

OPTIMAL TIME SHARING STRATEGIES FOR PARAMETER ESTIMATION AND CHANNEL SWITCHING PROBLEMS

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

By
Hamza Soğancı
December, 2014

OPTIMAL TIME SHARING STRATEGIES FOR PARAMETER
ESTIMATION AND CHANNEL SWITCHING PROBLEMS

By Hamza Soğancı

December, 2014

We certify that we have read this thesis and that in our opinion it is fully adequate,
in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Assoc. Prof. Dr. Sinan Gezici(Advisor)

Prof. Dr. Orhan Arıkan

Assoc. Prof. Dr. Selim Aksoy

Prof. Dr. A. Enis Çetin

Assoc. Prof. Dr. Ali Cafer Gürbüz

Approved for the Graduate School of Engineering and Science:

Prof. Dr. Levent Onural
Director of the Graduate School

ABSTRACT

OPTIMAL TIME SHARING STRATEGIES FOR PARAMETER ESTIMATION AND CHANNEL SWITCHING PROBLEMS

Hamza Soğancı

Ph.D. in Electrical and Electronics Engineering

Advisor: Assoc. Prof. Dr. Sinan Gezici

December, 2014

Time sharing (randomization) can offer considerable amount of performance improvement in various detection and estimation problems and communication systems. In the first three chapters of this dissertation, time sharing among different signal levels is considered for parametric estimation problems. In the final chapter, time sharing among different channels is investigated for an average power constrained communication system. In the first chapter, the aim is to improve the performance of a single fixed estimator by the optimal stochastic design of signal values corresponding to parameters. It is obtained that the optimal parameter design corresponds to time sharing between at most two different signal values. In the second chapter, the problem in the first chapter is generalized to a scenario where there are multiple parameters and multiple estimators. In this scenario, two different cost functions are considered. The first cost function is the total risk of all the estimators. The optimal solution for this case is time sharing between at most two different signal values. The second cost function is the maximum risk of all the estimators. For this case, it is shown that the optimal parameter design is time sharing among at most three different signal values. In the third chapter, the linear minimum mean squared error (LMMSE) estimator is considered. It is observed that time sharing is not needed for the LMMSE estimator, but still the performance can be improved by modifying the signal level. In the final chapter, the optimal channel switching problem is studied for Gaussian channels, and the optimal channel switching strategy is determined in the presence of average power and average cost constraints. It is shown that the optimal channel switching strategy is to switch among at most three channels.

Keywords: Time sharing, randomization, parameter estimation, stochastic parameter design, Bayes risk, LMMSE estimator, channel switching, Gaussian channel.

ÖZET

PARAMETRE KESTİRİM VE KANAL ATLAMA PROBLEMLERİ İÇİN EN İYİ ZAMAN PAYLAŞIMI STRATEJİLERİ

Hamza Soğancı

Elektrik-Elektronik Mühendisliği, Doktora

Tez Danışmanı: Doç. Dr. Sinan Gezici

Aralık, 2014

Zaman paylaşımı (rastgeleleştirme), sezim ve kestirim problemlerinde ve iletişim sistemlerinde başarımları önemli miktarda artırabilmektedir. Bu tezin ilk üç kısmında, kestirim problemleri için farklı sinyal seviyeleri arasında rastgeleleştirme ele alınmaktadır. Son kısımda ise, ortalama güç kısıtlı bir iletişim sistemi için farklı iletişim kanalları arasında zaman paylaşımı incelenmektedir. İlk kısımdaki amaç, parametrelere karşılık gelen sinyal değerlerinin optimal stokastik tasarımı yoluyla sabit tek bir kestiricinin performansını geliştirmektir. En iyi parametre tasarımının, en fazla iki sinyal değeri arasında zaman paylaşımına karşılık geldiği elde edilmektedir. İkinci kısımda, ilk kısımdaki problem çoklu parametre ve çoklu kestiricilerin bulunduğu bir senaryoya genelleştirilmektedir. Bu senaryoda iki farklı maliyet fonksiyonu ele alınmaktadır. İlk maliyet fonksiyonu, tüm kestiricilerin risklerinin toplamıdır. Bu durum için en iyi çözüm, en fazla iki sinyal değeri arasında zaman paylaşımıdır. İkinci maliyet fonksiyonu ise kestiricilerin riskleri arasında en yüksek olanıdır. Bu durum için, en iyi parametre tasarımının en fazla üç sinyal değeri arasında zaman paylaşımı olduğu gösterilmektedir. Üçüncü kısımda doğrusal en küçük ortalama karesel hata kestiricisi (DEKOKH) ele alınmaktadır. DEKOKH kestiricisi için zaman paylaşımına ihtiyaç olmadığı gözlemlenmektedir. Ancak yine de DEKOKH kestiricisinin başarımları sinyal seviyesinin değiştirilmesiyle artırılabilir. Son kısımda, Gauss kanallar için en iyi kanal değiştirme problemi çalışmakta ve en iyi kanal değiştirme stratejisi, ortalama güç ve ortalama maliyet kısıtları altında belirlenmektedir. En iyi kanal değiştirme stratejisinin, en fazla üç kanal arasında değiştirme yapmak olduğu gösterilmektedir.

Anahtar sözcükler: Zaman paylaşımı, rastgeleleştirme, parametre kestirimi, olasılıksal parametre tasarımı, Bayes riski, DEKOKH kestiricisi, kanal değiştirme, Gauss kanalı.

Acknowledgement

I would like to express my thanks and gratitude to my advisor Assoc. Prof. Sinan Gezici for his supervision, suggestions and invaluable encouragement throughout the development of this thesis. He has been a role model to me and it has always been an honour for me to be his student. Also, I would like to thank Prof. Dr. Orhan Arikan for his invaluable help and advices. In addition, I would like to extend my special thanks to Assoc. Prof. Dr. Selim Aksoy, Prof. Dr. A. Enis Çetin and Assoc. Prof. Dr. Ali Cafer Gürbüz for their valuable comments and suggestions on the thesis.

I wish to thank everyone in the Department of Electrical and Electronics Engineering at Bilkent University. I would like to thank in particular to my former colleagues Dr. Y. Kemal Alp, Dr. M. Emin Tutay and Asst. Prof. Dr. M. Burak GldoĖan for numerous technical and non-technical discussions.

This thesis study was supported in part by the National Young Researchers Career Development Programme (project no. 110E245) of the Scientific and Technological Research Council of Turkey (TBİTAK). I would like to thank to TBİTAK for their support. I would also like to thank all my current colleagues in Guidance Control and Navigation Division of TBİTAK SAGE for their help and support.

Finally, I would like to express my gratitude to my family. My parents Hseyin and Mukaddes, my sister Mine and my brother Murat. They always believed in me and helped me achieve my goals. My deepest gratitude goes to my wife, Betl, for her support and patience, which have been invaluable in helping me focus on my studies. My daughter Zehra was born towards the end of this dissertation and made my life filled with joy.

Contents

1	Introduction	1
1.1	Organization of the Dissertation	7
2	Optimal Stochastic Parameter Design for Estimation Problems	8
2.1	Stochastic Parameter Design	8
2.1.1	Unconstrained Optimization	10
2.1.2	Constrained Optimization	11
2.2	Optimality Conditions	14
2.3	Numerical Examples	18
2.4	Conclusions for Chapter 2	23
3	Optimal Stochastic Design of Multiple Parameters for Estimation Problems	24
3.1	Stochastic Design for Multi-Parameter Estimation	24
3.1.1	Unconstrained Optimization	26
3.1.2	Constrained Optimization	29
3.2	Characterization of Optimal Stochastic Parameter Design in the Presence of Average Power Constraint	32
3.3	Optimality Conditions	37
3.4	Numerical Examples	42
3.5	Conclusions for Chapter 3	52
4	Optimal Parameter Design for LMMSE Estimators	53
4.1	LMMSE Estimator	53
4.2	Optimal Parameter Design for LMMSE Estimator	55
4.3	Numerical Examples	62

4.4	Conclusions for Chapter 4	66
5	Optimal Channel Switching over Gaussian Channels under Average Power and Cost Constraints	67
5.1	System Model and Problem Formulation	68
5.2	Optimal Channel Switching	71
5.3	Optimal Channel Switching For Logarithmic Cost Functions . . .	83
5.4	Numerical Examples	92
5.5	Extensions	99
5.6	Conclusions for Chapter 5	100
6	Conclusion	101

List of Figures

2.1	System model. Device A transmits a stochastic signal $\mathbf{s}_\theta$ for each value of parameter θ , and Device B estimates θ based on the noise corrupted version of $\mathbf{s}_\theta$. One interpretation is to consider the dashed box as a measurement device, in which case $\mathbf{n}$ denotes the measurement noise.	9
2.2	Conditional risk versus θ for optimal deterministic and optimal stochastic approaches. Also results of convex relaxation is presented for two different values of D	19
2.3	$g_\theta(\mathbf{x})$ in (2.22) for various values of θ	21
2.4	Probability mass functions of the stochastic signal s_θ for $\theta = 5$ for unconstrained, optimal deterministic, optimal stochastic and conventional approaches. Also results of convex relaxation is presented for two different values of D	22
3.1	System model for $K = 2$. Devices A_1 and A_2 transmit stochastic signals $\mathbf{s}_{\theta_1}$ and $\mathbf{s}_{\theta_2}$ for each value of parameters θ_1 and θ_2 , respectively. Devices B_1 and B_2 estimate θ_1 and θ_2 based on the noise and interference corrupted version of $\mathbf{s}_{\theta_1}$ and $\mathbf{s}_{\theta_2}$, respectively. . .	25
3.2	Total Bayes risk versus θ_1 and θ_2 for the example defined in Section 3.4.	43
3.3	Total Bayes risk versus θ_1 and θ_2 for the example defined in Section 3.4.	44
3.4	Maximum Bayes risk versus θ_1 and θ_2 for the example defined in Section 3.4.	45

3.5	Maximum Bayes risk versus θ_1 and θ_2 for the example defined in Section 3.4.	46
3.6	Regions in which the optimality conditions stated in Proposition 3.6 are satisfied for different values of $\mathbf{z}$ for $\boldsymbol{\theta} = [-5, 5]$	49
3.7	The total Bayes risk $\tilde{g}_{\boldsymbol{\theta}}(\mathbf{x})$ for $\boldsymbol{\theta} = [-5, 5]$	50
3.8	The Bayes risk for the first estimator, $g_{\boldsymbol{\theta}_1}(\mathbf{x})$, and the Bayes risk for the second estimator, $g_{\boldsymbol{\theta}_2}(\mathbf{x})$, for $\boldsymbol{\theta} = [-5, 5]$	50
4.1	System model. Device A transmits a stochastic signal $\mathbf{s}_{\boldsymbol{\theta}}$ for each value of parameter $\boldsymbol{\theta}$, and Device B employs an LMMSE estimator and estimates $\boldsymbol{\theta}$ based on the noise corrupted version of $\mathbf{s}_{\boldsymbol{\theta}}$	55
4.2	MSE values of the first case for the conventional approach and the optimal parameter design.	63
4.3	MSE values of the second case for the conventional approach and the optimal parameter design.	64
4.4	MSE values of the third case for the conventional approach and the optimal parameter design.	65
5.1	A binary communication system that employs channel switching among K additive Gaussian noise channels, where C_i denotes the cost of using channel i , and N_i is the noise component at the i th channel. The transmitter employs antipodal signaling, $\{-s_i, s_i\}$, and the receiver performs sign detection, which is optimal for the considered scenario.	68
5.2	Average probability of error versus A_p for the optimal single channel and optimal channel switching strategies, where $K = 4$, $\boldsymbol{\sigma} = [0.4 \ 0.6 \ 0.8 \ 1]$, $\mathbf{C} = [7 \ 5 \ 3 \ 1]$, and $A_c = 2$	93
5.3	Average probability of error versus A_p for the optimal single channel and optimal channel switching strategies, where $K = 4$, $\boldsymbol{\sigma} = [0.4 \ 0.6 \ 0.8 \ 1]$, $\mathbf{C} = [7 \ 5 \ 3 \ 1]$, and $A_c = 5$	95
5.4	Average probability of error versus A_c for the optimal single channel and optimal channel switching strategies, where $K = 4$, $\boldsymbol{\sigma} = [0.4 \ 0.6 \ 0.8 \ 1]$, and $\mathbf{C} = [7 \ 5 \ 3 \ 1]$	96

- 5.5 Average probability of error versus A_p for the optimal single channel and optimal channel switching strategies, where $K = 5$, $\boldsymbol{\sigma} = [0.6 \ 0.7 \ 0.8 \ 0.9 \ 1]$, $\boldsymbol{C} = [1.329 \ 1.112 \ 0.941 \ 0.804 \ 0.6931]$, and $A_c = 0.9$ 97
- 5.6 Average probability of error versus A_c for the optimal single channel and optimal channel switching strategies, where $K = 5$, $\boldsymbol{\sigma} = [0.6 \ 0.7 \ 0.8 \ 0.9 \ 1]$, and $\boldsymbol{C} = [1.329 \ 1.112 \ 0.941 \ 0.804 \ 0.6931]$. . 98

List of Tables

2.1	Optimal stochastic solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \lambda_\theta \delta(\mathbf{x} - \mathbf{s}_{\theta,1}) + (1 - \lambda_\theta) \delta(\mathbf{x} - \mathbf{s}_{\theta,2})$, optimal deterministic solution $\mathbf{s}_\theta^{\text{opt}}$, and unconstrained solution $\mathbf{s}_\theta^{\text{unc}}$ for various values of $\boldsymbol{\theta}$	23
3.1	Unconstrained solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{unc}})$, optimal deterministic solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{opt}})$, and optimal stochastic solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \lambda_{\theta,1} \delta(\mathbf{x} - \mathbf{s}_{\theta,1}) + \lambda_{\theta,2} \delta(\mathbf{x} - \mathbf{s}_{\theta,2})$ for the total Bayes risk criterion for various values of $\boldsymbol{\theta}$	47
3.2	Unconstrained solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{unc}})$, optimal deterministic solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{opt}})$, and optimal stochastic solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \lambda_{\theta,1} \delta(\mathbf{x} - \mathbf{s}_{\theta,1}) + \lambda_{\theta,2} \delta(\mathbf{x} - \mathbf{s}_{\theta,2}) + \lambda_{\theta,3} \delta(\mathbf{x} - \mathbf{s}_{\theta,3})$ for the maximum Bayes risk criterion for various values of $\boldsymbol{\theta}$	48
4.1	Prior probabilities and $\mathbf{s}_\theta$ values for the three cases.	66
5.1	Parameters of the optimal channel switching strategy in Figure 5.2.	94
5.2	Parameters of the optimal channel switching strategy in Figure 5.3.	94
5.3	Parameters of the optimal channel switching strategy in Figure 5.5.	99

Dedicated to my family

Chapter 1

Introduction

Randomization or time sharing is a concept that can improve performance of certain detection and estimation systems; e.g., communication systems. Motivated by this fact, optimal time sharing strategies are studied in this dissertation for two different problems. In the first three chapters, time sharing (randomization) among different signal levels is investigated for parametric estimation problems. In the final chapter, time sharing among several channels is studied for a communication system in the context of optimal channel switching.

In parametric estimation problems, an unknown parameter is estimated based on observations, the probability distribution of which is known as a function of the unknown parameter [1, 2]. In the presence of prior information about the parameter, Bayesian estimators, such as the minimum mean-squared error (MMSE) estimator and the minimum mean-absolute error (MMAE) estimator, are commonly employed [1]. On the other hand, in the absence of prior information about the parameter, the minimum variance unbiased estimator (MVUE), if it exists, or the maximum likelihood estimator (MLE) can be used [2]. In these conventional formulations of the parameter estimation problem, the aim is to obtain an optimal estimator that minimizes a certain cost function, such as the mean-squared error. In this dissertation, we consider a different formulation in which the aim is to minimize the cost of a given estimator by performing randomization among

different signal values under certain constraints.

Randomization (time sharing) among different signal values has been utilized in various frameworks to improve performance of detection and estimation systems [3]-[16]. For example, performance of some detectors can be enhanced by the addition of a randomized noise component to the input (observation) without modifying the detector structure [3, 4, 5, 8, 9]. Such noise enhancement effects have been studied according to various criteria such as Neyman-Pearson (NP) [3, 4], Bayes [6], minimax [7], and restricted Bayes [8]. As another application of randomization, transmitting randomized signals for each information symbol can reduce the error probability of an average power constrained digital communication system in the presence of non-Gaussian noise [10, 11]. It is shown in [10] that the optimal strategy is to perform randomization among no more than three different transmitted signal values for each information symbol under second and fourth moment constraints. Randomization can be also utilized in jammer systems for improved jamming performance [17]-[19]. In [17], it is proved that a weak jammer employs on-off time sharing to maximize the average probability of error for a receiver operating in the presence of symmetric unimodal noise. On the other hand, for an average power constrained jammer that operates over an arbitrary additive noise channel, the detection probability of an instantaneously and fully adaptive receiver that employs the NP criterion is minimized via randomization between at most two different power levels [19]. In an estimation framework, benefits of randomization are observed in the context of noise enhanced estimation in [16], which proves that performance of some suboptimal estimators can be improved by adding randomized ‘noise’ to the observations before the estimation process.

Motivated by the investigation of signal randomization in recent works [3, 5, 10, 13, 16], we consider the concept of stochastic parameter design for estimation problems in the first chapter of this dissertation [20]. Specifically, we try to answer the following question: If a fixed estimator is used at the receiver, what should be the optimal distribution of the signal sent from the transmitter for each possible parameter value? Since there can exist power limits for transmitted signals in practice, this design problem needs to be solved under certain constraints.

As a specific example, consider a scenario in which the receiver employs the sample mean estimator to estimate a parameter θ based on a number of independent and identically distributed (i.i.d.) observations. The aim is to find the optimal random variable for each parameter value at the transmitter in order to minimize the Bayes risk of the sample mean estimator at the receiver. For instance, we would like to determine if sending i.i.d. Gaussian or Laplacian random variables with mean θ and variance 1 results in a lower Bayes risk. Or, more generally, among all continuous and discrete random variables, we would like to determine the one that minimizes the Bayes risk of the sample mean estimator.

In the second chapter, the aim is to propose a framework for the optimal stochastic design of *multiple* parameters [21, 22]. In this way, the approach in the first chapter for the single parameter case is extended to the multi-parameter scenario in which there exist multiple parameters (each can be a scalar or a vector) and corresponding fixed estimators. That is, the optimal stochastic design of multiple parameters is performed in order to optimize the performance of an array of fixed estimators. It should be emphasized that the difference of the multi-parameter case investigated in the second chapter from the single parameter case investigated in the first chapter is not only related to the number of parameters. The proposed multi-parameter formulation also takes into account the possible interference among parameter related signals (cf. Figure 3.1).

In the first two chapters, the optimal parameter design is performed for fixed (given) estimators. In the third chapter, the joint design of parameter dependent signals and the estimator is considered. Since the optimal joint design has high computational complexity and does not lead to closed-form expressions for optimal Bayesian estimators, we consider linear minimum mean squared error (LMMSE) estimators and aim to perform the optimal parameter design and the estimator design in this scenario.

In conventional estimation problems, the aim is to design an optimal estimator for a given distribution of the observations. However, motivations can also be provided for the stochastic parameter design problem investigated in this dissertation. For example, consider the design of a generic device (Device A in

Figure 2.1) which needs to output a certain parameter. This output is to be measured by a measurement device (the dashed box in Figure 2.1) which employs a certain estimation algorithm for determining the parameter (e.g., averages various measurements). Then, the aim is to design a stochastic signal $\mathbf{s}_\theta$ for each θ so that the accuracy (i.e., estimation performance) of the given measurement device is optimized. In other words, considering a certain type of a measurement device, the estimation performance of the overall system is to be optimized by designing stochastic signals for different parameters. Such a system model, in which estimation is performed based on measurements obtained by a number of measurement devices, is considered also in [23]. However, a different problem is considered in that study, and the optimal linear estimator is obtained in the presence of cost-constrained measurements. It should also be mentioned that most measurement devices are designed under a certain measurement noise assumption, such as Gaussian. They are typically non-adaptive devices, hence, in the presence of noise that deviates from the assumed noise distribution, their performance may degrade significantly. To improve the performance, the measurement device can be replaced with a more capable one; however, such a replacement may be very costly in some cases. To avoid the replacement cost and associated complications, the proposed stochastic parameter design approach can be used, which designs optimal signals for each parameter so that the performance of the suboptimal measurement device can be improved.

As another motivation of the setup in Figure 2.1, a wireless sensor network [24], in which a parameter value (such as temperature or pressure) is sent from one device to another, can be considered. When the transmitter (Device A) knows the probability distribution of the channel noise, $\mathbf{n}$ (which can be obtained via feedback), it can perform stochastic parameter design in order to optimize the performance of the estimator at the receiver (Device B). If the probability distribution of $\mathbf{n}$ is unknown, then the results can be considered to provide a theoretical upper bound on the estimation performance. It is important to note that the additive noise is used to model all the operations/effects between Device A and Device B in Figure 2.1. For example, signal values can be quantized, and encoded symbols can be sent via a specific digital communications method in

some cases. Then, the additive noise model in Figure 2.1 can be considered to provide an abstraction for all the blocks between Device A and Device B, such as quantizer, encoder/decoder, modulator/demodulator, and additive noise channel, as discussed in [17].

Time sharing is not only limited to randomization among different signal values. For communication systems that operate under average power constraints, time sharing among different detectors or channels can also provide performance improvements in the presence of additive time-invariant noise [4, 11, 13, 14, 17, 18, 19, 25, 26, 27, 28]. Time sharing among multiple detectors, which is also called *detector randomization*, presents an approach for improving error performance of average power constrained communication systems that operate over an additive time-invariant noise channel [4, 13, 14, 26, 29, 30]. In this approach, a receiver has multiple detectors and employs one of them at any given time according to a certain time sharing strategy. In [4], an average power constrained binary communication system is considered, and the optimal time sharing between two antipodal signal pairs and the corresponding maximum *a posteriori* probability (MAP) detectors is investigated. Significant performance improvements can be achieved as a result of the proposed approach in the presence of symmetric Gaussian mixture noise for a certain range of average power limits. In [13], the results in [4] and [11] are generalized by considering an average power constrained M -ary communication system that can employ time sharing among both signal levels and detectors over an additive noise channel with some known distribution. It is proved that the joint optimization of the transmitted signals and the detectors at the receiver results in time sharing between at most two MAP detectors corresponding to two deterministic signal constellations. [14] investigates the benefits of time sharing among multiple detectors for the downlink of a multiuser communication system and characterizes the optimal time sharing strategy. In a related study, the form of the optimal additive noise is obtained for variable detectors in the context of noise enhanced detection under both Neyman-Pearson and Bayesian frameworks [26].

In the presence of multiple channels between a transmitter and a receiver, performing time sharing among different channels, which is called *channel switching*,

can provide certain performance improvements [17, 27, 28, 31]. In the channel switching approach, communication occurs over one channel for a certain fraction of time, and then it switches to another channel during the next transmission. In [17], the channel switching problem is studied under an average power constraint for the optimal detection of binary antipodal signals over a number of channels that are subject to additive unimodal noise. It is shown that the optimal solution is either to communicate over one channel exclusively, or to switch between two channels with a certain time sharing factor. In [28], the channel switching problem is investigated for M -ary communication systems in the presence of additive noise channels with arbitrary probability distributions and by facilitating time sharing among multiple signal constellations over each channel. Under an average power constraint, the optimal solution that minimizes the average probability of error is obtained as one of the following strategies: deterministic signaling (i.e., use of one signal constellation) over a single channel; time sharing between two different signal constellations over a single channel; or switching (time sharing) between two channels with deterministic signaling over each channel [28]. In a different context, the concept of channel switching is employed for cognitive radio systems with opportunistic spectrum access, where a number of secondary users try to access the available frequency bands in the spectrum [32]-[34].

Although the channel switching problem has been investigated thoroughly under an average power constraint (e.g., [17, 28]), no studies have considered the cost of communications over different channels in obtaining the optimal channel switching strategy. In practical systems, each channel can be associated with a certain cost depending on its quality [23, 35, 36, 37, 38, 39]. For example, a channel that presents high signal-to-noise ratio (SNR) conditions has a high cost (price) compared to channels with low SNRs [36, 39]. Therefore, it is important to consider costs of different channels while designing a channel switching strategy. In the final chapter of this dissertation, the optimal channel switching problem is formulated for Gaussian channels in the presence of average power and average cost constraints, where each channel has a different cost depending on its quality [40]. In such a case optimal channel switching strategy should be identified not just by considering a certain power constraint but also by considering a certain

cost constraint.

1.1 Organization of the Dissertation

This dissertation is organized as follows. In Chapter 2, the optimal stochastic design of a single parameter is investigated for a fixed estimator. Optimality conditions are derived for the stochastic design. The problem in Chapter 2 is extended to multiple parameters in Chapter 3, where interference among the parameter dependent signals is also considered. In Chapter 4, the optimal parameter design is studied for LMMSE estimators, which is an important class of practical estimators. Finally in Chapter 5, the optimal channel switching strategy is investigated for Gaussian channels with different costs. This channel switching problem is formulated under both average power and average cost constraints.

Chapter 2

Optimal Stochastic Parameter Design for Estimation Problems

In this chapter, an optimal randomization (time sharing) strategy is investigated for parametric estimation problems. This chapter is organized as follows. In Section 2.1, the problem of optimal parameter design for estimation problems is defined. Then, theoretical results for the unconstrained case is obtained in Section 2.1.1. Next, theoretical results in the presence of an average power constraint is presented in Section 2.1.2. Sufficient conditions are derived in Section 2.2 in order to specify when the stochastic parameter design or the deterministic parameter design is optimal. A numerical example illustrating the improvement achieved via randomization is presented in Section 2.3. Finally concluding remarks are given in Section 2.4.

2.1 Stochastic Parameter Design

Consider a parameter estimation scenario as in Figure 2.1, where the aim is to send the information about parameter $\boldsymbol{\theta}$, which resides in $\mathbb{R}^M$, from Device A to Device B over an additive noise channel. For that purpose, Device A can

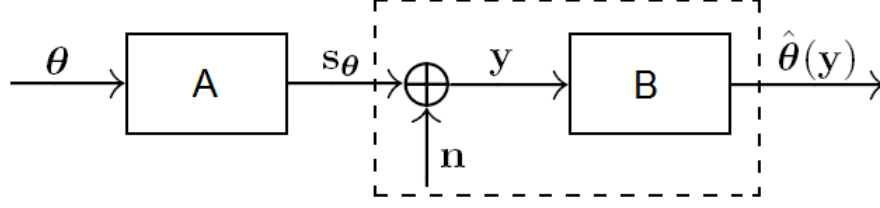


Figure 2.1: System model. Device A transmits a stochastic signal $\mathbf{s}_\theta$ for each value of parameter θ , and Device B estimates θ based on the noise corrupted version of $\mathbf{s}_\theta$. One interpretation is to consider the dashed box as a measurement device, in which case $\mathbf{n}$ denotes the measurement noise.

transmit a (random) function of θ , say $\mathbf{s}_\theta$, to Device B. Then, the received signal (observation) at Device B is expressed as

$$\mathbf{y} = \mathbf{s}_\theta + \mathbf{n} \quad (2.1)$$

where $\mathbf{n}$ denotes the channel noise, which has a probability density function (PDF) represented by $p_{\mathbf{n}}(\cdot)$. It is assumed that Device B employs a fixed estimator specified by $\hat{\theta}(\mathbf{y})$ in order to estimate θ . In addition, the prior distribution of θ is denoted by $w(\theta)$, and the parameter space in which θ resides is represented by Λ .

In this chapter, the problem is to find the optimal probability distribution of $\mathbf{s}_\theta$ for each $\theta \in \Lambda$ in order to minimize the Bayes risk of a given estimator. It should be noted that, in conventional estimation problems, the aim is to design the optimal estimator for a given probability distribution of the observation [2]. However, we consider a different problem in which the aim is to optimize the information carrying parameters in order to optimize the performance of a given estimator. Another important point is that unlike conventional estimation problems, $\mathbf{s}_\theta$ in (2.1) is modeled as a random variable for each value of θ ; that is, a *stochastic parameter design* approach is considered in this chapter.

2.1.1 Unconstrained Optimization

First, no constraints are considered in the selection of $\mathbf{s}_\theta$. Then, the optimal stochastic parameter design problem can be formulated as

$$\{p_{\mathbf{s}_\theta}^{\text{opt}}, \boldsymbol{\theta} \in \Lambda\} = \arg \min_{\{p_{\mathbf{s}_\theta}, \boldsymbol{\theta} \in \Lambda\}} r(\hat{\boldsymbol{\theta}}) \quad (2.2)$$

where $\{p_{\mathbf{s}_\theta}, \boldsymbol{\theta} \in \Lambda\}$ denotes the set of PDFs for $\mathbf{s}_\theta$ for all possible values of parameter $\boldsymbol{\theta}$, and $r(\hat{\boldsymbol{\theta}})$ is the Bayes risk of the estimator. In order to obtain a more explicit formulation of the problem, the Bayes risk can be expressed as

$$r(\hat{\boldsymbol{\theta}}) = \int_{\Lambda} w(\boldsymbol{\theta}) \int C[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}] p_{\boldsymbol{\theta}}(\mathbf{y}) d\mathbf{y} d\boldsymbol{\theta} \quad (2.3)$$

where $p_{\boldsymbol{\theta}}(\mathbf{y})$ denotes the PDF of $\mathbf{y}$, which is indexed by $\boldsymbol{\theta}$, and $C[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}]$ represents a cost function [2]. For example, $C[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}] = (\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})^2$ corresponds to the squared-error cost function, for which $r(\hat{\boldsymbol{\theta}})$ becomes the mean-squared error (MSE). In this chapter, a generic cost function $C[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}]$ is considered in all the derivations.

If $\mathbf{s}_\theta$ were modeled as a deterministic quantity for each value of $\boldsymbol{\theta}$, $p_{\boldsymbol{\theta}}(\mathbf{y})$ in (2.3) could be expressed in terms of the PDF of $\mathbf{n}$ as $p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_\theta)$ (see (2.1)). However, we consider a stochastic parameter design framework and model $\mathbf{s}_\theta$ as a stochastic variable for each $\boldsymbol{\theta}$. Then, assuming that the noise and $\mathbf{s}_\theta$ are independent, $p_{\boldsymbol{\theta}}(\mathbf{y})$ is calculated as $\int p_{\mathbf{s}_\theta}(\mathbf{x}) p_{\mathbf{n}}(\mathbf{y} - \mathbf{x}) d\mathbf{x}$. Therefore, (2.3) becomes

$$r(\hat{\boldsymbol{\theta}}) = \int_{\Lambda} w(\boldsymbol{\theta}) \int p_{\mathbf{s}_\theta}(\mathbf{x}) \int C[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}] p_{\mathbf{n}}(\mathbf{y} - \mathbf{x}) d\mathbf{y} d\mathbf{x} d\boldsymbol{\theta} . \quad (2.4)$$

Defining an auxiliary function $g_{\boldsymbol{\theta}}(\mathbf{x})$ as

$$g_{\boldsymbol{\theta}}(\mathbf{x}) \triangleq \int C[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}] p_{\mathbf{n}}(\mathbf{y} - \mathbf{x}) d\mathbf{y} , \quad (2.5)$$

the relation in (2.4) can be stated as

$$r(\hat{\boldsymbol{\theta}}) = \int_{\Lambda} w(\boldsymbol{\theta}) E\{g_{\boldsymbol{\theta}}(\mathbf{s}_\theta)\} d\boldsymbol{\theta} \quad (2.6)$$

where each expectation operation is over the PDF of $\mathbf{s}_\theta$ for a given value of $\boldsymbol{\theta}$. From (2.6), it is observed that $r(\hat{\boldsymbol{\theta}})$ can be minimized if, for each $\boldsymbol{\theta}$, the PDF of

$\mathbf{s}_\theta$ assigns all the probability to the minimizer of g_θ .¹ Namely, the solution of the optimization problem in (2.2) can be expressed as

$$p_{\mathbf{s}_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{unc}}) , \quad \mathbf{s}_\theta^{\text{unc}} = \arg \min_{\mathbf{x}} g_\theta(\mathbf{x}) \quad (2.7)$$

for all $\theta \in \Lambda$. Therefore, it is concluded that the optimal stochastic parameter design results in optimal PDFs that have single point masses. Hence, deterministic parameter design is optimal and no stochastic modeling is needed when there are no constraints in the design problem. However, in practice, the values of $\mathbf{s}_\theta$ cannot be chosen without any constraints (such as an average power constraint), and it will be shown in the next section that the stochastic parameter design can result in performance improvements in the presence of constraints on the moments of $\mathbf{s}_\theta$. Another important observation from (2.7) is that the solution does not require the knowledge of the prior distribution $w(\theta)$, since the optimal solution is obtained for each θ separately.

2.1.2 Constrained Optimization

In practical scenarios, the parameter design cannot be performed without any limitations. For example, in the absence of a power constraint, it would be possible to reduce the Bayes risk arbitrarily by transmitting signals with very high powers compared to the noise power.

In this section, a common design constraint in the form of an average power constraint is considered in the stochastic parameter design problem. Although a specific constraint type is used in the following, it will be discussed that other types of constraints can also be incorporated into the theoretical analysis.

Consider an average power constraint in the form of

$$\mathbb{E}\{\|\mathbf{s}_\theta\|^2\} \leq A_\theta \quad (2.8)$$

¹If there are multiple minimizers, any (combination) of them can be chosen for the optimal solution.

for $\boldsymbol{\theta} \in \Lambda$, where $\|\mathbf{s}_{\boldsymbol{\theta}}\|$ is the Euclidean norm of vector $\mathbf{s}_{\boldsymbol{\theta}}$, and $A_{\boldsymbol{\theta}}$ denotes the average power constraint for $\boldsymbol{\theta}$. It is noted from (2.8) that a generic model is considered for the constraint $A_{\boldsymbol{\theta}}$, which can depend on the value of $\boldsymbol{\theta}$ in general. For the special case in which the average power constraint is the same for all parameters, $A_{\boldsymbol{\theta}} = A$ for $\boldsymbol{\theta} \in \Lambda$ can be employed.

From (2.6) and (2.8), the optimal stochastic parameter design problem can be stated as

$$\begin{aligned} \min_{\{p_{\mathbf{s}_{\boldsymbol{\theta}}}, \boldsymbol{\theta} \in \Lambda\}} & \int_{\Lambda} w(\boldsymbol{\theta}) \mathbb{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\} d\boldsymbol{\theta} \\ \text{subject to} & \mathbb{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \leq A_{\boldsymbol{\theta}}, \forall \boldsymbol{\theta} \in \Lambda \end{aligned} \quad (2.9)$$

where $g_{\boldsymbol{\theta}}(\cdot)$ is as defined in (2.5). The investigation of the constrained optimization problem in (2.9) reveals that the problem can be solved separately for each $\boldsymbol{\theta}$ as follows:

$$\min_{p_{\mathbf{s}_{\boldsymbol{\theta}}}} \mathbb{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\} \quad \text{subject to} \quad \mathbb{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \leq A_{\boldsymbol{\theta}} \quad (2.10)$$

for $\boldsymbol{\theta} \in \Lambda$. In other words, the optimal PDF of $\mathbf{s}_{\boldsymbol{\theta}}$ can be obtained separately for each $\boldsymbol{\theta}$. Therefore, the result does not depend on the prior distribution $w(\boldsymbol{\theta})$, and the solution can be obtained in the absence of prior information.

Optimization problems in the form of (2.10) have been investigated in different studies in the literature [3, 10, 4]. Specifically, [3] and [4] aim to obtain the optimal additive “noise” PDF that maximizes the detection probability under a constraint on the false-alarm probability, and [10] investigates optimal signal PDFs in a power constrained binary communications systems. Based on similar arguments to those in [3, 10, 4], the following result can be obtained.

Proposition 2.1 *Suppose $g_{\boldsymbol{\theta}}$ is a continuous function and each component of $\mathbf{s}_{\boldsymbol{\theta}}$ resides in a finite closed interval. Then, an optimal solution to (2.10) can be expressed in the following form:*

$$p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}(\mathbf{x}) = \lambda_{\boldsymbol{\theta}} \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},1}) + (1 - \lambda_{\boldsymbol{\theta}}) \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},2}) \quad (2.11)$$

for $\lambda_{\boldsymbol{\theta}} \in [0, 1]$.

