance improvements in terms of average probability of error (c.f., Figure 1.4). This method has proven effective in reducing the average probability of error for power constrained communications systems over additive noise channels with multimodal probability density

[1] 1-D and 2-D constellations are almost universally employed in practice.

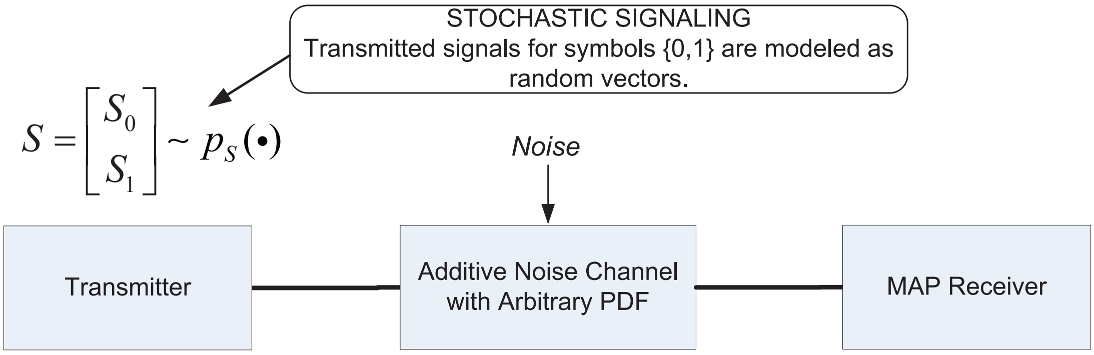

Figure 1.4: Stochastic signaling for a binary communications system operating over an additive noise channel with arbitrary noise PDF. Channel can assume multimodal noise PDF, e.g., Gaussian mixture noise.

functions [27]. It is shown in [25] that, for a given detector, an optimal stochastic signal can be represented by a randomization of no more than three different signal values under second and fourth moment constraints. Sufficient conditions are presented to determine whether stochastic signaling can help improve the correct decision performance over deterministic signaling methods. Joint optimization of signal structures and detectors in terms of error performance is investigated under an average power constraint in [17]. It is proven that the optimal performance can be achieved when the transmitted signal for each symbol is randomized between no more than two signal values and the corresponding MAP detector is employed at the receiver.

Another approach to improve the performance of communications systems over channels with multimodal noise PDFs is to perform *randomization among multiple detectors* [15, 28] as depicted in Figure 1.5. In that case, different detectors are employed at the receiver with certain probabilities. In [15], an average power constrained binary communications system is studied, and optimal randomization among antipodal signal pairs and the corresponding ML decision rules is investigated under the assumption that the receiver knows which deterministic pair is transmitted. It is concluded that randomization between at most two detectors is sufficient to maximize the correct decision probability. This result is illustrated in Figure 1.6 for the binary communications system given in

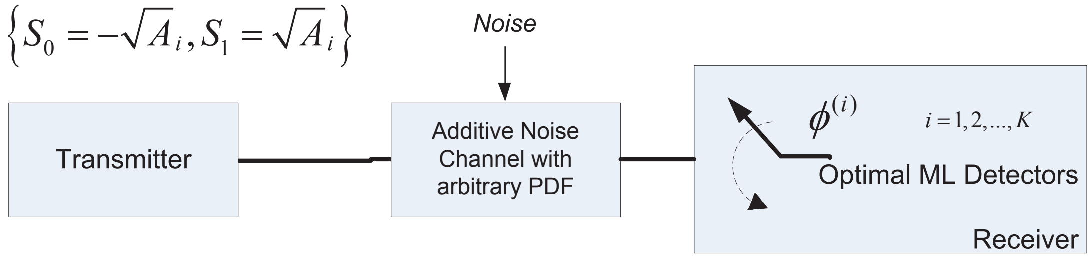

Figure 1.5: Detector randomization for a binary communications system operating over an additive noise channel with arbitrary noise PDF. Channel can assume multimodal noise PDF, e.g., Gaussian mixture noise.

Figure 1.5. In a related work, optimal additive noise components are studied for variable detectors in the context of stochastic resonance, and the optimal randomization between detector and additive noise pairs is investigated [10].

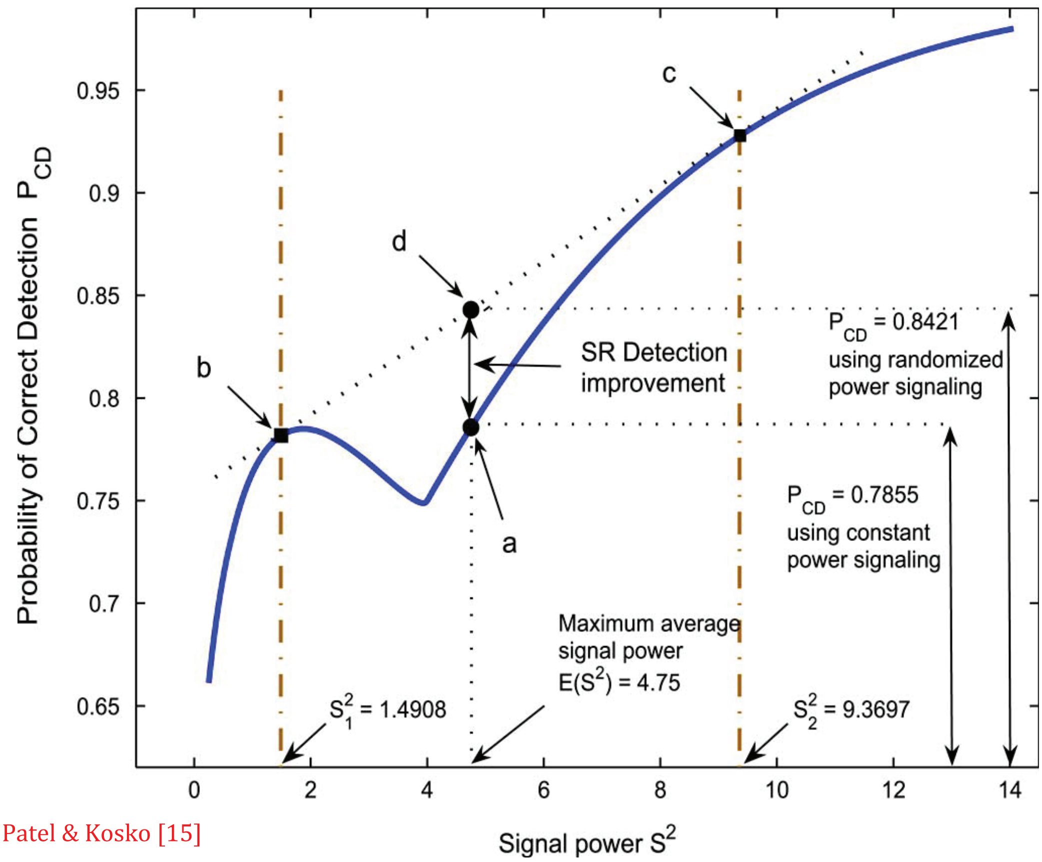

Figure 1.6: Detector randomization benefits for the binary communications system given in Figure 1.5 operating over an AWGN channel and employing antipodal signaling. Multimodal noise PDFs may result in detection performance improvements via signal strength randomization.

Similar theoretical approaches are adopted to tackle various problems in different research subjects. For example, in the context of wireless sensor networks (WSNs) the problem of pricing and transmission scheduling is examined for an access point in [29]. It is shown that an appropriate randomization between two business decisions and price pairs is sufficient to maximize time-average profit of the access point. In [30], through an information theoretic analysis, the worst-case noise distribution that maximizes the average probability of error and minimizes the channel capacity is found out to be a mixture of discrete lattices.

1.1.1 Performance Improvements under Minimum Probability of Error Criterion

Although the optimal design of stochastic signals and the corresponding MAP detector is analyzed in [17], and the optimal detector randomization and the corresponding MAP detectors are investigated in [15], no studies have considered the joint optimal design of detectors, stochastic signals, and detector randomization. Specifically, the study in [17] did not consider any detector randomization, and that in [15] assumed deterministic signals (no stochastic signaling). To that aim, in Chapter 2, both detector randomization and stochastic signaling are considered in a more generic formulation, and the problem of jointly optimizing detectors, stochastic signals, and detector randomization is addressed [31]. First, it is proven that stochastic signaling without detector randomization can never achieve a lower average probability of error than detector randomization with deterministic signaling for the same average power constraint and channel statistics. Then, based on this result and some additional analysis, the solution to the most generic optimization problem is obtained as the randomization between at most two MAP detectors corresponding to two deterministic signal vectors. Sufficient conditions for improvability and non-improvability of the correct decision performance via detector randomization are derived. Three detection examples are provided to compare various optimal and suboptimal signaling schemes.

1.1.2 Performance Improvements under Neyman-Pearson Criterion

Until recently, the discussions on the benefits of stochastic signaling were severely limited to the Bayesian formulation, specifically to the average probability of error criterion. Although the prior probabilities of the symbols are assumed to be equal in many communications systems, they can be unknown and nonequal in certain cases [24]. Furthermore, it may not be possible to impose cost structures on the decisions [20]. Under such scenarios, neither Bayesian nor minimax decision rules are applicable, and the Neyman-Pearson (NP) hypothesis testing provides a favorable alternative. For example, in WSN applications, a transmitter can send one bit of information (using on-off keying) about the presence of an event (e.g., fire), in which case the probabilities of detection and false alarm become the main performance metrics as in the NP approach.

In Chapter 3, we report an interesting and obviously overlooked fact for the problem of detecting the presence of a signal emitted from a power constrained transmitter operating over an additive Gaussian noise channel within the NP framework. Contrary to the average probability of error criterion [18], it is shown that for false alarm rates smaller than $Q(2)$, remarkable improvements in detection probability can be attained even in low dimensions by optimally distributing the fixed average power between two levels ($Q(\cdot)$ denotes the $Q-$function) [32]. More specifically, we study analytically the convexity/concavity properties of determining the presence of a power-limited signal immersed in additive Gaussian noise. It is proved that the detection probability corresponding to the $\alpha-$level likelihood ratio test (LRT) is either concave for $\alpha \geq Q(2)$ or has two inflection points such that the function is concave, convex and finally concave with respect to increasing values of the signal power for $\alpha < Q(2)$. Closed form expressions are provided to determine the regions over which power sharing/randomization enhances the detection performance over deterministic signaling at the average power level. In addition, the analysis is extended from scalar observations to multidimensional colored Gaussian noise corrupted signals. Based on the convexity/concavity results, optimal and near-optimal power sharing/randomization strategies are proposed for average/peak power constrained transmitters. For almost all practical applications, the required false alarm probability values are much smaller than $Q(2) \approx 0.02275$. As a consequence, power sharing can facilitate improved detection performance whenever the average power limitations are in the designated regions. Finally, the dual problem is considered from the perspective of a jammer to decrease the detection probability via power sharing/randomization. It is shown that the optimal strategy results in on-off jamming when the average noise power is below some critical value, a fact previously noted for spread spectrum communications systems [33].

In Chapter 4, we extend the discussion of the improvability of detection performance to channels with arbitrary noise PDFs under the NP framework. In the first part, the problem of designing the optimal signal distribution and the decision rule is addressed to maximize the detection probability without violating the constraints on the probability of false alarm and the average signal power. It is shown that the optimal solution can be obtained by randomizing between at most two signal vectors for the on-signal (symbol 1), and using the corresponding NP-type LRT at the receiver [34].

In the second part of Chapter 4, we investigate the same problem in the presence of multiple detectors at the receiver [35]. Specifically, we consider the joint optimal design of decision rules, stochastic signals, and detector randomization factors. Adopting a similar analysis strategy to that in Chapter 2, it is proven that the solution to the most generic optimization problem (i.e., employing both stochastic signaling and detector randomization) can be obtained as the randomization among no more than three NP decision rules corresponding to three deterministic signal vectors. As a result, the optimal parameters can be computed over a significantly reduced set instead of an infinite space of functions. Unfortunately even in that case, finding the optimal parameter set to maximize the detection probability may become a computationally cumbersome task necessitating the use of global optimization techniques.

1.2 Trade-offs between Measurement Device Cost and Estimation Accuracy

Although the statistical estimation problem in the presence of Gaussian noise is by far the most widely known and well-studied subject of estimation theory [20], approaches that consider the estimation performance jointly with system-resource constraints have become popular in recent years. Distributed detection and estimation problems took the first step by incorporating bandwidth and energy constraints due to data processing at the sensor nodes, and data transmission from sensor nodes to a fusion node in the context of WSNs [36]-[37]. Since then, the majority of the related studies have addressed the costs arising from similar system-level limitations with a relatively weak emphasis on the measurement costs due to amplitude resolution and dynamic range of the sensing apparatus. To begin with, we summarize the main aspects of the research that has been carried out in recent years to unfold the relationship between estimation capabilities and aforementioned costs of the sensing devices.

1.2.1 Related Work

In [36], detection problems are examined under a constraint on the expected cost resulting from measurement and transmission stages. It is found out that optimal detection performance can be achieved by a randomized on-off transmission scheme of the acquired measurements at a suitable rate. The distributed mean-location parameter estimation problem is considered in [38] for WSNs based on quantized observations. It is shown that when the dynamic range of the estimated parameter is small or comparable with the noise variance, a class of ML estimators exists with performance close to that of the sample mean estimator under stringent bandwidth constraint of one bit per sensor. When the dynamic range of the estimated parameter is comparable to or large than the noise variance, an optimum value for the quantization step results in the highest estimation accuracy possible for a given bandwidth constraint. In [39], a power scheduling strategy that minimizes the total energy consumption subject to a constraint on the worst mean-squared-error (MSE) distortion is derived for decentralized estimation in a heterogenous sensing environment. Assuming an uncoded quadrature amplitude modulation (QAM) transmission scheme and uniform randomized quantization at the sensor nodes, it is stated that depending on the corresponding channel quality, a sensor is either on or off completely. When a sensor is active, the optimal values for transmission power and quantization level for the sensor can be determined analytically in terms of the channel path losses and local observation noise levels.

In [40], distributed estimation of an unknown parameter is discussed for the case of independent additive observation noises with possibly different variances at the sensors and over nonideal fading wireless channels between the sensors and the fusion center. The concepts of estimation outage and estimation diversity are introduced. It is proven that the MSE distortion can be minimized under sum power constraints by turning off sensors transmitting over bad channels adaptively without degrading the diversity gain. In addition, performance decrease is reported when individual power constraints are also imposed at each sensor. In [37], the distributed estimation of a deterministic parameter immersed in uncorrelated noise in a WSN is targeted under a total bit rate constraint. The number of active sensors is determined together with the quantization bit rate of each active sensor in order to minimize the MSE.

The problem of estimating a spatially distributed, time-varying random field from noisy measurements collected by a WSN is investigated under bandwidth and energy constraints on the sensors in [41]. Using graph-theoretic techniques, it is shown that the energy consumption can be reduced by constructing reduced order Kalman-Bucy filters from only a subset of the sensors. In order to prevent degradation in the root-mean-squared (RMS) estimation error performance, efficient methods employing Pareto optimality criterion between the communication costs and RMS estimation error are presented. A power allocation problem for distributed parameter estimation is investigated under a total network power constraint for various topologies in [42]. It is shown that for the basic star topology, the optimal solution assumes either of the sensor selection, water-filling, or channel inversion forms depending on the measurement noise variance, and the corresponding analytical expressions are obtained. Asymptotically optimal power allocation strategies are derived for more complex branch, tree, and linear topologies assuming amplify-and-forward and estimate-and-forward transmission protocols. The decentralized WSN estimation is extended to incorporate the effects of imperfect data transmission from sensors to fusion center under stringent bandwidth constraints in [43].

Important results are also obtained for the sensor selection problem under various constraints on the system cost and estimation accuracy. The problem of choosing a set of k sensor measurements from a set of m available measurements so that the estimation error is minimized is addressed in [44] under a Gaussian assumption. It is shown that the combinatorial complexity of the solution can significantly be reduced without sacrificing much from the estimation accuracy by employing a heuristic based on convex optimization. In [45], a similar sensor selection problem is analyzed in a target detection framework when several classes of binary sensors with different discrimination performance and costs are available. Based on the conditional distributions of the observations at the fusion center, the performance of the corresponding optimal hypothesis tests is assessed using the symmetric Kullback-Leibler divergence. The solution of the resulting constrained maximization problem indicates that the sensor class with the best performance-to-cost ratio should be selected.

As outlined above, not much work has been performed, to the best of our knowledge, in the context of jointly designing the measurement stage from a cost-oriented perspective while performing estimation up to a predetermined level of accuracy. In other words, the trade-offs between measurement associated costs and estimation errors remain, to a large extent, undiscovered in the literature. On the other hand, if adopted, such an approach will inevitably require

- A general and reliable method of assessing the cost of measurements applicable to any real world phenomenon under consideration and
- An appropriate means of evaluating the best achievable estimation performance without reference to any specific estimator structure.

For the fulfilment of the first requirement, a novel measurement device model is proposed, and the problem of designing the optimal linear estimator and noise levels of measurement devices subject to a limited cost budget is addressed in [46]. Unlike previous studies, the cost of each device is determined with the accuracy of its measurements and expressed quantitatively in terms of the number of amplitude levels that can be resolved reliably. Intuitively, as the resolving power of a measurement device increases so does its cost. Furthermore, this method brings greater flexibility by enabling to work with variable precision over the acquired measurements. Based on this cost assignment scheme, the authors perform an optimization theoretic analysis to acquire the best measurements out of the observed quantities so that the estimation error is minimized for a given total cost constraint.

1.2.2 A Novel Measurement Device Cost Model

Before motivating our contributions to the estimation problem under cost constrained measurements, a brief overview of this novel measurement and cost-budget model is presented based on the discussions stated in [46]. Each measurement device is capable of sensing the value of a *scalar* physical quantity with some resolution in amplitude according to the measurement model

$$y = x + m\,, \tag{1.1}$$

where x denotes the observed random variable, m is the measurement noise associated with the employed measurement device, and y is the measurement value. Based on the measurement, the aim is to estimate the value of a (possibly random) parameter θ which is not directly accessible, but only accessible through the random variable x. It is assumed that m is a zero-mean random variable independent of x. As mentioned previously, the resolving power, specifically the number of amplitude levels that can be discriminated by the measurement device, solely determines the cost of each measurement under the proposed model. The dynamic range or scaling of the input to the measurement device is assumed to have no effect on the cost as long as the number of resolvable levels stays the same. In other words, range of the measurements does not contribute in assessing the cost of the measurements in this model. Under this scenario, the following quantitative expression is heuristically proposed in [46] to effectively determine the cost of making a single measurement

$$\mathrm{C} \triangleq \frac{1}{2}\log\left(1+\frac{\sigma_x^2}{\sigma_m^2}\right). \tag{1.2}$$

It is noted that the proposed cost function is stated in terms of the variances of the observation and measurement noise, which share the same motivations used by Hartley [47] to define the number of distinguishable signal levels at the receiver of an additive noise channel, and those of Shannon [48] to express the capacity of a Gaussian noise channel, where a message x is sent across a communications channel and is corrupted during transmission with additive Gaussian noise m.

For the sake of generality, it is stated in [46] that mutual information $I(x;y) = h(y) - h(m)$ can be employed as an alternative for the cost function proposed in (1.2) since it enables us to deal with non-Gaussian cases and helps to assess the value of a measurement more reasonably by revealing how many bits of information is actually conveyed in the measurement about the observed quantity. However, its computation requires explicit knowledge of the PDF $p_\theta(x)$ of the observed variable and may result in more involved formulation depending on the specific case under consideration (e.g., measurement noise with Gaussian mixture PDF). Moreover, when the measurement noise is Gaussian distributed with $\mathbb{E}[m^2] = \sigma_m^2$ independent of the observed signal distribution and simultaneously when there is an average power constraint on the observed signal variance as $\mathbb{E}[x^2] \leq \sigma_x^2$, the cost score obtained via (1.2) can be interpreted as a worst-case result. This is due to the fact that the Gaussian distribution maximizes the entropy over all distributions with the same variance which in turn maximizes $h(y)$, and finally $I(x;y)$. In other words, (1.2) becomes the solution of min-max problem $\min\limits_{p(m)}\max\limits_{p(x)} I(x;y)$ under the average power constraints mentioned above.[2] Therefore, by assuming that the errors introduced by the measurement devices are Gaussian distributed (an acceptable assumption), it is possible to handle a multitude of scenarios using the proposed cost function.

[2] A justification of this model from the viewpoint of economic theory is also presented in [46].

A deeper look into (1.2) reveals that it is a nonnegative, monotonically decreasing convex function of σ_m^2 for all $\sigma_x^2 > 0$ and $\sigma_m^2 > 0$ (in accordance with the properties of a valid rate-distortion function as mentioned below), and satisfies several properties that any meaningful cost function should possess as discussed in [46].

When the estimation is carried out using multiple (K) measurements, the mutual information between the actual random variable and its estimator can be upper bounded using the data-processing inequality [49] as $I(\theta, \hat{\theta}(\mathbf{y})) \leq I(\mathbf{x}, \mathbf{y})$. Assuming that the measurement noises are independent and using the properties of the joint entropy function, $I(\mathbf{x}, \mathbf{y}) \leq \sum_{i=1}^{K} I(x_i, y_i)$. Similar to the previous discussion, in the case of Gaussian measurement noises that are independent of the observed variables, $I(x_i, y_i)$'s are upper bounded with the cost function $(1/2)\log(1 + \sigma_{x_i}^2/\sigma_{m_i}^2)$. Therefore, when multiple observations are present, the total cost of measuring the observation vector $\mathbf{x}$ is defined in [46] as follows:

$$\mathrm{C_{tot}} \triangleq \sum_{i=1}^{K} \frac{1}{2} \log\left(1 + \frac{\sigma_{x_i}^2}{\sigma_{m_i}^2}\right). \tag{1.3}$$

As pointed out above and also by the authors of [46], the structure of the cost function reveals an immediate analogy with the results from the *rate-distortion* theory. More explicitly, an upper bound is imposed on the mutual information between the actual and estimated random variables $I(\theta, \hat{\theta}(\mathbf{y}))$ due to the data processing inequality and the total cost constraint $\mathrm{C_{tot}}$. This constraint can be interpreted as a *rate* constraint in the terminology of rate-distortion theory where the optimization problem can be cast as minimizing the average MSE distortion in the reconstruction of θ from a representation $\hat{\theta}(\mathbf{y})$ subject to the rate constraint $I(\theta, \hat{\theta}(\mathbf{y})) \leq \mathrm{C_{tot}}$. Then, the results from the rate-distortion theory manifests that for a given rate constraint $\mathrm{C_{tot}}$, it is not possible to reduce the MSE distortion denoted with $\|\theta - \hat{\theta}(\mathbf{y})\|_2^2$ beyond a certain value given by the corresponding *distortion-rate* function $D(\mathrm{C_{tot}})$. Finally, the above discussion generalizes in a straightforward manner when multiple parameters are estimated using multiple measurements without any change on the form of the proposed cost function in (1.3).

1.2.3 Average Fisher Information Metric for Scalar Parameter Estimation under Cost Constrained Measurements

Although the proposed model may lack in capturing the exact relationship between the cost and inner workings of any specific measurement hardware, it encompasses a sufficient amount of generality to remain useful under a multitude of circumstances. After formulating the measurement device model as outlined above, the optimal allocation of cost budget to the measurement devices is studied in [46] in order to minimize the estimation error, or equivalently in order to obtain the most favorable trade-off between the total cost and estimation accuracy. The estimation error is calculated by assuming that the observed variables are related to the unknown variables through a linear relation and for the estimation part, only linear minimum mean-squared-error (LMMSE) estimators are considered (as in the case of Wiener filtering problem in signal processing and channel equalization problem under intersymbol interference in communications systems). Although the proposed cost function is applicable to a wide variety of measurement problems with similar budget interpretations, the assumption of a linear relation between the observed and estimated (unknown) quantities and the restriction to an LMMSE estimator presents a major limitation against the generalization of similar analysis to a wider range of scenarios.

In estimation problems, the Cramer-Rao Bound (CRB) provides a lower bound on the MSEs of unbiased estimators. In addition, when the prior distribution of the estimated parameter is known, the Bayesian CRB (BCRB) can be calculated to obtain a lower bound on the MSE of any estimator [23]. The CRB and BCRB are quite useful in the analysis of estimation problems since

- They provide lower bounds that can be achieved (asymptotically) by certain estimators (e.g., MAP estimators),
- They are easier to calculate than the MSE as their formulations do not depend on the specific estimator structure under consideration.

Therefore, in this study we move beyond just minimizing the linear minimum mean-squared-error towards a more general performance metric: CRB for non-random parameter estimation and BCRB for random parameter estimation.

In Chapter 5, we focus on the scalar parameter estimation problem, and consider the problem of minimizing the BCRB (equivalently, maximizing the average Fisher information) at the outputs of measurement devices under the total cost constraint introduced in [46]. In other words, we propose a generic formulation for determining the optimal cost allocation among measurement devices in order to maximize the average Fisher information [50]. We also obtain a closed form expression for the Gaussian case, and present numerical examples.

1.2.4 Extension to Vector Parameter Estimation: Measurement Cost versus Estimation Accuracy

Although the optimal cost allocation problem is studied for the single parameter estimation case in [50] (also discussed in Chapter 5), and the signal recovery based on LMMSE estimators is investigated under cost constrained measurements using a linear system model in [46], no studies have analyzed the implications of the proposed measurement device model in a more general setting by considering both random and non-random parameter estimation under various estimation accuracy constraints and uncertainty in the linear system model. In Chapter 6, we propose novel measurement cost minimization problems under various constraints on estimation accuracy for a system characterized by a linear input-output relationship subject to Gaussian noise [51, 52]. For the measurement cost, we employ the recently proposed measurement device model in [46], and present a detailed treatment of the proposed measurement cost minimization problems. Main contributions of our study in Chapter 6 extend far beyond a multi-variate analysis of the discussion in Chapter 5, and can be summarized as follows:

- Formulated new convex optimization problems for the minimization of the total measurement cost by employing constraints on various estimation accuracy criteria (i.e., different functionals of the eigenvalues of the Fisher information matrix (FIM)) assuming a linear system model[3] in the presence of Gaussian noise.
- Studied system matrix uncertainty both from a general perspective and by employing a specific uncertainty model.
- Obtained closed form solutions for two of the proposed convex optimization problems in the case of invertible system matrix.
- Extended the results to the Bayesian estimation framework by treating the unknown estimated parameters as Gaussian distributed random variables.

In addition to the items listed above, simulation results are presented to discuss the theoretical results. Namely, we compare the performance of various estimation quality metrics through numerical examples using optimal and suboptimal cost allocation schemes, and simulate the effects of system matrix

[3]Such linear models have a multitude of application areas, a few examples of which are channel equalization, wave propagation, compressed sensing, and Wiener filtering problems [53, 54].

uncertainty. We also examine the behavior of the optimal solutions returned by various estimation accuracy criteria under scaling of the system noise variances, and identify the most robust criterion to variations in the average system noise power via numerical examples. The relationship between the number of effective measurements and the quality of estimation is also investigated under scaling of the system noise variances.

1.3 Organization of the Dissertation

This dissertation is organized as follows. Chapters 2, 3 and 4 are devoted to the analysis of how randomized signaling and detection approaches can help improve the performance of power constrained communications systems under Bayesian and Neyman-Pearson frameworks. In Chapter 2, optimal stochastic signaling and detector randomization is studied under an average transmit power constraint for the detection of vector-valued M-ary signals in arbitrary additive noise channels. In Chapter 3, the convexity/concavity properties of the detection probability are studied with respect to the transmitted signal and jammer power in the presence of additive Gaussian noise under the Neyman-Pearson framework. In Chapter 4, the analysis in the previous chapter is extended from the Gaussian case to noise channels with arbitrary distributions in the presence of single or multiple detectors at the receiver. Chapters 5 and 6 are devoted to the analysis of trade-offs between measurement device cost and estimation accuracy. In Chapter 5, the aim is to maximize the average Fisher information under a constraint on the total cost of measurement devices for arbitrary observation and measurement statistics. In Chapter 6, novel convex measurement cost minimization problems are proposed based on various estimation accuracy constraints for a linear system subject to additive Gaussian noise. Finally, Chapter 7 concludes this dissertation by providing an overall summary of the results along with some remarks on future work.

2

Detector Randomization and Stochastic Signaling for Minimum Probability of Error Receivers under Power Constraints

This chapter is organized as follows. Section 2.1 explains the optimal receiver design problem in the presence of stochastic signaling and detector randomization for an average power constrained M-ary communications system. The relation between the optimal error performances attainable by employing only stochastic signaling without detector randomization and only detector randomization with deterministic signaling is established. Furthermore, the optimal solution is provided in the form of an optimization problem. In Section 2.2, improvability and non-improvability of the correct decision performance via the optimal strategy is discussed. In Section 2.3, we present several methods for the numerical

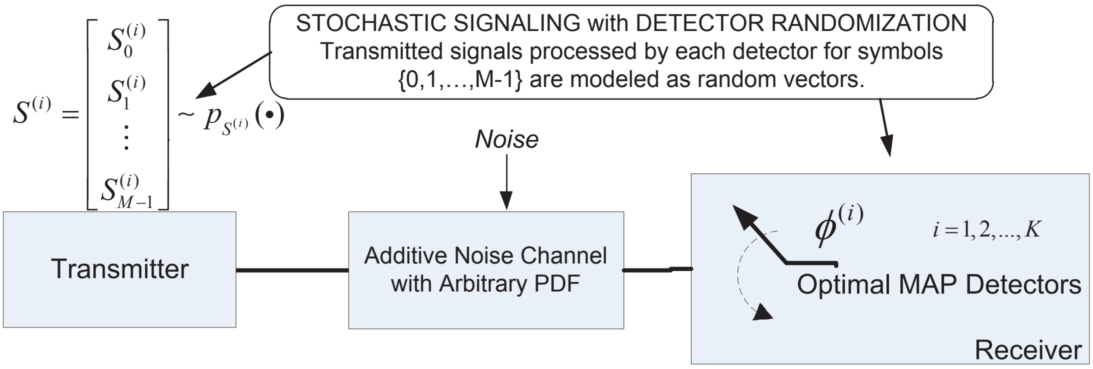

Figure 2.1: Stochastic signaling with detector randomization for an M-ary communications system operating over an additive noise channel with arbitrary noise PDF. Channel can assume multimodal noise PDF, e.g., Gaussian mixture noise.

solution of the optimization problem. Finally, numerical examples are given to corroborate theoretical results in Section 2.4.

2.1 Detector Randomization and Stochastic Signaling

Consider an M-ary communications system, in which the receiver acquires N-dimensional observations over an additive noise channel. The receiver is allowed to randomize or time-share among at most K different detectors (decision rules) to improve the detection performance, as shown in Figure 2.1. At any given time, only one of those K detectors can be employed at the receiver for the recognition of the transmitted symbol. The transmitter and the receiver are assumed to be synchronized in the sense that the transmitter knows which detector is currently in use at the receiver.[1] Furthermore, a stochastic signaling approach

[1] In practice, this can be achieved by employing a communications protocol that allocates the first $N_{s,1}$ symbols in the payload for detector 1, … , the last $N_{s,K}$ symbols for detector K. The information on the numbers of symbols for different detectors can be included in the header of a communications packet.

is adopted by treating the transmitted signals for each detector as random vectors. As investigated in [25] and [17] in the absence of detector randomization, employing stochastic signaling; that is, modeling signals for different symbols as random variables instead of deterministic quantities, can provide performance improvements in some scenarios.

Considering both detector randomization and stochastic signaling, the noisy observation vector $\mathbf{Y}$ received by the ith detector can be modeled as follows:

$$\mathbf{Y} = \mathbf{S}_j^{(i)} + \mathbf{N}\,, \quad j \in \{0, 1, \ldots, M-1\} \text{ and } i \in \{1, \ldots, K\}\,, \tag{2.1}$$

where $\mathbf{S}_j^{(i)}$ represents the transmitted signal vector for symbol j that is to be processed by detector i, and $\mathbf{N}$ is the noise component that is independent of $\mathbf{S}_j^{(i)}$. It should be emphasized that $\mathbf{S}_j^{(i)}$ is modeled as a random vector to facilitate stochastic signaling. Also, the prior probabilities of the symbols, represented by $\pi_0, \pi_1, \ldots, \pi_{M-1}$, are assumed to be known. In addition, although the signal model in (2.1) is in the form of a simple additive noise channel, it holds for flat-fading channels as well assuming perfect channel estimation [25].

At the receiver, K generic detectors (decision rules) are utilized to estimate the symbol specified in (2.1). That is, for a given observation vector $\mathbf{Y} = \mathbf{y}$, the ith detector $\phi^{(i)}(\mathbf{y})$ is described as

$$\phi^{(i)}(\mathbf{y}) = j\,, \quad \text{if } \mathbf{y} \in \Gamma_j^{(i)}\,, \tag{2.2}$$

for $j \in \{0, 1, \ldots, M-1\}$, where $\Gamma_0^{(i)}, \Gamma_1^{(i)}, \ldots, \Gamma_{M-1}^{(i)}$ form a partition of the observation space $\mathbb{R}^N$ for the ith detector [20]. The receiver can randomize among these K detectors in any manner in order to optimize its probability of error performance. Let v_i denote the randomization (or time-sharing) factor for detector $\phi^{(i)}$, where $\sum_{i=1}^{K} v_i = 1$ and $v_i \geq 0$ for $i = 1, \ldots, K$. Then, out of N_s symbols, $v_i N_\mathrm{s}$ of them are processed by detector $\phi^{(i)}$ for $i = 1, \ldots, K$.[2]

[2]It is assumed that $v_i N_\mathrm{s}$ is an integer for $i = 1, \ldots, K$. If not, the randomization factors can be achieved approximately. The approximation accuracy improves for larger N_s.

The aim of this study is to *jointly* optimize the randomization factors, detectors, and stochastic signals in order to achieve the minimum average probability of error, or equivalently, the maximum average probability of correct decision. The average probability of correct decision can be expressed as $\mathrm{P_c} = \sum_{i=1}^{K} v_i \, \mathrm{P_c^{(i)}}$, where v_i is the randomization factor for detector $\phi^{(i)}$, and $\mathrm{P_c^{(i)}}$ represents the corresponding probability of correct decision for that detector under M-ary signaling; that is,

$$\mathrm{P_c^{(i)}} = \sum_{j=0}^{M-1} \pi_j \int_{\Gamma_j^{(i)}} p_j^{(i)}(\mathbf{y}) \, \mathrm{d}\mathbf{y} \tag{2.3}$$

for $i = 1, 2, \ldots, K$, with $p_j^{(i)}(\mathbf{y})$ denoting the conditional probability density function (PDF) of the observation when the jth symbol that is to be received by the ith detector is transmitted. Since stochastic signaling is considered, $\mathbf{S}_j^{(i)}$ in (2.1) is modeled as a random vector. Recalling that the signals and the noise are independent, the conditional PDF of the observation can be calculated as $p_j^{(i)}(\mathbf{y}) = \int_{\mathbb{R}^N} p_{\mathbf{S}_j^{(i)}}(\mathbf{x}) \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) \, \mathrm{d}\mathbf{x} = \mathbb{E}\left\{ p_{\mathbf{N}}\big(\mathbf{y} - \mathbf{S}_j^{(i)}\big) \right\}$, where the expectation is over the PDF of $\mathbf{S}_j^{(i)}$. Then, the average probability of correct decision can be expressed as

$$\mathrm{P_c} = \sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j \, \mathbb{E}\left\{ p_{\mathbf{N}}\big(\mathbf{y} - \mathbf{S}_j^{(i)}\big) \right\} \, \mathrm{d}\mathbf{y} \right) . \tag{2.4}$$

In practical systems, there is a constraint on the average power emitted from the transmitter. Under the framework of stochastic signaling and detector randomization (or time-sharing), this constraint on the average power can be expressed in the following form [20]:

$$\sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \pi_j \, \mathbb{E}\left\{ \|\mathbf{S}_j^{(i)}\|_2^2 \right\} \right) \leq \mathrm{A} \; , \tag{2.5}$$

where A denotes the average power limit.

One of the main motivations behind this study is to understand how detector randomization and stochastic signaling can affect the error performance of an M-ary communications system. We look into the problem of jointly designing the optimal detectors, their randomization factors, and detector-specific signal PDFs employed at the transmitter in order to achieve the maximum (minimum) probability of correct decision (error) under the average power constraint given in (2.5). Mathematically stated, the optimization space is $\left\{\phi^{(i)},\ v_i,\ p_{\mathbf{S}_0^{(i)}},\ p_{\mathbf{S}_1^{(i)}},\ \ldots\ ,\ p_{\mathbf{S}_{M-1}^{(i)}}\right\}_{i=1}^{K}$ and the aim is to solve the following problem:

$$\begin{aligned}
\max_{\left\{\phi^{(i)}, v_i, p_{\mathbf{S}_0^{(i)}}, p_{\mathbf{S}_1^{(i)}}, \ldots, p_{\mathbf{S}_{M-1}^{(i)}}\right\}_{i=1}^{K}} \quad & \sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \int_{\Gamma_j^{(i)}} \pi_j\ \mathbb{E}\left\{p_{\mathbf{N}}\big(\mathbf{y}-\mathbf{S}_j^{(i)}\big)\right\} \mathrm{d}\mathbf{y} \right) \\
\text{subject to} \quad & \sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \pi_j\ \mathbb{E}\left\{\big\|\mathbf{S}_j^{(i)}\big\|_2^2\right\}\right) \leq \mathrm{A}\,, \\
& \sum_{i=1}^{K} v_i = 1\,, \quad v_i \geq 0\ , \quad \forall\, i \in \{1, 2, \ldots, K\}\,.
\end{aligned} \tag{2.6}$$

Note that there are also implicit constraints in the optimization problem in (2.6), since each $p_{\mathbf{S}_j^{(i)}}(\cdot)$ represents a PDF. Namely, $p_{\mathbf{S}_j^{(i)}}(\mathbf{x}) \geq 0\,,\ \forall \mathbf{x} \in \mathbb{R}^N$, and $\int_{\mathbb{R}^N} p_{\mathbf{S}_j^{(i)}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$ should also be satisfied $\forall\ j \in \{0, 1, \ldots, M-1\}$ and $\forall\ i \in \{1, \ldots, K\}$ by the optimal solution.

For a given detector i and the corresponding signal PDFs, $p_{\mathbf{S}_0^{(i)}},\ p_{\mathbf{S}_1^{(i)}},\ \ldots\ ,\ p_{\mathbf{S}_{M-1}^{(i)}}$, the conditional probability of observation $\mathbf{y}$ under hypothesis j (i.e., when symbol j is transmitted) is given by $p_j^{(i)}(\mathbf{y}) = \mathbb{E}\left\{p_{\mathbf{N}}\big(\mathbf{y}-\mathbf{S}_j^{(i)}\big)\right\}$. When deciding among M symbols based on observation $\mathbf{y}$, the MAP decision rule selects symbol k if $k = \underset{j\in\{0,1,\ldots,M-1\}}{\arg\max}\ \pi_j\, p_j^{(i)}(\mathbf{y})$, and it maximizes the average probability of correct decision [20]. Therefore, it is not necessary to search over all decision rules in (2.6); only the MAP decision rule should be determined for each detector and its corresponding average probability of correct decision should be considered [17]. The average probability of correct decision for a generic decision rule is given in (2.3). Using the decision region for the MAP detector; i.e., $\Gamma_j^{(i)} = \{\mathbf{y} \in \mathbb{R}^N \mid \pi_j\, p_j^{(i)}(\mathbf{y}) \geq \pi_l\, p_l^{(i)}(\mathbf{y})\,,\ \forall l \neq j\}$, the average probability of correct decision for detector i becomes

$$
\begin{aligned}
\mathrm{P}_{\mathrm{c,\,MAP}}^{(i)} &= \int_{\mathbb{R}^N} \max_{j\,\in\{0,\,1,\,\dots\,,\,M-1\}} \left\{ \pi_j \; p_j^{(i)}(\mathbf{y}) \right\} \,\mathrm{d}\mathbf{y} \\
&= \int_{\mathbb{R}^N} \max_{j\,\in\{0,\,1,\,\dots\,,\,M-1\}} \left\{ \pi_j \; \mathbb{E}\left\{ p_{\mathbf{N}}(\mathbf{y} - \mathbf{S}_j^{(i)}) \right\} \right\} \mathrm{d}\mathbf{y}\,. \qquad (2.7)
\end{aligned}
$$

Then, the optimal design problem in (2.6) can be stated as

$$
\begin{aligned}
\max_{\left\{ v_i,\, p_{\mathbf{S}_0^{(i)}},\, p_{\mathbf{S}_1^{(i)}},\, \dots\,,\, p_{\mathbf{S}_{M-1}^{(i)}} \right\}_{i=1}^{K}} \quad & \sum_{i=1}^{K} v_i \int_{\mathbb{R}^N} \max_{j\in\{0,\,1,\,\dots\,,\,M-1\}} \left\{ \pi_j \; \mathbb{E}\left\{ p_{\mathbf{N}}\big(\mathbf{y} - \mathbf{S}_j^{(i)}\big) \right\} \right\} \mathrm{d}\mathbf{y} \\
\text{subject to} \quad & \sum_{i=1}^{K} v_i \left(\sum_{j=0}^{M-1} \pi_j \; \mathbb{E}\left\{ \big\| \mathbf{S}_j^{(i)} \big\|_2^2 \right\} \right) \le \mathrm{A}\,, \\
& \sum_{i=1}^{K} v_i = 1\,, \quad v_i \ge 0\;, \quad \forall\, i \in \{1, 2, \dots, K\}\,. \qquad (2.8)
\end{aligned}
$$

It is noted that the optimization space is considerably reduced compared to that in (2.6) since there is no need to search over the detectors in (2.8).

The main idea behind considering only the MAP decision rules in evaluating the maximum average probability of correct decision in the optimization problem in (2.6) is that for any given set of stochastic signals and randomization factors, the maximum average probability of correct decision is achieved when the MAP rules are employed. Therefore, the optimization problem in (2.6) can be solved considering only the MAP decision rules which result in (2.8) via (2.7).

Another way to explain this approach can be as follows: Assume that the solution of the optimization problem in (2.6) is given by $\left\{ \hat{\phi}^{(i)},\ \hat{v}_i,\ \hat{p}_{\mathbf{S}_0^{(i)}},\ \hat{p}_{\mathbf{S}_1^{(i)}},\ \dots\ ,\ \hat{p}_{\mathbf{S}_{M-1}^{(i)}} \right\}_{i=1}^{K}$ where the decision rules $\hat{\phi}^{(i)}$'s are not MAP rules. Then, one can always achieve an equal or larger average probability of correct decision if he/she replaces $\hat{\phi}^{(i)}$'s with $\hat{\phi}_{\mathrm{MAP}}^{(i)}$'s, where $\hat{\phi}_{\mathrm{MAP}}^{(i)}$ denotes the MAP decision rule corresponding to $\hat{p}_{\mathbf{S}_0^{(i)}},\ \hat{p}_{\mathbf{S}_1^{(i)}},\ \dots\ ,\ \hat{p}_{\mathbf{S}_{M-1}^{(i)}}$. Hence, the optimal solution of (2.6) can always be obtained by considering the MAP decision rules only. Although this approach does not guarantee that the obtained solution is the unique one, it guarantees that the solution is optimal; that is, the largest average probability of correct decision is always achieved.

It is also noted that although the original optimization problem in (2.6), which performs an optimization over all possible detectors, is not the same as the simplified optimization problem in (2.8), which considers only the MAP decision rules, they are guaranteed to achieve the same maximum average probability of correct decision. Hence, the simplified problem in (2.8) can be considered instead of (2.6) to obtain the optimal solution.

The formulation in (2.8) generalizes the previous studies in the literature and covers them as special cases. For example, for $K = 1$ (i.e., no detector randomization), it reduces to the problem in [17] (hence, $K \geq 2$ is considered in this study). On the other hand, when deterministic signals are considered; that is, $p_{\mathbf{S}_j^{(i)}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_j^{(i)})$, $\forall i, j$, and when $M = 2$ (binary modulation), the problem in (2.8) reduces to that in [15].