Proof: Consider the set of all $(g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}), \|\mathbf{s}_{\boldsymbol{\theta}}\|^2)$ pairs and the set of all $(\mathbb{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\}, \mathbb{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\})$ pairs, and denote them as U and W , respectively. Namely, $U = \{(u_1, u_2) : u_1 = g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}), u_2 = \|\mathbf{s}_{\boldsymbol{\theta}}\|^2, \forall \mathbf{s}_{\boldsymbol{\theta}}\}$ and $W = \{(w_1, w_2) : w_1 = \mathbb{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\}, w_2 = \mathbb{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\}, \forall p_{\mathbf{s}_{\boldsymbol{\theta}}}\}$. As discussed in [3] and [10], the convex hull of U can be shown to be equal to W . Then, based on Carathéodory's theorem [41], it is concluded that any point in W can be obtained as a convex combination of at most three points in U . Also, since an optimal PDF should achieve the minimum value, it must correspond to the boundary of W , which results in a convex combination of at most two points in U . (The assumptions in the proposition imply that W is a closed set; therefore, it contains its boundary [10].) Hence, an optimal solution can be expressed as in (2.11) [19]. ■

Proposition 2.1 states that the optimal solution can be achieved by randomization between at most two different values for each $\boldsymbol{\theta}$. Based on this result, the optimal stochastic parameter design problem in (2.10) is expressed as

$$\begin{aligned} \min_{\lambda_{\boldsymbol{\theta}}, \mathbf{s}_{\boldsymbol{\theta},1}, \mathbf{s}_{\boldsymbol{\theta},2}} \quad & \lambda_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},1}) + (1 - \lambda_{\boldsymbol{\theta}}) g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},2}) \\ \text{subject to} \quad & \lambda_{\boldsymbol{\theta}} \|\mathbf{s}_{\boldsymbol{\theta},1}\|^2 + (1 - \lambda_{\boldsymbol{\theta}}) \|\mathbf{s}_{\boldsymbol{\theta},2}\|^2 \leq A_{\boldsymbol{\theta}}, \lambda_{\boldsymbol{\theta}} \in [0, 1] \end{aligned} \quad (2.12)$$

for $\boldsymbol{\theta} \in \Lambda$. Compared to (2.10), the formulation in (2.12) provides a significant simplification as it requires optimization over a finite number of variables instead of over all possible PDFs. Since generic cost functions and noise distributions are considered in the theoretical analysis, $g_{\boldsymbol{\theta}}$ in (2.5) is quite generic and the optimization problem in (2.12) can be nonconvex in general. Therefore, global optimization techniques such as particle swarm optimization (PSO) and differential evolution (DE) can be employed to obtain the solution [42, 43].

Remark 2.1 *Although the average power constraint in (2.8) is considered in obtaining the preceding results, the other types of constraints in the form of $\mathbb{E}\{h_i(\mathbf{s}_{\boldsymbol{\theta}})\} \leq A_{\boldsymbol{\theta},i}$ for $i = 1, \dots, N_c$ can also be incorporated. Specifically, assuming continuous h_i , the form of the optimal PDF in Proposition 1 becomes $p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}(\mathbf{x}) = \sum_{i=1}^{N_c} \lambda_{\boldsymbol{\theta},i} \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},i})$, with $\lambda_{\boldsymbol{\theta},i} \geq 0$ for $i = 1, \dots, N_c$ and $\sum_{i=1}^{N_c} \lambda_{\boldsymbol{\theta},i} = 1$, which can be proven by updating the definitions of sets U and W accordingly in the proof of Proposition 2.1.*

As an alternative approach, a convex relaxation technique can be employed to obtain an approximate solution of (2.10) in polynomial time [10, 44]. To that aim, it is assumed that $p_{\mathbf{s}_\theta}$ can be expressed as $p_{\mathbf{s}_\theta}(\mathbf{x}) = \sum_{l=1}^{N_m} \beta_l \delta(\mathbf{x} - \tilde{\mathbf{s}}_{\theta,l})$, where $\beta_l \geq 0$ for $l = 1, \dots, N_m$, $\sum_{l=1}^{N_m} \beta_l = 1$, and $\tilde{\mathbf{s}}_{\theta,1}, \dots, \tilde{\mathbf{s}}_{\theta,N_m}$ are known possible values for $\mathbf{s}_\theta$. Then, by defining $\boldsymbol{\beta} = [\beta_1 \cdots \beta_{N_m}]^T$, $\tilde{\mathbf{g}}_\theta = [g_\theta(\tilde{\mathbf{s}}_{\theta,1}) \cdots g_\theta(\tilde{\mathbf{s}}_{\theta,N_m})]^T$ and $\mathbf{c} = [\|\tilde{\mathbf{s}}_{\theta,1}\|^2 \cdots \|\tilde{\mathbf{s}}_{\theta,N_m}\|^2]^T$, the convex version of (2.10) can be obtained as

$$\min_{\boldsymbol{\beta}} \boldsymbol{\beta}^T \tilde{\mathbf{g}}_\theta \quad \text{subject to } \boldsymbol{\beta}^T \mathbf{c} \leq A_\theta, \quad \boldsymbol{\beta}^T \mathbf{1} = 1, \quad \boldsymbol{\beta} \succeq \mathbf{0} \quad (2.13)$$

where $\mathbf{1}$ and $\mathbf{0}$ denote the vectors of ones and zeros, respectively, and $\boldsymbol{\beta} \succeq \mathbf{0}$ means that each element of $\boldsymbol{\beta}$ is greater than or equal to zero. It is noted that (2.13) presents a linearly constrained linear optimization problem; hence, it can be solved efficiently in polynomial time [44]. In general, the solution of (2.13) provides an approximate solution, and the approximation accuracy can be improved by using a large value of N_m .

2.2 Optimality Conditions

The deterministic parameter design can be considered as a special case of the stochastic parameter design when $\mathbf{s}_\theta$ in (2.10) is modeled as a deterministic quantity for each θ . Namely, the deterministic parameter design problem can be formulated as

$$\min_{\mathbf{s}_\theta} g_\theta(\mathbf{s}_\theta) \quad \text{subject to } \|\mathbf{s}_\theta\|^2 \leq A_\theta \quad (2.14)$$

for $\theta \in \Lambda$ (c.f. (2.10)). Let $\mathbf{s}_\theta^{\text{opt}}$ denote the minimizer of the optimization problem in (2.14). Then, the minimum Bayes risk achieved by the optimal deterministic parameter design is given by $r_{\text{det}}(\hat{\theta}) = \int_{\Lambda} w(\theta) g_\theta(\mathbf{s}_\theta^{\text{opt}}) d\theta$ (see (2.6)). Similarly, let $r_{\text{sto}}(\hat{\theta}) = \int_{\Lambda} w(\theta) \int g_\theta(\mathbf{x}) p_{\mathbf{s}_\theta}^{\text{opt}}(\mathbf{x}) d\mathbf{x} d\theta$ represent the minimum Bayes risk achieved by the optimal stochastic parameter design, where $p_{\mathbf{s}_\theta}^{\text{opt}}$ denotes the optimal solution for θ . In order for the stochastic parameter design to improve over the deterministic parameter design, $r_{\text{sto}}(\hat{\theta})$ should be strictly smaller than $r_{\text{det}}(\hat{\theta})$. Otherwise, it is concluded that the deterministic parameter design cannot be improved via the stochastic approach; that is, $r_{\text{sto}}(\hat{\theta}) = r_{\text{det}}(\hat{\theta})$. In the following

proposition, sufficient conditions presented for the latter.

Proposition 2.2 *The deterministic parameter design cannot be improved via the stochastic approach if at least one of the following is satisfied for each θ :*

- g_θ is a convex function.
- The solution of the unconstrained problem (see (2.7)) satisfies the constraint; i.e., $\|\mathbf{s}_\theta^{\text{unc}}\|^2 \leq A_\theta$.

Proof: If the second condition is satisfied, that is, if $\|\mathbf{s}_\theta^{\text{unc}}\|^2 \leq A_\theta$, then the solution of (2.14) coincides with that of the unconstrained problem in Section 2.1.1; namely, $\mathbf{s}_\theta^{\text{opt}} = \mathbf{s}_\theta^{\text{unc}}$. Therefore, the solution of the optimal stochastic parameter design problem in (2.10) becomes $p_{\mathbf{s}_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{opt}})$. Hence, the deterministic design is optimal in such a scenario, and the stochastic approach is not needed.

In order to investigate the first condition, it is observed that, for any $\mathbf{s}_\theta$, $E\{\|\mathbf{s}_\theta\|^2\} \geq \|E\{\mathbf{s}_\theta\}\|^2$ is satisfied due to Jensen's inequality since norm is a convex function. Therefore, due to the constraint $E\{\|\mathbf{s}_\theta\|^2\} \leq A_\theta$ in (2.10), $\|E\{\mathbf{s}_\theta\}\|^2 \leq A_\theta$ must hold for any feasible PDF of $\mathbf{s}_\theta$. Let $E\{\mathbf{s}_\theta\}$ be defined as $\check{\mathbf{s}}_\theta \triangleq E\{\mathbf{s}_\theta\}$. As the minimizer of (2.14), $\mathbf{s}_\theta^{\text{opt}}$, achieves the minimum $g_\theta(\mathbf{s}_\theta)$ among all $\mathbf{s}_\theta$ that satisfy $\|\mathbf{s}_\theta\|^2 \leq A_\theta$, $\|E\{\mathbf{s}_\theta\}\|^2 = \|\check{\mathbf{s}}_\theta\|^2 \leq A_\theta$ implies that $g_\theta(E\{\mathbf{s}_\theta\}) = g_\theta(\check{\mathbf{s}}_\theta) \geq g_\theta(\mathbf{s}_\theta^{\text{opt}})$ is satisfied. When g_θ is a convex function as specified in the proposition, $E\{g_\theta(\mathbf{s}_\theta)\} \geq g_\theta(E\{\mathbf{s}_\theta\}) \geq g_\theta(\mathbf{s}_\theta^{\text{opt}})$ is obtained from Jensen's inequality and from the previous relation. Therefore, for convex g_θ , $E\{g_\theta(\mathbf{s}_\theta)\}$ can never be smaller than the minimum value of (2.14), $g_\theta(\mathbf{s}_\theta^{\text{opt}})$, for any PDF of $\mathbf{s}_\theta$ that satisfies the average power constraint. Hence, the minimum value of (2.10) cannot be smaller than $g_\theta(\mathbf{s}_\theta^{\text{opt}})$, meaning that it is always equal to $g_\theta(\mathbf{s}_\theta^{\text{opt}})$ (since (2.10) covers (2.14) as a special case).

All in all, when at least one of the conditions in the proposition are satisfied for all θ , the deterministic and the stochastic approaches achieve the same minimum

values for all parameters; that is, $g_{\theta}(\mathbf{s}_{\theta}^{\text{opt}}) = \int g_{\theta}(\mathbf{x})p_{\mathbf{s}_{\theta}}^{\text{opt}}(\mathbf{x})d\mathbf{x}$, $\forall \theta$. Therefore, $r_{\text{det}}(\hat{\theta}) = \int_{\Lambda} w(\theta)g_{\theta}(\mathbf{s}_{\theta}^{\text{opt}})d\theta$ and $r_{\text{sto}}(\hat{\theta}) = \int_{\Lambda} w(\theta) \int g_{\theta}(\mathbf{x})p_{\mathbf{s}_{\theta}}^{\text{opt}}(\mathbf{x})d\mathbf{x}d\theta$ become equal. $\blacksquare$

In order to present an example application of Proposition 2.2, consider a scenario in which a scalar parameter θ is to be estimated in the presence of zero-mean additive noise n . The average power constraint is in the generic form of $E\{|s_{\theta}|^2\} \leq A_{\theta}$ for all θ , and the estimator is specified by $\hat{\theta}(y) = y$. In addition, the cost function is modeled as $C[\hat{\theta}(y), \theta] = (\hat{\theta}(y) - \theta)^2$. In this scenario, g_{θ} in (2.5) can be calculated as

$$g_{\theta}(\mathbf{x}) = \int_{-\infty}^{\infty} (y - \theta)^2 p_n(y - \mathbf{x}) dy = \int_{-\infty}^{\infty} (y + \mathbf{x} - \theta)^2 p_n(y) dy = (\mathbf{x} - \theta)^2 + \text{Var}\{n\} \quad (2.15)$$

where $\text{Var}\{n\}$ denotes the variance of the noise. From (2.15), it is noted that g_{θ} is a convex function for any value of θ . Therefore, the first condition in Proposition 2.2 is satisfied for all θ , meaning that the performance of the deterministic parameter design cannot be improved via the stochastic approach.² Hence, the optimal solution can be obtained from (2.14), which yields

$$\mathbf{s}_{\theta}^{\text{opt}} = \arg \min_{|\mathbf{s}_{\theta}|^2 \leq A_{\theta}} (\mathbf{s}_{\theta} - \theta)^2.$$

For example, if $A_{\theta} = \theta^2$, then $\mathbf{s}_{\theta}^{\text{opt}} = \theta$ for all θ .

In the following proposition, sufficient conditions are presented to specify cases in which the stochastic parameter design provides improvements over the deterministic one.

Proposition 2.3 *The stochastic parameter design achieves a smaller Bayes risk than the deterministic one if there exists $\theta \in \Lambda$ for which $g_{\theta}(\mathbf{x})$ is second-order continuously differentiable around $\mathbf{s}_{\theta}^{\text{opt}}$ and a real vector $\mathbf{z}$ can be found such that*

$$(\mathbf{z}^T \mathbf{s}_{\theta}^{\text{opt}})(\mathbf{z}^T \nabla g_{\theta}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\theta}^{\text{opt}}}) < 0 \quad \text{and} \quad (2.16)$$

$$\|\mathbf{z}\|^2 < (\mathbf{z}^T \mathbf{s}_{\theta}^{\text{opt}})(\mathbf{z}^T \mathbf{H}_{\theta} \mathbf{z}) / (\mathbf{z}^T \nabla g_{\theta}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\theta}^{\text{opt}}}) \quad (2.17)$$

²It can be shown that g_{θ} is convex for all θ also for the absolute error cost function; i.e., $C[\hat{\theta}(y), \theta] = |\hat{\theta}(y) - \theta|$.

where $\mathbf{s}_\theta^{\text{opt}}$ is the solution of (2.14), $\nabla g_\theta(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_\theta^{\text{opt}}}$ denotes the gradient of $g_\theta(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_\theta^{\text{opt}}$, and $\mathbf{H}_\theta$ is the Hessian of $g_\theta(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_\theta^{\text{opt}}$.

Proof: In order to prove that a reduced Bayes risk can be achieved via the stochastic parameter design, consider a specific value of θ for which the conditions in the proposition are satisfied. Also consider two values $\mathbf{s}_{\theta,1}$ and $\mathbf{s}_{\theta,2}$ around $\mathbf{s}_\theta^{\text{opt}}$, which can be expressed as $\mathbf{s}_{\theta,i} = \mathbf{s}_\theta^{\text{opt}} + \boldsymbol{\epsilon}_i$ for $i = 1, 2$. Then, $g_\theta(\mathbf{s}_{\theta,i})$ can be approximated as $g_\theta(\mathbf{s}_{\theta,i}) \approx g_\theta(\mathbf{s}_\theta^{\text{opt}}) + \boldsymbol{\epsilon}_i^T \tilde{\mathbf{g}}_\theta + 0.5 \boldsymbol{\epsilon}_i^T \mathbf{H}_\theta \boldsymbol{\epsilon}_i$ for $i = 1, 2$, where $\tilde{\mathbf{g}}_\theta = \nabla g_\theta(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_\theta^{\text{opt}}}$ is the gradient and $\mathbf{H}_\theta$ is the Hessian of $g_\theta(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_\theta^{\text{opt}}$ [45]. Similarly, $\|\mathbf{s}_{\theta,i}\|^2$ can be expressed as $\|\mathbf{s}_{\theta,i}\|^2 \approx \|\mathbf{s}_\theta^{\text{opt}}\|^2 + 2 \boldsymbol{\epsilon}_i^T \mathbf{s}_\theta^{\text{opt}} + \|\boldsymbol{\epsilon}_i\|^2$ for $i = 1, 2$. In order to prove that employing $p_{\mathbf{s}_\theta}(\mathbf{x}) = \lambda \delta(\mathbf{x} - \mathbf{s}_{\theta,1}) + (1 - \lambda) \delta(\mathbf{x} - \mathbf{s}_{\theta,2})$ results in a lower risk than $g_\theta(\mathbf{s}_\theta^{\text{opt}})$, which is the one achieved by the deterministic parameter design (see (2.14)), it is sufficient to show that

$$\begin{aligned} \lambda g_\theta(\mathbf{s}_{\theta,1}) + (1 - \lambda) g_\theta(\mathbf{s}_{\theta,2}) &< g_\theta(\mathbf{s}_\theta^{\text{opt}}) \\ \lambda \|\mathbf{s}_{\theta,1}\|^2 + (1 - \lambda) \|\mathbf{s}_{\theta,2}\|^2 &< \|\mathbf{s}_\theta^{\text{opt}}\|^2 \leq A_\theta \end{aligned} \quad (2.18)$$

are satisfied for certain choice of parameters (see (2.10)). After inserting the expressions for $g_\theta(\mathbf{s}_{\theta,i})$ and $\|\mathbf{s}_{\theta,i}\|^2$ around $\mathbf{s}_\theta^{\text{opt}}$ into (2.18), it can be obtained that

$$\begin{aligned} \lambda \boldsymbol{\epsilon}_1^T \mathbf{H}_\theta \boldsymbol{\epsilon}_1 + (1 - \lambda) \boldsymbol{\epsilon}_2^T \mathbf{H}_\theta \boldsymbol{\epsilon}_2 + 2(\lambda \boldsymbol{\epsilon}_1 + (1 - \lambda) \boldsymbol{\epsilon}_2)^T \tilde{\mathbf{g}}_\theta &< 0 \\ \lambda \|\boldsymbol{\epsilon}_1\|^2 + (1 - \lambda) \|\boldsymbol{\epsilon}_2\|^2 + 2(\lambda \boldsymbol{\epsilon}_1 + (1 - \lambda) \boldsymbol{\epsilon}_2)^T \mathbf{s}_\theta^{\text{opt}} &< 0 \end{aligned} \quad (2.19)$$

Let $\boldsymbol{\epsilon}_1 = \eta \mathbf{z}$ and $\boldsymbol{\epsilon}_2 = \nu \mathbf{z}$. Then, (2.19) can be manipulated to obtain

$$\mathbf{z}^T \mathbf{H}_\theta \mathbf{z} + k(\mathbf{z}^T \tilde{\mathbf{g}}_\theta) < 0 \quad \text{and} \quad \|\mathbf{z}\|^2 + k(\mathbf{z}^T \mathbf{s}_\theta^{\text{opt}}) < 0 \quad (2.20)$$

with $k \triangleq 2(\lambda \eta + (1 - \lambda) \nu) / (\lambda \nu^2 + (1 - \lambda) \eta^2)$. If the first inequality in (2.20) is multiplied by $(\mathbf{z}^T \mathbf{s}_\theta^{\text{opt}}) / (\mathbf{z}^T \tilde{\mathbf{g}}_\theta)$, which is always negative due to the condition (2.16) in the proposition, (2.20) becomes

$$(\mathbf{z}^T \mathbf{H}_\theta \mathbf{z})(\mathbf{z}^T \mathbf{s}_\theta^{\text{opt}}) / (\mathbf{z}^T \tilde{\mathbf{g}}_\theta) + k(\mathbf{z}^T \mathbf{s}_\theta^{\text{opt}}) > 0 \quad \text{and} \quad \|\mathbf{z}\|^2 + k(\mathbf{z}^T \mathbf{s}_\theta^{\text{opt}}) < 0. \quad (2.21)$$

Since k can take any real value by adjusting $\lambda \in [0, 1]$ and infinitesimally small η and ν values, it is guaranteed that both inequalities in (2.21) can be satisfied if $(\mathbf{z}^T \mathbf{H}_\theta \mathbf{z})(\mathbf{z}^T \mathbf{s}_\theta^{\text{opt}}) / (\mathbf{z}^T \tilde{\mathbf{g}}_\theta)$ is larger than $\|\mathbf{z}\|^2$, which corresponds to (2.17). ■

Remark 2.2 For the conditions in (2.16) and (2.17) to be satisfied, $g_{\theta}(\mathbf{x})$ must be concave at $\mathbf{x} = \mathbf{s}_{\theta}^{\text{opt}}$ (i.e., $\mathbf{H}_{\theta}$ must be negative-definite) since $\|\mathbf{z}\|^2$ is always nonnegative and $(\mathbf{z}^T \mathbf{s}_{\theta}^{\text{opt}})/(\mathbf{z}^T \nabla g_{\theta}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\theta}^{\text{opt}}})$ is negative due to (2.16).

Proposition 2.3 provides a simple approach, based on the first and second order derivatives of g_{θ} , to determine if the stochastic parameter design can provide improvements over the deterministic one. If the conditions are satisfied, the improvements are guaranteed and the optimization problem in (2.12) or (2.13) can be solved to obtain the optimal solution. However, since the conditions are sufficient but not necessary, there can also exist certain scenarios in which improvements are observed although the conditions are not satisfied. Examples for various scenarios are provided in the next section.

2.3 Numerical Examples

In order to present examples of the theoretical results in the previous sections, consider an estimation problem in which a scalar parameter θ is estimated based on observation y that is modeled as $y = s_{\theta} + n$, with n denoting the additive noise component. (Although a scalar problem is considered for convenience, vector parameter estimation problems can be treated in a similar fashion (per component) when the noise components are independent and the cost function is additive [2].) The noise n is modeled by a Gaussian mixture distribution, specified as $p_n(n) = \sum_{l=1}^L \gamma_l \exp\{-(n - \mu_l)^2/(2\sigma_l^2)\}/(\sqrt{2\pi}\sigma_l)$, where the parameters are chosen in such a way to generate a zero-mean noise component. In addition, the estimator is given by $\hat{\theta}(y) = y$, and the cost function is selected as the uniform cost function, which is expressed as $C[\hat{\theta}(y), \theta] = 1$ if $|\hat{\theta}(y) - \theta| > \Delta$ and $C[\hat{\theta}(y), \theta] = 0$ otherwise. Based on this model, g_{θ} in (2.5) can be obtained as

$$g_{\theta}(\mathbf{x}) = \sum_{l=1}^L \gamma_l \left(Q\left(\frac{\mathbf{x} - \theta + \mu_l + \Delta}{\sigma_l}\right) + Q\left(\frac{-\mathbf{x} + \theta - \mu_l + \Delta}{\sigma_l}\right) \right) \quad (2.22)$$

where $Q(\mathbf{x}) = (1/\sqrt{2\pi}) \int_{\mathbf{x}}^{\infty} \exp\{-t^2/2\} dt$ denotes the Q -function. Regarding the constraint in (2.8), $E\{|s_{\theta}|^2\} \leq \theta^2$ is considered for each θ .

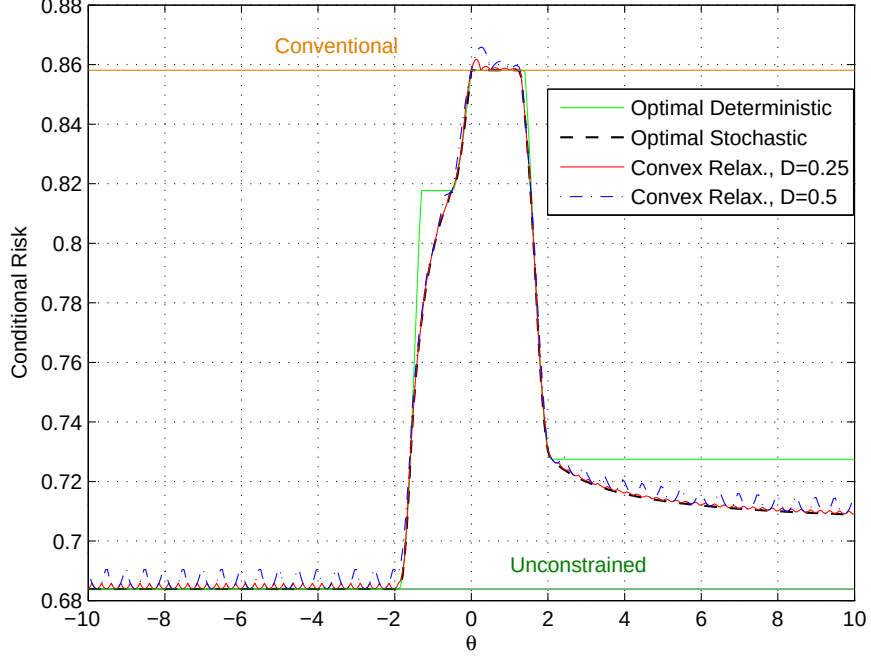


Figure 2.2: Conditional risk versus θ for optimal deterministic and optimal stochastic approaches. Also results of convex relaxation is presented for two different values of D .

For the numerical examples, parameter θ is modeled to lie between -10 and 10 ; that is, the parameter space is specified as $\Lambda = [-10, 10]$. Also, s_θ can take values in the interval $[-10, 10]$ under the average power constraint, $E\{|s_\theta|^2\} \leq \theta^2$. In addition, the parameters of the Gaussian mixture noise n are selected as $\gamma_1 = 0.33$, $\gamma_2 = 0.13$, $\gamma_3 = 0.08$, $\gamma_4 = 0.07$, $\gamma_5 = 0.11$, $\gamma_6 = 0.28$, $\mu_1 = -3.8$, $\mu_2 = -1.6$, $\mu_3 = -0.51$, $\mu_4 = 0.4657$, $\mu_5 = 2.42$, $\mu_6 = 4.3$, and $\sigma_l = 0.5$, $\forall l$. With this selection of the parameters, the noise becomes a zero-mean random variable so that $\hat{\theta}(y) = y$ can be regarded as a practical estimator.³ Finally, $\Delta = 1$ is considered for the uniform cost function described in the previous paragraph.

In Figure 2.2, the conditional risks (i.e., $E\{g_\theta(s_\theta)\}$ in (2.6)) are plotted versus

³Although this is not an optimal estimator, it can be used in practice due to its simplicity compared to the optimal estimator, which would have high complexity due to the multimodal noise structure.

θ for various parameter design approaches. For the optimal stochastic parameter design, both the exact solution obtained from (2.12) and the convex relaxation solutions obtained from (2.13) are plotted. In the convex relaxation approach, the set of possible values for s_θ are selected between -10 and 10 with an increment of D (in short, $-10 : D : 10$), and the results for $D = 0.25$ and $D = 0.5$ are illustrated in the figure. The results for the optimal deterministic parameter design are calculated from (2.14). In addition, the results obtained from the unconstrained problem (see (2.7)) and those obtained by using $p_{s_\theta}(x) = \delta(x - \theta)$ (labeled as “Conventional”) are shown in the figure to provide performance benchmarks. It is observed that the optimal stochastic parameter design achieves the minimum conditional risks for all θ values in the presence of the average power constraint. It provides performance improvements over the deterministic parameter design for certain range of parameter values, e.g., for $\theta > 2.1$. In addition, both the stochastic and the deterministic design approaches achieve the same conditional risks as the unconstrained solution for some θ values, which is due to the fact that the unconstrained solutions satisfy the average power constraint for those values of θ . Furthermore, the convex relaxation approaches (which provide low complexity solutions) perform very closely to the exact solutions of the optimal stochastic parameter design problem for small values of D .

In order to provide further explanations of the results in Figure 2.2, Figure 2.3 illustrates $g_\theta(x)$ in (2.22) for $\theta = -5$, $\theta = 0$, and $\theta = 5$. As expected from the expression in (2.22), each function in the figure is a shifted version of the others. Also, this figure can be used to determine when the unconstrained solution coincides with the solutions of the optimal stochastic and the optimal deterministic parameter designs. For example, for $\theta = -5$, the global minimum of $g_\theta(x)$ is achieved at -1.223 , which already satisfies the constraint. Therefore, all the three approaches yield the same conditional risk for that parameter (see Figure 2.2). On the other hand, for $\theta = 5$, the global minimum is at 8.777 ; hence, the conditional risk obtained from the unconstrained problem in (2.7) cannot be achieved by the constrained approaches. Specifically, the optimal deterministic approach in (2.14) chooses the minimum value in the interval $[-5, 5]$, which results in the optimal signal value of $s_\theta^{\text{opt}} = 0.81$. On the other hand, the solution of the optimal

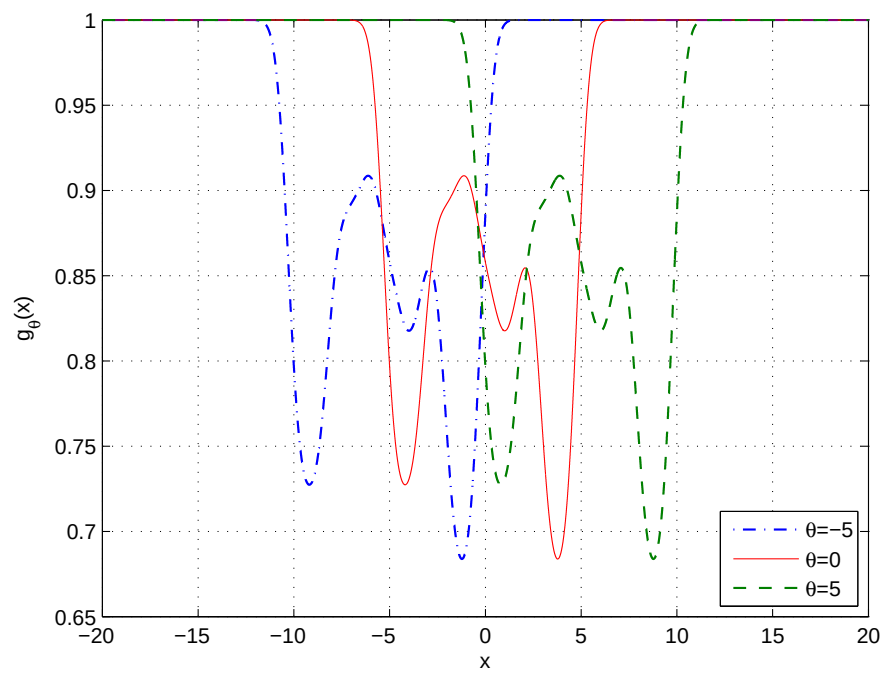


Figure 2.3: $g_\theta(x)$ in (2.22) for various values of θ .

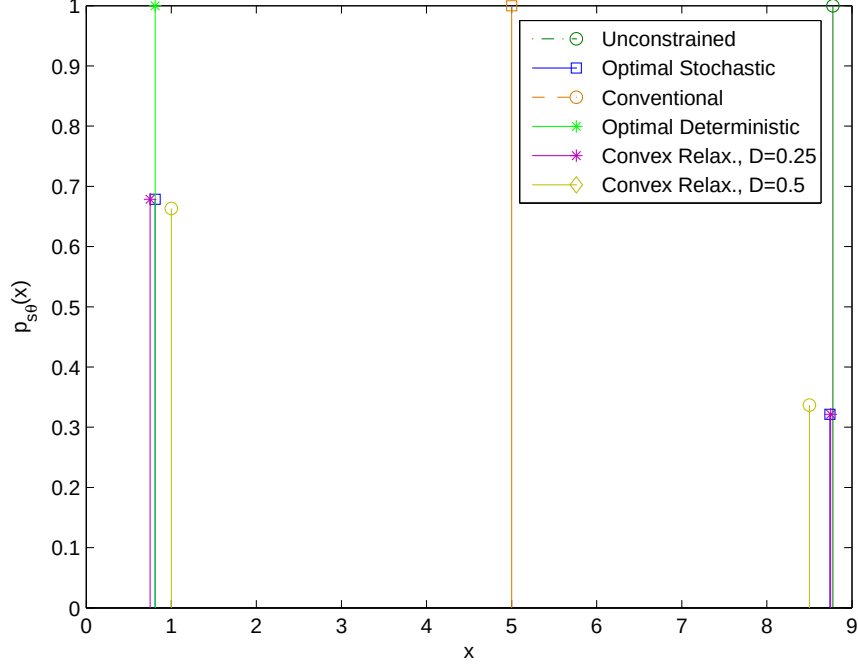


Figure 2.4: Probability mass functions of the stochastic signal s_θ for $\theta = 5$ for unconstrained, optimal deterministic, optimal stochastic and conventional approaches. Also results of convex relaxation is presented for two different values of D .

stochastic parameter design problem in (2.12) results in a randomization between 8.741 and 0.809 with probabilities of 0.321 and 0.679, respectively, and achieves a lower conditional risk than the deterministic approach (see Figure 2.2).

In Table 2.1, the optimal solutions for the optimal stochastic, the optimal deterministic and the unconstrained parameter design approaches are presented for various values of θ . Figure 2.4 shows the probability mass function of the stochastic signal s_θ for various parameter design approaches when $\theta = 5$. Figure 2.3 can also be used to explain the oscillatory behavior of the convex relaxation solutions in Figure 2.2. Since the convex relaxation approach considers possible s_θ values as $-10 : D : 10$ and since $g_\theta(x)$ shifts with θ , the signal values obtained from the convex optimization problem in (2.13) move around the optimal values of the exact

Table 2.1: Optimal stochastic solution $p_{s_\theta}^{\text{opt}}(x) = \lambda_\theta \delta(x - s_{\theta,1}) + (1 - \lambda_\theta) \delta(x - s_{\theta,2})$, optimal deterministic solution s_θ^{opt} , and unconstrained solution s_θ^{unc} for various values of θ .

θ	λ_θ	$s_{\theta,1}$	$s_{\theta,2}$	s_θ^{opt}	s_θ^{unc}
-5	1	-1.223	-	-1.223	-1.223
-3	1	0.777	-	0.777	0.777
-1.5	0.295	-0.331	1.774	1.5	2.277
0	1	0	-	0	3.777
1.5	0.42	-0.294	-1.954	-1.5	5.277
3	0.826	-1.177	6.719	-1.19	6.777
5	0.679	0.809	8.741	0.81	8.777

solution periodically. Finally, the conditions in Proposition 2.3 are evaluated for different θ values, and it is observed that they provide sufficient but not necessary conditions for specifying improvements via the stochastic parameter design over the deterministic one. For example, the calculations show that the conditions in Proposition 2.3 are satisfied for $\theta \in [-1.381, -1.31]$ and $\theta \in [1.397, 1.536]$, and improvements are observed in Figure 2.2 for those values of θ .

2.4 Conclusions for Chapter 2

In this chapter, the optimal stochastic design of a single parameter has been studied for a fixed estimator. It has been shown that the optimal stochastic parameter design results in time sharing (randomization) among at most two different signal values. In addition, a convex relaxation of the optimal parameter design problem, which results in linearly constrained linear programming, has been presented. Furthermore, sufficient conditions under which the stochastic parameter design can or cannot provide improvements over the deterministic parameter design have been obtained. Finally, the theoretical results have been illustrated through the numerical examples.

Chapter 3

Optimal Stochastic Design of Multiple Parameters for Estimation Problems

The problem considered in Chapter 2 is extended to the case of multiple parameters in this chapter. This chapter is organized as follows: In Section 3.1, the problem formulation is introduced and the optimal randomization strategies are obtained. In Section 3.2, some properties of the optimal stochastic parameter design approaches are discussed. Sufficient conditions are derived in Section 3.3 in order to specify when the stochastic parameter design or the deterministic parameter design is optimal. After the numerical examples in Section 3.4, concluding remarks are made in Section 3.5.

3.1 Stochastic Design for Multi-Parameter Estimation

In this section, we establish a framework for the stochastic design of multiple parameters for a given set of fixed estimators. Consider a parameter estimation

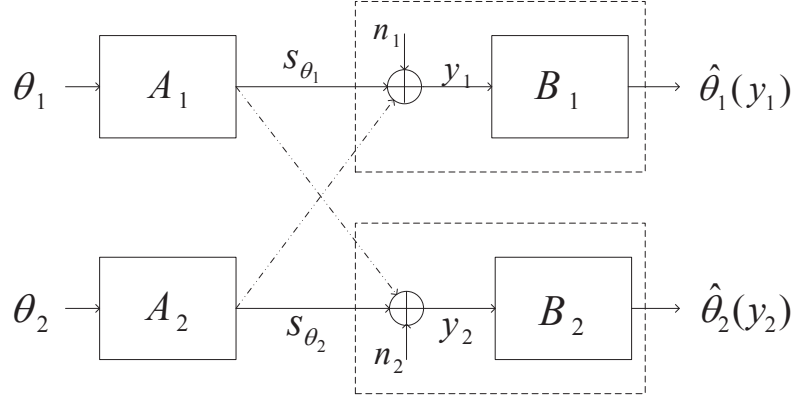


Figure 3.1: System model for $K = 2$. Devices A_1 and A_2 transmit stochastic signals $\mathbf{s}_{\theta_1}$ and $\mathbf{s}_{\theta_2}$ for each value of parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. Devices B_1 and B_2 estimate $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ based on the noise and interference corrupted version of $\mathbf{s}_{\theta_1}$ and $\mathbf{s}_{\theta_2}$, respectively.

scenario in which there exist K parameters denoted by $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K$, where each parameter resides in $\mathbb{R}^M$. Parameter $\boldsymbol{\theta}_i$ is transmitted by device A_i , which can transmit any signal $\mathbf{s}_{\theta_i}$ related to $\boldsymbol{\theta}_i$, where $i \in \{1, \dots, K\}$. The transmitted signal $\mathbf{s}_{\theta_i}$ is corrupted by both additive noise and the interference from other transmitted signals, and device B_i tries to estimate the unknown parameter $\boldsymbol{\theta}_i$ based on the noise and interference corrupted signal. An example system is depicted in Figure 3.1 for $K = 2$. It should be emphasized that parameter $\boldsymbol{\theta}_i$ is not necessarily transmitted as it is; instead, device A_i can transmit any function of $\boldsymbol{\theta}_i$, say $\mathbf{s}_{\theta_i}$. In addition, function $\mathbf{s}_{\theta_i}$ can be of any type; it can be a deterministic function of $\boldsymbol{\theta}_i$, or it can be a stochastic function. The aim of this study is to find the optimal $\mathbf{s}_{\theta_i}$, i.e., the optimal probability distribution of $\mathbf{s}_{\theta_i}$, for each $\boldsymbol{\theta}_i$.

It is noted that the difference between the single parameter case studied in Chapter 2 and the multi-parameter case investigated in this chapter is not only related to the number of parameters. The proposed multi-parameter formulation in this chapter also takes into account the possible interference among the parameter related signals, as shown by the dashed cross lines in Figure 3.1. Considering

K parameters, the received signal (observation) at device B_i can be expressed as

$$\mathbf{y}_i = \mathbf{s}_{\boldsymbol{\theta}_i} + \sum_{\substack{j=1 \\ j \neq i}}^K \rho_{ij} \mathbf{s}_{\boldsymbol{\theta}_j} + \mathbf{n}_i \quad (3.1)$$

for $i \in \{1, \dots, K\}$, where ρ_{ij} is the multiplier that is set according to the interference between the parameter related signals for the i^{th} and j^{th} parameters, and $\mathbf{n}_i$ represents the channel noise, which has a probability density function (PDF) denoted by $p_{\mathbf{n}_i}(\cdot)$. Each device B_i tries to estimate $\boldsymbol{\theta}_i$ based on the corresponding observation $\mathbf{y}_i$ in (3.1). It is assumed that the devices employ fixed estimators specified by $\hat{\boldsymbol{\theta}}_i(\mathbf{y}_i)$ in order to estimate $\boldsymbol{\theta}_i$. Let $\boldsymbol{\theta}$ denote the overall parameter vector, which is defined as $\boldsymbol{\theta} \triangleq [\boldsymbol{\theta}_1^T \cdots \boldsymbol{\theta}_K^T]^T$. The prior distribution of $\boldsymbol{\theta}$ is represented by $w(\boldsymbol{\theta})$, and the parameter space in which $\boldsymbol{\theta}$ resides is denoted by Λ .

The aim is to obtain the optimal probability distributions of $\mathbf{s}_{\boldsymbol{\theta}}$ for each $\boldsymbol{\theta} \in \Lambda$ in order to minimize a function of the Bayes risk for the given estimators, where $\mathbf{s}_{\boldsymbol{\theta}} \triangleq [\mathbf{s}_{\boldsymbol{\theta}_1}^T \cdots \mathbf{s}_{\boldsymbol{\theta}_K}^T]^T$. Since the parameters can interfere with each other, the optimization cannot be performed independently for each parameter in general; therefore, a joint optimization should be performed.

3.1.1 Unconstrained Optimization

In order to obtain the optimal probability distribution of $\mathbf{s}_{\boldsymbol{\theta}}$, a certain objective function is considered, and the optimization is performed by minimizing that function over the PDF of $\mathbf{s}_{\boldsymbol{\theta}}$. In this section, no constraints are considered in formulating the optimization problem. In this scenario, the optimal stochastic parameter design can be formulated as

$$\{p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}, \boldsymbol{\theta} \in \Lambda\} = \arg \min_{\{p_{\mathbf{s}_{\boldsymbol{\theta}}}, \boldsymbol{\theta} \in \Lambda\}} r(\hat{\boldsymbol{\theta}}) \quad (3.2)$$

where $\{p_{\mathbf{s}_{\boldsymbol{\theta}}}, \boldsymbol{\theta} \in \Lambda\}$ represents the set of PDFs for $\mathbf{s}_{\boldsymbol{\theta}}$ for all possible values of parameter $\boldsymbol{\theta}$, and $r(\hat{\boldsymbol{\theta}})$ is the objective function for the overall system. For the single parameter case considered in Chapter 2, the Bayes risk of the estimator

was a natural choice for this objective function. On the other hand, it is possible to consider various risk functions for the multi-parameter case. In this section, two different objective functions are considered. The first one is the sum of the Bayes risks of the K estimators in the system (called the *total Bayes risk*), and the second one is the maximum of the Bayes risks of the estimators (called the *maximum Bayes risk*). For both of these objective functions, the Bayes risk of each estimator should be calculated first. For the two parameter case, the Bayes risk of the first estimator is expressed as

$$r(\hat{\boldsymbol{\theta}}_1) = \int_{\Lambda_1} w(\boldsymbol{\theta}_1) \int p_{\mathbf{s}_{\boldsymbol{\theta}_1}}(\mathbf{x}_1) \int C[\hat{\boldsymbol{\theta}}(\mathbf{y}_1), \boldsymbol{\theta}_1] \\ \times \int p_{\mathbf{s}_{\boldsymbol{\theta}_2}}(\mathbf{x}_2) p_{\mathbf{n}_1}(\mathbf{y}_1 - \mathbf{x}_1 - \rho_{12}\mathbf{x}_2) d\mathbf{x}_2 d\mathbf{y}_1 d\mathbf{x}_1 d\boldsymbol{\theta}_1 \quad (3.3)$$

where $C[\hat{\boldsymbol{\theta}}(\mathbf{y}_1), \boldsymbol{\theta}_1]$ denotes the cost of estimating $\boldsymbol{\theta}_1$ as $\hat{\boldsymbol{\theta}}(\mathbf{y}_1)$ [2], and $p_{\mathbf{s}_{\boldsymbol{\theta}_i}}$ is the PDF of the signal related to parameter i . (The Bayes risk of the second estimator can be expressed in a similar fashion.)

Defining an auxiliary function $g_{\boldsymbol{\theta}_1}(\mathbf{x})$ for the first estimator as

$$g_{\boldsymbol{\theta}_1}(\mathbf{x}) \triangleq \int C[\hat{\boldsymbol{\theta}}_1(\mathbf{y}_1), \boldsymbol{\theta}_1] p_{\mathbf{n}_1}(\mathbf{y}_1 - \mathbf{x}_1 - \rho_{12}\mathbf{x}_2) d\mathbf{y}_1 \quad (3.4)$$

where $\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T]^T$, and a similar function for the second estimator, the total Bayes risk can be expressed as

$$r(\hat{\boldsymbol{\theta}}) = \int_{\Lambda} w(\boldsymbol{\theta}) \int p_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) (g_{\boldsymbol{\theta}_1}(\mathbf{x}) + g_{\boldsymbol{\theta}_2}(\mathbf{x})) d\mathbf{x} d\boldsymbol{\theta} \\ = \int_{\Lambda} w(\boldsymbol{\theta}) E\{\tilde{g}_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\} d\boldsymbol{\theta} \quad (3.5)$$

with $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_1^T \ \hat{\boldsymbol{\theta}}_2^T]^T$, $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T \ \boldsymbol{\theta}_2^T]^T$, $\mathbf{s}_{\boldsymbol{\theta}} = [\mathbf{s}_{\boldsymbol{\theta}_1}^T \ \mathbf{s}_{\boldsymbol{\theta}_2}^T]^T$, and

$$\tilde{g}_{\boldsymbol{\theta}}(\mathbf{x}) = g_{\boldsymbol{\theta}_1}(\mathbf{x}) + g_{\boldsymbol{\theta}_2}(\mathbf{x}). \quad (3.6)$$

For the K parameter case, similar expressions can be obtained by updating (3.3) and (3.4) in order to include the interference due to the other parameters as well. In that case, (3.5) still has the same form with the updated definition of $\tilde{g}_{\boldsymbol{\theta}}$ which is given by $\tilde{g}_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{i=1}^K g_{\boldsymbol{\theta}_i}(\mathbf{x})$.

Each expectation operation in (3.5) is over the PDF of $\mathbf{s}_\theta$ for a given value of θ . In the absence of constraints on the design of $\mathbf{s}_\theta$, $r(\hat{\theta})$ given by (3.5) can be minimized if the PDF of $\mathbf{s}_\theta$ assigns all the probability to the minimizer of $\tilde{g}_\theta$ in (3.6) for each θ .¹ In other words, the solution of the optimization problem in (3.2) for the total Bayes risk constraint in (3.5) can be obtained as

$$p_{\mathbf{s}_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{unc}}) , \quad \mathbf{s}_\theta^{\text{unc}} = \arg \min_{\mathbf{x}} \tilde{g}_\theta(\mathbf{x}) \quad (3.7)$$

for all $\theta \in \Lambda$, where δ denotes the Dirac delta function. As the solution corresponds to assigning all the probability to a single point, it is concluded that optimal PDFs for the stochastic parameter design are the ones with single point masses. Hence, the deterministic parameter design is optimal and there is no need for stochastic modeling in this scenario. Also it can be observed from (3.7) that the solution is independent of the prior distribution $w(\theta)$ as the optimal solution is obtained for each θ separately.

When the maximum Bayes risk criterion is considered, the objective function in (3.5) can be updated as

$$\begin{aligned} r(\hat{\theta}) &= \int_{\Lambda} w(\theta) \max_{i \in \{1, \dots, K\}} \left(\int p_{\mathbf{s}_\theta}(\mathbf{x}) g_{\theta_i}(\mathbf{x}) d\mathbf{x} \right) d\theta \\ &= \int_{\Lambda} w(\theta) \max_{i \in \{1, \dots, K\}} (\mathbb{E}\{g_{\theta_i}(\mathbf{s}_\theta)\}) d\theta . \end{aligned} \quad (3.8)$$

Based on similar arguments to those employed above for the total Bayes risk criterion, it can be observed that the solution is independent of the prior distribution $w(\theta)$ and the optimal solution can be obtained for each θ separately. Hence, the optimization problem for the maximum Bayes risk criterion can be formulated as follows:

$$p_{\mathbf{s}_\theta}^{\text{opt}} = \arg \min_{p_{\mathbf{s}_\theta}} \max_{i \in \{1, \dots, K\}} \mathbb{E}\{g_{\theta_i}(\mathbf{s}_\theta)\} . \quad (3.9)$$

A different problem that is in the same form as (3.9) is studied in [7]. Based on Proposition 1 in [7], it is concluded that the optimal solution of (3.9) corresponds to a discrete random variable with at most K point masses for each θ . Based on

¹In the case of multiple minimizers, any (combination) of them can be chosen for the optimal solution.

this result, the optimal stochastic parameter design problem for the maximum Bayes risk criterion can be expressed as

$$\begin{aligned} \min_{\{\lambda_{\boldsymbol{\theta},j}, \mathbf{s}_{\boldsymbol{\theta},j}\}_{j=1}^K} \max_{i \in \{1, \dots, K\}} \sum_{j=1}^K \lambda_{\boldsymbol{\theta},j} g_{\boldsymbol{\theta}_i}(\mathbf{s}_{\boldsymbol{\theta},j}) \\ \text{subject to } \sum_{j=1}^K \lambda_{\boldsymbol{\theta},j} = 1, \lambda_{\boldsymbol{\theta},j} \in [0, 1], \forall j \in \{1, \dots, K\} \end{aligned} \quad (3.10)$$

for $\boldsymbol{\theta} \in \Lambda$, where $\mathbf{s}_{\boldsymbol{\theta}}$ takes the value of $\mathbf{s}_{\boldsymbol{\theta},j}$ with probability $\lambda_{\boldsymbol{\theta},j}$ for $j = 1, \dots, K$. Compared to (3.9), the formulation in (3.10) provides a significant reduction in computational complexity as it requires optimization over a finite number of variables instead of over all possible PDFs. Since generic cost functions and noise distributions are considered in the theoretical analysis, function $g_{\boldsymbol{\theta}_i}$ in (3.4) is generic as well; hence, the optimization problem in (3.15) can be nonconvex in general. Therefore, global optimization techniques such as particle swarm optimization (PSO) or differential evolution can be employed to obtain the solution [42, 43].

3.1.2 Constrained Optimization

In this section, an average power constraint is considered in the formulation of the stochastic parameter design problem. Although this is a specific type of a constraint, other types of constraints can also be incorporated into the theoretical analysis in a similar fashion. It should be emphasized that there exist power constraints in almost all practical applications since otherwise it would be possible to transmit signals with very high power to reduce the objective function of the system arbitrarily.

Consider the average power constraint stated as

$$\mathbb{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \leq A_{\boldsymbol{\theta}} \quad (3.11)$$

for $\boldsymbol{\theta} \in \Lambda$, where $\|\mathbf{s}_{\boldsymbol{\theta}}\|$ is the Euclidean norm of vector $\mathbf{s}_{\boldsymbol{\theta}}$, and $A_{\boldsymbol{\theta}}$ represents the average power limit for $\boldsymbol{\theta}$. In general, constraint $A_{\boldsymbol{\theta}}$ can be a function of $\boldsymbol{\theta}$ as

well. From (3.5) and (3.11), the optimal stochastic parameter design problem for the total Bayes risk criterion can be expressed as

$$\begin{aligned} & \min_{\{p_{\mathbf{s}_\theta}, \boldsymbol{\theta} \in \Lambda\}} \int_{\Lambda} w(\boldsymbol{\theta}) \mathbb{E}\{\tilde{g}_\theta(\mathbf{s}_\theta)\} d\boldsymbol{\theta} \\ & \text{subject to } \mathbb{E}\{\|\mathbf{s}_\theta\|^2\} \leq A_\theta, \quad \forall \boldsymbol{\theta} \in \Lambda \end{aligned} \quad (3.12)$$

where $\tilde{g}_\theta(\cdot)$ is as defined in (3.6). Due to the structure of the objective function and the constraint, the constrained optimization problem in (3.12) can be solved individually for each $\boldsymbol{\theta}$ as

$$\min_{p_{\mathbf{s}_\theta}} \mathbb{E}\{\tilde{g}_\theta(\mathbf{s}_\theta)\} \quad \text{subject to } \mathbb{E}\{\|\mathbf{s}_\theta\|^2\} \leq A_\theta \quad (3.13)$$

for $\boldsymbol{\theta} \in \Lambda$. Therefore, the solution does not depend on the prior distribution $w(\boldsymbol{\theta})$.

When the maximum Bayes risk criterion is considered, it can be obtained from (3.8) and (3.11) that the problem becomes

$$\min_{p_{\mathbf{s}_\theta}} \max_{i \in \{1, \dots, K\}} \mathbb{E}\{g_{\theta_i}(\mathbf{s}_\theta)\} \quad \text{subject to } \mathbb{E}\{\|\mathbf{s}_\theta\|^2\} \leq A_\theta \quad (3.14)$$

for $\boldsymbol{\theta} \in \Lambda$. Similar optimization problems in the form of (3.13) and (3.14) have been investigated in the literature [3, 4, 10]. The problem in (3.13) has the same form as the one considered in Chapter 2. Therefore, the statistical behavior of the optimal solution is the same; that is, the optimal solution can be achieved by a randomization between at most two different values of $\mathbf{s}_\theta$ for each $\boldsymbol{\theta}$, as stated in Proposition 1 in Chapter 2. Then, the optimal solution can be obtained based on a similar approach to that in Chapter 2. Namely, the optimal stochastic parameter design problem for the total Bayes risk criterion can be expressed as

$$\begin{aligned} & \min_{\{\lambda_{\theta,j}, \mathbf{s}_{\theta,j}\}_{j=1}^2} \sum_{j=1}^2 \lambda_{\theta,j} \tilde{g}_\theta(\mathbf{s}_{\theta,j}) \\ & \text{subject to } \sum_{j=1}^2 \lambda_{\theta,j} \|\mathbf{s}_{\theta,j}\|^2 \leq A_\theta, \quad \sum_{j=1}^2 \lambda_{\theta,j} = 1, \\ & \lambda_{\theta,j} \in [0, 1], \quad j \in \{1, 2\} \end{aligned} \quad (3.15)$$

for $\boldsymbol{\theta} \in \Lambda$. On the other hand, the optimization problem in (3.14) has a different form than that in Chapter 2. Based on arguments similar to those in [46], the following result is obtained.

Proposition 3.1 *Suppose that functions g_{θ_i} for $i \in \{1, \dots, K\}$ are continuous, and each component of $\mathbf{s}_{\boldsymbol{\theta}}$ resides in a finite closed interval. Then, the optimal solution of (3.14) can be characterized by the following probability density:*

$$p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}(\mathbf{x}) = \sum_{j=1}^{K+1} \lambda_{\boldsymbol{\theta},j} \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},j}) \quad (3.16)$$

where $\lambda_{\boldsymbol{\theta},j} \geq 0$ and $\sum_{j=1}^{K+1} \lambda_{\boldsymbol{\theta},j} = 1$.

Proof: Consider the set of all possible $(g_{\theta_1}(\mathbf{s}_{\boldsymbol{\theta}}), \dots, g_{\theta_K}(\mathbf{s}_{\boldsymbol{\theta}}), \|\mathbf{s}_{\boldsymbol{\theta}}\|^2)$ values and denote it by U . Similarly, denote by W the set of all possible $(\mathbb{E}\{g_{\theta_1}(\mathbf{s}_{\boldsymbol{\theta}})\}, \dots, \mathbb{E}\{g_{\theta_K}(\mathbf{s}_{\boldsymbol{\theta}})\}, \mathbb{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\})$ values. That is, $U = \{(u_1, \dots, u_{K+1}) : u_1 = g_{\theta_1}(\mathbf{s}_{\boldsymbol{\theta}}), \dots, u_K = g_{\theta_K}(\mathbf{s}_{\boldsymbol{\theta}}), u_{K+1} = \|\mathbf{s}_{\boldsymbol{\theta}}\|^2, \forall \mathbf{s}_{\boldsymbol{\theta}}\}$ and $W = \{(w_1, \dots, w_{K+1}) : w_1 = \mathbb{E}\{g_{\theta_1}(\mathbf{s}_{\boldsymbol{\theta}})\}, \dots, w_K = \mathbb{E}\{g_{\theta_K}(\mathbf{s}_{\boldsymbol{\theta}})\}, w_{K+1} = \mathbb{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\}, \forall p_{\mathbf{s}_{\boldsymbol{\theta}}}\}$. As in [3], [10] and [46], it can be concluded that the convex hull of U is equal to W . Then, based on Carathéodory's theorem [41], it is stated that any point in W can be expressed as a convex combination of at most $(K+2)$ points in U . In addition, as an optimal PDF should achieve the minimum value, it must correspond to the boundary of W , which results in a convex combination of at most $(K+1)$ points in U . Therefore, an optimal solution can be expressed as stated in (3.16). ■

Proposition 3.1 states that the optimal solution can be achieved by a randomization among at most $K+1$ different values of $\mathbf{s}_{\boldsymbol{\theta}}$ for each $\boldsymbol{\theta}$. Based on this result, the optimal stochastic parameter design problem for the maximum Bayes risk criterion can be expressed as

$$\begin{aligned} \min_{\{\lambda_{\boldsymbol{\theta},j}, \mathbf{s}_{\boldsymbol{\theta},j}\}_{j=1}^{K+1}} \quad & \max_{i \in \{1, \dots, K\}} \sum_{j=1}^{K+1} \lambda_{\boldsymbol{\theta},j} g_{\theta_i}(\mathbf{s}_{\boldsymbol{\theta},j}) \\ \text{subject to} \quad & \sum_{j=1}^{K+1} \lambda_{\boldsymbol{\theta},j} \|\mathbf{s}_{\boldsymbol{\theta},j}\|^2 \leq A_{\boldsymbol{\theta}}, \quad \sum_{j=1}^{K+1} \lambda_{\boldsymbol{\theta},j} = 1, \\ & \lambda_{\boldsymbol{\theta},j} \in [0, 1], \quad j \in \{1, \dots, K+1\} \end{aligned} \quad (3.17)$$

for $\boldsymbol{\theta} \in \Lambda$.

From (3.15) and (3.17), it is concluded that randomization (time sharing) of transmitted signal values may offer improvements in the presence of an average

power constraint for both the total Bayes risk and the maximum Bayes risk criteria.

3.2 Characterization of Optimal Stochastic Parameter Design in the Presence of Average Power Constraint

In this section, we discuss some properties of the optimal stochastic parameter design approaches in the presence of average power constraints. Namely, the average transmitted signal powers corresponding to the optimal parameter design approaches are investigated, and the characterization of the optimal approaches is provided in various scenarios.

For the total Bayes risk criterion, the following two results are obtained when the *stochastic* parameter design is the solution of (3.15) (equivalently, (3.13)); that is, when the optimal solution involves randomization between two different signal values.²

Lemma 3.1 *Assume that the solution of (3.15) involves randomization between two different signals. Then, one of the signals has a power below the average power limit, and the other signal has a power above the average power limit.*

Proof: This result can be proved via contradiction as follows. First, assume that the powers of both signals are smaller than or equal to the average power limit. Then, the solution cannot be a randomization between these two signals since employing the signal with the lower risk (i.e., lower $\tilde{g}_{\theta}$) exclusively achieves a lower total Bayes risk (see (3.15)) than performing randomization between these signals. Hence, this scenario is not possible under the assumption in the

²In this study, the statement “the optimal solution involves randomization between two different signal values” is used to mean that there is no deterministic solution that achieves the same performance as the optimal stochastic solution.

lemma. Second, assume that either the powers of both signals are larger than the average power limit, or the power of one signal is equal to and that of the other is larger than the average power limit. In this scenario, the average power constraint in (3.15) is violated; hence, this cannot be a valid scenario. Therefore, it is concluded that if randomization between two different signals is the solution of (3.13), then one of the signals must have a power below the average power limit, and the other signal must have a power above the average power limit. ■

Lemma 3.2 *Assume that the solution of (3.15) involves randomization between two different signals. Then, the signal with the higher (lower) power has a lower (higher) risk than the other signal.*

Proof: This result can also be proved via contradiction. If the signal with the lower power has a risk which is smaller than or equal to the risk of the other signal, then there is no need for randomization. In that case, employing this signal exclusively yields a lower risk; hence, randomization between these signals cannot be optimal. Therefore, if randomization between two signals is the solution of (3.13), then the signal with the higher (lower) power must have a lower (higher) risk than the other signal. ■

Based on Lemma 3.1 and Lemma 3.2, the following result is obtained.

Proposition 3.2 *If the solution of (3.15) (equivalently, (3.13)) involves randomization between two different signals; that is, if stochastic parameter design is optimal, then the average signal power is equal to the average power limit; i.e., the solution operates at the average power limit.*

Proof: In order to prove the claim in the proposition, suppose that $\{\lambda_{\theta,j}, \mathbf{s}_{\theta,j}\}_{j=1}^2$ is an optimal solution and utilizes a power strictly lower than the average power limit; i.e., $\lambda_{\theta,1}\|\mathbf{s}_{\theta,1}\|^2 + (1 - \lambda_{\theta,1})\|\mathbf{s}_{\theta,2}\|^2 < A_{\theta}$. Without loss of generality, assume that $\|\mathbf{s}_{\theta,1}\|^2 > A_{\theta}$ and $\|\mathbf{s}_{\theta,2}\|^2 < A_{\theta}$ as a result of Lemma 3.1. According to Lemma 3.2, $\tilde{g}_{\theta}(\mathbf{s}_{\theta,1}) < \tilde{g}_{\theta}(\mathbf{s}_{\theta,2})$ is satisfied. Next, consider another

solution $\{\lambda'_{\theta,j}, \mathbf{s}_{\theta,j}\}_{j=1}^2$ with $\lambda'_{\theta,1} = (A_{\theta} - \|\mathbf{s}_{\theta,2}\|^2)/(\|\mathbf{s}_{\theta,1}\|^2 - \|\mathbf{s}_{\theta,2}\|^2)$. Note that the average power for this solution is equal to the average power limit; that is, $\lambda'_{\theta,1}\|\mathbf{s}_{\theta,1}\|^2 + (1 - \lambda'_{\theta,1})\|\mathbf{s}_{\theta,2}\|^2 = A_{\theta}$. In addition, it can be shown that $\lambda'_{\theta,1} > \lambda_{\theta,1}$ as $\|\mathbf{s}_{\theta,1}\|^2 > A_{\theta}$, $\|\mathbf{s}_{\theta,2}\|^2 < A_{\theta}$, and the average power of solution $\{\lambda'_{\theta,j}, \mathbf{s}_{\theta,j}\}_{j=1}^2$ is larger than that of solution $\{\lambda_{\theta,j}, \mathbf{s}_{\theta,j}\}_{j=1}^2$. Since $\tilde{g}_{\theta}(\mathbf{s}_{\theta,1}) < \tilde{g}_{\theta}(\mathbf{s}_{\theta,2})$ due to Lemma 3.2 and $\lambda'_{\theta,1} > \lambda_{\theta,1}$, it can be shown that solution $\{\lambda'_{\theta,j}, \mathbf{s}_{\theta,j}\}_{j=1}^2$ achieves a lower total Bayes risk than solution $\{\lambda_{\theta,j}, \mathbf{s}_{\theta,j}\}_{j=1}^2$; that is,

$$\lambda'_{\theta,1}\tilde{g}_{\theta}(\mathbf{s}_{\theta,1}) + (1 - \lambda'_{\theta,1})\tilde{g}_{\theta}(\mathbf{s}_{\theta,2}) < \lambda_{\theta,1}\tilde{g}_{\theta}(\mathbf{s}_{\theta,1}) + (1 - \lambda_{\theta,1})\tilde{g}_{\theta}(\mathbf{s}_{\theta,2}). \quad (3.18)$$

Based on (3.18), it is concluded that solution $\{\lambda_{\theta,j}, \mathbf{s}_{\theta,j}\}_{j=1}^2$ cannot be optimal, which results in a contradiction. Hence, it is concluded that a solution with an average power lower than the average power limit cannot be optimal for the scenario in the proposition. That is, the solution of (3.15) operates at the average power limit when the stochastic parameter design is optimal. ■

From Proposition 3.2, the solution of (3.15) can be obtained as stated in the following proposition.

Proposition 3.3 *The solution of (3.15) corresponds to either deterministic parameter design or stochastic parameter design, which can be obtained as follows:*

- **Deterministic Parameter Design:** Transmit $\mathbf{s}_{\theta}^{\text{det}}$ exclusively for $\theta \in \Lambda$, where

$$\mathbf{s}_{\theta}^{\text{det}} = \arg \min_{\|\mathbf{s}_{\theta}\|^2 \leq A_{\theta}} \tilde{g}_{\theta}(\mathbf{s}_{\theta}). \quad (3.19)$$

- **Stochastic Parameter Design:** Perform randomization between $\mathbf{s}_{\theta,1}^{\text{opt}}$ and $\mathbf{s}_{\theta,2}^{\text{opt}}$ with time sharing (randomization) factors $(A_{\theta} - \|\mathbf{s}_{\theta,2}^{\text{opt}}\|^2)/(\|\mathbf{s}_{\theta,1}^{\text{opt}}\|^2 - \|\mathbf{s}_{\theta,2}^{\text{opt}}\|^2)$ and $(\|\mathbf{s}_{\theta,1}^{\text{opt}}\|^2 - A_{\theta})/(\|\mathbf{s}_{\theta,1}^{\text{opt}}\|^2 - \|\mathbf{s}_{\theta,2}^{\text{opt}}\|^2)$, respectively, where

$$(\mathbf{s}_{\theta,1}^{\text{opt}}, \mathbf{s}_{\theta,2}^{\text{opt}}) = \arg \min_{\substack{\|\mathbf{s}_{\theta,1}\|^2 > A_{\theta} \\ \|\mathbf{s}_{\theta,2}\|^2 < A_{\theta}}} \frac{A_{\theta} - \|\mathbf{s}_{\theta,2}\|^2}{\|\mathbf{s}_{\theta,1}\|^2 - \|\mathbf{s}_{\theta,2}\|^2} \tilde{g}_{\theta}(\mathbf{s}_{\theta,1}) + \frac{\|\mathbf{s}_{\theta,1}\|^2 - A_{\theta}}{\|\mathbf{s}_{\theta,1}\|^2 - \|\mathbf{s}_{\theta,2}\|^2} \tilde{g}_{\theta}(\mathbf{s}_{\theta,2}) \quad (3.20)$$

for $\theta \in \Lambda$.

The solution of (3.15) is the one ((3.19) or (3.20)) that results in the lower total Bayes risk.

Proof: There exist two possible scenarios for the solution of (3.15). If no randomization is employed, the optimal solution can be obtained as in (3.19), which is called the deterministic parameter design. On the other hand, randomization between two signals can be performed. As stated in Proposition 3.2, the average signal power must be equal to the average power limit in this scenario; that is, $\lambda_{\theta,1}\|\mathbf{s}_{\theta,1}\|^2 + (1 - \lambda_{\theta,1})\|\mathbf{s}_{\theta,2}\|^2 = A_{\theta}$. Therefore, the time sharing (randomization) factors can be calculated as $\lambda_{\theta,1} = (A_{\theta} - \|\mathbf{s}_{\theta,2}\|^2)/(\|\mathbf{s}_{\theta,1}\|^2 - \|\mathbf{s}_{\theta,2}\|^2)$ and $\lambda_{\theta,2} = 1 - \lambda_{\theta,1}$. In addition, from Lemma 3.1, one signal has a power higher than the average power limit and the other signal has a power lower than the average power limit. Hence, the optimization problem in (3.15) can be simplified as the one in (3.20). Finally, it is observed that the solution that achieves the lower risk in (3.19) and (3.20) becomes the solution of (3.15). ■

Proposition 3.3 provides a simple approach for solving (3.15). Namely, the problems in (3.19) and (3.20) are solved, and the one that achieves the lower total Bayes risk becomes the solution of (3.15).

For the maximum Bayes risk criterion, the solution of (3.17) (equivalently, (3.14)) can be characterized as a special form under certain conditions. To that aim, the following lemma is presented first.

Lemma 3.3 *Consider a set of functions, f_i , for $i \in \{1, 2, \dots, K\}$. If minimum value of a certain function, say f_m , is strictly higher than the values of the other functions at the same point, then this point is the solution of the minimax problem; that is,*

$$\min_x \max_i f_i(x) = \min_x f_m(x). \quad (3.21)$$

Proof: Let x' denote the minimizer of $f_m(x)$ and $f_m(x') > f_i(x')$, $\forall i \in \{1, 2, \dots, K\} \setminus \{m\}$, as stated in the lemma. Suppose that x' is not the solution of the minimax problem, and consider another point x^* which yields a lower

value for the minimax problem; that is, $\max_i f_i(x^*) < f_m(x')$. By definition, $\max_i f_i(x^*) \geq f_m(x^*)$. Combining the last two inequalities, it is obtained that $f_m(x^*) < f_m(x')$, which contradicts the fact that the value of f_m is strictly higher than the values of the other functions at x' . Hence, it is concluded that no other point, x^* , can yield a lower value for the minimax problem than x' . ■

Based on Lemma 3.3, the following result is obtained about the solution of the optimal parameter design problem according to the maximum Bayes risk criterion.

Proposition 3.4 *Consider the probability distribution of $\mathbf{s}_\theta$ that minimizes the risk of the m^{th} estimator under the average power constraint, where $m \in \{1, \dots, K\}$. For that probability distribution, if the risk of the m^{th} estimator is strictly higher than the risks of the other estimators, then this distribution is the optimal solution of the minimax problem in (3.14) (equivalently, (3.17)) and it involves randomization between at most two signals.*

Proof: Consider the minimax problem in (3.17). Let the minimum risk of the m^{th} estimator be strictly higher than the risks of the other estimators for the distribution of $\mathbf{s}_\theta$ that minimizes the risk of the m^{th} estimator under the average power constraint; that is,

$$\min_{\{\lambda_{\theta,j}, \mathbf{s}_{\theta,j}\}} \sum_j \lambda_{\theta,j} g_{\theta_m}(\mathbf{s}_{\theta,j}) \triangleq \sum_j \lambda'_{\theta,j} g_{\theta_m}(\mathbf{s}'_{\theta,j}) > \sum_j \lambda'_{\theta,j} g_{\theta_i}(\mathbf{s}'_{\theta,j}) \quad (3.22)$$

for $i \in \{1, \dots, K\} \setminus \{m\}$, where $\{\lambda'_{\theta,j}, \mathbf{s}'_{\theta,j}\}$ denotes the probability distribution of $\mathbf{s}_\theta$ that minimizes the risk of the m^{th} estimator. In this scenario, Lemma 3 implies that the optimal solution for the m^{th} estimator is the solution of the minimax problem as well. Since the optimal solution for a single estimator corresponds to randomization between at most two signals (consider (3.14) and (3.16) as if $K = 1$), the solution of the minimax problem in (3.17) is obtained via randomization between at most two signals under the conditions in the proposition. ■

When the number of parameters is large, it can be difficult to solve the optimization problem in (3.17) since the dimension of the problem is high in that case.

Proposition 3.4 offers a relatively simple test based on the solution of several low dimensional optimization problems before trying to solve this high dimensional optimization problem. If the conditions stated in the proposition are satisfied then there is no need for solving the high dimensional optimization problem.

3.3 Optimality Conditions

In this section, various conditions are derived in order to specify when the stochastic parameter design or the deterministic parameter design is optimal. In order to investigate such optimality conditions, the objective function to be considered should be identified first. In this study, two different objective functions, the total Bayes risk and the maximum Bayes risk, are considered, and the optimality conditions differ for these functions. For the total Bayes risk, the problem can be simplified to minimizing the expectation of a single function, $\tilde{g}_{\theta}$, as given in (3.13). As it was stated in Section 3.1.2, this problem has the same form as the one studied in Chapter 2. Therefore, the optimality conditions proposed in Chapter 2 are valid for the total Bayes risk criterion in this study as well. However, for the maximum Bayes risk criterion, the problem has a different form as given in (3.14); hence, the optimality conditions are different in this scenario. In this section, the optimality conditions are investigated for the maximum Bayes risk criterion.

The optimal parameter design problem presented in (3.14) does not necessarily yield a stochastic solution in all cases. In certain scenarios, the deterministic design is the optimal solution and in such cases the problem in (3.14) can be reformulated as

$$\min_{\mathbf{s}_{\theta}} \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_{\theta}) \quad \text{subject to} \quad \|\mathbf{s}_{\theta}\|^2 \leq A_{\theta} \quad (3.23)$$

where $\mathbf{s}_{\theta}$ is modeled as a deterministic quantity for each θ . Let $\mathbf{s}_{\theta}^{\text{det}}$ represent the minimizer of the optimization problem in (3.23). Then, the minimum Bayes risk achieved by the *optimal deterministic parameter design* is expressed as $r_{\text{det}}(\hat{\theta}) = \int_{\Lambda} w(\theta) \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_{\theta}^{\text{det}}) d\theta$ (c.f. (3.8)). On the other hand, the

minimum Bayes risk achieved by the optimal parameter design is denoted by $r_{\text{sto}}(\hat{\boldsymbol{\theta}}) = \int_{\Lambda} w(\boldsymbol{\theta}) \int \max_{i \in \{1, \dots, K\}} g_{\boldsymbol{\theta}_i}(\mathbf{x}) p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}(\mathbf{x}) d\mathbf{x} d\boldsymbol{\theta}$, where $p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}$ is the optimal solution of (3.14) for a given $\boldsymbol{\theta}$. If the stochastic parameter design is the optimal solution of the problem in (3.14), then $r_{\text{sto}}(\hat{\boldsymbol{\theta}})$ is strictly smaller than $r_{\text{det}}(\hat{\boldsymbol{\theta}})$. Otherwise, it is concluded that the deterministic parameter design is the optimal solution and the stochastic design does not provide any improvements; that is, $r_{\text{sto}}(\hat{\boldsymbol{\theta}}) = r_{\text{det}}(\hat{\boldsymbol{\theta}})$. In the following proposition, sufficient conditions presented for the second case.

Proposition 3.5 *For the maximum Bayes risk criterion, the stochastic parameter design cannot provide any improvements over the deterministic parameter design if at least one of the following conditions is satisfied for each $\boldsymbol{\theta}$:*

- *The solution of the unconstrained problem (see (3.9) or (3.10)) is deterministic (denoted by $\mathbf{s}_{\boldsymbol{\theta}}^{\text{unc}}$) and satisfies the power constraint; i.e., $\|\mathbf{s}_{\boldsymbol{\theta}}^{\text{unc}}\|^2 \leq A_{\boldsymbol{\theta}}$.*
- *$g_{\boldsymbol{\theta}_i}$ is a convex function for $i \in \{1, \dots, K\}$.*

Proof: The first part of the proof can be obtained similarly to that of Proposition 2.2 in Chapter 2. If the first condition in the proposition is satisfied, i.e., if the unconstrained problem has a deterministic solution and $\|\mathbf{s}_{\boldsymbol{\theta}}^{\text{unc}}\|^2 \leq A_{\boldsymbol{\theta}}$, then the solution of (3.23) is the same as that of the unconstrained problem in Section 3.1.1; that is, $\mathbf{s}_{\boldsymbol{\theta}}^{\text{det}} = \mathbf{s}_{\boldsymbol{\theta}}^{\text{unc}}$. Therefore, the solution of the optimal stochastic parameter design problem in (3.14) is expressed as $p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta}}^{\text{unc}})$. Hence, the deterministic parameter design is optimal in this case, and the stochastic parameter design cannot provide any improvements.