The optimization problem in (2.8) provides a generic formulation that is valid for any noise PDF, and it is difficult to solve in general as the optimization needs to be performed over a space of signal PDFs. Let $\mathrm{P}_\mathrm{c}^\dagger$ denote the maximum average probability of correct decision obtained as the solution of the optimization problem in (2.8). To provide a simpler formulation of this problem, an upper bound on $\mathrm{P}_\mathrm{c}^\dagger$ will be derived first, and then the achievability of that bound will be investigated. To that aim, the following proposition is presented first.

Proposition 2.1.1 (Stochastic Signaling vs. Detector Randomization)**.** *Considering the same average power constraint and the same statistics for the additive noise, stochastic signaling (without detector randomization) can never achieve a larger average probability of correct decision than detector randomization (without stochastic signaling) when optimal MAP detectors are employed in both cases.*

Proof. Consider an M-ary communications system in which the transmitter employs stochastic signaling and the receiver uses the corresponding MAP rule for detection (no detector randomization is performed). Suppose that the transmitted signal for each symbol is characterized with the PDF $p_{\mathbf{X}_j}(\cdot)$, $\forall j \in \{0, 1, \ldots, M-1\}$. From (2.7), the average probability of correct decision for this system is given by

$$\mathrm{P}_{\mathrm{c,\,MAP}} = \int_{\mathbb{R}^N} \max_{j \in \{0,1,\ldots,M-1\}} \left\{ \pi_j \, \mathbb{E}_{\mathbf{X}_j} \{ p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_j) \} \right\} \mathrm{d}\mathbf{y}\,, \tag{2.9}$$

where the subscript of the expectation operator denotes that the expectation is taken with respect to the PDF of the corresponding random vector. The transmitted signals for all the M symbols can be expressed as the elements of a random vector $\mathbf{X}$ as follows: $\mathbf{X} \triangleq [\mathbf{X}_0 \; \mathbf{X}_1 \cdots \mathbf{X}_{M-1}] \in \mathbb{R}^{MN}$, where $\mathbf{X}_j$'s are N dimensional row vectors $\forall j \in \{0, 1, \ldots, M-1\}$. Then, the following inequality follows directly from the definitions of the 'max' and 'expectation' operations:

$$\begin{aligned} &\max_{j \in \{0,1,\ldots,M-1\}} \left\{ \pi_j \, \mathbb{E}_{\mathbf{X}_j} \{ p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_j) \} \right\} \\ &\qquad\qquad \leq \quad \mathbb{E}_{\mathbf{X}} \left\{ \max_{j \in \{0,1,\ldots,M-1\}} \{ \pi_j \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_j) \} \right\}. \end{aligned} \tag{2.10}$$

To see this, let $k = \underset{j \in \{0,1,\ldots,M-1\}}{\arg\max} \left\{ \pi_j \, \mathbb{E}_{\mathbf{X}_j} \{ p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_j) \} \right\}$ without loss of generality. Then,

$$\begin{aligned} \pi_k \, \mathbb{E}_{\mathbf{X}_k} \{ p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_k) \} &= \mathbb{E}_{\mathbf{X}} \{ \pi_k \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_k) \} \\ &\leq \mathbb{E}_{\mathbf{X}} \left\{ \max_{j \in \{0,1,\ldots,M-1\}} \{ \pi_j \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_j) \} \right\}. \end{aligned} \tag{2.11}$$

From (2.9) and (2.10), it is observed that

$$\begin{aligned} &\int_{\mathbb{R}^N} \max_{j \in \{0,1,\ldots,M-1\}} \left\{ \pi_j \, \mathbb{E}_{\mathbf{X}_j} \{ p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_j) \} \right\} \mathrm{d}\mathbf{y} \\ &\qquad\qquad \leq \; \mathbb{E}_{\mathbf{X}} \Bigg\{ \underbrace{\int_{\mathbb{R}^N} \max_{j \in \{0,1,\ldots,M-1\}} \{ \pi_j \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{X}_j) \} \, \mathrm{d}\mathbf{y}}_{\triangleq \, \mathrm{F}(\mathbf{X})} \Bigg\}. \end{aligned} \tag{2.12}$$

Looking more closely at (2.12), it is observed that $\mathrm{F}(\mathbf{x})$ represents the average probability of correct decision when the deterministic signal vector $\mathbf{x}$ is used for the transmission of M symbols over the additive noise channel and the corresponding MAP detector is employed at the receiver. Then, $\mathbb{E}_{\mathbf{X}}\{F(\mathbf{X})\}$ can be interpreted as a randomization (or time-sharing) among MAP detectors. The exact number of MAP detectors is determined by the number of distinct values that the random vector $\mathbf{X}$ can take.[3] Hence, assuming same average power constraint (see (2.8)), average probability of correct decision obtained by stochastic signaling with PDF $p_{\mathbf{X}}(\cdot)$ is always smaller than or equal to that of deterministic signaling and detector randomization according to the same PDF. $\square$

Similarly to the proof of Proposition 2.1.1, we can express the transmitted signals for all the M symbols that are to be received by detector i as the elements of a random vector: $\mathbf{S}^{(i)} \triangleq \left[\mathbf{S}_0^{(i)}\, \mathbf{S}_1^{(i)} \, \ldots \, \mathbf{S}_{M-1}^{(i)}\right] \in \mathbb{R}^{MN}$, where $\mathbf{S}_j^{(i)}$'s are N dimensional row vectors $\forall j \in \{0, 1, \ldots, M-1\}$. Then, the result in Proposition 2.1.1 can be employed to obtain a new optimization problem that provides an upper bound on the problem in (2.8). Specifically, instead of stochastic signals, consider detector randomization among deterministic signal values according to the joint signal PDF. Then, the inequality in (2.12) can be applied to the objective function in (2.8), and the following optimization problem can be obtained:

$$
\begin{aligned}
\max_{\left\{v_i,\, p_{\mathbf{S}^{(i)}}\right\}_{i=1}^{K}} \quad & \sum_{i=1}^{K} v_i \,\mathbb{E}\left\{ \int_{\mathbb{R}^N} \max_{j\in\{0,1,\ldots,M-1\}} \left\{ \pi_j \, p_{\mathbf{N}}\big(\mathbf{y} - \mathbf{S}_j^{(i)}\big)\right\} \, \mathrm{d}\mathbf{y} \right\} \\
\text{subject to} \quad & \sum_{i=1}^{K} v_i \,\mathbb{E}\left\{ \sum_{j=0}^{M-1} \pi_j \, \big\|\mathbf{S}_j^{(i)}\big\|_2^2 \right\} \leq \mathrm{A} \\
& \sum_{i=1}^{K} v_i = 1\,, \quad v_i \geq 0\,, \quad \forall\, i \in \{1, 2, \ldots, K\}
\end{aligned} \tag{2.13}
$$

where the expectations are taken with respect to the PDFs of $\mathbf{S}^{(i)}$'s. Proposition 2.1.1 implies that the solution to this optimization problem provides an upper bound on $\mathrm{P}_{\mathrm{c}}^{\dagger}$, which denotes the solution to the optimization problem in (2.8).

[3] In fact, a randomization among two MAP detectors is always sufficient in practice since optimal stochastic signals can be represented by a randomization of at most two different signal values under an average power constraint [17]. In other words, for any stochastic signal PDF, a corresponding discrete probability distribution with at most two mass points can be obtained, and the corresponding MAP detector randomization can be performed according to that distribution.

In order to achieve further simplification of the problem in (2.13), define $p_{\tilde{\mathbf{S}}}(\tilde{\mathbf{s}}) \triangleq \sum_{i=1}^{K} v_i \, p_{\mathbf{S}^{(i)}}(\tilde{\mathbf{s}})$, where $\tilde{\mathbf{s}} \triangleq [\tilde{\mathbf{s}}_0 \; \tilde{\mathbf{s}}_1 \; \cdots \; \tilde{\mathbf{s}}_{M-1}] \in \mathbb{R}^{MN}$. Since $\sum_{i=1}^{K} v_i = 1$, $v_i \geq 0$, $\forall i$ and $p_{\mathbf{S}^{(i)}}(\cdot)$'s are valid PDFs on $\mathbb{R}^{MN}$, $p_{\tilde{\mathbf{S}}}(\tilde{\mathbf{s}})$ satisfies the conditions to be a PDF. Then, the optimization problem in (2.13) can be written in the following equivalent form:

$$\begin{aligned} \max_{p_{\tilde{\mathbf{S}}}} \quad & \mathbb{E}\Bigg\{ \underbrace{\int_{\mathbb{R}^N} \max_{j \in \{0,1,\dots,M-1\}} \Big\{ \pi_j \, p_{\mathbf{N}}(\mathbf{y} - \tilde{\mathbf{S}}_j) \Big\} \, \mathrm{d}\mathbf{y}}_{\triangleq\, G(\tilde{\mathbf{S}})} \Bigg\} \\ \text{subject to} \quad & \mathbb{E}\Bigg\{ \underbrace{\sum_{j=0}^{M-1} \pi_j \, \big\|\tilde{\mathbf{S}}_j\big\|_2^2}_{\triangleq\, H(\tilde{\mathbf{S}})} \Bigg\} \leq \mathrm{A} \end{aligned} \tag{2.14}$$

where the expectations are taken with respect to $p_{\tilde{\mathbf{S}}}(\cdot)$, which denotes the joint PDF of transmitted signals for symbols $\{0, 1, \dots, M-1\}$. In (2.14), $G(\tilde{\mathbf{s}})$ represents the average probability of correct decision when the deterministic signal vector $\tilde{\mathbf{s}}$ is used for the transmission of M symbols over the additive noise channel and the corresponding MAP detector is employed at the receiver. Therefore, $\mathbb{E}\{G(\tilde{\mathbf{S}})\}$ can be interpreted as a randomization (or time-sharing) among possibly infinitely many MAP detectors.[4] A more compact version of the optimization problem in (2.14) can now be stated as follows:

$$\max_{p_{\tilde{\mathbf{S}}}} \quad \mathbb{E}\{G(\tilde{\mathbf{S}})\} \qquad \text{subject to} \quad \mathbb{E}\{H(\tilde{\mathbf{S}})\} \leq \mathrm{A}\,. \tag{2.15}$$

where the expectations are taken over $\tilde{\mathbf{S}}$ and $p_{\tilde{\mathbf{S}}}(\cdot)$ denotes the joint PDF of transmitted signals for symbols $\{0, 1, \dots, M-1\}$. Let $\mathrm{P}_{\mathrm{c}}^{\star}$ denote the maximum average probability of correct decision obtained as the solution to the optimization problem in (2.15). From Proposition 2.1.1, $\mathrm{P}_{\mathrm{c}}^{\star} \geq \mathrm{P}_{\mathrm{c}}^{\dagger}$ is always satisfied.

Optimization problems in the form of (2.15) have been investigated in various studies in the literature [9, 15, 17, 25]. Assuming that $G(\mathbf{s})$ in (2.14) is a continuous function and $\boldsymbol{a} \preceq \mathbf{s} \preceq \boldsymbol{b}$ is satisfied for some finite $\boldsymbol{a}$ and

[4] In the sequel, it will be shown that the optimal solution requires a randomization among at most two MAP detectors.

$\boldsymbol{b}$, the optimal solution of (2.15) can be represented by a randomization of at most two signal levels as a result of Carathéodory's theorem [19]; that is, $p_{\tilde{\mathbf{S}}}^{\text{opt}}(\tilde{\mathbf{s}}) = \lambda\,\delta(\tilde{\mathbf{s}} - \mathbf{s}_1) + (1-\lambda)\,\delta(\tilde{\mathbf{s}} - \mathbf{s}_2)$. Therefore, the problem in (2.15) can be solved over such signal PDFs, which results in the following optimization problem:

$$\begin{aligned} \max_{\{\lambda,\,\mathbf{s}_1,\,\mathbf{s}_2\}} \quad & \lambda\, G(\mathbf{s}_1) + (1-\lambda)\, G(\mathbf{s}_2) \\ \text{subject to} \quad & \lambda\, H(\mathbf{s}_1) + (1-\lambda)\, H(\mathbf{s}_2) \ \leq \mathrm{A}\,, \\ & \lambda \in [0,1] \end{aligned} \tag{2.16}$$

where $G(\mathbf{s}_k) = \int_{\mathbb{R}^N} \max\limits_{j\in\{0,1,\ldots,M-1\}} \{\pi_j\, p_{\mathbf{N}}(\mathbf{y} - \mathbf{s}_{k,j})\}\, \mathrm{d}\mathbf{y}$, $H(\mathbf{s}_k) = \sum_{j=0}^{M-1} \pi_j \left\|\mathbf{s}_{k,j}\right\|_2^2$, and $\mathbf{s}_k = [\mathbf{s}_{k,0} \ \ \mathbf{s}_{k,1}, \ \cdots \ \mathbf{s}_{k,M-1}] \in \mathbb{R}^{MN}$, with $\mathbf{s}_{k,j}$ being an N dimensional row vector $\forall j \in \{0,1,\ldots,M-1\}$. Therefore, it is observed that the solution of (2.15) can be obtained by optimizing over a significantly reduced optimization space via (2.16).

Finally, the following proposition states that the maximum average probabilities of correct decision achieved by the solutions of the optimization problems in (2.8) and (2.16) are equal.

Proposition 2.1.2. *The optimization problems in* (2.8) *and* (2.16) *result in the same maximum value.*

Proof. First consider the optimization problem in (2.8) when $K = 2$ detectors are used and deterministic signaling is employed for each detector, that is, $p_{\mathbf{S}^{(1)}}(\mathbf{s}^{(1)}) = \delta(\mathbf{s}^{(1)} - \mathbf{s}_1)$ and $p_{\mathbf{S}^{(2)}}(\mathbf{s}^{(2)}) = \delta(\mathbf{s}^{(2)} - \mathbf{s}_2)$. In that case, (2.8) reduces to the optimization problem in (2.16); hence, (2.8) covers (2.16) as a special case. Therefore, the maximum value of the objective function in (2.8) should be larger than or equal to that of (2.16); namely, $\mathrm{P}_\mathrm{c}^{\dagger} \geq \mathrm{P}_\mathrm{c}^{\star}$. On the other hand, Proposition 2.1.1 implies that (2.15) (equivalently (2.16)) provides an upper bound on (2.8); that is, $\mathrm{P}_\mathrm{c}^{\dagger} \leq \mathrm{P}_\mathrm{c}^{\star}$. Therefore, it is concluded that $\mathrm{P}_\mathrm{c}^{\dagger} = \mathrm{P}_\mathrm{c}^{\star}$. $\square$

Proposition 2.1.2 implies that the solution of the original optimization problem in (2.8), which considers the joint optimization of detectors, stochastic signals and detector randomization, can be obtained as the solution of the much simpler optimization problem specified in (2.16). This also means that when multiple detectors are available for randomization (i.e., $K \geq 2$), it is sufficient to employ detector randomization for two deterministic signal vectors; i.e., there is no need to employ stochastic signaling to achieve the optimal solution. On the other hand, when there is only one detector (i.e., $K = 1$), the optimal solution may involve stochastic signaling, as investigated in [17]. All in all, the optimal solution to the most generic optimization problem in (2.8) results in either detector randomization for two deterministic signal values (for $K \geq 2$), or stochastic signaling without detector randomization (for $K = 1$).

2.2 Improvability and Non-improvability Conditions

It should be noted that detector randomization with deterministic signaling may or may not improve detection performance over the conventional system (which does not perform any detector randomization or stochastic signaling, and transmits at the maximum power limit) in certain scenarios depending on the noise statistics. Before discussing the conditions for improvability and non-improvability, we need to introduce the objective we would like to improve upon. Conventional signaling strategies rely on using the available transmitter power at its limit by employing different signal constellations (e.g., antipodal signaling for binary communications with equal priors, PAM, QAM, etc.) without benefiting from possible gains of stochastic signaling and detector randomization. For any specific problem under consideration (any distribution of channel noise PDF, number of symbols, prior probabilities and average power constraint), the optimal deterministic signaling strategy can be investigated or a widely used signaling scheme can be adopted as the conventional approach.

In order to define improvability and non-improvability, we consider a generic conventional system as the reference, which is defined as the one that employs deterministic signaling at the power limit A and a single MAP detector at the receiver [17, 25]. Then, the system is called improvable if detector randomization with deterministic signaling can result in a higher average probability of correct decision under the average power constraint. Otherwise, it is called non-improvable. The average probability of correct decision for the conventional system can be expressed as (cf. (2.9))

$$\mathrm{P}_{\mathrm{c}}^{\mathrm{conv}} = \int \max_{j \in \{0,1,\ldots,M-1\}} \left\{ \pi_j \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{s}_j^{\mathrm{conv}}) \right\} \, \mathrm{d}\mathbf{y} \,, \tag{2.17}$$

where $\mathbf{s}^{\mathrm{conv}}$ represents the joint deterministic signal vector transmitted over the channel for all the M symbols and $\sum_{j=0}^{M-1} \pi_j \left\|\mathbf{s}_j^{\mathrm{conv}}\right\|_2^2 = \mathrm{A}$ is satisfied. The specific choice of $\mathbf{s}^{\mathrm{conv}}$ is determined by the properties of the problem under consideration and the aim is to improve upon $\mathrm{P}_{\mathrm{c}}^{\mathrm{conv}}$ under the average power constraint.

Improvability and non-improvability conditions can be derived for the problem studied above from the previous studies in the literature since the problem formulations in the form of (2.15) and (2.16) have been investigated in various studies such as [9, 17, 25, 55]. Namely, (2.15) is in the same form as the noise enhanced detection problem in which the aim is to maximize the detection probability under the false alarm constraint:

$$\max_{p_{\mathbf{N}}} \mathbb{E}\{F_1(\mathbf{N})\} \text{ subject to } \mathbb{E}\{F_0(\mathbf{N})\} \leq \alpha \,, \tag{2.18}$$

where $p_{\mathbf{N}}(\cdot)$ is the PDF of the additive noise, and $F_1(\mathbf{n})$ and $F_0(\mathbf{n})$ are, respectively, the detection and false alarm probabilities for a given additive noise value of $\mathbf{n}$. Since the signals cannot take infinitely large values in practice, the signal values can be considered to be in closed finite intervals in (2.15). Hence, the results for the problem above can be applied to the problem under consideration.

In order to obtain some sufficient conditions for the improvability and non-improvability of the average probability of correct decision, a similar approach to those of [9] and [56] can be pursued. To that aim, we begin by noting the relation among $p_{\tilde{\mathbf{S}}}$, $G(\tilde{\mathbf{S}})$ and $H(\tilde{\mathbf{S}})$ of the optimization problem in (2.15). $G(\tilde{\mathbf{s}})$ represents the average probability of correct decision when the deterministic signal vector $\tilde{\mathbf{s}}$ is used for the transmission of M symbols over the additive noise channel and the corresponding MAP detector is employed at the receiver, and $H(\tilde{\mathbf{s}})$ indicates the power of the same deterministic signal vector $\tilde{\mathbf{s}}$ averaged over the prior probabilities. For a given value h of H, we have $\tilde{\mathbf{s}} = H^{-1}(h)$, where H^{-1} is the inverse function of H. Since H is not a one-to-one mapping function in our case, we have a set of $\tilde{\mathbf{s}}$ values which satisfy $H(\tilde{\mathbf{s}}) = h$. A set of values g of G can be obtained correspondingly by $g = G(\tilde{\mathbf{s}}) = G(H^{-1}(h))$. By introducing the joint PDF $p_{\tilde{\mathbf{S}},h}(\cdot)$ for the signal distribution in the h domain, the optimization problem in (2.15) can be expressed equivalently as

$$\max_{p_{\tilde{\mathbf{S}},h}} \int_0^\infty g\, p_{\tilde{\mathbf{S}},h}(h)\,\mathrm{d}h \quad \text{subject to} \int_0^\infty h\, p_{\tilde{\mathbf{S}},h}(h)\,\mathrm{d}h \ \leq \mathrm{A} \tag{2.19}$$

where $p_{\tilde{\mathbf{S}},h}(\cdot)$ should satisfy the conditions to be a PDF (in fact, a finite upper limit can be used in the integrals instead of infinity considering practical scenarios). This approach enables us to continue the analysis in a single dimensional space instead of $\mathbb{R}^{MN}$. As stated by the authors of [9], even in the cases where the exact forms of g and h are hard to compute, the relationship between them can be learned by Monte-Carlo simulation using importance sampling. Similarly to [9], the function $J(t)$ is defined as the maximum value of g given h; i.e., $J(t) = \sup\{g : h = t\}$.

In the following, some of the previous results in the literature, namely those in [9] and [56], are adapted into our context:

2.2.1 Sufficient Conditions for Improvability

1) If $J''(\mathrm{A}) > 0$ when $J(t)$ is second-order continuously differentiable around A, then there exists at least one signaling process $\tilde{\mathbf{S}}$ with PDF $p_{\tilde{\mathbf{S}}}(\cdot)$ that can improve the correct decision performance from Theorem 1 of [9].

2) If there exists a joint signal vector $\mathbf{x}$ such that $G(\mathbf{x}) > \mathrm{P}_{\mathrm{c}}^{\mathrm{conv}}$ and $H(\mathbf{x}) \leq \mathrm{A}$, then the average probability of correct decision can be improved by using $p_{\tilde{\mathbf{S}}}(\tilde{\mathbf{s}}) = \delta(\tilde{\mathbf{s}} - \mathbf{x})$ from a generalization of Theorem 1 of [9].

3) The average probability of correct decision can be improved if there exists $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ that satisfy

$$\frac{[\mathrm{A} - H(\tilde{\mathbf{x}}_2)][G(\tilde{\mathbf{x}}_1) - G(\tilde{\mathbf{x}}_2)]}{H(\tilde{\mathbf{x}}_1) - H(\tilde{\mathbf{x}}_2)} \; > \; \mathrm{P}_{\mathrm{c}}^{\mathrm{conv}} - G(\tilde{\mathbf{x}}_2) \tag{2.20}$$

from Proposition 2 of [56].

4) Assume that $G(\mathbf{x})$ is second-order continuously differentiable around $\mathbf{x} = \mathbf{s}^{\mathrm{conv}}$. Define $p(\mathbf{x}, \mathbf{z}) \triangleq \sum_{k=1}^{MN} z_k \frac{\partial H(\mathbf{x})}{\partial x_k}$, $r(\mathbf{x}, \mathbf{z}) \triangleq \sum_{k=1}^{MN} z_k \frac{\partial G(\mathbf{x})}{\partial x_k}$, $u(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^{MN} \sum_{k=1}^{MN} z_l z_k \frac{\partial^2 H(\mathbf{x})}{\partial x_l \partial x_k}$, and $w(\mathbf{x}, \mathbf{z}) \triangleq \sum_{l=1}^{MN} \sum_{k=1}^{MN} z_l z_k \frac{\partial^2 G(\mathbf{x})}{\partial x_l \partial x_k}$, where x_l and z_l represent the lth components of $\mathbf{x}$ and $\mathbf{z}$, respectively. Then, the average probability of correct decision can be improved if there exists an MN-dimensional vector $\mathbf{z}$ such that $p(\mathbf{x}, \mathbf{z}) > 0$, $r(\mathbf{x}, \mathbf{z}) > 0$ and

$$w(\mathbf{x}, \mathbf{z})\, p(\mathbf{x}, \mathbf{z}) \; > \; u(\mathbf{x}, \mathbf{z})\, r(\mathbf{x}, \mathbf{z}) \tag{2.21}$$

are satisfied at $\mathbf{x} = \mathbf{s}^{\mathrm{conv}}$ from Proposition 3 of [56].

5) The average probability of correct decision can be improved if $G(\mathbf{x})$ and $-H(\mathbf{x})$ are strictly convex at $\mathbf{x} = \mathbf{s}^{\mathrm{conv}}$ from Proposition 4 of [56].

6) Assume that $G(\mathbf{x})$ is second-order continuously differentiable around $\mathbf{x} = \mathbf{s}^{\mathrm{conv}}$. The average probability of correct decision can be improved if there exists an MN-dimensional vector $\mathbf{v}$ such that

$\left(\sum_{l=1}^{MN} v_l \frac{\partial G(\mathbf{x})}{\partial x_l}\right) \left(\sum_{l=1}^{MN} v_l \frac{\partial H(\mathbf{x})}{\partial x_l}\right) < 0$ is satisfied at $\mathbf{x} = \mathbf{s}^{\text{conv}}$, where v_l represents the lth component of $\mathbf{v}$ from Proposition 5 of [56].

2.2.2 Sufficient Conditions for Non-improvability

1) If there exists a non-decreasing concave function $\Psi(t)$ that satisfies $\Psi(t) \geq J(t)$ $\forall t$ and $\Psi(\mathrm{A}) = J(\mathrm{A}) = \mathrm{P}_{\mathrm{c}}^{\mathrm{conv}}$, then the average probability of correct decision is non-improvable from Theorem 2 of [9].

2) Assume that $H(\tilde{\mathbf{s}}) \leq \mathrm{A}$ implies $G(\tilde{\mathbf{s}}) \leq \mathrm{P}_{\mathrm{c}}^{\mathrm{conv}}$ for all $\tilde{\mathbf{s}} \in \mathcal{C}$, where $\mathcal{C}$ is a convex set consisting of all possible values of transmitted joint signal vector $\tilde{\mathbf{s}}$. If $H(\tilde{\mathbf{s}})$ is a convex function and $G(\tilde{\mathbf{s}})$ is a concave function over $\mathcal{C}$, then the average probability of correct decision is non-improvable from Proposition 1 of [56].

2.3 Details on Optimization

The optimization problems in the form of (2.16) have been investigated in various studies in the literature, such as [9, 15–17]. The main approaches in solving (2.16) include the analytical techniques as in [15] and [9], the convex relaxation technique to obtain an approximate solution in polynomial time as employed in [16], and the global optimization algorithms such as differential evolution and particle swarm optimization [16]. In this study, a global optimization technique based on multistart and pattern search algorithms from MATLAB's Global Optimization Toolbox [57–59], are used to obtain the solution of (2.16).

2.3.1 Global Optimization Algorithms

Since the optimization problem in (2.16) may not be a convex problem in general, global optimization algorithms, such as particle swarm optimization (PSO) and differential evolution (DE), can be employed. Although these approaches have a random nature and do not guarantee finding the global optimal, they work quite efficiently in many practical problems (e.g., [16, 17]). In our study, we have used the global optimization technique based on multistart and pattern search algorithms from MATLAB's Global Optimization Toolbox.

2.3.2 Convex Relaxation Approach

The convex relaxation approach can be employed to obtain an approximate solution in an efficient manner. Specifically, a set of candidate signal values, say $\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_L$ are considered for $\tilde{\mathbf{S}}$ in (2.15), and the weights for those possible signal values, $\lambda_1, \ldots, \lambda_L$, can be searched for. In other words, the optimization problem in (2.15) is approximated as

$$\max_{\lambda_1,\ldots,\lambda_L} \sum_{i=1}^{L} \lambda_i G(\tilde{\mathbf{s}}_i) \quad \text{subject to} \quad \sum_{i=1}^{L} \lambda_i H(\tilde{\mathbf{s}}_i) \leq \mathrm{A}\,, \quad \sum_{i=1}^{L} \lambda_i = 1\,, \quad \lambda_i \geq 0\,, \tag{2.22}$$

which is a linearly constrained linear optimization problem that can be solved very efficiently in polynomial time.

The accuracy of the approximation increases as more candidate signal values are considered (as L increases). In fact, for digital systems, since the possible signal values are discrete, it may be possible to obtain the exact solution via (2.22) above if all the possible signal values are considered. This convex relaxation approach is employed in [16] for a similar problem in the context of noise enhanced detection.

2.3.3 Analytical Approach

The third approach in obtaining the solution of (2.16) is the analytical approach as in [9, 15]. In those studies, noise enhanced detection is studied in the Neyman-Pearson framework; that is, the optimal probability density function (PDF) of noise is searched under a constraint on the false alarm probability. This problem is in the form of (2.18). Assuming that the signals cannot take infinitely large values (e.g., they are in finite closed intervals), the optimization problem in (2.15) is in the same form as the problem formulation in (2.18). Therefore, the analytical approaches in [15] and [9] can be employed to obtain the solution. Specifically, the approach in Section III.C of [9] (which is also employed in [16]) and the SR noise finding algorithm in Section III of [15] can be adopted for the problem.

2.4 Simulation Results

In this section, three numerical examples are presented to compare the optimal solution obtained in the previous section and various signaling techniques in terms of probability of error performance. A communications system specified as in (2.1) is considered with scalar observations and equal priors. First two examples involve a binary communications system. Whereas, the third example considers a quaternary communications system employing symmetric signaling. It is assumed that the receiver is able to implement multiple detectors ($K \geq 2$) and to randomize among them. Gaussian mixture models with equal weights and variances are assumed for the noise in all the examples, the PDF of which can be expressed as $p_N(n) = \sum_{i=1}^{L} \exp\{-(n-\mu_i)^2/(2\sigma^2)\}/(\sqrt{2\pi}\,\sigma L)$ [27]. Note that the average power of the noise can be calculated from $\mathbb{E}\{N^2\} = \sigma^2 + (1/L)\sum_{i=1}^{L} \mu_i^2$. Similar to those introduced in [17], three different signaling schemes are considered:

Gaussian Solution (Conventional): Lacking any information about the noise PDF, the transmitter employs antipodal signaling, which is known to be optimal in the presence zero-mean Gaussian noise and equal priors [20]. For power constraint A , $\left\{-\sqrt{\mathrm{A}}, \sqrt{\mathrm{A}}\right\}$ is selected as signal levels for binary communication. For quaternary communication ($M = 4$), $\left\{\frac{-3\sqrt{\mathrm{A}}}{\sqrt{5}}, \frac{-\sqrt{\mathrm{A}}}{\sqrt{5}}, \frac{\sqrt{\mathrm{A}}}{\sqrt{5}}, \frac{3\sqrt{\mathrm{A}}}{\sqrt{5}}\right\}$ is considered as the conventional signaling scheme. On the other hand, MAP decision rule is used at the receiver.

Optimal−Stochastic: Goken et al. [17] showed that joint optimization of signaling structure and detector rule can be performed over a number of parameters instead of functions if the transmitter has some means of estimating the noise PDF at the receiver. It is proved that optimal signal for each symbol can be characterized by a discrete random variable with at most two mass points. Under this setting, only a single detector is considered at the receiver (that is, no detector randomization is employed), and it uses the MAP decision rule corresponding to the optimal signals obtained from the solution of the following optimization problem, as stated in (9) of [17]:

$$\begin{aligned}
\max_{\{\lambda, \mathbf{s}_1, \mathbf{s}_2\}} \quad & \int_{\mathbb{R}^N} \max_{j \in \{0,1,\ldots,M-1\}} \left\{ \pi_j \left(\lambda \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{s}_{1,j}) + (1-\lambda) \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{s}_{2,j}) \right) \right\} \mathrm{d}\mathbf{y} \\
\text{subject to} \quad & \lambda \left(\sum_{j=0}^{M-1} \pi_j \left\| \mathbf{s}_{1,j} \right\|_2^2 \right) + (1-\lambda) \left(\sum_{j=0}^{M-1} \pi_j \left\| \mathbf{s}_{2,j} \right\|_2^2 \right) \le \mathrm{A}\,, \\
& \lambda \in [0,1]
\end{aligned} \tag{2.23}$$

Optimal−Deterministic: A simplified version of the optimal solution in (2.23) for single detector case can be obtained by assuming that the transmitted signal for each symbol is deterministic; i.e., it is not a randomization of two distinct signal levels. The optimization problem in (2.23) reduces to

$$\begin{aligned}
\max_{\mathbf{s}} \quad & \int_{\mathbb{R}^N} \max_{j \in \{0,1,\ldots,M-1\}} \left\{ \pi_j \, p_{\mathbf{N}}(\mathbf{y} - \mathbf{s}_j) \right\} \mathrm{d}\mathbf{y} \\
\text{subject to} \quad & \sum_{j=0}^{M-1} \pi_j \left\| \mathbf{s}_j \right\|_2^2 \le \mathrm{A}
\end{aligned} \tag{2.24}$$

This scheme does not employ any detector randomization or stochastic signaling, and obtains the optimal deterministic signal levels and the corresponding MAP detector [17].

In addition to the approaches described above, the following scheme investigated in the previous section is considered as the overall optimal solution:

Optimal Detector Randomization with Optimal Deterministic Signaling: This case refers to the solution of the most generic optimization problem in (2.6), which can be obtained from (2.16) as studied in the previous section.

2.4.1 Example 1

The average error probabilities of the schemes described above are plotted versus A/σ^2 for $\mathrm{A} = 1$ in Figure 2.2, where the parameters of the Gaussian mixture noise are given by $L = 6$ and $\boldsymbol{\mu} = [-1.08\ -0.81\ -0.27\ 0.27\ 0.81\ 1.08]$ as in [17]. However, unlike [17], symmetric signaling assumption is not employed. From Figure 2.2, it is observed that the Gaussian solution has the worst performance as expected since it is optimized for zero-mean Gaussian noise. Optimizing deterministic signal levels improves the performance of the Gaussian solution, as observed from the Optimal–Deterministic curve. Further performance improvements are obtained when optimal stochastic signals are considered instead of deterministic signals (see Optimal–Stochastic). However, the best probability of error performance is achieved by the optimal solution of the most generic optimization problem investigated in the previous section, which performs optimal detector randomization among two MAP detectors corresponding to two deterministic signal pairs (see Optimal–Detector Randomization). In accordance with Proposition 2.1.1, stochastic signaling without detector randomization cannot perform better than detector randomization with deterministic signaling.

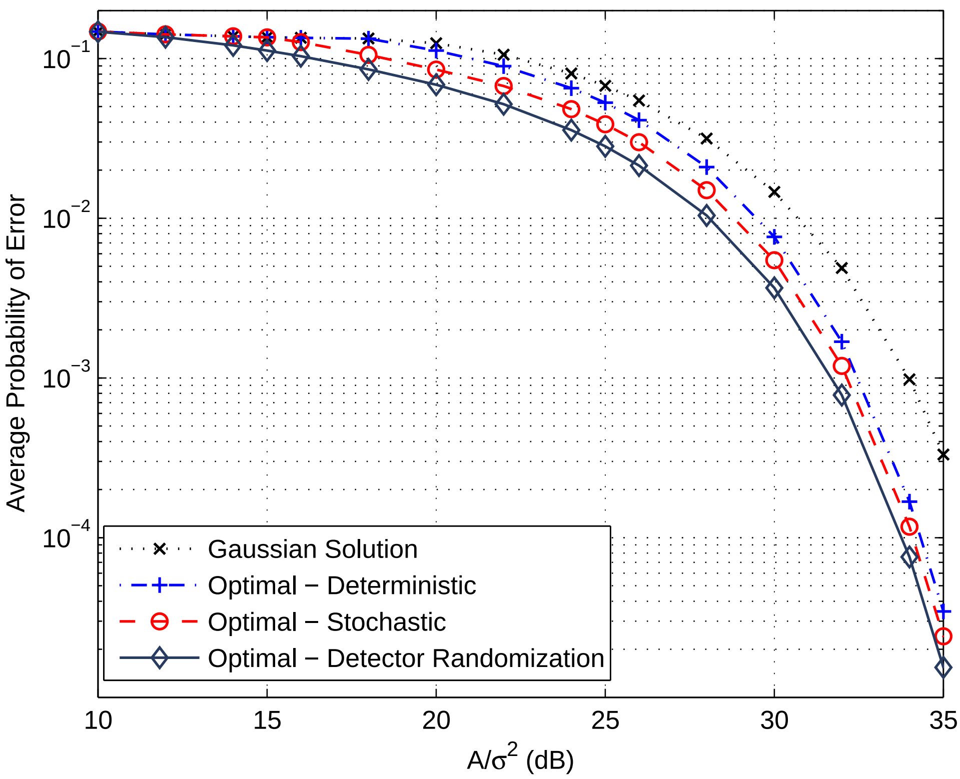

Figure 2.2: Average probability of error versus A/σ^2 for various approaches. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.27\ 0.81\ 1.08]$ with equal weights is considered.

In Tables 2.1, 2.2 and 2.3, some optimal signals are presented for the Optimal–Deterministic, Optimal–Stochastic and Optimal–Detector Randomization schemes in Figure 2.2, respectively. For optimal deterministic signaling, $\mathbf{s}_0$ and $\mathbf{s}_1$ denote the optimal deterministic signal levels for symbol 0 and symbol 1 in Table 2.1. On the other hand, the optimal signal vector for symbols 0 and 1 has the PDF in the form of $p_S(s) = \beta\,\delta(s - s_{1,i}) + (1-\beta)\,\delta(s - s_{2,i})$ for optimal stochastic signaling as shown in Table 2.2. Finally, the optimal solution obtained in the previous section (Optimal–Detector Randomization) employs the signal pair $[s_{1,0} \quad s_{1,1}]$ and the corresponding MAP detector with probability λ, and the signal pair $[s_{2,0} \quad s_{2,1}]$ and the corresponding MAP detector with probability $1-\lambda$ as presented in Table 2.3. It is observed that all optimal signaling

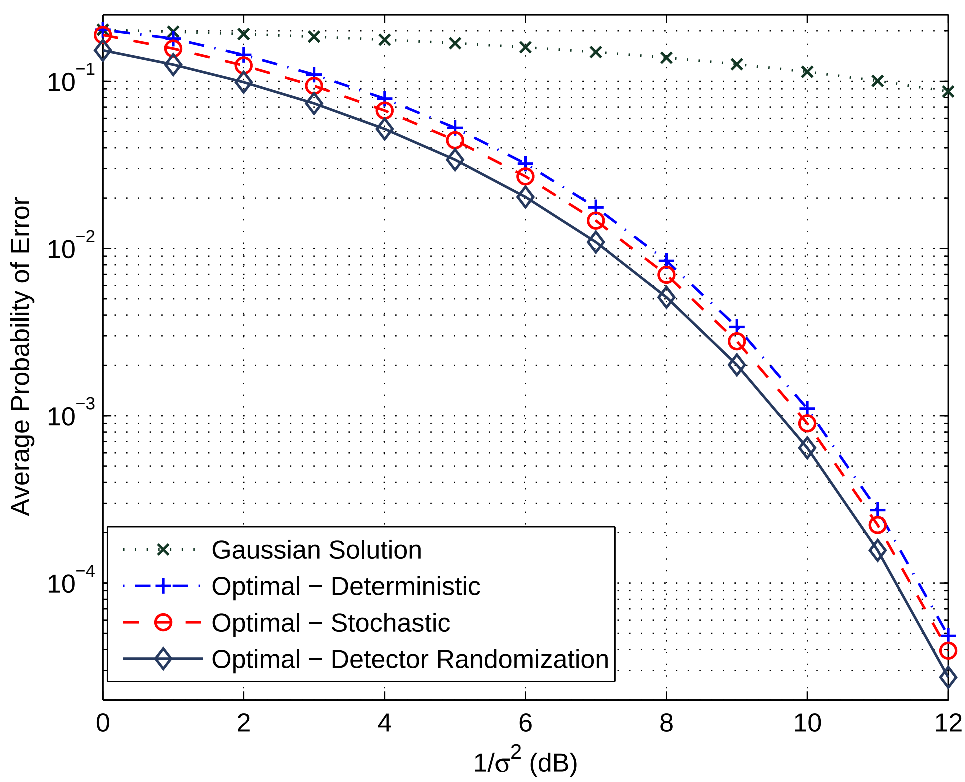

Figure 2.3: Average probability of error versus A/σ^2. A symmetric Gaussian mixture noise, which has its mass points at [-2 2] with equal weights is considered.

schemes get close to deterministic signaling for small A/σ^2, which is also verified from Figure 2.2. However, the signaling schemes become quite different as A/σ^2 increases from 10 dB, which results in differences in probability of error performance.

2.4.2 Example 2

In this example, the parameters of the Gaussian mixture noise are given by $L = 2$ and $\boldsymbol{\mu} = [-2\ 2]$, the average power constraint is $\mathrm{A} = 5$, and the average error probabilities of the schemes described above are plotted versus $1/\sigma^2$ in Figure 2.3. As before, a binary communications system is assumed and symmetric signaling is not employed. Compared to Example 1, the worst performance of the conventional signaling scheme (Gaussian solution) is much more evident in

Example 2. This is because, the overlaps among the components of the Gaussian mixture noise is more severe for conventional signaling case in this example. On the other hand, the superior performance of optimal detector randomization with optimal deterministic signaling is verified once again. Decrease in the average probability of error can be explained by the introduction of the second MAP detector and the optimal randomization of signal levels between two detectors for each symbol. Similarly to the previous example, improvements disappear as σ^2 increases. Tables 2.4, 2.5 and 2.6 present the optimal signals for different values of $1/\sigma^2$. The notation is the same as in the previous example.

2.4.3 Example 3

Another example is constructed to investigate the benefits that can be obtained via detector randomization in M-ary communication systems when $M > 2$. For this purpose, a quaternary communications system ($M = 4$) with symmetric signaling is considered, that is $\mathbf{s}_2 = -\mathbf{s}_0$ and $\mathbf{s}_3 = -\mathbf{s}_1$. The parameters of the Gaussian mixture noise are given by $L = 6$ and $\boldsymbol{\mu} = [-0.432\ \ -0.324\ \ -0.108\ 0.108\ 0.324\ 0.432]$. The average error probabilities of the schemes described previously are plotted versus A/σ^2 for $\mathrm{A} = 1$ in Figure 2.4 with the corresponding optimal signals given in Tables 2.7, 2.8 and 2.9. Note that the results for symbols 0 and 1 are listed in the tables, and the results for symbols 2 and 3 are the negatives of them respectively since symmetric signaling is considered. It is evident from the figure that optimal detector randomization with deterministic signals achieves the best probability of error performance. Performance improvements among different signaling schemes deteriorate as A/σ^2 drops below 20 dB. A final observation is that optimal stochastic signaling approach cannot improve upon optimal deterministic signaling for this example. This is possibly due to the fact that signal space is overcrowded with the PDFs of four distinct symbols (each with six Gaussian mixture components) and there is no room left for

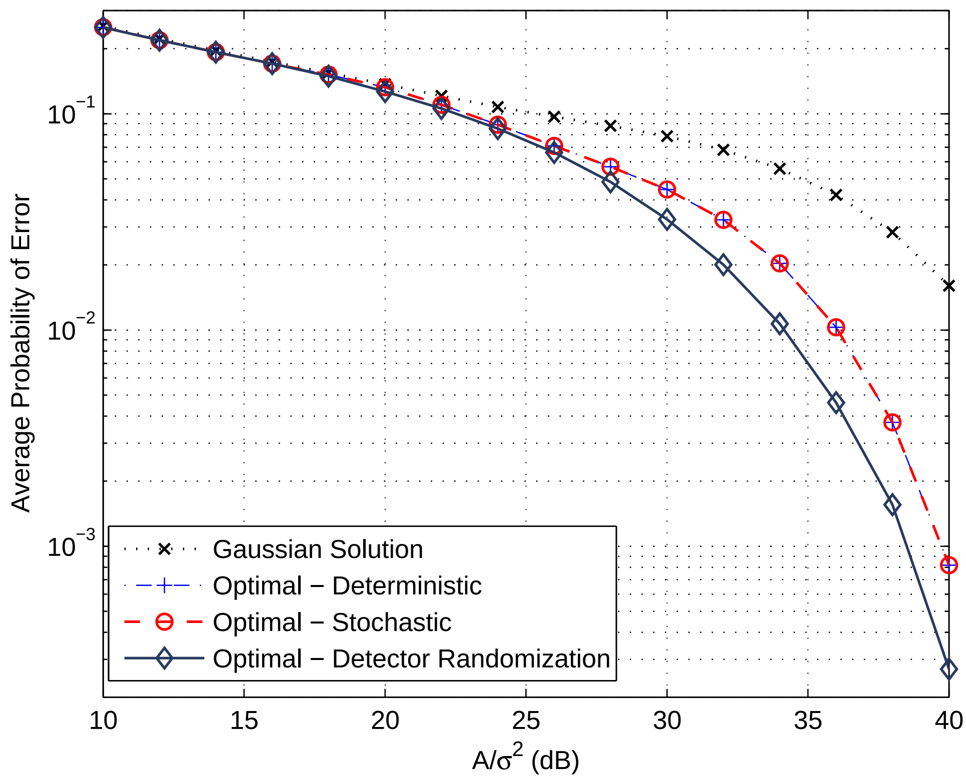

Figure 2.4: Average probability of error versus A/σ^2. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.108\ 0.324\ 0.432]$ with equal weights is considered.

any performance improvement via randomization after the optimal allocation of deterministic signal values.