For the second condition in the proposition, it is note that, for any $\mathbf{s}_{\boldsymbol{\theta}}$, $E\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \geq \|E\{\mathbf{s}_{\boldsymbol{\theta}}\}\|^2$ holds due to Jensen's inequality as norm is a convex function. Therefore, $E\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \leq A_{\boldsymbol{\theta}}$ in (3.14) implies that $\|E\{\mathbf{s}_{\boldsymbol{\theta}}\}\|^2 \leq A_{\boldsymbol{\theta}}$ must hold for any feasible PDF of $\mathbf{s}_{\boldsymbol{\theta}}$. Let $E\{\mathbf{s}_{\boldsymbol{\theta}}\}$ be denoted by $\check{\mathbf{s}}_{\boldsymbol{\theta}}$; that is, $\check{\mathbf{s}}_{\boldsymbol{\theta}} \triangleq E\{\mathbf{s}_{\boldsymbol{\theta}}\}$. Since the minimizer of (3.23), $\mathbf{s}_{\boldsymbol{\theta}}^{\text{det}}$, achieves the minimum value of $\max_{i \in \{1, \dots, K\}} g_{\boldsymbol{\theta}_i}(\mathbf{s}_{\boldsymbol{\theta}})$

among all $\mathbf{s}_\theta$ that satisfy $\|\mathbf{s}_\theta\|^2 \leq A_\theta$, the inequality $\|\mathbf{E}\{\mathbf{s}_\theta\}\|^2 = \|\check{\mathbf{s}}_\theta\|^2 \leq A_\theta$ implies that

$$\max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{E}\{\mathbf{s}_\theta\}) = \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\check{\mathbf{s}}_\theta) \geq \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_\theta^{\text{det}}) \quad (3.24)$$

holds. If g_{θ_i} 's are convex functions, then

$$\max_{i \in \{1, \dots, K\}} \mathbf{E}\{g_{\theta_i}(\mathbf{s}_\theta)\} \geq \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{E}\{\mathbf{s}_\theta\}) \geq \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_\theta^{\text{det}}) \quad (3.25)$$

is obtained from Jensen's inequality and from (3.24). Therefore, when g_{θ_i} 's are convex, $\max_{i \in \{1, \dots, K\}} \mathbf{E}\{g_{\theta_i}(\mathbf{s}_\theta)\}$ is never smaller than the minimum value of (3.23), $\max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_\theta^{\text{det}})$, for any PDF of $\mathbf{s}_\theta$ that satisfies the average power constraint. For this reason, the minimum value of (3.14) cannot be smaller than $\max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_\theta^{\text{opt}})$, which means that it is always equal to $\max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_\theta^{\text{det}})$ as (3.14) covers (3.23) as a special case.

Overall, if at least one of the conditions in the proposition is satisfied for all θ , the deterministic and stochastic parameter design approaches achieve the same minimum values for all parameters; that is, $\max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_\theta^{\text{det}}) = \int \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{x}) p_{\mathbf{s}_\theta^{\text{opt}}}(\mathbf{x}) d\mathbf{x}$, $\forall \theta$. Hence, $r_{\text{det}}(\hat{\theta}) = \int_{\Lambda} w(\theta) \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_\theta^{\text{det}}) d\theta$ and $r_{\text{sto}}(\hat{\theta}) = \int_{\Lambda} w(\theta) \int \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{x}) p_{\mathbf{s}_\theta^{\text{opt}}}(\mathbf{x}) d\mathbf{x} d\theta$ are equal. ■

For an example of Proposition 3.5, consider a scenario in which two scalar parameters θ_1 and θ_2 are to be estimated in the presence of zero-mean additive noise $\mathbf{n}$. The average power constraint is in the form of $\mathbf{E}\{|\mathbf{s}_\theta|^2\} \leq A_\theta$ for all θ , and the estimator is specified by $\hat{\theta}(\mathbf{y}) = \mathbf{y}$. Also, the cost function is modeled as $C[\hat{\theta}(\mathbf{y}), \theta] = (\hat{\theta}_1(y_1) - \theta_1)^2 + (\hat{\theta}_2(y_2) - \theta_2)^2$. In this case, g_{θ_1} in (3.4) can be calculated as

$$\begin{aligned} g_{\theta_1}(x_1) &= \int_{-\infty}^{\infty} (y_1 - \theta_1)^2 p_{n_1}(y_1 - x_1 - \rho_{12}x_2) dy_1 \\ &= \int_{-\infty}^{\infty} (y + x_1 + \rho_{12}x_2 - \theta_1)^2 p_{n_1}(y) dy \\ &= (x_1 + \rho_{12}x_2 - \theta_1)^2 + \text{Var}\{n_1\} \end{aligned} \quad (3.26)$$

where $\text{Var}\{n_1\}$ is the variance of the noise component for the first parameter. From (3.26), it is observed that g_{θ_1} is a convex function for any value of θ .

Similarly it is possible to show that g_{θ_2} is also a convex function for any θ . Therefore, the second condition in Proposition 3.5 is satisfied for all θ , which implies that the performance of the deterministic parameter design cannot be improved via the stochastic approach in this scenario.

In the following proposition, a modified version of Proposition 2.3 in Chapter 2 is obtained in order to present sufficient conditions that specify scenarios in which the stochastic parameter design provides improvements over the deterministic one.

Proposition 3.6 *For the maximum Bayes risk criterion, the stochastic parameter design achieves a lower Bayes risk than the deterministic parameter design if there exists $\theta \in \Lambda$ for which all $g_{\theta_i}(\mathbf{x})$'s are second-order continuously differentiable around $\mathbf{s}_{\theta}^{\text{det}}$, and a real vector $\mathbf{z}$ and a positive number k can be found such that*

$$(\mathbf{z}^T \mathbf{s}_{\theta}^{\text{det}}) (\mathbf{z}^T \nabla g_{\theta_i}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\theta}^{\text{det}}}) < 0 \quad \text{and} \quad (3.27)$$

$$\|\mathbf{z}\|^2 < (\mathbf{z}^T \mathbf{s}_{\theta}^{\text{det}}) (\mathbf{z}^T \mathbf{H}_{\theta_i} \mathbf{z} - \gamma_i/k) / (\mathbf{z}^T \nabla g_{\theta_i}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\theta}^{\text{det}}}) \quad (3.28)$$

for $i \in \{1, \dots, K\}$, where $\mathbf{s}_{\theta}^{\text{det}}$ is the solution of (3.23), $\nabla g_{\theta_i}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\theta}^{\text{det}}}$ denotes the gradient of $g_{\theta_i}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_{\theta}^{\text{det}}$, $\mathbf{H}_{\theta_i}$ is the Hessian of $g_{\theta_i}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_{\theta}^{\text{det}}$, and $\gamma_i \triangleq \max_i g_{\theta_i}(\mathbf{s}_{\theta}^{\text{det}}) - g_{\theta_i}(\mathbf{s}_{\theta}^{\text{det}})$.

Proof: In order to prove that the stochastic parameter design can achieve a reduced Bayes risk, consider a specific value of θ for which the conditions in the proposition are satisfied. In addition, consider L values $\mathbf{s}_{\theta,1}$ through $\mathbf{s}_{\theta,L}$ around $\mathbf{s}_{\theta}^{\text{det}}$, which can be expressed as $\mathbf{s}_{\theta,j} = \mathbf{s}_{\theta}^{\text{det}} + \epsilon_j$ for $j \in \{1, \dots, L\}$, where $2 \leq L \leq K+1$. Then, $g_{\theta_i}(\mathbf{s}_{\theta,j})$ is approximated as $g_{\theta_i}(\mathbf{s}_{\theta,j}) \approx g_{\theta_i}(\mathbf{s}_{\theta}^{\text{det}}) + \epsilon_j^T \hat{\mathbf{g}}_{\theta_i} + 0.5 \epsilon_j^T \mathbf{H}_{\theta_i} \epsilon_j$ for $i \in \{1, \dots, K\}$, where $\hat{\mathbf{g}}_{\theta_i} = \nabla g_{\theta_i}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\theta}^{\text{det}}}$ is the gradient and $\mathbf{H}_{\theta_i}$ is the Hessian of $g_{\theta_i}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_{\theta}^{\text{det}}$ [45]. Similarly, $\|\mathbf{s}_{\theta,j}\|^2$ is expressed as $\|\mathbf{s}_{\theta,j}\|^2 \approx \|\mathbf{s}_{\theta}^{\text{det}}\|^2 + 2 \epsilon_j^T \mathbf{s}_{\theta}^{\text{det}} + \|\epsilon_j\|^2$ for $j \in \{1, \dots, L\}$. To prove that the use of $p_{\mathbf{s}_{\theta}}(\mathbf{x}) = \sum_{j=1}^L \lambda_{\theta,j} \delta(\mathbf{x} - \mathbf{s}_{\theta,j})$ results in a lower risk than $\max_{i \in \{1,2\}} g_{\theta_i}(\mathbf{s}_{\theta}^{\text{det}})$, which is the risk achieved by the deterministic parameter design (see (3.23)), it is sufficient

to show that

$$\begin{aligned} \max_{i \in \{1, \dots, K\}} \sum_{j=1}^L \lambda_{\theta, j} g_{\theta_i}(\mathbf{s}_{\theta, j}) &< \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_{\theta}^{\det}) \\ \sum_{j=1}^L \lambda_{\theta, j} \|\mathbf{s}_{\theta, j}\|^2 &\leq \|\mathbf{s}_{\theta}^{\det}\|^2 \leq A_{\theta} \end{aligned} \quad (3.29)$$

hold for certain selection of parameters (see (3.13)). Inserting the expressions for $g_{\theta_i}(\mathbf{s}_{\theta, j})$ and $\|\mathbf{s}_{\theta, j}\|^2$ around $\mathbf{s}_{\theta}^{\det}$ into (3.29), the following inequalities are obtained:

$$\begin{aligned} \max_{i \in \{1, \dots, K\}} \left\{ g_{\theta_i}(\mathbf{s}_{\theta}^{\det}) + 0.5 \sum_{j=1}^L \lambda_{\theta, j} \boldsymbol{\epsilon}_j^T \mathbf{H}_{\theta_i} \boldsymbol{\epsilon}_j + \sum_{j=1}^L \lambda_{\theta, j} \boldsymbol{\epsilon}_j^T \hat{\mathbf{g}}_{\theta} \right\} &< \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_{\theta}^{\det}) \\ \sum_{j=1}^L \lambda_{\theta, j} \|\boldsymbol{\epsilon}_j\|^2 + 2 \sum_{j=1}^L \lambda_{\theta, j} \boldsymbol{\epsilon}_j^T \mathbf{s}_{\theta}^{\det} &< 0 \end{aligned} \quad (3.30)$$

Let $\boldsymbol{\epsilon}_j$ be expressed as $\boldsymbol{\epsilon}_j = \eta_j \mathbf{z}$ for $j \in \{1, \dots, L\}$. Then, (3.30) can be stated after some manipulation as

$$\begin{aligned} \max_{i \in \{1, \dots, K\}} \left\{ g_{\theta_i}(\mathbf{s}_{\theta}^{\det}) + k_1(\mathbf{z}^T \mathbf{H}_{\theta_i} \mathbf{z}) + k_2(\mathbf{z}^T \hat{\mathbf{g}}_{\theta_i}) \right\} &< \max_{i \in \{1, \dots, K\}} g_{\theta_i}(\mathbf{s}_{\theta}^{\det}) \\ k_1 \|\mathbf{z}\|^2 + k_2(\mathbf{z}^T \mathbf{s}_{\theta}^{\det}) &< 0 \end{aligned} \quad (3.31)$$

where $k_1 \triangleq 0.5 \sum_{j=1}^L \lambda_{\theta, j} \eta_j^2$ and $k_2 \triangleq \sum_{j=1}^L \lambda_{\theta, j} \eta_j$.

The first inequality in (3.31) can be separated into K inequalities using the fact that the left hand side of all the K inequalities are smaller than the right hand side. Also these K inequalities can further be simplified depending on the value of i that maximizes the right hand side:

$$\begin{aligned} k_1(\mathbf{z}^T \mathbf{H}_{\theta_i} \mathbf{z}) + k_2(\mathbf{z}^T \hat{\mathbf{g}}_{\theta_i}) &< \gamma_i \\ k_1 \|\mathbf{z}\|^2 + k_2(\mathbf{z}^T \mathbf{s}_{\theta}^{\det}) &< 0 \end{aligned} \quad (3.32)$$

for $i \in \{1, \dots, K\}$, where γ_i is as defined in the proposition. If the first K inequalities in (3.32) are multiplied by $(\mathbf{z}^T \mathbf{s}_{\theta}^{\det})/(\mathbf{z}^T \hat{\mathbf{g}}_{\theta_i})$, which is always negative due to the condition (3.27) in the proposition, (3.32) becomes

$$\begin{aligned} k_1(\mathbf{z}^T \mathbf{H}_{\theta_i} \mathbf{z}) \frac{(\mathbf{z}^T \mathbf{s}_{\theta}^{\det})}{(\mathbf{z}^T \hat{\mathbf{g}}_{\theta_i})} + k_2(\mathbf{z}^T \mathbf{s}_{\theta}^{\det}) &> \gamma_i \frac{(\mathbf{z}^T \mathbf{s}_{\theta}^{\det})}{(\mathbf{z}^T \hat{\mathbf{g}}_{\theta_i})} \\ k_1 \|\mathbf{z}\|^2 + k_2(\mathbf{z}^T \mathbf{s}_{\theta}^{\det}) &< 0. \end{aligned} \quad (3.33)$$

It is noted that k_2 can take any real value by adjusting $\lambda_{\theta,j} \in [0, 1]$ and infinitesimally small η_j values. Therefore, it is guaranteed that both inequalities in (3.33) are satisfied if $(\mathbf{z}^T \mathbf{H}_{\theta_i} \mathbf{z} - \gamma_i/k)(\mathbf{z}^T \mathbf{s}_{\theta}^{\text{det}})/(\mathbf{z}^T \hat{\mathbf{g}}_{\theta_i})$ is larger than $\|\mathbf{z}\|^2$, which is the condition in (3.28). $\blacksquare$

The conditions in Proposition 3.6 provide a relatively simple technique, which is based on the first and second order derivatives of g_{θ_i} , for determining if the stochastic parameter design can provide improvements over the deterministic one. If the conditions are satisfied, the stochastic parameter design is guaranteed to outperform the deterministic parameter design, in which case the optimization problem in (3.15) can be solved to obtain the optimal solution. It should also be noted that there may exist scenarios in which the stochastic parameter design provides improvements over the deterministic one even though the conditions in Proposition 6 are not satisfied, which is due to the fact that the conditions are sufficient but not necessary. In the next section, examples are presented for various scenarios.

3.4 Numerical Examples

In this section, numerical examples are presented in order to investigate the performance of the optimal parameter design approach in various scenarios. Consider an estimation problem in which parameter vector $\boldsymbol{\theta} = [\theta_1 \ \theta_2]^T$ is to be estimated based on observation vector $\mathbf{y} = [y_1 \ y_2]^T$, which is modeled as

$$\mathbf{y} = \mathbf{s}_{\boldsymbol{\theta}} + \rho(\mathbf{1} - \mathbf{I})\mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n} \quad (3.34)$$

where $\mathbf{s}_{\boldsymbol{\theta}} = [s_{\theta_1} \ s_{\theta_2}]^T$, $\mathbf{n} = [n_1 \ n_2]^T$, $\rho = \rho_{12} = \rho_{21}$ denotes the multiplier term related to interference (see (3.1)), $\mathbf{I}$ is the identity matrix of size 2×2 , and $\mathbf{1}$ is the matrix of ones with the same size. The components n_1 and n_2 of the additive noise $\mathbf{n}$ are independent and identically distributed Gaussian random variables, specified by PDFs $p_{n_1}(\mathbf{n}) = p_{n_2}(\mathbf{n}) = \exp\{-(\mathbf{n} - \boldsymbol{\mu})^2/(2\sigma^2)\}/(\sqrt{2\pi}\sigma)$. The estimator is specified by $\hat{\boldsymbol{\theta}}(\mathbf{y}) = \mathbf{y}$, which estimates each parameter independently based on the corresponding observation. The cost function for each parameter

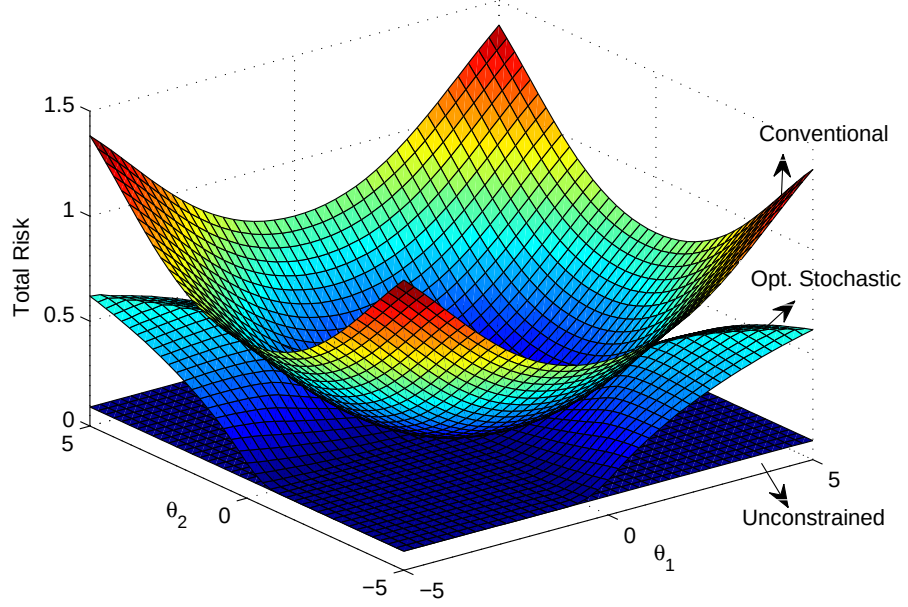


Figure 3.2: Total Bayes risk versus θ_1 and θ_2 for the example defined in Section 3.4.

is chosen as the uniform cost function, which is calculated as $C[\hat{\theta}_i(y_i), \theta_i] = 1$ if $|\hat{\theta}_i(y_i) - \theta_i| > \Delta$ and $C[\hat{\theta}_i(y_i), \theta_i] = 0$ otherwise for $i = 1, 2$. Based on this model, g_{θ_1} in (3.4) can be obtained as

$$g_{\theta_1}(\mathbf{x}) = Q\left(\frac{x_1 + \rho x_2 - \theta_1 + \mu + \Delta}{\sigma}\right) + Q\left(\frac{-x_1 - \rho x_2 + \theta_1 - \mu + \Delta}{\sigma}\right) \quad (3.35)$$

where $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty \exp\{-t^2/2\} dt$ denotes the Q -function. For each $\boldsymbol{\theta}$, $E\{\|\mathbf{s}_\boldsymbol{\theta}\|^2\} \leq \|\boldsymbol{\theta}\|^2$ is employed as the constraint stated in (3.11). Similarly to (3.35), g_{θ_2} for the second parameter can be obtained.

In the numerical examples, the parameter spaces for both parameters are specified as $\Lambda_1 = \Lambda_2 = [-10, 10]$. Also, s_{θ_1} and s_{θ_2} can take values in the interval $[-10, 10]$ subject to the average power constraint, $E\{\|\mathbf{s}_\boldsymbol{\theta}\|^2\} \leq \|\boldsymbol{\theta}\|^2$. In addition, the Gaussian distribution of the noise is taken to be zero mean with $\sigma = 0.5$ and ρ is chosen to be 0.25. Since the noise is a zero-mean random variable, $\hat{\boldsymbol{\theta}}(\mathbf{y}) = \mathbf{y}$

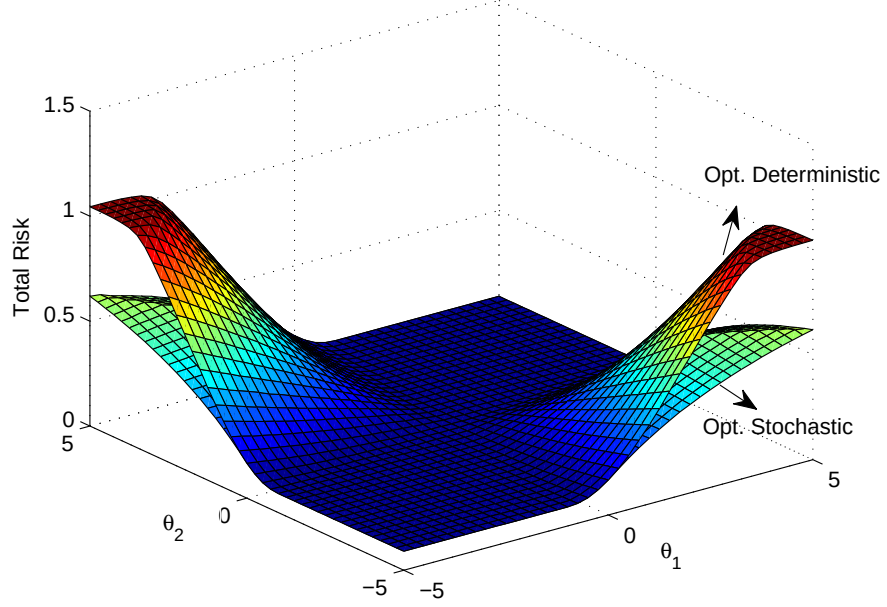


Figure 3.3: Total Bayes risk versus θ_1 and θ_2 for the example defined in Section 3.4.

can be considered as a practical estimator.³ Finally, $\Delta = 1$ is used for the uniform cost function described in the previous paragraph.

In Figure 3.2, the total Bayes risks for the stochastic parameter design, the unconstrained parameter design and the conventional parameter design (which transmits the parameters as they are; that is, employs $s_{\theta_i} = \theta_i$) are illustrated. Also in Figure 3.3, the total Bayes risks for the stochastic parameter design and the deterministic parameter design are compared. It is observed that the stochastic parameter design achieves improvements over the deterministic and conventional designs. Also, for some values of θ_1 and θ_2 , the performance of the stochastic design is the same as the unconstrained one.

In Figure 3.4, the maximum Bayes risks for the stochastic parameter design,

³Although this is not the optimal estimator, it can be used in practice due to its simplicity compared to the optimal estimator.

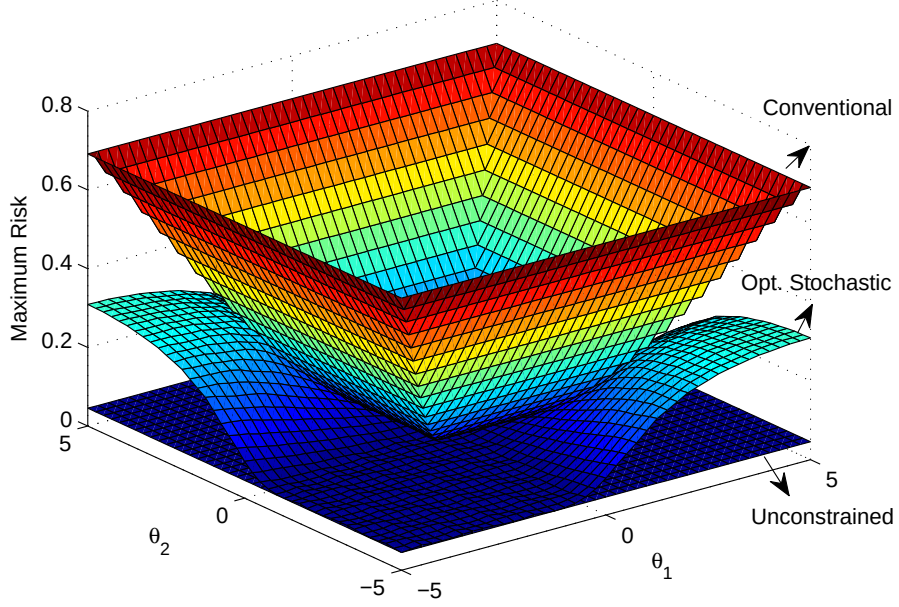


Figure 3.4: Maximum Bayes risk versus θ_1 and θ_2 for the example defined in Section 3.4.

the unconstrained parameter design and the conventional parameter design are plotted. Also, in Figure 3.5, the maximum Bayes risks for stochastic parameter design and the deterministic parameter design are illustrated. Similar to the previous scenario, it is observed that the stochastic parameter design provides improvements over the conventional and deterministic parameter design approaches for certain range of parameter values.

In Table 3.1, the total Bayes risk criterion is considered, and the optimal solutions for the stochastic, the deterministic and the unconstrained parameter design approaches are presented for various values of θ . It is observed from the table that the optimal stochastic parameter design can involve randomization between two different signals for certain values of θ , which corresponds to the cases in which the stochastic approach outperforms the deterministic parameter design, as can be verified from Figure 3.3. Similarly, Table 3.2 presents the optimal solutions for the stochastic, the deterministic and the unconstrained parameter design

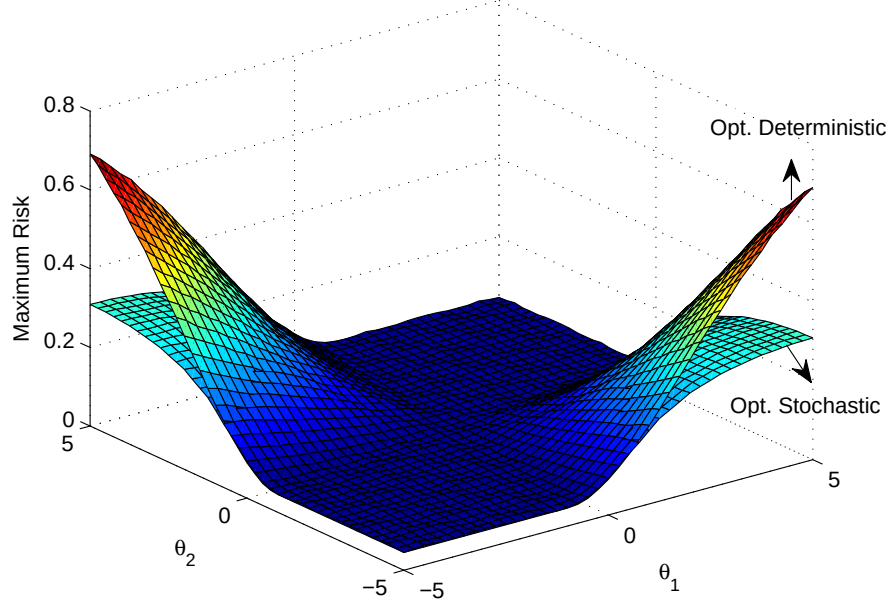


Figure 3.5: Maximum Bayes risk versus θ_1 and θ_2 for the example defined in Section 3.4.

approaches for the maximum Bayes risk criterion. The main difference in this scenario is that randomization (time sharing) among up to three different signals can be performed for the optimal stochastic parameter design in accordance with Proposition 3.1.

Next, the maximum Bayes risk criterion is considered, and the conditions in Proposition 3.6 are studied. Namely, the existence of a real vector $\mathbf{z}$ and a positive number k that satisfy the conditions in Proposition 3.6 for a certain value of $\boldsymbol{\theta}$ is investigated. Consider the parameter value $\boldsymbol{\theta} = [-5, 5]$. If all the four conditions (two conditions for each estimator) are satisfied for this value of $\boldsymbol{\theta}$, then it is guaranteed that the stochastic parameter design yields a lower maximum Bayes risk than the deterministic design. To test the first condition for each estimator, we need the value of $\mathbf{s}_{\boldsymbol{\theta}}^{\text{det}}$ and the gradients of $g_{\theta_1}(\mathbf{x})$ and $g_{\theta_2}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}}^{\text{det}}$. As it can be observed from Table 3.2, $\mathbf{s}_{\boldsymbol{\theta}}^{\text{det}} = [-5, 5]$ for $\boldsymbol{\theta} = [-5, 5]$, and the gradients of

Table 3.1: Unconstrained solution $p_{\mathbf{s}_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{unc}})$, optimal deterministic solution $p_{\mathbf{s}_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{opt}})$, and optimal stochastic solution $p_{\mathbf{s}_\theta}^{\text{opt}}(\mathbf{x}) = \lambda_{\theta,1} \delta(\mathbf{x} - \mathbf{s}_{\theta,1}) + \lambda_{\theta,2} \delta(\mathbf{x} - \mathbf{s}_{\theta,2})$ for the total Bayes risk criterion for various values of θ .

θ_1	θ_2	$\mathbf{s}_\theta^{\text{unc}}$	$\mathbf{s}_\theta^{\text{opt}}$	$\lambda_{\theta,1}$	$\mathbf{s}_{\theta,1}$	$\lambda_{\theta,2}$	$\mathbf{s}_{\theta,2}$
-5	-5	(-4,-4)	(-4,-4)	1	(-4,-4)	-	-
-5	-2.5	(-4.667,-1.333)	(-4.667,-1.333)	1	(-4.667,-1.333)	-	-
-5	0	(-5.333,1.333)	(-4.889,1.049)	1	(-4.889,1.049)	-	-
-5	2.5	(-6,4)	(-4.694,3.035)	0.669	(-5.565,3.622)	0.331	(0.561,2.243)
-5	5	(-6.667,6.667)	(-5.950,3.820)	0.508	(1.138,4.553)	0.492	(-6.280,6.280)
-2.5	-5	(-1.333,-4.667)	(-1.333,-4.667)	1	(-1.333,-4.667)	-	-
-2.5	-2.5	(-2,-2)	(-2,-2)	1	(-2,-2)	-	-
-2.5	0	(-2.667,0.667)	(-2.444,0.527)	1	(-2.444,0.527)	-	-
-2.5	2.5	(-3.333,3.333)	(-2.5,2.5)	0.918	(-2.572,2.572)	0.082	(-2.023,-0.206)
-2.5	5	(-4,6)	(-3.035,4.694)	0.669	(-3.622,5.565)	0.331	(-2.243,-0.561)
0	-5	(1.333,-5.333)	(1.049,-4.889)	1	(1.049,-4.889)	-	-
0	-2.5	(0.667,-2.667)	(0.527,-2.444)	1	(0.527,-2.444)	-	-
0	2.5	(-0.667,2.667)	(-0.527,2.444)	1	(-0.527,2.444)	-	-
0	5	(-1.333,5.333)	(-1.049,4.889)	1	(-1.049,4.889)	-	-

$g_{\theta_1}(\mathbf{x})$ and $g_{\theta_2}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_\theta^{\text{det}}$ can be calculated based on the following equations:

$$\begin{aligned} \nabla g_{\theta_1}(\mathbf{x}) &= \frac{-e^{-\frac{a_1^2}{2}} + e^{-\frac{a_2^2}{2}}}{\sqrt{2\pi}} [1, \rho]^T \\ \nabla g_{\theta_2}(\mathbf{x}) &= \frac{-e^{-\frac{b_1^2}{2}} + e^{-\frac{b_2^2}{2}}}{\sqrt{2\pi}} [1, \rho]^T \end{aligned} \quad (3.36)$$

where $a_1 = (\mathbf{x}_1 - \rho \mathbf{x}_2 - \theta_1 + \mu + \Delta)/\sigma$, $a_2 = (-\mathbf{x}_1 + \rho \mathbf{x}_2 + \theta_1 - \mu + \Delta)/\sigma$, $b_1 = (\mathbf{x}_2 - \rho \mathbf{x}_1 - \theta_2 + \mu + \Delta)/\sigma$, and $b_2 = (-\mathbf{x}_2 + \rho \mathbf{x}_1 + \theta_2 + \mu + \Delta)/\sigma$. Based on these equations, the first condition in Proposition 3.6 can be evaluated for each estimator. The first two plots in Figure 3.6 illustrate the values of $\mathbf{z}$ for which the first condition in Proposition 3.6 is satisfied for the first and the second estimator, respectively, for $k = 1$ and $\theta = [-5, 5]$. Namely, in the white (black) regions, the conditions are satisfied (not satisfied). As observed from the figure, there are certain regions in which the first condition is satisfied for each estimator. Next, the second condition in Proposition 3.6 is tested. To that aim, the Hessians of $g_{\theta_1}(\mathbf{x})$ and $g_{\theta_2}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_\theta^{\text{det}}$ are calculated. The Hessians of these functions can

Table 3.2: Unconstrained solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{unc}})$, optimal deterministic solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}_\theta^{\text{opt}})$, and optimal stochastic solution $p_{s_\theta}^{\text{opt}}(\mathbf{x}) = \lambda_{\theta,1} \delta(\mathbf{x} - \mathbf{s}_{\theta,1}) + \lambda_{\theta,2} \delta(\mathbf{x} - \mathbf{s}_{\theta,2}) + \lambda_{\theta,3} \delta(\mathbf{x} - \mathbf{s}_{\theta,3})$ for the maximum Bayes risk criterion for various values of θ .

θ_1	θ_2	$\mathbf{s}_\theta^{\text{unc}}$	$\mathbf{s}_\theta^{\text{opt}}$	$\lambda_{\theta,1}$	$\mathbf{s}_{\theta,1}$	$\lambda_{\theta,2}$	$\mathbf{s}_{\theta,2}$	$\lambda_{\theta,3}$	$\mathbf{s}_{\theta,3}$
-5	-5	(-4,-4)	(-4,-4)	1	(-4,-4)	-	-	-	-
-5	-2.5	(-4.667,-1.333)	(-4.667,-1.333)	1	(-4.667,-1.333)	-	-	-	-
-5	0	(-5.333,1.333)	(-4.915,0.915)	1	(-4.915,0.915)	-	-	-	-
-5	2.5	(-6,4)	(-4.824,2.824)	0.637	(-5.498,3.416)	0.194	(0.503,2.167)	0.169	(-4.481,-1.094)
-5	5	(-6.667,6.667)	(-5,5)	0.487	(-6.264,6.313)	0.261	(-4.607,-0.995)	0.252	(0.794,4.682)
-2.5	-5	(-1.333,-4.667)	(-1.333,-4.667)	1	(-1.333,-4.667)	-	-	-	-
-2.5	-2.5	(-2,-2)	(-2,-2)	1	(-2,-2)	-	-	-	-
-2.5	0	(-2.667,0.667)	(-2.458,0.458)	0.998	(-2.458,0.458)	0.002	(-2.464,0.465)	-	-
-2.5	2.5	(-3.333,3.333)	(-2.5,2.5)	0.879	(-2.617,2.599)	0.066	(0.415,2.093)	0.055	(-2.011,-0.553)
-2.5	5	(-4,6)	(-2.824,4.824)	0.623	(-3.458,5.502)	0.194	(-2.211,-0.498)	0.183	(1.158,4.493)
0	-5	(1.333,-5.333)	(0.915,-4.915)	1	(0.915,-4.915)	-	-	-	-
0	-2.5	(0.667,-2.667)	(0.458,-2.458)	1	(0.458,-2.458)	-	-	-	-
0	2.5	(-0.667,2.667)	(-0.458,2.458)	1	(-0.458,2.458)	-	-	-	-
0	5	(-1.333,5.333)	(-0.915,4.915)	1	(-0.915,4.915)	-	-	-	-

be found as follows:

$$\begin{aligned}
\mathbf{H}_{\theta_1}(\mathbf{x}) &= \frac{a_1 e^{-\frac{a_1^2}{2}} + a_2 e^{-\frac{a_2^2}{2}}}{\sqrt{2\pi}} \begin{bmatrix} 1 & \rho \\ \rho & \rho^2 \end{bmatrix} \\
\mathbf{H}_{\theta_2}(\mathbf{x}) &= \frac{b_1 e^{-\frac{b_1^2}{2}} + b_2 e^{-\frac{b_2^2}{2}}}{\sqrt{2\pi}} \begin{bmatrix} \rho^2 & \rho \\ \rho & 1 \end{bmatrix}
\end{aligned} \tag{3.37}$$

where a_1 , a_2 , b_1 , and b_2 are as defined previously. Based on (3.36) and (3.37), the second condition in Proposition 3.6 can be evaluated for each estimator. The results of these evaluations are shown in the second and third plots in Figure 3.6 for different values of $\mathbf{z}$. Similar to the first condition, the second condition is satisfied in certain range of $\mathbf{z}$ values (white regions). The last plot in Figure 3.6 shows the intersection of the regions in which the conditions are satisfied. As observed from the figure, the intersection is not an empty set, hence we can conclude that there exist a real vector $\mathbf{z}$ and a positive number k for which all the conditions in Proposition 3.6 are satisfied. Therefore, in this scenario, it is guaranteed that the stochastic parameter design achieves a lower Bayes risk than the deterministic design as a result of Proposition 3.6.