2.5 Concluding Remarks

In this chapter, optimal receiver design is studied for a communications system in which both detector randomization and stochastic signaling can be performed [31]. First, it is proven that stochastic signaling without detector randomization cannot achieve a smaller average probability of error than detector randomization with deterministic signaling for the same average power constraint and noise statistics. Then, it is shown that the optimal receiver design results in a randomization between at most two MAP detectors corresponding to two deterministic signal vectors. In addition, sufficient conditions are derived for improvability and non-improvability of the correct decision performance via detector randomization. Three numerical examples are provided to explain the results.

Table 2.1: Optimal deterministic signaling for the scenario in Figure 2.2 . A symmetric Gaussian mixture noise, which has its mass points at $\pm$[0.27 0.81 1.08] with equal weights is considered.

A/σ^2 (dB)	S_0	S_1
10	-0.9992	1.0008
12	-0.9997	1.0003
14	-0.9992	1.0008
15	-0.9997	1.0003
16	-0.9996	1.0004
18	-0.4968	1.1298
20	-1.3302	0.2552
22	-1.0224	0.5274
24	-1.1174	0.4084
25	-0.7622	0.7552
26	-0.9599	0.5509
28	-1.0141	0.4871
30	-0.8191	0.6761
32	-0.6872	0.8042
34	-0.4520	1.2990
35	-0.4886	0.9996
36	-1.3153	0.1723

Table 2.2: Optimal stochastic signaling for the scenario in Figure 2.2 . A symmetric Gaussian mixture noise, which has its mass points at $\pm$[0.27 0.81 1.08] with equal weights is considered.

A/σ^2 (dB)	λ	$S_{1,0}$	$S_{2,0}$	$S_{1,1}$	$S_{2,1}$
10	1	-0.9993	N/A	1.0007	N/A
12	0.5890	-0.9998	-0.9986	1.0017	0.9997
14	0.9474	-1.0000	-0.9925	1.0000	0.9944
15	0.7212	-1.1417	-0.5766	0.9835	0.9836
16	0.5992	-0.6402	-1.4368	0.9623	0.9634
18	0.6119	-0.6412	-1.4596	0.9600	0.9600
20	0.3927	-1.4543	-0.6388	0.9600	0.9600
22	0.4261	-1.4414	-0.5934	0.9553	0.9553
24	0.4496	-1.4259	-0.5700	0.9523	0.9523
25	0.4594	-1.4135	-0.5590	0.9556	0.9556
26	0.4680	-0.9558	-0.9558	1.4049	0.5528
28	0.5169	-0.9599	-0.9599	0.5403	1.3857
30	0.5099	-0.8934	-0.8934	0.6014	1.4408
32	0.5023	-1.3206	-0.4886	1.0026	1.0026
34	0.4878	-0.5371	-1.3641	0.9519	0.9519
35	0.4892	-1.0150	-1.0150	0.4731	1.2977
36	0.4686	-1.1337	-1.1337	1.1761	0.3538

Table 2.3: Optimal detector randomization with optimal deterministic signaling for the scenario in Figure 2.2 . A symmetric Gaussian mixture noise, which has its mass points at $\pm$[0.27 0.81 1.08] with equal weights is considered.

A/σ^2 (dB)	λ	$S_{1,0}$	$S_{2,0}$	$S_{1,1}$	$S_{2,1}$
10	0.0386	-0.6439	-1.0119	0.6441	1.0115
12	0.6962	-1.1077	-0.6892	1.1084	0.6936
14	0.4473	-0.7603	-1.1576	0.7613	1.1583
15	0.4878	-0.7712	-1.1755	0.7755	1.1763
16	0.4850	-1.1916	-0.7783	1.1907	0.7780
18	0.4582	-1.2095	-0.7811	1.2086	0.7801
20	0.5417	-0.7746	-1.2127	0.7758	1.2137
22	0.5276	-0.7620	-1.2053	0.7709	1.2120
24	0.4895	-1.1974	-0.7556	1.2042	0.7626
25	0.4980	-1.1930	-0.7606	1.1998	0.7517
26	0.4935	-0.7552	-1.1876	0.7522	1.1964
28	0.5220	-1.1860	-0.7444	1.1810	0.7554
30	0.4644	-0.7348	-1.1724	0.7598	1.1793
32	0.5467	-1.1877	-0.7191	1.1509	0.7721
34	0.4432	-0.7420	-1.1554	0.7470	1.1724
35	0.4399	-0.7406	-1.1272	0.7476	1.1960
36	0.5637	-1.1949	-0.7443	1.1243	0.7433

Table 2.4: Optimal deterministic signaling for the scenario in Figure 2.3 . A symmetric Gaussian mixture noise, which has its mass points at [-2 2] with equal weights is considered.

$1/\sigma^2$ (dB)	S_0	S_1
0	-2.2359	2.2362
1	-1.3463	1.2139
2	-1.8662	0.5732
3	-2.2893	0.0583
4	-2.0515	0.2245
5	-0.3911	1.8281
6	-2.1446	0.0296
7	-0.5547	1.5837
8	-0.0524	2.0574
9	-0.7411	1.3461
10	-1.4442	0.6252
11	-2.0190	0.0360
12	-0.8504	1.1934

Table 2.5: Optimal stochastic signaling for the scenario in Figure 2.3 . A symmetric Gaussian mixture noise, which has its mass points at [-2 2] with equal weights is considered.

$1/\sigma^2$ (dB)	λ	$S_{1,0}$	$S_{2,0}$	$S_{1,1}$	$S_{2,1}$
0	0.7078	-1.8619	-1.8618	0.7223	4.5932
1	0.2906	-1.8217	-1.8217	4.6888	0.6433
2	0.2940	-4.7150	-0.5793	1.7966	1.7966
3	0.2993	-4.7102	-0.5263	1.7787	1.7788
4	0.6945	-1.7674	-1.7672	0.4798	4.6887
5	0.3120	-4.6580	-0.4399	1.7597	1.7599
6	0.6815	-1.7525	-1.7525	0.4086	4.6255
7	0.6750	-1.7483	-1.7483	0.3813	4.5897
8	0.6689	-1.7469	-1.7468	0.3570	4.5528
9	0.6632	-1.7521	-1.7511	0.3326	4.5126
10	0.6583	-1.7971	-1.7993	0.2699	4.4346
11	0.3473	-1.7618	-1.7618	4.4383	0.2916
12	0.6608	-0.6452	-4.7865	1.3975	1.3974

Table 2.6: Optimal detector randomization with optimal deterministic signaling for the scenario in Figure 2.3 . A symmetric Gaussian mixture noise, which has its mass points at [-2 2] with equal weights is considered.

$1/\sigma^2$ (dB)	λ	$S_{1,0}$	$S_{2,0}$	$S_{1,1}$	$S_{2,1}$
0	0.4453	-3.0614	-1.2207	3.0610	1.2216
1	0.5778	-1.1881	-3.1479	1.1908	3.1471
2	0.5889	-1.1602	-3.1981	1.1617	3.1990
3	0.4072	-3.2258	-1.1355	3.2250	1.1351
4	0.5925	-1.1121	-3.2361	1.1131	3.2354
5	0.4106	-3.2342	-1.0934	3.2351	1.0924
6	0.4153	-3.2257	-1.0745	3.2267	1.0776
7	0.5789	-1.0614	-3.2117	1.0624	3.2145
8	0.5726	-1.0537	-3.1966	1.0467	3.1976
9	0.4339	-3.1802	-1.0428	3.1794	1.0380
10	0.5598	-1.0280	-3.1556	1.0409	3.1684
11	0.5549	-1.0278	-3.1336	1.0270	3.1650
12	0.5515	-1.0166	-3.0838	1.0262	3.1971

Table 2.7: Optimal deterministic signaling for the scenario in Figure 2.4 . A symmetric Gaussian mixture noise, which has its mass points at $\pm$[0.108 0.324 0.432] with equal weights is considered.

A/σ^2 (dB)	S_0	S_1
10	0.3658	1.3661
12	0.3762	1.3633
14	0.3869	1.3603
16	0.3977	1.3571
18	1.3553	0.4041
20	0.3066	1.3806
22	0.3152	1.3786
24	1.3780	0.3180
26	1.3778	0.3188
28	0.3154	1.3785
30	0.3097	1.3796
32	1.3809	0.3051
34	1.3816	0.3021
36	0.3002	1.3820
38	1.3675	0.2990
40	0.2983	1.3553

Table 2.8: Optimal stochastic signaling for the scenario in Figure 2.4 . A symmetric Gaussian mixture noise, which has its mass points at $\pm$[0.108 0.324 0.432] with equal weights is considered.

A/σ^2 (dB)	λ	$S_{1,0}$	$S_{2,0}$	$S_{1,1}$	$S_{2,1}$
10	1	1.3667	N/A	0.3636	N/A
12	1	1.3633	N/A	0.3762	N/A
14	1	1.3603	N/A	0.3868	N/A
16	1	0.3956	N/A	1.3578	N/A
18	1	1.3554	N/A	0.4029	N/A
20	1	1.3806	N/A	0.3066	N/A
22	1	1.3786	N/A	0.3152	N/A
24	1	1.3778	N/A	0.3180	N/A
26	1	1.3778	N/A	0.3188	N/A
28	1	1.3785	N/A	0.3154	N/A
30	1	1.3799	N/A	0.3097	N/A
32	1	1.3809	N/A	0.3051	N/A
34	1	1.3816	N/A	0.3021	N/A
36	1	0.3002	N/A	1.3820	N/A
38	1	0.2990	N/A	1.3626	N/A
40	1	0.3497	N/A	1.3683	N/A

Table 2.9: Optimal detector randomization with optimal deterministic signaling for the scenario in Figure 2.4 . A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.108\ 0.324\ 0.432]$ with equal weights is considered.

A/σ^2 (dB)	λ	$S_{1,0}$	$S_{2,0}$	$S_{1,1}$	$S_{2,1}$
10	0.7559	0.3633	1.3682	1.3662	0.3642
12	1	1.3633	N/A	0.3762	N/A
14	1	1.3612	N/A	0.3837	N/A
16	1	1.3581	N/A	0.3940	N/A
18	0.3653	0.2105	1.4295	1.2420	0.4408
20	0.5058	1.2609	0.4659	0.2323	1.4654
22	0.5303	1.2599	0.4764	0.2360	1.4752
24	0.2622	0.4836	1.3224	1.4854	0.3081
26	0.2678	0.4885	0.3104	1.4882	1.3183
28	0.6570	0.4889	0.0720	1.4814	1.0790
30	0.3178	0.0736	0.4866	1.0689	1.4698
32	0.7143	0.4826	1.0553	1.4547	0.0758
34	0.2520	0.0777	1.4404	1.0450	0.4786
36	0.2206	0.0790	0.4745	1.0356	1.4268
38	0.2131	0.0798	0.4712	1.0271	1.4158
40	0.1720	0.0802	0.4684	1.0204	1.4066

3

Convexity Properties of Detection Probability under Additive Gaussian Noise: Optimal Signaling and Jamming Strategies

In this chapter, we study the convexity/concavity properties for the problem of detecting the presence of a signal emitted from a power constrained transmitter in the presence of additive Gaussian noise under the Neyman-Pearson framework. Section 3.1 introduces the problem. In Section 3.2, it is proved that the detection probability corresponding to the $\alpha-$level likelihood ratio test is either concave or has two inflection points such that the function is concave, convex and finally concave with respect to increasing values of the signal power. In addition, the analysis is extended from scalar observations to multidimensional colored Gaussian noise corrupted signals. Based on the convexity/concavity results, optimal and near-optimal power sharing/randomization strategies are proposed for average/peak power constrained transmitters. In Section 3.3, a similar analysis is carried out for the case of a power constrained jammer.

3.1 Problem Formulation

Consider the problem of detecting the presence of a target signal, where the receiver needs to decide between the two hypotheses $\mathcal{H}_0$ or $\mathcal{H}_1$ based on a real-valued scalar observation Y acquired over an AWGN channel.

$$\mathcal{H}_0 : \;\; Y = \sigma N \;\; , \qquad \mathcal{H}_1 \; : \; Y = \sqrt{S} + \sigma N \tag{3.1}$$

Here, $N \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable with zero mean and unit variance, $\sigma > 0$ is the noise standard deviation at the receiver, $\sqrt{S}$ represents the transmitted signal for the alternative hypothesis $\mathcal{H}_1$, and $S > 0$ is the corresponding signal power. The additive noise N is statistically independent of the signal $\sqrt{S}$. The scalar channel model in (3.1) provides an abstraction for a continuous-time system that passes the received signal through a correlator (matched filter) and samples it once per symbol interval, thereby capturing the effects of modulator, additive noise channel and receiver front-end processing. In addition, although the above model is in the form of a simple additive noise channel, it may be sufficient to incorporate various effects such as thermal noise, multiple-access interference, inter-symbol interference and jamming [2].

It is well-known that the NP detector gives the most powerful α-level test of $\mathcal{H}_0$ versus $\mathcal{H}_1$ [20]. In other words, when the aim is to maximize the probability of detection such that the probability of false alarm does not exceed a predetermined value α, the NP detector is the optimal choice and takes the following form of an LRT for continuous PDFs:

$$\delta_{NP}(y) = \begin{cases} 1\,, & \text{if } \; p_1(y) \geq \eta\, p_0(y) \\ 0\,, & \text{if } \; p_1(y) < \eta\, p_0(y) \end{cases} \tag{3.2}$$

where the threshold $\eta \geq 0$ is chosen such that the probability of false alarm satisfies $\mathrm{P_{FA}} = \mathrm{P}_0\left(p_1(y) \geq \eta\, p_0(y)\right) = \alpha$, with subscript 0 denoting that the probability is calculated conditioned on the null hypothesis $\mathcal{H}_0$. Then, the NP decision rule is the optimal one among all α-level decision rules, i.e., $\mathrm{P_D} = \mathrm{P}_1\left(p_1(y) \geq \eta\, p_0(y)\right)$ is maximized, where the probability is calculated under the condition that the alternative hypothesis $\mathcal{H}_1$ is true.

The hypothesis pair in (3.1) can be restated in terms of the distributions on the observation space as

$$\mathcal{H}_0 : \; Y \sim \mathcal{N}(0, \sigma^2) \quad , \qquad \mathcal{H}_1 \; : \; Y \sim \mathcal{N}(\sqrt{S}, \sigma^2)\,. \tag{3.3}$$

The likelihood ratio for (3.1) is then given by

$$L(y) = \frac{p_1(y)}{p_0(y)} = \exp\left\{\frac{\sqrt{S}}{\sigma^2}\left(y - \frac{\sqrt{S}}{2}\right)\right\}\,. \tag{3.4}$$

Since $S > 0$, the likelihood ratio $L(y)$ is a strictly increasing function of the observation y. Therefore, comparing $L(y)$ to the threshold η is equivalent to comparing y to another threshold $\eta' = L^{-1}(\eta)$, where L^{-1} is the inverse function of L. Then, the probability of false alarm is expressed as

$$\mathrm{P_{FA}} = \mathbb{E}_0\{\delta_{NP}(Y)\} = \mathrm{P}_0\left(L(Y) \geq \eta\right) = \mathrm{P}_0\left(Y \geq \eta'\right) \tag{3.5}$$

$$= Q\left(\frac{\eta'}{\sigma}\right)\,, \tag{3.6}$$

where $Q-$function is the tail probability of the standard Gaussian distribution, i.e., $Q(x) = (1/\sqrt{2\pi}\,)\int_x^\infty \mathrm{e}^{-t^2/2}dt$. It is noted that any value of false alarm probability α can be attained by choosing the threshold $\eta' = \sigma Q^{-1}(\alpha)$, where Q^{-1} is the inverse $Q-$function. Then, for fixed S, the optimal $\alpha-$level NP decision rule employed at the receiver is given by

$$\delta_{NP}(y) = \begin{cases} 1\,, & \text{if } y \geq \sigma Q^{-1}(\alpha) \\ 0\,, & \text{if } y < \sigma Q^{-1}(\alpha) \end{cases} \tag{3.7}$$

which also possesses the constant false alarm rate (CFAR) property. Let $\gamma \triangleq S/\sigma^2$ denote the normalized signal power at the receiver. Then, the detection probability achieved by δ_{NP} is obtained as

$$\mathrm{P_D}(\gamma) = \mathrm{P}_1\left(Y \geq \sigma Q^{-1}(\alpha)\right) = Q\left(Q^{-1}(\alpha) - \sqrt{\gamma}\right). \tag{3.8}$$

For fixed α, the relationship between the detection probability and γ is known as the power function of the test in radar terminology.

We will first discuss the convexity/concavity properties of the detection probability with respect to the signal power for the NP test given in (3.7). This is motivated by the possibility of enhancing the detection performance via time sharing/randomizing between two signal power levels while satisfying an average power constraint. In the absence of fading, the received power is a deterministically scaled version of the transmitted power for non-varying AWGN channels. Hence, any constraint on the transmitted power can be related to one on the received power and consecutively to one in the normalized form, and vice versa. In addition to the average power constraint, a hard limit on the peak transmitted power can be imposed as well in accordance with practical considerations.

3.2 Convexity Properties in Signal Power

3.2.1 Convexity/Concavity Results

In the following analysis, the endpoints are excluded from the set of feasible false alarm probabilities. Specifically, α is confined in the interval $(0, 1)$ excluding the trivial cases of $\alpha \in \{0, 1\}$. We first note the limits of the detection probability, i.e., $\lim_{\gamma \to 0} \mathrm{P_D}(\gamma) = \alpha$ and $\lim_{\gamma \to \infty} \mathrm{P_D}(\gamma) = 1$. Differentiating with respect to γ yields

$$\mathrm{P}'_\mathrm{D}(\gamma) = \frac{1}{2\sqrt{2\pi\gamma}} \exp\left\{ -\frac{\left(Q^{-1}(\alpha) - \sqrt{\gamma}\right)^2}{2} \right\} \tag{3.9}$$

which is positive $\forall \gamma > 0$ indicating that $\mathrm{P_D}(\gamma)$ is a strictly increasing function of γ. Similarly, the limits for the first derivative is given as $\lim_{\gamma \to 0} \mathrm{P}'_\mathrm{D}(\gamma) = \infty$ and $\lim_{\gamma \to \infty} \mathrm{P}'_\mathrm{D}(\gamma) = 0$.

Proposition 3.2.1. *For $\alpha \in [Q(2), 1)$, $\mathrm{P_D}(\gamma)$ is a monotonically increasing concave function of $\gamma \in (0, \infty)$. For $\alpha \in (0, Q(2))$, $\mathrm{P_D}(\gamma)$ is a monotonically increasing function with two inflection points such that the function is concave, convex and finally concave with respect to increasing values of γ.*

Proof. It suffices to consider the second derivative of the detection probability with respect to γ, i.e.,

$$\mathrm{P}''_\mathrm{D}(\gamma) = \frac{1}{4\sqrt{2\pi}\,\gamma} \exp\left\{ -\frac{\left(Q^{-1}(\alpha) - \sqrt{\gamma}\right)^2}{2} \right\} \left(Q^{-1}(\alpha) - \sqrt{\gamma} - \frac{1}{\sqrt{\gamma}} \right). \tag{3.10}$$

Since the first two terms in (3.10) are positive $\forall \gamma > 0$, the sign of the second derivative is determined by the third term, i.e., $(Q^{-1}(\alpha) - \sqrt{\gamma} - 1/\sqrt{\gamma})$. First, it is noted that for $\alpha \geq Q(0) = 0.5$, we have $Q^{-1}(\alpha) \leq 0$ which implies $\mathrm{P}''_\mathrm{D}(\gamma) < 0$ for all $\gamma > 0$ and the detection probability is concave. Next, let $x \triangleq \sqrt{\gamma}$. The third term in (3.10) has the reversed sign of $f(x) = x^2 - Q^{-1}(\alpha)x + 1$ for $x > 0$. The discriminant of the quadratic polynomial is given by $\Delta = (Q^{-1}(\alpha))^2 - 4$. When $\alpha \in [Q(2), Q(-2)]$, the discriminant is nonpositive $\Delta \leq 0$, and we have $f(x) \geq 0\ \forall x$ implying that $\mathrm{P}''_\mathrm{D}(\gamma) \leq 0$. Together with the previous result $\alpha \geq Q(0)$, it is concluded that $\mathrm{P_D}(\gamma)$ is concave for $\alpha \geq Q(2) \approx 0.02275013$. For $\alpha < Q(2)$, $f(x)$ has two distinct roots corresponding to the inflection points of $\mathrm{P_D}(\gamma)$, which are given as

$$\gamma_{1,2} = 0.25 \left(Q^{-1}(\alpha) \mp \sqrt{(Q^{-1}(\alpha))^2 - 4} \right)^2 \tag{3.11}$$

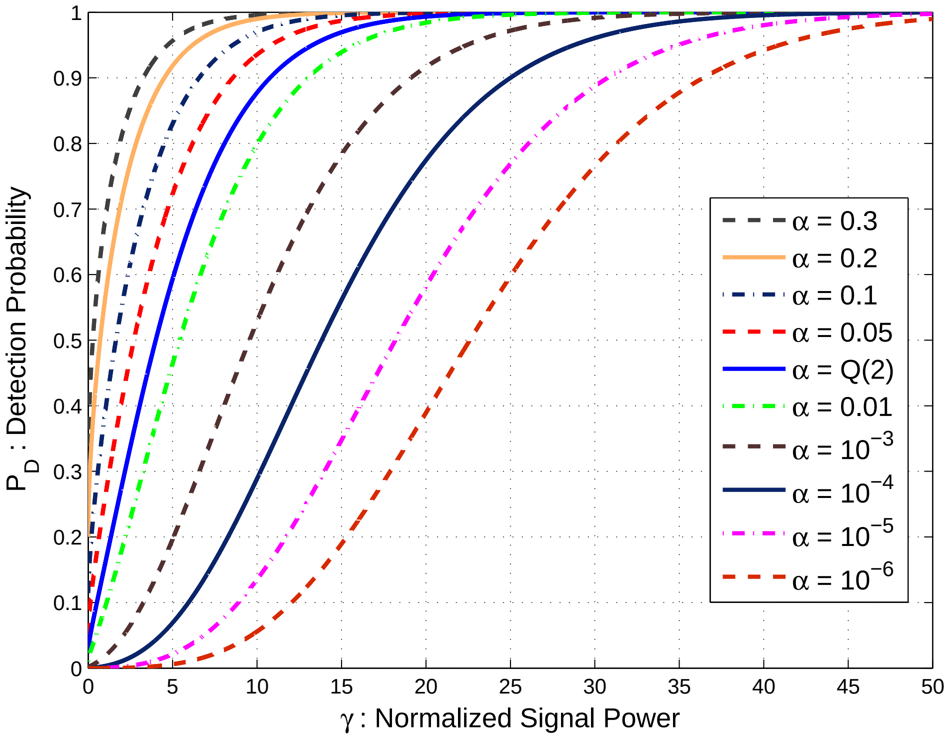

Figure 3.1: Detection probability of the NP decision rule in (3.7) is plotted versus normalized signal power γ for various values of the false alarm probability α. As an example, when $\alpha = 10^{-4}$, the inflection points are located at $\gamma_1 \approx 0.0851$ and $\gamma_2 \approx 11.7459$ with $\mathrm{P_D}(\gamma_1) \approx 0.0003$ and $\mathrm{P_D}(\gamma_2) \approx 0.3852$.

suggesting that $\mathrm{P_D}(\gamma)$ is concave for $\gamma \in (0, \gamma_1) \cup (\gamma_2, \infty)$ and convex for $\gamma \in [\gamma_1, \gamma_2]$. $\square$

Figure 3.1 depicts the detection probability of the NP decision rule in (3.7) versus γ for various values of the false alarm probability α. As expected, $\mathrm{P_D}(\gamma)$ is concave for $\alpha \in [Q(2), 1)$, and consists of concave, convex and finally concave intervals for $\alpha \in (0, Q(2))$. For the latter case, even though its existence is guaranteed, the effect of the first inflection point is far less obvious than the second inflection point. This can be attributed to the fact that for small values of α, $\gamma_1 \approx 0$ and $\mathrm{P_D}(\gamma_1) \approx \alpha$ whereas $\gamma_2 \approx (Q^{-1}(\alpha))^2$ and $\mathrm{P_D}(\gamma_2) \approx 0.5$, where the approximations are obtained using the first order Taylor series expansion.

3.2.2 Optimal Signaling

The concavity for $\alpha \in [Q(2), 1)$ stated in Proposition 3.2.1 indicates that the detection performance of an average power-limited transmitter cannot be improved by time sharing/randomizing between different power levels. Fortunately, the range of false alarm probabilities facilitating improved detection performance have higher practical significance. In order to obtain the optimal power sharing strategy, we first present the following lemma.

Lemma 3.2.1. *Let $\alpha < Q(2)$, and γ_1 and γ_2 be the inflection points of $\mathrm{P_D}(\gamma)$ as given in (3.11). There exist unique points $\gamma_{C1} \in (0, \gamma_1]$ and $\gamma_{C2} \geq \gamma_2$ such that the tangent to $\mathrm{P_D}(\gamma)$ at γ_{C1} is also tangent at γ_{C2} and this tangent lies above $\mathrm{P_D}(\gamma)$ for all $\gamma > 0$.*

Proof. In Proposition 3.2.1, it is proved that $\mathrm{P_D}(\gamma)$ is strictly convex and increasing over the interval (γ_1, γ_2), which implies that $\mathrm{P}'_\mathrm{D}(\gamma_1) < \mathrm{P}'_\mathrm{D}(\gamma_2)$. On the contrary, $\mathrm{P}'_\mathrm{D}(\gamma)$ is monotonically decreasing over the intervals $(0, \gamma_1)$ and (γ_2, ∞). Furthermore, $\mathrm{P}'_\mathrm{D}(\gamma)$ is continuous $\forall \gamma > 0$. Since $\lim_{\gamma \to 0} \mathrm{P}'_\mathrm{D}(\gamma) = \infty$, there exists a unique point $\gamma_{1x} \in (0, \gamma_1]$ such that $\mathrm{P}'_\mathrm{D}(\gamma_{1x}) = \mathrm{P}'_\mathrm{D}(\gamma_2)$. Similarly, there exists a unique point $\gamma_{2x} \in [\gamma_2, \infty)$ such that $\mathrm{P}'_\mathrm{D}(\gamma_{2x}) = \mathrm{P}'_\mathrm{D}(\gamma_1)$ since $\lim_{\gamma \to \infty} \mathrm{P}'_\mathrm{D}(\gamma) = 0$. As a result, for every $\hat{\gamma}_1 \in [\gamma_{1x}, \gamma_1]$ there exists a unique point $\hat{\gamma}_2 \in [\gamma_2, \gamma_{2x}]$ such that $\mathrm{P}'_\mathrm{D}(\hat{\gamma}_1) = \mathrm{P}'_\mathrm{D}(\hat{\gamma}_2)$. In other words, we can define a one-to-one continuous function as $\hat{\gamma}_2(\hat{\gamma}_1) = (\mathrm{P}'_\mathrm{D})^{-1}\left(\mathrm{P}'_\mathrm{D}(\hat{\gamma}_1)\right)$. Now, consider the function $f(\gamma, \hat{\gamma}_1) = \mathrm{P_D}(\gamma) - \left(\mathrm{P}'_\mathrm{D}(\hat{\gamma}_1)(\gamma - \hat{\gamma}_1) + \mathrm{P_D}(\hat{\gamma}_1)\right)$, which gives the vertical difference between $\mathrm{P_D}(\gamma)$ and the tangent to $\mathrm{P_D}(\gamma)$ at $\hat{\gamma}_1$. Recall that $\partial f / \partial \gamma = \mathrm{P}'_\mathrm{D}(\gamma) - \mathrm{P}'_\mathrm{D}(\hat{\gamma}_1)$ has a zero at some point $\hat{\gamma}_2 \in [\gamma_2, \gamma_{2x}]$. We can define the following continuous function: $h(\hat{\gamma}_1) \triangleq f(\hat{\gamma}_2(\hat{\gamma}_1), \hat{\gamma}_1) = \mathrm{P_D}(\hat{\gamma}_2) - \mathrm{P_D}(\hat{\gamma}_1) - \mathrm{P}'_\mathrm{D}(\hat{\gamma}_1)(\hat{\gamma}_2 - \hat{\gamma}_1)$. By differentiation, it is observed that $h(\cdot)$ is an increasing function: $\partial h(\hat{\gamma}_1)/\partial \hat{\gamma}_1 = \mathrm{P}'_\mathrm{D}(\hat{\gamma}_2)\hat{\gamma}'_2 - \mathrm{P}'_\mathrm{D}(\hat{\gamma}_1) - \mathrm{P}''_\mathrm{D}(\hat{\gamma}_1)(\hat{\gamma}_2 - \hat{\gamma}_1) - \mathrm{P}'_\mathrm{D}(\hat{\gamma}_1)(\hat{\gamma}'_2 - 1) = -\mathrm{P}''_\mathrm{D}(\hat{\gamma}_1)(\hat{\gamma}_2 - \hat{\gamma}_1) > 0$ where the last equality follows from $\mathrm{P}'_\mathrm{D}(\hat{\gamma}_1) = \mathrm{P}'_\mathrm{D}(\hat{\gamma}_2)$ and the inequality is due to the strict concavity of $\mathrm{P_D}(\hat{\gamma}_1)$ over $\hat{\gamma}_1 \in [\gamma_{1x}, \gamma_1]$ and $\hat{\gamma}_2 > \hat{\gamma}_1$. By substituting

$\hat{\gamma}_1 = \gamma_{1x}$, we have $\hat{\gamma}_2 = \gamma_2$ and $h(\gamma_{1x}) = \mathrm{P_D}(\gamma_2) - \mathrm{P_D}(\gamma_{1x}) - \mathrm{P}'_{\mathrm{D}}(\gamma_{1x})(\gamma_2 - \gamma_{1x}) \leq 0$. The last inequality follows by noting that $\mathrm{P}'_{\mathrm{D}}(\gamma) \leq \mathrm{P}'_{\mathrm{D}}(\gamma_{1x})$ for $\gamma \in [\gamma_{1x}, \gamma_2]$ and $\mathrm{P_D}(\gamma_2) = \mathrm{P_D}(\gamma_{1x}) + \int_{\gamma_{1x}}^{\gamma_2} \mathrm{P}'_{\mathrm{D}}(\gamma) d\gamma$. On the other extreme, when $\hat{\gamma}_1 = \gamma_1$, we have $\hat{\gamma}_2 = \gamma_{2x}$ and $h(\gamma_1) = \mathrm{P_D}(\gamma_{2x}) - \mathrm{P_D}(\gamma_1) - \mathrm{P}'_{\mathrm{D}}(\gamma_1)(\gamma_{2x} - \gamma_1) \geq 0$. Again, the inequality follows from $\mathrm{P}'_{\mathrm{D}}(\gamma) \geq \mathrm{P}'_{\mathrm{D}}(\gamma_1)$ for $\gamma \in [\gamma_1, \gamma_{2x}]$ and $\mathrm{P_D}(\gamma_{2x}) = \mathrm{P_D}(\gamma_1) + \int_{\gamma_1}^{\gamma_{2x}} \mathrm{P}'_{\mathrm{D}}(\gamma) d\gamma$. Since $h(\cdot)$ is a continuous increasing function, it must have a unique root $\gamma_{C1} \in [\gamma_{1x}, \gamma_1]$. Consequently, tangent to $\mathrm{P_D}(\gamma)$ at γ_{C1} is also tangent at some point $\gamma_{C2} = (\mathrm{P}'_{\mathrm{D}})^{-1}\left(\mathrm{P}'_{\mathrm{D}}(\gamma_{C1})\right) \in [\gamma_2, \gamma_{2x}]$.

Next, we prove that the tangent lies above $\mathrm{P_D}(\gamma)$ for all $\gamma > 0$. Since $\mathrm{P_D}(\gamma)$ is strictly concave for $(0, \gamma_1)$, the tangent at γ_{C1} lies above $\mathrm{P_D}(\gamma)$ for $\gamma \in (0, \gamma_1)$. Recall that the same line is also tangent to $\mathrm{P_D}(\gamma)$ at γ_{C2} and as a result, it lies above $\mathrm{P_D}(\gamma)$ for $\gamma > \gamma_2$. Subsequently, the line connecting the points $(\gamma_1, \mathrm{P_D}(\gamma_1))$ and $(\gamma_2, \mathrm{P_D}(\gamma_2))$ lies above $\mathrm{P_D}(\gamma)$ for $\gamma \in [\gamma_1, \gamma_2]$ since $\mathrm{P_D}(\gamma)$ is convex over this interval. Since the inflection points $(\gamma_1, \mathrm{P_D}(\gamma_1))$ and $(\gamma_2, \mathrm{P_D}(\gamma_2))$ are below the tangent line, the line connecting them also lies below the tangent line. This proves that the tangent line is above $\mathrm{P_D}(\gamma)$ for all $\gamma > 0$. $\square$

Using a similar analysis to that in the proof of Lemma 3.2.1, we can also obtain the following lemma.

Lemma 3.2.2. *Let $\alpha < Q(2)$, and γ_1 and γ_2 be the inflection points of $\mathrm{P_D}(\gamma)$. Suppose also that γ_{C1} and γ_{C2} are the contact points of the tangent line as described in Lemma 1. Given a point $\hat{\gamma} \in [\gamma_1, \gamma_{C2}]$, there exists a unique point $\gamma_C(\hat{\gamma}) \in [\gamma_{C1}, \gamma_1]$ such that the tangent at $\gamma_C(\hat{\gamma})$ passes through the point $(\hat{\gamma}, \mathrm{P_D}(\hat{\gamma}))$ and lies above $\mathrm{P_D}(\gamma)$ for all $\gamma \in (0, \hat{\gamma})$.*[1]

[1] The dependence of tangent point γ_C to $\hat{\gamma}$ is explicitly emphasized by writing it as a function, i.e., $\gamma_C(\hat{\gamma})$.

Based on Lemma 3.2.1 and Lemma 3.2.2, we state the optimal signaling strategy for the communications system in (3.1) operating under peak power constraint Γ_{peak} and average power constraint Γ_{avg} ($\Gamma_{\text{avg}} \leq \Gamma_{\text{peak}}$).

Proposition 3.2.2. *Let $\alpha < Q(2)$. For $\Gamma_{\text{avg}} \leq \gamma_{C1}$ or $\Gamma_{\text{avg}} \geq \gamma_{C2}$ or $\Gamma_{\text{peak}} \leq \gamma_1$, the best strategy is to exclusively transmit at the average power Γ_{avg}, i.e., power sharing/randomization does not help. When $\Gamma_{\text{avg}} \in (\gamma_{C1}, \gamma_{C2})$ and $\gamma_{C2} \leq \Gamma_{\text{peak}}$, the optimal strategy is to time share/randomize between powers γ_{C1} and γ_{C2} with the fraction of time $(\gamma_{C2} - \Gamma_{\text{avg}})/(\gamma_{C2} - \gamma_{C1})$ allocated to the power γ_{C1}. On the contrary if $\Gamma_{\text{avg}} \in [\gamma_C(\Gamma_{\text{peak}}), \Gamma_{\text{peak}}]$ while $\Gamma_{\text{peak}} \in (\gamma_1, \gamma_{C2})$, the optimal strategy is to time share/randomize between powers $\gamma_C(\Gamma_{\text{peak}})$ and the peak power Γ_{peak} with the fraction of time $(\Gamma_{\text{peak}} - \Gamma_{\text{avg}})/(\Gamma_{\text{peak}} - \gamma_C(\Gamma_{\text{peak}}))$ allocated to the power $\gamma_C(\Gamma_{\text{peak}})$. Consequently, if $\Gamma_{\text{avg}} < \gamma_C(\Gamma_{\text{peak}})$ while $\Gamma_{\text{peak}} \in (\gamma_1, \gamma_{C2})$, transmitting continuously at Γ_{avg} is the optimal strategy.*

The proof can be established in a straightforward manner by showing that the proposed strategy results in the smallest concave function that is larger than $\mathrm{P_D}(\gamma)$ for $\gamma \in (0, \Gamma_{\text{peak}}]$ which corresponds to the upper boundary of the convex hull of $\mathrm{P_D}(\gamma)$ within the same interval [19]. If we do not pay attention to the peak power constraint for a second, these results indicate that very weak and strong transmitters should operate continuously at their average power while transmitters with moderate power can benefit significantly from power sharing strategies.

The critical points γ_{C1} and γ_{C2} can be obtained as the unique pair that satisfies $\mathrm{P}'_{\mathrm{D}}(\gamma_{C1}) = \mathrm{P}'_{\mathrm{D}}(\gamma_{C2}) = (\mathrm{P_D}(\gamma_{C2}) - \mathrm{P_D}(\gamma_{C1}))\,/\,(\gamma_{C2} - \gamma_{C1})$, which can be solved numerically by plugging in the corresponding expressions. Since the simultaneous solution of these equality constraints can be difficult due to terms involving exponentials and $Q-$functions, we propose the following problem to construct the optimal signaling strategy:

$$\begin{aligned} \max_{\lambda,\gamma_{C1},\gamma_{C2}} \quad & \lambda\, Q\left(Q^{-1}(\alpha)-\sqrt{\gamma_{C1}}\right)+(1-\lambda)Q\left(Q^{-1}(\alpha)-\sqrt{\gamma_{C2}}\right) \\ \text{s.t.} \quad & \lambda\,\gamma_{C1}+(1-\lambda)\gamma_{C2}\le \Gamma_{\text{avg}} \end{aligned} \tag{3.12}$$

where $\gamma_{C1} \in (0, \gamma_1]$, $\gamma_{C2} \in [\gamma_2, \Gamma_{\text{peak}}]$, and $\lambda \in [0, 1]$ denotes the fraction of time power γ_{C1} is used assuming $\Gamma_{\text{peak}} \geq \gamma_{C2}$ and $\Gamma_{\text{avg}} \in [\gamma_{C1}, \gamma_{C2}]$. The employed solver can be initialized with $\gamma_{C1} = \gamma_1$, $\gamma_{C2} = \gamma_2$, and $\lambda = (\gamma_2 - \Gamma_{\text{avg}})/(\gamma_2 - \gamma_1)$. As an example, for $\alpha = 10^{-4}$, $\Gamma_{\text{avg}} = 5$, and $\Gamma_{\text{peak}} = 20$, the optimal strategy can achieve a detection probability of 0.1946 by employing power $\gamma_{C1} = 2.69 \times 10^{-5}$ with probability 0.7307 and power $\gamma_{C2} = 18.5664$ with probability 0.2693, whereas by exclusively transmitting at the average power, the detection probability remains at 0.0690. [2] If the peak power constraint is lowered to $\Gamma_{\text{peak}} = 10$, the optimal strategy can still increase the detection probability to 0.1445 by randomizing between $\gamma_C = 4.99 \times 10^{-5}$ and peak power $\Gamma_{\text{peak}} = 10$ with approximately equal probabilities as suggested by the solution of $\mathrm{P}'_{\mathrm{D}}(\gamma_C) = (\mathrm{P}_{\mathrm{D}}(\Gamma_{\text{peak}}) - \mathrm{P}_{\mathrm{D}}(\gamma_C)) \,/\, (\Gamma_{\text{peak}} - \gamma_C)$. Finally, it should be emphasized that the detection probability can be improved even further by designing the optimal signaling scheme jointly with the detector employed at the receiver, which will be discussed in the next chapter. However, in that case we need to sacrifice from the simplistic structure of the threshold detector which is also easier to update if the channel statistics change slowly over time.

3.2.3 Near-optimal Strategy

We recall from the previous discussion that for small values of the false alarm probability, the first inflection point γ_1 gets close to zero. It is also stated above that the value of $\mathrm{P}_{\mathrm{D}}(\gamma_1)$ equals approximately to α in that case. Since the critical

[2] Numerical results are obtained using MATLAB's *multistart* method and *sqp* algorithm together with the local solver *fmincon*.

points γ_C and γ_{C1} are located inside the interval $(0, \gamma_1]$, they get close to zero as well while the corresponding detection probabilities approach α. Also evident from the example above, this observation gives clues of a suboptimal approach. We make a simplifying assumption and suppose that $\mathrm{P_D}(\gamma)$ is convex over the interval $(0, \gamma_2)$. Using arguments similar to those in the Appendix, it is then possible to show that there exists a unique point $\gamma_{\text{on}} \geq \gamma_2$ such that the tangent to $\mathrm{P_D}(\gamma)$ at γ_{on} passes through the point $(0, \alpha)$. Then, γ_{on} can be obtained from $\mathrm{P_D}(\gamma_{\text{on}}) - \gamma_{\text{on}}\mathrm{P}'_{\mathrm{D}}(\gamma_{\text{on}}) = \alpha$. More explicitly, we need to solve for $\hat{x}$ such that

$$\hat{x} = Q^{-1}\left(\frac{Q^{-1}(\alpha) - \hat{x}}{2\sqrt{2\pi}} \exp\left\{-\frac{\hat{x}^2}{2}\right\} + \alpha\right) \tag{3.13}$$

and the contact point can be obtained by substituting $\gamma_{\text{on}} = \left(Q^{-1}(\alpha) - \hat{x}\right)^2$. The form of the equation in (3.13) suggests that a fixed point iteration can be employed to obtain the solution [60]. This observation leads to the following near-optimal strategy in the case of strict false alarm requirements.

Near-optimal strategy: Let $\alpha < Q(2)$. A suboptimal strategy with reasonable performance is to switch between powers 0 and γ_{on} with the fraction of on-power time $\Gamma_{\text{avg}}/\gamma_{\text{on}}$ when $\Gamma_{\text{avg}} < \gamma_{\text{on}} < \Gamma_{\text{peak}}$. For $\gamma_{\text{on}} \geq \Gamma_{\text{peak}}$, the proposed suboptimal strategy randomizes between powers 0 and Γ_{peak} with the fraction of on-power time $\Gamma_{\text{avg}}/\Gamma_{\text{peak}}$. For $\Gamma_{\text{avg}} > \gamma_{\text{on}}$, the transmission is conducted exclusively at the average power.

Figure 3.2 provides more insight about the near-optimal performance of the proposed approach. For various values of the false alarm probability α, we have computed the inflection points γ_1 and γ_2 from (3.11), evaluated the corresponding detection probabilities $\mathrm{P_D}(\gamma_1)$ and $\mathrm{P_D}(\gamma_2)$, respectively, and plotted the resulting detection performance curves with respect to α. As the false alarm constraint is tightened (smaller values), it is observed that the vertical gap between the detection performances calculated at the respective inflection points becomes much more pronounced. Since $\mathrm{P_D}(\gamma)$ is monotonically increasing and $\gamma_{C1} \leq \gamma_1$ is assured from Lemma 3.2.1, $\mathrm{P_D}(\gamma_{C1})$ always takes values smaller than $\mathrm{P_D}(\gamma_1)$,

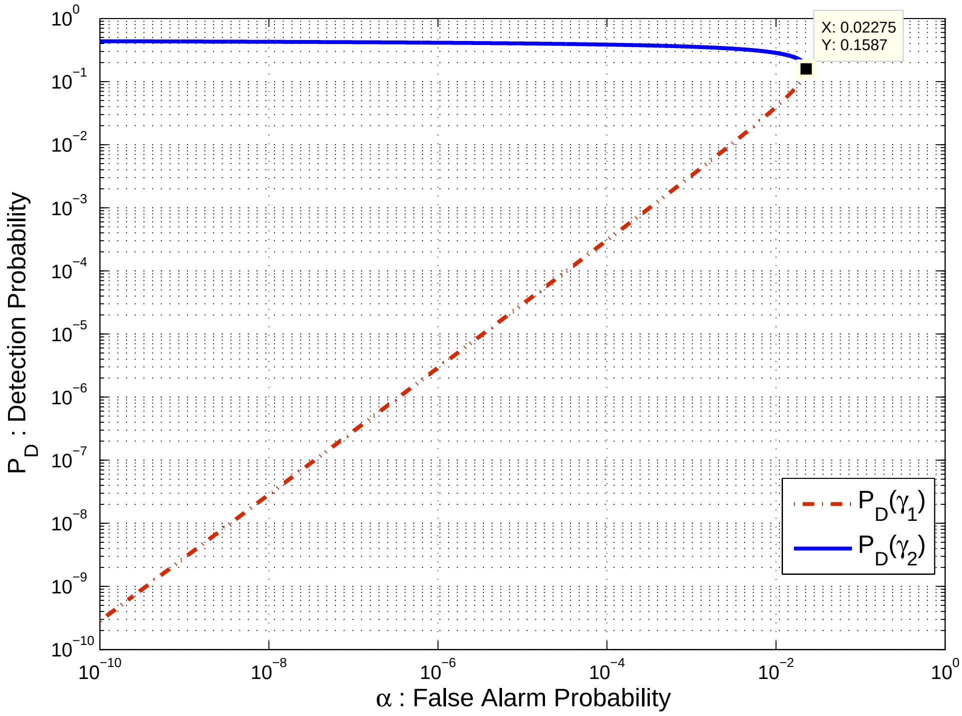

Figure 3.2: Detection probability of the NP decision rule in (3.7) is evaluated at the inflection points γ_1 and γ_2.

which is denoted with the red curve. On the contrary, the detection probability corresponding to the larger contact point γ_{C2} results in $\mathrm{P_D}(\gamma_{C2}) \geq \mathrm{P_D}(\gamma_2)$, which is represented by the blue curve. For a given α, the optimal strategy stated in Proposition 3.2.2 randomizes between γ_{C1} and γ_{C2}, whose contributions to the detection performance should therefore lie below the red curve and above the blue curve, respectively. As a result, the contribution from the smaller contact point γ_{C1} can safely be ignored over a large set of false alarm probabilities without sacrificing from the detection performance claimed by the optimal strategy stated in Proposition 3.2.2.

3.2.4 Extension to Multidimensional Case

As mentioned earlier in the Introduction, when the observations acquired by the receiver are corrupted with colored Gaussian noise, the detection probability can be maximized by transmitting along the eigenvector corresponding to the minimum eigenvalue of the noise covariance matrix [20]. More specifically, we consider the following hypothesis-testing problem where, given an M dimensional data vector, we have to decide between $\mathcal{H}_0 : \mathbf{Y} = \mathbf{N}$ and $\mathcal{H}_1 : \mathbf{Y} = \sqrt{S}\mathbf{v}_{min} + \mathbf{N}$, where $\mathbf{N} \sim \mathcal{N}(0, \mathbf{\Sigma})$ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{\Sigma}$, and $\mathbf{v}_{min}$ is the normalized eigenvector corresponding to the minimum eigenvalue of $\mathbf{\Sigma}$ with $|\mathbf{v}_{min}|^2 = 1$. It should be pointed out that a feedback mechanism is required from the receiver to the transmitter in order to facilitate signaling along the least noisy direction. In the absence of such a mechanism, the following analysis provides an upper bound on the detection performance.

At the receiver, the optimal correlation detector employs the decision statistics $T(\mathbf{y}) = \mathbf{v}_{min}^T \cdot \mathbf{y}$, which is a linear combination of jointly Gaussian random variables. Hence, the hypotheses can be rewritten as $\mathcal{H}_0 : T(\mathbf{Y}) \sim \mathcal{N}(0, \lambda_{min})$ and $\mathcal{H}_1 : T(\mathbf{Y}) \sim \mathcal{N}(\sqrt{S}, \lambda_{min})$, where λ_{min} denotes the minimum eigenvalue of $\mathbf{\Sigma}$ [20]. From the false alarm constraint, the detector threshold can be obtained as $\mathrm{P_{FA}} = \mathrm{P}_0\left(T(\mathbf{Y}) \geq \eta\right) = Q\left(\eta/\sqrt{\lambda_{min}}\right) = \alpha$ and $\eta = \sqrt{\lambda_{min}}Q^{-1}(\alpha)$. The corresponding optimal NP decision rule is given as

$$\delta_{NP}(\mathbf{Y}) = \begin{cases} 1 & \text{if } \mathbf{v}_{min}^T \cdot \mathbf{y} \geq \sqrt{\lambda_{min}}Q^{-1}(\alpha) \\ 0 & \text{if } \mathbf{v}_{min}^T \cdot \mathbf{y} < \sqrt{\lambda_{min}}Q^{-1}(\alpha) \end{cases} \tag{3.14}$$

By defining $\gamma \triangleq S/\lambda_{min}$, the detection probability attained by δ_{NP} is computed from $\mathrm{P_D}(\gamma) = \mathrm{P}_1(T(\mathbf{Y}) \geq \sqrt{\lambda_{min}}\, Q^{-1}(\alpha)) = Q\left(Q^{-1}(\alpha) - \sqrt{\gamma}\right)$. Notice that this expression is exactly in the same form as (3.8) after replacing σ^2 with λ_{min} and similar results to those in Section 3.2 can be obtained in this multidimensional setting.

3.3 Convexity Properties in Noise Power

In this section, we investigate the binary hypothesis testing problem stated in (3.1) from the perspective of a power constrained jammer. By assuming signal power S to be fixed, we aim to determine the optimal power allocation strategy for a Gaussian jammer in order to minimize the detection probability at the receiver. The power of the jammer is controlled over time through the variable σ. Considering a smart receiver, it is assumed that the value of σ is learned instantly and the detection threshold in (3.7) is updated to maintain a constant false alarm probability α. It should be pointed out that the receiver can improve its detection performance by employing the optimal NP decision rule corresponding to the power distribution of the stochastic Gaussian jammer. However, in this study we assume that the receiver keeps the threshold detector to exploit reduced costs and ease of adaptability. On the other hand, jamming would be performed more effectively if the receiver could not adapt to varying noise power instantaneously.

Under constant transmit power S, the detection probability as a function of the normalized jammer power, $\beta \triangleq \sigma^2/S$, can be expressed as $\mathrm{P_D}(\beta) = Q\left(Q^{-1}(\alpha) - \beta^{-1/2}\right)$. The limits can be computed as $\lim_{\beta\to 0}\mathrm{P_D}(\beta) = 1$ and $\lim_{\beta\to\infty}\mathrm{P_D}(\beta) = \alpha$. Differentiating with respect to β yields $\mathrm{P}'_\mathrm{D}(\beta) = -(2\sqrt{2\pi})^{-1}\beta^{-3/2}\exp\Big\{-0.5\big(Q^{-1}(\alpha)-\beta^{-1/2}\big)^2\Big\}$, which is negative $\forall\, \beta > 0$. The limits for the first derivative are $\lim_{\beta\to 0}\mathrm{P}'_\mathrm{D}(\beta) = 0$ and $\lim_{\beta\to\infty}\mathrm{P}'_\mathrm{D}(\beta) = 0$.

Proposition 3.3.1. $\mathrm{P_D}(\beta)$ *is a monotonically decreasing function of* $\beta \in (0,\infty)$ *with a single inflection point at*

$$\beta^* = \left(\frac{\sqrt{(Q^{-1}(\alpha))^2+12} - Q^{-1}(\alpha)}{6}\right)^2. \tag{3.15}$$

Proof. The second derivative of the detection probability is $\mathrm{P}''_\mathrm{D}(\beta) = (4\sqrt{2\pi})^{-1}\beta^{-7/2}\exp\Big\{-0.5\big(Q^{-1}(\alpha)-\beta^{-1/2}\big)^2\Big\}\big(3\beta + Q^{-1}(\alpha)\sqrt{\beta} - 1\big)$. As before, the sign of the second derivative is determined by the left most expression in

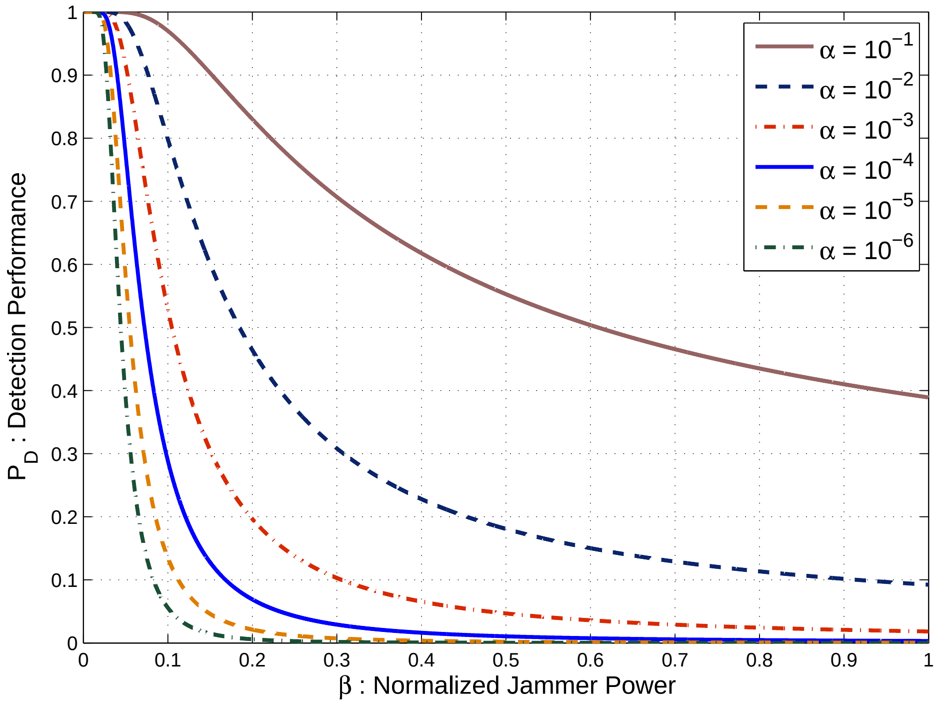

Figure 3.3: Detection probability of the NP decision rule in (3.7) is plotted versus normalized jammer power β for various values of the false alarm probability α. As an example, when $\alpha = 10^{-4}$, the inflection point is located at $\beta^* \approx 0.05164$ with $\mathrm{P_D}(\beta^*) \approx 0.7523$.

parentheses. By substituting $x \triangleq \sqrt{\beta}$, the roots of the resulting quadratic polynomial are obtained as $\big(-Q^{-1}(\alpha) \pm \sqrt{(Q^{-1}(\alpha))^2 + 12}\big)/6$. Since $x = \sqrt{\beta} > 0$, the positive root results in the inflection point given in (3.15) indicating that $\mathrm{P_D}(\beta)$ is strictly concave for $\beta < \beta^*$ and strictly convex for $\beta > \beta^*$. $\square$

The detection performance of the NP detector given by (3.7) is depicted in Figure 3.3 versus β for various values of the false alarm probability α, which point out the possibility of decreasing the detection probability via time-sharing of the jammer noise power. In order to obtain the optimal power sharing strategy for the jammer, we first present the following lemma which can be proved using a similar approach to that of Lemma 3.2.1.

Lemma 3.3.1. *Let* β^* *be the inflection point of* $\mathrm{P_D}(\beta)$ *as given in* (3.15)*. There exists a unique point* $\beta_C \geq \beta^*$ *such that the tangent to* $\mathrm{P_D}(\beta)$ *at* β_C *lies below* $\mathrm{P_D}(\beta)$ *and passes through the point* $(0, 1)$*.*

The contact point β_C can be obtained from $\mathrm{P_D}(\beta_C) - \beta_C \mathrm{P}'_{\mathrm{D}}(\beta_C) = 1$, or equivalently solving for $\hat{x}$ in

$$\hat{x} = Q^{-1}\left(1 - \frac{Q^{-1}(\alpha) - \hat{x}}{2\sqrt{2\pi}} \exp\left\{-\frac{\hat{x}^2}{2}\right\}\right) \tag{3.16}$$

and then substituting into $\beta_C = (Q^{-1}(\alpha) - \hat{x})^{-2}$. In addition to the fixed point iteration approach, *fzero* function provided in MATLAB which implements Brent's method can successfully return β_C [61, Chapter 4].

Next, we present the optimal strategy for a Gaussian jammer operating under peak power constraint $\mathrm{J_{peak}}$ and average power constraint $\mathrm{J_{avg}}$ ($\mathrm{J_{avg}} \leq \mathrm{J_{peak}}$) towards a smart receiver employing the adaptable threshold detector given in (3.7).

Proposition 3.3.2. *The jammer's optimal strategy is to switch between powers 0 and* β_C *with the fraction of on-power time* $\mathrm{J_{avg}}/\beta_C$ *when* $\mathrm{J_{avg}} < \beta_C < \mathrm{J_{peak}}$*. For* $\beta_C \geq \mathrm{J_{peak}}$*, the optimal strategy randomizes between powers 0 and* $\mathrm{J_{peak}}$ *with the fraction of on-power time* $\mathrm{J_{avg}}/\mathrm{J_{peak}}$*. For* $\mathrm{J_{avg}} > \beta_C$*, jamming is performed continuously at the average power.*

Again the proof follows by noting that the stated strategy results in the largest convex function that is smaller than $\mathrm{P_D}(\beta)$ for $\beta \in [0, \mathrm{J_{peak}}]$. Finally as an example, for $\alpha = 10^{-4}$, $\mathrm{J_{avg}} = 0.04$, and $\mathrm{J_{peak}} = 0.1$, on-off Gaussian jamming can reduce the detection probability from 0.8999 down to 0.7109 by transmitting with power $\beta_C = 0.08779$ for approximately 45.56 percent of the time and aborting jamming for 54.44 percent of the time. If the peak power constraint is lowered to $\mathrm{J_{peak}} = 0.06$, the optimal strategy can still decrease the detection probability to 0.7612 by randomizing between 0 and peak power $\mathrm{J_{peak}} = 0.06$ with two-thirds of on-power time fraction.

3.4 Concluding Remarks

In this chapter, we have examined the convexity/concavity properties of the detection probability for the problem of determining the presence of a target signal immersed in additive Gaussian noise. Unnoticed in the previous literature, we have found out that the detection performance of a power constrained transmitter can be increased via time-sharing between different levels whenever the false alarm requirement is smaller than $Q(2) \approx 0.02275$. Although the optimal strategy indicates a randomization between two nonzero power levels for moderate values of the power constraint, it is shown that the on-off signaling strategy can well approximate the optimal performance. Next, we have considered the dual problem for a power constrained jammer and proved the existence of a critical power level up to which on-off jamming can be employed to degrade the detection performance of a smart receiver.

4

Optimal Stochastic Signal Design and Detector Randomization for Power Constrained On-Off Keying Systems in Neyman-Pearson Framework

In this chapter, we extend the work conducted in the previous chapter to Neyman-Pearson detection over channels with arbitrary noise PDFs. Section 4.1 discusses the joint optimal design of the signaling scheme and the decision rule for the case of a single detector at the receiver. Section 4.2 considers the case of multiple detectors at the receiver and states the solution to to the most generic problem which requires the joint optimal design of decision rules, stochastic signals, and detector randomization factors. A detection example is presented to justify the performance improvements due to stochastic signaling and detector randomization in Section 4.3.

4.1 Case 1: Single Detector at the Receiver

Consider an on-off keying communications system, in which the receiver acquires M-dimensional observations over an additive noise channel and decides between the two hypotheses H_0 or H_1, which are modeled as

$$H_0 \,:\, \mathbf{Y} = \mathbf{N} \;\;, \quad H_1 \,:\, \mathbf{Y} = \mathbf{S} + \mathbf{N} \tag{4.1}$$

where $\mathbf{Y}$ is the noisy observation vector, $\mathbf{S}$ represents the transmitted signal for the alternative hypothesis (H_1), and $\mathbf{N}$ is the noise component that is independent of $\mathbf{S}$. Instead of using a constant level for $\mathbf{S}$ as in the conventional case, one can consider a more generic scenario in which the signal $\mathbf{S}$ can be stochastic. Then, the aim is to find the optimal PDF for $\mathbf{S}$ in (4.1) and the corresponding decision rule that maximize the probability of detection under the constraints on the probability of false alarm and average transmit power. A feedback mechanism from the receiver to the transmitter is assumed to facilitate the joint optimization of the signaling structure and the decision rule, which is a reasonable assumption, for example, for cognitive radio (CR) systems.

Note that the probability distribution of the noise component in (4.1) is not necessarily Gaussian. Due to interference, such as multiple-access interference, the noise component can have a significantly different probability distribution from the Gaussian distribution [62].

At the receiver, the structure of a randomized test is assumed to choose between the two hypotheses. Such a test is completely characterized by a decision rule ϕ. For a given observation vector $\mathbf{y}$, this test accepts hypothesis H_1 with probability $\phi(\mathbf{y})$, and rejects it with probability $1 - \phi(\mathbf{y})$ where $0 \leq \phi(\mathbf{y}) \leq 1$ for all $\mathbf{y}$. If ϕ takes on only the values 0 and 1, the test reduces to a nonrandomized one, and ϕ simply becomes the indicator function of the decision region [28].

In the NP framework, for a given value of $\alpha \in (0,\, 1)$, the aim is to maximize the probability of detection such that the probability of false alarm does not exceed α. In other words, the tradeoff between type-I and type-II errors is taken into account in the NP approach [20]. Given the decision rule (detector) ϕ, the two probabilities of interest, the probability of detection $\mathrm{P_D}$ and the probability of false alarm $\mathrm{P_{FA}}$, can be calculated as follows:

$$\begin{aligned} \mathrm{P_D} &= \mathbb{E}_1\left\{\phi(\mathbf{Y})\right\} = \int_{\mathbb{R}^M} \phi(\mathbf{y})\, p_1(\mathbf{y})\, \mathrm{d}\mathbf{y} \\ \mathrm{P_{FA}} &= \mathbb{E}_0\left\{\phi(\mathbf{Y})\right\} = \int_{\mathbb{R}^M} \phi(\mathbf{y})\, p_0(\mathbf{y})\, \mathrm{d}\mathbf{y} \end{aligned} \tag{4.2}$$

where $p_i(\mathbf{y})$ denotes the conditional PDF of the observation when hypothesis H_i is assumed to be true for $i \in \{0,\, 1\}$, and the subscripts on the expectation operators indicate the corresponding hypotheses. Since stochastic signaling is considered, $\mathbf{S}$ in (4.1) is modeled as a random vector. Recalling that the signal and the noise are independent, the conditional PDF of the observation under the alternative hypothesis H_1 can be calculated as $p_1(\mathbf{y}) = \int_{\mathbb{R}^M} p_{\mathbf{S}}(\mathbf{x})\, p_{\mathbf{N}}(\mathbf{y}-\mathbf{x})\, \mathrm{d}\mathbf{x} = \mathbb{E}\left\{p_{\mathbf{N}}\left(\mathbf{y}-\mathbf{S}\right)\right\}$, where the expectation is taken over the PDF of $\mathbf{S}$. On the other hand, the conditional PDF of the observation under the null hypothesis H_0 is given simply by $p_0(\mathbf{y}) = p_{\mathbf{N}}(\mathbf{y})$. Then, using the linearity of the expectation operator, the probability of detection can be expressed as

$$\begin{aligned} \mathrm{P_D} &= \int_{\mathbb{R}^M} \phi(\mathbf{y})\, \mathbb{E}\left\{p_{\mathbf{N}}\left(\mathbf{y}-\mathbf{S}\right)\right\}\, \mathrm{d}\mathbf{y} \\ &= \mathbb{E}\left\{\int_{\mathbb{R}^M} \phi(\mathbf{y})\, p_{\mathbf{N}}\left(\mathbf{y}-\mathbf{S}\right)\, \mathrm{d}\mathbf{y}\right\} \triangleq \mathbb{E}\left\{h(\phi;\mathbf{S})\right\} \end{aligned} \tag{4.3}$$

and the probability of false alarm $\mathrm{P_{FA}}$ is given by

$$\mathrm{P_{FA}} = \int_{\mathbb{R}^M} \phi(\mathbf{y})\, p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y}\,. \tag{4.4}$$

In practical systems, there is a constraint on the average power emitted from the transmitter. Under the framework of stochastic signaling, this constraint on the average power can be expressed in the following form [20]:

$$\mathbb{E}\left\{\|\mathbf{S}\|_2^2\right\} \leq \mathrm{A} \tag{4.5}$$

where A denotes the average power limit.

One of the main motivations behind this study is to understand how stochastic signaling can help improve the detection performance of an on-off keying system without violating the constraint on the false alarm probability. Under the NP decision criterion, the optimal signaling and detector design problem can then be stated as

$$\begin{aligned} \max_{\{\phi,\, p_{\mathbf{S}}\}} \quad & \mathbb{E}\left\{h(\phi;\mathbf{S})\right\} \\ \text{subject to} \quad & \mathrm{P_{FA}} \leq \alpha \quad \text{and} \quad \mathbb{E}\left\{\|\mathbf{S}\|_2^2\right\} \leq \mathrm{A} \end{aligned} \tag{4.6}$$

where the expectations are taken over the PDF of $\mathbf{S}$, $\alpha \in (0,\, 1)$, and $h(\phi;\mathbf{S})$ and $\mathrm{P_{FA}}$ are as in (4.3) and (4.4), respectively. Note that there are also implicit constraints in the optimization problem in (4.6), since $p_{\mathbf{S}}(\cdot)$ represents a PDF. Namely, $p_{\mathbf{S}}(\mathbf{x}) \geq 0\,,\ \forall \mathbf{x} \in \mathbb{R}^M$, and $\int_{\mathbb{R}^M} p_{\mathbf{S}}(\mathbf{x})\, \mathrm{d}\mathbf{x} = 1$ should also be satisfied by the optimal solution.

Based on (4.3) and (4.4), a more explicit version of (4.6) can be expressed as

$$\begin{aligned} \max_{\{\phi,\, p_{\mathbf{S}}\}} \quad & \mathbb{E}\left\{\int_{\mathbb{R}^M} \phi(\mathbf{y})\, p_{\mathbf{N}}\left(\mathbf{y}-\mathbf{S}\right)\, \mathrm{d}\mathbf{y}\right\} \\ \text{subject to} \quad & \int_{\mathbb{R}^M} \phi(\mathbf{y})\, p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y} \quad \leq \alpha \\ & \mathbb{E}\left\{\|\mathbf{S}\|_2^2\right\} \leq \mathrm{A} \end{aligned} \tag{4.7}$$

where $\alpha \in (0,\, 1)$, and similarly to (4.6) the expectations are taken over the PDF of $\mathbf{S}$.

Although the optimization problem in (4.6) provides a generic formulation that is valid for any noise PDF, it is difficult to solve in general as the optimization needs to be performed over a space of signal PDFs and decision rules. In the following analysis, it is proven that optimizing over a set of variables (instead of functions) is sufficient to obtain the optimal signal PDF and the decision rule. To that aim, the following lemma is presented first.

Lemma 4.1.1. *Given a decision rule ϕ that satisfies the false alarm constraint; if $h(\phi; \mathbf{s})$ in (4.3) is a continuous function of $\mathbf{s}$ defined on a compact subset of $\mathbb{R}^M$, then an optimal solution to (4.6) can be expressed in the form of*

$$p_{\mathbf{S}}^{\mathrm{opt}}(\mathbf{x}) = \lambda\, \delta(\mathbf{x} - \mathbf{s}_1) + (1 - \lambda)\, \delta(\mathbf{x} - \mathbf{s}_2)\,, \tag{4.8}$$

where $\lambda \in [0, 1]$.

Proof. Suppose that a decision rule $\tilde{\phi}$ is given such that the constraint on the probability of false alarm is satisfied, i.e., $\int_{\mathbb{R}^M} \tilde{\phi}(\mathbf{y})\, p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y} \le \alpha$. Then, $h(\tilde{\phi}; \mathbf{s}) = \int_{\mathbb{R}^M} \tilde{\phi}(\mathbf{y})\, p_{\mathbf{N}}\, (\mathbf{y} - \mathbf{s})\ \mathrm{d}\mathbf{y}$ in (4.3) becomes a function of $\mathbf{s}$ only. Formally, $\mathrm{P_D}$ in (4.3) can be represented as $\mathrm{P_D} = \mathbb{E}\left\{h(\mathbf{S})\right\}$ and the optimization problem in (4.6) can be stated as

$$\max_{p_{\mathbf{S}}}\ \mathbb{E}\left\{h(\mathbf{S})\right\} \quad \text{subject to} \quad \mathbb{E}\left\{\|\mathbf{S}\|_2^2\right\} \le \mathrm{A}\,. \tag{4.9}$$

Similar optimization problems have been studied extensively in the literature under various frameworks [9, 15, 17, 25]. Given the conditions in the lemma, Carathéodory's theorem from convex analysis [19] implies that the optimal solution of (4.9) can be expressed by a randomization of at most two signal vectors. Therefore, for any decision rule $\tilde{\phi}$ satisfying the false alarm constraint, the optimal signal PDF can be represented as in (4.8). □

At this point, it should be emphasized that the above lemma points out to a significant reduction on the complexity of the optimization problem under certain conditions. Namely, the optimal signal design no longer involves a search over all possible signal PDFs; but instead a randomization between at most two different signal vectors suffices. Hence, the problem in (4.6) can be solved over the signal PDFs that are in the form of (4.8). Led by this observation, a further simplification of the optimization problem is presented.

Proposition 4.1.1. *Under the conditions in Lemma 4.1.1, the optimization problem in* (4.6) *can be expressed as follows:*

$$\begin{aligned}
&\max_{\{\lambda,\mathbf{s}_1,\mathbf{s}_2,\eta\}} \int_{\Gamma(\lambda,\mathbf{s}_1,\mathbf{s}_2,\eta)} \{\lambda\, p_{\mathbf{N}}(\mathbf{y}-\mathbf{s}_1) + (1-\lambda)\, p_{\mathbf{N}}(\mathbf{y}-\mathbf{s}_2)\}\, \mathrm{d}\mathbf{y} \\
&\textit{subject to} \int_{\Gamma(\lambda,\mathbf{s}_1,\mathbf{s}_2,\eta)} p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y} \;= \alpha \\
&\qquad \lambda\, \|\mathbf{s}_1\|_2^2 + (1-\lambda)\, \|\mathbf{s}_2\|_2^2 \le \mathrm{A} \\
&\qquad \lambda \in [0,1] \quad \textit{and} \quad \eta \ge 0
\end{aligned} \tag{4.10}$$

where $\Gamma(\lambda,\mathbf{s}_1,\mathbf{s}_2,\eta) = \{\mathbf{y}\in\mathbb{R}^M : \lambda\, p_{\mathbf{N}}(\mathbf{y}-\mathbf{s}_1) + (1-\lambda)\, p_{\mathbf{N}}(\mathbf{y}-\mathbf{s}_2) > \eta\, p_{\mathbf{N}}(\mathbf{y})\}$, *and* $\alpha \in (0,\, 1)$ *.*

Proof. It is known that the NP detector gives the most powerful α-level test of H_0 versus H_1 [20]. In other words, when the aim is to maximize the probability of detection such that the probability of false alarm does not exceed a predetermined value α, the NP detector is the optimal choice. When deciding between two simple hypotheses H_0 versus H_1 based on observation $\mathbf{y}$, the NP decision rule takes the following form of an LRT:

$$\tilde{\phi}_{NP}(\mathbf{y}) = \begin{cases} 1\,, & \text{if } \; p_1(\mathbf{y}) > \eta\, p_0(\mathbf{y}) \\ \gamma(\mathbf{y})\,, & \text{if } \; p_1(\mathbf{y}) = \eta\, p_0(\mathbf{y}) \\ 0\,, & \text{if } \; p_1(\mathbf{y}) < \eta\, p_0(\mathbf{y}) \end{cases} \tag{4.11}$$

where $\eta \ge 0$ and $0 \le \gamma(\mathbf{y}) \le 1$ are chosen such that the probability of false alarm satisfies $\mathrm{P_{FA}} = \mathbb{E}_0\{\tilde{\phi}_{NP}(\mathbf{Y})\} = \alpha$, where the expectation is taken with respect to the null hypothesis H_0. Then, the NP decision rule is the optimal one among all α-level decision rules, i.e. $\mathrm{P_D} = \mathbb{E}_1\{\tilde{\phi}_{NP}(\mathbf{Y})\}$ is maximized, where the expectation is taken with respect to the alternative hypothesis H_1. It can be proven that such a rule always exists for all $\alpha \in (0,\, 1)$ and is unique [20]. Let $L(\mathbf{Y}) = p_1(\mathbf{Y})/p_0(\mathbf{Y})$ be the likelihood function. For continuous $L(\mathbf{Y})$, $\gamma(\mathbf{Y})$ can be chosen arbitrarily since the probability of the event $\{p_1(\mathbf{Y}) = \eta\, p_0(\mathbf{Y})\}$ is equal to 0 under both H_0 and H_1 [20, 28].

To keep the formulation simpler, the PDF of the channel noise is assumed to be continuous which gives rise to a continuous likelihood function. An extension for the discrete case is straightforward by incorporating the parameter $\gamma(\mathbf{y})$ into the calculations for the detection and false alarm probabilities. Under the continuity assumption, while deciding between two simple hypotheses based on observation $\mathbf{y}$, the NP decision rule, which selects hypothesis H_1 if $p_1(\mathbf{y}) > \eta\, p_0(\mathbf{y})$ and selects hypothesis H_0 otherwise, maximizes the probability of detection under the false alarm constraint. Therefore, when the signal PDF $p_{\mathbf{S}}(\cdot)$ is specified, it is sufficient to consider only the detection probability of the NP rule instead of a search over all the decision rules.

As the NP decision rule assigns observation $\mathbf{y}$ to hypothesis H_1 if $p_1(\mathbf{y}) > \eta\, p_0(\mathbf{y})$ and decides hypothesis H_0 otherwise, the probability of detection and false alarm expressions in (4.2) can be expressed for an NP decision rule as

$$\begin{aligned} \mathrm{P_D} &= \mathbb{E}_1\left[\tilde{\phi}_{NP}(\mathbf{Y})\right] = \int_{\Gamma} p_1(\mathbf{y})\,\mathrm{d}\mathbf{y} \\ \mathrm{P_{FA}} &= \mathbb{E}_0\left[\tilde{\phi}_{NP}(\mathbf{Y})\right] = \int_{\Gamma} p_0(\mathbf{y})\,\mathrm{d}\mathbf{y} \end{aligned} \tag{4.12}$$

where $\Gamma = \left\{\mathbf{y} \in \mathbb{R}^M : p_1(\mathbf{y}) > \eta\, p_0(\mathbf{y})\right\}$.

In Lemma 4.1.1, it is shown that an optimal signal PDF is in the form of (4.8). As a result, the conditional PDF of the observation under hypothesis H_1 can be written as $p_1(\mathbf{y}) = \mathbb{E}\{p_{\mathbf{N}}\,(\mathbf{y}-\mathbf{S})\} = \int_{\mathbb{R}^M} p_{\mathbf{S}}(\mathbf{x})\, p_{\mathbf{N}}(\mathbf{y}-\mathbf{x})\,\mathrm{d}\mathbf{x} = \lambda\, p_{\mathbf{N}}\,(\mathbf{y}-\mathbf{s}_1) + (1-\lambda)\, p_{\mathbf{N}}\,(\mathbf{y}-\mathbf{s}_2)$. Similarly, the average power constraint in (4.6) becomes $\lambda\, \|\mathbf{s}_1\|_2^2 + (1-\lambda)\, \|\mathbf{s}_2\|_2^2 \leq \mathrm{A}$. Therefore, the expressions for $\mathrm{P_D}$ and $\mathrm{P_{FA}}$ at the end of the previous paragraph imply that the optimization problems in (4.6) and (4.10) are equivalent as stated in the proposition. □

Proposition 4.1.1 implies that the solution of the original optimization problem in (4.6), which considers the joint optimization of the stochastic signal PDF and the detector, can be obtained as the solution of the much simpler optimization problem specified in (4.10).

Comparing the formulations in (4.6) and (4.10), it is noted that a significant complexity reduction is obtained in the representation of the problem by optimizing over a set of variables instead of a set of functions. The solution of the optimization problem in (4.10) can be obtained via global optimization techniques (since it is not a convex problem in general), or a convex relaxation approach as in [16] can be employed to obtain approximate solutions in polynomial time. In this study, the multistart and patternsearch methods from MATLAB's Global Optimization Toolbox are used to obtain the solution of (4.10).

Assuming that the selected optimization algorithm successfully returns the parameters $\left\{\lambda^{\mathrm{opt}}, \mathbf{s}_1^{\mathrm{opt}}, \mathbf{s}_2^{\mathrm{opt}}, \eta^{\mathrm{opt}}\right\}$ for the problem in (4.10), the optimal signal PDF can be constructed as $p_{\mathbf{S}}^{\mathrm{opt}}(\mathbf{x}) = \lambda^{\mathrm{opt}}\delta(\mathbf{x}-\mathbf{s}_1^{\mathrm{opt}})+(1-\lambda^{\mathrm{opt}})\delta(\mathbf{x}-\mathbf{s}_2^{\mathrm{opt}})$, and the optimal decision rule assumes the form of the corresponding NP decision rule that decides hypothesis H_1 if $\lambda^{\mathrm{opt}}p_{\mathbf{N}}\left(\mathbf{y}-\mathbf{s}_1^{\mathrm{opt}}\right)+(1-\lambda^{\mathrm{opt}})p_{\mathbf{N}}\left(\mathbf{y}-\mathbf{s}_2^{\mathrm{opt}}\right) > \eta^{\mathrm{opt}}\, p_{\mathbf{N}}(\mathbf{y})$ and decides hypothesis H_0 otherwise.

4.2 Case 2: Multiple Detectors at the Receiver

We consider an average power constrained on-off keying communications system operating over an additive noise channel. The receiver can randomize among at most K different detectors (decision rules) in any manner to improve the average detection performance, as shown in Figure 4.1. At any given time, only a single detector is employed at the receiver to conclude the presence/absence of a signal level embedded in noise. Via a communications protocol, the transmitter is informed of the detector currently active at the receiver. As pointed out in Section 4.1, in the absence of detector randomization, employing stochastic signaling; that is, modeling the on-signal as a random variable instead of assuming a constant level, can help improve the detection performance without violating the constraints on the false alarm probability and average signal power.

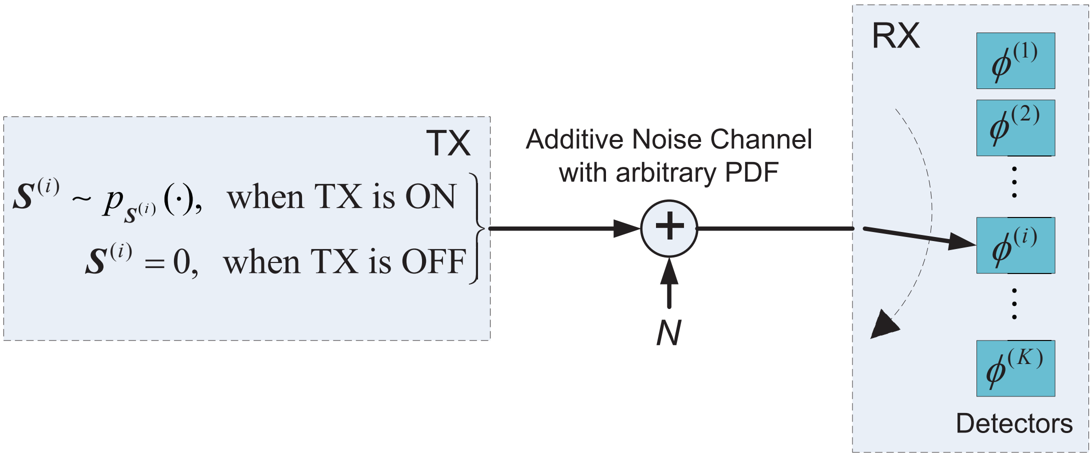

Figure 4.1: On-off keying communications system model for joint stochastic signaling and detector randomization.

Given an N-dimensional observation vector, the receiver has to decide between two hypotheses H_0 or H_1 specified as

$$H_0 : \mathbf{Y} = \mathbf{N}\,, \quad H_1 : \mathbf{Y} = \mathbf{S}^{(i)} + \mathbf{N}\,, \; i \in \{1, \ldots, K\} \tag{4.13}$$

where $\mathbf{Y}$ is the noisy observation vector, $\mathbf{S}^{(i)}$ represents the transmitted signal vector for the on-signal destined for detector i, and $\mathbf{N}$ is the noise component that is independent of $\mathbf{S}^{(i)}$. Furthermore, $\mathbf{S}^{(i)}$ is modeled as a random vector to facilitate stochastic signaling.

Let v_i denote the randomization factor for detector i, where $\sum_{i=1}^{K} v_i = 1$ and $v_i \geq 0$ for $i = 1, \ldots, K$. The two probabilities of interest in the NP framework, the average probability of detection $\mathrm{P_D}$ and the average probability of false alarm $\mathrm{P_{FA}}$, can be calculated as $\mathrm{P_D} = \sum_{i=1}^{K} v_i \, \mathrm{P_D^{(i)}}$ and $\mathrm{P_{FA}} = \sum_{i=1}^{K} v_i \, \mathrm{P_{FA}^{(i)}}$. $\mathrm{P_D^{(i)}}$ and $\mathrm{P_{FA}^{(i)}}$ represent the detection and false alarm probabilities for detector i, respectively; and are specified by

$$\mathrm{P_D^{(i)}} = \int_{\mathbb{R}^N} \phi^{(i)}(\mathbf{y})\, p_1^{(i)}(\mathbf{y})\, \mathrm{d}\mathbf{y} \tag{4.14}$$

$$\mathrm{P_{FA}^{(i)}} = \int_{\mathbb{R}^N} \phi^{(i)}(\mathbf{y})\, p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y} \tag{4.15}$$

where $\phi^{(i)}$ is the decision rule for detector i, and $p_1^{(i)}(\mathbf{y})$ denotes the conditional PDF of the observation received by detector i under the alternative hypothesis H_1. Recalling that signal and noise are independent, $p_1^{(i)}(\mathbf{y}) = \int_{\mathbb{R}^N} p_{\mathbf{S}^{(i)}}(\mathbf{s})\, p_{\mathbf{N}}(\mathbf{y}-\mathbf{s})\, \mathrm{d}\mathbf{s} = \mathbb{E}\{p_{\mathbf{N}}\big(\mathbf{y}-\mathbf{S}^{(i)}\big)\}$, where the expectation is taken over the PDF of $\mathbf{S}^{(i)}$. Similarly, under the framework of stochastic signaling and detector randomization, the constraint on the average signal power can be expressed as [20]: $\sum_{i=1}^{K} v_i\, \mathbb{E}\{\|\mathbf{S}^{(i)}\|_2^2\} \leq \mathrm{A}$, where A denotes the average power limit.

For a given detector i and the corresponding signal PDFs, the probability of detection is maximized under the false alarm constraint using the NP decision rule [20, 23], which takes the form of an LRT

$$\phi_{\mathrm{NP}}^{(i)}(\mathbf{y}) = \begin{cases} 1\,, & \text{if } \; p_1^{(i)}(\mathbf{y}) \geq \eta^{(i)}\, p_{\mathbf{N}}(\mathbf{y}) \\ 0\,, & \text{if } \; p_1^{(i)}(\mathbf{y}) < \eta^{(i)}\, p_{\mathbf{N}}(\mathbf{y}) \end{cases}, \tag{4.16}$$

where the decision threshold $\eta^{(i)} \geq 0$ is chosen such that the probability of false alarm satisfies $\mathrm{P}_{\mathrm{FA}}^{(i)} = \int_{\mathbb{R}^N} \phi_{\mathrm{NP}}^{(i)}(\mathbf{y}) p_{\mathbf{N}}(\mathbf{y}) \mathrm{d}\mathbf{y} = \alpha_i$ for some value $\alpha_i \in (0,\, 1)$. Then, the NP rule is the optimal one among all α_i-level decision rules for detector i, i.e., $\mathrm{P}_{\mathrm{D}}^{(i)} = \int_{\mathbb{R}^N} \phi_{\mathrm{NP}}^{(i)}(\mathbf{y}) p_1^{(i)}(\mathbf{y}) \mathrm{d}\mathbf{y}$ is maximized [20, 23]. Therefore, it is not necessary to search over all decision rules; only the NP decision rule should be determined for each detector and the corresponding average detection and false alarm probabilities should be considered. Using the decision region for the NP detector, $\Gamma_{\mathrm{NP}}^{(i)}(p_{\mathbf{S}^{(i)}}, \eta^{(i)}) = \{\mathbf{y} \in \mathbb{R}^N : \mathbb{E}\{p_{\mathbf{N}}(\mathbf{y}-\mathbf{S}^{(i)})\} \geq \eta^{(i)}\, p_{\mathbf{N}}(\mathbf{y})\}$, detection and false alarm probabilities for detector i can be expressed as

$$\mathrm{P}_{\mathrm{D,NP}}^{(i)} = \int_{\Gamma_{\mathrm{NP}}^{(i)}} \mathbb{E}\big\{p_{\mathbf{N}}(\mathbf{y}-\mathbf{S}^{(i)})\big\}\, \mathrm{d}\mathbf{y} \tag{4.17}$$

$$\mathrm{P}_{\mathrm{FA,NP}}^{(i)} = \int_{\Gamma_{\mathrm{NP}}^{(i)}} p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y}\,. \tag{4.18}$$

By adapting stochastic signaling and detector randomization into the NP framework, we aim to *jointly* optimize the randomization factors, decision thresholds and signal PDFs in order to maximize the average probability of detection under the constraints on the average probability of false alarm and average signal power (Joint optimization can be facilitated via a feedback mechanism from the receiver to the transmitter, such as those in cognitive radio (CR) systems). Then, by denoting the optimization space as $\mathcal{S} \triangleq \{v_i, \eta^{(i)}, p_{\mathbf{S}^{(i)}}\}_{i=1}^{K}$, the optimal design problem can be solved from

$$\begin{aligned}
&\max_{\mathcal{S}} \sum_{i=1}^{K} v_i \int_{\Gamma_{\mathrm{NP}}^{(i)}} \mathbb{E}\big\{p_{\mathbf{N}}(\mathbf{y}-\mathbf{S}^{(i)})\big\}\,\mathrm{d}\mathbf{y} \\
&\text{subject to} \sum_{i=1}^{K} v_i \int_{\Gamma_{\mathrm{NP}}^{(i)}} p_{\mathbf{N}}(\mathbf{y})\,\mathrm{d}\mathbf{y} \le \alpha \\
&\sum_{i=1}^{K} v_i\, \mathbb{E}\left\{\left\|\mathbf{S}^{(i)}\right\|_2^2\right\} \le \mathrm{A}\,, \quad \sum_{i=1}^{K} v_i = 1\,, \quad \boldsymbol{v} \succeq \mathbf{0}
\end{aligned} \tag{4.19}$$

where $\alpha \in (0,\, 1)$ is the average false alarm constraint, $\boldsymbol{v} \succeq \mathbf{0}$ means that $v_i \ge 0\ \forall i \in \{1, 2, \ldots, K\}$, and expectations are taken over the signal PDFs $p_{\mathbf{S}^{(i)}}$. Implicit constraints are also present in (4.19) due to each $p_{\mathbf{S}^{(i)}}$ representing a PDF.

A direct evaluation of (4.19) requires an exhaustive search over the space of randomization factors, decision thresholds and signal PDFs, which is inherently a difficult procedure. Let $\mathrm{P}_{\mathrm{c}}^{\dagger}$ denote the maximum average probability of detection obtained from the solution of (4.19). In the sequel, an upper bound on this problem with a simpler solution is derived, and then the achievability of this bound is demonstrated. To this end, the following observations are stated first.