In order to gain intuition and further understanding, Figure 3.7 and Figure 3.8 visualize how the stochastic design approach provides performance improvements over the deterministic one for the total Bayes risk and the maximum Bayes risk

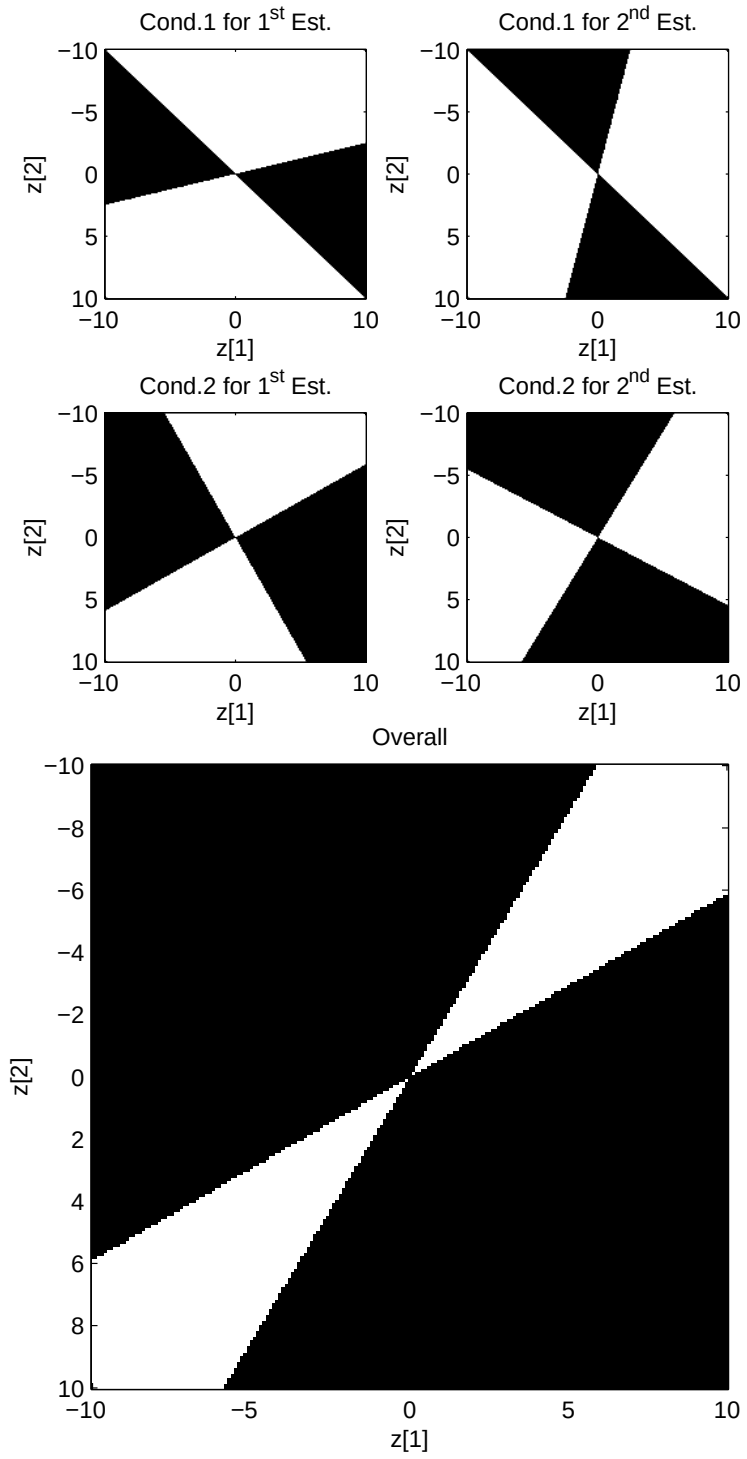


Figure 3.6: Regions in which the optimality conditions stated in Proposition 3.6 are satisfied for different values of $\mathbf{z}$ for $\boldsymbol{\theta} = [-5, 5]$.

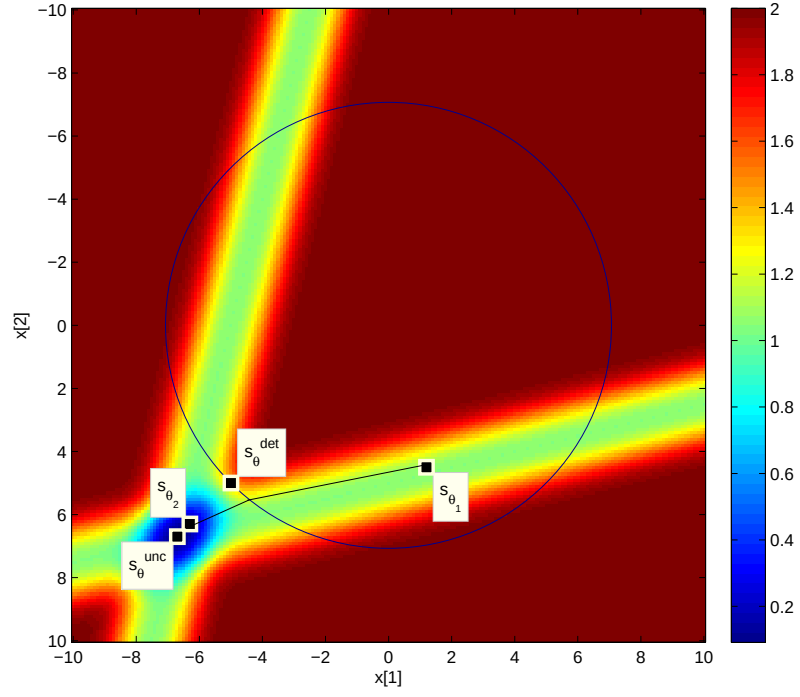


Figure 3.7: The total Bayes risk $\tilde{g}_\theta(\mathbf{x})$ for $\theta = [-5, 5]$.

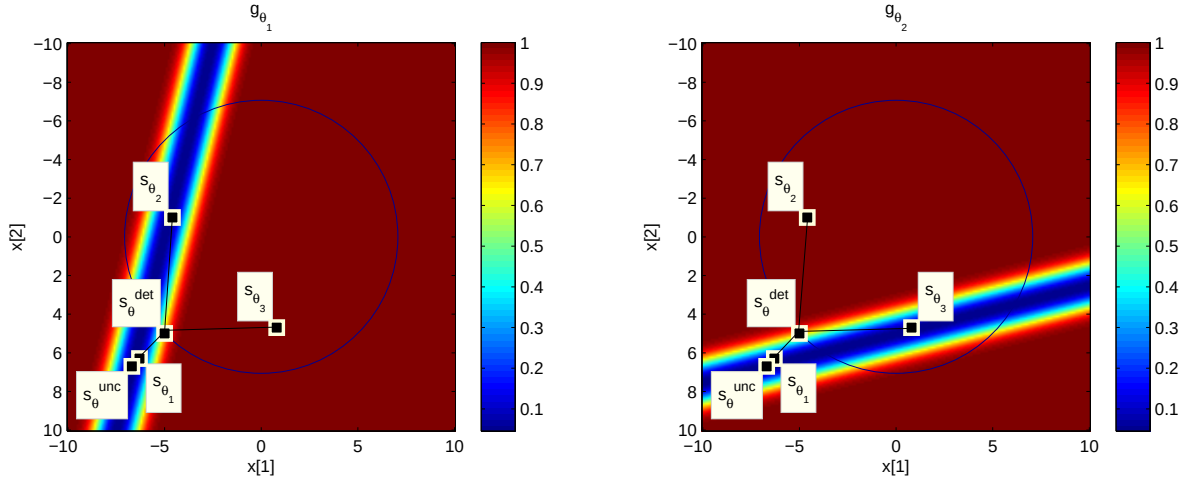


Figure 3.8: The Bayes risk for the first estimator, $g_{\theta_1}(\mathbf{x})$, and the Bayes risk for the second estimator, $g_{\theta_2}(\mathbf{x})$, for $\theta = [-5, 5]$.

criteria, respectively. In Figure 3.7, the total Bayes risk $\tilde{g}_{\boldsymbol{\theta}}(\mathbf{x})$ is illustrated for $\boldsymbol{\theta} = [-5, 5]$. The region inside the circle corresponds to the values of $\mathbf{x}$ that satisfy the power constraint individually. Here it can be seen that the minimum value of $\tilde{g}_{\boldsymbol{\theta}}(\mathbf{x})$ is observed at a value of $\mathbf{x}$ which does not satisfy the power constraint. In that case, the unconstrained solution simply picks that value of $\mathbf{x}$ as $\mathbf{s}_{\boldsymbol{\theta}}^{\text{unc}}$. On the other hand, the optimal deterministic solution, picks the value of $\mathbf{x}$ residing on or inside the circle, which minimizes the value of $\tilde{g}_{\boldsymbol{\theta}}(\mathbf{x})$, denoted as $\mathbf{s}_{\boldsymbol{\theta}}^{\text{det}}$. Obviously, there is a performance gap between the unconstrained solution and the optimal deterministic solution, as can be observed from Figure 3.2 and Figure 3.3. The proposed stochastic design approach aims to achieve improvements over the deterministic solution. To that aim, the stochastic design perform randomization (time sharing) between two signals, i.e., one that satisfies the power constraint with low performance, the other that does not satisfy the constraint but has high performance, as can be seen from Figure 3.7. By randomizing between these two signals, it is possible to satisfy the power constraint on the average and to achieve a better performance than the optimal deterministic design.

In Figure 3.8, the Bayes risks for the estimators, $g_{\boldsymbol{\theta}_1}(\mathbf{x})$ and $g_{\boldsymbol{\theta}_2}(\mathbf{x})$, are illustrated for $\boldsymbol{\theta} = [-5, 5]$. As stated in (3.14), the stochastic design aims to minimize the maximum of the expectations of $g_{\boldsymbol{\theta}_1}(\mathbf{x})$ and $g_{\boldsymbol{\theta}_2}(\mathbf{x})$. It is observed from Figure 3.8 that the minimum values of $g_{\boldsymbol{\theta}_1}(\mathbf{x})$ and $g_{\boldsymbol{\theta}_2}(\mathbf{x})$ coincide outside the power constraint. Therefore, it is not possible to pick a single point which minimizes the maximum Bayes risk and satisfy the constraint at the same time. Similar to the total Bayes risk case, the stochastic design should perform randomization between some signals inside and outside the constraint to achieve improvements over the deterministic design. However, as seen in Figure 3.8, the minimum values of $g_{\boldsymbol{\theta}_1}(\mathbf{x})$ and $g_{\boldsymbol{\theta}_2}(\mathbf{x})$ do not coincide inside the constraint and as a result any signal minimizing one of $g_{\boldsymbol{\theta}_1}(\mathbf{x})$ and $g_{\boldsymbol{\theta}_2}(\mathbf{x})$ maximizes the other one. Hence it is not possible to pick just one signal that satisfies the constraint and randomize it with another signal that does not satisfy the constraint. To overcome this problem, two signals satisfying the constraint should be chosen and these two signals should be randomized with a signal that does not satisfy the power constraint. As a result, the expectations of both $g_{\boldsymbol{\theta}_1}(\mathbf{x})$ and $g_{\boldsymbol{\theta}_2}(\mathbf{x})$ are minimized to a certain

point, which makes it possible to minimize the maximum of these expectations.

3.5 Conclusions for Chapter 3

In this chapter, the optimal stochastic design of multiple parameters has been studied for a given set of fixed estimators. Two different performance criteria have been considered; namely, the total Bayes risk criterion and the maximum Bayes risk criterion. It has been shown that, in the presence of K parameters, the optimal stochastic parameter design results in time sharing (randomization) among at most two and $(K + 1)$ different signals values for the total Bayes risk and the maximum Bayes risk criteria, respectively. In addition, the average transmitted signal powers corresponding to the optimal parameter design approaches have been specified, and the characterization of the optimal approaches has been provided in various scenarios. Furthermore, various conditions have been derived in order to specify when the stochastic parameter design or the deterministic parameter design is optimal. Finally, the numerical examples have been presented to investigate the theoretical results, and to illustrate the amount of improvements achieved via the proposed approach.

Chapter 4

Optimal Parameter Design for LMMSE Estimators

In this chapter, the jointly optimal design of parameter dependent signals and the estimator is studied considering the class of LMMSE estimators. This chapter is organized as follows: In Section 4.1, a brief overview of LMMSE estimators is presented. The optimal parameter design is formulated and some properties of the optimal design are obtained in Section 4.2. Numerical examples are presented in Section 4.3 and finally the concluding remarks are made in Section 4.4.

4.1 LMMSE Estimator

Consider a parameter estimation scenario, in which the parameter to be estimated is denoted by $\boldsymbol{\theta}$, and the estimator observes this parameter over an additive noise channel. Then, the received signal (observation) at the estimator is expressed as

$$\mathbf{y} = \boldsymbol{\theta} + \mathbf{n} \quad (4.1)$$

where $\mathbf{n}$ denotes the channel noise, which has a probability density function (PDF) represented by $p_{\mathbf{n}}(\cdot)$. In order to estimate $\boldsymbol{\theta}$ based on $\mathbf{y}$, the LMMSE estimator is employed, which is represented by $\hat{\boldsymbol{\theta}}(\mathbf{y})$.

First assume that the parameter $\boldsymbol{\theta}$ to be estimated is scalar and the estimation is to be performed based on a set of observations $\mathbf{y}[0], \mathbf{y}[1], \dots, \mathbf{y}[K-1]$. Since we consider a linear estimator, the estimation of the unknown parameter based on the observations is performed as follows:

$$\hat{\boldsymbol{\theta}}(\mathbf{y}) = \sum_{k=0}^{K-1} a_k \mathbf{y}[k] + a_K . \quad (4.2)$$

For the MMSE criterion, the estimator can be specified by finding the weighting coefficients a_k 's which minimize the Bayesian mean squared error (MSE):

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = E\{(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})^2\} \quad (4.3)$$

where the expectation is taken over the joint probability density function of $\boldsymbol{\theta}$ and $\mathbf{y}$ [1].

To find the optimal weighting coefficients; hence, the optimal estimator, we substitute (4.2) into (4.3), differentiate with respect to a_K , and set it to zero, which yields

$$a_K = E\{\boldsymbol{\theta}\} - \sum_{k=0}^{K-1} a_k E\{\mathbf{y}[k]\} \quad (4.4)$$

Then, using this value of a_K , the MSE should be minimized over a_k 's. Defining $\mathbf{a} = [a_0, a_1, \dots, a_{K-1}]^T$, the MSE can be written as

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = \mathbf{a}^T C_{\mathbf{y}\mathbf{y}} \mathbf{a} - \mathbf{a}^T C_{\mathbf{y}\boldsymbol{\theta}} - C_{\boldsymbol{\theta}\mathbf{y}} \mathbf{a} + C_{\boldsymbol{\theta}\boldsymbol{\theta}} \quad (4.5)$$

where $C_{\mathbf{y}\mathbf{y}}$ is the $K \times K$ covariance matrix of $\mathbf{y}$, $C_{\boldsymbol{\theta}\mathbf{y}}$ is the $1 \times K$ cross covariance vector, such that $C_{\boldsymbol{\theta}\mathbf{y}}^T = C_{\mathbf{y}\boldsymbol{\theta}}$, and $C_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the variance of $\boldsymbol{\theta}$ [1]. To minimize (4.5) we should take the gradient of it with respect to $\mathbf{a}$ and set it to zero, which yields

$$\mathbf{a} = C_{\mathbf{y}\mathbf{y}}^{-1} C_{\mathbf{y}\boldsymbol{\theta}} . \quad (4.6)$$

Using these optimal coefficients, the LMMSE is found as

$$\hat{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} + C_{\boldsymbol{\theta}\mathbf{y}} C_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{y} - E\{\mathbf{y}\}) \quad (4.7)$$

and the minimum risk achieved by this estimator is given by

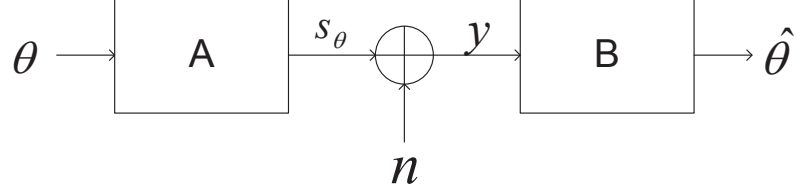


Figure 4.1: System model. Device A transmits a stochastic signal $\mathbf{s}_\theta$ for each value of parameter θ , and Device B employs an LMMSE estimator and estimates θ based on the noise corrupted version of $\mathbf{s}_\theta$.

$$\text{MSE}(\hat{\theta}) = C_{\theta\theta} - C_{\theta\mathbf{y}}C_{\mathbf{y}\mathbf{y}}^{-1}C_{\mathbf{y}\theta}. \quad (4.8)$$

When the unknown parameter θ is a p dimensional vector instead of a scalar, the optimal estimator and the corresponding risk can be found via similar arguments. The optimal estimator is given by (4.7) again but this time $C_{\theta\mathbf{y}}$ is not a $1 \times N$ cross covariance vector but a $p \times N$ covariance matrix. Similarly, the minimum risk calculated by (4.8) is not a scalar but a $p \times p$ matrix, and its diagonal elements give the minimum risk corresponding to the estimation of each element of the parameter vector θ [1].

4.2 Optimal Parameter Design for LMMSE Estimator

Consider a parameter estimation scenario as in Figure 4.1, where the aim is to send the information about parameter θ to Device A over an additive noise channel. Then, the received signal (observation) at Device A is expressed as

$$\mathbf{y} = \mathbf{s}_\theta + \mathbf{n} \quad (4.9)$$

where $\mathbf{n}$ denotes the channel noise, which has a probability density function (PDF) represented by $p_{\mathbf{n}}(\cdot)$. It is assumed that Device B employs an LMMSE estimator specified by $\hat{\boldsymbol{\theta}}(\mathbf{y})$ in order to estimate $\boldsymbol{\theta}$.

In this section, the problem is investigated for the case where the parameter $\boldsymbol{\theta}$, hence, the observation $\mathbf{y}$, are single scalar parameters. For this case, the MSE in (4.8) can be written as follows:

$$\text{MSE} = \text{E}[(\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))^2] - \frac{\text{E}[(\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))(\mathbf{y} - \text{E}(\mathbf{y}))]^2}{\text{E}[(\mathbf{y} - \text{E}(\mathbf{y}))^2]} . \quad (4.10)$$

Here it can be seen that this error depends on two variables, $\boldsymbol{\theta}$ and $\mathbf{y}$. To achieve a smaller error we can only change the observation $\mathbf{y}$ since we have no control over $\boldsymbol{\theta}$. It is noted from (4.9) that the observation consists of two parts, one of which is the noise that we cannot control either. Therefore, we can try to achieve the minimum risk only by designing an appropriate $\mathbf{s}_{\boldsymbol{\theta}}$.

To minimize the MSE in (4.10), it is sufficient to maximize the second term as the MSE is always non-negative by definition, which can also be shown via Cauchy-Schwarz inequality as follows:

$$\begin{aligned} \text{MSE} &= \text{E}[(\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))^2] - \frac{\text{E}[(\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))(\mathbf{y} - \text{E}(\mathbf{y}))]^2}{\text{E}[(\mathbf{y} - \text{E}(\mathbf{y}))^2]} \\ &\geq \text{E}[(\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))^2] - \frac{\text{E}[(\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))^2]\text{E}[(\mathbf{y} - \text{E}(\mathbf{y}))^2]}{\text{E}[(\mathbf{y} - \text{E}(\mathbf{y}))^2]} \\ &\geq 0 . \end{aligned} \quad (4.11)$$

Therefore, we can just consider the second term in (4.10) and minimize the MSE by maximizing this term. Then, the problem takes the following form:

$$\{p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}, \boldsymbol{\theta} \in \Lambda\} = \arg \max_{\{p_{\mathbf{s}_{\boldsymbol{\theta}}}, \boldsymbol{\theta} \in \Lambda\}} \frac{\text{E}[(\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))(\mathbf{y} - \text{E}(\mathbf{y}))]^2}{\text{E}[(\mathbf{y} - \text{E}(\mathbf{y}))^2]} . \quad (4.12)$$

Substituting (4.9) into (4.12) we can write down the problem as follows:

$$\begin{aligned} \{p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\text{opt}}, \boldsymbol{\theta} \in \Lambda\} &= \arg \max_{\{p_{\mathbf{s}_{\boldsymbol{\theta}}}, \boldsymbol{\theta} \in \Lambda\}} \frac{\text{E}[(\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))(\mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n} - \text{E}(\mathbf{s}_{\boldsymbol{\theta}}) - \text{E}(\mathbf{n}))]^2}{\text{E}[(\mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n} - \text{E}(\mathbf{s}_{\boldsymbol{\theta}}) - \text{E}(\mathbf{n}))^2]} \\ &= \arg \max_{\{p_{\mathbf{s}_{\boldsymbol{\theta}}}, \boldsymbol{\theta} \in \Lambda\}} \frac{[\text{E}(\boldsymbol{\theta}\mathbf{s}_{\boldsymbol{\theta}}) - \text{E}(\boldsymbol{\theta})\text{E}(\mathbf{s}_{\boldsymbol{\theta}})]^2}{\text{E}(\mathbf{s}_{\boldsymbol{\theta}}^2) - [\text{E}(\mathbf{s}_{\boldsymbol{\theta}})]^2 + \sigma_n^2} \\ &= \arg \max_{\{p_{\mathbf{s}_{\boldsymbol{\theta}}}, \boldsymbol{\theta} \in \Lambda\}} \frac{[\int (\boldsymbol{\theta} - \text{E}(\boldsymbol{\theta}))\omega(\boldsymbol{\theta})\text{E}(\mathbf{s}_{\boldsymbol{\theta}}|\boldsymbol{\theta})d\boldsymbol{\theta}]^2}{\int \omega(\boldsymbol{\theta})\text{E}(\mathbf{s}_{\boldsymbol{\theta}}^2|\boldsymbol{\theta})d\boldsymbol{\theta} - [\int \omega(\boldsymbol{\theta})\text{E}(\mathbf{s}_{\boldsymbol{\theta}}|\boldsymbol{\theta})d\boldsymbol{\theta}]^2 + \sigma_n^2} , \end{aligned} \quad (4.13)$$

which shows that the design depends only on the first and second order moments of p_{s_θ} . Hence, any distribution with the same values of the first and second order moments yields the same result.

To simplify the problem, consider a discrete set of values for θ , i.e., $\theta \in \{\theta_1, \theta_2, \dots, \theta_L\}$ and define the following vectors:

$$\begin{aligned}\mathbf{a} &\triangleq [(\theta_1 - E(\theta))\omega(\theta_1) \cdots (\theta_L - E(\theta))\omega(\theta_L)] \\ \mathbf{b} &\triangleq [\omega(\theta_1) \cdots \omega(\theta_L)] \\ \mathbf{x} &\triangleq [E(s_\theta|\theta = \theta_1) \cdots E(s_\theta|\theta = \theta_L)] \\ \mathbf{z} &\triangleq [E(s_\theta^2|\theta = \theta_1) \cdots E(s_\theta^2|\theta = \theta_L)]\end{aligned}\tag{4.14}$$

Based on these vectors, the problem in (4.12) can be restated as:

$$\begin{aligned}\{\mathbf{x}^*, \mathbf{z}^*\} &= \arg \max_{\{\mathbf{x}, \mathbf{z}\}} \frac{(\mathbf{a}^T \mathbf{x})^2}{\mathbf{b}^T \mathbf{z} - (\mathbf{b}^T \mathbf{x})^2 + \sigma_n^2} \\ &= \arg \max_{\{\mathbf{x}, \mathbf{z}\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{b}^T \mathbf{z} + \sigma_n^2 - \mathbf{x}^T \mathbf{B} \mathbf{x}}\end{aligned}\tag{4.15}$$

where $\mathbf{A} \triangleq \mathbf{a}\mathbf{a}^T$ and $\mathbf{B} \triangleq \mathbf{b}\mathbf{b}^T$. Here we end up with an optimization problem with two vectors which depend on each other. We need to find the optimal $\{\mathbf{x}^*, \mathbf{z}^*\}$ pair but these vectors depend on each other. Besides any constraint that can be imposed on this problem, the optimal $\{\mathbf{x}^*, \mathbf{z}^*\}$ pair should satisfy $(\mathbf{x}_i^*)^2 \leq \mathbf{z}_i^*$ for all elements of these vectors due to Jensen's inequality, which states that for any random variable Y , $f(E(Y)) \leq E(f(Y))$ for any convex function f ; hence, $(E(Y))^2 \leq E(Y^2)$ in our case. It is possible to find a solution to (4.15) which does not satisfy this condition but that solution will not be a legitimate one because it will not be possible to find a probability distribution for that $\{\mathbf{x}^*, \mathbf{z}^*\}$ pair. From this point on, this condition will be considered for each solution but will not be stated over and over again. The equality case of Jensen's inequality will be important in our analysis since the equality occurs, i.e., $(E(Y))^2 = E(Y^2)$ when the random variable Y has a probability distribution with one point mass. Therefore, if all the elements of the optimal $\{\mathbf{x}^*, \mathbf{z}^*\}$ pair satisfy $(\mathbf{x}_i^*)^2 = \mathbf{z}_i^*$, then it means that no stochastic design is necessary and the deterministic design is optimal.

Since the optimization problem in (4.15) is quite complicated, it can be useful to investigate its general characteristics before trying to obtain the solution. First of all, none of the four terms in (4.15) is negative. Namely, $\mathbf{x}^T \mathbf{A} \mathbf{x}$, $\mathbf{x}^T \mathbf{B} \mathbf{x}$ and σ_n^2 are all non-negative since they are the squares of some other terms. On the other hand, $\mathbf{b}^T \mathbf{z}$ is strictly positive since $\mathbf{b}^T \mathbf{z} = \mathbb{E}(\mathbf{s}_\theta^2)$, which is the expectation of the square of a random variable. Using the condition stated above, $\mathbf{x}_i^2 \leq \mathbf{z}_i$, it is possible to simplify the $\mathbf{b}^T \mathbf{z}$ term. Since $\mathbf{b}$ is a non-negative vector and each element of vector $\mathbf{z}$ can be written as $\mathbf{z}_i = \mathbf{x}_i^2 + \epsilon_i$, where $\epsilon_i \geq 0$, $\mathbf{b}^T \mathbf{z}$ can be expressed as

$$\mathbf{b}^T \mathbf{z} = \mathbf{x}^T \mathbf{D} \mathbf{x} + \epsilon \quad (4.16)$$

where

$$\mathbf{D} \triangleq \text{diag}\{\omega(\theta_1), \dots, \omega(\theta_L)\} \quad (4.17)$$

and $\epsilon \geq 0$. From (4.16), (4.15) can be updated as follows:

$$\begin{aligned} \{\mathbf{x}^*, \epsilon^*\} &= \arg \max_{\{\mathbf{x}, \epsilon \geq 0\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{D} \mathbf{x} + \epsilon + \sigma_n^2 - \mathbf{x}^T \mathbf{B} \mathbf{x}} \\ &= \arg \max_{\{\mathbf{x}, \epsilon \geq 0\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T (\mathbf{D} - \mathbf{B}) \mathbf{x} + \epsilon + \sigma_n^2}. \end{aligned} \quad (4.18)$$

With this modification, the optimization problem in (4.15) is simplified by reducing the dimension of the optimization problem.

Another observation from (4.15) is that the denominator of the objective function is strictly positive. Due to Jensen's inequality, $\mathbf{x}^T \mathbf{B} \mathbf{x} \leq \mathbf{b}^T \mathbf{z}$, where $\mathbf{x}^T \mathbf{B} \mathbf{x} = [\mathbb{E}(\mathbf{s}_\theta)]^2$ and $\mathbf{b}^T \mathbf{z} = \mathbb{E}(\mathbf{s}_\theta^2)$. Hence, (4.15) is always non-negative and never goes to infinity since the denominator does not become zero. Therefore, the formulation in (4.18) assures that the condition stated above, $\mathbf{x}_i^2 \leq \mathbf{z}_i$, is satisfied for all elements of these vectors and as the result the solution of (4.18) always yields a legitimate solution.

In practical scenarios, the parameter design cannot be performed without any limitations. For example, in the absence of a power constraint, it would be possible to reduce the Bayes risk arbitrarily by transmitting signals with very

high powers compared to the noise power. Consider an average power constraint in the form of

$$\mathbb{E}\{\|\mathbf{s}_\theta\|^2\} \leq A_\theta \quad (4.19)$$

which is equivalent to

$$\mathbf{b}^T \mathbf{z} \leq A_\theta \quad (4.20)$$

where $\|\mathbf{s}_\theta\|$ is the Euclidean norm of vector $\mathbf{s}_\theta$, and A_θ denotes the average power constraint for θ . It is noted from (4.19) that a generic model is considered for the constraint A_θ , which can depend on the value of θ in general. Together with the constraint, the problem in (4.18) can be restated as

$$\begin{aligned} \{\mathbf{x}^*, \epsilon^*\} = \arg \max_{\{\mathbf{x}, \epsilon\}} & \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T (\mathbf{D} - \mathbf{B}) \mathbf{x} + \epsilon + \sigma_n^2} \\ \text{subject to } & \mathbf{x}^T \mathbf{D} \mathbf{x} + \epsilon \leq A_\theta, \quad \epsilon \geq 0 \end{aligned} \quad (4.21)$$

As stated before, if the stochastic design is the optimal solution of (4.15) then at least for one element of the $\{\mathbf{x}^*, \mathbf{z}^*\}$ pair, the condition $(\mathbf{x}_i^*)^2 < \mathbf{z}_i^*$ should be satisfied, which means that $\epsilon^* > 0$ in (4.21). Suppose that the stochastic design is the optimal solution of (4.21), i.e., $\epsilon^* > 0$. Considering the arguments in the previous section, since none of the terms in (4.21) is negative, decreasing the value of ϵ always increases the value of (4.21); hence, decreases the error of the system. Therefore, the optimal solution of (4.21) requires the minimum value of ϵ^* , which is 0, implying that the optimal solution of (4.21) corresponds to the deterministic design. Therefore, it is concluded that when there is a power constraint in the optimization, the deterministic design is the optimal solution. Since the deterministic design is optimal, it is possible to simplify (4.21) further using the fact that $\epsilon^* = 0$ as follows:

$$\begin{aligned} \mathbf{x}^* = \arg \max_{\{\mathbf{x}\}} & \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T (\mathbf{D} - \mathbf{B}) \mathbf{x} + \sigma_n^2} \\ \text{subject to } & \mathbf{x}^T \mathbf{D} \mathbf{x} \leq A_\theta. \end{aligned} \quad (4.22)$$

This result is not only valid for the cases with a power constraint. Based on similar arguments, it is possible to show that the deterministic design is optimal in the absence of power constraint as well.

Proposition 4.1 *The solution $\mathbf{x}^*$ of the problem in (4.22) satisfies the constraint with equality, that is, $\mathbf{x}^{*T} \mathbf{D} \mathbf{x}^* = A_\theta$.*

Proof: Assume that the solution of (4.22) satisfies the constraint as follows: $\mathbf{x}^{*T} \mathbf{D} \mathbf{x}^* = A_\theta^*$, where $A_\theta = \alpha A_\theta^*$ with $\alpha > 1$. In that case, this optimal solution should yield a greater cost value than the solution at the boundary as follows:

$$\frac{\mathbf{x}^{*T} \mathbf{A} \mathbf{x}^*}{\mathbf{x}^{*T} (\mathbf{D} - \mathbf{B}) \mathbf{x}^* + \sigma_n^2} > \frac{\sqrt{\alpha} \mathbf{x}^{*T} \mathbf{A} \sqrt{\alpha} \mathbf{x}^*}{\sqrt{\alpha} \mathbf{x}^{*T} (\mathbf{D} - \mathbf{B}) \sqrt{\alpha} \mathbf{x}^* + \sigma_n^2} . \quad (4.23)$$

This function can be simplified as

$$\frac{\mathbf{x}^{*T} \mathbf{A} \mathbf{x}^*}{\mathbf{x}^{*T} (\mathbf{D} - \mathbf{B}) \mathbf{x}^* + \sigma_n^2} > \alpha \frac{\mathbf{x}^{*T} \mathbf{A} \mathbf{x}^*}{\mathbf{x}^{*T} (\mathbf{D} - \mathbf{B}) \mathbf{x}^* + \sigma_n^2 / \alpha} . \quad (4.24)$$

Previously it was shown that each term in (4.24) is positive and as a result of that

$$\frac{\mathbf{x}^{*T} \mathbf{A} \mathbf{x}^*}{\mathbf{x}^{*T} (\mathbf{D} - \mathbf{B}) \mathbf{x}^* + \sigma_n^2} < \frac{\mathbf{x}^{*T} \mathbf{A} \mathbf{x}^*}{\mathbf{x}^{*T} (\mathbf{D} - \mathbf{B}) \mathbf{x}^* + \sigma_n^2 / \alpha} . \quad (4.25)$$

Considering the fact that $\alpha > 1$, (4.25) contradicts with (4.24) and the assumption made at the beginning of the proof as well. Therefore, it is concluded that the optimal solution is always at the boundary of the constraint; that is, $\mathbf{x}^{*T} \mathbf{D} \mathbf{x}^* = A_\theta$. ■

Based on Proposition 4.1, the problem in (4.22) can be formulated as the Rayleigh quotient as follows:

$$\begin{aligned} \mathbf{x}^* &= \arg \max_{\{\mathbf{x}\}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}} \\ &\text{subject to } \mathbf{x}^T \mathbf{D} \mathbf{x} = A_\theta \end{aligned} \quad (4.26)$$

where $\mathbf{G} \triangleq \mathbf{D}(1 + \sigma_n^2 / A_\theta) - \mathbf{B}$. The problem in (4.26) can be written as a generalized eigenvalue problem if $\mathbf{A}$ and $\mathbf{G}$ are symmetric matrices and $\mathbf{G}$ is also a positive definite matrix [47]. Here $\mathbf{A}$ is a symmetric matrix since it is defined as $\mathbf{A} = \mathbf{a} \mathbf{a}^T$. $\mathbf{G}$ is also a symmetric matrix as $\mathbf{D}$ is a diagonal matrix and $\mathbf{B}$ is defined as $\mathbf{B} = \mathbf{b} \mathbf{b}^T$. It is also possible to show that $\mathbf{G}$ is a positive definite matrix. Since $\mathbf{x}^T \mathbf{G} \mathbf{x} = \mathbf{x}^T (\mathbf{D} - \mathbf{B}) \mathbf{x} + \sigma_n^2$, if we can show that $\mathbf{x}^T (\mathbf{D} - \mathbf{B}) \mathbf{x} + \sigma_n^2 > 0$,

it can then be concluded that $\mathbf{G}$ is a positive definite matrix. Due to Jensen's inequality $\mathbf{x}^T(\mathbf{D} - \mathbf{B})\mathbf{x} > 0$, unless $\mathbf{s}_\theta$ has just a single value, since $\mathbf{x}^T \mathbf{D} \mathbf{x} = E(\mathbf{s}_\theta^2)$ and $\mathbf{x}^T \mathbf{B} \mathbf{x} = E(\mathbf{s}_\theta)^2$. Also σ_n^2 has a non-negative value. Therefore, as long as $\sigma_n^2 \neq 0$ or $\mathbf{s}_\theta$ has more than one possible value, $\mathbf{G}$ is positive definite. In that case, the solution of the problem in (4.26) can be modified as the generalized eigenvalue problem as $\mathbf{A}\mathbf{x} = \lambda \mathbf{G}\mathbf{x}$, and the solution to (4.26) is the generalized eigenvector with the maximum eigenvalue. This eigenvector should be scaled to satisfy the power constraint as well.

Proposition 4.2 *The solution of the problem in (4.26) does not depend on the value of σ_n , i.e., the optimal solution is independent of the noise variance of the channel.*

Proof: Matrix $\mathbf{G}$ in (4.26) is defined as $\mathbf{G} = \mathbf{D}(1 + \sigma_n^2/A_\theta) - \mathbf{B}$. Here $\mathbf{D}$ is a diagonal matrix, hence it is full rank. As a result, $\mathbf{G}$ is an invertible matrix and using this fact, the problem in (4.26) can be written as an eigenvalue problem as $\mathbf{G}^{-1}\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$. $\mathbf{G}$ can be rewritten as $\mathbf{G} = \underline{\mathbf{D}} + \underline{\mathbf{B}}$, where $\underline{\mathbf{D}} = \alpha \mathbf{D}$, $\underline{\mathbf{B}} = -\mathbf{B}$ and $\alpha = 1 + \sigma_n^2/A_\theta$. In that case, the inverse of $\mathbf{G}$ can be expressed as [48]

$$\mathbf{G}^{-1} = \underline{\mathbf{D}}^{-1} - \frac{\underline{\mathbf{D}}^{-1} \underline{\mathbf{B}} \underline{\mathbf{D}}^{-1}}{1 + \text{trace}(\underline{\mathbf{B}} \underline{\mathbf{D}}^{-1})} . \quad (4.27)$$

Using the definitions given above, $\underline{\mathbf{D}}^{-1}$ and $\underline{\mathbf{B}}$ can be written as

$$\underline{\mathbf{D}}^{-1} = \frac{1}{\alpha} \begin{bmatrix} 1/(\omega(\boldsymbol{\theta}_1)) & 0 & . & . & . & 0 \\ 0 & 1/(\omega(\boldsymbol{\theta}_2)) & 0 & . & . & 0 \\ . & 0 & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & . & . & 1/(\omega(\boldsymbol{\theta}_L)) \end{bmatrix} \quad (4.28)$$

$$\underline{\mathbf{B}} = - \begin{bmatrix} (\omega(\boldsymbol{\theta}_1))^2 & (\omega(\boldsymbol{\theta}_1))(\omega(\boldsymbol{\theta}_2)) & . & . & . & (\omega(\boldsymbol{\theta}_1))(\omega(\boldsymbol{\theta}_L)) \\ (\omega(\boldsymbol{\theta}_1))(\omega(\boldsymbol{\theta}_2)) & (\omega(\boldsymbol{\theta}_2))^2 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ (\omega(\boldsymbol{\theta}_1))(\omega(\boldsymbol{\theta}_L)) & . & . & . & . & (\omega(\boldsymbol{\theta}_L))^2 \end{bmatrix} . \quad (4.29)$$

Using these matrices, the $\underline{\mathbf{D}}^{-1}\underline{\mathbf{B}}\underline{\mathbf{D}}^{-1}$ term in (4.27) can be calculated as follows:

$$\underline{\mathbf{D}}^{-1}\underline{\mathbf{B}}\underline{\mathbf{D}}^{-1} = \frac{-1}{\alpha^2} \begin{bmatrix} 1 & 1 & . & . & . & 1 \\ 1 & 1 & . & . & . & 1 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & . & . & . & . & 1 \end{bmatrix} . \quad (4.30)$$

Since the problem in (4.26) can be simplified to the eigenvalue problem as $\mathbf{G}^{-1}\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, it is necessary to find the multiplication of $\mathbf{G}^{-1}$ and $\mathbf{A}$. Using the definition of $\mathbf{G}^{-1}$ given in (4.27), this multiplication can be calculated as

$$\mathbf{G}^{-1}\mathbf{A} = \underline{\mathbf{D}}^{-1}\mathbf{A} - \frac{\underline{\mathbf{D}}^{-1}\underline{\mathbf{B}}\underline{\mathbf{D}}^{-1}\mathbf{A}}{1 + \text{trace}(\underline{\mathbf{B}}\underline{\mathbf{D}}^{-1})} . \quad (4.31)$$

Using the definition of $\mathbf{A}$ in the previous section, the $\underline{\mathbf{D}}^{-1}\underline{\mathbf{B}}\underline{\mathbf{D}}^{-1}\mathbf{A}$ term in (4.31) can be calculated as

$$\underline{\mathbf{D}}^{-1}\underline{\mathbf{B}}\underline{\mathbf{D}}^{-1}\mathbf{A} = \frac{-1}{\alpha^2} \begin{bmatrix} 0 & 0 & . & . & . & 0 \\ 0 & 0 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & . & . & 0 \end{bmatrix} .$$

Based on this result, it can be found that the second term in (4.31) is always zero. As a result of this, the eigenvalue problem of $\mathbf{G}^{-1}\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ can be modified as $\underline{\mathbf{D}}^{-1}\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Using the definition of $\underline{\mathbf{D}}$, this eigenvalue problem can be modified further as $(1 + \sigma_n^2/A_\theta)^{-1}\mathbf{D}^{-1}\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Here σ_n appears as a scaling factor and does not effect the solution of this eigenvalue problem. Therefore it can be concluded that the optimal solution of (4.26) is independent of the channel characteristics. ■

4.3 Numerical Examples

In this section, numerical examples are presented to investigate the performance of the optimal parameter design for an LMMSE estimation problem. Parameter

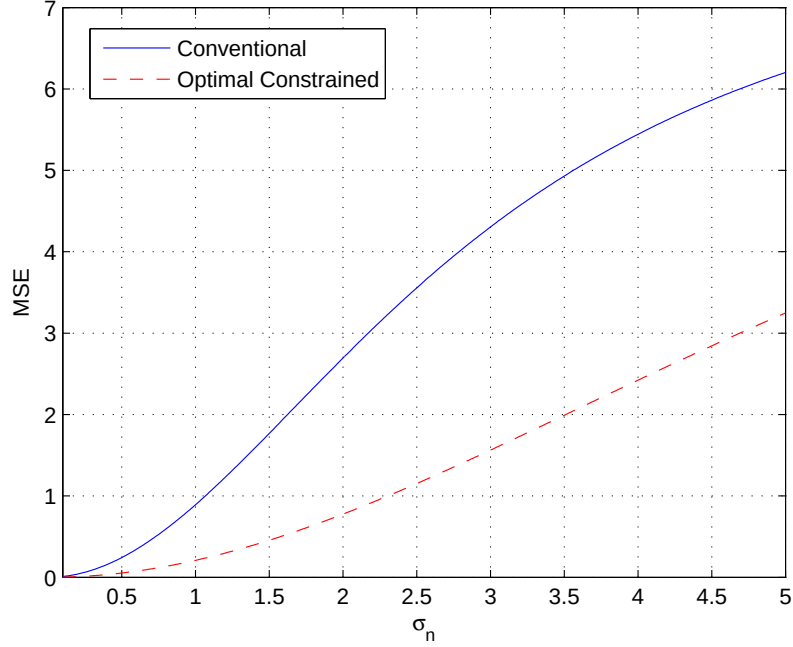


Figure 4.2: MSE values of the first case for the conventional approach and the optimal parameter design.

θ is assumed to take all integer values between 1 and 10, i.e., $\theta \in \{1, 2, \dots, 10\}$. The power constraint is set to be equal to the average power of θ , i.e., $A_\theta = E[||\theta||^2]$. For this set of θ values and the power constraint, error performances of the conventional approach and the optimal parameter design are compared for different sets of prior probabilities of θ and various values of σ_n .

Three different cases are investigated for three different sets of prior probabilities that are considered for θ . In the first case, equal prior probabilities are assigned to the possible of θ . In the second case, linearly decreasing prior probabilities are considered, whereas for the final case linearly increasing prior probabilities are considered. For each case, the prior probabilities of the θ values can be seen in Table 4.1. For each case, the MSE values of the LMMSE estimator for both the conventional approach and the optimal parameter design are presented in Figure 4.2, Figure 4.3, and Figure 4.4. As it can be observed from all these figures, the optimal parameter design performs significantly better than

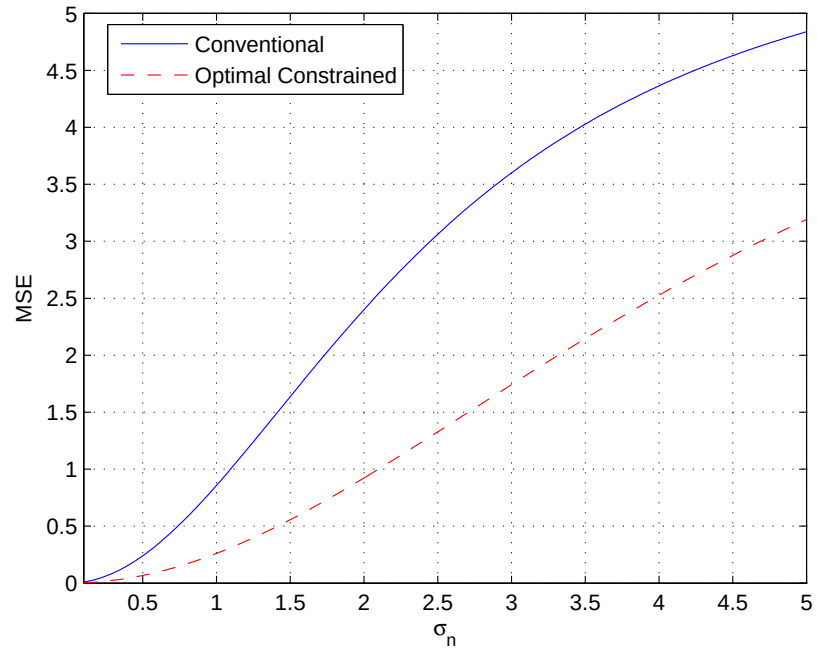


Figure 4.3: MSE values of the second case for the conventional approach and the optimal parameter design.

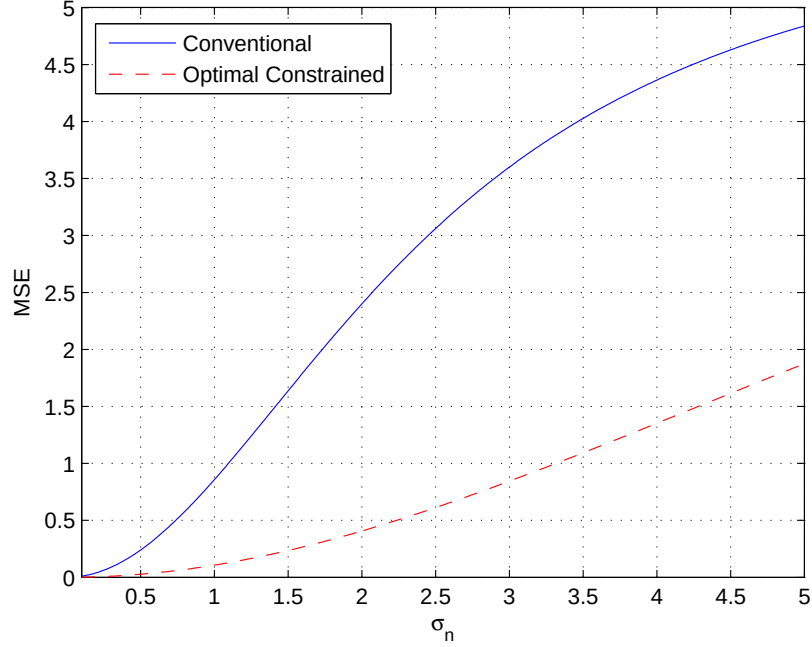


Figure 4.4: MSE values of the third case for the conventional approach and the optimal parameter design.

the conventional approach. Although there is significant improvement in each of these cases, the highest improvement is observed in the third case and the lowest improvement is observed in the second case. This result can be explained by the fact that the power constraint of the third case is the highest, whereas for the second case the power constraint is the lowest. In Table 4.1, the $\mathbf{s}_\theta$ values for each value of θ are given for each case. Here it can be observed that the $\mathbf{s}_\theta$ values are spread to a wider interval than the θ values. Therefore, the MSE values for $\mathbf{s}_\theta$ are lower than the MSE values for θ . As stated previously in Proposition 4.2, the solution does not depend on the value of σ_n , hence, the $\mathbf{s}_\theta$ values given in Table 4.1 are valid for all values of σ_n .

Table 4.1: Prior probabilities and $\mathbf{s}_\theta$ values for the three cases.

-	-	First Case		Second Case		Third Case	
-	θ	$\omega(\theta)$	$\mathbf{s}_\theta$	$\omega(\theta)$	$\mathbf{s}_\theta$	$\omega(\theta)$	$\mathbf{s}_\theta$
θ_1	1	0.1	-9.7211	0.1818	-5.7446	0.0182	-18.1659
θ_2	2	0.1	-7.5609	0.1636	-3.8297	0.0364	-15.1383
θ_3	3	0.1	-5.4006	0.1455	-1.9149	0.0545	-12.1106
θ_4	4	0.1	-3.2404	0.1273	0	0.0727	-9.0830
θ_5	5	0.1	-1.0801	0.1091	1.9149	0.0909	-6.0553
θ_6	6	0.1	1.0801	0.0909	3.8297	0.1091	-3.0277
θ_7	7	0.1	3.2404	0.0727	5.7446	0.1273	0
θ_8	8	0.1	5.4006	0.0545	7.6594	0.1455	3.0277
θ_9	9	0.1	7.5609	0.0364	9.5743	0.1636	6.0553
θ_{10}	10	0.1	9.7211	0.0182	11.4891	0.1818	9.0830

4.4 Conclusions for Chapter 4

In this chapter, the jointly optimal design of parameter dependent signals and the estimator has been investigated for LMMSE estimators. It has been shown that the optimal strategy corresponds to the deterministic design when there is an average power constraint. It has also been concluded that the deterministic design is optimal when there is no constraint as well. In addition, it has been proved that the optimal parameter value is independent of the channel characteristics. Furthermore, it has been shown that the average transmitted power is always equal to the average power limit. Finally, numerical examples have been presented to investigate the theoretical results.

Chapter 5

Optimal Channel Switching over Gaussian Channels under Average Power and Cost Constraints

In this chapter, the optimal channel switching problem is formulated for Gaussian channels in the presence of average power and average cost constraints. This chapter is organized as follows: The system model and problem formulation are introduced in Section 5.1, and the optimal channel switching problem is studied for generic cost functions in Section 5.2. In Section 5.3, logarithmic functions are considered for the optimal channel switching problem, and numerical examples are presented in Section 5.4. Finally, extensions to M -ary systems are presented in Section 5.5 and some concluding remarks are provided in Section 5.6.

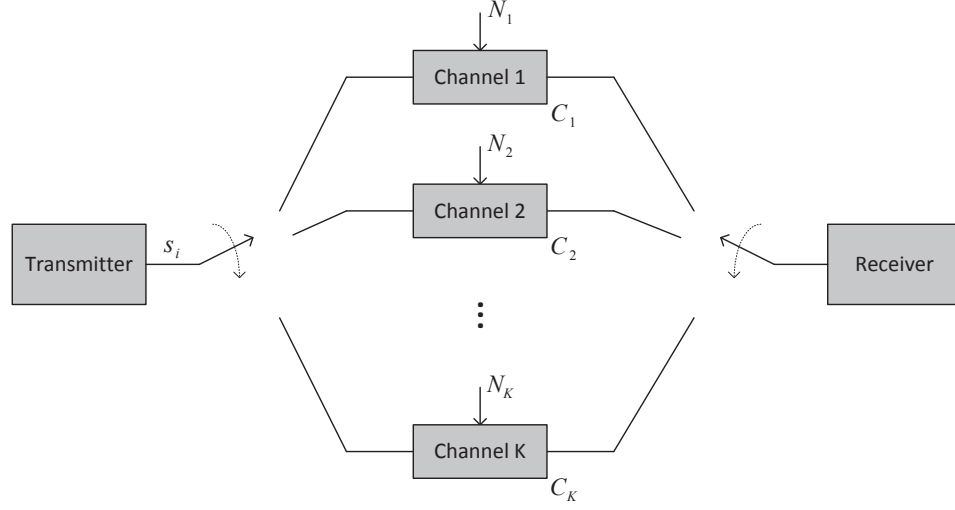


Figure 5.1: A binary communication system that employs channel switching among K additive Gaussian noise channels, where C_i denotes the cost of using channel i , and N_i is the noise component at the i th channel. The transmitter employs antipodal signaling, $\{-s_i, s_i\}$, and the receiver performs sign detection, which is optimal for the considered scenario.

5.1 System Model and Problem Formulation

Consider a communication system in which K additive noise channels are available between the transmitter and the receiver, as shown in Figure 5.1. The transmitter-receiver pair can synchronously switch among those K channels over time; i.e., they can perform time sharing among different channels by employing only one channel at a given time [17, 28]. The channels are corrupted by independent zero-mean Gaussian noise components, denoted by N_i , where $N_i \sim \mathcal{N}(0, \sigma_i^2)$ for $i \in \{1, \dots, K\}$, with σ_i^2 denoting the noise variance of channel i . In addition, there is a cost associated with the usage of each channel, denoted by C_i for $i \in \{1, \dots, K\}$. The cost values are specified by nonnegative numbers and they satisfy $C_i > C_j$ if $\sigma_i^2 < \sigma_j^2$ for all $i \neq j$. In other words, if a channel has a smaller (larger) average noise power, it has a higher (lower) cost. Assigning costs to different channels or measurement devices has various motivations and implications, as discussed for example in [35]-[23].

A binary communication system that employs antipodal signaling is considered; that is, $s_i^{(j)}$ is transmitted over channel i for the j th information symbol, where $j \in \{0, 1\}$ and $s_i^{(0)} = -s_i^{(1)}$ for $i \in \{1 \dots K\}$ (extensions to the M -ary case are discussed in Section 5.6). The received signal corresponding to the i th channel can be expressed as

$$Y = s_i^{(j)} + N_i, \quad j \in \{0, 1\} \quad \text{and} \quad i \in \{1 \dots K\}, \quad (5.1)$$

where N_i is the Gaussian noise component over channel i , which is independent of $s_i^{(j)}$.

It is assumed that the symbols are equally likely; that is, the prior probabilities of signals $s_i^{(0)}$ and $s_i^{(1)}$ are equal to 0.5 for $i \in \{1, \dots, K\}$. Then, the optimal receiver to decide between the two symbols is the sign detector for all channels [2]. Specifically, when the received signal in (5.1) is given by $Y = y$, the optimal decision rule for channel i is expressed as $\phi_i(y) = 0$ if $y < 0$ and $\phi_i(y) = 1$ if $y \geq 0$ for $i \in \{1 \dots K\}$. It is important to note that the receiver is the same, namely, the sign detector, for all channels due to antipodal signaling and equally likely symbols. Therefore, the receiver does not need to be adapted as channel switching occurs, which can be useful for practical applications.

For notational simplicity, the superscript in $s_i^{(1)}$ is dropped in the remainder of the manuscript; i.e., for channel i , the transmitted signal for information symbol 1 is denoted by s_i , where $s_i > 0$ without loss of generality. Similarly, the transmitted signal for symbol 0 is denoted by $-s_i$ for channel i . Let λ_i denote the fraction of time during which channel i is employed for transmission, which is called the *channel switching factor* for channel i . The channel switching factors satisfy $\sum_{i=1}^K \lambda_i = 1$ and $\lambda_i \geq 0 \forall i \in \{1 \dots K\}$. In practice, channel switching is performed by utilizing the i th channel for $100\lambda_i$ percent of time for $i = 1, \dots, K$.

The aim of this study is to jointly optimize the channel switching factors and signal powers in order to minimize the average probability of error under average power and cost constraints. The error probability of channel i can be expressed for the signal model in (5.1) and the sign detector as $Q(s_i/\sigma_i)$, where $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty e^{-t^2/2} dt$ is the Q -function. Therefore, the average probability of error

can be expressed as $\sum_{i=1}^K \lambda_i Q(s_i/\sigma_i)$, where λ_i 's are the channel switching factors. By defining a power term for each channel as $P_i = |s_i|^2$, the average probability of error can also be expressed as $\sum_{i=1}^K \lambda_i g_i(P_i)$, where $g_i(P) \triangleq Q(\sqrt{P}/\sigma_i)$. It is noted that for any two channels, the one with a higher cost always results in a lower error probability for the same power level; that is, if $C_i > C_j$ (which implies $\sigma_i^2 < \sigma_j^2$), then $g_i(P) < g_j(P)$ for all $P > 0$. In practical systems, there exists an average power constraint, which can be expressed as $\sum_{i=1}^K \lambda_i P_i \leq A_p$, where A_p represents the average power limit. In addition, an average transmission cost constraint can be stated as $\sum_{i=1}^K \lambda_i C_i \leq A_c$, where A_c denotes the average cost limit (budget). Then, the proposed optimization problem can be expressed as

$$\begin{aligned} & \min_{\{\lambda_i, P_i\}_{i=1}^K} \sum_{i=1}^K \lambda_i g_i(P_i) \\ \text{subject to } & \sum_{i=1}^K \lambda_i P_i \leq A_p, \quad \sum_{i=1}^K \lambda_i C_i \leq A_c, \\ & \sum_{i=1}^K \lambda_i = 1, \quad \lambda_i \geq 0, \quad \forall i \in \{1 \dots K\}. \end{aligned} \quad (5.2)$$

In other words, the aim is to obtain the optimal channel switching strategy that minimizes the average probability of error under average power and cost constraints. For a motivational example of this formulation, one can consider a cognitive radio system in which there are multiple available frequency bands (channels) in the spectrum, each associated with a certain cost [35]-[39]. Then, the cognitive radio system can perform optimal channel switching in order to minimize the average probability of error under average power and cost constraints.

It is noted from the formulation in (5.2) that the power levels, P_i 's, are modeled as deterministic quantities for each channel; that is, time sharing among different power levels is not considered for a given channel. The reason for this is that since the error probability $g_i(P)$ is a strictly convex function of P , time sharing among different power levels over a given channel can only degrade the performance (i.e., increase the error probability) under an average power constraint. In addition, it is assumed in the remainder of the manuscript that the noise variances σ_i^2 's of the channels are distinct without loss of generality. This is mainly because of the fact that if there are multiple channels with the same noise variances, it is always

better to employ only one of them due to the convexity of the error probability.¹ Hence, the problem formulation that considers only the channels with distinct noise variances is sufficient to achieve the overall optimal solution.

Finally, it is worth mentioning that the results in this study can also be applied to multipath channels with block *frequency-flat* fading under the assumption of perfect channel estimation at the receiver. In that case, the proposed channel switching approach can be employed for each fading state.

5.2 Optimal Channel Switching

In this section, a detailed theoretical investigation of the optimal channel switching problem in (5.2) is presented. In the following analysis, it is assumed without loss of generality that the noise variances of the channels satisfy $\sigma_1^2 < \sigma_2^2 < \dots < \sigma_K^2$, which implies that the cost values are ordered as $C_1 > C_2 > \dots > C_K$. In addition, the average cost limit A_c in (5.2) is assumed to be larger than or equal to the minimum of the cost values; i.e., $A_c \geq C_K$, since (5.2) would yield no solution otherwise. Then, the following remark, which specifies two simple cases, is presented first.

Remark 5.1 (i) *If $A_c = C_K$, the optimal solution of (5.2) is to transmit over channel K exclusively with power A_p .*

(ii) *If $A_c \geq C_1$, the optimal solution of (5.2) is to transmit over channel 1 exclusively with power A_p .*

¹In order to verify this statement, let channel i and channel j have the same noise variances specified by $\sigma_i^2 = \sigma_j^2 = \sigma^2$, and let $g(P) = Q(\sqrt{P}/\sigma)$ denote their error probability expression. Then, it can be shown that instead of employing channel i and channel j with powers P_i and P_j and channel switching (time sharing) factors of λ_i and λ_j , respectively, it is always better to employ only one of these channels with power $(\lambda_i P_i + \lambda_j P_j)/(\lambda_i + \lambda_j)$ and a time sharing factor of $(\lambda_i + \lambda_j)$. This is because of the convexity of $g(P)$ for $P > 0$, which implies that $\lambda_i g(P_i) + \lambda_j g(P_j) > (\lambda_i + \lambda_j)g((\lambda_i P_i + \lambda_j P_j)/(\lambda_i + \lambda_j))$ for all $\lambda_i, \lambda_j \in (0, 1)$ and $P_i, P_j > 0$.

Proof: The first part is obvious since the use of another channel apart from channel K would violate the average cost constraint in (5.2) as $C_1 > C_2 > \dots > C_K = A_c$. Also, since $g_i(P)$ is a monotone decreasing function of P , the optimal strategy operates at the average power limit A_p in (5.2).

To prove the second part, consider a generic strategy that employs channel switching with power P_i and channel switching factor λ_i for channel i , which achieves an average probability of error given by $\sum_{i=1}^K \lambda_i g_i(P_i)$. Then, the following inequalities can be obtained:

$$\sum_{i=1}^K \lambda_i g_i(P_i) > \sum_{i=1}^K \lambda_i g_1(P_i) > g_1\left(\sum_{i=1}^K \lambda_i P_i\right) \quad (5.3)$$

The first inequality follows from the fact that $g_1(P) < g_i(P)$ for all $P \geq 0$ and $i \in \{2, \dots, K\}$ since channel 1 has the smallest noise variance (the largest cost). On the other hand, the second inequality follows from the strict convexity of g_1 for positive arguments. It is noted that the expression on the right-hand-side of (5.3) is the probability of error that is achieved by employing channel 1 exclusively with power $\sum_{i=1}^K \lambda_i P_i$. Therefore, it is concluded from (5.3) that employing channel 1 exclusively always achieves a smaller average probability of error than any strategy that employs channel switching. (Since $A_c \geq C_1$, it is possible to employ channel 1 exclusively.) In addition, since $g_1(P)$ is a monotone decreasing function of P , the optimal strategy operates at the average power limit A_p . ■

Remark 5.1 presents intuitive results for two simple cases, which can be summarized as follows: If the budget (average cost limit) allows the use of the worst (cheapest) channel only, then the only feasible approach is to employ that channel exclusively, which becomes the optimal solution of (5.2). On the other hand, if the budget allows the use of any channel with any switching factors, then the optimal solution is to employ the best channel all the time by using all the available power; that is, channel switching can only degrade the performance in this scenario. Since the solutions in these two special cases are obtained in a simple manner, we focus on the other scenarios for which the average cost limit satisfies $C_K < A_c < C_1$ in the remainder of this study.

Instead of trying to solve the problem in (5.2) directly for obtaining the optimal channel switching strategy, the properties of the optimal solution are investigated first in order to propose alternative approaches that yield the optimal channel switching strategy with reduced computational complexity. To that aim, the following proposition states that the optimal solution of (5.2) always satisfies the average power and average cost constraints with equality.

Proposition 5.1 *Assume that $C_K < A_c < C_1$ and let $\{\lambda_i^*, P_i^*\}_{i=1}^K$ denote the solution of the optimization problem in (5.2). Then, $\sum_{i=1}^K \lambda_i^* P_i^* = A_p$ and $\sum_{i=1}^K \lambda_i^* C_i = A_c$; that is, the optimal channel switching strategy utilizes the maximum average power and the maximum average cost.*

Proof: The claims in the proposition can be proved via contradiction. In order to prove the claim about the utilization of the maximum average power, suppose that $\{\lambda_l, P_l\}_{l=1}^K$ is an optimal solution of (5.2) with $\sum_{l=1}^K \lambda_l P_l < A_p$ and channel i is one of the employed channels; i.e., $\lambda_i > 0$. Next, define another solution as $\{\lambda_l, P'_l\}_{l=1}^K$, where $P'_l = P_l + \Delta P_l$ with $\Delta P_l \geq 0$ and $\sum_{l=1}^K \lambda_l P'_l = A_p$. In particular, let $\Delta P_l = 0, \forall l \neq i$ and $\Delta P_i = (A_p - \sum_{l=1}^K \lambda_l P_l) / \lambda_i$. Then, the following relation can be derived:

$$\sum_{l=1}^K \lambda_l g_l(P_l) = \lambda_i g_i(P_i) + \sum_{\substack{l=1 \\ l \neq i}}^K \lambda_l g_l(P_l) > \lambda_i g_i(P'_i) + \sum_{\substack{l=1 \\ l \neq i}}^K \lambda_l g_l(P'_l) = \sum_{l=1}^K \lambda_l g_l(P'_l) \quad (5.4)$$

where the inequality is obtained due to the facts that g_i is monotone decreasing, $P'_l = P_l, \forall l \neq i$, and $P'_i > P_i$. From (5.4), it is concluded that the solution $\{\lambda_l, P_l\}_{l=1}^K$, which operates at an average power below A_p , has a higher average probability of error than $\{\lambda_l, P'_l\}_{l=1}^K$, which utilizes an average power of A_p . This leads to a contradiction since $\{\lambda_l, P_l\}_{l=1}^K$ was assumed to be an optimal solution of (5.2). Therefore, a solution that operates at an average power below A_p cannot be optimal. In other words, an optimal solution must utilize all the available power; i.e., operate at the average power limit, A_p .

In order to prove the claim about the operation at the maximum average cost, first suppose that the optimal solution employs at least two different channels,

say channel i and channel j with channel switching factors λ_i and λ_j , respectively, where $i < j$ (hence, $C_i > C_j$), and it operates at an average cost of A'_c , which is strictly less than A_c ; that is, $A'_c < A_c$. For notational convenience, define $P_{ij} \triangleq \lambda_i P_i + \lambda_j P_j$ and $C_\lambda \triangleq \lambda_i C_i + \lambda_j C_j$. Then, consider an alternative solution which employs a similar strategy to the optimal solution except that it uses channel i with power P'_i and channel switching factor γ , and channel j with power P_j and channel switching factor $\lambda_i + \lambda_j - \gamma$, where $\lambda_i < \gamma < \lambda_i + \lambda_j$ with $\gamma P'_i + (\lambda_i + \lambda_j - \gamma)P_j = P_{ij}$ (the same average power as the optimal one) and $\gamma C_i + (\lambda_i + \lambda_j - \gamma)C_j = C_\lambda + A_c - A'_c$ (larger average cost than the optimal one). By equating the average power terms (that is, $\gamma P'_i + (\lambda_i + \lambda_j - \gamma)P_j = \lambda_i P_i + \lambda_j P_j$), P'_i can be obtained as $P'_i = \lambda_i P_i / \gamma + (1 - \lambda_i / \gamma)P_j$. Then, the following relations can be obtained:

$$\gamma g_i(P'_i) + (\lambda_i + \lambda_j - \gamma)g_j(P_j) = \gamma g_i(\lambda_i P_i / \gamma + (1 - \lambda_i / \gamma)P_j) + (\lambda_i + \lambda_j - \gamma)g_j(P_j) \quad (5.5)$$

$$< \lambda_i g_i(P_i) + (\gamma - \lambda_i) g_i(P_j) + (\lambda_i + \lambda_j - \gamma)g_j(P_j) \quad (5.6)$$

$$\leq \lambda_i g_i(P_i) + \lambda_j g_j(P_j) \quad (5.7)$$

where the first inequality is obtained from the strict convexity of g_i and the second inequality follows from the fact that $g_i(P_j) \leq g_j(P_j)$ since channel i has a smaller noise variance (higher cost) than channel j . The inequality in (5.5)-(5.7), namely, $\lambda_i g_i(P_i) + \lambda_j g_j(P_j) > \gamma g_i(P'_i) + (\lambda_i + \lambda_j - \gamma)g_j(P_j)$, leads to a contradiction since the optimal solution results in a higher average probability of error than the alternative solution, which uses the same average power but operates at the maximum average cost. Therefore, it is concluded that a solution that employs at least two channels and operates below the average cost limit A_c cannot be optimal. In order to complete the proof, suppose that an optimal solution employs a single channel (say, channel i) and operates below A_c ; that is, channel i is employed exclusively with power P_i and its cost C_i is strictly smaller than A_c ; that is, $C_i < A_c < C_1$. Next, consider an alternative solution that employs channel i and channel 1 with channel switching factors λ'_i and $1 - \lambda'_i$, respectively, and with the same power P_i , where $\lambda'_i \in (0, 1)$. Then,

$$\lambda'_i g_i(P_i) + (1 - \lambda'_i)g_1(P_i) < \lambda'_i g_i(P_i) + (1 - \lambda'_i)g_i(P_i) = g_i(P_i) \quad (5.8)$$

where the inequality follows from the fact that $g_1(P) < g_i(P)$, $\forall P$, by definition (note that $C_1 > C_i$). The inequality in (5.8) leads to a contradiction since the alternative solution achieves a smaller average probability of error than the optimal one by using the same average power. Therefore, a solution that employs a single channel and operates below the maximum average cost cannot be optimal. Overall, since any channel switching strategy either uses a single channel or switches among multiple channels, the previous arguments prove that an optimal channel switching strategy must always operate at the maximum average cost. ■

Proposition 5.1 states that the optimal channel switching strategy utilizes all the average power and average cost. Therefore, the optimization problem in (5.2) can be solved by considering equality constraints (instead of inequality constraints) for the average power and average cost, which leads to an important reduction in computational complexity. Another implication of Proposition 5.1 is presented in the following corollary.

Corollary 5.1 *Assume that $C_K < A_c < C_1$. If $C_i \neq A_c$, $\forall i \in \{1, \dots, K\}$, then the optimal solution of (5.2) involves channel switching among multiple channels; that is, transmission over a single channel is not optimal.*

Proof: Let $\{\lambda_i^*, P_i^*\}_{i=1}^K$ denote the solution of the optimization problem in (5.2). Proposition 1 states that $\sum_{i=1}^K \lambda_i^* C_i = A_c$ must hold. If $C_i \neq A_c$, $\forall i \in \{1, \dots, K\}$, then the condition of $\sum_{i=1}^K \lambda_i^* C_i = A_c$ cannot be satisfied unless at least two of λ_i^* 's are nonzero, which implies switching among multiple channels. ■

It should be noted that the converse of Corollary 5.1 is not necessarily true. That is, when $C_i = A_c$ for some $i \in \{1, \dots, K\}$, the structure of the optimal solution depends on the cost values and the average power constraint. In other words, either transmission over a single channel or channel switching can be optimal depending on the system parameters.

Although the optimization problem in (5.2) is formulated to search over strategies that involve channel switching among up to K channels, the following proposition states that the optimal solution of (5.2) can be expressed as channel switching among $\min\{K, 3\}$ or fewer channels.

Proposition 5.2 *Assume that the power levels satisfy $P_i \in [0, P_{\max}]$ for some finite $P_{\max}$. Then, the optimal channel switching strategy is to switch among at most $\min\{K, 3\}$ channels.*

Proof: If $K \leq 3$, the statement in the proposition is satisfied trivially. Assume that $K > 3$ and define the following sets:

$$\mathcal{W} = \left\{ \left(\sum_{i=1}^K \lambda_i P_i, \sum_{i=1}^K \lambda_i g_i(P_i), \sum_{i=1}^K \lambda_i C_i \right), \forall \lambda_i \geq 0, \sum_{i=1}^K \lambda_i = 1, \forall P_i \in [0, P_{\max}] \right\} \quad (5.9)$$

$$\mathcal{U} = \{(P, g_i(P), C_i), \forall i \in \{1, \dots, K\}, \forall P \in [0, P_{\max}]\} \quad (5.10)$$

By definition, set $\mathcal{W}$ contains the optimal solution of (5.2) since it consists of all possible average power, average probability of error and average cost triples. Also, it is observed from (5.9) and (5.10) that $\mathcal{W}$ is a subset of the convex hull of set $\mathcal{U}$; i.e., $\mathcal{W} \subset \text{hull}(\mathcal{U})$. This is because of the fact that all the triples in $\mathcal{W}$ can be obtained as the convex combinations of K elements in $\mathcal{U}$ whereas some convex combinations of the elements of $\mathcal{U}$, which involve the use of at least one channel multiple times,² are not included in $\mathcal{W}$. Since $\mathcal{W}$ is contained in the convex hull of set $\mathcal{U}$, any element of $\mathcal{W}$ can be expressed as a convex combination of $\dim(\mathcal{U}) + 1 = 4$ elements in $\mathcal{U}$ as a result of Carathéodory's theorem [49], where $\dim(\mathcal{U})$ denotes the dimension of the space in which $\mathcal{U}$ resides. (Note that $\mathcal{U} \subset \mathbb{R}^3$.) In addition, since the aim is to achieve the minimum average probability of error (see (5.2)), the optimal solution corresponds to a point on the boundary of $\text{hull}(\mathcal{U})$, which can be achieved by a convex combination of $\dim(\mathcal{U}) = 3$ elements in $\mathcal{U}$ by Carathéodory's theorem [49]. Finally, it is noted that all such convex combinations are guaranteed to be elements of set $\mathcal{W}$ due

²For example, the convex combination of $(P_1, g_1(P_1), C_1)$ and $(P_2, g_1(P_2), C_1)$ is not included in $\mathcal{W}$, which involves the use of channel 1 twice.

to the following reason: The difference of $\text{hull}(\mathcal{U})$ from $\mathcal{W}$ (that is, $\text{hull}(\mathcal{U}) \setminus \mathcal{W}$) consists of the points corresponding to strategies that use at least one channel multiple times. However, such strategies cannot be optimal solutions since the use of a channel multiple times always increases the average probability of error compared to the use of that channel once with the same average power (which can be proved by an argument similar to that in Footnote 1). Therefore, the optimal solution cannot be in $\text{hull}(\mathcal{U}) \setminus \mathcal{W}$; i.e., it is always in $\mathcal{W}$, which implies that the optimal solution can be expressed as a convex combination of up to 3 elements in $\mathcal{U}$ that correspond to different channel indices (see index i in (5.10)). Hence, channel switching among up to 3 different channels is optimal. ■

Based on Proposition 5.1 and Proposition 5.2, the optimal channel switching corresponds to one of the following three strategies:

Strategy 1 – Transmission over a Single Channel: In this case, one of the channels is employed exclusively. Based on Corollary 5.1, this strategy cannot be an optimal solution of (5.2) unless there exists a channel with cost A_c . If there exists such a channel and i^* denotes the index of that channel (that is, $C_{i^*} = A_c$), then the minimum average probability of error achieved by this strategy is given by $g_{i^*}(A_p)$, which corresponds to transmission over channel i^* exclusively by utilizing the maximum available power (cf. (5.2)). Note that this strategy may or may not be the optimal solution of the problem in (5.2) depending on the system parameters.