Suppose that the decision rule $\tilde{\phi}_{\mathrm{NP}}$ (i.e., threshold $\tilde{\eta}$) and the signal PDF $\tilde{p}_{\mathbf{S}}(\cdot)$ are specified for one of the detectors employed at the receiver. The corresponding detection probability can be written as $\tilde{\mathrm{P}}_{\mathrm{D}} = \int_{\mathbb{R}^N} \tilde{\phi}_{\mathrm{NP}}(\mathbf{y})\, \mathbb{E}\{p_{\mathbf{N}}(\mathbf{y}-\mathbf{S})\}\,\mathrm{d}\mathbf{y} = \mathbb{E}\{\int_{\mathbb{R}^N} \tilde{\phi}_{\mathrm{NP}}(\mathbf{y})\, p_{\mathbf{N}}(\mathbf{y}-\mathbf{S})\,\mathrm{d}\mathbf{y}\}$, where the linearity of the expectation operator is imposed over the fixed decision rule $\tilde{\phi}_{\mathrm{NP}}$. Recall that the expression inside the expectation operator is the probability of detection when the deterministic signal vector $\mathbf{s}$ is used for the transmission of on-symbol over the additive noise channel and decision rule $\tilde{\phi}_{\mathrm{NP}}$ is employed at the receiver. Although the detector $\tilde{\phi}_{\mathrm{NP}}$ is in the optimal form for the signal distribution $\mathbb{E}\{p_{\mathbf{N}}(\mathbf{y}-\mathbf{S})\}$, it can be suboptimal for each component $p_{\mathbf{N}}(\mathbf{y}-\mathbf{s})$. By applying the NP criterion to each signal component $p_{\mathbf{N}}(\mathbf{y}-\mathbf{s})$ that make up the received signal distribution for the on-symbol, the probability of detection can be increased even further without violating the false alarm constraint. More specifically,

$$\hat{\phi}_{\text{NP}}(\mathbf{y},\mathbf{s}) = \begin{cases} 1, & \text{if } p_{\mathbf{N}}(\mathbf{y}-\mathbf{s}) \geq \eta(\mathbf{s})\, p_{\mathbf{N}}(\mathbf{y}) \\ 0, & \text{if } p_{\mathbf{N}}(\mathbf{y}-\mathbf{s}) < \eta(\mathbf{s})\, p_{\mathbf{N}}(\mathbf{y}) \end{cases} \tag{4.20}$$

where $\eta(\mathbf{s}) \geq 0$ is determined as a function of $\mathbf{s}$ from the false alarm constraint via $\int_{\mathbb{R}^N} \hat{\phi}_{\text{NP}}(\mathbf{y},\mathbf{s}) p_{\mathbf{N}}(\mathbf{y}) \mathrm{d}\mathbf{y} = \int_{\mathbb{R}^N} \tilde{\phi}_{\text{NP}}(\mathbf{y})\, p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y}$. As a result, the decision rule $\tilde{\phi}_{\text{NP}}$ for the given detector can be replaced with a set of decision rules $\hat{\phi}_{\text{NP}}$ indexed by parameter $\mathbf{s}$ such that

$$\mathbb{E}\left\{\int_{\mathbb{R}^N} \hat{\phi}_{\text{NP}}(\mathbf{y},\mathbf{S}) p_{\mathbf{N}}(\mathbf{y}-\mathbf{S}) \mathrm{d}\mathbf{y}\right\} \geq \int_{\mathbb{R}^N} \tilde{\phi}_{\text{NP}}(\mathbf{y}) \mathbb{E}\left\{p_{\mathbf{N}}(\mathbf{y}-\mathbf{S})\right\} \mathrm{d}\mathbf{y} \tag{4.21}$$

is always satisfied while guaranteeing the false alarm constraint due to the increased number of optimal NP decision rules in the new formulation (in contrast with the limited number of detectors in the original problem).

In accordance with the terminology in Chapter 2 and Section 4.1 , the left side of the inequality in (4.21) can be interpreted as a randomization among NP detectors corresponding to deterministic signal vectors, while the right hand side can be understood as stochastic signaling using a single detector. Hence, assuming the same average power and false alarm constraints, the average probability of detection obtained by stochastic signaling with PDF $\tilde{p}_{\mathbf{S}}(\cdot)$ is always smaller than or equal to that of deterministic signaling and detector randomization according to the same PDF when optimal NP detectors are employed in both cases under the same statistics for the additive noise.

Notice that a new decision rule is added for each $\mathbf{s}$ in the support of $\tilde{p}_{\mathbf{S}^{(i)}}$ to obtain the upper bound for a given detector i in the previous analysis. This procedure can be extended safely across multiple detectors by assuming that the supports of $\tilde{p}_{\mathbf{S}^{(i)}}$, $i = 1, 2, \ldots, K$ are non-overlapping. If there were overlapping supports, then $\exists\, \tilde{\mathbf{s}} \in \mathbb{R}^N$ such that $\tilde{p}_{\mathbf{S}^{(i)}}(\tilde{\mathbf{s}}) \neq 0$ and $\tilde{p}_{\mathbf{S}^{(j)}}(\tilde{\mathbf{s}}) \neq 0$ for $i \neq j$. After applying the procedure described above, there would be contributions in the overall average false alarm probability as $\tilde{v}_i \tilde{p}_{\mathbf{S}^{(i)}}(\tilde{\mathbf{s}}) \int_{\mathbb{R}^N} \hat{\phi}^{(i)}_{\mathrm{NP}}(\mathbf{y}, \tilde{\mathbf{s}})\, p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y} + \tilde{v}_j \tilde{p}_{\mathbf{S}^{(j)}}(\tilde{\mathbf{s}}) \int_{\mathbb{R}^N} \hat{\phi}^{(j)}_{\mathrm{NP}}(\mathbf{y}, \tilde{\mathbf{s}}) p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y} \triangleq \alpha_{ij}$. Similarly, the contributions from these terms to the average detection probability would be $\tilde{v}_i \tilde{p}_{\mathbf{S}^{(i)}}(\tilde{\mathbf{s}}) \int_{\mathbb{R}^N} \hat{\phi}^{(i)}_{\mathrm{NP}}(\mathbf{y}, \tilde{\mathbf{s}})\, p_{\mathbf{N}}(\mathbf{y} - \tilde{\mathbf{s}})\, \mathrm{d}\mathbf{y} + \tilde{v}_j \tilde{p}_{\mathbf{S}^{(j)}}(\tilde{\mathbf{s}}) \int_{\mathbb{R}^N} \hat{\phi}^{(j)}_{\mathrm{NP}}(\mathbf{y}, \tilde{\mathbf{s}})\, p_{\mathbf{N}}(\mathbf{y} - \tilde{\mathbf{s}})\, \mathrm{d}\mathbf{y}$. Then, the contributions from detectors i and j can be replaced in the respective expressions with a single term corresponding to the NP decision rule $\hat{\phi}_{\mathrm{NP}}(\mathbf{y}, \tilde{\mathbf{s}})$ with the false alarm probability $\alpha_{ij}/(\tilde{v}_i \tilde{p}_{\mathbf{S}^{(i)}}(\tilde{\mathbf{s}}) + \tilde{v}_j \tilde{p}_{\mathbf{S}^{(j)}}(\tilde{\mathbf{s}}))$ and the corresponding weight coefficient would become $\tilde{v}_i \tilde{p}_{\mathbf{S}^{(i)}}(\tilde{\mathbf{s}}) + \tilde{v}_j \tilde{p}_{\mathbf{S}^{(j)}}(\tilde{\mathbf{s}})$. Since the receiver operating characteristics (ROC) curve corresponding to an NP decision rule is concave for any given $\tilde{\mathbf{s}}$, the resulting system would have an even higher average detection probability while possessing the same average false alarm probability and average signal power as the former case [23].

In the light of these observations and the inequality in (4.21), an upper bound on the problem in (4.19) can be obtained as

$$\begin{aligned} &\max_{p_{\mathbf{S},\eta}} \quad \mathbb{E}\{D(\mathbf{S}, \eta)\} \\ &\text{subject to} \quad \mathbb{E}\{F(\mathbf{S}, \eta)\} \leq \alpha \quad \text{and} \quad \mathbb{E}\left\{\|\mathbf{S}\|_2^2\right\} \leq \mathrm{A} \end{aligned} \tag{4.22}$$

with $D(\mathbf{S}, \eta) \triangleq \int_{\Gamma(\mathbf{S},\eta)} p_{\mathbf{N}}(\mathbf{y} - \mathbf{S})\, \mathrm{d}\mathbf{y}$, and $F(\mathbf{S}, \eta) \triangleq \int_{\Gamma(\mathbf{S},\eta)} p_{\mathbf{N}}(\mathbf{y})\, \mathrm{d}\mathbf{y}$, where $\Gamma(\mathbf{s}, \eta) = \left\{\mathbf{y} \in \mathbb{R}^N : p_{\mathbf{N}}(\mathbf{y} - \mathbf{s}) \geq \eta\, p_{\mathbf{N}}(\mathbf{y})\right\}$ and the expectations are taken with respect to the joint PDF $p_{\mathbf{S},\eta}(\mathbf{s}, \eta)$ by treating both $\mathbf{S}$ and η as random variables. Let $\mathrm{P}_{\mathrm{c}}^{\star}$ denote the maximum average probability of detection obtained as the solution to the optimization problem in (4.22). Since this is an upper bound, $\mathrm{P}_{\mathrm{c}}^{\star} \geq \mathrm{P}_{\mathrm{c}}^{\dagger}$ is always satisfied.

Assuming that $D(\mathbf{s}, \eta)$ and $F(\mathbf{s}, \eta)$ are continuous functions defined on a compact subset of $\mathbb{R}^{N+1}$, then an optimal solution to (4.22) can be expressed by a convex combination among at most three components due to Carathéodory's theorem [19]; that is, $p^{\mathrm{opt}}_{\mathbf{S},\eta}(\mathbf{s}, \eta) = \lambda_1\, \delta(\mathbf{s} - \mathbf{s}_1, \eta - \eta_1) + \lambda_2\, \delta(\mathbf{s} - \mathbf{s}_2, \eta - \eta_2) + \lambda_3\, \delta(\mathbf{s} - \mathbf{s}_3, \eta - \eta_3)$. Motivated by this observation, we state the following proposition.

Proposition 4.2.1. *The solution of the optimization problem in* (4.19) *can be obtained as follows:*

$$\begin{aligned} \max_{\{\lambda_i,\mathbf{s}_i,\eta_i\}_{i=1}^3} \quad & \sum_{i=1}^{3} \lambda_i \int_{\Gamma(\mathbf{s}_i,\,\eta_i)} p_{\mathbf{N}}(\mathbf{y}-\mathbf{s}_i)\,\mathrm{d}\mathbf{y} \\ \textit{subject to} \quad & \sum_{i=1}^{3} \lambda_i \int_{\Gamma(\mathbf{s}_i,\,\eta_i)} p_{\mathbf{N}}(\mathbf{y})\,\mathrm{d}\mathbf{y} \le \alpha \\ & \sum_{i=1}^{3} \lambda_i\, \|\mathbf{s}_i\|_2^2 \le \mathrm{A}\ , \ \sum_{i=1}^{3} \lambda_i = 1 \\ & \lambda_i \ge 0 \ \ \textit{and} \ \ \eta_i \ge 0 \ \ \forall i \in \{1,2,3\} \end{aligned} \tag{4.23}$$

where $\Gamma(\mathbf{s}_i,\,\eta_i) = \big\{\mathbf{y}\in\mathbb{R}^N : p_{\mathbf{N}}(\mathbf{y}-\mathbf{s}_i) \ge \eta_i\, p_{\mathbf{N}}(\mathbf{y})\big\}$ $\forall i \in \{1,2,3\}$, *and* $\alpha \in (0,\,1)$ *.*

Proof. The optimization problem in (4.23) is obtained by substituting the form of the optimal PDF $p^{\mathrm{opt}}_{\mathbf{S},\eta}(\mathbf{s},\eta)$ into the optimization problem in (4.22). Now, we show that the optimization problems in (4.19) and (4.23) result in the same maximum value. Since (4.22) and equivalently (4.23), provide an upper bound on (4.19), $\mathrm{P}_\mathrm{c}^\dagger \le \mathrm{P}_\mathrm{c}^\star$. Next, consider the optimization problem in (4.19) when $K = 3$ detectors are used and deterministic signaling is employed for each detector, that is, $p_{\mathbf{S}^{(i)}}(\mathbf{s}) = \delta(\mathbf{s}-\mathbf{s}_i)$, $i = 1,2,3$. In that case, (4.19) reduces to the optimization problem in (4.23). As a result, the maximum value of the objective function in (4.19) should be larger than or equal to that of (4.23); namely, $\mathrm{P}_\mathrm{c}^\dagger \ge \mathrm{P}_\mathrm{c}^\star$. Therefore, $\mathrm{P}_\mathrm{c}^\dagger = \mathrm{P}_\mathrm{c}^\star$ must be satisfied. □

A few conclusions can be drawn from Proposition 4.2.1. Firstly, when multiple detectors are available for randomization ($K \ge 3$), it is sufficient to employ detector randomization among three deterministic signal vectors; i.e., there is no need to employ stochastic signaling to achieve the optimal solution. Secondly, the solution of (4.19) can be obtained by optimizing over a significantly reduced optimization space via (4.23).

4.3 Simulation Results

In this section, a numerical example is presented to compare the detection performances of the optimal solutions obtained in Section 4.1 and Section 4.2 against various suboptimal signalling techniques. A binary hypotheses-testing problem specified as in (4.13) is considered with scalar observations. Such a scenario is well suited for binary communications systems that transmit no signal for bit 0 and a signal (or a randomization of two signal values as discussed above) for bit 1 (i.e., on-off keying). It is assumed that the receiver is capable of randomizing among multiple detectors ($K \geq 3$). The noise N in (4.13) is assumed to have a symmetric Gaussian mixture distribution with equal variances as follows: $p_N(n) = \sum_{i=1}^{L} l_i \exp\{-(n-\mu_i)^2/(2\sigma^2)\}/(\sqrt{2\pi}\,\sigma)$, where $\boldsymbol{l} = [0.1492\ 0.1088\ 0.2420\ 0.2420\ 0.1088\ 0.1492]$, and $\boldsymbol{\mu} = [-1.211\ \ -0.755\ \ -0.3\ 0.3\ 0.755\ 1.211]$. It is noted that the average power of the noise can be calculated from $\mathbb{E}\{N^2\} = \sigma^2 + \sum_{i=1}^{L} l_i\,\mu_i^2$. The average signal power and average false alarm constraints are selected as $\mathrm{A} = 1$ and $\alpha = 0.05$, respectively.

The following signaling schemes, which employ a single detector at the receiver, will be considered:

Gaussian Solution: Lacking any information about the noise PDF in the channel, the transmitter employs signaling at the maximum permitted power level, and the receiver employs the corresponding NP detector as if the noise present in the channel were Gaussian distributed. We assume that the receiver has a limited capability in the sense that it can only measure first and second order statistics of the channel noise but cannot extract higher order statistics, e.g. the form of the PDF with complete knowledge of its parameters is unavailable. Before moving forward, let us demonstrate the optimal NP detector structure for Gaussian channel noise.

Assuming arbitrary values μ_0 and σ^2 respectively for the mean and variance of the noise and $S > 0$, α-level NP test at the receiver results in the following one sample optimal detection scheme

$$\tilde{\phi}_{NP}(y) = \begin{cases} 1\,, & \text{if } \; y \geq \sigma \mathcal{Q}^{-1}(\alpha) + \mu_0 \\ 0\,, & \text{if } \; y < \sigma \mathcal{Q}^{-1}(\alpha) + \mu_0 \end{cases} \tag{4.24}$$

where $\mathcal{Q}(x) = (\int_x^\infty e^{-t^2/2}\,\mathrm{d}t)/\sqrt{2\pi}$ is the tail probability of the standard normal distribution. The corresponding probability of detection is expressed as $\mathrm{P_D}(\tilde{\phi}) = \mathcal{Q}(\mathcal{Q}^{-1}(\alpha) - (S - \mu_0)/\sigma)$. Since the peak power that can be emitted from the transmitter is limited with A, probability of detection is maximized when all the available power is utilized, that is $S = \sqrt{\mathrm{A}}$. To prevent any bias due to average noise power, a zero mean Gaussian noise with variance $\hat{\sigma}^2 = \sigma^2 + \sum_{i=1}^{L} l_i \cdot \mu_i^2$ is assumed in the following analysis for the conventional case. Here, σ^2 and μ_i's are the same as those given in the Gaussian mixture PDF, which is assumed to be the actual channel noise.

Conventional Solution: In this case, the transmitter employs deterministic signaling at the maximum permitted power level, which is known to be optimal if the noise present in the channel were Gaussian distributed. Unlike the previous case, to mitigate the effects of non-Gaussian channel noise, the receiver is assumed to know the channel statistics and allowed to design the optimal NP decision rule corresponding to the deterministic signaling at the power limit. This optimization problem can be expressed as follows:

$$\max_{\eta} \int_{\Gamma} p_N(y - \sqrt{\mathrm{A}})\,\mathrm{d}y \;\text{ s.t. } \int_{\Gamma} p_N(y)\,\mathrm{d}y = \alpha\;,\; \eta \geq 0 \tag{4.25}$$

where $\Gamma = \{y \in \mathbb{R} : p_N(y - \sqrt{\mathrm{A}}) > \eta\, p_N(y)\}$ and $\alpha \in (0,\, 1)$.

Optimal−Stochastic: This approach refers to the joint design of the signaling structure and the decision rule formulated in (4.6), which can also be obtained from (4.10) as studied in the previous section.

Optimal−Deterministic: A simplified version of the optimal solution in (4.10) can be obtained by assuming that the transmitted signal is deterministic; i.e., it is not a randomization of two distinct signal levels. The optimization problem in (4.10) reduces to

$$\begin{aligned} &\max_{\{s,\,\eta\}} \int_{\Gamma} p_N(y-s)\,\mathrm{d}y \\ &\text{subject to} \int_{\Gamma} p_N(y)\,\mathrm{d}y \;=\alpha\,,\;\; |s|^2 \leq \mathrm{A}\,,\;\;\eta \geq 0 \end{aligned} \tag{4.26}$$

where $\Gamma = \{y \in \mathbb{R} : p_N(y-s) > \eta\, p_N(y)\}$ and $\alpha \in (0,\, 1)$.

Finally, the following scheme is considered as the overall optimal solution when detector randomization is allowed at the receiver as well:

Optimal Detector Randomization with Deterministic Signaling: This case refers to the solution of the most generic optimization problem in (4.19), which can be obtained from (4.23) as studied in the previous section.

In obtaining the optimal solutions for the global optimization problems stated above, MATLAB's *multistart* method is employed with 500 random start points and *sqp* algorithm is used together with the local solver *fmincon*. The extrema returned by the method are cross-checked with the results from the *patternsearch* method. This procedure is repeated for all values of σ in the set $\{0.01 : 0.005 : 0.30\}$.

In Figure 4.2, the detection probabilities of the schemes described above are plotted versus σ for $\mathrm{A} = 1$ and $\alpha = 0.05$. From the figure, it is observed that the Gaussian solution has the worst performance as expected since neither the signaling scheme nor the detector is optimized according to the channel noise PDF. Respectively, conventional solution presents a poor performance as well since no optimization is performed for the signaling scheme employed at the transmitter even though the detector is optimized by taking into account the actual noise PDF. As mentioned above, signaling at the maximum permitted power level is not necessarily optimal for non-Gaussian cases. Having a multimodal PDF,

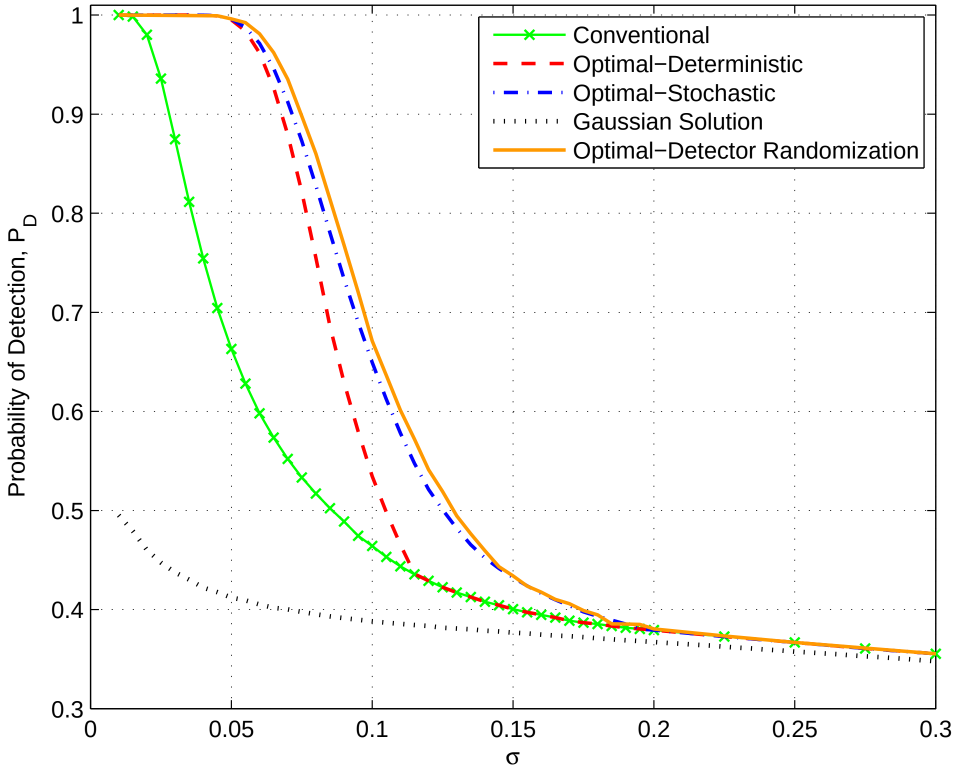

Figure 4.2: Probability of detection $\mathrm{P_D}$ as a function of σ for different approaches when $\mathrm{A} = 1$ and $\alpha = 0.05$. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.3\ 0.755\ 1.211]$ with respective weights $[0.2420\ 0.1088\ 0.1492]$ is considered.

channel noise degrades the performance of the communications system when the on-signal (symbol 1) is transmitted at the power limit. Optimizing deterministic signal levels improves over the performance of the conventional solution for $0.01 \leq \sigma \leq 0.115$ as observed from the Optimal–Deterministic curve by avoiding the overlaps among the components of the Gaussian mixture noise more effectively. Further performance improvements are obtained over a larger interval $0.01 \leq \sigma \leq 0.20$ when optimal stochastic signals are considered instead of conventional signaling (see Optimal–Stochastic). The superior performance of optimal stochastic signaling over optimal deterministic signaling is also evident from the values assumed by the probability of detection curves for $0.04 \leq \sigma \leq 0.20$.

In contrast to finding the single signal value that best avoids the overlaps among mixture components, stochastic signaling scheme allots the available power in such a way that a large portion of the power is allocated to the signal component that results in less overlap between the original and the shifted noise PDF on average. Optimal–Stochastic strategy performs a randomization between two signal values for symbol 1, and employs the corresponding α-level NP decision rule at the receiver. For example, at $\sigma = 0.1$, the optimal stochastic signal is a randomization of $s_1 = 0.2732$ and $s_2 = 1.2460$ with $\lambda = 0.3739$, achieving a detection probability of 0.6494. On the other hand, the optimal deterministic solution sets $s = 0.7684$, resulting in a detection probability of 0.5345.

However, the highest detection performance is achieved by the solution of the most generic joint optimization problem given in (4.23), which performs randomization among NP detectors corresponding to three or fewer deterministic signal values for the on-symbol (see Optimal–Detector Randomization). For example, at $\sigma = 0.1$, a detection probability of 0.671 can be achieved by transmitting $s_1 = 1.211$ with probability $\lambda_1 = 0.665$ and $s_2 = 0.265$ with probability $\lambda_2 = 0.335$, and employing the corresponding NP detectors with false alarm probabilities $\alpha_1 = 0.0368$ and $\alpha_2 = 0.0763$ (see Table 4.2 for more results). It is seen in Table 4.2 that randomization between two NP decision rules achieves the highest detection performance for most values of σ. Since Proposition 4.2.1 states that *at most* three detectors are sufficient to obtain the optimal solution via randomization, one can find examples where optimal performance can be achieved using fewer detectors as in this case. On the contrary, there may be cases where randomization among three detectors becomes a necessity for optimality (e.g., some multivariate noise PDFs, $N > 1$).

As σ is increased beyond 0.20, it is observed that both optimal signaling schemes converge to conventional signaling which in turn converges to Gaussian solution for increasing values of σ. This is mainly due to the fact that the overlap among mixture components of the noise PDF becomes significant for large values of σ, and there is not enough freedom left for the randomization to become effective over transmitting at the power limit. It is also concluded from the results of the previous section that the performance figure achieved via detector randomization is the global optimum; that is, it cannot be beaten by the combination of any different signaling schemes with a single detector as long as the problem formulation stays the same. In order to explain the results depicted in Figure 4.2, Table 4.1 presents the solutions of the optimization problems in (4.25), (4.26), and (4.10) for the Conventional, Optimal–Deterministic and Optimal–Stochastic approaches, respectively. Additionally, Table 4.2 presents the solution of the optimization problem in (4.23) for the Optimal–Detector Randomization scheme. The optimal solution for detector randomization employs signal S_i and the corresponding NP detector characterized with the threshold parameter η_i with probability λ_i for $i = 1, 2, 3$.

4.4 Concluding Remarks

In this chapter, power constrained on-off keying communications systems are investigated in the presence of stochastic signaling and detector randomization under the Neyman-Pearson framework. First, the case with a single detector at the receiver is investigated [34]. The problem of jointly designing the signaling scheme and the decision rule is addressed in order to maximize the probability of detection without violating the constraints on the probability of false alarm and the average transmit power. Based on a theoretical analysis, it is shown that the optimal solution can be obtained by employing randomization between at most two signal values for the on-signal (symbol 1) and using the corresponding NP-type likelihood ratio test at the receiver. As a result, the optimal parameters can be computed over a significantly reduced optimization space instead of an infinite set of functions using global optimization techniques. Next, the case with multiple detectors at the receiver is analyzed [35]. The joint optimal design of decision rules, stochastic signals, and detector randomization factors is performed. It is shown that the solution to the most generic optimization problem that employs both stochastic signaling and detector randomization can be obtained as the randomization among no more than three NP decision rules corresponding to three deterministic signal vectors. Finally, a detection example is provided to illustrate how stochastic signaling and detector randomization can help improve detection performance over various optimal and suboptimal signaling schemes.

Table 4.1: Conventional, Optimal-Deterministic and Optimal-Stochastic signaling parameters for the scenario in Figure 4.2 . A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.3\ 0.755\ 1.211]$ with respective weights [0.2420 0.1088 0.1492] is considered.

	Conventional	Deterministic		Stochastic			
σ	η	S	η	λ	S_1	S_2	η
0.0100	0.0001	0.2905	2.3945	1	0.2939	N/A	3.9357
0.0150	0.0636	0.2253	0	0.0819	2.7722	0.2244	0.0001
0.0200	0.5013	0.2295	0	0	N/A	0.2311	0
0.0250	1.2181	0.2307	0	0.1123	2.9136	0.2285	0
0.0300	1.8979	0.2333	0	0.6002	1.2771	0.2295	0
0.0350	2.4083	0.2375	0.0001	0.3161	1.7452	0.2332	0.0001
0.0400	2.6936	0.2390	0.0033	0	N/A	0.2407	0.0035
0.0450	2.9353	0.2395	0.0317	0.5999	1.2766	0.2359	0.0239
0.0500	3.0207	0.2404	0.1336	0.3133	1.7505	0.2408	0.1121
0.0550	3.0596	0.2425	0.3816	0.3937	0.2390	1.2697	0.2850
0.0600	3.1017	0.2440	0.8064	0.6071	1.2688	0.2424	0.5813
0.0650	3.0847	0.2437	1.3624	0.6081	1.2672	0.2451	0.9688
0.0700	3.0653	0.2499	2.0143	0.6029	1.2662	0.2469	1.4271
0.0750	3.0335	0.2512	2.7082	0.3863	0.2505	1.2607	1.8904
0.0800	3.0124	0.2566	3.3672	0.3837	0.2559	1.2577	2.3098
0.0850	2.9840	0.2540	3.9014	0.6190	1.2548	0.2575	2.6797
0.0900	2.9514	0.7536	3.6711	0.6203	1.2524	0.2675	2.7853
0.0950	2.9290	0.7527	3.7640	0.3780	0.2718	1.2501	3.0160
0.1000	2.8703	0.7684	3.6386	0.3739	0.2732	1.2460	3.0092
0.1050	2.8200	0.7826	3.4175	0.6307	1.2411	0.2780	3.1266
0.1100	2.7471	0.7831	3.3962	0.3704	0.2908	1.2404	3.0855
0.1150	2.6453	0.9998	2.6473	0.3642	0.2933	1.2343	2.9496
0.1200	2.5919	1	2.5922	0.4622	0.6651	1.2162	2.8010
0.1250	2.5379	1	2.5404	0.5177	1.2269	0.6765	2.7790
0.1300	2.4731	1	2.4715	0.4772	0.6704	1.2258	2.7375
0.1350	2.4119	1	2.4117	0.5255	1.2269	0.6637	2.6332
0.1400	2.3517	1	2.3506	0.5309	1.2277	0.6526	2.4677
0.1450	2.2923	1	2.2925	0.5385	1.2253	0.6441	2.3372
0.1500	2.2350	1	2.2358	0.4552	0.6395	1.2222	2.3147
0.1550	2.1818	1	2.1817	0.5640	1.2136	0.6231	2.2941
0.1600	2.1298	1	2.1290	0.4238	0.6165	1.2066	2.2580
0.1650	2.0807	1	2.0807	0.5747	1.2068	0.6190	2.2029
0.1700	2.0350	1	2.0348	0.5857	1.2006	0.6130	2.1547
0.1750	1.9888	1	1.9885	0.3896	0.6008	1.1866	2.1124
0.1800	1.9439	1	1.9439	0.6222	1.1803	0.5940	2.0610
0.1850	1.9008	1	1.9007	0.3516	0.5844	1.1649	2.0133
0.1900	1.8588	1	1.8588	0.3270	0.5703	1.1523	1.9692
0.1950	1.8182	1	1.8182	0	N/A	1	1.8182
0.2000	1.8243	1	1.8222	0.3330	1.0009	0.9996	1.8222
0.2250	1.8735	1	1.8735	0	N/A	1	1.8764
0.2500	1.9319	1	1.9308	0	N/A	1	1.9314
0.2750	1.9999	1	2.0028	0	N/A	1	2.0028
0.3000	2.0707	1	2.0704	1	1	N/A	2.0715
0.3250	2.1397	1	2.1406	0	N/A	1	2.1401
0.3500	2.2160	1	2.2155	1	1	N/A	2.2150
0.3750	2.2811	1	2.2817	1	1	N/A	2.2815
0.4000	2.3491	1	2.3477	1	1	N/A	2.3477
0.4250	2.4161	1	2.4148	1	1	N/A	2.4148
0.4500	2.4744	1	2.4725	1	1	N/A	2.4726
0.4750	2.5256	1	2.5259	1	1	N/A	2.5271
0.5000	2.5693	1	2.5686	1	1	N/A	2.5688

Table 4.2: Optimal-Detector Randomization parameters for the scenario in Figure 4.2. A symmetric Gaussian mixture noise, which has its mass points at $\pm[0.3\ 0.755\ 1.211]$ with respective weights $[0.2420\ 0.1088\ 0.1492]$ is considered.

	Detector I			Detector II			Detector III		
σ	λ_1	S_1	η_1	λ_2	S_2	η_2	λ_3	S_3	η_3
0.010	0.609	1.216	2.556	0.389	0.201	3.433	0.002	0.169	1.931
0.0450	0	N/A	N/A	0.8743	0.2363	0.0231	0.1257	2.7492	0.0015
0.050	0.609	1.267	0.021	0.391	0.237	0.328	0	N/A	N/A
0.0550	0	N/A	N/A	0.6098	1.2661	0.1523	0.3902	0.2400	0.2527
0.0600	0.3945	0.2453	0.4878	0.6055	1.2693	0.3902	0	1.9733	2.3581
0.0650	0.6126	1.2625	0.6475	0	N/A	N/A	0.3874	0.2467	0.9437
0.0700	0.3866	0.2489	1.3807	0	N/A	N/A	0.6134	1.2615	1.0801
0.075	0.617	1.258	2.511	0.383	0.251	1.274	0	N/A	N/A
0.0800	0	N/A	N/A	0.3840	0.2545	1.9900	0.6160	1.2582	2.4770
0.0850	0.2149	0.2593	3.4355	0.1652	0.2571	2.3744	0.6199	1.2539	2.4075
0.0900	0.6145	1.2586	2.9294	0	N/A	N/A	0.3855	0.2628	3.1239
0.0950	0.1892	1.2502	3.8253	0.4352	1.2479	3.1038	0.3756	0.2666	3.2774
0.100	0.665	1.211	3.380	0.335	0.265	3.014	0	N/A	N/A
0.1050	0.0009	1.9957	0.5287	0.3583	0.2739	3.2265	0.6408	1.2300	3.1442
0.1100	0.6187	1.2392	3.7124	0.0129	1.2352	1.8192	0.3684	0.2869	2.9296
0.1150	0.6370	1.2340	3.0692	0	N/A	N/A	0.3630	0.2879	3.3199
0.1200	0.3474	0.3005	2.7437	0.6392	1.2300	3.6973	0.0134	0.3396	3.4611
0.125	0.639	1.228	2.972	0.218	0.319	3.254	0.143	0.315	2.879
0.1300	0.6404	1.2189	2.7439	0.3530	0.3302	3.2656	0.0066	1.2319	0.9840
0.1350	0.3587	0.3253	2.6340	0.5848	1.2245	2.9967	0.0565	1.2277	3.4942
0.1400	0.3618	0.3557	2.8739	0	N/A	N/A	0.6382	1.2228	2.4916
0.1450	0.3550	0.3618	2.7743	0.1275	1.2257	3.1870	0.5175	1.2134	2.2397
0.150	0.551	1.212	2.362	0.449	0.651	2.226	0	N/A	N/A
0.1550	0.6523	1.2060	2.4224	0.3423	0.3591	2.2676	0.0054	1.1399	1.0528
0.1600	0.4208	0.6664	2.0964	0.0851	1.1833	2.5406	0.4941	1.1851	2.2690
0.1650	0.5520	1.1807	2.3030	0.0584	1.1902	3.4638	0.3896	0.6158	1.9163
0.1700	0.0001	1.2001	0.1978	0.4260	0.6376	2.1375	0.5739	1.2002	2.1428
0.175	0.686	1.153	1.993	0.314	0.530	2.552	0	N/A	N/A
0.1800	0	N/A	N/A	0.7098	1.1363	1.9788	0.2902	0.5362	2.6315
0.1850	0.4966	1.1902	1.7432	0.5034	0.7675	2.5772	0	N/A	N/A
0.1900	0.5563	1.1769	1.8276	0	N/A	N/A	0.4437	0.7190	2.1363
0.1950	0.3430	0.6293	2.2603	0.6570	1.1469	1.8653	0	N/A	N/A
0.200	0.724	1.101	1.863	0.247	0.594	2.644	0.029	1.118	1.194
0.2250	0.9013	1.0338	1.7973	0	N/A	N/A	0.0987	0.6104	1.7969
0.250	0.979	1.007	1.917	0.021	0.636	3.482	0	N/A	N/A
0.300	0.751	1.005	2.033	0.249	0.984	2.190	0	N/A	N/A
0.3500	0.3509	1.0194	2.0869	0.6491	0.9894	2.2912	0	N/A	N/A
0.400	0.9994	1	2.3509	0.0006	1.0541	0.8901	0	N/A	N/A
0.4500	0.0184	0	2.9887	0.9816	1.0093	2.4560	0	N/A	N/A
0.500	0.9243	1.0401	2.5001	0.0757	0	1.5707	0	N/A	N/A

5

Average Fisher Information Maximization in the Presence of Cost Constrained Measurements

From this chapter on, we begin to investigate the trade-offs between measurement device cost and estimation accuracy. This chapter is organized as follows. In Section 5.1, we state the average Fisher information maximization problem under a constraint on the total cost of measurement devices, which are assumed to introduce additive random measurement noises to observations with arbitrary statistics. In Section 5.2, closed form solutions are obtained for the case of independent Gaussian observations and measurement noises. Section 5.3 considers the case of Gaussian observations with arbitrary covariance matrix when the cost budget is not stringent. Section 5.4 addresses the dual problem for the Gaussian case, in which the total measurement cost is minimized to attain a target minimum average Fisher information score.

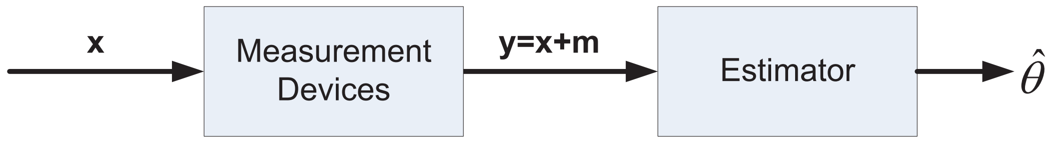

Figure 5.1: Observation vector $\mathbf{x}$ is measured by K measurement devices, and the measurements $\mathbf{x}+\mathbf{m}$ are used by an estimator to estimate the value of an unknown parameter θ.

5.1 Problem Statement and Optimal Solution

Consider a scenario as in Figure 5.1 in which noisy measurements of an observation vector $\mathbf{x}$ are acquired by K measurement devices, and then the measured values in vector $\mathbf{y}$ are processed to estimate the value of parameter θ. The measurement devices are modeled to introduce additive random measurement noises denoted by $\mathbf{m}$. In other words, the PDF of $\mathbf{x}$ is indexed by parameter θ, and the aim is to estimate that parameter based on the outputs of measurement devices. Various motivations for a similar system model can be found in [20, 23, 53]. It should be emphasized that the model in Figure 5.1 presents a generic estimation framework in which measurements are processed by an estimator in order to determine the value of an unknown parameter. For example, in a wireless sensor network application, measurement devices correspond to sensors, which are used to estimate a parameter in the system, such as the temperature.

Given a fixed budget, it is not possible to employ measurement devices that operate with arbitrarily high precision. Due to the ubiquitous trade-off between device cost and measurement accuracy in practical scenarios, we assume that there is a constraint on the total cost of measurement devices, and express it using the cost function proposed in [46]. Specifically, for a given overall cost budget C, this constraint translates into

$$\sum_{i=1}^{K} \frac{1}{2} \log\left(1 + \frac{\sigma_{x_i}^2}{\sigma_{m_i}^2}\right) \leq \mathrm{C}\,, \tag{5.1}$$

where $\sigma_{x_i}^2$ denotes the variance of the i^{th} component of observation vector $\mathbf{x}$ (i.e., variance of the input to the i^{th} measurement device), and $\sigma_{m_i}^2$ is the variance of the

i^{th} component of $\mathbf{m}$ (i.e., variance of the noise introduced by the i^{th} measurement device). A careful inspection of (5.1) reveals that a measurement device has a higher cost if it can perform measurements with a lower measurement variance (i.e., with higher accuracy).

In order to maximize the estimation accuracy, we consider the maximization of average Fisher information, or equivalently the minimization of the BCRB at the output of the measurement devices. The main motivation for the suggested approach is that an optimal cost assignment strategy can be obtained by solving the corresponding optimization problem without assuming any specific estimator structure. In addition, it is known that certain estimators, such as the MAP estimator, can (asymptotically) achieve the BCRB; hence, the minimization of the BCRB corresponds to the (approximate) minimization of the MSE for some estimators. Specifically, a necessary and sufficient condition for the MAP estimator to be efficient (i.e., achieve the BCRB with equality) is that the a-posteriori probability density of the parameter θ, i.e., $p(\theta|\mathbf{y})$ must be Gaussian for all $\mathbf{y}$. In the case of nonrandom parameter estimation, similar conditions exist for the efficiency (i.e., achieve the CRB with equality) of the ML estimator within regularity conditions [20, Page 173], [23, Page 67]. In the special case of nonrandom estimation for one-parameter exponential family of PDFs, CRB is achieved if and only if $\hat{\theta}(\mathbf{y}) = T(\mathbf{y})$ where $T(\mathbf{y})$ denotes the corresponding complete sufficient statistics for θ.

For an arbitrary estimator $\hat{\theta}$, the BCRB on the MSE is expressed as [23]

$$\mathrm{MSE}\left\{\hat{\theta}\right\} = \mathbb{E}\left\{\left(\hat{\theta}(\mathbf{Y}) - \theta\right)^2\right\} \geq (\mathrm{J_D} + \mathrm{J_P})^{-1} , \tag{5.2}$$

where $\mathrm{J_D}$ and $\mathrm{J_P}$ denote the information obtained from the observations and the prior knowledge, respectively, which are stated as

$$\mathrm{J_D} = \mathbb{E}_{\mathbf{Y},\theta}\left\{\left(\frac{\partial \log p^{\theta}_{\mathbf{Y}}(\mathbf{Y})}{\partial\theta}\right)^2\right\} = \mathbb{E}_{\theta}\left\{\mathrm{J_S}\right\} , \quad \mathrm{J_P} = \mathbb{E}_{\theta}\left\{\left(\frac{\partial \log w(\theta)}{\partial\theta}\right)^2\right\} \tag{5.3}$$

where $\mathrm{J_S} = \mathbb{E}_{\mathbf{Y}|\theta}\left\{\left(\frac{\partial \log p^\theta_{\mathbf{Y}}(\mathbf{Y})}{\partial\theta}\right)^2\right\}$ is the standard Fisher information with $p^\theta_{\mathbf{Y}}(\mathbf{y})$ and $w(\theta)$ representing the PDF of measurement vector $\mathbf{Y}$ and the prior PDF of parameter θ, respectively. As $\mathrm{J_P}$ depends only on the prior distribution, it is independent of the cost of the measurement devices. Therefore, the aim is to maximize $\mathrm{J_D}$, which is defined as the average Fisher information, under the cost constraint in (5.1). Then, based on (5.1) and (5.3), the optimal cost assignment problem can be formulated as

$$\begin{aligned} &\max_{\{\sigma^2_{m_i}\}_{i=1}^K} \int_{-\infty}^{\infty} w(\theta) \int_{\mathbb{R}^K} \frac{1}{p^\theta_{\mathbf{Y}}(\mathbf{y})} \left(\frac{\partial p^\theta_{\mathbf{Y}}(\mathbf{y})}{\partial\theta}\right)^2 \mathrm{d}\mathbf{y} \\ &\text{subject to} \quad \sum_{i=1}^K \frac{1}{2}\log\left(1+\frac{\sigma^2_{x_i}}{\sigma^2_{m_i}}\right) \leq \mathrm{C}\,. \end{aligned} \tag{5.4}$$

It is noted that the expectation operation for the calculation of $\mathrm{J_D}$ in (5.3) is taken over both random variables θ and $\mathbf{Y}$; resulting in the objective function in (5.4).