Strategy 2 – Channel Switching between Two Channels: In this strategy, channel switching is performed between two different channels. Let channel i and channel j denote those channels. Then, based on Proposition 5.1, the problem in (5.2) can be formulated under Strategy 2 as

$$\min_{\lambda, P_i, P_j} \lambda g_i(P_i) + (1 - \lambda) g_j(P_j)$$

$$\text{subject to } \lambda P_i + (1 - \lambda) P_j = A_p, \quad \lambda C_i + (1 - \lambda) C_j = A_c, \quad \lambda \in [0, 1]. \quad (5.11)$$

It is observed from the average cost constraint in (5.11) that, for the optimal channel switching between two channels, one of the channels should have a cost higher than A_c and the other channel should have a cost lower than A_c . Therefore,

in order to obtain the optimal solution for Strategy 2, the problem in (5.11) should be solved for $K_s K_g$ channel pairs, where K_s (K_g) is the number of channels the costs of which are lower (higher) than A_c . In other words, the problem in (5.11) should be solved for all channel pairs $(i, j) \in \mathcal{S}_2$, where $\mathcal{S}_2 = \{(i, j) : C_i > A_c > C_j \text{ and } i, j \in \{1, \dots, K\}\}$.

Based on the argument in the previous paragraph, assume, without loss of generality, that $C_i > A_c > C_j$ for the problem in (5.11). Then, the optimal value of λ can be obtained from the average cost constraint as $\lambda^* = (A_c - C_j)/(C_i - C_j)$. Also, due to the average power constraint, the powers are related as $P_j = (A_p - \lambda^* P_i)/(1 - \lambda^*)$. In addition, by considering the definition of function g_i , that is, $g_i(P_i) = Q(s_i/\sigma_i)$, where $P_i = s_i^2$, the optimization problem in (5.11) can be expressed as follows:

$$\min_{s_i \in (0, \sqrt{A_p/\lambda^*})} \lambda^* Q\left(\frac{s_i}{\sigma_i}\right) + (1 - \lambda^*) Q\left(\frac{1}{\sigma_j} \sqrt{\frac{A_p - \lambda^* s_i^2}{1 - \lambda^*}}\right) \quad (5.12)$$

where the constraint for s_i is obtained from the relation $\lambda^* s_i^2 + (1 - \lambda^*) s_j^2 = A_p$. From (5.12), it is observed that the optimal solution for Strategy 2 requires a search over a one-dimensional space only (for each possible channel pair). In addition, it can be shown that the objective function in (5.12) is strictly convex for $s_i \in (0, \sqrt{A_p/\lambda^*})$.³ Therefore, convex optimization algorithms can be employed to obtain the result in polynomial time [44]. In fact, as stated in the following proposition, the structure of the objective function also leads to additional properties, which result in further simplifications.

Proposition 5.3 *Suppose that $C_i > A_c > C_j$, and define $A_{ij} \triangleq \frac{\sigma_i^2 \sigma_j^2}{\sigma_j^2 - \sigma_i^2} \log\left(\frac{\sigma_j^2}{\sigma_i^2}\right)$, where $\log$ denotes the natural logarithm. Then, the optimal solution of (5.11), denoted by $\{\lambda^*, P_i^*, P_j^*\}$, satisfies the following relations depending on the average power limit:*

(i) *If $A_p = A_{ij}$, then $P_i^* = P_j^* = A_{ij}$.*

³The first-order derivative of the objective function is presented in (5.14), which is a monotone increasing function of s_i for $s_i \in (0, \sqrt{A_p/\lambda^*})$.

(ii) If $A_p > A_{ij}$, then $P_j^* > A_p > P_i^* > A_{ij}$.

(iii) If $A_p < A_{ij}$, then $A_{ij} > P_i^* > A_p > P_j^*$.

In addition, the ratio between the optimal power levels cannot exceed σ_j^2/σ_i^2 ; that is,

$$\max \left\{ \frac{P_j^*}{P_i^*}, \frac{P_i^*}{P_j^*} \right\} < \frac{\sigma_j^2}{\sigma_i^2}. \quad (5.13)$$

Proof: Since the problem in (5.11) can be solved based on (5.12), consider the first-order derivative of the objective function in (5.12) with respect to s_i , which is expressed as

$$-\frac{\lambda^*}{\sqrt{2\pi} \sigma_i} e^{-\frac{s_i^2}{2\sigma_i^2}} + \frac{\lambda^* \sqrt{1-\lambda^*} s_i}{\sqrt{2\pi} \sigma_j \sqrt{A_p - \lambda^* s_i^2}} e^{-\frac{(A_p - \lambda^* s_i^2)}{2(1-\lambda^*)\sigma_j^2}}. \quad (5.14)$$

It is observed that the first-order derivative in (5.14) is a monotone increasing function of s_i for $s_i \in (0, \sqrt{A_p/\lambda^*})$, which starts from $-\lambda^*/(\sqrt{2\pi} \sigma_i)$ at $s_i = 0$ and increases monotonically towards infinity as s_i goes to $\sqrt{A_p/\lambda^*}$. Therefore, there is a unique minimizer s_i^* for the optimization problem in (5.12), which corresponds to the point at which the first-order derivative is zero. Equating the first-order derivative in (5.14) to zero yields the following necessary and sufficient condition for the optimal solution of (5.12):

$$e^{\frac{s_i^2}{\sigma_i^2} - \frac{(A_p - \lambda^* s_i^2)}{(1-\lambda^*)\sigma_j^2}} = \frac{\sigma_j^2 (A_p - \lambda^* s_i^2)}{(1-\lambda^*)\sigma_i^2 s_i^2}. \quad (5.15)$$

Since $\lambda^* s_i^2 + (1-\lambda^*) s_j^2 = A_p$, the condition in (5.15) can also be expressed as

$$e^{\frac{s_i^2}{\sigma_i^2} - \frac{s_j^2}{\sigma_j^2}} = \frac{\sigma_j^2 s_j^2}{\sigma_i^2 s_i^2}. \quad (5.16)$$

If $A_p = A_{ij}$, it can be shown, by using the definition of A_{ij} in the proposition, that $s_i^2 = s_j^2 = A_{ij}$ satisfies the condition in (5.16). Since the solution of (5.16) is unique, the optimal solution of (5.12) is obtained as $s_i^* = s_j^* = \sqrt{A_{ij}}$. Therefore, the optimal solution of the problem in (5.11), which yields the same solution as (5.12), is given by $P_i^* = P_j^* = A_{ij}$, as stated in the first part of Proposition 5.3.

In order to prove the second part of the proposition, it is first observed that the first-order derivative in (5.14) is a monotone decreasing function of A_p and a monotone increasing function of s_i . Therefore, the value of s_i at which the first-order derivative becomes zero gets larger as A_p increases. Since the first-order derivative becomes zero at $s_i = \sqrt{A_{ij}}$ when $A_p = A_{ij}$ (as proved in the first part), the first-order derivative becomes zero at a value larger than $\sqrt{A_{ij}}$ when $A_p > A_{ij}$. Hence, the optimal solution of (5.12) satisfies $s_i^* > \sqrt{A_{ij}}$ for $A_p > A_{ij}$. In addition, it is concluded from (5.16) that as s_i^* increases, the optimal value of s_j should also increase in order for the optimality condition in (5.16) to be satisfied. In other words, $s_i^* > \sqrt{A_{ij}}$ also implies $s_j^* > \sqrt{A_{ij}}$ based on the relation in (5.16). Next, the ordering between s_i^* and s_j^* should be determined. To that aim, the optimal signal values are expressed as $(s_i^*)^2 = \alpha A_{ij}$ and $(s_j^*)^2 = \beta A_{ij}$, where α and β are some positive numbers that are larger than one. Then, the optimality condition in (5.16) becomes $e^{A_{ij}(\alpha/\sigma_i^2 - \beta/\sigma_j^2)} = \beta\sigma_j^2/(\alpha\sigma_i^2)$. From the definition of A_{ij} in the proposition, σ_j^2/σ_i^2 can be expressed as $\sigma_j^2/\sigma_i^2 = e^{A_{ij}(1/\sigma_i^2 - 1/\sigma_j^2)}$. Then, the optimality condition is stated as

$$\frac{\alpha}{\beta} = e^{A_{ij}\left(\frac{\beta-1}{\sigma_j^2} - \frac{\alpha-1}{\sigma_i^2}\right)}. \quad (5.17)$$

If it is assumed that $\alpha > \beta$, then (5.17) implies that $\frac{\beta-1}{\alpha-1} > \frac{\sigma_j^2}{\sigma_i^2}$. However, since $\sigma_i^2 < \sigma_j^2$ (as $C_i > C_j$), this inequality leads to a contradiction. Therefore, α cannot be larger than β . On the other hand, if it is assumed that $\alpha < \beta$, then (5.17) becomes $\frac{\alpha-1}{\beta-1} > \frac{\sigma_i^2}{\sigma_j^2}$, which is not a contradiction. Therefore, it is obtained that $\alpha < \beta$, that is, $s_i^* < s_j^*$, when $A_p > A_{ij}$. Furthermore, due to the average power constraint, $\lambda^*(s_i^*)^2 + (1 - \lambda^*)(s_j^*)^2 = A_p$, it is concluded that $s_j^* > \sqrt{A_p} > s_i^*$. Combining this result with the first result in this paragraph, it is obtained that when $A_p > A_{ij}$, the optimal signal values satisfy $s_j^* > \sqrt{A_p} > s_i^* > \sqrt{A_{ij}}$. Hence, the second part of Proposition 5.3 is proved. The third part of the proposition can be proved in a similar manner to the proof of the second part, and it can be shown that $\sqrt{A_{ij}} > s_i^* > \sqrt{A_p} > s_j^*$ based on (5.14) and (5.16).

The final statement in the proposition can be proved as follows: For $A_p > A_{ij}$, it is obtained in the previous paragraph that $\frac{\alpha-1}{\beta-1} > \frac{\sigma_i^2}{\sigma_j^2}$, where $\beta > \alpha > 1$ with

$(s_i^*)^2 = \alpha A_{ij}$ and $(s_j^*)^2 = \beta A_{ij}$. The inequality can be manipulated as follows:

$$\frac{\sigma_j^2}{\sigma_i^2} > \frac{\beta - 1}{\alpha - 1} > \frac{\beta}{\alpha} = \frac{(s_j^*)^2}{(s_i^*)^2} = \frac{P_j^*}{P_i^*} \quad (5.18)$$

where the second inequality is obtained from the relation $\beta > \alpha > 1$. For $A_p < A_{ij}$, the second part of the proposition states that $(s_i^*)^2 > (s_j^*)^2$. Since $\sigma_i^2 < \sigma_j^2$ by definition (as $C_i > C_j$), it is obtained that $(s_i^*)^2/\sigma_i^2 > (s_j^*)^2/\sigma_j^2$. Therefore, the relation in (5.16) yields

$$\frac{\sigma_j^2 (s_j^*)^2}{\sigma_i^2 (s_i^*)^2} = e^{\frac{(s_i^*)^2}{\sigma_i^2} - \frac{(s_j^*)^2}{\sigma_j^2}} > 1, \quad (5.19)$$

which results in $P_i^*/P_j^* < \sigma_j^2/\sigma_i^2$. Finally, for $A_p = A_{ij}$, $P_i^*/P_j^* = 1$ as stated in the first part of the proposition. Overall, the ratio between the optimal power levels is upper bounded by σ_j^2/σ_i^2 for any value of A_p , as stated in the proposition. ■

Proposition 5.3 is helpful in reducing the search space for the optimization problem in (5.12). Specifically, for each channel pair (i, j) with $C_i > C_j$, the value of A_{ij} is calculated first, as defined in the proposition. Then, the optimal signal values are obtained as follows:

- If $A_p = A_{ij}$, the optimal solution is given by $s_i^* = s_j^* = \sqrt{A_{ij}}$.
- If $A_p > A_{ij}$, the optimization problem in (5.12) is solved for $s_i \in (\max\{\sqrt{A_{ij}}, \sigma_i \sqrt{A_p}/\sigma_j\}, \sqrt{A_p})$, which is obtained from (5.13) and the relation in the second part of the proposition.
- If $A_p < A_{ij}$, the problem in (5.12) which is obtained from (5.13) and the relation in the third part of the proposition is solved for $s_i \in (\sqrt{A_p}, \min\{\sqrt{A_{ij}}, \sqrt{A_p/\lambda^*}, \sigma_j \sqrt{A_p}/\sigma_i\})$.

Once the optimal value of s_i , denoted by s_i^* , is obtained, the optimal value of s_j is calculated as $s_j^* = \sqrt{(A_p - \lambda^*(s_i^*)^2)/(1 - \lambda^*)}$, where $\lambda^* = (A_c - C_j)/(C_i - C_j)$.

Strategy 3 – Channel Switching among Three Channels: In this strategy, channel switching is performed among three different channels. Let channel

i , channel j , and channel k denote those channels. Then, based on Proposition 1, the problem in (5.2) can be formulated under Strategy 3 as

$$\begin{aligned}
& \min_{\lambda_i, \lambda_j, \lambda_k, P_i, P_j, P_k} \lambda_i g_i(P_i) + \lambda_j g_j(P_j) + \lambda_k g_k(P_k) \\
& \text{subject to } \lambda_i P_i + \lambda_j P_j + \lambda_k P_k = A_p, \\
& \lambda_i C_i + \lambda_j C_j + \lambda_k C_k = A_c, \\
& \lambda_i + \lambda_j + \lambda_k = 1, \lambda_i, \lambda_j, \lambda_k \geq 0.
\end{aligned} \tag{5.20}$$

Due to the strict average cost constraint, it is required that at least one of the channels must have a cost lower than A_c and at least one of the channels must have a cost higher than A_c . Therefore, in order to obtain the optimal solution for Strategy 3, the problem in (5.20) should be solved for $K_s K_g (K - 2)$ channel triples, where K_s (K_g) is the number of channels the costs of which are lower (higher) than A_c , and K is the total number of channels. In other words, the problem in (5.20) should be solved for all channel triples $(i, j, k) \in \mathcal{S}_3$, where $\mathcal{S}_3 = \{(i, j, k) : C_i > A_c > C_j \text{ and } i, j, k \in \{1, \dots, K\}\}$. In addition, it is observed that the solution of (5.20) can be obtained via optimization over a *three-dimensional* space instead of six by utilizing the three equality constraints.

It is noted from Proposition 5.2 that Strategy 3 is guaranteed to provide the optimal solution of the channel switching problem in (5.2). In addition, it covers Strategy 2 and Strategy 1 as special cases, which may be suboptimal in general. Therefore, in order to obtain the optimal channel switching solution, it can be necessary in general to solve the optimization problem in (5.20), which is computationally more complex than obtaining the optimal solutions under Strategy 1 and Strategy 2. However, in some cases (see Proposition 5.5), it is guaranteed that Strategy 1 or Strategy 2 can provide the optimal solution of the channel switching problem in (5.2); that is, it is not necessary to solve the optimization problem in (5.20) for obtaining the optimal channel switching solution. Therefore, whenever the conditions under which Strategy 1 or Strategy 2 is optimal are satisfied, the optimal channel switching solution can be obtained in a low-complexity manner as follows: If there exist no channels with cost A_c , then Strategy 2 provides the optimal solution. If there exists a channel with cost A_c , then the optimal solution is either to employ that channel exclusively with the maximum power (Strategy

1), or to switch between two channels as specified by the solution of (5.12) (Strategy 2). In that case, the strategy that achieves the smaller average probability of error becomes the optimal solution of (5.2).

5.3 Optimal Channel Switching For Logarithmic Cost Functions

In this section, specific theoretical results are obtained by considering a suitable cost function for the channels. Since each channel can be regarded as a measurement device, a cost function similar to that proposed in [23] can be adopted for relating the noise power of each channel to a cost value as follows:

$$C_i = \log \left(1 + \frac{b}{\sigma_i^2} \right), \quad i \in \{1 \dots K\}, \quad (5.21)$$

where $b > 0$ is a given system parameter (a constant). It is noted that the function in (5.21) has the desirable property that it assigns higher (lower) cost values to less (more) noisy channels; that is, $\sigma_i^2 < \sigma_j^2$ implies $C_i > C_j$. In addition, $\lim_{\sigma_i \rightarrow \infty} C_i = 0$ and $\lim_{\sigma_i \rightarrow 0} C_i = \infty$. As in the previous section, it is assumed without loss of generality that the noise variances of the channels satisfy $\sigma_1^2 < \sigma_2^2 < \dots < \sigma_K^2$, which implies that the cost values are ordered as $C_1 > C_2 > \dots > C_K$.

Based on the cost function in (5.21), the following result is obtained first.

Lemma 5.1 *Consider infinitely many channels and assume that the channels take a continuum of cost values in the interval $[C_{\min}, C_{\max}]$ based on the cost function in (5.21), where $0 < C_{\min} < C_{\max} < \infty$. Let $g(P, C)$ denote the error probability when transmission is performed by utilizing a power level of P over a channel with cost C . Then, $g(P, C)$ is a strictly convex function over set $\mathcal{S}_c$, which is a convex set defined as $\mathcal{S}_c \triangleq \{(P, C) : P > b/(e^C + 1), C \in (C_{\min}, C_{\max})\}$.*

Proof: As discussed in Section 5.1, the error probability for a transmission power of P is expressed as $Q(\sqrt{P}/\sigma)$, where σ is the standard deviation

of the channel noise. Based on the cost function in (5.21), σ is expressed as $\sigma = \sqrt{b/(e^C - 1)}$, which leads to the following expression for the error probability:

$$g(P, C) = Q(h(P)f(C)) , \quad (5.22)$$

where $h(P) \triangleq \sqrt{P}$ and $f(C) \triangleq \sqrt{(e^C - 1)/b}$. In order to investigate the convexity of (5.22), the derivatives of h and f are calculated first, which are expressed as

$$h' = \frac{1}{2h} , \quad f' = \frac{bf^2 + 1}{2bf} , \quad h'' = -\frac{1}{4h^3} , \quad f'' = \frac{(bf^2 - 1)(bf^2 + 1)}{4b^2 f^3} . \quad (5.23)$$

Then, the first-order partial derivatives of $g(P, C)$ with respect to P and C are given by

$$\frac{\partial g(P, C)}{\partial P} = h' f Q'(hf) \quad \text{and} \quad \frac{\partial g(P, C)}{\partial C} = h f' Q'(hf) , \quad (5.24)$$

where $Q'(x)$ denotes the first derivative of the Q -function. (From (5.24), it is observed that the error probability is a monotone decreasing function of power and cost as expected since Q -function is monotone decreasing.) Next, the second-order partial derivatives are calculated as

$$\frac{\partial^2 g(P, C)}{\partial P^2} = (h' f)^2 Q''(hf) + h'' f Q'(hf) = (f h'' - f^3 h'(h')^2) Q'(hf) \quad (5.25)$$

$$\frac{\partial^2 g(P, C)}{\partial C^2} = (h f')^2 Q''(hf) + h f'' Q'(hf) = (h f'' - h^3 f'(f')^2) Q'(hf) \quad (5.26)$$

$$\frac{\partial^2 g(P, C)}{\partial P \partial C} = h h' f f' Q''(hf) + h' f' Q'(hf) = (1 - h^2 f^2) h' f' Q'(hf) \quad (5.27)$$

where the relation $Q''(x) = -x Q'(x)$ is employed to obtain the final expressions. From (5.25)-(5.27), the 2×2 Hessian matrix can be formed for $g(P, C)$, and the convexity of $g(P, C)$ can be investigated based on the positive definiteness of the Hessian matrix, which requires the leading principal minors to be positive [47]. It noted from (5.25) that the second-order derivative with respect to P is always positive. Therefore, the only condition for positive definiteness becomes the determinant of the Hessian matrix to be positive, which leads, after some manipulation, to the following inequality:

$$P > \frac{b}{b f^2 + 2} = \frac{b}{e^C + 1} , \quad (5.28)$$

where the final expression is obtained based on the definition of f ; i.e., $f(C) = \sqrt{(e^C - 1)/b}$. Therefore, the convexity of $g(P, C)$ requires that power P should be larger than $b/(e^C + 1)$ as stated in the lemma.

Finally, it is shown that $\mathcal{S}_c$, as defined in the lemma, is a convex set. Let (P_i, C_i) and (P_j, C_j) denote any two elements from set $\mathcal{S}_c$. Then, their convex combination is given by $(\lambda P_i + (1 - \lambda)P_j, \lambda C_i + (1 - \lambda)C_j)$, where $\lambda \in [0, 1]$. Since both C_i and C_j are in $(C_{\min}, C_{\max})$, their convex combination resides in the same interval as well. In addition, the convex combination of the powers satisfies the condition for set $\mathcal{S}_c$ due to the following inequalities:

$$\lambda P_i + (1 - \lambda)P_j > \lambda \frac{b}{e^{C_i} + 1} + (1 - \lambda) \frac{b}{e^{C_j} + 1} > \frac{b}{e^{\lambda C_i + (1 - \lambda)C_j} + 1} \quad (5.29)$$

where the second inequality follows from the strict convexity of $b/(e^C + 1)$. Therefore, $\mathcal{S}_c$ is a convex set, and $g(P, C)$ is a strictly convex function over set $\mathcal{S}_c$. ■

Lemma 5.1 describes the convexity properties of the error probability, which is considered as a function of power and cost. Based on Lemma 5.1, the solutions of the optimal channel switching problem can be specified in certain scenarios. To that aim, the following proposition presents the optimal solution when channel switching is performed between two channels (i.e., Strategy 2).

Proposition 5.4 *Suppose there exist K channels and each channel has a cost value obtained from the cost function in (5.21). If the average power limit satisfies $A_p \geq \frac{b\sigma_K^4}{\sigma_1^2(2\sigma_K^2 + b)}$, then the optimal solution for Strategy 2 employs channel i and channel j , where*

$$i = \arg \min_{k \in \{1, \dots, K\}} C_k \text{ subject to } C_k > A_c, \quad (5.30)$$

$$j = \arg \max_{k \in \{1, \dots, K\}} C_k \text{ subject to } C_k < A_c. \quad (5.31)$$

Proof: From Proposition 5.1, it is known that the optimal channel switching solution utilizes the maximum average cost. Therefore, for Strategy 2, the optimal pair of channels, say (k, l) , must satisfy $k \leq i$ and $l \geq j$, where i and j are as defined in (5.30) and (5.31), respectively. (Note that the cost values are ordered

as $C_1 > C_2 > \dots > C_K$.) For simplicity of notation, define $\mathbf{z}_k \triangleq (P_k, C_k)$, for $k = 1, \dots, K$, and $\mathbf{a} \triangleq (A_p, A_c)$. In order to prove that the optimal channel pair for Strategy 2 is (i, j) , first consider channel pair (k, l) , where $k = i$ and $l > j$. The optimal solution for channel pair (i, l) must utilize the maximum average power and cost due to Proposition 5.1. In addition, consider an alternative solution that employs channel pair (i, j) and operates at the average power and cost limits. Then, the following inequalities are obtained:

$$\lambda \mathbf{z}_i + (1 - \lambda) \mathbf{z}_j = \mathbf{a} \quad \text{and} \quad \gamma \mathbf{z}_i + (1 - \gamma) \mathbf{z}_l = \mathbf{a}, \quad (5.32)$$

where $\lambda = (A_c - C_j)/(C_i - C_j)$ and $\gamma = (A_c - C_l)/(C_i - C_l)$, which are obtained from the average cost constraint. Since $C_i > A_c > C_j > C_l$, it can be shown that $\gamma > \lambda$. Therefore, $\mathbf{z}_j$ can be expressed as

$$\mathbf{z}_j = \frac{\gamma - \lambda}{1 - \lambda} \mathbf{z}_i + \frac{1 - \gamma}{1 - \lambda} \mathbf{z}_l. \quad (5.33)$$

Then, it is shown in the following that channel pair (i, l) cannot be optimal since it results in a higher average probability of error than channel (i, j) :

$$\lambda g(\mathbf{z}_i) + (1 - \lambda)g(\mathbf{z}_j) = \lambda g(\mathbf{z}_i) + (1 - \lambda)g\left(\frac{\gamma - \lambda}{1 - \lambda} \mathbf{z}_i + \frac{1 - \gamma}{1 - \lambda} \mathbf{z}_l\right) \quad (5.34)$$

$$< \lambda g(\mathbf{z}_i) + (1 - \lambda)\left(\frac{\gamma - \lambda}{1 - \lambda} g(\mathbf{z}_i) + \frac{1 - \gamma}{1 - \lambda} g(\mathbf{z}_l)\right) \quad (5.35)$$

$$= \gamma g(\mathbf{z}_i) + (1 - \gamma) g(\mathbf{z}_l) \quad (5.36)$$

where $g(\mathbf{z}_i) = g(P_i, C_i)$ denotes the average probability of error as a function of power and cost, as defined in (5.22). In obtaining the equality in (5.34), the expression in (5.33) is employed, and the inequality in (5.35) follows from the strict convexity of g , which is guaranteed under the assumption in the proposition, namely, $A_p \geq \frac{b\sigma_K^4}{\sigma_1^2(2\sigma_K^2 + b)}$. In order to verify the convexity of g in this scenario, Lemma 1 is considered first, which states that the power levels should satisfy $P > b/(e^C + 1)$ for strict convexity. Since the cost values are ordered as $C_1 > C_2 > \dots > C_K$ (equivalently, $\sigma_1^2 < \sigma_2^2 < \dots < \sigma_K^2$), $P > b/(e^{C_K} + 1)$ is required to guarantee that function g is strictly convex for all channels. From Proposition 3, it is concluded that the optimal power levels under Strategy 2 always satisfy $\min\{P_k, P_l\} > \sigma_k^2 A_p / \sigma_l^2$ for channel pair (k, l) with $C_k > A_c > C_l$.⁴ Therefore, if

⁴This result is obtained by combining the inequality in (5.13) with the three possible scenarios in Proposition 3.

$\sigma_1^2 A_p / \sigma_K^2 > b / (e^{C_K} + 1)$ holds, then it is guaranteed that the optimal power levels for any channel pair for Strategy 2 satisfy the convexity condition in Lemma 5.1. Mathematically stated,

$$\min\{P_k, P_l\} > \frac{\sigma_k^2 A_p}{\sigma_l^2} > \frac{\sigma_1^2 A_p}{\sigma_K^2} > \frac{b}{e^{C_K} + 1} = \frac{b}{2 + b/\sigma_K^2} \quad (5.37)$$

for all $k \leq i$ and $l \geq j$, where first inequality is obtained from Proposition 5.3, the second one follows from the relation $\sigma_1^2 < \sigma_2^2 < \dots < \sigma_K^2$, the third one is imposed in order to guarantee that the power levels satisfy the convexity condition in Lemma 5.1, and the equality is obtained from (5.28). From (5.37), it is deduced that the condition $A_p > b \sigma_K^2 / (2 \sigma_1^2 + b \sigma_1^2 / \sigma_K^2)$ guarantees the strict convexity of g .

Similar arguments to those in (5.32)-(5.36) can be used to prove that channel pair (k, j) with $k < i$ results in a larger average probability of error than channel pair (i, j) . Then, it can be concluded that channel pair (k, l) cannot be optimal if $k < i$ and/or $l > j$. Hence, the optimal channel pair for Strategy 2 is shown to be the channel pair (i, j) as defined in the proposition when the average power limit A_p is larger than the specified value. ■

Proposition 5.4 states that if the average power limit is larger than a certain value, then the optimal solution for Strategy 2 is to switch between the two channels, one of which has the lowest cost among the channels with costs higher than A_c , and the other has the highest cost among the channels with costs lower than A_c . In other words, among all the channel pairs, where each pair has one channel with a cost higher than A_c and another channel with a cost lower than A_c , the one that has the *minimum cost difference* is selected in order to achieve the minimum average probability of error, which is mainly due to the convexity of the error probability, as specified in Lemma 1. Thanks to Proposition 5.4, it is not necessary to search over all feasible channel pairs to obtain the optimal solution for Strategy 2 when the average power limit is larger than the specified value.

Remark 5.2 *Under the condition in Proposition 5.4, if there exists a channel with cost A_c , then it outperforms the channel pair (i, j) specified in (5.30) and*

(5.31); that is, Strategy 1 outperforms Strategy 2 in that scenario. This is due to the strict convexity of g , which results in $\lambda g(P_i, C_i) + (1-\lambda) g(P_j, C_j) > g(A_p, A_c)$. In other words, if $A_p \geq \frac{b\sigma_K^4}{\sigma_1^2(2\sigma_K^2+b)}$, transmission over a single channel with cost A_c at the maximum power level A_p achieves a smaller average probability of error than performing optimal channel switching between two channels.

Based on Lemma 5.1, it is also possible to describe scenarios in which Strategy 1 or Strategy 2 is the optimal solution of the channel switching problem; that is, switching among more than two channels is not needed. The following proposition presents such a scenario:

Proposition 5.5 *Consider the optimal channel switching problem in (5.2) with the cost values as defined in (5.21). The optimal channel switching strategy involves at most two channels if the average power limit satisfies $A_p \geq \frac{2b\sigma_K^4}{\sigma_1^2(2\sigma_K^2+b)}$.*

Proof: The statement in the proposition can be proved via contradiction. Suppose that the optimal solution is to switch among three different channels, and let the channel indices, channel switching factors, and power levels for that optimal solution be denoted by (i, j, k) , $(\lambda_i, \lambda_j, \lambda_k)$ and (P_i, P_j, P_k) , respectively, where $C_i > C_j > C_k$ without loss of generality. Since the optimal solution must utilize the maximum average cost A_c (see Proposition 5.1), either $C_i > A_c > C_j > C_k$ or $C_i > C_j > A_c > C_k$ must hold. Assume that $C_i > A_c > C_j > C_k$. (The proof for the other scenario can be obtained in a similar manner.) As stated in Proposition 5.1, the optimal solution operates at the maximum average power and cost, which leads to the following equality:

$$\lambda_i \mathbf{z}_i + \lambda_j \mathbf{z}_j + (1 - \lambda_i - \lambda_j) \mathbf{z}_k = \mathbf{a} , \quad (5.38)$$

where $\mathbf{z}_i = (P_i, C_i)$ and $\mathbf{a} = (A_p, A_c)$, as in the proof of Proposition 5.4. Consider an alternative solution that switches between two channels, channel i and channel j , with channel switching factors γ and $(1-\gamma)$ and powers P_i and P_j , respectively, and utilizes the maximum average power and cost; that is,

$$\gamma \mathbf{z}_i + (1 - \gamma) \mathbf{z}_j = \mathbf{a} . \quad (5.39)$$

Based on (5.38) and (5.39), γ and λ_i can be obtained from the average cost constraint as $\gamma = (A_c - C_j)/(C_i - C_j)$ and $\lambda_i = (A_c - C_k - \lambda_j(C_j - C_k))/(C_i - C_k)$. First, it is shown that $\lambda_i > \gamma$. To that aim, the following inequality is obtained from the condition $\lambda_i > \gamma$ based on the definition of λ_i and γ : $(A_c - C_k - \lambda_j(C_j - C_k))(C_i - C_j) > (A_c - C_j)(C_i - C_k)$, which reduces, after some manipulation, to $\lambda_j C_j + (1 - \lambda_j) C_i > A_c$. Since $(1 - \lambda_j) = \lambda_i + \lambda_k$, $C_i > C_k$, and $\lambda_i C_i + \lambda_j C_j + \lambda_k C_k = A_c$, the inequality $\lambda_j C_j + (1 - \lambda_j) C_i > A_c$ always holds, which verifies that $\lambda_i > \gamma$. Then, from (5.38) and (5.39), $\mathbf{z}_j$ can be expressed as

$$\mathbf{z}_j = \frac{\lambda_i - \gamma}{1 - \gamma - \lambda_j} \mathbf{z}_i + \frac{1 - \lambda_i - \lambda_j}{1 - \gamma - \lambda_j} \mathbf{z}_k. \quad (5.40)$$

The remaining part of the proof depends on the values of powers P_i , P_j , and P_k .

Case 1: If all the power levels satisfy the convexity condition in Lemma 5.1, then the following inequality can be obtained:

$$\gamma g(\mathbf{z}_i) + (1 - \gamma) g(\mathbf{z}_j) = \quad (5.41)$$

$$= \gamma g(\mathbf{z}_i) + \lambda_j g(\mathbf{z}_j) + (1 - \gamma - \lambda_j) g(\mathbf{z}_j) \quad (5.42)$$

$$< \gamma g(\mathbf{z}_i) + \lambda_j g(\mathbf{z}_j) + (1 - \gamma - \lambda_j) \left(\frac{\lambda_i - \gamma}{1 - \gamma - \lambda_j} g(\mathbf{z}_i) + \frac{1 - \lambda_i - \lambda_j}{1 - \gamma - \lambda_j} g(\mathbf{z}_k) \right) \quad (5.43)$$

$$= \lambda_i g(\mathbf{z}_i) + \lambda_j g(\mathbf{z}_j) + (1 - \lambda_i - \lambda_j) g(\mathbf{z}_k) \quad (5.44)$$

where $g(\mathbf{z}_i) = g(P_i, C_i)$ denotes the average probability of error as a function of power and cost, as defined in (5.22). In obtaining the inequality in (5.43), the definition of $\mathbf{z}_j$ in (5.40) and the strict convexity of g are employed. Note that g is strictly convex when the power levels satisfy the condition in Lemma 5.1, which is the assumption in Case 1. In addition, it is noted that the $(1 - \gamma - \lambda_j)$ term in (5.41) is never negative since $\gamma < \lambda_i$ as proved in the previous paragraph and $\lambda_i < 1 - \lambda_j$ by definition. The inequality in (5.41)-(5.44) implies that the channel switching between channel i and channel j with channel switching factors γ and $(1 - \gamma)$, respectively, achieves a lower average probability of error than the optimal solution, which switches among channels i , j , and k with channel switching factors λ_i , λ_j , and λ_k , respectively. Hence, a contradiction arises. Therefore, the strategy that switches among three channels cannot be optimal. In other words,

for any strategy that switches among three channels, there exist a strategy that performs channel switching between two channels and achieves a smaller average probability of error.

Case 2: Suppose that some of the power levels do not satisfy the convexity condition in Lemma 5.1. Since the average power should be equal to A_p due to Proposition 5.1, at least one power level should be below A_p . Assume without loss of generality that $P_i < A_p$. Then, the average probability of error for the optimal solution that switches among three different channels can be bounded from below as follows:

$$\lambda_i g(\mathbf{z}_i) + \lambda_j g(\mathbf{z}_j) + \lambda_k g(\mathbf{z}_k) = \lambda_i g(\mathbf{z}_i) + (\lambda_j + \lambda_k) \left(\tilde{\lambda}_j g(\mathbf{z}_j) + \tilde{\lambda}_k g(\mathbf{z}_k) \right) \quad (5.45)$$

$$\geq \lambda_i g(\mathbf{z}_i) + (\lambda_j + \lambda_k) (\nu_j g(\tilde{\mathbf{z}}_j) + \nu_k g(\tilde{\mathbf{z}}_k)) \quad (5.46)$$

$$= \lambda_i g(\mathbf{z}_i) + \tilde{\nu}_j g(\tilde{\mathbf{z}}_j) + \tilde{\nu}_k g(\tilde{\mathbf{z}}_k) \quad (5.47)$$

with $\tilde{\lambda}_j \triangleq \lambda_j/(\lambda_j + \lambda_k)$, $\tilde{\lambda}_k \triangleq \lambda_k/(\lambda_j + \lambda_k)$, $\tilde{\nu}_j \triangleq (\lambda_j + \lambda_k)\nu_j$, $\tilde{\nu}_k \triangleq (\lambda_j + \lambda_k)\nu_k$, $\tilde{\mathbf{z}}_j \triangleq (\tilde{P}_j, C_j)$, and $\tilde{\mathbf{z}}_k \triangleq (\tilde{P}_k, C_k)$, where $\tilde{P}_j$ and $\tilde{P}_k$ are the optimal power levels and ν_j and ν_k are the corresponding optimal channel switching factors when the channel switching is performed between channel j and channel k *only* under the average cost limit $\tilde{A}_c \triangleq \tilde{\lambda}_j C_j + \tilde{\lambda}_k C_k$ and the average power limit $\tilde{A}_p \triangleq \tilde{\lambda}_j P_j + \tilde{\lambda}_k P_k > A_p$.⁵ Since $(\tilde{P}_j, \tilde{P}_k)$ and (ν_j, ν_k) are the solution of the optimal channel switching problem in the presence of channel j and channel k only under the average power limit $\tilde{A}_p$ and the average cost limit $\tilde{A}_c$, the average probability of error is bounded from below by the expression in (5.46). From Proposition 5.3, $\min\{\tilde{P}_j, \tilde{P}_k\} > \sigma_j^2 \tilde{A}_p / \sigma_k^2$. Then, based on a similar argument to that in (5.37), it can be shown that the convexity condition in Lemma 5.1 is satisfied for power levels $\tilde{P}_j$ and $\tilde{P}_k$ if $\tilde{A}_p > b \sigma_K^4 / (2 \sigma_1^2 \sigma_K^2 + b \sigma_1^2)$, which always holds due to the assumption in the proposition and the fact that $\tilde{A}_p > A_p$. Next assume without loss of generality that $\tilde{\nu}_j \tilde{P}_j \geq \tilde{\nu}_k \tilde{P}_k$. Then, the lower bound in (5.47) can be

⁵The inequality $\tilde{A}_p > A_p$ follows from the assumption that $P_i < A_p$.

improved as follows:

$$\lambda_i g(\mathbf{z}_i) + \tilde{\nu}_j g(\tilde{\mathbf{z}}_j) + \tilde{\nu}_k g(\tilde{\mathbf{z}}_k) = (\lambda_i + \tilde{\nu}_j) (\lambda_i^* g(\mathbf{z}_i) + \lambda_j^* g(\tilde{\mathbf{z}}_j)) + \tilde{\nu}_k g(\tilde{\mathbf{z}}_k) \quad (5.48)$$

$$\geq (\lambda_i + \tilde{\nu}_j) (\nu_i^* g(\mathbf{z}_i^*) + \nu_j^* g(\mathbf{z}_j^*)) + \tilde{\nu}_k g(\tilde{\mathbf{z}}_k) \quad (5.49)$$

$$= \hat{\nu}_i g(\mathbf{z}_i^*) + \hat{\nu}_j g(\mathbf{z}_j^*) + \tilde{\nu}_k g(\tilde{\mathbf{z}}_k) \quad (5.50)$$

with $\lambda_i^* \triangleq \lambda_i/(\lambda_i + \tilde{\nu}_j)$, $\lambda_j^* \triangleq \tilde{\nu}_j/(\lambda_i + \tilde{\nu}_j)$, $\hat{\nu}_i \triangleq (\lambda_i + \tilde{\nu}_j)\nu_i^*$, $\hat{\nu}_j \triangleq (\lambda_i + \tilde{\nu}_j)\nu_j^*$, $\mathbf{z}_i^* \triangleq (P_i^*, C_i)$, and $\mathbf{z}_j^* \triangleq (P_j^*, C_j)$, where P_i^* and P_j^* are the optimal power levels and ν_i^* and ν_j^* are the corresponding optimal channel switching factors when the channel switching is performed between channel i and channel j *only* under the average cost limit $A_c^* \triangleq \lambda_i^* C_i + \lambda_j^* C_j$ and the average power limit $A_p^* \triangleq \lambda_i^* P_i + \lambda_j^* \tilde{P}_j > 0.5 A_p$.⁶ From Proposition 5.3, $\min\{P_i^*, P_j^*\} > \sigma_i^2 A_p^*/\sigma_j^2$. Then, similar to (5.37), it can be shown that the convexity condition in Lemma 5.1 is satisfied for power levels P_i^* and P_j^* if $A_p^* > b\sigma_K^4/(2\sigma_1^2\sigma_K^2 + b\sigma_1^2)$, which is true due to the assumption in the proposition and the fact that $A_p^* > 0.5A_p$. From (5.45)-(5.50), it is concluded that when the assumption in the proposition holds, for any strategy that performs channel switching among three different channels, there exists another strategy that switches among the same channels with power levels that satisfy the convexity condition in Lemma 5.1, and achieves a smaller average probability of error. Therefore, the arguments in the previous part of the proof (Case 1) can be employed to show that there exists a strategy that performs channel switching between two channels and achieves a smaller average probability of error than the lower bound in (5.50). Therefore, channel switching among three channels cannot be optimal under the condition in the proposition. ■

Proposition 5.5 states that if the average power limit is larger than a certain value, then the optimal channel switching strategy is to use a single channel exclusively or to switch between two channels; that is, Strategy 3 is not optimal. In such a scenario, the optimal solution is either to transmit over a single channel with cost A_c if such a channel exists, or to switch between channel i and channel j as specified in Proposition 5.4 if there exists no channels with cost A_c .