In order to specify this optimization problem, it is assumed that the observations are independent from the measurement noises; hence, $p^\theta_{\mathbf{Y}}(\mathbf{y})$ in (5.3) can be expressed more explicitly as a convolution between the PDFs of $\mathbf{x}$ and $\mathbf{m}$; that is, $p^\theta_{\mathbf{Y}}(\mathbf{y}) = \int_{\mathbb{R}^K} p^\theta_{\mathbf{X}}(\mathbf{y}-\mathbf{m})\, p_{\mathbf{M}}(\mathbf{m})\,\mathrm{d}\mathbf{m}$. In addition, it is reasonable to assume that the noises introduced by the measurement devices are independent, in which case $p_{\mathbf{M}}(\mathbf{m})$ becomes $p_{\mathbf{M}}(\mathbf{m}) = p_{M_1}(m_1)\cdots p_{M_K}(m_K)$. As discussed in [46], it is well-justified to express the cost of a measurement device as a function of its measurement noise variance (see (5.1)). Therefore, each measurement noise component can be modeled as $m_i = \sigma_{m_i}\tilde{m}_i$, where $\tilde{m}_i$ denotes a zero-mean, unit-variance random variable with a known PDF $p_{\tilde{M}_i}$, and $\sigma^2_{m_i}$ represents the variance of the measurement device, which determines its cost as defined in (5.1). Hence, the PDF of the i^{th} measurement noise can be expressed as $p_{M_i}(m_i) = \sigma^{-1}_{m_i} p_{\tilde{M}_i}(m_i/\sigma_{m_i})$. From the preceding, $p^\theta_{\mathbf{Y}}(\mathbf{y})$ is given by

$p_{\mathbf{Y}}^{\theta}(\mathbf{y}) = \int_{\mathbb{R}^K} p_{\mathbf{X}}^{\theta}(\mathbf{y}-\mathbf{m}) \prod_{i=1}^{K} \sigma_{m_i}^{-1} p_{\tilde{M}_i}(m_i/\sigma_{m_i}) \mathrm{d}m_i$, which becomes

$$\begin{aligned} p_{\mathbf{Y}}^{\theta}(\mathbf{y}) &= \prod_{i=1}^{K} \sigma_{m_i}^{-1} \int_{-\infty}^{\infty} p_{X_i}^{\theta}(y_i - m_i) p_{\tilde{M}_i}(m_i/\sigma_{m_i}) \mathrm{d}m_i \\ &= \prod_{i=1}^{K} \int_{-\infty}^{\infty} p_{X_i}^{\theta}(y_i - \sigma_{m_i} m_i) p_{\tilde{M}_i}(m_i) \mathrm{d}m_i \,, \end{aligned} \tag{5.5}$$

in the case of independent observations. In fact, the objective function in (5.4) can be written as the sum of K components in that case (see (5.3)) as

$$\sum_{i=1}^{K} \int_{-\infty}^{\infty} w(\theta) \int_{-\infty}^{\infty} \frac{1}{p_{Y_i}^{\theta}(y)} \left(\frac{\partial p_{Y_i}^{\theta}(y)}{\partial \theta} \right)^2 \mathrm{d}y \,, \tag{5.6}$$

where $p_{Y_i}^{\theta}(y) = \int_{-\infty}^{\infty} p_{X_i}^{\theta}(y - \sigma_{m_i} m) p_{\tilde{M}_i}(m) \mathrm{d}m$. Since the optimization problem in (5.4) provides a generic formulation that is valid for any observation PDF, the problem can be non-concave in general. Hence, global optimization tools such as particle swarm optimization and differential evolution can be used to obtain the solution [63].

5.2 Special Case 1: Independent Gaussian Observations and Measurement Noises

In the case of independent Gaussian observations and measurement noises, it is possible to obtain closed-form solutions of the optimization problem stated in (5.4). To that aim, let the observation vector $\mathbf{X}$ has independent Gaussian components distributed with $\mathbf{X} \sim N\left(\theta \cdot \mathbf{1}, diag\{\sigma_{x1}^2, \ldots, \sigma_{xK}^2\}\right)$, and let the measurement noise vector $\mathbf{M}$ has zero-mean Gaussian distribution with independent components as $\mathbf{M} \sim N\left(\mathbf{0}, diag\{\sigma_{m_1}^2, \ldots, \sigma_{m_K}^2\}\right)$, where $\mathbf{1}$ and $\mathbf{0}$ denote the all-ones and all-zeros vector of length K, respectively. In that case, the average Fisher information $\mathrm{J_D}$ can be calculated as $\sum_{i=1}^{K} \left(\sigma_{m_i}^2 + \sigma_{x_i}^2\right)^{-1}$ irrespective of the prior distribution of θ, i.e., $\mathrm{J_D} = \mathrm{J_S}$. Hence, the aim reduces to the maximization of $\sum_{i=1}^{K} \left(\sigma_{m_i}^2 + \sigma_{x_i}^2\right)^{-1}$ over $\left\{\sigma_{m_1}^2, \ldots, \sigma_{m_K}^2\right\}$ under the constraint in

(5.1) and we have the following optimization problem:

$$\begin{aligned}
\max_{\{\sigma_{m_i}^2\}_{i=1}^K} \quad & \sum_{i=1}^{K} \frac{1}{\sigma_{m_i}^2 + \sigma_{x_i}^2} \\
\text{subject to} \quad & \sum_{i=1}^{K} \frac{1}{2} \log\left(1 + \frac{\sigma_{x_i}^2}{\sigma_{m_i}^2}\right) \leq \text{C}.
\end{aligned} \tag{5.7}$$

From (5.7), it is noted that both the objective function and the constraint are convex. Since the maximum of convex functions over convex sets has to occur at the boundary [19], we can take the cost constraint as equality in (5.7). This is a standard optimization problem that can be solved using Lagrange multipliers. Hence, we can write the Lagrange functional as

$$J(\sigma_{m_1}^2, \ldots, \sigma_{m_K}^2) = \sum_{i=1}^{K} \frac{1}{\sigma_{m_i}^2 + \sigma_{x_i}^2} + \lambda \left(\sum_{i=1}^{K} \frac{1}{2} \log\left(1 + \frac{\sigma_{x_i}^2}{\sigma_{m_i}^2}\right) - \text{C} \right) \tag{5.8}$$

and differentiating with respect to $\sigma_{m_i}^2$, we have

$$\sigma_{m_i}^2 = \frac{\sigma_{x_i}^4}{\gamma - \sigma_{x_i}^2}. \tag{5.9}$$

However, $\sigma_{m_i}^2$'s are positive $\forall i \in \{1, 2, \ldots, K\}$ and it may not always be possible to find a solution of this form. In this case, we use the Karush-Kuhn-Tucker (KKT) conditions to verify that the optimal cost allocation strategy can be achieved via the following assignment of the noise variances to the measurement devices:

$$\sigma_{m_i}^2 = \begin{cases} \frac{\sigma_{x_i}^4}{\gamma - \sigma_{x_i}^2} & \text{if } \sigma_{x_i}^2 < \gamma \\ \infty & \text{if } \sigma_{x_i}^2 \geq \gamma \end{cases} \quad \text{with} \quad \gamma = \left(2^{2C} \prod_{i \in S_K} \sigma_{x_i}^2 \right)^{1/|S_K|} \tag{5.10}$$

where $S_K = \big\{ i \in \{1, \ldots, K\} : \sigma_{m_i}^2 \neq \infty \big\}$ and $|S_K|$ denotes the number of elements in the set S_K.[1] We choose a constant γ and only measure those observations with variances smaller than γ. No bits are used to measure the random variables with variances greater than γ. In other words, if the variance of the

[1]Notice that the formulation in (5.10) is closely related to the "water-filling" solutions common in information theory [49, Page 276 and 314].

observed variable is greater than a threshold γ, a measurement device with infinite variance; that is, with zero cost, is considered; namely, that observation is not measured at all. On the other hand, for observations with variances smaller than γ, noise variance of the corresponding measurement device is determined from the formulation in (5.10), which assigns low measurement variances (high costs) to observations with low variances. Below we comment on some properties of the obtained solution:

- The assignment function in (5.10) is smooth and the derivatives of all orders with respect to $\sigma_{x_i}^2$ exist and are positive for $0 < \sigma_{x_i}^2 < \gamma$. Therefore, as $\sigma_{x_i}^2$ increases towards γ, $\sigma_{m_i}^2$ increases much much faster.
- For a fixed value of K (available number of observations), relaxing the cost constraint (increasing the value of C) results in higher Fisher information scores with a limiting value of $\sum_{i=1}^{K} 1/\sigma_{x_i}^2$, which is the unconstrained Fisher information of the observations about the parameter θ.

5.2.1 Alternative Strategies

Instead of the optimal cost assignment strategy specified in (5.10), one can also consider the following simple alternatives:

Strategy-1 (Equal measurement device variances): In this strategy, it is assumed that identical measurement device variances are employed for all the observations; that is, $\sigma_{m_i}^2 = \sigma_m^2,\ i = 1, \ldots, K$. Then, the cost constraint in (5.4) can be solved for equality, and σ_m^2 can simply be obtained as the smallest positive real root of the K^{th} degree polynomial described with the equation $\prod_{i=1}^{K}\left(1 + \sigma_{x_i}^2/\sigma_m^2\right) = 2^{2\mathrm{C}}$. If the observation variances are equal; that is, $\sigma_{x_i}^2 = \sigma_x^2,\ i = 1, \ldots, K$, this strategy becomes equal to the optimal solution given in (5.10) where σ_m^2 are calculated from $\sigma_m^2 = \sigma_x^2/\left(2^{2\mathrm{C}/K} - 1\right)$, and the corresponding Fisher information score is expressed as $K\left(1 - 2^{-2\mathrm{C}/K}\right)/\sigma_x^2$, which

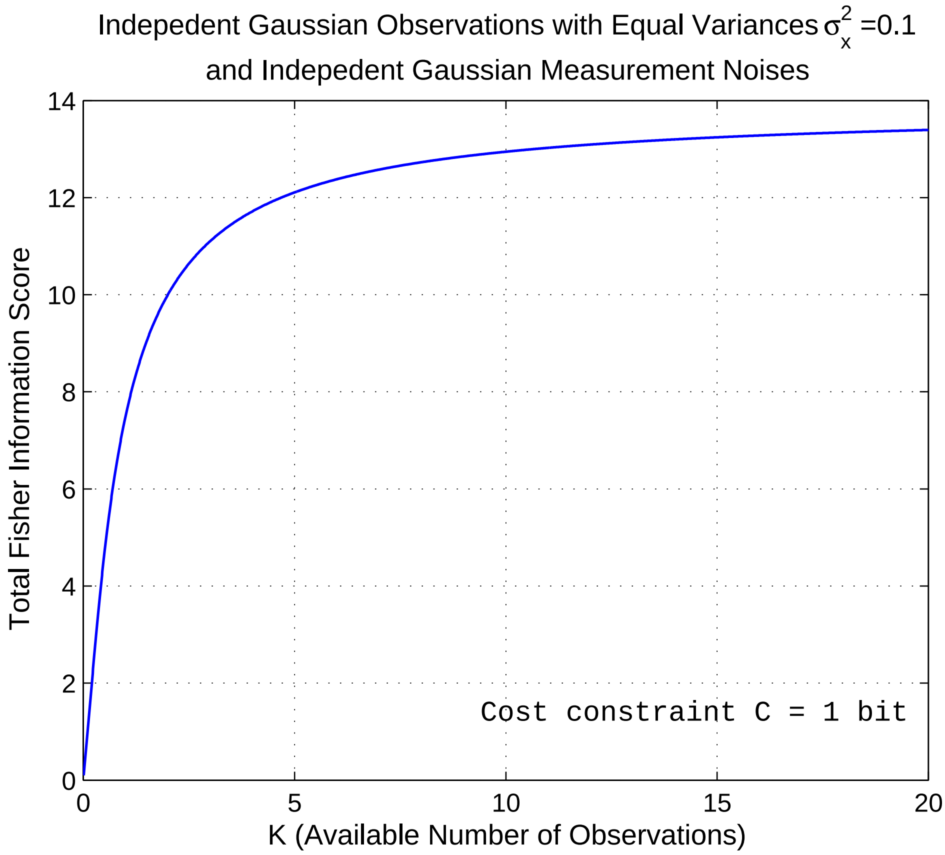

Figure 5.2: Independent Gaussian observations with variances $\sigma_x^2 = 0.1$ and independent Gaussian measurement noises are considered. Strategy-1, assigning equal measurement variances to all measurement devices is the optimal strategy. The total Fisher information score under cost constrained measurements with $\mathrm{C} = 1$ is depicted with respect to the number of available measurements. Under this scenario, Fisher information increases with each additional measurement to a limiting value of $20 \ln 2$ where as the unconstrained Fisher information goes to infinity.

is an increasing function of K for fixed C with a steep ascent (on the order of $K \sim 10\mathrm{C}$ measurements, see Figure 5.2) to its limiting value of

$$\lim_{K \to \infty} \frac{K\left(1 - 2^{-2\mathrm{C}/K}\right)}{\sigma_x^2} = \frac{2\mathrm{C} \ln 2}{\sigma_x^2}\,, \tag{5.11}$$

and similarly an increasing function of C for fixed K with a limiting value of K/σ_x^2, which denotes the unconstrained Fisher information in the case of observations with equal variances.

Strategy-2 (All cost to the best observation): In this case, the total budget C is spent on the best observation, which has the smallest variance. If the

k^{th} observation is the best one, the cost constraint in (5.4) can be used to calculate the variance of the measurement noise for that observation alone from $\sigma_{m_k}^2 = \sigma_{x_k}^2 / \left(2^{2\mathrm{C}} - 1\right)$. For all the other observations, the corresponding measurement variances are set to infinity (i.e., no measurements are taken from those observations). In this case, the corresponding Fisher information score is $\left(1 - 2^{-2\mathrm{C}}\right) / \sigma_{x_k}^2$, which is an increasing function of C with a limiting value of $1/\sigma_{x_k}^2$.

5.2.2 Numerical Results for Special Case 1

In order to provide numerical examples of the results derived in Section 5.2, consider a scenario with 4 pairs of independent Gaussian observations and measurement noises. Let $\sigma_{x_1}^2 = 0.1$, $\sigma_{x_2}^2 = 0.5$, $\sigma_{x_3}^2 = 0.9$ and $\sigma_{x_4}^2 = 1.3$. The variances of the measurement devices are calculated using the proposed optimal strategy stated in (5.10), Strategy 1, and Strategy 2. These variances and the corresponding Fisher information values are presented for different values of C in Table 5.1 . It is observed that the optimal strategy assigns smaller variances (larger costs) to observations with smaller variances, and achieves the maximum Fisher information score as expected. For further investigations, Figure 5.3 illustrates the Fisher information versus the total budget C for different strategies. It is observed that the Fisher information in Strategy 2, which assigns all the cost to the best observation, converges to $1/\sigma_{x_1}^2$ as expected (since $\sigma_{m_1}^2$ converges to zero as C increases). On the other hand, Strategy 2 and the optimal strategy converge for small values of C since the optimal strategy involves assigning all the cost to the best observation if C is small. Regarding Strategy 1, it converges to the optimal strategy for large C, where both reach to the unconstrained Fisher information score of $\sum_{i=1}^{K} 1/\sigma_{x_i}^2$, even though significant deviations are observed for intermediate values of C. Overall, the optimal cost assignment strategy yields

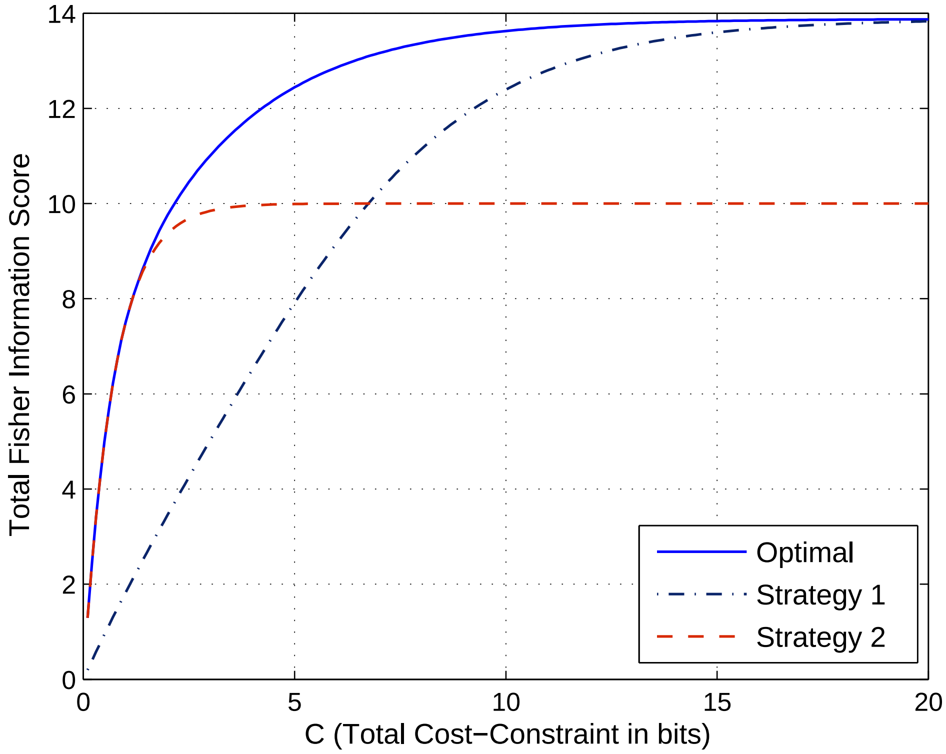

Figure 5.3: Independent Gaussian observations with variances $\sigma^2_{x_1} = 0.1$, $\sigma^2_{x_2} = 0.5$, $\sigma^2_{x_3} = 0.9$, $\sigma^2_{x_4} = 1.3$ and independent Gaussian measurement noises are considered. The performance of the optimal cost allocation strategy is depicted together with the results from Strategies 1 and 2.

the highest Fisher information in all the cases, and indicates the opportunity to achieve high estimation accuracy.

5.3 Special Case 2: Gaussian Observations with Arbitrary Covariance Matrix and Independent Gaussian Measurement Noises - High Budget Case

It is also possible to obtain a closed-form solution of the optimization problem stated in (5.4) in the case of colored Gaussian observations and independent

Gaussian measurement noises when the cost-budget is not very stringent. To that aim, let the observation vector $\mathbf{X}$ has Gaussian components distributed with $\mathbf{X} \sim N(\theta \cdot \mathbf{1}, \mathbf{\Sigma}_x)$ with $\{\sigma^2_{x_1}, \ldots, \sigma^2_{x_K}\}$ constituting the diagonal components, and let the measurement noise vector $\mathbf{M}$ has zero-mean Gaussian distribution with independent components as $\mathbf{M} \sim N(\mathbf{0}, \mathbf{D}_m)$ with $\mathbf{D}_m = diag\{\sigma^2_{m_1}, \ldots, \sigma^2_{m_K}\}$, where $\mathbf{1}$ and $\mathbf{0}$ denote the all-ones and all-zeros vector of length K, respectively. Under these assumptions, $\mathbf{Y} \sim N(\theta \cdot \mathbf{1}, \mathbf{\Sigma}_x + \mathbf{D}_m)$ and the average Fisher information $\mathrm{J_D}$ can once again be calculated irrespective of the prior distribution of θ, i.e., $\mathrm{J_D} = \mathrm{J_S} = \mathbf{1}^T (\mathbf{\Sigma}_x + \mathbf{D}_m)^{-1} \mathbf{1}$ which is the sum of all matrix elements in $(\mathbf{\Sigma}_x + \mathbf{D}_m)^{-1}$. Let $EleSum\{\cdot\}$ denote the operator whose output is the sum of all the elements of its input argument. It is easy to see that $EleSum\{\cdot\}$ is a linear operator. In high-budget scenarios, we can safely assume that the perturbations caused by the measurements are small compared to the range of $\mathbf{x}$. By employing the first order Taylor series approximation for the inverse of a positive definite

Table 5.1: The measurement variances and the corresponding Fisher information scores for the Optimal Strategy (5.7), Strategy 1 (Equal measurement variances for all devices), and Strategy 2 (All cost to the best observation) corresponding to scenario in Figure 5.3.

		$\sigma^2_{m_1}$	$\sigma^2_{m_2}$	$\sigma^2_{m_3}$	$\sigma^2_{m_4}$	Fisher Info.
C = 1	Optimal	0.0333	∞	∞	∞	7.5000
	Strategy 1	1.5817	1.5817	1.5817	1.5817	1.8250
	Strategy 2	0.0333	∞	∞	∞	7.5000
C = 2	Optimal	0.0126	0.6338	∞	∞	9.7639
	Strategy 1	0.6185	0.6185	0.6185	0.6185	3.4657
	Strategy 2	0.0067	∞	∞	∞	9.3750
C = 2.5	Optimal	0.0097	0.3973	3.5334	∞	10.4545
	Strategy 1	0.4373	0.4373	0.4373	0.4373	4.2516
	Strategy 2	0.0032	∞	∞	∞	9.6875
C = 5	Optimal	0.0037	0.1096	0.4304	1.1403	12.4425
	Strategy 1	0.1172	0.1172	0.1172	0.1172	7.9127
	Strategy 2	9.775e-4	∞	∞	∞	9.9902
C = 10	Optimal	0.0006	0.0164	0.0546	0.1171	13.6262
	Strategy 1	0.0162	0.0162	0.0162	0.0162	12.3938
	Strategy 2	9.537e-8	∞	∞	∞	10.0000
C = 15	Optimal	0.0001	0.0028	0.0092	0.0193	13.8354
	Strategy 1	0.0027	0.0027	0.0027	0.0027	13.5975
	Strategy 2	9.31e-11	∞	∞	∞	10.0000

symmetric matrix, we can write $(\mathbf{\Sigma}_x + \mathbf{D}_m)^{-1} \approx \mathbf{\Sigma}_x^{-1} - \mathbf{\Sigma}_x^{-1}\mathbf{D}_m\mathbf{\Sigma}_x^{-1}$. Hence, the aim reduces to the minimization of $EleSum\ \{\mathbf{\Sigma}_x^{-1}\mathbf{D}_m\mathbf{\Sigma}_x^{-1}\}$ over $\left\{\sigma^2_{m_1}, \ldots, \sigma^2_{m_K}\right\}$ under the constraint in (5.1). It is possible to simplify the objective function further by defining $\mathbf{\Sigma}_x^{-1} = [\mathbf{e}_1\, \mathbf{e}_2\, \cdots\, \mathbf{e}_K]$ where $\mathbf{e}_i$ denotes the i^{th} column of the inverse of the observation covariance matrix. Let c_i denote the square of the sum of the elements in $\mathbf{e}_i$, that is $c_i = (Elesum\{\mathbf{e}_i\})^2$. Then the optimization problem can be expressed as follows:

$$\min_{\left\{\sigma^2_{m_i}\right\}_{i=1}^{K}} \sum_{i=1}^{K} c_i\, \sigma^2_{m_i} \text{ subject to } \sum_{i=1}^{K} \frac{1}{2} \log\left(1 + \frac{\sigma^2_{x_i}}{\sigma^2_{m_i}}\right) \leq \mathrm{C}\,,$$
$$\sigma^2_{m_i} \geq 0 \quad \forall i \in \{1, 2, \ldots, K\}\ . \tag{5.12}$$

From (5.12), it is noted that the objective function is linear, the constraint is convex, and both functions are continuously differentiable which indicate that Slater's condition holds. Therefore, KKT conditions are necessary and sufficient for optimality. Then, the optimal measurement noise variances can be calculated from

$$\sigma^2_{m_i} = -\frac{\sigma^2_{x_i}}{2} + \sqrt{\frac{\sigma^4_{x_i}}{4} + \gamma\frac{\sigma^2_{x_i}}{c_i}} \tag{5.13}$$

where $\gamma > 0$ is obtained by substituting (5.13) into the cost constraint $\sum_{i=1}^{K} \frac{1}{2} \log\left(1 + \sigma^2_{x_i}/\sigma^2_{m_i}\right) = \mathrm{C}$.

5.4 Converse to Special Case 1

In some scenarios, it may be more desirable to minimize the cost of the measurement devices while the average Fisher information stays above a certain value. By putting a lower bound on the average Fisher information, we are effectively restricting the BCRB to stay below a predetermined value. Assuming independent Gaussian observations and measurement noises, the converse problem can be formulated as

$$\begin{aligned}&\min_{\{\sigma^2_{m_i}\}_{i=1}^K} \sum_{i=1}^{K} \frac{1}{2}\log\left(1+\frac{\sigma^2_{x_i}}{\sigma^2_{m_i}}\right)\\ &\text{subject to} \sum_{i=1}^{K} \frac{1}{\sigma^2_{m_i}+\sigma^2_{x_i}} \geq \mathrm{I_F}.\end{aligned} \tag{5.14}$$

Notice that although the objective function is convex, the constraint is not a convex set. In fact, the constraint set is what is left after the convex set $\mathcal{C} = \left\{\boldsymbol{\sigma^2_m} \succeq 0 : \sum_{i=1}^{K} \frac{1}{\sigma^2_{m_i}+\sigma^2_{x_i}} < \mathrm{I_F}\right\}$ is subtracted from $\{\boldsymbol{\sigma^2_m} \succeq 0\}$. Since the global minimum of the unconstrained objective function is achieved for $\boldsymbol{\sigma^2_m} = \boldsymbol{\infty}$ which is contained in the set $\mathcal{C}$ and the objective function is convex, it is concluded that the minimum of the objective function has to occur at the boundary, i.e., $\sum_{i=1}^{K} \frac{1}{\sigma^2_{m_i}+\sigma^2_{x_i}} = \mathrm{I_F}$ must be satisfied. Then, by similar arguments to those in Section 5.2, optimal values of the measurement noise variances can be obtained from

$$\sigma^2_{m_i} = \begin{cases} \frac{\sigma^4_{x_i}}{\gamma-\sigma^2_{x_i}} & \text{if } \sigma^2_{x_i} < \gamma \\ \infty & \text{if } \sigma^2_{x_i} \geq \gamma \end{cases} \quad \text{with} \quad \gamma = \frac{|S_K|}{\left(\sum_{i\in S_K} \frac{1}{\sigma^2_{x_i}} - \mathrm{I_F}\right)} \tag{5.15}$$

where $S_K = \left\{i \in \{1,\ldots,K\} : \sigma^2_{m_i} \neq \infty\right\}$ and $|S_K|$ denotes the number of elements in the set S_K. Notice that the average Fisher information constraint should be assigned a quantity that is smaller than the unconstrained Fisher information $\sum_{i=1}^{K} 1/\sigma^2_{x_i}$ (total Fisher information at the input to the measurement devices) since additional processing cannot increase the Fisher information content.

When the observation variances are equal $\sigma^2_{x_i} = \sigma^2_x \; \forall\, i$, the optimal measurement variances are also equal and can be calculated from $\sigma^2_{m_i} = \sigma^2_m = \frac{K}{\mathrm{I_F}} - \sigma^2_x$. The corresponding minimized measurement cost is given by $\mathrm{C} = \frac{K}{2}\log\left(\frac{K}{K-\mathrm{I_F}\sigma^2_x}\right)$ which is a decreasing function of K (See Figure 5.4) with a limiting value of

$$\lim_{K\to\infty} \frac{K}{2}\log\left(\frac{K}{K-\mathrm{I_F}\sigma^2_x}\right) = \frac{\mathrm{I_F}\sigma^2_x}{2\ln 2}. \tag{5.16}$$

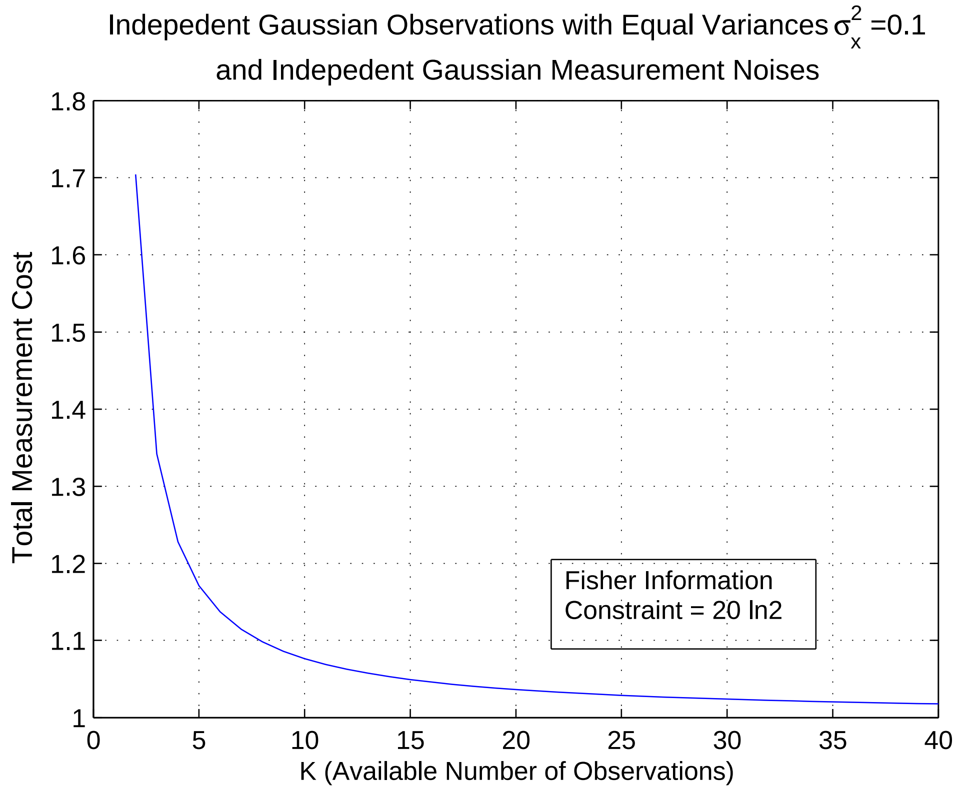

Figure 5.4: Independent Gaussian observations with variances $\sigma_x^2 = 0.1$ and independent Gaussian measurement noises are considered. The total measurement cost under Fisher information constraint with $\mathrm{I_F} = 20 \ln 2$ is depicted with respect to the number of available measurements. Under this scenario, total cost decreases with each additional measurement to a limiting value of $\mathrm{C} = 1$. Notice that it is not possible to achieve the Fisher information constraint using a single observation.

5.5 Concluding Remarks

In this chapter, an optimal estimation framework is considered in the presence of cost constrained measurements [50]. The aim is to maximize the average Fisher information under a constraint on the total cost of measurement devices. An optimization problem is formulated to calculate the optimal costs of measurement devices that maximize the average Fisher information for arbitrary observation and measurement statistics. In addition, closed form expressions are obtained in the case of Gaussian observations and measurement noises. The converse problem is also addressed for the Gaussian case where we consider minimizing the cost of the measurements such that the average Fisher information is not smaller than a predetermined value. Numerical examples are presented to explain the results.

6

Cost Minimization of Measurement Devices under Estimation Accuracy Constraints in the Presence of Gaussian Noise

This chapter is organized as follows. In Section 6.1, novel convex measurement cost minimization problems are proposed based on various estimation accuracy constraints for a linear system subject to additive Gaussian noise. In Section 6.2, we modify the proposed optimization problems to handle the worst-case scenarios under system matrix uncertainty. Next, we take a specific but nevertheless practical uncertainty model, and discuss how the optimization problems are altered while preserving convexity. In Section 6.3, we focus on two optimization problems proposed in Section 6.1, and simplify them to obtain closed form solutions in the case of invertible system matrix. In Section 6.4, we provide several

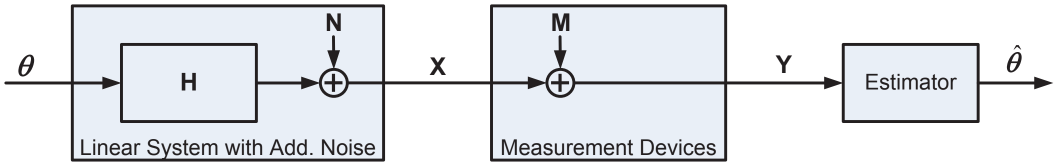

Figure 6.1: Measurement and estimation systems model block diagram for a linear system with additive noise.

numerical examples to illustrate the results presented in this chapter. Extensions to Bayesian estimation with Gaussian priors are discussed in Section 6.5.

6.1 Optimal Cost Allocation under Estimation Accuracy Constraints

Consider a discrete-time system model as in Figure 6.1 in which noisy measurements are obtained at the output of a linear system, and then the measurements are processed to estimate the value of a non-random parameter vector $\boldsymbol{\theta}$. The observation vector $\mathbf{X}$ at the output of the linear system can be represented by $\mathbf{X} = \mathbf{H}^T\boldsymbol{\theta} + \mathbf{N}$, where $\boldsymbol{\theta} \in \mathbb{R}^L$ denotes a vector of parameters to estimate, $\mathbf{N} \in \mathbb{R}^K$ is the inherent random system noise, and $\mathbf{X} \in \mathbb{R}^K$ is the observation vector at the output of the linear system. The system noise $\mathbf{N}$ is assumed to be a Gaussian distributed random vector with zero-mean, independent but not necessarily identical components, i.e., $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_\mathbf{N})$, where $\mathbf{D}_\mathbf{N} = \mathrm{diag}\{\sigma_{n_1}^2, \sigma_{n_2}^2, \ldots, \sigma_{n_K}^2\}$ is a diagonal covariance matrix, and $\mathbf{0}$ denotes the all-zeros vector of length K. We also assume that the number of observations is at least equal to the number of estimated parameters (i.e., $K \geq L$) and the system matrix $\mathbf{H}$ is an $L \times K$ matrix with full row rank L so that the columns of $\mathbf{H}$ span $\mathbb{R}^L$.

Noisy measurements of the observation vector $\mathbf{X}$ are made by K measurement devices at the output of the linear system, and then the measured values in vector $\mathbf{Y} \in \mathbb{R}^K$ are processed to estimate the parameter vector $\boldsymbol{\theta}$. It is assumed that each measurement device is capable of sensing the value of a *scalar* physical quantity with some resolution in amplitude according to the measurement model $y_i = x_i + m_i$, where m_i denotes the measurement noise associated with the i^{th} measurement device. In other words, measurement devices are modeled to introduce additive random measurement noise which can be expressed as $\mathbf{Y} = \mathbf{X}+\mathbf{M}$. It is also reasonable to assume that measurement noise vector $\mathbf{M}$ is independent of the inherent system noise $\mathbf{N}$. In addition, the noise components introduced by the measurement devices (the elements of $\mathbf{M}$) are assumed to be zero-mean independent Gaussian random variables with possibly distinct variances[1], i.e., $\mathbf{M} \sim \mathcal{N}(\mathbf{0}, \mathbf{D_M})$, where $\mathbf{D_M}$ is a diagonal covariance matrix given by $\mathbf{D_M} = \text{diag}\{\sigma^2_{m_1}, \sigma^2_{n_2}, \ldots, \sigma^2_{m_K}\}$. Based on the outputs of the measurements devices, unknown parameter vector $\boldsymbol{\theta}$ is estimated.

In practical scenarios, a major issue is the cost of performing measurements. The cost of a measurement device is primarily assessed with its resolution, more specifically with the number of amplitude levels that the device can reliably discriminate. Intuitively, as the accuracy of a measurement device increases so does its cost. Therefore, it may not always be possible to make high resolution measurements with a limited budget. In a recent work [46], a novel measurement device model is proposed where the cost of each device is expressed quantitatively in terms of the number of amplitude levels that can be resolved reliably. In this model, the amplitude resolution of the measurement devices solely determines the cost of each measurement. The dynamic range or scaling of the input to the measurement device is assumed to have no effect on the cost as long as the number of resolvable levels stays the same. More explicitly, in [46], the cost associated with measuring the i^{th} component of the observation vector $\mathbf{x}$

[1]Since Gaussian distribution maximizes the differential entropy over all distributions with the same variance, the assumption that the errors introduced by the measurement devices are Gaussian distributed handles the worst-case scenario.

is given by $\mathrm{C}_i = 0.5\log_2\left(1+\sigma^2_{x_i}/\sigma^2_{m_i}\right)$, where $\sigma^2_{x_i}$ denotes the variance of the i^{th} component of observation vector $\mathbf{x}$ (i.e., the variance of the input to the i^{th} measurement device), and $\sigma^2_{m_i}$ is the variance of the i^{th} component of $\mathbf{m}$ (i.e., the variance of the noise introduced by the i^{th} measurement device)[2]. Notice that $\sigma^2_{x_i} = \sigma^2_{n_i}$, $\forall i \in \{1, 2, \ldots, K\}$, since $\boldsymbol{\theta}$ is a deterministic parameter vector. Then, the overall cost of measuring all the components of the observation vector $\mathbf{x}$ is expressed as

$$\mathrm{C} = \sum_{i=1}^{K} \mathrm{C}_i = \sum_{i=1}^{K} \frac{1}{2}\log_2\left(1 + \frac{\sigma^2_{n_i}}{\sigma^2_{m_i}}\right) . \tag{6.1}$$

A closer look into (6.1) reveals that it is a nonnegative, monotonically decreasing and convex function of $\sigma^2_{m_i}$, $\forall\,\sigma^2_{n_i} > 0$ and $\forall\,\sigma^2_{m_i} > 0$. It is also noted that a measurement device has a higher cost if it can perform measurements with a lower measurement variance (i.e., with higher accuracy). Such an approach brings great flexibility by enabling to work with variable precision over the acquired measurements. After formulating the measurement device model as outlined above, our objective is to minimize the total cost of the measurement devices under a constraint on estimation accuracy. In other words, we are allowed to design the noise levels of the measurement devices such that the overall cost is minimized under a constraint on the minimum acceptable estimation performance.

In non-random parameter estimation problems, the Cramer-Rao bound (CRB) provides a lower bound on the mean-squared-errors (MSEs) of unbiased estimators under some regularity conditions [23]. Specifically, the CRB on the estimation error for an arbitrary unbiased estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ is expressed as

$$\mathbb{E}\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^T\right\} \succeq \mathbf{J}^{-1}(\mathbf{Y},\boldsymbol{\theta}) \triangleq \mathrm{CRB}\,, \tag{6.2}$$

[2]For an in-depth discussion on the plausibility of this measurement device model and its relation to the number of distinguishable amplitude levels, we refer the reader to [46].

where $\mathbf{J}(\mathbf{Y},\boldsymbol{\theta})$ is the Fisher information matrix (FIM) of the measurement $\mathbf{Y}$ relative to the parameter vector $\boldsymbol{\theta}$, which is defined as

$$\mathbf{J}(\mathbf{Y},\boldsymbol{\theta}) \triangleq \int \frac{1}{p_{\mathbf{Y}}^{\boldsymbol{\theta}}(\mathbf{y})} \left(\frac{\partial p_{\mathbf{Y}}^{\boldsymbol{\theta}}(\mathbf{y})}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial p_{\mathbf{Y}}^{\boldsymbol{\theta}}(\mathbf{y})}{\partial \boldsymbol{\theta}}\right)^T \mathrm{d}\mathbf{y}\,, \tag{6.3}$$

where $\partial/\partial\boldsymbol{\theta}$ denotes the gradient (i.e., a column vector of partial derivatives) with respect to parameters $\theta_1, \ldots, \theta_K$. Or, equivalently, the elements of the FIM can be calculated from [23]

$$J_{ij} = -\mathbb{E}_{\mathbf{Y}|\boldsymbol{\theta}} \left\{ \frac{\partial^2 \log p_{\mathbf{Y}}^{\boldsymbol{\theta}}(\mathbf{Y})}{\partial\theta_i \partial\theta_j} \right\} . \tag{6.4}$$

The symbol $\succeq$ between nonnegative definite matrices in (6.2) represents the inequality with respect to the positive semidefinite matrix cone. Specifically, it indicates that the difference matrix obtained by subtracting the right hand side of the inequality from the left hand side is nonnegative definite. Assuming independent Gaussian distributions for $\mathbf{N}$ and $\mathbf{M}$, it can be shown that the CRB is given as follows [64]

$$\mathrm{CRB} = \mathbf{J}^{-1}(\mathbf{Y},\boldsymbol{\theta}) = \left(\mathbf{H}\mathrm{Cov}^{-1}(\mathbf{N}+\mathbf{M})\mathbf{H}^T\right)^{-1} , \tag{6.5}$$

where $\mathrm{Cov}(\cdot)$ denotes the covariance matrix of the random vector $\mathbf{N}+\mathbf{M}$ and $\mathrm{Cov}(\mathbf{N}+\mathbf{M}) = \mathbf{D_N} + \mathbf{D_M} = \mathrm{diag}\{\sigma_{n_1}^2 + \sigma_{m_1}^2\,,\, \sigma_{n_2}^2 + \sigma_{m_2}^2\,,\, \ldots\,,\, \sigma_{n_K}^2 + \sigma_{m_K}^2\}$ due to independence. Then, $\mathbf{D} \triangleq \mathrm{Cov}^{-1}(\mathbf{N}+\mathbf{M}) =$ $\mathrm{diag}\left\{1/\left(\sigma_{n_1}^2 + \sigma_{m_1}^2\right)\,,\, 1/\left(\sigma_{n_2}^2 + \sigma_{m_2}^2\right)\,,\, \ldots\,,\, 1/\left(\sigma_{n_K}^2 + \sigma_{m_K}^2\right)\right\}$, where $\mathrm{Cov}^{-1}(\cdot)$ represents the inverse of the covariance matrix. Notice that the CRB can actually be attained in this case by employing the maximum likelihood (ML) estimator (also the best linear unbiased estimator (BLUE) in this case), $\hat{\boldsymbol{\theta}}(\mathbf{y}) = \left(\mathbf{HDH}^T\right)^{-1}\mathbf{HDy}$, where the efficiency of the estimator follows from linearity of the system and due to the assumption of Gaussian distributions [23]. Specifically, the covariance matrix of the estimator equals the inverse of the FIM, i.e., $\mathrm{Cov}\left(\hat{\boldsymbol{\theta}}(\mathbf{y})\right) = \left(\mathbf{HDH}^T\right)^{-1}$.

Remark: When non-Gaussian distributions are assumed, we can utilize the preceding observation to obtain an upper bound on the CRB. To see this, a few preliminaries are needed. First, the FIM of a random vector $\mathbf{Z}$ with respect to a translation parameter is defined as follows [64]

$$\mathbf{J}(\mathbf{Z}) \triangleq \mathbf{J}(\boldsymbol{\theta}+\mathbf{Z},\boldsymbol{\theta}) = \int \frac{1}{p_{\mathbf{Z}}(\mathbf{z})} \left(\frac{\partial p_{\mathbf{Z}}(\mathbf{z})}{\partial \mathbf{z}}\right) \left(\frac{\partial p_{\mathbf{Z}}(\mathbf{z})}{\partial \mathbf{z}}\right)^T \mathrm{d}\mathbf{z}\,, \tag{6.6}$$

where $p_{\mathbf{Z}}(\mathbf{z})$ is the probability density function of $\mathbf{Z}$ that is independent of $\boldsymbol{\theta}$. A well-known property of the FIM under translation is $\mathbf{J}(\mathbf{Z}) \succeq \mathrm{Cov}^{-1}(\mathbf{Z})$ with equality if and only if $\mathbf{Z}$ is Gaussian [64].

Based on these preliminaries, for linear models in the form of Figure 6.1 but with arbitrary probability distributions for $\mathbf{N}$ and $\mathbf{M}$, it can be shown that $\mathbf{J}(\mathbf{Y},\boldsymbol{\theta}) = \mathbf{H}\mathbf{J}(\mathbf{N}{+}\mathbf{M})\mathbf{H}^T$, where $\mathbf{J}(\mathbf{N}{+}\mathbf{M})$ indicates the FIM under a translation parameter of random vector $\mathbf{N}+\mathbf{M}$ [64]. In order to upper bound the CRB, it is first observed that $\mathbf{J}(\mathbf{N}+\mathbf{M}) \succeq \mathrm{Cov}^{-1}(\mathbf{N}+\mathbf{M})$. Using the properties of nonnegative definite matrices, we have

$$\mathrm{CRB} = \mathbf{J}^{-1}(\mathbf{Y},\boldsymbol{\theta}) = \left(\mathbf{H}\mathbf{J}(\mathbf{N}+\mathbf{M})\mathbf{H}^T\right)^{-1} \preceq \left(\mathbf{H}\mathrm{Cov}^{-1}(\mathbf{N}+\mathbf{M})\mathbf{H}^T\right)^{-1}, \tag{6.7}$$

which naturally indicates that the difference matrix obtained by subtracting the CRB from the covariance matrix of the linear estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ must be nonnegative definite. Correspondingly, it is also possible to lower bound the CRB for independent random vectors $\mathbf{N}$ and $\mathbf{M}$. To that aim, we can revert to the Fisher Information Inequality (FII) [65]. FII states that $\mathbf{J}^{-1}(\mathbf{N}+\mathbf{M}) \succeq \mathbf{J}^{-1}(\mathbf{N}) + \mathbf{J}^{-1}(\mathbf{M})$ with equality if and only if $\mathbf{N}$ and $\mathbf{M}$ are Gaussian. Therefore,

$$\mathrm{CRB} = \mathbf{J}^{-1}(\mathbf{Y},\boldsymbol{\theta}) \succeq \left(\mathbf{H}\left(\mathbf{J}^{-1}(\mathbf{N}) + \mathbf{J}^{-1}(\mathbf{M})\right)^{-1}\mathbf{H}^T\right)^{-1}. \tag{6.8}$$

As a result, a lower bound on the CRB can also be obtained in terms of the FIMs under translation parameters (6.6) of random vectors $\mathbf{N}$ and $\mathbf{M}$ with arbitrary probability distributions. □

Returning to our case of independent Gaussian system noise and measurement noise, the CRB is equal to the covariance matrix (i.e., estimation error covariance) of the ML estimator $\hat{\boldsymbol{\theta}}(\mathbf{y}) = \left(\mathbf{H}\mathbf{D}\mathbf{H}^T\right)^{-1}\mathbf{H}\mathbf{D}\mathbf{y}$ as mentioned in the paragraph following (6.5). Furthermore, when the system and measurement noise distributions are not restricted to Gaussian, the covariance matrix of the linear estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ can also be used as an upper bound to the CRB as shown in (6.7). For this reason, in the following analysis we employ several performance metrics based on the CRB given in (6.5) in order to assess the quality of estimation. In other words, we propose measurement cost minimization formulations under various estimation accuracy constraints based on the CRB expression in (6.5). However, before that analysis, we first express the CRB in a more familiar form in the optimization theoretic sense

$$\mathrm{CRB} = \mathbf{J}^{-1}(\mathbf{Y},\boldsymbol{\theta}) = \left(\sum_{i=1}^{K} \frac{1}{\sigma_{n_i}^2 + \sigma_{m_i}^2} \mathbf{h}_i \mathbf{h}_i^T\right)^{-1}, \tag{6.9}$$

and the corresponding ML estimator that achieves this bound becomes

$$\hat{\boldsymbol{\theta}}(\mathbf{y}) = \left(\mathbf{H}\mathbf{D}\mathbf{H}^T\right)^{-1}\mathbf{H}\mathbf{D}\mathbf{y} = \left(\sum_{i=1}^{K} \frac{1}{\sigma_{n_i}^2 + \sigma_{m_i}^2} \mathbf{h}_i \mathbf{h}_i^T\right)^{-1} \sum_{i=1}^{K} \frac{y_i}{\sigma_{n_i}^2 + \sigma_{m_i}^2} \mathbf{h}_i \,. \tag{6.10}$$

6.1.1 Average Mean-Squared-Error

The diagonal components of the CRB provide a lower bound on the MSE while estimating the components of parameter $\boldsymbol{\theta}$. Specifically,

$$\mathbb{E}_{\mathbf{Y}|\boldsymbol{\theta}}\left\{\left\|\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta}\right\|_2^2\right\} \geq \mathrm{tr}\left\{\mathbf{J}^{-1}(\mathbf{Y},\boldsymbol{\theta})\right\}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace operator [23]. In other words, the harmonic average of the eigenvalues of the FIM is taken as the performance metric. Based on this metric, the following measurement cost minimization problem is proposed:

$$\begin{aligned}
\min_{\{\sigma_{m_i}^2\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K} \log_2\left(1 + \frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right) \\
\text{subject to} \quad & \mathrm{tr}\left\{\left(\sum_{i=1}^{K} \frac{1}{\sigma_{n_i}^2 + \sigma_{m_i}^2} \mathbf{h}_i \mathbf{h}_i^T\right)^{-1}\right\} \leq \mathrm{E}\,,
\end{aligned} \tag{6.11}$$

where E denotes a constraint on the maximum allowable average estimation error. Due to the inevitable intrinsic system noise, the design criterion E must satisfy $\mathrm{E} > \mathrm{tr}\left\{\left(\mathbf{H}\mathbf{D}_{\mathbf{N}}^{-1}\mathbf{H}^T\right)^{-1}\right\} = \mathrm{tr}\left\{\left(\sum_{i=1}^K \frac{\mathbf{h}_i\mathbf{h}_i^T}{\sigma_{n_i}^2}\right)^{-1}\right\}$. Substituting $\mu_i = 1/\left(\sigma_{n_i}^2 + \sigma_{m_i}^2\right)$, (6.11) becomes

$$\begin{aligned} \max_{\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^K \log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\ \text{subject to} \quad & \mathrm{tr}\left\{\left(\sum_{i=1}^K \mu_i\,\mathbf{h}_i\mathbf{h}_i^T\right)^{-1}\right\} \le \mathrm{E}\,. \end{aligned} \tag{6.12}$$

It is noted that the objective function is smooth and concave for $\forall\,\mu_i \in [0, 1/\sigma_{n_i}^2)$. Since the constraint is also a convex function of μ_i's for $\forall\,\mu_i \ge 0$, this is a convex optimization problem [66, Sec. 7.5.2]. Consequently, it can be efficiently solved in polynomial time using interior point methods and the numerical convergence is assured. It is also possible to express this optimization problem using linear matrix inequalities (LMIs) as follows:

$$\begin{aligned} \max_{\{z_i\}_{i=1}^L,\,\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^K \log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\ \text{subject to} \quad & \begin{bmatrix} \sum_{i=1}^K \mu_i\,\mathbf{h}_i\mathbf{h}_i^T & \mathbf{e}_j \\ \mathbf{e}_j^T & z_i \end{bmatrix} \succeq 0, \quad j = 1,\ldots,L \\ & \sum_{i=1}^K z_i \le \mathrm{E}\,, \end{aligned} \tag{6.13}$$

where $\mathbf{e}_j$ denotes the column vector of length L with a 1 in the $j^{\,\mathrm{th}}$ coordinate and 0's elsewhere. Or equivalently,

$$\begin{aligned} \max_{\mathbf{Z}\in\mathbf{S}_L,\,\{\mu_i\}_{i=1}^K} \quad \frac{1}{2}\sum_{i=1}^K \log_2\left(1-\sigma_{n_i}^2\mu_i\right) \quad \text{subject to} \quad & \begin{bmatrix} \mathbf{Z} & \mathbf{I} \\ \mathbf{I} & \sum_{i=1}^K \mu_i\,\mathbf{h}_i\mathbf{h}_i^T \end{bmatrix} \succeq 0\,, \\ & \mathrm{tr}(\mathbf{Z}) \le \mathrm{E}\,, \end{aligned} \tag{6.14}$$

where $\mathbf{S}_L$ denotes the set of symmetric $L \times L$ matrices.

6.1.2 Shannon Information

An alternative measure of the estimation accuracy considers the Shannon (mutual) information content between the unknown parameter vector $\boldsymbol{\theta}$ and the measurement vector $\mathbf{Y}$. More explicitly, the interest is to place a constraint on the log volume of the η-confidence ellipsoid which is defined as the minimum ellipsoid that contains the estimation error with probability η [66, Sec. 7.5.2]. As shown in [44], the η-confidence ellipsoid is given by

$$\varepsilon_\alpha = \left\{\mathbf{z} \,|\, \mathbf{z}^T \mathbf{J}(\mathbf{Y}, \boldsymbol{\theta})\mathbf{z} \,\le\, \alpha\right\}, \tag{6.15}$$

where $\alpha = F^{-1}_{\chi^2_K}(\eta)$ is obtained from the cumulative distribution function of a chi-squared random variable with K degrees of freedom. Then, the log volume of the η-confidence ellipsoid is obtained as[3]

$$\log \mathbf{vol}(\varepsilon_\alpha) = \beta - \frac{1}{2} \log\det\left(\sum_{i=1}^{K} \frac{1}{\sigma^2_{n_i} + \sigma^2_{m_i}} \mathbf{h}_i \mathbf{h}_i^T\right), \tag{6.16}$$

where $\beta = \frac{n}{2}\log(\alpha\pi) - \log\left(\Gamma\left(\frac{n}{2}+1\right)\right)$, with Γ denoting the Gamma function. Notice that the design criterion is related to the geometric mean of the eigenvalues of the FIM. Based on this metric, the following measurement cost optimization problem can be obtained:

$$\begin{aligned} \max_{\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K} \log_2\left(1 - \sigma^2_{n_i}\mu_i\right) \\ \text{subject to} \quad & \log\det\left(\sum_{i=1}^{K} \mu_i\, \mathbf{h}_i\mathbf{h}_i^T\right) \,\ge\, 2(\beta - \mathrm{S}), \end{aligned} \tag{6.17}$$

where μ_i is as defined in (6.12) and S is a constraint on the log volume of η-confidence ellipsoid satisfying $\mathrm{S} > \beta - 0.5 \log\det\left(\mathbf{H}\mathbf{D}_{\mathbf{N}}^{-1}\mathbf{H}^T\right) = \beta - 0.5\log\det\left(\sum_{i=1}^K \frac{\mathbf{h}_i\mathbf{h}_i^T}{\sigma^2_{n_i}}\right)$. Since $\log\det\left(\sum_{i=1}^K \mu_i\,\mathbf{h}_i\mathbf{h}_i^T\right)$ is a smooth concave function of μ_i for $\mu_i \ge 0$, the resulting optimization problem is convex [66, Sec. 3.1.5]. The smoothness property of the problem is also very helpful for obtaining the solution via numerical methods.

[3] We use 'log' without a subscript to denote the natural logarithm.

By introducing a lower triangular non-singular matrix $\mathbf{L}$ and utilizing Cholesky decomposition of positive definite matrices, it is possible to rewrite the constraint in terms of a lower bound. To that aim, let $\sum_{i=1}^{K} \mu_i \, \mathbf{h}_i \mathbf{h}_i^T \succeq \mathbf{L}\mathbf{L}^T$. Then, the optimization problem can be expressed equivalently as

$$\max_{\mathbf{L} \in \mathbf{U}_L, \, \{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^{K} \log_2 \left(1 - \sigma_{n_i}^2 \mu_i\right) \quad \text{subject to} \; \begin{bmatrix} \mathbf{I} & \mathbf{L}^T \\ \mathbf{L} & \sum_{i=1}^{K} \mu_i \, \mathbf{h}_i \mathbf{h}_i^T \end{bmatrix} \succeq 0 \,, \quad \sum_{i=1}^{L} \log \, L_{i,\,i} \; \geq \; (\beta - \mathrm{S}) \,, \tag{6.18}$$

where $\mathbf{U}_L$ denotes the set of lower triangular non-singular $L \times L$ square matrices, $L_{i,\,i}$ represents the $i^{\,\text{th}}$ diagonal coefficient of $\mathbf{L}$, and L is the dimension of $\mathbf{L}$.

6.1.3 Worst-Case Error Variance

When the primary concern shifts from accuracy requirements towards robust behavior, it may be more desirable to have a constraint on the worst-case variance of the estimation error, which is associated with the maximum (minimum) eigenvalue of the CRB (FIM) [44, 67–69]. The corresponding optimization problem is stated as follows:

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^{K} \log_2 \left(1 - \sigma_{n_i}^2 \mu_i\right) \quad \text{subject to} \quad \lambda_{\min} \left\{ \sum_{i=1}^{K} \mu_i \, \mathbf{h}_i \mathbf{h}_i^T \right\} \; \geq \; \Lambda \,, \tag{6.19}$$

where $\lambda_{\min}\{\cdot\}$ represents the minimum eigenvalue of its argument, and Λ is a predetermined lower bound on the minimum eigenvalue of the FIM satisfying $\Lambda < \lambda_{\min} \left\{ \mathbf{H} \mathbf{D}_{\mathbf{N}}^{-1} \mathbf{H}^T \right\} = \lambda_{\min} \left\{ \sum_{i=1}^{K} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\sigma_{n_i}^2} \right\}$. Since the constraint can be represented in the form of an LMI, this problem can equivalently be expressed as

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^{K} \log_2 \left(1 - \sigma_{n_i}^2 \mu_i\right) \quad \text{subject to} \quad \sum_{i=1}^{K} \mu_i \, \mathbf{h}_i \mathbf{h}_i^T \; \succeq \; \Lambda \mathbf{I} \,, \tag{6.20}$$

where $\mathbf{I}$ is the $L \times L$ identity matrix. The resulting problem is also convex [66, Sec. 7.5.2].

6.1.4 Worst-Case Coordinate Error Variance

Another variation of the worst-case error criteria can be obtained by placing a constraint on the maximum error variance among all the individual estimator components, i.e., restricting the largest diagonal entry of the CRB. Using this performance criterion, we have the following optimization problem

$$\begin{aligned}
\max_{\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K} \log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\
\text{subject to} \quad & \max_{j=1,\ldots,K} \left(\left(\sum_{i=1}^{K} \mu_i\, \mathbf{h}_i\mathbf{h}_i^T\right)^{-1}\right)_{j,j} \le \varrho\,,
\end{aligned} \tag{6.21}$$

where ϱ is a constraint on the maximum allowable diagonal entry of the CRB (estimation error covariance matrix) satisfying $\varrho > \max_{j=1,\ldots,K}\left(\left(\mathbf{H}\mathbf{D}_{\mathbf{N}}^{-1}\mathbf{H}^T\right)^{-1}\right)_{j,j} = \max_{j=1,\ldots,K}\left(\left(\sum_{i=1}^{K}\frac{\mathbf{h}_i\mathbf{h}_i^T}{\sigma_{n_i}^2}\right)^{-1}\right)_{j,j}$. This problem can equivalently be expressed as

$$\begin{aligned}
\max_{\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K} \log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\
\text{subject to} \quad & \begin{bmatrix} \varrho & \mathbf{e}_j^T \\ \mathbf{e}_j & \sum_{i=1}^{K} \mu_i\, \mathbf{h}_i\mathbf{h}_i^T \end{bmatrix} \succeq 0, \quad j=1,\ldots,L
\end{aligned} \tag{6.22}$$

where $\mathbf{e}_j$ denotes the column vector of length L with a 1 in the j^{th} coordinate and 0's elsewhere. This is also a convex optimization problem [66, Sec. 7.5.2].

6.2 Extensions to Cases with System Matrix Uncertainty - Robust Measurement

It may also be the case that there exists some uncertainty concerning the elements in the system matrix $\mathbf{H}$ [44]. Suppose that the system matrix $\mathbf{H}$ can take values from a given finite set $\mathcal{H}$. In the robust measurement problem, we consider the optimization over the worst-case scenario. Specifically, we choose the matrix from the family of system matrices $\mathcal{H}$ resulting in the worst estimation accuracy constraint, and perform the optimization accordingly. Recalling that the infimum (supremum) preserves concavity (convexity), it is possible to restate the measurement cost optimization problems given in Section 6.1, and still maintain convex optimization problems. Then, the resulting optimization problems with respect to each criterion are expressed as follows

6.2.1 Average Mean-Squared-Error

$$\begin{aligned} \max_{\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K} \log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\ \text{subject to} \quad & \sup_{\mathbf{H}\in\mathcal{H}} \operatorname{tr}\left\{\sum_{i=1}^{K} \mu_i\,\mathbf{h}_i\mathbf{h}_i^T\right\}^{-1} \leq \mathrm{E}\,, \end{aligned} \tag{6.23}$$

or equivalently,

$$\begin{aligned} \max_{\mathbf{Z}\in\mathbf{S}_L,\,\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K} \log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\ \text{subject to} \quad & \begin{bmatrix} \mathbf{Z} & \mathbf{I} \\ \mathbf{I} & \sum_{i=1}^K \mu_i\,\mathbf{h}_i\mathbf{h}_i^T \end{bmatrix} \succeq 0 \quad \text{for all } \mathbf{H}\in\mathcal{H} \\ & \operatorname{tr}(\mathbf{Z}) \leq \mathrm{E}\,. \end{aligned} \tag{6.24}$$

6.2.2 Shannon Information

$$\begin{aligned} \max_{\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K} \log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\ \text{subject to} \quad & \inf_{\mathbf{H}\in\mathcal{H}} \log\det\left\{\sum_{i=1}^{K} \mu_i\,\mathbf{h}_i\mathbf{h}_i^T\right\} \geq 2(\beta-\mathrm{S})\,, \end{aligned} \tag{6.25}$$

or equivalently,

$$\begin{aligned}
\max_{\mathbf{L}\in\mathbf{U}_L,\,\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\
\text{subject to} \quad & \begin{bmatrix} \mathbf{I} & \mathbf{L}^T \\ \mathbf{L} & \sum_{i=1}^K \mu_i\,\mathbf{h}_i\mathbf{h}_i^T \end{bmatrix} \succeq 0 \quad \text{for all } \mathbf{H}\in\mathcal{H}\,, \\
& \sum_{i=1}^{L}\log\, L_{i,i} \,\geq\, (\beta-\mathrm{S})\,.
\end{aligned} \tag{6.26}$$

6.2.3 Worst-Case Error Variance

$$\begin{aligned}
\max_{\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\
\text{subject to} \quad & \sum_{i=1}^{K}\mu_i\,\mathbf{h}_i\mathbf{h}_i^T \,\succeq\, \Lambda\mathbf{I} \quad \text{for all } \mathbf{H}\in\mathcal{H}\,.
\end{aligned} \tag{6.27}$$

6.2.4 Worst-Case Coordinate Error Variance

$$\begin{aligned}
\max_{\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1-\sigma_{n_i}^2\mu_i\right) \\
\text{subject to} \quad & \sup_{\mathbf{H}\in\mathcal{H}}\ \max_{j=1,\ldots,K}\left(\left\{\sum_{i=1}^{K}\mu_i\,\mathbf{h}_i\mathbf{h}_i^T\right\}^{-1}\right)_{j,j} \leq \varrho\,.
\end{aligned} \tag{6.28}$$

When the set $\mathcal{H}$ is finite, the problem can be solved using standard arguments from convex optimization. However, the set $\mathcal{H}$ is in general not finite, and the solutions of the above optimization problems require general techniques from semi-infinite convex optimization such as those explained in [70, 71]. In the following, a specific uncertainty model is considered where it is possible to further simplify the optimization problems given in (6.26) and (6.27) by expressing the constraints as LMIs. To that aim, let $\mathbf{H}\in\mathcal{H}=\left\{\bar{\mathbf{H}}+\mathbf{\Delta}\,:\,\left\|\mathbf{\Delta}^T\right\|_2\leq\epsilon\right\}$, where $\|\cdot\|_2$ denotes the spectral norm (i.e., the square root of the largest eigenvalue of the positive semidefinite matrix $\mathbf{\Delta}\mathbf{\Delta}^T$). It is possible to express this constraint as an LMI, $\boldsymbol{\Delta}\boldsymbol{\Delta}^T \preceq \epsilon^2\mathbf{I}$. Suppose also that $\boldsymbol{\mu}$ is defined as the following diagonal matrix $\boldsymbol{\mu} \triangleq \operatorname{diag}\{\mu_1\,,\,\mu_2\,,\,\ldots\,,\,\mu_K\}$, and $\mathbf{W} \triangleq \mathbf{L}\mathbf{L}^T$ is a symmetric positive definite matrix. Then, the constraint in (6.26) can be expressed in terms of $\bar{\mathbf{H}}$ and $\boldsymbol{\Delta}$ as

$$\mathbf{W} \preceq \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T + \bar{\mathbf{H}}\boldsymbol{\mu}\boldsymbol{\Delta}^T + \boldsymbol{\Delta}\boldsymbol{\mu}\bar{\mathbf{H}}^T + \boldsymbol{\Delta}\boldsymbol{\mu}\boldsymbol{\Delta}^T, \quad \text{for all } \boldsymbol{\Delta}\boldsymbol{\Delta}^T \preceq \epsilon^2\mathbf{I}\,. \tag{6.29}$$

Similarly, the constraint in (6.27) is given by

$$\Lambda\mathbf{I} \preceq \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T + \bar{\mathbf{H}}\boldsymbol{\mu}\boldsymbol{\Delta}^T + \boldsymbol{\Delta}\boldsymbol{\mu}\bar{\mathbf{H}}^T + \boldsymbol{\Delta}\boldsymbol{\mu}\boldsymbol{\Delta}^T, \quad \text{for all } \boldsymbol{\Delta}\boldsymbol{\Delta}^T \preceq \epsilon^2\mathbf{I}\,. \tag{6.30}$$

In [72, Theorem 3.3], a necessary and sufficient condition is derived for quadratic matrix inequalities in the form of (6.29) and (6.30) to be true. In the light of this theorem, (6.29) holds if and only if there exists $t \geq 0$ such that

$$\begin{bmatrix} \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T - \mathbf{W} - t\mathbf{I} & \bar{\mathbf{H}}\boldsymbol{\mu} \\ \boldsymbol{\mu}\bar{\mathbf{H}}^T & \boldsymbol{\mu} + \frac{t}{\epsilon^2}\mathbf{I} \end{bmatrix} \succeq 0\,, \tag{6.31}$$

and (6.30) holds if and only if there exists $t \geq 0$ such that

$$\begin{bmatrix} \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T - (\Lambda + t)\mathbf{I} & \bar{\mathbf{H}}\boldsymbol{\mu} \\ \boldsymbol{\mu}\bar{\mathbf{H}}^T & \boldsymbol{\mu} + \frac{t}{\epsilon^2}\mathbf{I} \end{bmatrix} \succeq 0\,. \tag{6.32}$$

Notice that (6.31) and (6.32) are both linear in $\boldsymbol{\mu}$, $\mathbf{W}$ and t. Hence, under this specific uncertainty model, we can express the optimization problem in (6.26) as

$$\begin{aligned} \max_{t,\,\mathbf{W}\in\mathbf{S}_{++}^L,\,\{\mu_i\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1 - \sigma_{n_i}^2\mu_i\right) \\ \text{subject to} \quad & \begin{bmatrix} \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T - \mathbf{W} - t\mathbf{I} & \bar{\mathbf{H}}\boldsymbol{\mu} \\ \boldsymbol{\mu}\bar{\mathbf{H}}^T & \boldsymbol{\mu} + \frac{t}{\epsilon^2}\mathbf{I} \end{bmatrix} \succeq 0\,, \\ & \log\det(\mathbf{W}) \geq 2(\beta - \mathrm{S})\,, \\ & t \geq 0\,, \end{aligned} \tag{6.33}$$

where $\mathbf{S}_{++}^L$ denotes symmetric positive-definite $L \times L$ matrices. Similarly, it is possible to write the optimization problem in (6.27) as

$$\max_{t,\,\{\mu_i\}_{i=1}^K} \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1 - \sigma_{n_i}^2\mu_i\right) \quad \text{subject to} \quad \begin{bmatrix} \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T - (\Lambda + t)\mathbf{I} & \bar{\mathbf{H}}\boldsymbol{\mu} \\ \boldsymbol{\mu}\bar{\mathbf{H}}^T & \boldsymbol{\mu} + \frac{t}{\epsilon^2}\mathbf{I} \end{bmatrix} \succeq 0\,,$$

$$t \geq 0\,. \tag{6.34}$$

6.3 Special Case - Invertible System Matrix H

When the system matrix $\mathbf{H}$ is a $K \times K$ invertible matrix meaning that the number of unknown parameters is equal to the number of observations, it is possible to obtain closed-form solutions of the optimization problems stated in (6.11) and (6.17). Moreover, for the solution of (6.11), it is not necessary to assume that the components of the system noise $\mathbf{N}$ are independent; it is sufficient to have $\mathbf{N}$ as a Gaussian distributed random vector with zero-mean and arbitrary covariance matrix (possibly colored), i.e., $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_N})$ with $\{\sigma^2_{n_1}, \sigma^2_{n_2}, \ldots, \sigma^2_{n_K}\}$ constituting the diagonal components of $\mathbf{\Sigma_N}$, and $\mathbf{0}$ denoting the all-zeros vector of length K as before. To that aim, assuming independent Gaussian distributions for $\mathbf{N}$ and $\mathbf{M}$, and square $\mathbf{H}$ with full-rank (invertible), it is observed that

$$\begin{aligned} \mathrm{CRB} = \mathbf{J}^{-1}(\mathbf{Y}, \boldsymbol{\theta}) &= \left(\mathbf{H}\mathrm{Cov}^{-1}(\mathbf{N}+\mathbf{M})\mathbf{H}^T\right)^{-1} = \left(\mathbf{H}^{-1}\right)^T \mathrm{Cov}(\mathbf{N}+\mathbf{M})\mathbf{H}^{-1} \\ &= \left(\mathbf{H}^{-1}\right)^T \mathbf{\Sigma_N}\mathbf{H}^{-1} + \left(\mathbf{H}^{-1}\right)^T \mathbf{D_M}\mathbf{H}^{-1}, \end{aligned} \tag{6.35}$$

where the first part of the CRB, $(\mathbf{H}^{-1})^T \mathbf{\Sigma_N}\mathbf{H}^{-1}$ is a known quantity, and the second part $(\mathbf{H}^{-1})^T \mathbf{D_M}\mathbf{H}^{-1}$ will be subject to design while assessing the quality of the estimation. Similar to the previous discussion, CRB can be achieved in this case by employing the corresponding linear unbiased estimator which turns out simply to be a multiplication of the measurement vector with the inverse of the system matrix, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{y}) = (\mathbf{H}^{-1})^T\mathbf{y}$. Returning to two commonly used performance metrics introduced in Section 6.1, we next examine the closed-form solutions of the corresponding cost minimization problems.

6.3.1 Average Mean-Squared-Error

Due to the CRB, it is known that the average MSE while estimating the components of the parameter $\boldsymbol{\theta}$ is bounded from below as

$$\begin{aligned}\mathbb{E}_{\mathbf{Y}|\boldsymbol{\theta}}\left\{\left\|\hat{\boldsymbol{\theta}}(\mathbf{Y})-\boldsymbol{\theta}\right\|_2^2\right\} &\geq \operatorname{tr}\left\{\mathbf{J}^{-1}(\mathbf{Y},\boldsymbol{\theta})\right\} \\ &= \operatorname{tr}\left\{\left(\mathbf{H}^{-1}\right)^T \boldsymbol{\Sigma}_{\mathbf{N}}\mathbf{H}^{-1}\right\} + \operatorname{tr}\left\{\left(\mathbf{H}^{-1}\right)^T \mathbf{D}_{\mathbf{M}}\mathbf{H}^{-1}\right\}\ ,\end{aligned} \tag{6.36}$$

where the last equality follows from the linearity of the trace operator and the invertibility of $\mathbf{H}$. Since $\left(\mathbf{H}^{-1}\right)^T \boldsymbol{\Sigma}_{\mathbf{N}}\mathbf{H}^{-1}$ is known, let $t = \operatorname{tr}\left\{\left(\mathbf{H}^{-1}\right)^T \boldsymbol{\Sigma}_{\mathbf{N}}\mathbf{H}^{-1}\right\}$. When the aim is to minimize the measurement cost subject to a constraint on the lower bound for the average MSE (achievable in the case of Gaussian distributions), the optimization problem can be expressed similarly to (6.11) as follows:

$$\begin{aligned}\min_{\left\{\sigma_{m_i}^2\right\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1+\frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right) \\ \text{subject to} \quad & \operatorname{tr}\left\{\left(\mathbf{H}^{-1}\right)^T \mathbf{D}_{\mathbf{M}}\mathbf{H}^{-1}\right\} \leq \mathrm{E}-t\,,\end{aligned} \tag{6.37}$$

where E denotes a constraint for the overall average estimation error suggested by the CRB (achievable in this case), and t represents the unavoidable estimation error due to intrinsic system noise $\mathbf{N}$. Notice that for consistency, the design parameter E should be selected as $\mathrm{E} > t$.

From the independence of the measurement noise components, $\mathbf{D}_{\mathbf{M}} = \operatorname{diag}\{\sigma_{m_1}^2, \sigma_{m_2}^2, \ldots, \sigma_{m_K}^2\}$ is a diagonal covariance matrix with $\sigma_{m_i}^2 > 0$, $\forall i \in \{1, 2, \ldots, K\}$. In the view of this observation, it is possible to simplify the objective function further by defining $\mathbf{F} \triangleq \left(\mathbf{H}^{-1}\right)^T = [\mathbf{f}_1\ \mathbf{f}_2\ \ldots\ \mathbf{f}_K]$, where $\mathbf{f}_i$ represents the i^{th} row of the inverse of the system matrix $\mathbf{H}$. Let $f_i \triangleq \|\mathbf{f}_i\|_2^2$ denote the square of the Euclidean norm of the vector $\mathbf{f}_i$, that is, the sum of squares of the elements in $\mathbf{f}_i$. It is noted that f_i is always positive for invertible $\mathbf{H}$, and is constant for fixed $\mathbf{H}$. Then the optimization problem in (6.37) can be expressed as follows:

$$
\begin{aligned}
\min_{\{\sigma^2_{m_i}\}_{i=1}^K} \quad & \frac{1}{2}\sum_{i=1}^{K} \log_2\left(1+\frac{\sigma^2_{n_i}}{\sigma^2_{m_i}}\right) \\
\text{subject to} \quad & \sum_{i=1}^{K} f_i\,\sigma^2_{m_i} \;\le\; \mathrm{E}-t\,, \\
& \sigma^2_{m_i} \ge 0 \quad \forall\, i \in \{1,2,\ldots,K\}\;.
\end{aligned}
\tag{6.38}
$$

From (6.38), it is noted that the constraint function is linear in $\sigma^2_{m_i}$'s, the objective function is convex, and both functions are continuously differentiable which altogether indicate that Slater's condition holds. Therefore, Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality. Then, the optimal measurement noise variances can be calculated from

$$
\sigma^2_{m_i} = -\frac{\sigma^2_{n_i}}{2} + \sqrt{\frac{\sigma^4_{n_i}}{4} + \gamma\frac{\sigma^2_{n_i}}{f_i}}\;, \tag{6.39}
$$

where $\gamma > 0$ is obtained by substituting (6.39) into the average MSE constraint, that is $\sum_{i=1}^{K} f_i\,\sigma^2_{m_i} \;=\; \mathrm{E}-t$.

Special Case: When the inverse of the system matrix has normalized rows, i.e., $f_i = 1$, and the components of the system noise are independent zero-mean Gaussian random variables, the optimal measurement noise variances should satisfy $\sum_{i=1}^{K} \sigma^2_{m_i} = \mathrm{E} - \sum_{i=1}^{K} \sigma^2_{n_i}$. If identical system noise components are assumed as well, i.e., $\sigma^2_{n_i} = \sigma^2_n$, $i = 1,\ldots,K$, then the optimal solution results in $\sigma^2_{m_i} = \sigma^2_m$, $i = 1,\ldots,K$, where $\sigma^2_m = \mathrm{E}/K - \sigma^2_n$ is obtained from the average MSE constraint. The corresponding optimal cost is given by $(K/2)\log_2\left(\mathrm{E}/(\mathrm{E}-K\sigma^2_n)\right)$. This is an increasing function of K for fixed E. Furthermore, the derivatives of all orders with respect to K exist, and are positive for $K < \mathrm{E}/\sigma^2_n$. Therefore, estimating more parameters under an average error constraint based on the CRB requires even more accurate measurement devices with higher costs as long as $K < \mathrm{E}/\sigma^2_n$ is satisfied.

6.3.2 Shannon Information

Another measure of estimation accuracy that results in a closed form solution in the case of invertible system matrix $\mathbf{H}$ is the Shannon information criterion. Using this metric as the constraint function, we are effectively restricting the log volume of the η-confidence ellipsoid to stay below a predetermined value S. Using similar arguments to Section 6.1.2 and the invertibility of $\mathbf{H}$,

$$\begin{aligned}\log\,\det\left(\mathbf{H}\mathrm{Cov}^{-1}(\mathbf{N}+\mathbf{M})\mathbf{H}^T\right) &= \log\left(\det\mathbf{H}\cdot\det\left(\mathrm{Cov}^{-1}(\mathbf{N}+\mathbf{M})\right)\cdot\det\mathbf{H}^T\right)\\ &= 2\log\,|\det\mathbf{H}| - \sum_{i=1}^{K}\log\left(\sigma_{n_i}^2+\sigma_{m_i}^2\right)\,, \qquad (6.40)\end{aligned}$$

where the second equality follows the properties of the determinant and logarithm, i.e., $\det\mathbf{H} = \det\mathbf{H}^T$, $\det\left(\mathrm{Cov}^{-1}(\mathbf{N}+\mathbf{M})\right) = 1/\det\left(\mathrm{Cov}(\mathbf{N}+\mathbf{M})\right)$, and $\mathrm{Cov}(\mathbf{N}+\mathbf{M}) = \mathbf{D_N}+\mathbf{D_M} = \mathrm{diag}\{\sigma_{n_1}^2+\sigma_{m_1}^2\,,\,\sigma_{n_2}^2+\sigma_{n_2}^2\,,\,\ldots\,,\,\sigma_{n_K}^2+\sigma_{m_K}^2\}$ due to Gaussian distributed independent system and measurement noises with independent components. Since the system matrix $\mathbf{H}$ is known, let $\alpha \triangleq \log\,|\det\mathbf{H}|$. Under these conditions, the optimization problem in (6.17) can be stated as

$$\begin{aligned}\min_{\{\sigma_{m_i}^2\}_{i=1}^K} &\quad \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1+\frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right)\\ \text{subject to} &\quad \sum_{i=1}^{K}\log\left(\sigma_{n_i}^2+\sigma_{m_i}^2\right)\;\le\;2(\mathrm{S}+\alpha-\beta)\,, \qquad (6.41)\end{aligned}$$

where S and β are as defined in (6.17).

Notice that although the objective in (6.41) is a convex function of $\sigma_{m_i}^2$'s, the constraint is not a convex set. In fact, the constraint set is what is left after the convex set

$$\mathcal{C} = \left\{\boldsymbol{\sigma_m^2}\succeq 0\,:\,\sum_{i=1}^{K}\log\left(\sigma_{n_i}^2+\sigma_{m_i}^2\right)\;>\;2(\mathrm{S}+\alpha-\beta)\right\}$$

is subtracted from $\left\{\boldsymbol{\sigma_m^2}\succeq 0\right\}$. Since the global minimum of the unconstrained objective function is achieved for $\sigma_{\mathbf{m}}^{\mathbf{2}} = \boldsymbol{\infty}$ which is contained in set $\mathcal{C}$ and the objective function is convex, it is concluded that the minimum of the objective function has to occur at the boundary, i.e., $\sum_{i=1}^{K} \log\left(\sigma_{n_i}^2 + \sigma_{m_i}^2\right) = 2(\mathrm{S} + \alpha - \beta)$ must be satisfied [19]. Therefore, we can take the constraint as equality in (6.41). This is a standard optimization problem that can be solved using Lagrange multipliers. Hence, by defining $\varrho \triangleq 2(\mathrm{S} + \alpha - \beta)$, we can write the Lagrange functional as

$$J(\sigma_{m_1}^2, \ldots, \sigma_{m_K}^2) = \frac{1}{2} \sum_{i=1}^{K} \log_2 \left(1 + \frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right) + \lambda \left(\sum_{i=1}^{K} \log\left(\sigma_{n_i}^2 + \sigma_{m_i}^2\right) - \varrho \right), \tag{6.42}$$

and differentiating with respect to $\sigma_{m_i}^2$, we have the following assignment of the noise variances to the measurement devices

$$\sigma_{m_i}^2 = (\gamma^{1/K} - 1)\, \sigma_{n_i}^2 , \quad \text{where} \quad \gamma = \frac{2^{\varrho}}{\prod_{j=1}^{K} \sigma_{n_j}^2} . \tag{6.43}$$

For consistency, the design parameter S should be selected as $\varrho = 2(\mathrm{S} + \alpha - \beta) > \sum_{i=1}^{K} \log\left(\sigma_{n_i}^2\right)$ since the intrinsic system noise puts a lower bound on the minimum attainable volume of the confidence ellipsoid. Some properties of the obtained solution can be summarized as follows

- For given ϱ, K and $\sigma_{n_i}^2$'s, the minimum achievable cost is $(K/2)\log_2\left(\frac{\gamma^{1/K}}{\gamma^{1/K}-1}\right)$, where γ is computed as in (6.43).
- For a fixed value of K (available number of observations), relaxing the constraint on the volume of the η-confidence ellipsoid (increasing the value of ϱ) results in smaller measurement device costs with a limiting value of 0, as expected.
- If the observation variances are equal; that is, $\sigma_{n_i}^2 = \sigma_n^2$, $i = 1, \ldots, K$, employing identical measurement devices for all the observations; that is, $\sigma_{m_i}^2 = \sigma_m^2$, $i = 1, \ldots, K$, is the optimal strategy. From (6.43), the optimal value of the measurement noise variances is calculated as $\sigma_{m,\mathrm{opt}}^2 = e^{\varrho/K} - \sigma_n^2$, and the corresponding minimum total measurement cost is given as $\varrho/(2\log 2) - (K/2)\log_2\left(e^{\varrho/K} - \sigma_n^2\right)$ which is an increasing function of K

for $\varrho > K \log \sigma_n^2$. Intuitively, this result as well indicates that estimating more parameters under a fixed constraint on the volume of the ellipsoid containing the estimation errors requires a higher total measurement device cost.

6.4 Numerical Results

In this section, we present an example that illustrates several theoretical results developed in the previous section. To that aim, a discrete-time linear system as depicted in Figure 6.1 is considered

$$\mathbf{Y} = \mathbf{H}^T \boldsymbol{\theta} + \mathbf{N} + \mathbf{M}\,, \tag{6.44}$$

where $\boldsymbol{\theta}$ is a length-20 vector containing the unknown parameters to be estimated, $\mathbf{H}$ is a 20×100 system matrix with full row rank, the intrinsic system noise $\mathbf{N}$ and the measurement noise $\mathbf{M}$ are length-100 Gaussian distributed random vectors with independent components. The entries of the system matrix $\mathbf{H}$ are generated from a process of i.i.d. uniform random variables in the interval $[-0.1,\, 0.1]$. Also, the components of the system noise vector $\mathbf{N}$ are independently Gaussian distributed with zero mean, and it is assumed that their variances come from a uniform distribution defined in the interval $[0.05,\, 1]$. The implication of this assumption is that the observations at the output of the linear system possess uniformly varying degrees of accuracy. In other words, it is assured that observations corrupted by weak, moderate and strong levels of Gaussian noise are available with similar proportions for the estimation stage. In the following, we look into the problem of optimally assigning costs to measurement devices under various estimation accuracy constraints when the variances of the intrinsic system noise components are uniformly distributed as explained above. Note that our results obtained in the previous section are still valid for Gaussian system noise processes with arbitrary diagonal covariance matrices (i.e., the non-zero components of the diagonal covariance matrix need not be uniformly distributed as in this example). In obtaining the optimal solutions for the convex optimization problems stated above, *fmincon* method from MATLAB's Optimization Toolbox and the CVX software [73] are used.

6.4.1 Performance of Various Estimation Quality Metrics under Perfect System State Information

First, we investigate the cost assignment problem under perfect information on the system matrix and intrinsic noise variances. Recall that four different performance constraints are proposed for that purpose in Section 6.1. In the following four experiments, we analyze the behavior of the total measurement cost while each constraint metric is varied between its extreme values. The total cost is measured in bits by taking logarithms with respect to base 2. The constraint metric is expressed as the ratio of its current value to the value it attains for the limiting case when zero measurement noise variances are assumed. As an example, for average mean-squared-error criterion, the total measurement cost C will be tabulated versus $\mathrm{E}/\mathrm{tr}\left\{\left(\mathbf{H}\mathbf{D}_{\mathbf{N}}^{-1}\mathbf{H}^T\right)^{-1}\right\}$.

In addition to the optimal cost allocation scheme proposed in this study, we also consider two suboptimal cost allocation strategies:

Equal cost to all measurement devices: In this strategy, it is assumed that a single set of measurement devices with identical costs is employed for all observations so that $\mathrm{C}_i = \mathrm{C}$, $i = 1, 2, \ldots, K$. This, in turn, implies that the ratio of the measurement noise variance to the intrinsic system noise variance, $x \triangleq \sigma^2_{m_i}/\sigma^2_{n_i}$, is constant for all measurement devices. Then, the total cost can be expressed in terms of x as $\mathrm{C} = 0.5K \log_2(1 + 1/x)$, and similarly the FIM becomes $\mathbf{J}(\mathbf{Y}, \boldsymbol{\theta}) = \frac{\mathbf{H}\mathbf{D}_{\mathbf{N}}^{-1}\mathbf{H}^T}{x+1} = \frac{1}{x+1}\sum_{i=1}^{K} \frac{\mathbf{h}_i\mathbf{h}_i^T}{\sigma^2_{n_i}}$. Using this observation, the constraint functions provided for different performance metrics in the optimization problems (6.11), (6.17), (6.19), and (6.21) can be algebraically solved for equality to determine the value of x without applying any convex optimization techniques, and the corresponding measurement variances and cost assignments can be obtained.

Equal measurement noise variances: In this case, measurement devices are assumed to introduce random errors with equal noise variances, that is, $\sigma_{m_i}^2 = \sigma_m^2,\ i = 1, 2, \ldots, K$. In other words, all observations are assumed to be corrupted with identical noise processes, and the best measurement noise variance value that minimizes the overall measurement cost while satisfying the estimation accuracy constraint is selected. Accordingly, the objective function in the proposed optimization problems simplifies to $\mathrm{C} = 0.5 \sum_{i=1}^{K} \log_2 \left(1 + \sigma_{n_i}^2 / \sigma_m^2\right)$ and the FIM employed in the constraint functions takes the form $\mathbf{J}(\mathbf{Y}, \boldsymbol{\theta}) = \sum_{i=1}^{K} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\sigma_{n_i}^2 + \sigma_m^2}$. By substituting these expressions into the various optimization approaches provided in Section 6.1, these problems can be solved rapidly over a *single* parameter σ_m^2 using the tools of convex analysis, and the optimal cost allocations can be obtained for the case of equal measurement noise variances.

Average Mean-Squared-Error Criterion

In this experiment, we study the effects of the average MSE constraint on the total measurement device cost. Starting from the minimum achievable value for the average MSE due to intrinsic system noise (i.e., $\mathrm{tr}\{(\mathbf{H}\mathbf{D}_{\mathbf{N}}^{-1}\mathbf{H}^T)^{-1}\}$), we increase the constraint up to 100 times this minimal value, as depicted in Figure 6.2. Three curves are presented corresponding to the optimal cost allocation strategy and two suboptimal strategies, one employing equal cost and the other employing equal noise variance among the measurement devices. It is noted that the optimal strategy results in the minimum cost for all values of the MSE constraint as expected. Its performance is followed by the equal cost assignment

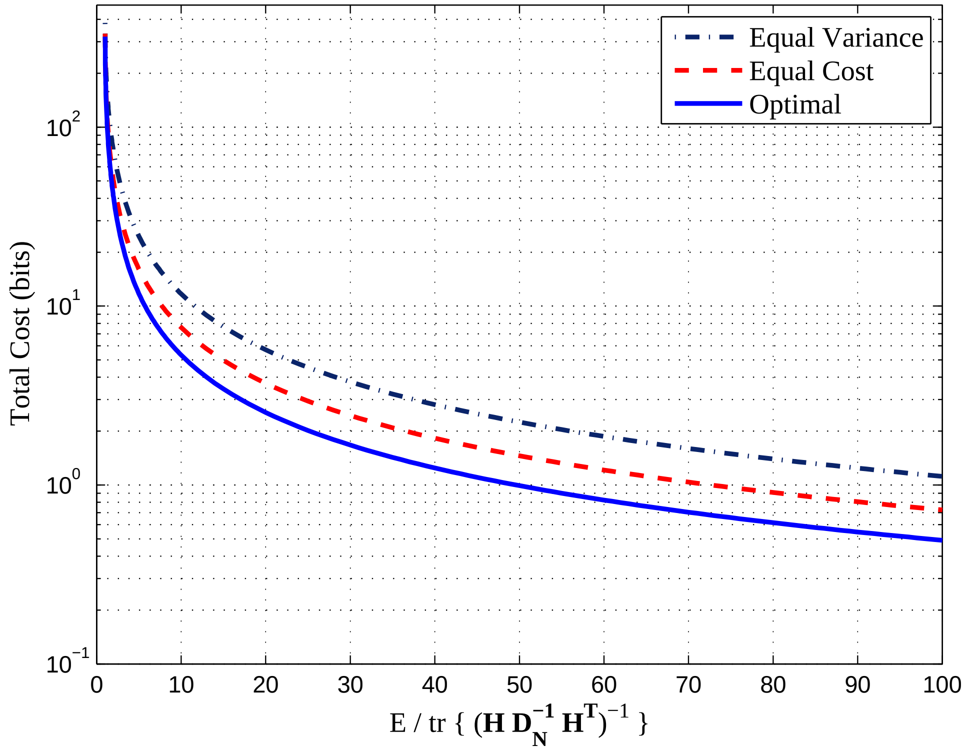

Figure 6.2: Total cost versus normalized average MSE constraint.

scheme, and the worst performing strategy is the one that assigns equal measurement noise variances to all the devices. When the average MSE criterion is stringent (for smaller values of E), all the strategies require increasingly more accurate measurements (hence higher costs) to satisfy the constraint. As the MSE constraint is relaxed (i.e., for larger values of E), the measurement costs of three different strategies start to drop down to zero but become less responsive as they move along.