⁶The inequality $A_p^* > 0.5 A_p$ follows from the assumptions that $\tilde{\nu}_j \tilde{P}_j \geq \tilde{\nu}_k \tilde{P}_k$ and $P_i < A_p$.

5.4 Numerical Examples

In this section, various numerical examples are presented in order to provide illustrations of the theoretical results and to investigate performance gains that can be achieved via channel switching. The following strategies are compared in the numerical examples:

Optimal Single Channel: In this strategy, channel switching is not allowed, and only one channel is employed exclusively. The optimal solution for this approach is obtained by using Strategy 1 in Section 5.2.

Optimal Channel Switching: In this strategy, channel switching is allowed, and the optimal solution of the channel switching problem in (5.2) is obtained based on Strategy 3. (Since Strategy 3 covers Strategy 2 as a special case, Strategy 2 is not considered separately.)

A scenario with K Gaussian channels is considered, and the standard deviations and the costs of the channels are represented, for notational simplicity, in the vector form as $\boldsymbol{\sigma} = [\sigma_1 \cdots \sigma_K]$ and $\mathbf{C} = [C_1 \cdots C_K]$, respectively. First, a four-channel system is studied, where $\boldsymbol{\sigma} = [0.4 \ 0.6 \ 0.8 \ 1]$, $\mathbf{C} = [7 \ 5 \ 3 \ 1]$, and the average cost limit is equal to 2; that is, $A_c = 2$. In Figure 5.2, the average probabilities of error are plotted versus the average power limit A_p for the optimal single channel and optimal channel switching approaches. It is observed that the optimal channel switching strategy outperforms the optimal single channel strategy for *all* values of A_p . This is an expected result since the optimal single channel approach cannot be the optimal solution of the channel switching problem in this scenario as there exists no channel with a cost of A_c and $A_c < C_1$ (Corollary 5.1). In order to provide further investigations of the results in Figure 5.2, the parameters of the optimal channel switching strategy are presented in Table 5.1 for some values of A_p . In the table, the optimal channel switching solution is represented by channel switching factors $(\lambda_i, \lambda_j, \lambda_k)$ and power levels (P_i, P_j, P_k) , where $i < j < k$. The channels that are not employed in the optimal solution are marked with “–” in the table. Since at most three channels

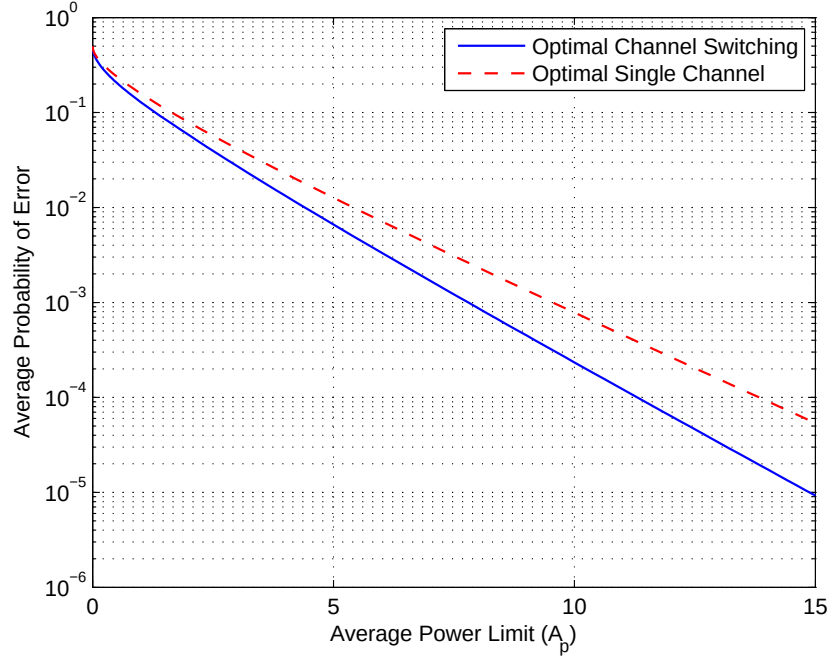


Figure 5.2: Average probability of error versus A_p for the optimal single channel and optimal channel switching strategies, where $K = 4$, $\sigma = [0.4 \ 0.6 \ 0.8 \ 1]$, $C = [7 \ 5 \ 3 \ 1]$, and $A_c = 2$.

can be utilized in the optimal solution according to Proposition 5.2, only two of the channel switching factors are shown in the table, and the remaining one can be calculated as $\lambda_k = 1 - \lambda_i - \lambda_j$. It should be noted that λ_i , λ_j , and λ_k correspond to the channel switching factors of the *employed channels* with the smallest index, the second smallest index, and the third smallest index, respectively. For example, for $A_p = 0.1$, channel 1 is employed with channel switching factor 0.1667 and power 0.1821 and channel 4 is employed with channel switching factor 0.8333 and power 0.0836. (In this case, $\lambda_k = 0$, meaning that only two channels are employed in the optimal solution). It is observed from Table 5.1 that the optimal channel switching strategy performs channel switching between two channels, which in compliance with Proposition 5.2. In addition, the calculations show that the optimal channel switching solution utilizes the maximum average power and maximum average cost as claimed in Proposition 5.1. In addition, the statements in Proposition 5.3 are verified, which can be exemplified as follows:

Table 5.1: Parameters of the optimal channel switching strategy in Figure 5.2.

A_p	λ_i	λ_j	P_1	P_2	P_3	P_4
0.1	0.1667	0.8333	0.1821	–	–	0.0836
0.2	0.1667	0.8333	0.2663	–	–	0.1867
2	0.25	0.75	–	1.3461	–	2.2180
4	0.5	0.5	–	–	3.4117	4.5883
6	0.5	0.5	–	–	4.9898	7.0102
8	0.5	0.5	–	–	6.5601	9.4399
10	0.5	0.5	–	–	8.1270	11.873

Table 5.2: Parameters of the optimal channel switching strategy in Figure 5.3.

A_p	λ_i	λ_j	P_1	P_2	P_3	P_4
0.1	0.6667	0.3333	0.1281	–	–	0.0437
0.2	0.6667	0.3333	0.2314	–	–	0.1371
1	0.6667	0.3333	0.6894	–	–	1.6212
2	1	–	–	2	–	–
3	1	–	–	3	–	–
4	1	–	–	4	–	–
5	1	–	–	5	–	–

Parameter A_{ij} in Proposition 5.3 can be calculated for channel 3 and channel 4 as $A_{34} = 0.7934$. As observed from Table 5.1, when channel 3 and channel 4 are employed, $A_p > A_{34}$ is satisfied and the conditions in Part (ii) of Proposition 5.3 hold; that is, $P_4 > A_p > P_3 > A_{34}$. In addition, the ratio of the optimal power levels is always smaller than the ratio of the noise variances, $1/(0.8)^2 = 1.5625$, as stated in (5.13) in Proposition 5.3. Compared to the optimal channel switching strategy, which performs channel switching between channel 4 and another channel, the optimal single channel solution always utilizes channel 4 at the maximum power limit A_p since it is the only channel with a cost that is lower than the average cost limit A_c . However, as observed from Figure 5.2 and Table 5.1, performing time sharing between channel 4 and a channel with a higher cost (lower error probability) reduces the average probability of error in this scenario.

Next, the same channel configuration is considered with a different average cost limit, which is given by $A_c = 5$, and the average probability of error curves

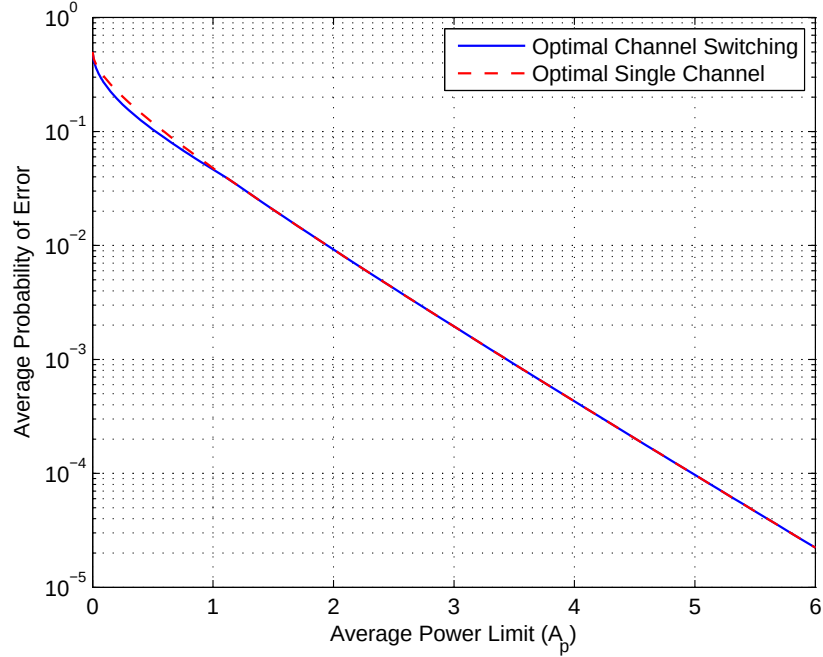


Figure 5.3: Average probability of error versus A_p for the optimal single channel and optimal channel switching strategies, where $K = 4$, $\sigma = [0.4 \ 0.6 \ 0.8 \ 1]$, $C = [7 \ 5 \ 3 \ 1]$, and $A_c = 5$.

are presented in Figure 5.3. In this case, since there is a channel with a cost that is equal to A_c , Corollary 1 does not apply; i.e., channel switching is not necessarily optimal. As observed from the figure, for small values of the average power limit A_p , the optimal channel switching strategy outperforms the optimal single channel strategy, whereas both strategies achieve the same performance as A_p increases. Table 5.2 presents the parameters of the optimal channel switching solution, which indicates that employing channel 2 exclusively at the power limit (which is the optimal single channel solution) becomes optimal when A_p is larger than a certain value whereas switching between channel 1 and channel 4 is optimal for small values of A_p . Hence, it is concluded that when there exists a channel with a cost equal to A_c , employing a single channel exclusively may or may not be the optimal solution depending on the system parameters.

In order to investigate the effects of the average cost limit in more detail,

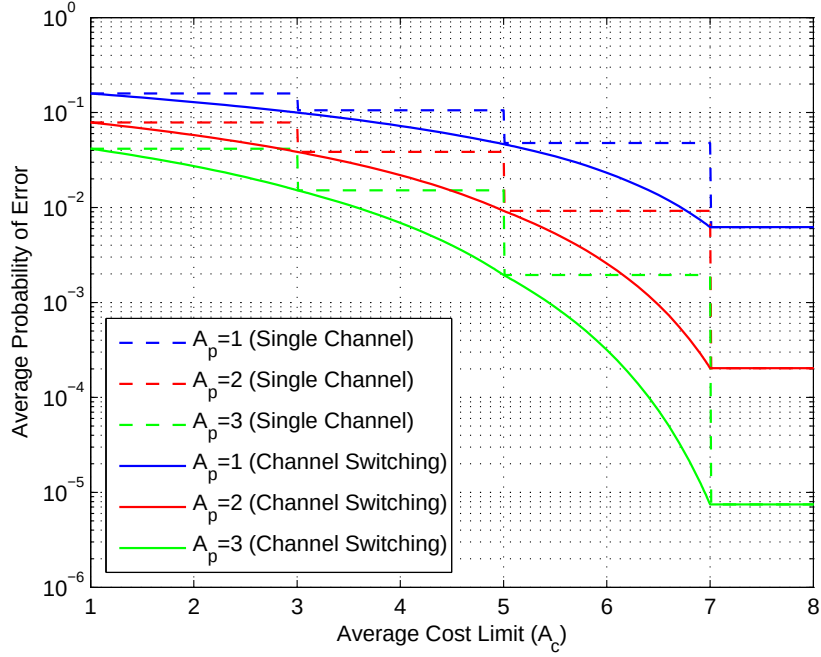


Figure 5.4: Average probability of error versus A_c for the optimal single channel and optimal channel switching strategies, where $K = 4$, $\sigma = [0.4 \ 0.6 \ 0.8 \ 1]$, and $C = [7 \ 5 \ 3 \ 1]$.

the average probabilities of error are plotted versus A_c in Figure 5.4 for various values of A_p based on the same channel configuration as in the previous scenario. As expected, the average probability of error is a non-increasing function of the average cost limit A_c . Also, in accordance with Part (ii) of Remark 5.1, the average probability of error converges to the error probability of the best channel (channel 1) at the average power limit A_p when A_c is larger than or equal to the cost of the best channel; i.e., when $A_c \geq 7$. In addition, it is observed that the optimal single channel strategy results in piecewise constant average probabilities of error, which is due to the fact that the optimal single channel solution corresponds the use of the best channel that has a cost lower than or equal to A_c . Specifically, the optimal single channel strategy achieves the error probabilities of $g_4(A_p)$, $g_3(A_p)$, $g_2(A_p)$, and $g_1(A_p)$ for $A_c \in [1, 3)$, $A_c \in [3, 5)$, $A_c \in [5, 7)$, and $A_c \geq 7$, respectively, where $g_i(A_p) = Q(\sqrt{A_p}/\sigma_i)$ denotes the error probability of channel i at power level A_p . Furthermore, Figure 5.4 verifies

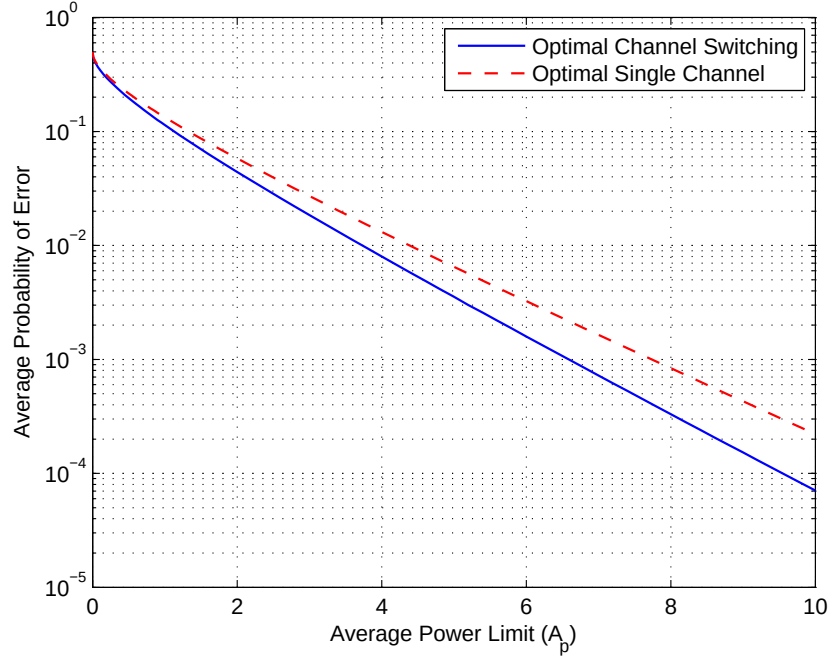


Figure 5.5: Average probability of error versus A_p for the optimal single channel and optimal channel switching strategies, where $K = 5$, $\sigma = [0.6 \ 0.7 \ 0.8 \ 0.9 \ 1]$, $C = [1.329 \ 1.112 \ 0.941 \ 0.804 \ 0.6931]$, and $A_c = 0.9$.

the argument in Corollary 5.1 that, for $C_K < A_c < C_1$, channel switching is guaranteed to outperform the optimal single channel strategy if A_c is not equal to the cost of one of the channels.

As another scenario, a five-channel system is considered, and the cost values are calculated based on the logarithmic cost function in (5.21) with $b = 1$. The standard deviations of the channels are set to $\sigma = [0.6 \ 0.7 \ 0.8 \ 0.9 \ 1]$, and the average cost limit is given by $A_c = 0.9$. In Figure 5.5, the average probability of error is plotted versus A_p for the optimal single channel and optimal channel switching strategies. Similar to the scenario in Figure 5.2, it is observed that channel switching outperforms the single channel approach for all values of A_p as a consequence of Corollary 5.1 as there exists no channel with a cost equal to A_c . The parameters of the optimal channel switching strategy in Figure 5.5 are presented in Table 5.3 for some values of A_p . It is noted that the optimal

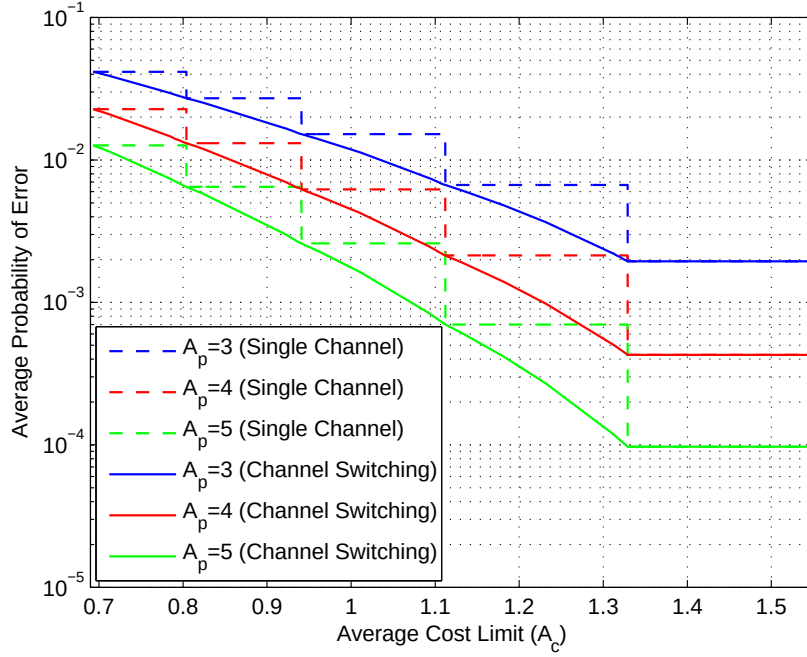


Figure 5.6: Average probability of error versus A_c for the optimal single channel and optimal channel switching strategies, where $K = 5$, $\sigma = [0.6 \ 0.7 \ 0.8 \ 0.9 \ 1]$, and $\mathbf{C} = [1.329 \ 1.112 \ 0.941 \ 0.804 \ 0.6931]$.

solution performs channel switching among at most three channels in compliance with Proposition 5.2. Also, numerical calculations show that the results in Proposition 5.1 and Proposition 5.3 are satisfied. In addition, as stated in Proposition 5.4, when $A_p \geq b\sigma_5^4/(\sigma_1^2(2\sigma_5^2 + b)) = 0.926$, the optimal channel switching between two channels is performed between channel 3 and channel 4, which is in accordance with (5.30) and (5.31). Furthermore, channel switching among three channels is not optimal for $A_p \geq 2b\sigma_5^4/(\sigma_1^2(2\sigma_5^2 + b)) = 1.852$ as claimed in Proposition 5.5.

Finally, the average probabilities of error are plotted versus A_c in Figure 5.6 for various values of A_p based on the scenario in Figure 5.5. Similar observations to those related to Figure 5.4 can be made. Namely, if A_c is smaller than the cost of the best channel, which is equal to 1.329, the optimal channel switching strategy outperforms the single channel one when A_c is not equal to the cost of

Table 5.3: Parameters of the optimal channel switching strategy in Figure 5.5.

A_p	λ_i	λ_j	P_1	P_2	P_3	P_4	P_5
0.05	0.3028	0.0347	0.0811	0.0652	—	—	0.0352
0.2	0.3254	0.6746	0.2652	—	—	—	0.1685
1	0.7007	0.2993	—	—	0.9881	1.0278	—
2	0.7007	0.2993	—	—	1.9310	2.1617	—
3	0.7007	0.2993	—	—	2.8648	3.3164	—
4	0.7007	0.2993	—	—	3.7955	4.4788	—
5	0.7007	0.2993	—	—	4.7247	5.6447	—

a channel. On the other hand, for $A_c \geq 1.329$, both strategies achieve an average probability of error that is equal to the error probability of the best channel at the power limit.

5.5 Extensions

Although the results in this chapter are obtained for binary communications, most of these results can be extended to M -ary communications as well under certain conditions. In particular, the results in Remark 5.1 and Proposition 5.1 are valid as long as the error probability $g_k(P)$ is a monotone decreasing and convex function of signal power P for $k \in \{1, \dots, K\}$, and $g_i(P) < g_j(P) \forall P$ whenever $C_i > C_j$ (or, whenever $\sigma_i^2 < \sigma_j^2$). As studied in [25], for maximum likelihood (ML) detection over additive Gaussian channels, the error probability is a convex function of the signal power for all 1-dimensional and 2-dimensional constellations (such as BPSK, PAM, QPSK, and QAM, which are commonly employed in practice), and it is convex also for higher dimensional constellations at high SNRs. In addition, for Gaussian channels, the error probability is a monotone decreasing function of the signal power and a monotone increasing function of the noise power. Hence, the results in Remark 1 and Proposition 1 are valid for M -ary communication systems for 1-dimensional and 2-dimensional constellations at all SNRs and for higher order constellations at high SNRs. Regarding Proposition 5.2, since there is no specific assumption about the modulation order

in the proof, it also covers scenarios in which any type of M -ary modulation is employed. Therefore, the optimal channel switching strategy is to switch among at most three different channels for M -ary communication systems as well.

Lemma 5.1 can also be extended to M -ary communication systems under certain conditions. As discussed in [50], the error probability for many coherent modulation schemes can be represented either exactly or approximately in the form of $\eta Q(\kappa\sqrt{P}/\sigma)$, where P is the average symbol energy, and η and κ are some positive constants that depend on the modulation type and order. For this error probability expression, it can be shown that the convexity condition in Lemma 5.1 becomes $P > b/(\kappa^2(e^C + 1))$. In addition, the results in Proposition 5.4 and Proposition 5.5 can be extended in a similar manner; namely, the thresholds for A_p in the propositions are scaled by $1/\kappa^2$, as can be verified based on (5.37).

5.6 Conclusions for Chapter 5

In this study, optimal channel switching has been investigated for Gaussian channels in the presence of average power and average cost constraints. For generic cost functions, it has been shown that the optimal channel switching strategy performs time sharing among at most three channels and operates at the average power and average cost limits. Also, for channel switching between two channels, it has been proved that the ratio of the optimal power levels is upper bounded by the ratio of the larger noise variance to the smaller one. In addition, for logarithmic cost functions, the convexity properties of the error probability have been characterized as a function of power and cost, and the optimal channel switching strategy has been shown to employ at most two channels, which can be determined based on specific formulas, when the average power limit is larger than a certain threshold. Finally, numerical examples have provided illustrations of the theoretical results.

Chapter 6

Conclusion

Optimal randomization (time sharing) strategies for parametric estimation problems and communication systems have been studied in this dissertation. In Chapter 2, an estimation problem with a single estimator has been considered. The estimator has been assumed to be fixed. However, it is possible to send any (stochastic) function of the parameter to be estimated to the estimator. The optimal design of the parameter that minimizes the Bayes risk of the estimator has been derived and it has been shown that the optimal parameter design is to time share (randomize) between at most two different signal levels; that is, the optimal parameter design can be deterministic or stochastic. In order to specify when the stochastic or the deterministic approach is optimal, the optimality conditions have been derived.

The discussion in Chapter 2 has been extended to cover multiple parameters and multiple estimators in Chapter 3. Interference among parameter dependent signals has also been considered in this chapter. The optimal parameter design has been obtained for two different cost functions. The first cost function is the total risk of all the estimators. For this cost function it has been shown that the optimal parameter design is to randomize between at most two different signal levels. The second cost function is the maximum risk of all the estimators. For this cost function it has been shown that, in the presence of K parameters, the

optimal parameter design is to randomize among at most $(K + 1)$ different signal levels. The optimality conditions and characterization of the optimal solution have also been provided.

In Chapter 4, the jointly optimal design of parameter dependent signals and the estimator has been studied considering the class of LMMSE estimators. It has been shown that randomization is not needed in this scenario; that is, the deterministic parameter design is the optimal solution for LMMSE estimators. Furthermore, it has been observed that the optimal solution does not depend on the channel characteristics, and the average power employed by the optimal design is equal to the power limit.

Finally, the optimal channel switching strategies for Gaussian channels have been studied under average power and average cost constraints in Chapter 5. It has been shown that the optimal channel switching strategy is to switch among at most three channels for generic cost functions. In addition it has been observed that the optimal strategy operates at the power and cost limits. Furthermore logarithmic cost functions have been considered and it has been shown that the optimal channel switching strategy is to switch among at most two channels for logarithmic cost functions.

For the parameter design problem, the future work involves the application of the theoretical results obtained in this dissertation to practical and widely used estimators. On the other hand, the future work for the channel switching problem is related to the incorporation of switching costs (delays) [34] in the design of optimal channel switching strategies.

Bibliography

- [1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, Inc., 1993.
- [2] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.
- [3] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, “Theory of the stochastic resonance effect in signal detection: Part I—Fixed detectors,” *IEEE Trans. Sig. Processing*, vol. 55, pp. 3172–3184, July 2007.
- [4] A. Patel and B. Kosko, “Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection,” *IEEE Trans. Sig. Processing*, vol. 57, pp. 1655–1669, May 2009.
- [5] S. M. Kay, “Noise enhanced detection as a special case of randomization,” *IEEE Sig. Processing Lett.*, vol. 15, pp. 709–712, 2008.
- [6] S. M. Kay, J. H. Michels, H. Chen, and P. K. Varshney, “Reducing probability of decision error using stochastic resonance,” *IEEE Sig. Processing Lett.*, vol. 13, pp. 695–698, Nov. 2006.
- [7] S. Bayram and S. Gezici, “Noise-enhanced M -ary hypothesis-testing in the minimax framework,” in *Proc. International Conference on Signal Processing and Commun. Systems*, (Omaha, Nebraska), pp. 31–36, Sep. 2009.
- [8] S. Bayram, S. Gezici, and H. V. Poor, “Noise enhanced hypothesis-testing in the restricted Bayesian framework,” *IEEE Trans. Sig. Processing*, vol. 58, pp. 3972–3989, Aug. 2010.

- [9] S. Bayram and S. Gezici, “Noise enhanced M-ary composite hypothesis-testing in the presence of partial prior information,” *IEEE Trans. Sig. Processing*, vol. 59, pp. 1292–1297, Mar. 2011.
- [10] C. Goken, S. Gezici, and O. Arikan, “Optimal stochastic signaling for power-constrained binary communications systems,” *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3650–3661, Dec. 2010.
- [11] C. Goken, S. Gezici, and O. Arikan, “Optimal signaling and detector design for power-constrained binary communications systems over non-Gaussian channels,” *IEEE Commun. Letters*, vol. 14, pp. 100–102, Feb. 2010.
- [12] M. Tutay, S. Gezici, and O. Arikan, “Optimal randomization of signal constellations on the downlink of a multiuser DS-CDMA system,” *IEEE Trans. Wireless Commun.*, vol. 12, pp. 4878–4891, Oct. 2013.
- [13] B. Dulek and S. Gezici, “Detector randomization and stochastic signaling for minimum probability of error receivers,” *IEEE Trans. Commun.*, vol. 60, pp. 923–928, Apr. 2012.
- [14] M. E. Tutay, S. Gezici, and O. Arikan, “Optimal detector randomization for multiuser communications systems,” *IEEE Trans. Commun.*, vol. 61, pp. 2876–2889, July 2013.
- [15] B. Dulek, N. D. Vanli, S. Gezici, and P. K. Varshney, “Optimum power randomization for the minimization of outage probability,” *IEEE Trans. Wireless Commun.*, vol. 12, pp. 4627–4637, Sep. 2013.
- [16] H. Chen, P. K. Varshney, and J. H. Michels, “Noise enhanced parameter estimation,” *IEEE Trans. Sig. Processing*, vol. 56, pp. 5074–5081, Oct. 2008.
- [17] M. Azizoglu, “Convexity properties in binary detection problems,” *IEEE Trans. Inform. Theory*, vol. 42, pp. 1316–1321, July 1996.
- [18] B. Dulek, S. Gezici, and O. Arikan, “Convexity properties of detection probability under additive Gaussian noise: Optimal signaling and jamming strategies,” *IEEE Trans. Signal Process.*, vol. 61, pp. 3303–3310, July 2013.

- [19] S. Bayram, N. D. Vanli, B. Dulek, I. Sezer, and S. Gezici, "Optimum power allocation for average power constrained jammers in the presence of non-Gaussian noise," *IEEE Commun. Letters*, vol. 16, pp. 1153–1156, Aug. 2012.
- [20] H. Soganci, S. Gezici, and O. Arikan, "Optimal stochastic parameter design for estimation problems," *IEEE Trans. Sig. Processing*, vol. 60, no. 9, pp. 4950–4956, 2012.
- [21] H. Soganci, S. Gezici, and O. Arikan, "Optimal stochastic design for multi-parameter estimation problems," in *2014 IEEE Int. Conf. Acous., Speech, Signal Process. (ICASSP)*, pp. 5651–5655, 2014.
- [22] H. Soganci, S. Gezici, and O. Arikan, "Optimal stochastic design of multiple parameters for estimation problems," submitted, 2014.
- [23] A. Ozcelikkale, H. M. Ozaktas, and E. Arikan, "Signal recovery with cost-constrained measurements," *IEEE Trans. on Signal Process.*, vol. 58, no. 7, pp. 3607–3617, 2010.
- [24] K. Sohraby, D. Minoli, and T. Znati, *Wireless Sensor Networks: Technology, Protocols, and Applications*. Hoboken, New Jersey: John Wiley & Sons, 2007.
- [25] S. Loyka, V. Kostina, and F. Gagnon, "Error rates of the maximum-likelihood detector for arbitrary constellations: Convex/concave behavior and applications," *IEEE Trans. Inform. Theory*, vol. 56, pp. 1948–1960, April 2010.
- [26] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part II–Variable detectors," *IEEE Trans. Sig. Processing*, vol. 56, pp. 5031–5041, Oct. 2007.
- [27] J. A. Ritcey and M. Azizoglu, "Performance analysis of generalized selection combining with switching constraints," *IEEE Commun. Letters*, vol. 4, pp. 152–154, May 2000.
- [28] B. Dulek, M. E. Tutay, S. Gezici, and P. K. Varshney, "Optimal signaling and detector design for M-ary communication systems in the presence of multiple

- additive noise channels,” *Digital Signal Processing*, vol. 26, pp. 153–168, Mar. 2014.
- [29] B. Dulek and S. Gezici, “Optimal stochastic signal design and detector randomization in the Neyman-Pearson framework,” in *37th IEEE Int. Conf. Acous., Speech, Signal Process. (ICASSP)*, (Kyoto, Japan), pp. 3025–3028, Mar. 25–30 2012.
 - [30] E. L. Lehmann, *Testing Statistical Hypotheses*. New York: Chapman & Hall, 2 ed., 1986.
 - [31] Y. Ma and C. C. Chai, “Unified error probability analysis for generalized selection combining in Nakagami fading channels,” *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2198–2210, Nov. 2000.
 - [32] F. Gaaloul, H.-C. Yang, R. Radaydeh, and M.-S. Alouini, “Switch based opportunistic spectrum access for general primary user traffic model,” *IEEE Wireless Commun. Letters*, vol. 1, pp. 424–427, Oct. 2012.
 - [33] Y. Liu and M. Liu, “To stay or to switch: Multiuser dynamic channel access,” in *Proc. IEEE INFOCOM*, pp. 1249–1257, Apr. 2013.
 - [34] L. Chen, S. Iellamo, and M. Coupechoux, “Opportunistic spectrum access with channel switching cost for cognitive radio networks,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, (Kyoto, Japan), June 2011.
 - [35] D. Niyato and E. Hossain, “Spectrum trading in cognitive radio networks: A market-equilibrium-based approach,” *IEEE Wireless Commun.*, vol. 15, pp. 71–80, Dec. 2008.
 - [36] J. Huang, R. Berry, and M. L. Honig, “Auction-based spectrum sharing,” *ACM/Springer Mobile Networks and Apps.*, pp. 405–418, 2006.
 - [37] D. Niyato and E. Hossain, “Competitive pricing for spectrum sharing in cognitive radio networks: Dynamic game, inefficiency of Nash equilibrium, and collusion,” *IEEE Journal on Selected Areas in Commun.*, vol. 26, pp. 192–202, Jan. 2008.

- [38] D. Niyato and E. Hossain, “Competitive spectrum sharing in cognitive radio networks: a dynamic game approach,” *IEEE Trans. on Wireless Commun.*, vol. 7, pp. 2651–2660, July 2008.
- [39] Z. Ji and K. Liu, “Dynamic spectrum sharing: A game theoretical overview,” *IEEE Commun. Magazine*, vol. 45, pp. 88–94, May 2007.
- [40] M. E. Tutay, S. Gezici, H. Soganci, and O. Arikan, “Optimal channel switching over Gaussian channels under average power and cost constraints,” *IEEE Trans. Commun.*, submitted, 2014.
- [41] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Boston, MA: Athena Specific, 2003.
- [42] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization method for constrained optimization problems*, pp. 214–220. IOS Press, 2002. in *Intelligent Technologies—Theory and Applications: New Trends in Intelligent Technologies*.
- [43] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. New York: Springer, 2005.
- [44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [45] S. Bayram and S. Gezici, “On the improvability and nonimprovability of detection via additional independent noise,” *IEEE Sig. Processing Lett.*, vol. 16, pp. 1001–1004, Nov. 2009.
- [46] S. Bayram and S. Gezici, “Stochastic resonance in binary composite hypothesis-testing problems in the Neyman-Pearson framework,” *Digital Signal Processing*, vol. 22, pp. 391–406, May 2012.
- [47] G. Strang, *Introduction to Linear Algebra*. Wellesley Cambridge Press, 4 ed., 2009.
- [48] K. S. Miller, “On the inverse of the sum of matrices,” *Mathematics Magazine*, pp. 67–72, 1981.

- [49] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1968.
- [50] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.