Shannon Information Criterion

This experiment aims to discover the relationship between Shannon information constraint and total measurement device cost. Since the constraint is expressed as a 'greater than' inequality, we begin with the maximum attainable value of $\log\det(\mathbf{H}\mathbf{D}_{\mathbf{N}}^{-1}\mathbf{H}^T)$ and loosen the constraint by decreasing towards the negative

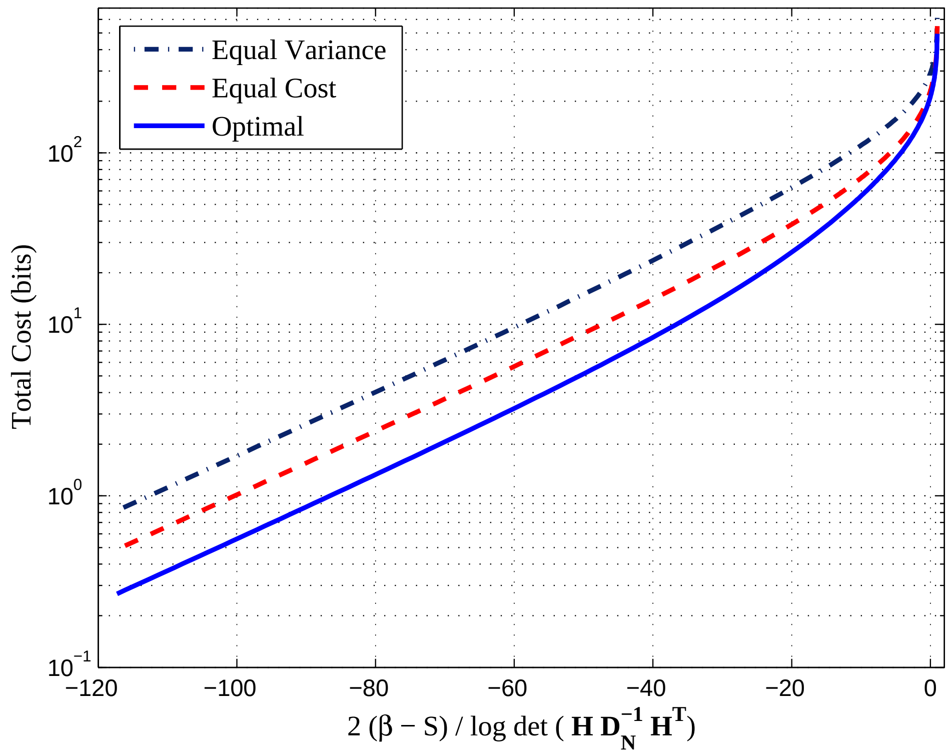

Figure 6.3: Total cost versus normalized Shannon information constraint.

multiples of this quantity as shown in Figure 6.3. When the constraint is very restrictive (corresponding to high values of $2(\beta - \mathrm{S})$), the differences among the performances of optimal and suboptimal strategies disappear. As the constraint is relaxed away from the maximum attainable value, it is observed that the decrease in the total cost is less responsive with respect to the average MSE. However, as the relaxation continues we see that the drop in the total cost for the Shannon information criterion maintains its pace for a longer time while the drop in the average MSE criterion seems to saturate. Again similar to the previous case, the performance of the optimal strategy is superior to the equal measurement device cost strategy, and the worst performance belongs to the equal measurement variance scheme.

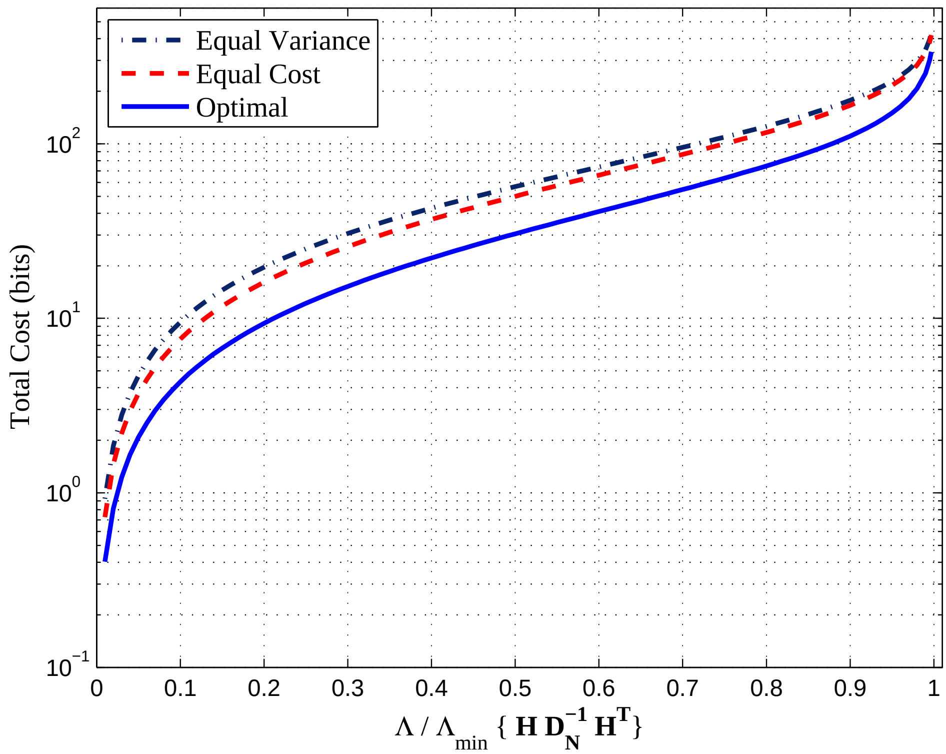

Figure 6.4: Total cost versus normalized worst-case error variance constraint.

Worst-Case Error Variance

In this experiment, we investigate the effects of the worst-case error variance criterion on the total measurement device cost under different cost allocation strategies. Similar conclusions to the previous experiments can be drawn by examining Figure 6.4.

Worst-Case Coordinate Error Variance

This experiment focuses on the relationship between the constraint on the largest diagonal entry of the CRB and the total measurement device costs achievable via different cost allocation strategies. The results are illustrated in Figure 6.5. It is noted that the plots depicted in Figure 6.2 embody a large degree of resemblance to those given in Figure 6.5. This similarity is anticipated and can be attributed

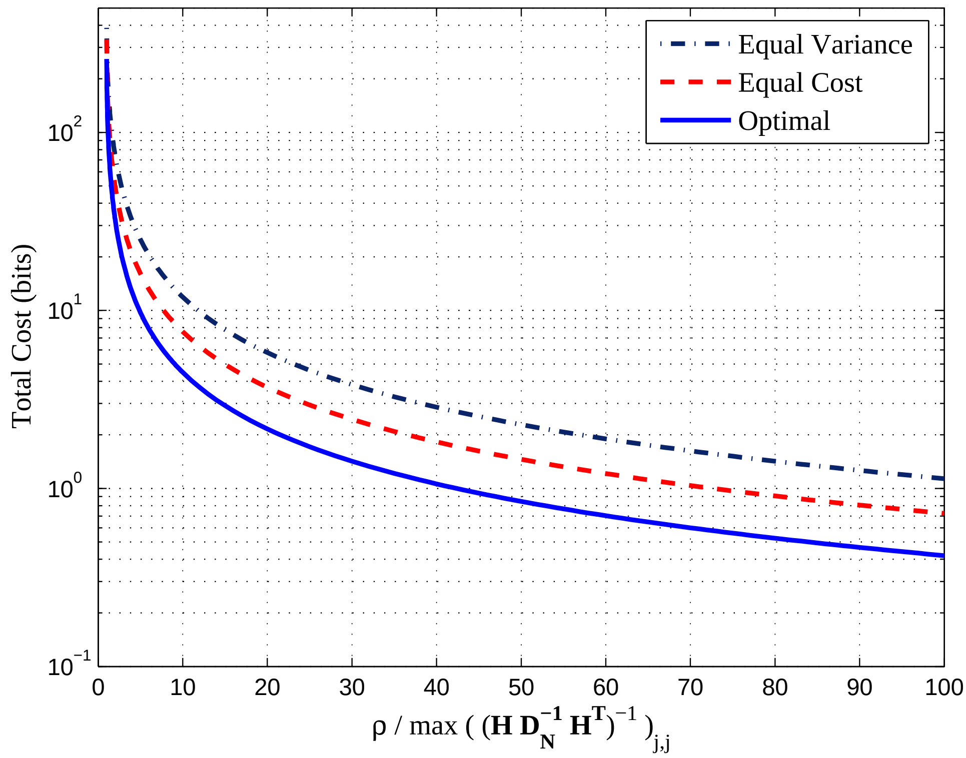

Figure 6.5: Total cost versus normalized worst-case coordinate error variance constraint.

to the fact that the former criterion puts a constraint on the average of the diagonal entries of the CRB whereas the latter places a similar constraint on their maximum.

Finally, we can stress a few more points. It is necessary that the intrinsic system noise variances and the system matrix are jointly evaluated to compute the optimal measurement noise variances and the corresponding cost allocations. In other words, in order to assign more cost to a specific observation, it is not sufficient to just know that the particular observation is reliable (i.e., has smaller variance) but we also need to know its intrinsic combinations with the other observations due to linear system matrix. Furthermore, the performance figures are quite useful in the sense that they provide the minimum cost necessary to obtain a desired level of estimation accuracy.

6.4.2 Performance Comparison of Estimation Quality Metrics under Scaling of the System Noise Variances

In this section, we devise a new experiment in order to jointly assess the performance of the proposed optimal cost assignment strategies under different estimation quality metrics. Using the same set of system noise variances employed in the previous experiments, we scale them with a factor c that varies inside the interval $[0.1, 1]$ with 0.01 increments. Specifically, $\hat{\sigma}^2_{n_i} = c\,\sigma^2_{n_i},\ i = 1, \ldots, K$, where $c \in \{0.1 : 0.01 : 1\}$. For such a comparison to make sense, the constraints on the estimation quality metrics are selected so that the optimal total measurement costs returned by the various approaches are equal for a certain value of the scale parameter c. Then, using the same value as the constraint, we evaluate the performance of each optimal cost allocation strategy for the rest of the scale parameter values. To that aim, we construct two examples. In the first one, the performances of the optimal schemes under four different performance metrics are equated for $c = 0.5$, producing an optimal total cost of 40.11. The corresponding constraint function values are $\mathrm{E} = 23.1371$ for the average MSE criterion, $2(\beta - \mathrm{S}) = 1.9389$ for Shannon Information criterion, $\Lambda = 0.4364$ for the worst-case error variance criterion and $\varrho = 1.3646$ for the worst-case coordinate error variance criterion. The results are illustrated in Figure 6.6. Intuitively, as the intrinsic system noise variances are increased, more reliable measurements (higher costs) are required to satisfy the same level of accuracy. Comparing the performances in Figure 6.6, where all the costs are equated for $c = 0.5$, we observe that the average MSE criterion results in the least (i.e., the best) optimal cost score for increasing values of the scale parameter c. Its performance is followed by the Shannon information criterion, next by the worst-case coordinate error variance criterion, and finally by the worst-case error variance criterion. In other words, the effects of increasing system noise variances are much more pronounced for the

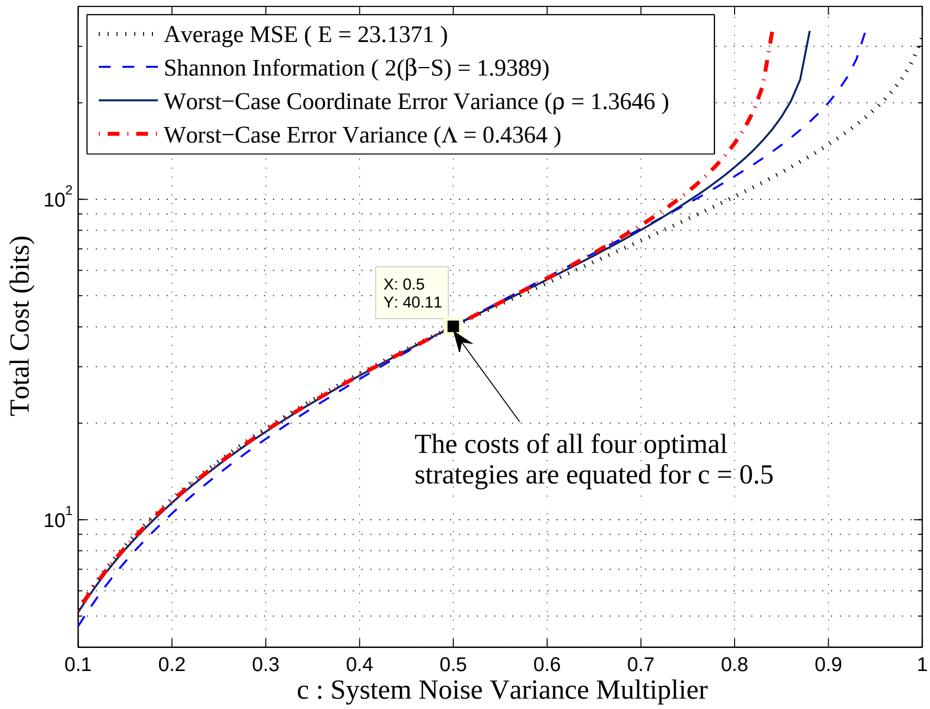

Figure 6.6: The performance of various optimal cost allocation strategies under scaling of the system noise variances. All costs are equal for $c = 0.5$.

worst-case error variance criterion, which operates by setting a constraint on the minimum eigenvalue of the FIM, than the remaining criteria. If the noise scale parameter c is decreased below 0.5, it is observed that the Shannon information criterion produces the lowest measurement cost followed by the worst-case coordinate error variance criterion, worst-case error variance criterion, and finally average MSE criterion in the order of increasing costs. It is noted that, except for the average MSE criterion, the performance of the remaining three metrics stays in the same order for values of c above and below 0.5. Another important observation is that among the four estimation quality metrics, the performance of the MSE criterion is the one that is least susceptible to changes in the system noise variance. That is, as c is increased beyond 0.5 and decreased below 0.5, the least varying performance metric corresponds to the average MSE criterion. Therefore, in applications where the level of the system noise variance are likely to fluctuate around a nominal value and a predetermined value of the estimation accuracy has to be satisfied, the average MSE criterion provides the most robust alternative in terms of the measurement device selection. However, even in this case, a small change of order 0.01 in the value of the scale parameter disturbs the total cost by more than 1 bit for the average MSE metric.

In the second example, the performances of the estimation quality metrics are equated for $c = 1$, resulting in a total cost score of 320.8. We employ the same constraint value ($\mathrm{E} = 23.1371$) for the average MSE criterion, and the adjustments are applied to the remaining metrics. The corresponding constraint function values are calculated as $2(\beta - \mathrm{S}) = 0.66$ for the Shannon Information criterion, $\Lambda = 0.3664$ for the worst-case error variance criterion, and $\varrho = 1.5519$ for the worst-case coordinate error variance criterion. The results are illustrated in Figure 6.7. In accordance with the observations for high values of c in the previous example, the worst-case error variance metric quickly responds to the drop in the level of the system noise variance values. Hence, the lowest cost is provided by the worst-case error variance criterion for $c < 1$. On the other hand, the optimal cost value for the average MSE criterion exhibits the slowest descent for decreasing values of c. Also noted from the figure is that the performance curve for the Shannon information criterion down-crosses the curve corresponding to the worst-case coordinate error variance criterion at around $c = 0.21$.

6.4.3 The Relationship between the Number of Effective Measurements and the Quality of Estimation under Scaling of the System Noise Variances

In this experiment we discuss the relationship between the number of effective measurements K_{eff} and various estimation quality metrics under scaling of the system noise variances. A measurement is assessed as effective whenever the cost of that measurement exceeds a certain fraction of the optimal value of the

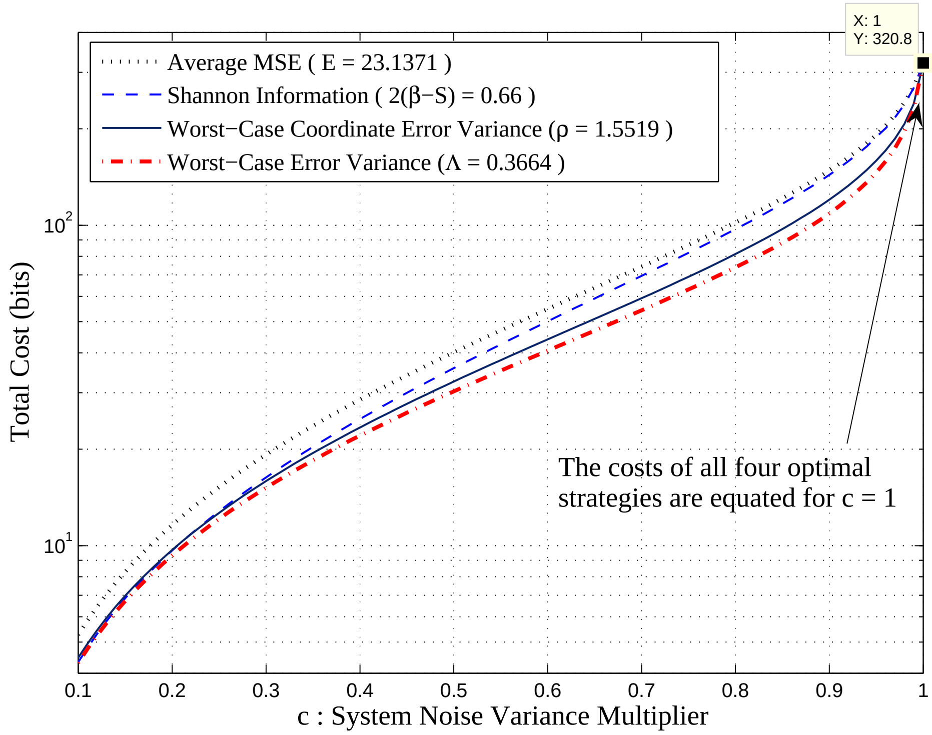

Figure 6.7: The performance of various optimal cost allocation strategies under scaling of the system noise variances. All costs are equal for $c = 1$.

total measurement cost. More specifically, we require that $\mathrm{C}_i > p\,(\mathrm{C}/K)$ where K represents the total number of measurements. With this construction, it is assured that the total cost of the effective measurements is greater than $(1-p)\mathrm{C}$, from which a suitable value for p can be determined [46]. For small values of p, we can safely assume that the remaining measurements do not cause a significant change on the total cost or provide any significant contribution to the estimation accuracy. Similar to the study in [46], $p = 0.125$ is selected. The same constraint values as in Figure 6.7 are employed for the estimation accuracy metrics. Since the performances of all four estimation accuracy criteria are fixed to a high cost score of 320.8 for $c = 1$, it is noted from Figure 6.8 that most of the observations are utilized at this value of the scale parameter in order to satisfy the strict constraints. As the average system noise power is reduced by assigning smaller values to the system noise variance multiplier c, the number of effective measurements decreases for all the four cases in accordance with decreasing measurement costs.

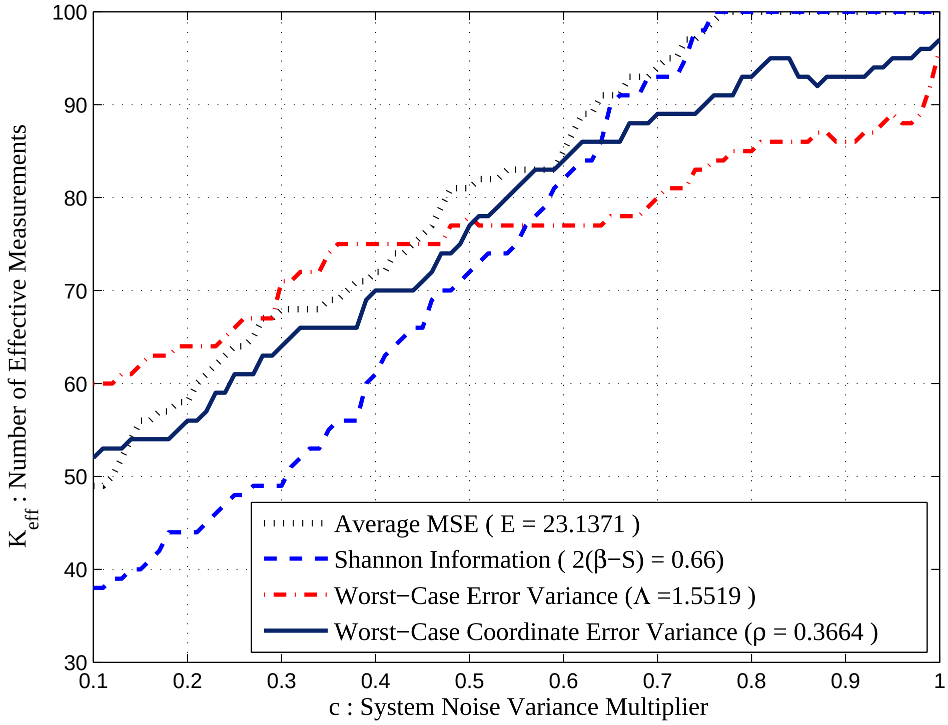

Figure 6.8: Number of effective measurements under scaling of the system noise variances for various estimation accuracy metrics.

In other words, lower noise variances result in looser constraints which can be achieved by using fewer number of high resolution (costly) measurements. For small values of c, the worst-case error variance requires the largest number of measurements followed by the average MSE criterion, the worst-case coordinate error variance criterion, and finally the Shannon information criterion. For higher values of c, the situation is reversed apart from the average MSE criterion which requires the largest number of effective measurements. When $c \leq 0.56$, a relatively small number of accurate measurements is sufficient to conduct a reliable estimation using the Shannon information criterion with respect to the remaining criteria.

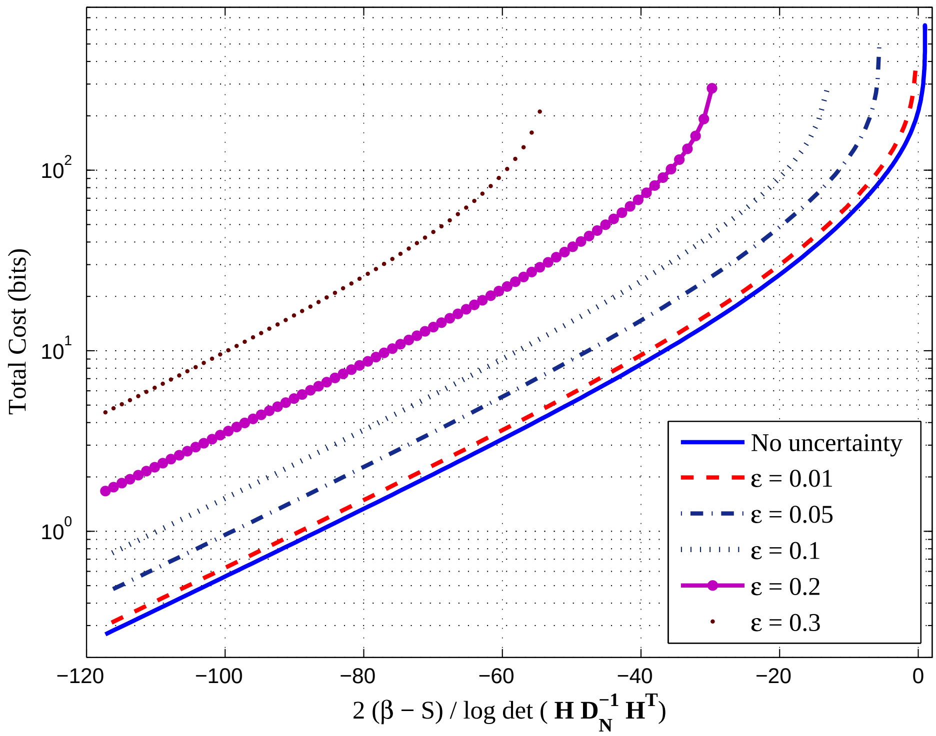

Figure 6.9: Effects of system matrix uncertainty on the total measurement cost for Shannon information criterion.

6.4.4 Effects of System Matrix Uncertainty

So far, we have assumed that the system matrix is known perfectly at the measurement stage. In this experiment, we consider the case in which the measurement system can only have partial knowledge about the system matrix according to the specific uncertainty model introduced in Section 6.2. That is, the system matrix is represented as the sum of a known matrix plus a random disturbance matrix $\mathbf{H} \in \mathcal{H} = \left\{\bar{\mathbf{H}} + \mathbf{\Delta} \,:\, \left\|\mathbf{\Delta}^T\right\|_2 \leq \epsilon\right\}$, where the degree of uncertainty is controlled with the spectral norm of the disturbance matrix $\mathbf{\Delta}$. Below, we present the results concerning the effects of system uncertainty on the optimal cost allocation problem for the Shannon information and the worst-case error variance criteria in Figure 6.9 and Figure 6.10, respectively. For both cases, it is observed that the total cost increases as the amount of uncertainty in the system matrix increases for a given value of the constraint. The increase in the system matrix

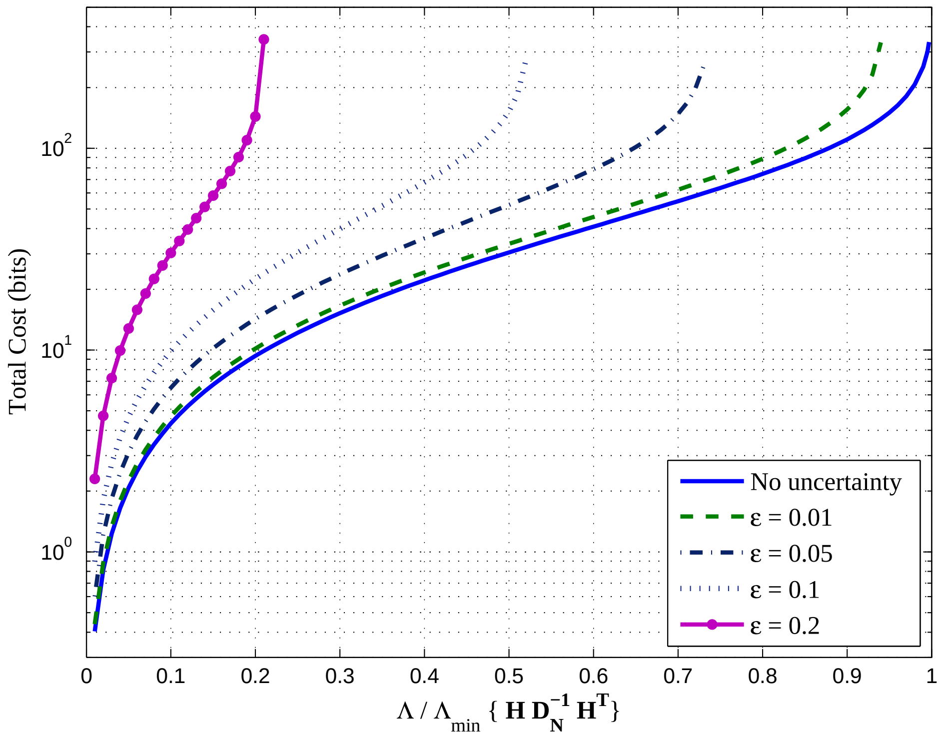

Figure 6.10: Effects of system matrix uncertainty on the total measurement cost for worst-case error variance criterion.

uncertainty also leads to smaller values of the maximum attainable estimation accuracy measures (the asymptotes where the total cost increases unboundedly).

6.5 Extensions to Bayesian Framework

In Section 6.1, parameter $\boldsymbol{\theta}$ is modeled as a deterministic unknown parameter. Whenever prior information is available about the distribution of the unknown parameter, this additional information can be utilized at the estimation stage. As a result, a more refined metric to assess the quality of the estimator performance is employed which is commonly known as the Bayesian CRB (BCRB) and expressed as follows:

$$\mathbb{E}\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^T\right\} \geq (\mathbf{J_D}+\mathbf{J_P})^{-1} \triangleq \mathrm{BCRB}\,, \tag{6.45}$$

where $\mathbf{J_D}$ represents data information matrix and $\mathbf{J_P}$ represents prior information matrix, whose elements are [23]

$$\begin{aligned} J_{D_{ij}} &= -\mathbb{E}_{\mathbf{Y},\boldsymbol{\theta}}\left\{\frac{\partial^2 \log p_{\mathbf{Y}}^{\boldsymbol{\theta}}(\mathbf{Y})}{\partial\theta_i\partial\theta_j}\right\} = \mathbb{E}_{\boldsymbol{\theta}}\left\{\mathbf{J}(\mathbf{Y},\boldsymbol{\theta})\right\} \text{ and} \\ J_{P_{ij}} &= -\mathbb{E}_{\boldsymbol{\theta}}\left\{\frac{\partial^2 \log w(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right\}\,, \end{aligned} \tag{6.46}$$

where $\mathbf{J}(\mathbf{Y},\boldsymbol{\theta})$ is the standard Fisher information matrix defined in (6.3).

When the prior probability of the parameter is Gaussian with $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma_\theta})$, under the same assumptions regarding the independence of $\mathbf{N} \sim \mathcal{N}(\mathbf{0},\mathbf{D_N})$ and $\mathbf{M} \sim \mathcal{N}(\mathbf{0},\mathbf{D_M})$, the BCRB for the linear system given in Figure 6.1 can be obtained as

$$\mathrm{BCRB} = \left(\sum_{i=1}^{K} \frac{1}{\sigma_{n_i}^2+\sigma_{m_i}^2}\,\mathbf{h}_i\mathbf{h}_i^T + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\,. \tag{6.47}$$

Correspondingly, the total cost function should be restated to incorporate the change in the variance of the input to each measurement noise device as follows:

$$\mathrm{C} = \sum_{i=1}^{K}\mathrm{C}_i = \sum_{i=1}^{K}\frac{1}{2}\log_2\left(1+\frac{\sigma_{x_i}^2}{\sigma_{m_i}^2}\right)\,, \tag{6.48}$$

where $\sigma_{x_i}^2$ is the i^{th} diagonal entry of the observation covariance matrix $\mathrm{Cov}(\mathbf{X}) = \mathbf{H}^T\boldsymbol{\Sigma_\theta}\mathbf{H}+\mathbf{D_N}$.

Based on these expressions, all the proposed cost minimization formulations in Section 6.1 can be modified accordingly to obtain the optimal cost assignment strategies in the presence of prior information. Specifically, the CRB is replaced with the BCRB, and the cost function stated in (6.48) is substituted as the objective function inside the optimization problems given in (6.14), (6.18), (6.20), and (6.22). However, the modified optimization problems are not necessarily convex. It is also noted that the problem formulation constructed by employing the LMMSE estimator in [46] is equivalent to the dual of the Bayesian estimation case under the average MSE criterion given in (6.11) when Gaussian priors are assumed.

6.6 Concluding Remarks

In this chapter, we have studied the measurement cost minimization problem for a linear system in the presence of Gaussian noise based on the measurement device model introduced in [46]. By considering the non-random parameter estimation case, novel convex optimization problems have been obtained under various estimation accuracy constraints [51, 52]. Uncertainty in the system matrix has been modeled both under general terms and by using a specific uncertainty model. It has been indicated that the convexity properties of the proposed optimization problems are preserved under uncertainty. When the system matrix is invertible, closed form expressions have been presented for two different estimation accuracy metrics which enable a quick assessment of the corresponding cost allocation strategies analytically or via simpler numerical techniques. It has been shown that the prior information can be incorporated into the optimization problems but the resulting problems need no longer be convex. Through numerical examples, the relationships among various criteria have been analyzed in depth.

7

Conclusions and Future Work

In this dissertation, we have derived optimal stochastic procedures for signal detection and estimation problems under certain inequality constraints that arise from practical considerations. More specifically, in the first three chapters the emphasis has been on enhancing the performance of average power constrained communications systems by employing stochastic signaling at the transmitter and/or detector randomization at the receiver. In the first chapter, we have shown that for M-ary communications the average probability of error is minimized by randomizing between at most two MAP detectors corresponding to two deterministic signal vectors. The performance improvements are much more evident for multimodal noise distributions as suggested by various numerical examples. In the second chapter, we have switched to the Neyman-Pearson framework and considered the well-known problem of detecting the presence of a target signal immersed in additive Gaussian noise. Through a rigorous treatment, we have proved that the probability of detection is a concave function of the transmit signal power when the false alarm probability is larger than $Q(2) \approx 0.02275$. Evidently, this condition renders a performance improvement via stochastic signaling impossible under an average transmit power constraint. However, for false alarm probabilities smaller than $Q(2)$ as is usually the case in practice, the detection probability is concave, convex and finally concave for increasing values of the transmit power. By additional analysis, we have determined the conditions on peak and average power constraints so that the detection probability can be increased via time-sharing between two non-zero power levels. Furthermore, the optimal transmit signal power distribution is stated, and a more practical power allocation scheme that employs on-off signaling is demonstrated to produce near-optimal performance. Next, the dual problem for a power constrained Gaussian jammer attacking a smart receiver that can adapt its decision threshold based on the received jamming power is considered. Even in this case, it is shown that there exists a critical power level up to which on-off jamming can be employed to degrade the detection performance beyond that can be achieved by the constant power jamming scenario. Motivated by the results, we have studied stochastic signaling and detector randomization for the Neyman-Pearson detection over channels with arbitrary noise distributions. For the case of a single detector at the receiver, the detection probability under an average transmit power constraint is maximized by randomizing between at most two signal values for the on-signal and using the corresponding NP-type likelihood ratio test at the receiver. When multiple detectors are available at the receiver, randomization among no more than three NP decision rules corresponding to three deterministic signal vectors is sufficient to attain the optimal performance. It is imperative to notice that in all the cases discussed so far, optimization over an infinite set of functions is reduced down to that over a few variables involving randomization factors, signal/power values, and additionally decision thresholds in the case of NP detectors.

In the last two chapters, we have analyzed the relationship between estimation accuracy and measurement cost based on a recently proposed measurement device cost model. In order to provide a generic framework that is independent of any specific estimator structure, average Fisher information is utilized as the estimation accuracy metric. In the fourth chapter, an optimization problem is formulated to calculate the optimal costs of measurement devices that maximize the average Fisher information for arbitrary observation and measurement statistics. Closed form solutions are given in the case of Gaussian observations and measurement noises. In the fifth chapter, the subject is treated in more detail for the non-random parameter estimation case by assuming a linear system model subject to additive Gaussian noise. Novel convex optimization problems are obtained under various estimation accuracy constraints that depend on different manipulations of the Fisher information matrix. Closed form solutions are provided for two of these problems when the system matrix is invertible. It is shown that a certain form of uncertainty in the system matrix also leads to convex optimization problems. When extended to the Bayesian estimation theory by assuming Gaussian priors, it is observed that the resulting optimization problems are no longer convex.

For the first part of the dissertation, a future work is to investigate how the optimal strategy of the transmitter-receiver pair changes with the jammer's power randomization/sharing. Equilibrium conditions can be sought in a game-theoretic setting by allowing both parties to transmit stochastically while the receiver can randomize among multiple detectors. In addition, the convexity properties of the outage probability and outage capacity with respect to the transmit power can be studied for fading channels to determine whether practical improvements can be obtained via power randomization. Another direction is to assess the effects of stochastic signaling in decreasing the error probability for communications systems with relays.

Bibliography

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

[2] M. Azizoglu, "Convexity properties in binary detection problems," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1316–1321, July 1996.

[3] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni, "Stochastic resonance," *Reviews of Modern Physics*, vol. 70, no. 1, pp. 223–287, Jan. 1998.

[4] S. M. Kay, "Can detectability be improved by adding noise?" *IEEE Signal Processing Letters*, vol. 7, no. 1, pp. 8–10, Jan. 2000.

[5] G. P. Harmer, B. R. Davis, and D. Abbott, "A review of stochastic resonance: Circuits and measurement," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 2, pp. 299–309, Apr. 2002.

[6] S. Zozor and P.-O. Amblard, "Stochastic resonance in locally optimal detectors," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3177–3181, Dec. 2003.

[7] 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.

[8] S. M. Kay, J. H. Michels, H. Chen, and P. K. Varshney, “Reducing probability of decision error using stochastic resonance,” *IEEE Signal Processing Letters*, vol. 13, no. 11, pp. 695–698, Nov. 2006.

[9] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, “Theory of the stochastic resonance effect in signal detection: Part I–Fixed detectors,” *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3172–3184, July 2007.

[10] ——, “Theory of the stochastic resonance effect in signal detection: Part II–Variable detectors,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5031–5041, Oct. 2007.

[11] H. Chen, P. K. Varshney, and J. H. Michels, “Noise enhanced parameter estimation,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5074–5081, Oct. 2008.

[12] H. Chen, P. K. Varshney, S. Kay, and J. H. Michels, “Noise enhanced nonparametric detection,” *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 499–506, Feb. 2009.

[13] H. Chen, P. K. Varshney, and J. H. Michels, “Improving sequential detection performance via stochastic resonance,” *IEEE Signal Processing Letters*, vol. 15, pp. 685–688, Dec. 2008.

[14] S. M. Kay, “Noise enhanced detection as a special case of randomization,” *IEEE Signal Processing Letters*, vol. 15, pp. 709–712, 2008.

[15] A. Patel and B. Kosko, “Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection,” *IEEE Transactions on Signal Processing*, vol. 57, no. 5, pp. 1655–1669, May 2009.

[16] S. Bayram, S. Gezici, and H. V. Poor, “Noise enhanced hypothesis-testing in the restricted Bayesian framework,” *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 3972–3989, Aug. 2010.

[17] C. Goken, S. Gezici, and O. Arikan, "Optimal signaling and detector design for power-constrained binary communications systems over non-Gaussian channels," *IEEE Communications Letters*, vol. 14, no. 2, pp. 100–102, Feb. 2010.

[18] S. Loyka, V. Kostina, and F. Gagnon, "Error rates of the maximum-likelihood detector for arbitrary constellations: Convex/concave behavior and applications," *IEEE Transactions on Information Theory*, vol. 56, pp. 1948–1960, Apr. 2010.

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

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

[21] R. G. Gallager, *Principles of Digital Communication.* Cambridge, UK: Cambridge University Press, 2008.

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

[23] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Part I*, 2nd ed. New York, NY: John Wiley & Sons, Inc., 2001.

[24] I. Korn, J. P. Fonseka, and S. Xing, "Optimal binary communication with nonequal probabilities," *IEEE Transactions on Communications*, vol. 51, no. 9, pp. 1435–1438, Sep. 2003.

[25] C. Goken, S. Gezici, and O. Arikan, "Optimal stochastic signaling for power-constrained binary communications systems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3650 –3661, Dec. 2010.

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

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

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

[29] 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.

[30] S. Shamai and S. Verdu, "Worst-case power-constrained noise for binary-input channels," *IEEE Transactions on Information Theory*, vol. 38, pp. 1494–1511, Sep. 1992.

[31] B. Dulek and S. Gezici, "Detector randomization and stochastic signaling for minimum probability of error receivers," *IEEE Transactions on Communications*, vol. 60, no. 4, pp. 923–928, April 2012.

[32] ——, "Convexity properties of detection probability under additive Gaussian noise: Optimal signaling and jamming strategies," *IEEE Transactions on Signal Processing*, (Submitted on Feb. 2012).

[33] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications*. Rockville, MD: Computer Science Press, 1985, vol. 1.

[34] B. Dulek and S. Gezici, "Optimal signaling and detector design for power constrained on-off keying systems in Neyman-Pearson framework," in *Proc. 16th IEEE Workshop on Statistical Signal Processing (SSP 2011)*, Nice, France, June 2011, pp. 93–96.

[35] ——, "Optimal stochastic signal design and detector randomization in the Neyman-Pearson framework," in *37th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012)*, Kyoto, Japan, March 2012.

[36] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Energy-efficient detection in sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 693–702, April 2005.

[37] J. Li and G. AlRegib, "Rate-constrained distributed estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1634–1643, May 2007.

[38] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, March 2006.

[39] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, Feb. 2006.

[40] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.

[41] H. Zhang, J. Moura, and B. Krogh, "Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2383–2395, June 2009.

[42] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2649–2661, July 2008.

[43] T. C. Aysal and K. E. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1398–1410, April 2008.

[44] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 451–462, 2009.

[45] M. Lazaro, M. Sanchez-Fernandez, and A. Artes-Rodriguez, "Optimal sensor selection in binary heterogeneous sensor networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1577–1587, April 2009.

[46] A. Ozcelikkale, H. M. Ozaktas, and E. Arikan, "Signal recovery with cost-constrained measurements," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3607–3617, 2010.

[47] R. V. L. Hartley, "Transmission of information," *Bell System Technical Journal*, vol. 7, pp. 535–563, 1928.

[48] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423,623–656, July, October 1948.

[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 2006.

[50] B. Dulek and S. Gezici, "Average Fisher information maximisation in presence of cost-constrained measurements," *IET Electronics Letters*, vol. 47, no. 11, pp. 654–656, May 2011.

[51] ——, "Cost minimization of measurement devices under estimation accuracy constraints in the presence of gaussian noise," *Digital Signal Processing*, vol. 22, no. 5, pp. 828–840, Sept. 2012.

[52] ——, "A confidence ellipsoid approach for measurement cost minimization under Gaussian noise," in *13th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2012)*, Cesme, Turkey, June 2012.

[53] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, New Jersey: Prentice Hall, 1993.

[54] M. H. Hayes, *Statistical Digital Signal Processing and Modeling.* John Wiley & Sons, 1996.

[55] C. Goken, S. Gezici, and O. Arikan, "On the optimality of stochastic signaling under an average power constraint," in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing*, Monticello,Illinois, Sep. 2010.

[56] S. Bayram and S. Gezici, "On the improvability and nonimprovability of detection via additional independent noise," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 1001–1004, Nov. 2009.

[57] T. G. Kolda, R. M. Lewis, and V. Torczon, "A generating set direct search augmented Lagrangian algorithm for optimization with a combination of general and linear constraints," Sandia National Laboratories, Albuquerque, NM and Livermore, CA, Tech. Rep. SAND2006-5315, Aug. 2006.

[58] A. R. Conn, N. Gould, and P. L. Toint, "A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds," *Mathematics of Computation*, vol. 66, pp. 261–288, January 1997.

[59] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Marti, "Scatter search and local nlp solvers: A multistart framework for global optimization," *INFORMS Journal on Computing*, vol. 19, no. 3, pp. 328–340, 2007.

[60] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Singapore: Athena Scientific, 1995.

[61] R. P. Brent, *Algorithms for Minimization without Derivatives.* Englewood Cliffs, NJ: Prentice-Hall, 1973.

[62] T. Erseghe, V. Cellini, and G. Dona, "On UWB impulse radio receivers derived by modeling MAI as a Gaussian mixture process," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2388–2396, June 2008.

[63] R. Storn and K. Price, "Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341–359, Dec. 1997.

[64] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1246–1250, May 1998.

[65] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1501–1518, 1991.

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

[67] Y. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Transactions Signal Processing*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[68] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," *IEEE Transactions Signal Processing*, vol. 53, no. 8, pp. 3172–3182, Aug. 2005.

[69] A. Das and D. Kempe, "Sensor selection for minimizing worst-case prediction error," in *International Conference on Information Processing in Sensor Networks (IPSN'08)*, April 2008, pp. 97–108.

[70] R. Hettich and K. Kortanek, "Semi-infinite programming: Theory, methods, and applications," *SIAM Review*, vol. 35, no. 5, pp. 380–429, 1993.

[71] A. Mutapcic and S. Boyd, "Cutting-set methods for robust convex optimization with pessimizing oracles," *Optimization Methods and Software*, vol. 24, no. 3, pp. 381–406, 2009.

[72] Z.-Q. Luo, J. F. Sturm, and S. Zhang, "Multivariate nonnegative quadratic mappings," *SIAM Journal on Optimization*, vol. 14, pp. 1140–1162, 2002.

[73] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, April 2011.